

Proceedings of the  
**1st International Workshop on  
eLearning Approaches for the  
Linked Data Age  
(Linked Learning 2011)**

collocated with the  
**8th Extended Semantic Web Conference  
(ESWC2011)**

29 May 2011, Heraklion, Greece



## Preface

While sharing of educational resources on the Web became common practice throughout the last years, a large amount of research was dedicated to interoperable eLearning repositories based on semantic technologies. Data interoperability is even more crucial, since sharing of online resources at Web-scale is widely facilitated by established APIs, such as OAI-PMH or SSI. Moreover, adoption of social computing aspects within personal learning environments has become a dominant paradigm building on principles such as user-centred identity management, service-orientation, and social participation. However, it remains an unresolved challenge to provide a meaningful, automated and personalized integration of diverse learning resources, e.g., formal and informal ones as found on the Web.

Though the eLearning area has brought up a number of comprehensive metadata standards (e.g., ADL SCORM, IEEE LOM, IMS LD) aiming at interoperability across eLearning environments, actual take-up is still fragmented. This can be attributed to their merely XML-driven approaches, the lack of established controlled vocabularies and the incompatibility of individual schemas. Several research efforts tried to address these issues by using Semantic Web technologies and ontology-based approaches. However, these efforts often failed to attract a critical mass of adopters. This is due to reasons such as inherent complexity, the lack of scalable and high-performance tool support when following complex reasoning-based approaches and the inavailability of vocabularies when following proprietary representation schemes. In the meantime, the Semantic Web has redefined itself throughout the last years as a Web of “Linked Data” by establishing principles which support sharing of large datasets on the Web together with a technology stack (use of URIs, RDF, and SPARQL) aimed at their realisation. The huge success and widespread adoption of the Linked Data approach has led to the availability of vast amounts of public data such as DBPedia, WordNet RDF or the data.gov.uk initiative.

Although the Linked Data approach is not yet adopted widely within the eLearning domain, this workshop has emerged on the fundamental belief that the Linked Data approach has the potential to fulfill the eLearning vision of Web-scale interoperability of eLearning resources and highly personalised and adaptive eLearning applications. The workshop has been established to become a highly interactive research forum for exploring the promises of the Web of Linked Data in technology-enhanced learning by gathering researchers from the areas of the Semantic Web and technology-enhanced learning.

The response to the call for papers was overwhelming for the first edition of this workshop. Therefore, after extensive peer review (each submission was reviewed by at least two independent reviewers) we were able to select 13 papers for presentation in the program of the workshop. In addition, the workshop program also had an excellent keynote speaker – Vania Dimitrova of the University of Leeds – a well-known researcher in the areas of technology enhanced learning and Semantic Web.

The workshop would not be possible without contributions of many persons and institutions. We are very thankful to the organisers of the ESWC 2011 conference for providing us with an opportunity to organize the workshop, for their excellent collaboration, and for looking after many important logistic issues. We are also very grateful to the members of the program committee for their commitment in reviewing the papers and assuring the good quality of the workshop program. We also thank the authors for their invaluable contributions to the workshop by writing, revising and presenting their papers. Of course, great appreciation of her time and expertise goes to our keynote speaker Vania Dimitrova. We also want to express our strong gratitude to the publishers of CEUR for publishing the Linked Learning 2011 workshop proceedings, to the European Commission (EC) and the EC-funded research project mEducator for sponsoring the best paper award and to the EasyChair developers for supporting the submission and review process.

May 2011,

Stefan Dietze, Mathieu d'Aquin, Dragan Gasevic, Miguel-Angel Sicilia

## **Organisers**

- Stefan Dietze, The Open University, UK
- Mathieu d'Aquin, The Open University, UK
- Dragan Gasevic, Athabasca University, Canada,
- Miguel-Angel Sicilia, University of Alcalá, Spain

## **Program Committee**

- Lora Aroyo, Free University of Amsterdam, The Netherlands
- Soeren Auer, University of Leipzig, Germany
- Panagiotis Bamidis, Aristotle University of Thessaloniki, Greece
- Charalampos Bratsas, Aristotle University of Thessaloniki, Greece
- Dan Brickley, W3C & Free University of Amsterdam, The Netherlands
- Vania Dimitrova, University of Leeds, UK
- John Domingue, The Open University, UK & Semantic Technologies Institute International, Austria.
- Nikolas Dovrolis, Democritus University of Thrace, Greece
- Marek Hatala, Simon Fraser University, Canada
- Jelena Jovanovic, University of Belgrade, Serbia
- Eleni Kaldoudi, Democritus University of Thrace, Greece
- Tomi Kauppinen, University of Münster, Germany
- Carsten Keßler, University of Münster, Germany
- Effie Lai-Chong Law, Leicester University, UK & ETH, Zurich, Switzerland
- Nikos Manouselis, Greek Research and Technology Network, Greece
- Dave Millard, University of Southampton, UK
- Evangelia Mitsopoulou, St George's University London, UK
- Wolfgang Nejdl, L3S Research Center, Germany
- Mikael Nilsson, Royal Institute of Technology, Sweden
- Carlos Pedrinaci, The Open University, UK
- Davide Taibi, Institute for Educational Technologies, Italian National Research Council, Italy.
- Vlad Tanasescu, University of Edinburgh, UK
- Fridolin Wild, The Open University, UK
- Martin Wolpers, Fraunhofer FIT.ICON, Germany
- Hong Qing Yu, The Open University, UK

## **Reviewers**

- Dhaval Thakker, University of Leeds, UK

# The OU Linked Open Data: Production and Consumption

Fouad Zablith, Miriam Fernandez and Matthew Rowe

Knowledge Media Institute (KMi), The Open University  
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom  
{f.zablith, m.fernandez, m.c.rowe}@open.ac.uk

**Abstract.** The aim of this paper is to introduce the current efforts toward the release and exploitation of The Open University's (OU) Linked Open Data (LOD). We introduce the work that has been done within the LUCERO project in order to select, extract and structure subsets of information contained within the OU data sources and migrate and expose this information as part of the LOD cloud. To show the potential of such exposure we also introduce three different prototypes that exploit this new educational resource: (1) the OU expert search system, a tool focused on finding the best experts for a certain topic within the OU staff; (2) the Buddy Study system, a tool that relies on Facebook information to identify common interest among friends and recommend potential courses within the OU that 'buddies' can study together, and; (3) Linked OpenLearn, an application that enables exploring linked courses, Podcasts and tags to OpenLearn units. Its aim is to enhance the browsing experience for students, by detecting relevant educational resources on the fly while reading an OpenLearn unit.

**Keywords:** Linked Open Data, education, expert search, social networks.

## 1 Introduction

The explosion of the Linked Open Data (LOD) movement in the last few years has produced a large number of interconnected datasets containing information about a large variety of topics, including geography, music and research publications among others. [2]

The movement is receiving worldwide support from public and private sectors like the UK<sup>1</sup> and US<sup>2</sup> governments, international media outlets, such as the BBC [5] or the New York Times [1], and companies with a social base like Facebook.<sup>3</sup> Such organisations are supporting the movement either by releasing

---

<sup>1</sup> <http://data.gov.uk>

<sup>2</sup> <http://www.data.gov/semantic/index>

<sup>3</sup> <http://developers.facebook.com/docs/opengraph>

large datasets of information or by generating applications that exploit it to connect data across different locations.

Despite its relevance and the support received in the last few years, very few pieces of work have either released or exploited LOD in the context of education. One of these few examples is the DBLP Bibliography Server Berlin,<sup>4</sup> which provides bibliographic information about scientific papers. However, education is principally one of the main sectors where the application of the LOD technologies can provoke a higher impact.

When performing learning and investigation tasks, students and academics have to go through the tedious and laborious task of browsing different information resources, analysing them, extracting their key concepts and mentally linking data across resources to generate their own conceptual schema about the topic. Educational resources are generally duplicated and dispersed among different systems and databases, and the key concepts within these resources as well as their inter and intra connections are not explicitly shown to users. We believe that the application of LOD technologies within and across educational institutions can explicitly generate the necessary structure and connections among educational resources, providing better support to users in their learning and investigation tasks.

In this context, the paper presents the work that has been done within The Open University (OU) towards the release and exploitation of several educational and institutional resources as part of the LOD cloud. First, we introduce the work that has been done within the LUCERO project to select, extract and structure subsets of OU information as LOD. Second, we present the potential of this data exposure and interlinking by presenting three different prototypes: (1) the OU expert search system, a tool focused on finding the best experts for a certain topic within the OU staff; (2) the Buddy Study system, a tool focused on exploiting Facebook information to identify common interests among friends and recommend potential courses within the OU that ‘buddies’ can study together, and; (3) Linked Open Learn, an application that enables exploring linked courses, Podcasts and tags to OpenLearn units.

The rest of the paper is organised as follows: Section 2 presents the state of the art in the areas of LOD within the education context. Section 3 presents the work that has been done within the LUCERO project to expose OU data as part of the LOD cloud. Sections 4, 5 and 6 present example prototype applications that consume the OU’s LOD for Expert Search, Buddy Study and Linked OpenLearn respectively. Section 7 describes the conclusions that we have drawn from this work, and section 8 presents our plans for future work.

## 2 Related Work

While LOD is being embraced in various sectors as mentioned in the previous section, we are currently witnessing a substantial increase in universities adopting

---

<sup>4</sup> <http://www4.wiwiw.fu-berlin.de/dblp/>

the Linked Data initiative. For example, the University of Sheffield's Department of Computer Science<sup>5</sup> provides a Linked Data service describing research groups, staff and publications, all semantically linked together[6]. Similarly the University of Southampton has recently announced the release of their LOD portal (<http://data.southampton.ac.uk>), where more data will become available in the near future. Furthermore, the University of Manchester's library catalogue records can now be accessed in RDF format<sup>6</sup>. In addition, other universities are currently working on transforming and linking their data: University of Bristol,<sup>7</sup> Edinburgh (e.g., the university's buildings information is now generated in LOD<sup>8</sup>), and Oxford<sup>9</sup>. Furthermore the University of Muenster announced a funded project, LODUM, the aim of which is to release the university's research information as Linked Data. This includes information related to people, projects, publications, prizes and patents.<sup>10</sup>

With the increase of the adoption of LOD publishing standards, the exchange of data will be much easier, not only within one university, but also across the LOD ready ones. This enables, for example, the comparison of specific qualifications offered by different universities in terms of courses required, pricing and availability.

### 3 The Open University Linked Open Data

The Open University is the first UK University to expose and publish its organizational information in LOD.<sup>11</sup> This is accomplished as part of the LUCERO project (Linking University Content for Education and Research Online)<sup>12</sup>, where the data extraction, transformation and maintenance are performed. This enables having multiple hybrid datasets accessible in an open way through the online access point: <http://data.open.ac.uk>.

The main purpose of releasing all this data as part of the LOD cloud is that members of the public, students, researchers and organisations will be able to easily search, extract and, more importantly, reuse the OU's information and data.

#### 3.1 Creating the OU LOD

Detailed information about the process of LOD generation within the OU is available at the LUCERO project website.<sup>12</sup> We briefly discuss in this section

---

<sup>5</sup> <http://data.dcs.shef.ac.uk>

<sup>6</sup> <http://prism.talis.com/manchester-ac>

<sup>7</sup> <https://mmb.ilt.bris.ac.uk/display/ldw2011/University+of+Bristol+data>

<sup>8</sup> <http://ldfocus.blogs.edina.ac.uk/2011/03/03/university-buildings-as-linked-data-with-scrapewiki>

<sup>9</sup> <http://data.ox.ac.uk>

<sup>10</sup> <http://www.lodum.de>

<sup>11</sup> <http://www3.open.ac.uk/media/fullstory.aspx?id=20073>

<sup>12</sup> <http://lucero-project.info>

the steps involved in the creation of Linked Data. To achieve that, the main requirement is to have a set of tools that generate RDF data from existing data sources, load such RDF into a triple store, and make it accessible through a web access point.

Given the fact that the OU's data repositories are scattered across many departments, using different platforms, and subject to constant update, a well-defined overflow needs to be put in place. The initial workflow is depicted in Figure 1, and is designed to be efficient in terms of time, flexibility and reusability. The workflow is component based, and the datasets characteristics played a major role in the implementation and setup of the components. For example, when the data sources are available in XML format, the XML updater will handle the process of identifying new XML entities and pass them to the RDF extractor, where the RDF data is generated, and ready to be added to (or removed from) the triple store. Finally the data is exposed to the web, and can be queried through a SPARQL endpoint.<sup>13</sup>

The scheduler component takes care of initiating the extraction/update process at specific time intervals. This update process is responsible for checking what was added, modified, or removed from the dataset, and accordingly applies to the triple store the appropriate action. Having such a process in place is important in the OU scenario where the data sources are continuously changing. Another point worth mentioning is the linking process that links entities coming from different OU datasets (e.g., courses mentioned in Podcast data and library records), in addition to linking external entities (e.g., course offerings in a GeoNames defined location<sup>14</sup>). To achieve interlinking OU entities, independently from which dataset the extraction is done, we rely on an Entity Named System, which generates a unique URI (e.g., based on a course code) depending on the specified entity (this idea was inspired from the Okkam project<sup>15</sup>). Such unique URIs enable a seamless integration and extraction of linked entities within common objects that exist in the triple store and beyond, one of the core Linked Data requirements [3].

## 3.2 The Data

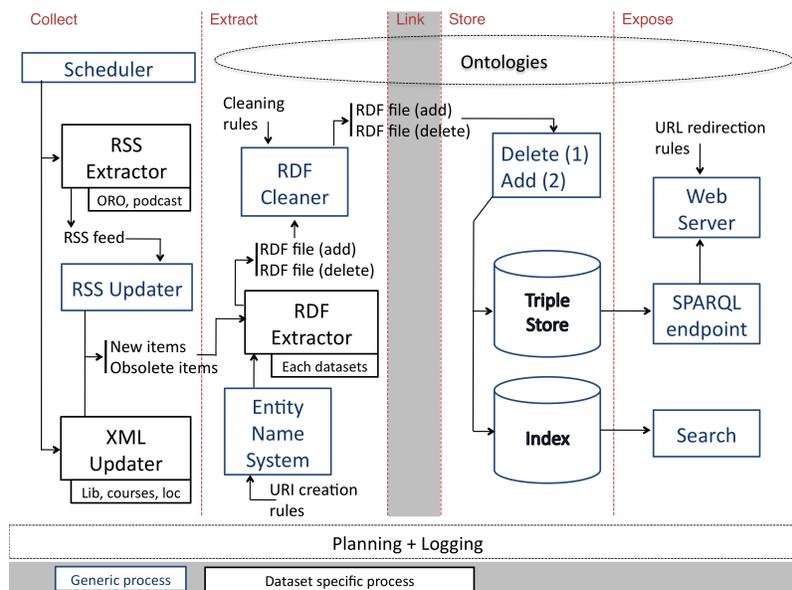
Data about the OU courses, Podcasts and academic publications is already available to be queried and explored, and the team is now working to bring together educational and research content from the university's campus information, OpenLearn (already available for testing purposes) and library material. More concretely, data.open.ac.uk offers a simple browsing mechanism, and a SPARQL endpoint to access the following data:

---

<sup>13</sup> <http://data.open.ac.uk/query>

<sup>14</sup> <http://www.geonames.org>

<sup>15</sup> <http://www.okkam.org>



**Fig. 1.** The LUCERO Workflow

- The Open Research Online (ORO) system<sup>16</sup>, which contains information about academic publications of OU research. For that, the Bibliographic Ontology (bibo)<sup>17</sup> is mainly used to model the data.
- The OU Podcasts,<sup>18</sup> which contain Podcast material related to courses and research interests. A variety of ontologies are used to model this data, including the W3C Media Ontology,<sup>19</sup> in addition to a specialised SKOS<sup>20</sup> representation of the iTunesU topic categories.
- A subset of the courses from the Study at the OU website,<sup>21</sup> which provides courses information and registration details for students. We model this data by relying on the Courseware,<sup>22</sup> AIISO<sup>23</sup> and GoodRelations ontologies [4], in addition to extensions that reflect OU specific information (e.g., course assessment types).

Furthermore, there are other sources of data that are currently being processed. This includes for example the OU list of provided publications, the

<sup>16</sup> <http://oro.open.ac.uk>

<sup>17</sup> <http://bibliontology.com/specification>

<sup>18</sup> <http://podcast.open.ac.uk>

<sup>19</sup> <http://www.w3.org/TR/mediaont-10>

<sup>20</sup> <http://www.w3.org/2004/02/skos>

<sup>21</sup> <http://www3.open.ac.uk/study>

<sup>22</sup> <http://courseware.rkbexplorer.com/ontologies/courseware>

<sup>23</sup> <http://vocab.org/aiiso/schema>

library catalogue, and public information about locations on the OU campus (e.g., buildings) and university staff.

## 4 The OU Expert Search

Expert search can be defined as the task of identifying people who have relevant expertise in a topic of interest. This task is key for every enterprise, but especially for universities, where interdisciplinary collaborations among research areas is considered a high success factor. Typical user scenarios in which expert search is needed within the university context include: a) finding colleagues from whom to learn, or with whom to discuss ideas about a particular subject; b) assembling a consortium with the necessary range of skills for a project proposal, and; c) finding the most adequate reviewers to establish a program committee.

As discussed by Yimam-Seid and Kobsa [7], developing and manually updating an expert system database is time consuming and hard to maintain. However, valuable information can be identified from documents generated within an organisation [8]. Automating expert finding from such documents provides an efficient and sustainable approach to expertise discovery.

OU researchers, students and lecturers constantly produce a plethora of documents, including for example conference articles, journal papers, thesis, books, reports and project proposals. As part of the LUCERO project, these documents have been pre-processed and made accessible as LOD. The purpose of this application is therefore to exploit such information so that OU students and researchers can find the most appropriate experts starting from a topic of interest.<sup>24</sup>

### 4.1 Consumed Data

This application is based on two main sources of information: (a) LOD from the Open Research Online system, and (b) additional information extracted from the OU staff directory. The first information source is exploited in order to extract the most suitable experts about a certain topic. The second information source complements the previous recommended set of experts by providing their corresponding contact information within the OU. Note that sometimes, ex-OU members and external collaborators or OU researchers may appear in the ranking of recommended experts. However, for those individuals, no contact information is provided, indicating that those experts are not part of the OU staff.

As previously mentioned, the information provided by Open Research Online contains data that describe publications originating from OU researchers. In particular, among the properties provided for each publication, this system exploits the following ones: a) the title, b) the abstract, c) the date, d) the authors and, e) the type of publication, i.e., conference paper, book, thesis, journal paper, etc.

---

<sup>24</sup> The OU Expert Search is accessible to OU staff at: <http://kmi-web15.open.ac.uk:8080/ExpertSearchClient>

To exploit this information the system performs two main steps. Firstly when the system receives the user's query, i.e., the area of expertise where a set of experts need to be found (e.g., "semantic search"), the system uses the title and abstract of the publications to find the top- $n$  documents related to that area of expertise. At the moment  $n$  has been empirically set to 10.

Secondly, once the top- $n$  documents have been selected, the authors of these documents are extracted and ranked according to five different criteria: (a) original score of their publications, (b) number of publications, (c) type of publications, (d) date of the publications and, (e) other authors of the publication.

The initial score of the publications is obtained by matching the user's keyword query against the title and the abstract of the OU publications. Publications that provide a better match within their title and abstract against the keywords of the query are ranked higher. This matching is performed and computed using the Lucene<sup>25</sup> text search engine. Regarding the number of publications, authors with a higher number of publications (among the top- $n$  previously retrieved) are ranked higher. Regarding the type of publication, these are ranked first, then books, then journal papers, and finally conference articles. The rationality behind this is that an author writing a thesis or a book holds a higher level of expertise than an author who has only written conference papers. Regarding the date of the publication, we consider the 'freshness' of the publications and continuity of an author's publications within the same area. More recent publications are ranked higher than older ones, and authors publishing in consecutive years about a certain topic are also ranked higher than authors that have sporadic publications about the topic. Regarding other authors, experts sharing a publication with fewer colleagues are ranked higher. The rationality behind this is that the total knowledge of a publication should be divided among the expertise brought into it, i.e., the number of authors. Additionally we also consider the order of authors in the publication. Main authors are considered to have a higher level of expertise and are therefore ranked higher.

To perform the first step (i.e., retrieving the top- $n$  documents related to the user's query) we could have used the SPARQL endpoint and, at run-time, searched for those keywords within the title and abstract properties of the publications. However, to speed the search process up, and to enhance the query-document matching process, we have decided to pre-process and index the title and abstract information of the publications using the popular Lucene search engine. In this way, the fuzzy and spelling check query processing and ranking capabilities of the Lucene search engine are exploited to optimise the initial document search process.

To perform the second step, once the top- $n$  documents have been selected, the rest of the properties of the document (authors, type, and date) are obtained at run-time using the SPARQL endpoint.

Finally, once the set of authors have been ranked, we look for them in the OU staff directory (using the information about their first name and last name). If the author is included in the directory, the system provides related information about

---

<sup>25</sup> <http://lucene.apache.org/java/docs/index.html>

the job title, department within the OU, e-mail address and phone number. By exploiting the OU staff directory we are able to identify which experts are members of the OU and which of them are external collaborators, or old members not further working for the institution.

Without the structure and conceptual information provided by the OU LOD, the implementation of the previously described ranking criteria, as well as the interlinking of data with the OU staff directory, would have required a huge data pre-processing effort. The OU LOD provides the information with a fine-grained structure that facilitates the design of ranking criteria based on multiple concepts, as well as the interlinking of information with other repositories.

## 4.2 System Implementation

The system is based on lightweight client server architecture. The back end (or server side) is implemented as a Java Servlet, and accesses the OU LOD information by means of HTTP requests to the SPARQL endpoint. Some of the properties provided by the LOD information (more particularly the title and the abstract of the publications) are periodically indexed using Lucene to speed-up and enhance the search process by means of the exploitation of its fuzzy and spell checker query processing, and ranking capabilities. The rest of the properties (authors, date, and type of publications) are accessed at run time, once the top- $n$  publications have been selected.

The front end is a thin client implemented as a web application using only HTML, CSS and Javascript (jQuery).<sup>26</sup> The client doesn't handle any processing of the data, it only takes care of the visualisation of the search results and the search input. It communicates with the back-end by means of an HTTP request that passes as a parameter the user's query and retrieves the ranking of authors and their corresponding associated information by means of a JSON object.

## 4.3 Example and Screenshots

In this section, we provide an example of how to use the OU expert search system. As shown in Figure 2, the system receives as a keyword query input “*semantic search*”, with the topic for which the user aims to find an expert. As a result, the system provides a list of authors (“*Enrico Motta*”, “*Vanessa Lopez*”, etc), who are considered to be the top OU experts in the topic. For each expert, if available, the system provides the contact details (department, e-mail, phone extension) and the top publications about the topic. For each publication, the system shows its title, the type of document, and its date. If the user passes the cursor on the top of the title of the publication, the summary is also visualised (see the example in Figure 2 for the publication “*Reflections of five years of evaluating semantic search systems*”). In addition the title of the publication also constitutes a link to its information in the open.ac.uk domain.

---

<sup>26</sup> <http://www.jquery.com>

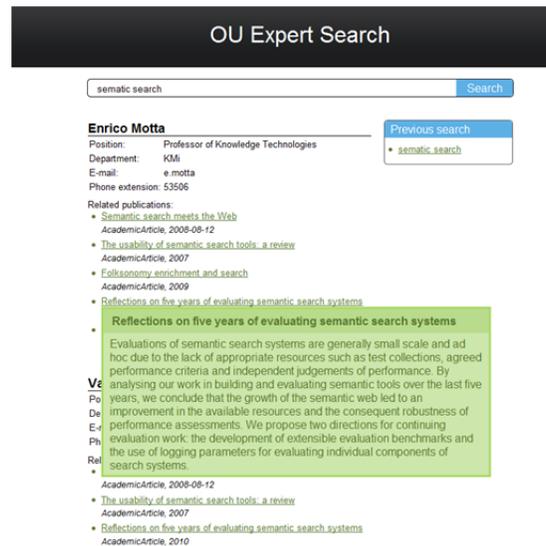


Fig. 2. The OU Expert Search system

## 5 Buddy Study

The Open University is a well-established institution in the United Kingdom, offering distance-learning courses covering a plethora of subject areas. A key factor in enabling learning and understanding of course materials is support for students, provided in the form of an on-hand tutor for each studied module, where interactions with the tutor are facilitated via the Web and/or email exchanges. An alternative method of support could be provided through peers, in a similar manner to a classroom environment, where working together and explanations of problems from disparate viewpoints enhances understanding.

Based on this thesis, Buddy Study<sup>27</sup> combines the popular social networking platform Facebook with the OU Linked Data service, the goal being to suggest learning partners – so called ‘Study Buddies’ – from a person’s social network on the site together with possible courses that could be pursued together.

### 5.1 Consumed Data

Buddy Study combines information extracted from Facebook with Linked Data offered by The Open University, where the former contains ‘wall posts’ – messages posted publicly on a person’s profile page – and comments on such wall posts, while the latter contains structured, machine-readable information describing courses offered by The Open University.

<sup>27</sup> <http://www.matthew-rowe.com/BuddyStudy>

Combining the two information sources, in the form of a ‘mashup’, is performed using the following approach. First the user logs into the application – using Facebook Connect – and grants access to their information. The application then extracts the most recent  $n$  wall posts and the comments on those posts –  $n$  can be varied, thereby affecting the later recommendations. Given the extracted content, cleaning is then performed by removing all the stop words, thus reducing the wall posts and comments to their basic terms.

A bag of words model is compiled for each person in the user’s social network as follows: for each wall post or comment posted by a given person all the terms are placed in the bag, maintaining duplicates and therefore frequencies. This model maintains information of the association between a user and his/her social network members in the form of shared terms. A bag of words model is then compiled for each OU course in a similar manner: first we query the SPARQL endpoint of the OU’s Linked Data asking for the title and description for each course. For the returned information, stop words are removed and the title and description – containing the remaining terms – are then used to build the bag of words model for the course.

The goal of Buddy Study is to recommend study partners to support course learning. Therefore we compare the bag of words model of each person with the bag of words model of each course, recording the frequency and terms that overlap. The user’s social network members are then ranked based on the number of overlapping terms – the intuition being that the greater the number of common terms with courses, the greater the likelihood of a course being correlated with the user. Variance of  $n$  will therefore affect this ranking, given that the inclusion of a greater number of posts will increase the number of possible study partners, while smaller values for  $n$  will yield more recently interacted with social network members. Variance of this parameter is provided in the application.

The application is not finished yet; we still need to recommend possible courses that could be studied with each possible study buddy. This is performed in a similar fashion, by comparing the bag of words model of the social network member with the model of each course, counting the frequencies of overlapping terms for each course, and then ranking accordingly. Due to space restrictions, and to avoid information overload, we only show the top-10 courses. For each social network user, and for each course that is suggested, Buddy Study displays the common terms, thereby providing the reasons for the course suggestion.

If for a moment we assume a scenario where Linked Data is not provided by the OU, then the function of Buddy Study could, in theory continue, by consuming information provided in an alternative form. However, this application forms the prototype upon which for future work – explained in greater detail within the conclusions of this paper – is to be based. Such advancements will utilise concepts for study partner recommendation rather than merely terms, the reasoning behind this extension is to alleviate the noisy form that terms take. By leveraging concepts from collections of terms, recommendations would be generated that are more accurate and better suited to the user in question. Without Linked Data, this is not possible.

## 5.2 System Implementation

The application is live and available online at the previously cited URL. It is built using PHP, and uses the Facebook PHP Software Development Kit (SDK)<sup>28</sup>. Authentication is provided via Facebook Connect,<sup>29</sup> enabling access to Facebook information via the Graph API. The ARC2 framework<sup>30</sup> is implemented to query the remote SPARQL endpoint containing The Open University's Linked Data, and parse the returned information accordingly.

## 5.3 Example and Screenshots

To ground the use of Buddy Study, Figure 3 shows an example screenshot from the application when recommending study partners for Matthew Rowe – one of the authors of this paper. At this rank position in the results, the possible study mate is shown together with the courses that could be studied together. The courses are hyperlinked to their resource within the OU Linked Open Data service, and in the proceeding brackets the terms that correlate with the courses are shown. In this instance the top-ranked course is identified by the common terms ‘API’ and ‘Info’.

```
7. Dan Brickley
  • Keeping ahead in information and communication technologies (api, info, )
  • Communication and information technologies (api, info, )
  • Investigating the psychological world (info, )
  • Fractal geometry (terms, )
  • Diabetes care (info, )
  • Engineering: an active introduction (info, )
  • Issues in international finance and investment (api, )
  • Management and business research (youre, )
  • Postgraduate foundation module in music (info, )
  • Engineering small worlds: micro and nano technologies (terms, )
```

Fig. 3. Buddy Study showing the 7th ranked social network member

## 6 Linked OpenLearn

The Open University offers a set of free learning material through the OpenLearn website.<sup>31</sup> Such material cover various topics ranging from Arts<sup>32</sup>, to Sciences and Engineering.<sup>33</sup> In addition to that, the OU has other learning resources published in the form of Podcasts, along with courses offered at specific presentations during the year. While all these resources are accessible online, connections are

<sup>28</sup> <https://github.com/facebook/php-sdk>

<sup>29</sup> <http://developers.facebook.com/docs/authentication>

<sup>30</sup> <http://arc.semsol.org>

<sup>31</sup> <http://openlearn.open.ac.uk>

<sup>32</sup> OpenLearn unit example in Arts: [http://data.open.ac.uk/page/openlearn/a216\\_1](http://data.open.ac.uk/page/openlearn/a216_1)

<sup>33</sup> A list of units and topics is available at: <http://openlearn.open.ac.uk/course>

not always explicitly available, making it hard for students to easily exploit all the available resources. For example, while there exists a link between specific Podcasts and related courses, such links do not exist between OpenLearn units and Podcasts. This leaves it to the user to infer and find the appropriate and relevant material to the topic of interest.

Linked OpenLearn<sup>34</sup> is an application that enables exploring linked courses, Podcasts and tags to OpenLearn units. It aims to facilitate the browsing experience for students, who can identify on the spot relevant material without leaving the OpenLearn page. With this in place, students are able, for example, to easily find a linked Podcast, and play it directly without having to go through the Podcast website.

## 6.1 Consumed Data

Linked OpenLearn relies on The Open University's Linked Data to achieve what was previously considered very costly to do. Within large organizations, it's very common to have systems developed by different departments, creating a set of disconnected data silos. This was the case of Podcasts and OpenLearn units at the OU. While courses were initially linked to both Podcasts and OpenLearn in their original repositories, it was practically hard to generate the links between Podcasts and OpenLearn material. However, with the deployment of Linked Data, such links are made possible through the use of coherent and common URIs of represented entities.

To achieve our goals of generating relevant learning material, we make use of the courses, Podcasts, and OpenLearn datasets in `data.open.ac.uk`. As a first step, while the user is browsing an OpenLearn unit, the system identifies the unique reference number of the unit from the URL. Then this unique number is used in the query passed to the OU Linked Data SPARQL endpoint (`http://data.open.ac.uk/query`), to generate the list of related courses including their titles and links to the study at the OU pages.

In the second step, another query is sent to retrieve the list of Podcasts related to the courses fetched above. At this level we get the Podcasts' titles, as well as their corresponding downloadable media material (e.g., video or audio files), which enable users to play the content directly within the application. Finally the list of related tags are fetched, along with an embedded query that generates the set of related OpenLearn units, displayed in a separate window. The user at this level has the option to explore a new unit, and the corresponding related entities will be updated accordingly. The application is still a prototype, and there is surely room for further data to extract. For example, once the library catalogue is made available, a much richer interface can be explored by students with related books, recordings, computer files, etc.

---

<sup>34</sup> <http://fouad.zablith.org/apps/openlearnlinkeddata>

## 6.2 System Implementation

We implemented the Linked OpenLearn application in PHP, and used the ARC2 library to query the OU Linked Data endpoint. To visualise the data on top of the web page, we relied on the jQuery User Interface library,<sup>35</sup> and used the dialog windows for displaying the parsed SPARQL results. The application is operational at present, and is launched through a Javascript bookmarklet, which detects the OpenLearn unit that the user is currently browsing, and opens it in a new iFrame, along with the linked entities visualised in the jQuery boxes.

## 6.3 Example and Screenshot

To install the application, the user has to drag the applications' bookmarklet<sup>36</sup> to the browser's toolbar. Then, whenever viewing an OpenLearn unit, the user clicks on the bookmarklet to have the related entities displayed on top of the unit page. Figure 4 illustrates one arts related OpenLearn unit, with the connected entities displayed on the right, and a running Podcast selected from the "Linked Podcasts" window. The user has the option to click on the related course to go directly to the course described in the Study at the OU webpage, or click on linked tags to see the list of other related OpenLearn units, which can be browsed within the same window.

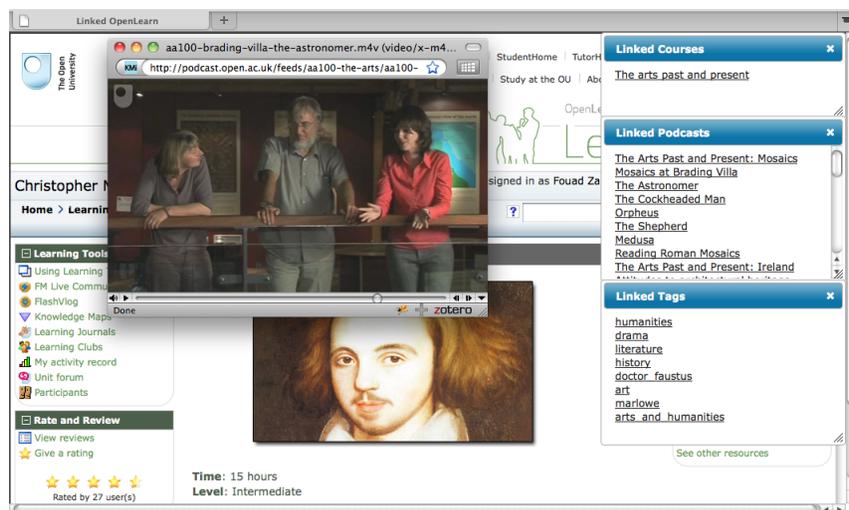


Fig. 4. Linked OpenLearn Screenshot

<sup>35</sup> <http://www.jqueryui.com>

<sup>36</sup> The bookmarklet is available at: <http://fouad.zablith.org/apps/openlearnlinkeddata>, and has been tested in Firefox, Safari and Google Chrome

## 7 Conclusions

In this section we report on our experiences when generating and exploiting LOD within the context of an educational institution. Regarding our experience on transforming information distributed in several OU repositories and exposing it as LOD, the process complexity was mainly dependent on the datasets in terms of type, structure and cleanliness. Initially, before any data transformation can be done, it was required to decide on the vocabulary to use. This is where the type of data to model plays a major role. With the goal to reuse, as much as possible, already existing ontologies, it was challenging to find the adequate ones for all our data. While some vocabularies are already available, for example to represent courses, it required more effort to model OU specific terminologies (e.g., at the qualifications level). To assure maximum interoperability, we chose to use multiple terminologies (when available) to represent the same entities. For example, courses are represented as modules from the AIISO ontology, and at the same time as courses from the Courseware ontology. Other factors that affected the transformation of the data are the structure and cleanliness of the data sources. During the transformation process, we faced many cases where duplication, and information not abiding to the imposed data structure, hampered the transformation stage. However, this initiated the need to generate the data following well-defined patterns and standards, in order to get easily processable data to add to the LOD.

Regarding our experiences exploiting the data, we have identified three main advantages of relying on the LOD platform within the context of education. Firstly the exposure of all these material as free Web resources have open opportunities for the development of novel and interesting applications like the three presented in this paper. The second main advantage is the structure provided by the data. This is apparent in the OU Expert Search system, where the different properties of articles are exploited to generate different ranking criteria, which when combined, provide much stronger support when finding the appropriate expertise. Finally, the links generated across the different educational resources have provided a new dimension to the way users can access, browse and use the provided educational resources. A clear example of this is the exploitation of LOD technology within the OpenLearn system, where OpenLearn units are now linked to courses and Podcasts, allowing students to easily find in a single site, all the information they are looking for.

We believe that universities need to evolve the way they expose knowledge, share content and engage with learners. We see LOD as an exciting opportunity that can be exploited within the education community, especially by interlinking people and educational resources within and across institutions. This interlinking of information will facilitate the learning and investigation process of students and research staff, enhancing the global productivity and satisfaction of the academic community. We hope that, in the near future, more researchers and developers will embrace LOD approach, by creating new applications and learning from previous experiences to expose more and more educational data in a way that is directly linkable and reusable.

## 8 Future Work

The application of Linked Data within the OU has opened multiple research paths. Regarding the production of Linked Data, in addition to transforming the library records to LOD, the LUCERO team is currently working on connecting the OU's Reading Experience Database (RED)<sup>37</sup> to the Web of Data. Such database aims to provide access and information about reading experiences around the world. It helps the readership for books issued in new editions for new audiences in different countries to be tracked. Its publication as LOD is an interesting example about how the integration of Linked Data technology can open new investigation paths to different research areas, in this case humanities.

Regarding the consumption of LOD, we envision, on the one hand, to enhance the three previously mentioned applications and, on the other hand to generate new applications as soon as more information is available and interconnected. As example of the former, for the Buddy Study application we plan to extend the current approach for identifying common terms between social network members and courses to instead utilise common concepts. At present the use of online messages results in the inclusion of abbreviated and slang terms, resulting in recommendations that are generated from noise. By instead using concepts, we believe that the suggested courses would be more accurate and suitable for studying. As an example of the latter, we aim to generate a search application over the RED database, able to display search results on an interactive map and link them not just to relevant records within the RED database, but also with relevant objects of the LOD cloud.

## References

1. C. Bizer. The emerging web of linked data. *IEEE Int. Systems*, pages 87–92, 2009.
2. C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.
3. T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. 2011.
4. M. Hepp. GoodRelations: an ontology for describing products and services offers on the web. *Knowledge Engineering: Practice and Patterns*, pages 329–346, 2008.
5. G. Kobilarov, T. Scott, Y. Raimond, S. Oliver, C. Sizemore, M. Smethurst, C. Bizer, and R. Lee. Media meets semantic webhow the bbc uses dbpedia and linked data to make connections. pages 723–737, 2009.
6. M. Rowe. Data.dcs: Converting legacy data into linked data. In *Linked Data on the Web Workshop, WWW2010*, 2010.
7. D. Yimam-Seid and A. Kobsa. Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, 13(1):1–24, 2003.
8. J. Zhu, X. Huang, D. Song, and S. Rüger. Integrating multiple document features in language models for expert finding. *Knowledge and Information Systems*, 23(1):29–54, 2010.

---

<sup>37</sup> <http://www.open.ac.uk/Arts/reading>

# Using Linked Data to Reduce Learning Latency for e-Book Readers

Julien Robinson, Johann Stan, and Myriam Ribière

Alcatel-Lucent Bell Labs France, 91620 Nozay, France,  
Julien.Robinson@alcatel-lucent.com

**Abstract.** Web 2.0 technologies provide an opportunity to transform learning into a social experience: social learning can directly benefit from user-generated content analysis. The e-book device is an object central to the learning process; it becomes an ideal medium to promote innovative learning tools. In this paper, we describe how we leverage user generated content in e-books, generally in the form of unstructured annotations, to reduce the latency between professor knowledge, book content and student assimilation.

**Keywords:** Social Learning, Semantic Web, Linked Data, e-Book

## 1 Introduction

The set of people learning with a given e-book form a de-facto community, which is even more relevant if they are reading at the same time, with the same purpose or during the same activity, e.g. a class. Currently, very few platforms allow social interactions between such people that are dedicated to their reading activity; readers can connect but at a very high level and generally outside the platform (e.g. forums in LibraryThing<sup>1</sup>, chatting in Bookglutton<sup>2</sup> or more general social cataloging features GoodReads<sup>3</sup> and Shelfari<sup>4</sup>).

The main problem with these platforms is that interactions cannot be directly linked to the context of reading and thus, do not allow fine-grained interactions about a specific part of the book. Bookglutton goes a step further by letting users chat while reading, but does not exploit the contents and context of annotations in the e-book. This issue was already explored [1] [2], but no solution was proposed based on the semantic analysis of the annotations.

An important observation for our work is the fact that in most cases, annotations in e-books are in the form of short, unstructured textual messages, associated to images, paragraphs or chapters. These can include transcriptions, reformulations, additional comments, questions or explanations (such as student

---

<sup>1</sup> <http://www.librarything.com/> - visited September 2010

<sup>2</sup> <http://www.bookglutton.com/> - visited January 2011

<sup>3</sup> <http://www.goodreads.com/> - visited October 2010

<sup>4</sup> <http://www.shelfari.com/> - visited March 2011

annotations in BookGlutton<sup>5</sup>). Clearly, services that leverage these annotations can increase the social dimension of reading. They can help readers determine a common environment to socially learn by seeing others' annotations or to generate interactions between students during a same activity.

In this paper we present our ongoing work on services that leverage semantic technologies and especially Linked Data for providing a social dimension for reading and implicitly, learning.

## 2 Management of Shared Annotations using Linked Data

Linked Data is an evolving component of the Semantic Web. It can be defined as a community effort to extract or publish data from various sources, interlink it and store it in a semantic web format, such as RDF. The content of these online linked datasets is diverse. It consists mostly of data about people, companies, books, geographic locations, films, music, television and radio programs, information about medicine (genes, proteins, clinical trials) and online communities (statistical data). The most used online dataset is certainly DBPedia[3].

An important source of social knowledge is the set of annotations produced by the user. In our framework, our primary hypothesis is to consider these annotations on a conceptual level. More concretely, this means that we associate each annotation to one or several concepts in a Linked Data source (e.g. DBPedia). There are several reasons to make this association: (i) provide a rich approximation of their meaning, (ii) the capability to retrieve related concepts and category terms, operation that we call semantic expansion<sup>6</sup> and (iii) the fact that annotations are generally short and that contextual cues are not always available. Therefore, Linked Data sources can enrich the description of such annotations with additional metadata and related concepts.

The association of annotations to concepts in Linked Data is performed by a disambiguation process that leverages the context of the annotation to match it with the most relevant concept in the knowledge base. Each annotation provided by the user is associated to a context, which is constructed from different sources, such as previous annotations shared by the user in the e-book and the set of annotations shared by others that are related to the given paragraph. Parallel to this, a set of candidate concepts are retrieved from DBPedia, that contain the given annotation in their label. The candidate concepts are ranked using cosine similarity between the vector constructed from their abstracts and the contextual cue vector. Once this disambiguation has been performed, the second use of Linked Data is the expansion of the concept by exploring both its semantic neighborhood and hierarchical links.

<sup>5</sup> e.g. <http://www.bookglutton.com/glutton/melissthangx0x0/14459>

<sup>6</sup> As an example, here is the semantic expansion of the annotation "Facebook": [Global Internet Community, Student Culture, Social Information Processing, Social Networking Service, Social Media, Web 2.0, Internet Technologies]

### 3 Easy Social Navigation with Heatmaps

Key to the exploitation of Linked Data results in social learning is the way the user will perceive the analysis results and whether it will be useful or not. In [1], Jae K. Kim et al. describe how they let students place symbolic evaluations as annotations, and then graphically represent the synthesis of these annotations as colors in the e-book, thus enabling what is called **Social Navigation**. We follow the same investigation track, but instead of focusing on design choices and symbolic evaluations, we analyze free text annotations using Semantic Web techniques.

The user, a learning reader (our target is post-graduate students), opens an e-book and states their interest using freely chosen keywords; the keywords express their goal for the reading session. The keyword is semantically analyzed and a global similarity computation is run on the annotations of the book. Each annotation is attributed a heat value based on its relevance to the user's goal in reading the book. The appearance of the book is modified to display in a colored way the zones where the annotations are the most heated.

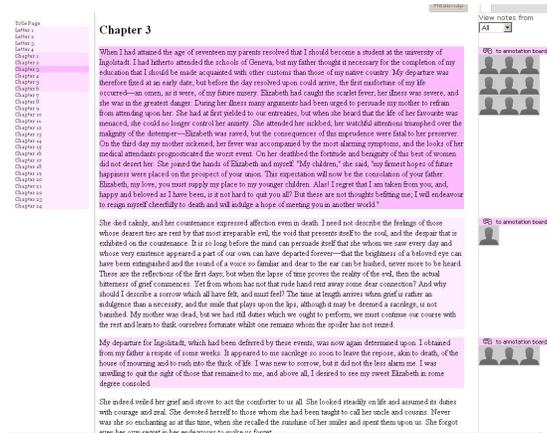
Figure 1(a) presents an example of the heatmap feature. This was produced by our prototype of the sBook [4] using data gathered from annotations from BookGlutton (229 annotations by 39 different users on the e-book "Frankenstein" by Mary Shelley).

The choice was made to use the paragraph as the finest precision of annotation heat. The heat of a paragraph is the total sum of the heat of each annotation on the paragraph. Note that the heatmap is a navigation indication for the user, it is not meant to be the final step for a learning reader; it is completed by the display of authors of annotations (on the right-hand panel) and the use of what we call annotation boards: these are separate windows that browse through all the annotations of a paragraph and the comments inspired by these annotations. The annotation board is where the user may find valuable information and participate in discussions.

The choice of the algorithm for heating an annotation is the focus of our current research. Figure 1(b) presents the administration interface that enables to choose different implementations for the user to experiment with. Annotations are heated based on their proximity to the resulting vector of words. We currently implement three expansion methods:

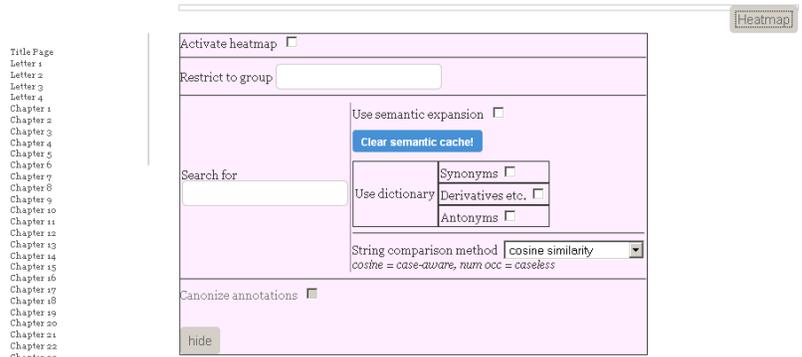
- no expansion, we simply search for the keyword
- a dictionary-based heatmap, that may take into account synonyms, antonyms and derived words (e.g. adjectives for a noun)
- a Linked-Data-based heatmap, that takes into account the expansion of the search keyword as described in the present paper.

Our first experiments show that named entities (e.g. "Paris") are best expanded using Linked Data, while literary theory terms (e.g. "poetry", "characterization") are best expanded using dictionary techniques. This is why we keep both options open and also separate, so as to validate their usefulness with user experimentation.



(a) sBook with Heatmap activated (overview)

Frankenstein, by Mary Wollstonecraft Shelley



(b) Administration Heatmap Configuration

Fig. 1. Heatmap screenshots

In a first step the annotations are processed as text; currently being implemented, we first process each annotation by associating it to a cluster of tags, obtained by semantic evaluation techniques. The similarity measure will be computed between the cluster of tags of the annotation and the vector of words expanded from the user keyword.

It may be noted that, while we explore keyword-based heatmaps, we also intend to explore other options: time-based heatmap, where the heat of an annotation is related to its creation date; location-based heatmap, taking into account the location of the annotation's author; or more generally context-based and activity-based heatmaps (which present intersections). These will be used for further user experimentations, first to compare their effectiveness with keyword-based heatmaps, second because they are relevant to our general research on the link between the digital and physical worlds.

## 4 User Interaction Profiles

Each user in the collaborative learning environment associated to the e-book is represented by an interaction profile, constructed from their shared annotations in the system. The objective of the interaction profile is to provide a digital representation of the user's expertise. Each concept in the profile can be manually annotated by users. More specifically, we propose an interactive tool for users to declare their level of expertise and to choose how much of this expertise to share in different social spheres and situations (Figure 2).

Thus, a user can for example say that they have high expertise in "European Union Member States" but that they wish to share this expertise only with classmates and teachers and that they are ready to interact on this topic only when they are at the library or at school. With the help of DBPedia, we also retrieve the hierarchical tree associated to the concept, which allows the user to share the same concept with other social spheres, but with less granularity. In this way, the user can share different levels of granularity about the same knowledge with multiple social spheres (e.g. detailed knowledge for the class, less detailed for friends).

Linked Data allows to retrieve the hierarchical tree and semantic neighborhood associated to a given concept. In the case of collaborative learning, we leverage this feature by offering users a more granular way of sharing their interaction profile, constructed from the shared annotations. Such interaction profiles make it possible for the system to select the right person to ask a question on a specific topic. In order to overcome the problem of data sparsity in case of new members in such learning communities, we are currently investigating the possibility to connect this service to other Social Platforms and to build a user interaction profile from annotations shared in them. In this way, the user will be recommended people to interact with even if they did not share sufficient annotations. More specifically, each time the user opens the e-book, they will be presented with the most relevant readers to interact with, based on complementary expertise in specific topics of interest and declared sharing preferences related to that topic. This provides students with a seamless way of linking to co-learners and thus form spontaneous learning communities on a specific part of a lecture.

## 5 Conclusion

In this paper we have presented our research, which focuses on integrating Linked Data results into e-book navigation tools for students in order to enhance social learning. We base our work on user annotations, we process these annotations with Linked Data technologies, we use the produced synthesized results to modify our e-book display; we hope to maximize reading and learning efficiency, to create common environments and implicitly, communication opportunities. At the basis of our innovation is the hypothesis that the huge amount of shared content in different Social Platforms offers exciting opportunities to enrich the learning



Fig. 2. Interaction Profile Declaration Tool

experience and reduce the knowledge latency in the class, which we define as the delay between the encoding of the knowledge of a professor on a specific subject into information (e.g. a lecture or slides) and the assimilation of this information as knowledge by students.

## 6 Acknowledgment

We would like to thank BookGlutton for providing open data that we were able to exploit, and in particular Aaron Miller (CTO). The set of ideas about the future of e-books we have presented in this paper are currently being addressed in the scope of a collaborative project between Alcatel-Lucent, Abilene Christian University and Cambridge University Press <sup>7</sup>.

## References

1. Jae K. Kim, Rosta Farzan, and Peter Brusilovsky. Social navigation and annotation for electronic books. In *BooksOnline '08: Proceeding of the 2008 ACM workshop on Research advances in large digital book repositories*, pages 25–28, New York, NY, USA, 2008. ACM.
2. Jill Freyne, Rosta Farzan, Peter Brusilovsky, Barry Smyth, and Maurice Coyle. Collecting community wisdom: integrating social search & social navigation. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces*, pages 52–61, New York, NY, USA, 2007. ACM Press.
3. Jens Lehmann, Chris Bizer, Georgi Kobilarov, Sren Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia - a crystallization point for the web of data. *Journal of Web Semantics*, 7(3):154–165, 2009.
4. Myriam Ribière, Jérôme Picault, and Sylvain Squedin. The ebook: towards social and personalized learning experiences. In *Proceedings of the third workshop on Research advances in large digital book repositories and complementary media*, BooksOnline '10, pages 3–8, New York, NY, USA, 2010. ACM.

<sup>7</sup> Cambridge University - <http://www.cambridge.org/press/article.asp?artid=154961>

# A Linked Knowledge Base for Simulation Learning

Irene Celino and Daniele Dell'Aglio

CEFRIEL – Politecnico of Milano, Via Fucini 2, 20133 Milano, Italy  
{irene.celino,daniele.dellaglio}@cefriel.it

**Abstract.** Simulation Learning is a frequent practice to conduct near-real, immersive and engaging training sessions. AI Planning and Scheduling systems are used to automatically create and supervise learning sessions; to this end, they need to manage a large amount of knowledge about the simulated situation, the learning objectives, the participants' behaviour, etc.

In this paper, we explain how Linked Data and Semantic Web technologies can help the creation and management of knowledge bases for Simulation Learning. We also present our experience in building such a knowledge base in the context of Crisis Management Training.

## 1 Introduction

Traditional research on Semantic Web in e-learning [1, 2] are aimed at promoting interoperability between training systems, thus usually the core investigation targets are standards and schemata to describe learning objects [3, 4].

Our research is focused on a different kind of e-learning system, i.e. Simulation Training to improve soft skills [5]. In this context, not only it is needed to describe learning objects, but also to fully plan simulation sessions; those sessions should be interactive and engaging to challenge the trainees to improve their skills. Simulation Learning systems generally re-create a near-real environment for training sessions, in which learners are subject to stimuli: they have to learn how to deal with the simulated situation and how to react to it.. Such simulations need to be effective and engaging, so that the learners do not simply memorise notions about the specific matter, question or theme, but they actively and permanently acquire skills, practice and knowledge.

The scenario production is therefore the core and critical activity when building a Simulation Learning system. Knowledge technologies are needed to model and manage all the required information, often generated and managed by different and independent sources: scenario descriptions, events and stimuli for the trainees, storyboards for the learning sessions, multimedia assets, supporting documents and guidelines, trainees description and behaviour/decisions, learning session monitoring, etc. Such a wealth of information makes the Simulation Learning a knowledge-intensive context, which requires smart solutions.

We decided to adopt Linked Data and Semantic Web technologies to address the requirements of Simulation Learning. The knowledge diversity and scale

calls for a solution which provides interlinking between different datasets while preserving possibly independent information sources; moreover, the knowledge coherence and consistency must be assured to guarantee the significance, meaningfulness and correctness of simulation scenarios and storyboards presented to trainees.

In this paper, we present our current effort in exploiting Linked Data and Semantic Web technologies to build a Knowledge Base for a Simulation Learning environment. We explain why we believe that the selected technologies not only offer a suitable means to knowledge representation and management, but they are specifically required to address the challenges of such an environment.

Section 2 introduces the basic concepts of Simulation Learning systems and a concrete scenario in Crisis Management Training; Section 3 details our exploration in the use of Linked Data and Semantic Web to build a Simulation Learning Knowledge Base illustrating the gained benefits; Section 4 specifies our modelling choices, while Section 5 suggests that such modelling could benefit from provenance tracking; finally, Section 6 concludes the paper.

## 2 Simulation Learning

Learning should be relevant to people’s workplace and lives: learning content should be truly understood, remembered and applied to actual practices. Only in this way, by actively engaging participants in experiential training, learners can apply their knowledge and learn the best practices [5]; more and more often, indeed, it is not enough to read information and listen to a frontal lecture.

In this section, we introduce the theme of Simulation Learning for Decision-making, we draw a generic architecture of a system to support Simulation Learning and we describe a concrete scenario that we will use throughout the paper to exemplify our approach.

### 2.1 Simulation for Decision-making

Training plays an important function in the preparation of professional practitioners. Currently, there are two main modalities for such training: table-top exercises and real-world simulations. *Table-top exercises* are low cost and can be easily and frequently organised. However, they cannot create a believable atmosphere of stress and confusion, which is prevailing in real-life situations and is crucial to the training of timely and effective decision making. On the other hand, training through *simulation exercises* on the field can be very effective [6], but it is considerably more expensive, it can require specialist equipment and it can be difficult to organise.

Simulation exercises require an Exercise Director (or *trainer*) who plays a key role in every form of exercise: the trainer has access to the whole exercise programme, ensures that it proceeds according to a plan, often feeds information to the “players” (the *trainees*) to let them make informed decisions in response (verbally or by written messages). Sometimes information fed to the trainees

is timed in advance at pre-set intervals, regardless of the previous responses. However, flexibility allows a trainer to use judgement and experience in timing the inputs: his/her role should be aimed to facilitate rather than orchestrate the exercise, thus intervention should be minimal and trainees should be given time to recognise and correct problems. Nevertheless, usually it is up to the trainer to decide, for example, how much advice to give to trainees.

## 2.2 Architecture of a Simulation Learning System

The architecture of a Simulation Learning System is depicted in Figure 1. In the picture, we can identify the involved actors, which are the *trainees* – the learning participants engaged in the simulation – and the *trainer* – who activates the exercise and monitors the progress of actions during the training session.

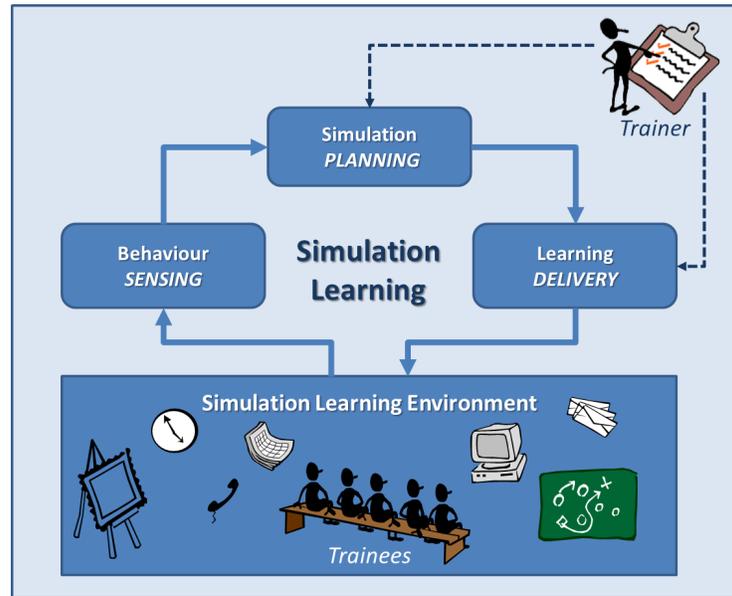
The figure also shows the four main modules of such an architecture, the first three following the usual AI sense-think-act cycle:

- *Behaviour Sensing*: this module is aimed to create and update a model of each trainee from sensors information (e.g. heart rate, blood pressure, respiration); the model represents trainee’s future and actual behaviour and provides indications on how to personalise the training path.
- *Simulation Planning*: this module is aimed to create and simulate a training scenario and its evolution, by combining the information in the behavioural model with knowledge about the learning scenarios; the output of this module is the actual simulation storyboard presented to the trainees.
- *Learning Delivery*: this module is aimed to effectively represent the simulation storyboard in the learning environment, including the rendering of audio-video inputs or Non-Player Characters (NPC, cf. Section 4.3).
- *Simulation Learning Environment*: this is the “place” where the training is conducted; the location can be a physical room or a virtual environment where the trainees interact and receive stimuli during a learning session.

The core of such system is therefore the Simulation Planning module, which contains the basic engine for creating active exercises for classes of trainees. The module is responsible for deciding which stimuli are sent to trainees and how they should be coordinated to create a meaningful and effective lesson plan. In broad terms, it is responsible for allocating over time the set of lesson stimuli indexed according to differences in presentation media, emotional characterization, personalization needs, etc.

## 2.3 Crisis Management Training Scenario

There is increasing recognition for the need to train non-technical skills like control and decision making for Crisis Management in national emergencies, high-reliability industries, as well as in industrial workplaces [7, 8]. In the happening of a catastrophic event, it is human behaviour – and often human behaviour alone – that determines the speed and efficacy of the crisis management effects [9].



**Fig. 1.** High-level architecture of a Simulation Learning System (from the classical sense-think-act cycle of AI)

The Pandora project<sup>1</sup> aims to provide a framework to bridge the gap between table-top exercises and real-world simulation exercises for Crisis Management, providing a near-real training environment at affordable cost. Its training system captures the good practice tenets of experiential learning but with greater efficiency and focuses on real, rather than abstract learning environments. The effective use of integrated ICT reduces the high dependence upon the trainer that is currently required to deliver exercises. Moreover, the Pandora framework supports the measurement and performance assessment of Crisis Managers, the key decision makers participating in a training exercise event as trainees.

As such, Pandora is developing an enabling technology to simulate believable dynamic elements of an entire disaster environment by emulating a crisis room (the Simulation Learning Environment). In this context, we are developing a Knowledge Base that makes use of Linked Data and Semantic Web technologies to model and interlink the pieces of data needed in the training simulation sessions. In the rest of the paper, we will use the Crisis Management scenario to exemplify our approach.

<sup>1</sup> Cf. <http://www.pandoraproject.eu/>.

### 3 Our Simulation Learning Linked Knowledge Base

Within a Simulation Learning system, knowledge exchange plays a central role. In this section we give some details about the Simulation Planning module, focusing on the requirements, design and implementation principles of its Knowledge Base. All the technical details are related to the choices made in the Pandora framework.

#### 3.1 Knowledge required to Plan a Simulation

To formalize the lesson plan, it is natural to choose a basic representation from timeline-based planning [10]. A plan is represented as a set of events having a temporal duration, distributed over a time horizon and indexed according to distinct features which should be planned for. This set of events is organized inside a data structure called *Event Network*, very common in current state of the art planning technology. The Event Network is a temporal plan of multimedia communicative acts toward trainees (e.g., e-mail messages, video news from an emergency location, etc.).

The Event Network can be generated by a *Simulation Planner*. This planner compiles static information into the Event Network, and then adapts the events configuration according to the actions of the trainees, thus simulating different courses of action of the world. The planner can be adapted from a generic AI Timeline-based Planning and Scheduling module [10].

The core information item elaborated by a Simulation Planner is the so-called *synchronization*. Synchronizations are the causal rules that regulate the transitions between values on the same planning feature and the synchronization of values among different planning features. In the Crisis Management scenario, synchronizations are used to influence the Crisis Managers' decisions, e.g. to generate changes in the emergency conditions.

When adopting Planning and Scheduling technologies to simulate a scenario, it is worth highlighting how a great effort and amount of time is necessary to understand the problem, capturing all its specificity, and to create a model of the relevant aspects of the domains and the problem [11]. This consideration suggests, on the one hand, the need for identifying commonalities and similarities among the different domains and problems to operate in a more systematic way and, on the other hand, the opportunity to exploit Semantic Web technologies to ease and support the knowledge modelling task.

For those reasons, we have built a *Knowledge Base* with Linked Data and Semantic Web technologies. This KB is a central component in the Simulation Learning system, responsible for collecting and maintaining the “knowledge” about scenarios and training sessions. As such, the KB is the core information source for the simulation: it contains all the knowledge required by the Simulation Planner to “orchestrate” the events during the training sessions. All the causality in a simulation domain is modelled and stored in the KB; this knowledge is then converted by the Simulation Planner into the suitable data

structures to synthesize the Event Network configurations for the lesson plan goals.

### 3.2 Requirements for the Knowledge Base

The Knowledge Base [12] was carefully designed to fulfil a pressing requirement: containing and managing all the knowledge needed to model and run the simulation scenarios, the training events, the trainees’ behaviour, the time sequence, and so on.

To fulfil such a requirement, the KB must reuse pre-existing information (e.g., in the Crisis Management scenario, training procedures, emergency management guidelines) and, in the meantime, it must allow for customization and diversification of training knowledge (e.g., emergency policies and legislation change from country to country). Furthermore, since most of the related information can be pre-existing in a variety of formats, the KB must be able to gather information from heterogeneous sources (e.g., location data from geographic datasets, audio and video inputs from multimedia archives, participants profiles) and to synthesize and interlink such knowledge into a coherent base.

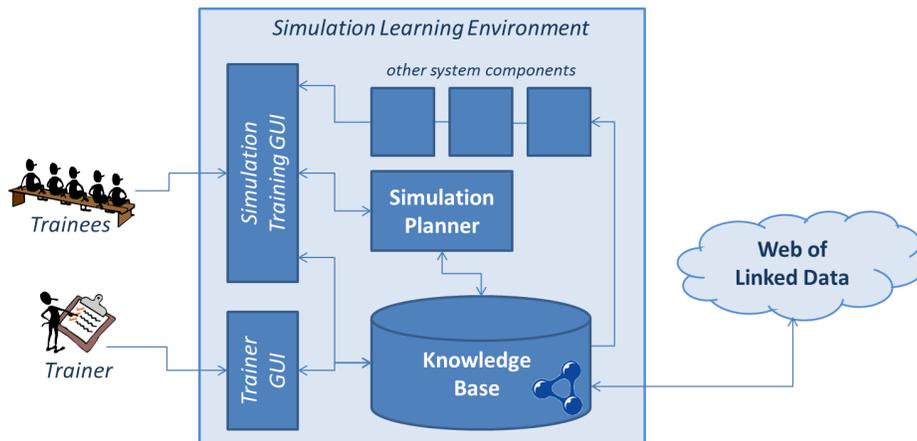


Fig. 2. Role of the Knowledge Base in a Simulation Learning Environment

The role of the KB in the Simulation Learning Environment and its interactions with other components is depicted in Figure 2:

- The KB is “initialized” by the trainer who *models the simulation scenarios* and the training path alternative options;
- It is accessed by the Simulation Planner that needs to understand what *“events” should be triggered* and presented to the trainees during the learning sessions;

- It is also accessed by other system components that need to get/give *information about the training session* and the knowledge exchanged during or after its delivery (cf. Section 4);
- It is used to record the events and decisions taken during each training session, in order to enable the semi-automatically *creation of an individual trainee debriefing report* at the end of the training session.

To cope with such challenges, we adopted Linked Data and Semantic Web technologies for the design and development of our Knowledge Base.

### 3.3 Benefits from the adoption of Linked Data

The choice of Linked Data and Semantic Web technologies in our KB is motivated by the need for an easy access, (re)use and integration of data and knowledge [13].

The *ease of access* to the KB is implicit in the use of Web technologies, which represent a mature and established technology stack. Following the Linked Data principles [14], we provide a standard access means to the data and knowledge stored in the KB. Moreover, Linked Data and Semantic Web facilitate and enable an entity-centric design of Web APIs: in our implementation, on top of the KB, we have developed a RESTful service<sup>2</sup> with specific methods to get details about certain entities on the basis of the concepts (entity types) defined in the KB ontologies and models (cf. Section 4). The RESTful service is also employed to abstract from the physical location of data, as explained further on.

The *reuse of pre-existing datasets* is also enabled by our technological choice. Several useful data sources are already present on the Web of Data and, thus, immediately exploitable by the KB. For example, in the Crisis Management scenario, environment characteristics of crisis settings are retrieved from GeoNames<sup>3</sup>, the geographical database containing over 10 million geographical names, 7.5 million unique features, 2.8 million populated places and 5.5 million alternate names. For example, a scenario about a river flood or a earthquake benefits from the retrieval of localized information from GeoNames. As a pragmatic solution, we are “caching” the relevant features from GeoNames locally to the KB. However, the reuse of GeoNames URIs constitutes a link to the remote dataset and allows for further knowledge retrieval. In the same way, we can connect the KB to other knowledge bases like Freebase<sup>4</sup> or DBpedia<sup>5</sup> [15] to get information on a number of general-purpose topics and entities. The linkage to the latter sources is still in progress.

But this re-usability benefit applies also to the knowledge explicitly modelled for domain-specific learning scenarios: the choice of RDF to encode the data and of RDFS/OWL to model their structure pays, since those data are partially published on the open Web, thus enriching the Web of Linked Data

<sup>2</sup> Cf. <http://pandoratest01.xlab.si:8080/pandora-ckb/>.

<sup>3</sup> Cf. <http://www.geonames.org/>.

<sup>4</sup> Cf. <http://freebase.com/>.

<sup>5</sup> Cf. <http://dbpedia.org/>.

and becoming available for other Simulation Learning systems or for different tools. To this end, in our Crisis Management scenario, we decided to store the schemata and data generated by Pandora components natively as RDF triples in the KB; the knowledge coming from pre-existing sources in different formats (e.g., taxonomies, spreadsheets, guidelines) have been converted – manually or, whenever possible, semi-automatically – to a structured RDF format. The benefits of this approach are: the general Crisis Management knowledge is available to the whole community; the simulation scenarios can be reused by any installation of the training system; the further enhancements and extensions of the core knowledge are immediately “reflected” in all systems that make use of our KB.

The *ease of integration* comes from the native interlinking capability of Linked Data technologies. RDF provides the basic mechanism to specify the existence and meaning of connections between items through RDF links [16]. In other words, through the adoption of RDF, we not only give a structure to the data stored in the KB, but we also interlink the entities described by such data. Moreover, the links drawn between knowledge items are typed, thus conveying the “semantics” of such relationships and enabling the inference of additional knowledge. The information sources of the KB can be maintained and evolve over time in an independent way, but, in the meantime, can be connected via the Linked Data lightweight integration means.

The KB contains different (although interlinked) datasets, which also require diverse confidentiality/security levels for management and access. To this end, the KB is designed as a set of federated RDF stores<sup>6</sup>: the shared knowledge (e.g. general Crisis Management information, basic scenarios) should be “centralised”, to let all training system instances access and use it, while the installation-specific knowledge (e.g., detailed or customized scenarios, trainees information, personalizations) is managed in a local triple store, not accessible from outside the system (see Figure 3). The RESTful service on top of the KB, as explained earlier, provides a uniform access to the KB and hides to the other Pandora components the existence of the various “realms” of distinct Linked Data sources.

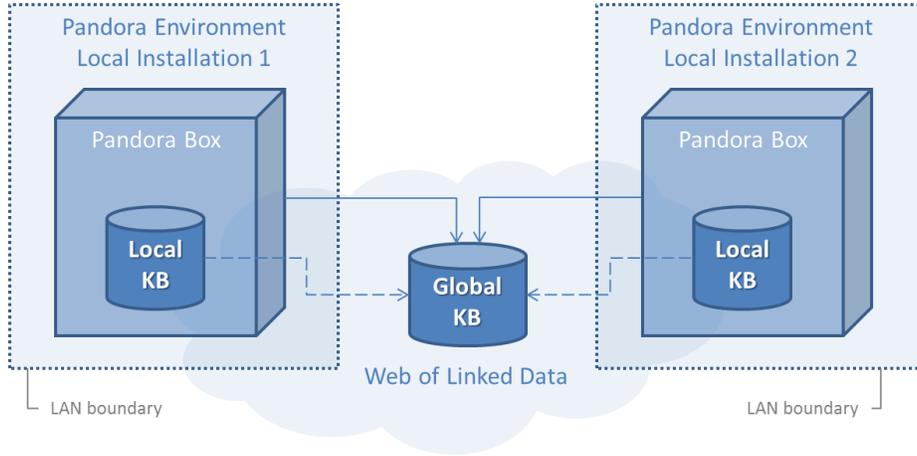
Finally, the adoption of Semantic Web technologies in the form of ontologies and rules provides a further gain, since we can exploit *reasoning and inference* for knowledge creation and consistency checking, as explained in next section.

## 4 Modelling and Retrieval in our Knowledge Base

As previously mentioned, our Knowledge Base manages several different and interlinked types of information. In this section, we introduce three “families” of data included in the KB and explain their modelling choices. We also illustrate their use in the Crisis Management training scenario within the Pandora Integrated Environment.

---

<sup>6</sup> In the Pandora project, since the work is still in progress and for now we have one single system installation, the current initial release of the KB consists of a unique triple store with all the integrated knowledge.



**Fig. 3.** The KB as federation of different triple stores to preserve security and confidentiality while benefitting from interlinking.

#### 4.1 User Modelling

As introduced in Section 2.2, a Behaviour Sensing module is devoted to the “detection” of trainees’ performance in order to create individual models that help in tailoring the learning strategy of each participant to the simulation. Prior to the training session, dedicated psychological tests and physiological assessment at rest (e.g., through a Holter that measures the heart rate activity at rest) are used to measure some relevant variables (like personality traits, leadership style, background experience, self-efficacy, stress and anxiety). Those variables are then updated during the training session, through self-assessment measurements (i.e., asking the trainee about his performance) or through the elaboration of the raw data recorded by the sensors.

Those data about trainees’ behaviour are stored and updated in our KB, as instances of ontology concepts that represent the “affective factors” that influence the decision-making of the trainees. Due to the sensitivity of such information, the individual performances of the trainees are modelled in RDF and stored in the “local” triple store (cf. Figure 3) for apparent privacy reasons. We are also investigating the possibility to exploit Named Graphs [17] for access control: if the training session recordings are “stored” in the KB as separated named graphs, a named graph-aware access control component could grant admission to the allowed users (e.g., the trainer) and could deny the access of the malicious or occasional users (e.g., the other trainees).

In the specific scenario of the Pandora Integrated Environment, the learning sessions are targeted to the training of Crisis Managers. Therefore, the KB stores and manages also a set of specific information about them.

The Crisis Managers are the so-called Gold Commanders, who are responsible for the strategic development of responses to crisis situations. The trainee group

is usually composed of the representatives of the “command team”, i.e. the core agencies involved in the strategic Crisis Management (e.g., police, local authority, fire brigade, ambulance); sometimes, other trainees can come from other utility companies (e.g. electricity, road transportation, environmental agency).

In our KB, therefore, we modelled the basic knowledge about those Gold Commanders by creating classes to represent the different trainees typologies. Those classes are “instantiated” per each training session, by adding the individual trainees to the KB. This lets the system record the training of each participant in relation to his/her role in the simulation; this knowledge is very precious for both the debriefing phase – when the trainer summarizes the performance results of each trainee (see also below) – and for a general analysis and mining of the achieved objectives and learning needs of the different agencies.

The initial version of the user modelling is part of the Pandora Ontology<sup>7</sup>.

## 4.2 Training Simulation Modelling

The core module of the simulation learning system is the Simulation Planning (cf. Section 2.2). Our KB therefore must be able to manage the knowledge required for the planning, in terms of the basic entities used by AI Planning Applications based on Timeline Representations.

In literature, several attempts tried to formalize the semantics of planners [18, 19]. However, those approaches, on the one hand, tried to specify a generic planning ontology and, on the other hand, were specifically tailored to some application domains.

Building on their experience, we decided to make our own formalization to encompass the family of techniques known under the name of Timeline-based Planning and Scheduling. In fact, current AI planning literature shows that timeline-based planning can be an effective alternative to classical planning for complex domains which require the use of both temporal reasoning and scheduling features [10]. Moreover, our modelling aims to become the foundation for the investigation on the interplay between Semantic Web Technologies and Planning and Scheduling research [12]; Semantic Web knowledge bases, in fact, can represent a good alternative to the current domain modelling in the planning area, which encompasses a multitude of custom and not interoperable languages.

Our modelling is formalized in a Timeline-based Planning Ontology<sup>8</sup>. As in classical Control Theory, the planning problem is modelled by identifying a set of relevant features (called *components*) which are the primitive entities for knowledge modelling. Components represent logical or physical subsystems whose properties may vary in time; in the simulation learning, components are either trainees behavioural traits or learning scenario variables. Their *temporal evolutions* is controlled by the planner to obtain a desired *behaviour*. Therefore, our ontology includes a set of *time functions* that describe the evolution over temporal intervals. The evolution is modelled by *events* happening on modelled

<sup>7</sup> Cf. <http://swa.cefriel.it/ontologies/pandora>.

<sup>8</sup> Cf. <http://swa.cefriel.it/ontologies/tplanning>.

components. To this end, a set of *planning rules* (or synchronizations) specifies what events can be triggered to modify these evolutions. The task of the Simulation Planner is to find a sequence of events that brings the system entities into a desired final state.

The core concept of the Timeline-based Planning Ontology is therefore the planning rule: each rule puts in relation a “reference” event – which is the potential *cause* of some phenomena in the simulation – with a “target” event – which is the possible *consequence* –, under a set of conditions called *rule relations*. We modelled such conditions as SPARQL FILTER or LET clauses<sup>9</sup>; therefore, we reused the modelling of such clauses and functions included in the SPIN Modeling Vocabulary [20] and extended it with regards to temporal conditions.

At learning *design time* – i.e. prior to the simulation sessions –, the trainer has to model the possible training scenarios, by instantiating in the KB the ontology concepts, in particular the planning rules and the related events. The choice of Linked Data and Semantic Web technologies for our modelling is not only useful for reusing and exploiting pre-existing knowledge. In this case, we can also exploit the semantics of such ontology for the *consistency checking* of the simulation scenarios: by automatic means, we can check if all the planning rules are satisfiable, if they represent possible “states” of the world simulated during the sessions, if all the events can happen under opportune conditions, and so on.

At *run-time* – i.e. during the simulation learning sessions –, all the events and decisions taken by the trainees during their learning are recorded in the KB. The KB is therefore used by the Simulation Planner to create and update the simulation plan. SPARQL-based querying is used to perform the knowledge retrieval required in this step: based on the actual recorded events, only the admissible planning rules are returned to let the planner decide what events to trigger.

After the learning session, at *debriefing time*, the recording of trainees’ behaviour and decision-taking is exploited to summarize the session progress. Also in this case, SPARQL-based querying on the KB is exploited to retrieve all the events and situations that involved each trainee; this knowledge is immediately at disposal of the trainer to produce a debriefing report for each participant and can be used to highlight personal performance, achieved training goals and attention points for improvement or further training.

### 4.3 Asset Modelling

The Learning Delivery module (cf. Figure 1) takes as input the simulation plan and “execute” it by sending the opportune stimuli to the trainee. To do this, it needs to recreate the actual simulation conditions, by pretending a near-real situation. For example, in the Crisis Management training scenario, the participants must be solicited by phone calls, mail, news, videos, etc. that give them

<sup>9</sup> The SPARQL LET clause is defined in some implementations, like the Jena Semantic Web Framework <http://openjena.org/>

updates on the evolution of the emergency. To this end, the Learning Delivery module manages two types of “learning objects” that are described in the KB.

The first type of simulation objects consists in *audio and video assets*, which give information to the trainees about what happens outside the simulation room. In the Pandora scenario, those assets are pre-canned recording of simulated video news or audio inputs – like phone calls from the crisis setting – which are used to put pressure on the trainees and, in the meantime, to give them further inputs on which they must base their decisions. To model such assets, it is possible to re-use existing learning objects modelling, such as [4, 21]. In the Pandora project, we are still in the process of selecting the most suitable modelling for our purpose.

There is a second type of stimuli for the simulation trainees. Since the sensing system records the “performance” of each participant also in terms of stress and anxiety, the simulation can be adapted to the specific conditions and deliver tailored inputs for the individual trainees. For example, if the purpose is to augment the pressure on a participant, the input could be made more dramatic. To this end, the Learning Delivery module makes use of *Non-Player Characters* (NPC): in games terminology, elements that act as a fictional agents and that are animated and controlled by the system. Those NPCs simulate additional actors from outside the learning environment and are used to deliver information to the trainees.

Our KB, therefore, includes also the modelling of NPC descriptions, in terms of their role in the simulation, their basic characteristics (e.g. gender, ethnicity, disability), their profiles (expertise, experience, emotional type, communication skills, etc.), their multimedia rendering mode (from the simplest text representation to fully rendered 3D avatar), etc. For this modelling, Linked Data are exploited for the reuse of pre-existing descriptions and Semantic Web technologies are leveraged to retrieve and select the most suitable NPC to simulate a desired stress or anxiety situation.

## 5 Towards Provenance Tracking

As detailed in the previous section, our Linked Knowledge Base is used to manage the knowledge required to produce simulation-based learning sessions. We think that Simulation Learning can be seen as a special case of the Open Provenance Model (OPM) [22]. The sessions are our main *process*, the trainees, as well as the simulated external characters, are our *agents* and the events and the decisions taken by the trainees are the *artifacts* of the learning sessions.

Our future investigation will focus on the definition of the suitable OPM Profiles for Simulation Learning systems; specifically, we aim at mapping our Timeline-based Planning Ontology to the Open Provenance Model Vocabulary Specification [23]. While this is still work in progress, hereafter we give some hints on how we can build on the Open Provenance Model and why it is useful.

The provenance tracking in simulation learning can be done at two levels: at design time – when the learning scenarios are modelled in the KB with their

possible planning rules –, and after the learning sessions – when the results of the simulations are analysed.

At design time, provenance can be used to trace the cause-consequence chains between the possible simulation events. As explained in Section 4.2, planning rules are used to model the admissible transitions between events in the simulation; the *completion and inference rules* defined in OPM [22] can be exploited for the consistency checking of the simulation modelling. On the one hand, those rules can help in refining the modelling, by eliminating useless entities, combining eventual repetitions and introducing missing entities; on the other hand, OPM rules can help in examining the possible decision-trees (i.e., the possible alternative planning options) to identify unreachable states or decision bottlenecks.

After the learning sessions, the simulation records can be analysed to understand and synthesise the learning outcomes. Tracking the provenance of trainees’ decisions and mining the most popular causal chains across several sessions delivery can be of great help for identifying learning needs, common behaviours (as well as common trainees’ mistakes), wide-spread procedures, etc. This information can become of considerable importance: on the one hand, to improve the learning simulations and better address learners requirements and, on the other hand, to better study and interpret learning outcomes for individual participants or for entire classes of trainees.

## 6 Conclusions

In this paper, we presented our approach and experience in building a Linked Knowledge Base to support Simulation Learning systems. We introduced the general architecture of such a system together with a concrete scenario in Crisis Management training; we illustrated the benefits of the use of Linked Data and Semantic Web technologies and we summarised our modelling choices. We also suggested the introduction of provenance tracking, to further enrich and better analyse the contents of a Knowledge Base for Simulation Learning.

Our approach is being integrated in the Pandora Environment, which, in the second half of 2011, will be tested at the UK Emergency Planning College in their “Emergency Response and Recovery” training courses.

## Acknowledgments

This research is partially funded by the EU PANDORA project (FP7-ICT-2007-1-225387). We would like to thank the project partner for their collaboration.

## References

1. Stojanovic, L., Staab, S., Studer, R.: elearning based on the semantic web. In: In WebNet2001 - World Conference on the WWW and Internet. (2001) 23–27

2. Tiropanis, T., Davis, H.C., Millard, D.E., Weal, M.J.: Semantic technologies for learning and teaching in the web 2.0 era. *IEEE Intelligent Systems* **24**(6) (2009) 49–53
3. Learning Technology Standards Committee of the IEEE: Draft standard for learning technology - learning object metadata. Technical report, IEEE Standards Department, New York (2002)
4. Brase, J., Nejdil, W.: Ontologies and Metadata for eLearning. In Staab, S., Studer, R., eds.: *Handbook on Ontologies*. International Handbooks on Information Systems. Springer (2004) 555–574
5. Aldrich, C.: *Simulations and the Future of Learning: An Innovative (and Perhaps Revolutionary) Approach to e-Learning*. Pfeiffer (2003)
6. Caird-Daley, A., Harris, D., Bessell, K., Lowe, M.: *Training Decision Making using Serious Games*. Technical report, Human Factors Integration Defence Technology Centre (2007)
7. Sniezek, J., Wilkins, D., Wadlington, P., Baumann, M.: Training for Crisis Decision-Making: Psychological Issues and Computer-Based Solutions. *Journal of Management Information Systems* **18**(4) (2002) 147–168
8. Stern, E., Sundelius, B.: Crisis Management Europe: An Integrated Regional Research and Training Program. *International Studies Perspective* **3**(1) (2002) 71–88
9. Lehto, M., Nah, F.: Decision-making Models and Decision Support. In: *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, Inc., NY (2006)
10. Cesta, A., Fratini, S.: The timeline representation framework as a planning and scheduling software development environment. In: *27th Workshop of the UK Planning and Scheduling SIG*. (2008)
11. Cesta, A., Cortellessa, G., Fratini, S., Oddi, A.: Developing an end-to-end planning application from a timeline representation framework. In: *21st Applications of Artificial Intelligence Conference*. (2009)
12. Celino, I., Dell’Aglia, D., De Benedictis, R., Grilli, S., Cesta, A.: Ontologies, rules and linked data to support crisis managers training. *IEEE Learning Technology Newsletter, Special Issue "Semantic Web Technologies for Technology Enhanced Learning"* **13** (2011) Issue 1
13. Heath, T., Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*. 1st edn. Volume 1 of *Synthesis Lectures on the Semantic Web: Theory and Technology*. Morgan & Claypool (2011)
14. Berners-Lee, T.: *Linked Data – W3C Design Issues, Architectural and philosophical points*, Online at <http://www.w3.org/DesignIssues/LinkedData.html>. (2006)
15. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia – a crystallization point for the web of data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* **7** (2009) 154–165
16. Bizer, C., Heath, T., Berners-Lee, T.: *Linked Data – The Story So Far*. *International Journal on Semantic Web and Information Systems* **5** (2009) 1–22
17. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named graphs, provenance and trust. In: *WWW ’05: Proceedings of the 14th international conference on World Wide Web*, ACM (2005) 613–622
18. Rajpathak, D., Motta, E.: An ontological formalization of the planning task. In: *International Conference on Formal Ontology in Information Systems (FOIS’04)*. (2004) 305–316
19. Gil, Y., Blythe, J.: Planet: A sharable and reusable ontology for representing plans. In: *the AAI - Workshop on Representational Issues for Real-World Planning Systems*. (2000) 28–33

20. Knublauch, H.: SPIN Modeling Vocabulary. <http://spinrdf.org/spin.html> (October 20, 2009)
21. Meta-information for Multimedia eLearning. In: Computer Science in Perspective. (2003)
22. Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., den Bussche, J.V.: The Open Provenance Model core specification (v1.1). Future Generation Computer Systems (2010)
23. Zhao, J.: Open Provenance Model Vocabulary Specification. Online at <http://purl.org/net/opmv/ns> (2010)

# Connecting medical educational resources to the Linked Data cloud: the mEducator RDF Schema, store and API

Evangelia Mitsopoulou<sup>1</sup>, Davide Taibi<sup>2</sup>, Daniela Giordano<sup>3</sup>, Stefan Dietze<sup>4</sup>, Hong Qing Yu<sup>4</sup>, Panagiotis Bamidis<sup>5</sup>, Charalampos Bratsas<sup>5</sup>, Luke Woodham<sup>1</sup>

<sup>1</sup> St George's University of London, Center for Medical and Healthcare Education, Cranmer Terrace, SW17 0RE London, United Kingdom  
{emitsopo, lwoodham}@sgul.ac.uk

<sup>2</sup> Consiglio Nazionale delle Ricerche, Istituto per le Tecnologie Didattiche, Via Ugo La Malfa 153, 90146, Palermo, Italy  
davide.taibi@itd.cnr.it

<sup>3</sup> Università di Catania, Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Viale A.Doria 6, 95125, Catania, Italy  
dgiordan@dieei.unict.it

<sup>4</sup> The Open University, Knowledge Media Institute, MK7 6AA, Milton Keynes, UK  
{s.dietze, h.q.yu}@open.ac.uk

<sup>5</sup> Aristotle University of Thessaloniki, PO Box 323, 54124, Thessaloniki, Greece  
{bamidis, mpampis}@med.auth.gr

**Abstract.** The existence of non-interoperable metadata schemas and limited use of shared vocabularies means that retrieving and processing educational resources across the Web represents a challenge. The emerging Linked Data paradigm has provided the tools and methods to share and expose metadata in a more unified and well interlinked manner, permitting both humans and machines to process Web data. The availability of vast amounts of RDF-based Linked Data is offering a worthwhile alternative to the isolated and heterogeneous data silos which previously dominated the Web. In the EC-funded project mEducator a standardised approach is proposed to describing and exposing medical educational resources. In this paper are described firstly the design considerations, and conceptual model upon which mEducator's metadata scheme approach was based. Afterwards, follows the description of the serialization of the scheme in RDF/XML, and in turn follows an example showing how medical educational resources are exposed on the Web using an RDF endpoint, and discuss the potential advantages of this approach.

**Keywords:** Metadata, Linked Data, RDF vocabularies, medical learning resources, mEducator, Conceptual Model, RDF Store

## 1 Introduction

Medical and Healthcare Education has been modernised lately by means of Information and Communication Technologies (ICT). Many European academic institutions use their own Learning Content Management Systems (LCMS) in order to deliver educational resources in their curriculum. These resources can have different formats, such as images, text, videos, podcasts, virtual patients, serious games about health, diagnostic tests such as electrocardiograms (ECGs) etc. Currently, it is difficult to exchange resources between different platforms due to lack of standardised content sharing mechanisms. An attempt to solve the above problem is being made by the mEducator Best Practise Network (BPN) [1], funded by the EC in the eContentplus programme, which aims to analyze the use of existing standards and reference models in the e-learning field to discover, retrieve, share and re-use medical educational resources. The goal of the current paper is to present this effort and to show how this effort could be further examined, evaluated and reused by a wider community.

The diffusion of learning experiences based on the Web has increased over the years, and the application of Semantic Web technologies to e-learning has been widely studied by researchers. In 2001 Tim Berners Lee presented the Semantic Web as a Web interpretable by machines [2]. Many research studies propose the application of Semantic Web technologies in different e-learning situations. In [3] and [4] ontologies are used to describe the contents of learning resources. The importance of the use of the Semantic Web approach to learning object metadata representation is discussed in [5], [6]. The use of Semantic Web services architecture for e-learning was proposed by [7] and [8]. The modeling of an e-learning environment by means of a multi agents system was studied by [9]. Virtual learning environments based on Semantic Web technologies were discussed in [10]. Recent developments in the social Semantic Web have focused on informal learning experiences [11], as well as the diffusion of Linked Data, which has opened up new opportunities for improving e-learning experiences.

Linked data can be useful to support interoperability, accessibility and reusability of learning resources. For this reason, in the mEducator project, a standardised metadata description scheme has been proposed, and an RDF triple-store repository compliant with Linked Data requirements has been created to store educational resources. By using a standardised format to describe resources a basis for interoperability and a framework for discovery will be provided [12].

The rest of this paper is structured as follows: Section 2 describes the development and structure of the conceptual model and the schema, along with its RDF serialization. The mEducator implementation and its application of Linked Data principles, including the RDF instance file, is discussed in Section 3. Details of the REST API and rdf repository are provided, as well as a description of the approach taken by the mEducator RDF schema and RDF endpoint to interlink educational resources with data sets as part of the Linked Data cloud.

## **2 A general schema for educational resources**

### **2.1 In General**

The first step when developing a metadata scheme is to specify what kind of metadata is needed in order to solve a problem in a specific domain of interest. In this case, the domain of interest is medical and healthcare education, and the problem is how to describe different types of medical resources in order to facilitate content sharing and re-use across institutions. In particular the description must allow resources to be efficiently repurposed, a process that involves all or part of a resource being modified and re-used to suit it to an audience or purpose different than was originally intended. Repurposing is a term first used in [13] to refer to the adaptation of existing online learning resources to different educational levels. Later in [14] the term repurposing was used in a wider sense to refer to the conversion of a Virtual Patient (VP) created for one purpose into a VP fit for another purpose i.e. a different subject, discipline or healthcare profession. Virtual Patient as defined in [13] is a specific type of medical educational resource that refers to an interactive computer simulation of real-life clinical scenarios for the purpose of medical training, education, or assessment.

In order to best determine the metadata requirements for the project a conceptual model for the description of medical educational resources was formulated. The needs of mEducator's target group were identified by conducting a survey of mEducator's partners. The results of the survey indicated the nature of current perceptions about the concept of educational content, educational content sharing and repurposing. Based on the survey's results and following an investigation of existing e-learning standards, an initial framework for the metadata concepts was set that allowed the above educational material to be annotated semantically and described in a machine readable format. A conceptual model of the schema was then developed based on this initial framework. Taking into consideration the fact that the intended users are not professional indexers and would tend to fill in the minimum required information, the generally agreed approach was to keep the schema as lean as possible and to maximise the re-use of existing e-learning standards.

### **2.2 Research on existing Standards**

Existing standards for describing learning objects, in particular IEEE Learning Object Metadata (LOM) [15] and its extension Healthcare LOM [16] - designed by MedBiquitous to provide for the specific demands of medical and healthcare education - were considered as the starting point for the development of the schema. This starting point was also based on previous research done by [17] whose input to the current work was fundamental. To continue, the XML binding of IEEE LOM was initially reused by mEducator in order to implement the scheme in XML. However, as the need to ensure that the metadata was compliant with the principles of Linked Data became clear, RDF [18] was identified as the most appropriate framework to use. The notion of linking between resources and ontologies through RDF URIs offered a way of benefitting from the many advantages of Linked Data.

Work towards an implementation of IEEE LOM in RDF has been started by [19], but this work was discontinued around 2004. For this reason Dublin Core [20] was considered and Dublin Core's expression in RDF/XML [21], along with a further investigation of existing RDF Vocabularies, made up the final phase of the process of structuring mEducator's conceptual model. Examples of these RDF Vocabularies includes SKOS[22], which was used to provide a standard way of representing controlled vocabularies in RDF defined within mEducator, and FOAF [23], which was used to describe people in RDF.

Those properties of Dublin Core that met mEducator's conceptual model needs were reused in the scheme as super properties. They were not used in the same manner as they are defined in the Dublin Core element namespace [24], because this approach was considered to be too general. In contrast, the approach of the Dublin Core Terms Namespace [25], in which all properties are sub-properties of the ones defined in [24], was adopted i.e. `mdc:creator`, `mdc:identifier`, `mdc:title`, `mdc:rights`, `mdc:subject`, `mdc:language` are sub-properties of `dc:creator`, `dc:identifier` etc respectively, where `mdc` is the prefix for mEducator Namespace [26]. From the FOAF specification, some basic properties were reused i.e. those that are used to describe a person: `foaf:name`, `foaf:Organisation`, `foaf:mbox_sha1sum`.

Where existing standards did not meet mEducator requirements, additional properties were proposed. For example, a property was needed in the schema to describe to which organisation a person belongs to. The lack of such a property in FOAF necessitated the proposal of a new property within mEducator's namespace i.e. `mdc:memberOf`.

### 2.3 Conceptual Model

The conceptual model for the Learning Resource that was developed to be the basis of the mEducator schema is intended to:

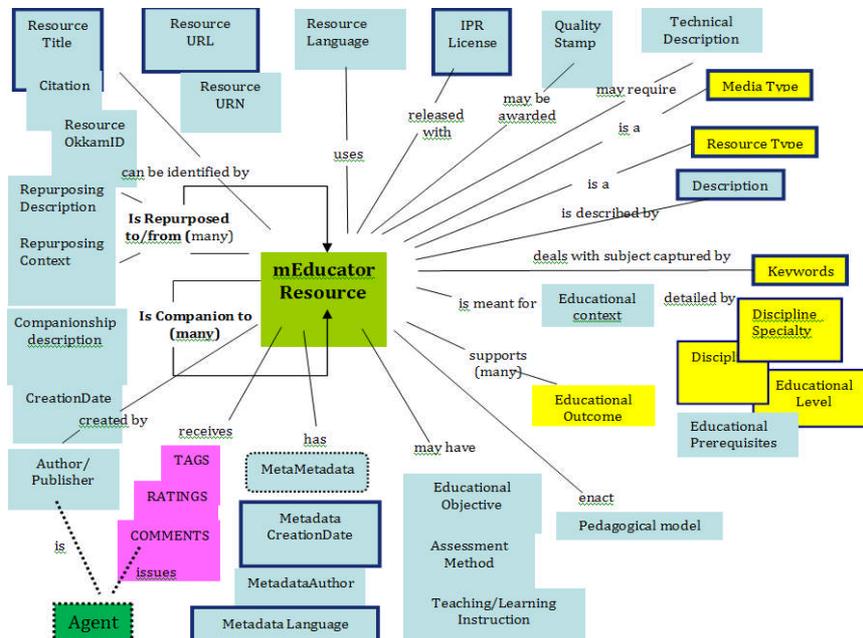
- 1) Capture the semantics of the relationships (if any) among the metadata fields and the Learning Resource
- 2) Make evident the rationale of the design decisions in the schema
- 3) Identify potentially overlapping semantics in the metadata fields, and remove them in order to support the modular construction of metadata
- 4) Clearly locate the appropriate place for the development of mEducator-specific controlled vocabularies and taxonomies.

As depicted below (Figure 1), the basic metadata of a medical learning resource consists of: its title, its identifier (i.e. url), the language in which it is created, the language of its metadata, the date of resource creation, the date of metadata creation the resource author, the author of the metadata, a brief description, a technical description, its IPR license, its citation and any further information that certifies its quality. In addition, further metadata has been proposed that provides pedagogical value such as: educational objectives, assessment methods, teaching instructions, educational level, educational prerequisites and educational outcomes. Further metadata will be used to provide keywords describing the resource, and to describe the discipline and discipline specialty to which a resource relates.

Another category of metadata fields included in the proposed scheme capture the repurposing history of the resource. The definition of the repurposing process is given in [14] and the metadata provided may include the title and identifier of the parent resource from which the current resource has been created, any further resource(s) that the current resource has been repurposed to, the type of repurposing that took place (i.e. change of language, change of target etc.) and a brief description about the repurposing process that took place.

There are different types of medical educational resources, and this information should be captured and described. Within mEducator two controlled vocabularies have been created to enable the categorisation of the resources based on two different criteria: the informational quality (including the nature of the information provided, how it is organized, presented or collected) and the media format of the resource (or the media formats included in the resource if this is complex and includes multiple media types). These use the resourceType metadata field and the mediaType metadata field.

Metadata fields for describing accompanied resources are also included. The rationale behind accompanied resources is that sometimes an educational resource is accompanied by other resources such as pdfs, videos, images etc in order it to be fully understood and better taught. In order to describe this information the isAccompaniedResource metadata field was proposed.



**Fig. 1.** The conceptual model used as a basis for developing the mEducator metadata application profile, summarising the metadata information proposed to describe a medical educational resource. Mandatory fields are in bold, fields using controlled vocabularies are in yellow, fields capturing user-generated information are in pink.

In order to make the schema as simple and flexible as possible, the majority of the fields are optional. Those fields which are mandatory include the title of the resource, the unique identifier (i.e. URL, URN, OkkamID etc.), the IPR license, the language of its content, the language of its metadata, a description and relevant keywords. In addition to defining a metadata field as mandatory or optional, the multiplicity for each field has been specified. For instance, each educational resource should have one title only and its metadata should be written in a single language. Other fields are permitted to have multiple instances.

## 2.4 RDF Serialisation

Linked Open Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets [27] (Bizer, 2009). At the heart of the Linked Data approach are its core technologies, such as RDF for data representation and SPARQL as a standardised means to query RDF stores via HTTP.

The first step towards implementing mEducator's metadata schema in RDF was to define its RDF Model [28]. This RDF model comes out of the conceptual model described in the previous section. More specifically, all the metadata fields included

in the conceptual model become properties in the RDF Model. Each property should have a value i.e. the property title can have the value “cardiology”. Generally speaking, the type of value a property can take varies from a plain string, (datatype properties) to a more complex object composed of a group of properties and values (object properties). In order to describe more complicated values or to describe a value whose type is not described by an existing class (e.g. String) it is necessary to define classes. Thus the need to define a new class for a property is dependent on the type of the value that property can take.

Another component that was defined within the RDF Model is controlled vocabularies. A controlled vocabulary is a vocabulary consisting of a “prescribed list of terms or headings, each one having an assigned meaning, and provide a way to organize knowledge for subsequent retrieval" [29]. Controlled vocabularies are, in other words, another type of value that a property can take.

Below are the classes, properties and controlled vocabularies that were defined as part of the RDF Model for mEducator.

Classes
Resource, RepurposingResource, IPRTType, RepurposingContext, EducationalLevel, Discipline, DisciplineSpeciality, Subject, ExternalTerm

Properties
creator, identifier, repurposingIdentifier, description, language, rights, created, metadataCreated, memberOf, profileURI, assessmentMethods, educationalOutcomes, educationalObjectives, teachingLearningInstructions, citation, discipline, disciplineSpeciality, externalTerm, externalSource, conceptID, educationalPrerequisites, educationalContext, educationalLevel, title, repurposingTitle, isAccompaniedBy, toRepurposingContext, fromRepurposingContext, isRepurposedTo, isRepurposedFrom, resourceType, mediaType, technicalDescription, metadataCreator, metadataLanguage, quality, repurposingDescription

Controlled Vocabularies
ResourceType, MediaType, EducationalOutcome, IPRTType, EducationalLevel, RepurposingContext

### **3 Deployment: an RDF store and API compliant with the mEducator schema**

The Linked Data cloud is based on publishing structured data on the Web using RDF, SPARQL and interconnected URIs to reference data entities. That has led to a large

amount of well-connected and interlinked data sets and has emerged as an extremely useful approach for exposing data.

In the mEducator project, a sharing mechanism which exposes educational resource descriptions as rich and well-interlinked Linked Data has been proposed. This approach requires the implementation and application of the mEducator metadata scheme and the provision of an API that enables access to a repository containing this data. This is part of a more general semantic solution based on Linked Services [37] that allows distributed searches and queries across the Web. The detailed description of this solution is outside the scope of the present paper (a more complete description can be found in [30]). While external resource metadata is discovered and processed on the fly, an RDF store complying with the schema proposed in this paper is used to expose retrieved metadata as LOD and enrich it with links to structured LOD data sets. The store's API could be directly accessed by a 3<sup>rd</sup> party application allowing users to insert new resources or search resources using their metadata, likewise software agents can collect resource metadata using web service endpoint of compliant LCMS, and use the API to store them in the repository.

The solution proposed respects the four principles of Linked Data [31]:

1. The first principle requires the use of URI as a namespace to identify resources
2. The second principle proposes the use of HTTP URI in order to look for information about the resources.
3. The third principle recommends to provide useful information, using RDF when someone looks up a URI.
4. The fourth principle consists in creating links to other resources on the Linked Data cloud

### **3.1 RDF Schema Level**

As shown in the previous section, mEducator's metadata description scheme uses an RDF model to represent its data. Generally speaking, the usage of URIs and RDF in the schema demonstrates the implementation of the basic principles of Linked Data. PURLs [32] along with targeted locations were used in order to develop mEducator's namespaces, and the domain <http://www.purl.org/meducator> has been registered for mEducator. In turn, further subdomains have been registered to define namespaces for both the schema and the controlled vocabularies defined within mEducator. The namespace of the schema is <http://purl.org/meducator/ns/> and the namespaces for the controlled vocabularies are named similarly.

RDF entities – classes and instances - are described using URIs, and most specifically, via HTTP URIs, conforming to the first two design principles of Linked Data. By referencing classes using HTTP URIs and also providing further information regarding their definition, their meaning can be understood even in the event that they become dereferenced.

The definition of properties is done in a similar fashion, with HTTP URIs and additional information regarding their meaning also being provided. In addition to this, property definitions include links to other properties or to classes of other standards. This is one of the strengths of RDF and Linked Data. By connecting

existing standards, and linking to other widely used standards such as Dublin Core the interoperability of the mEducator scheme is maintained.

The use of controlled vocabularies also complies with Linked Data principles, and is implemented in RDF/XML. Some vocabularies were implemented using SKOS i.e. `mediaType`, `ResourceType` and `educationalOutcomes`, while others were implemented using just RDF i.e. `IPR Type` and `Repurposing Context`. In both cases, HTTP URIs have been used to refer to each concept of a controlled vocabulary, e.g. <http://purl.org/meducator/licenses#Attribution> for an IPR license and <http://purl.org/meducator/mediaType#image> for a media type. All of them include the `<rdfs:label>` element that provides useful information in the event that they become dereferenced.

### 3.2 RDF Instance Level

The incorporation of Linked Data principles at the RDF instance level is similar to that at the schema level. Firstly, each RDF instance is described by an HTTP URI. Secondly, all the properties whose value will be taken from a controlled vocabulary use HTTP URIs. For example, the `ResourceType` property may have the value <http://purl.org/meducator/resourceType#virtualPatient>.

A critical advantage of Linked Data is the interlinking of resources by considering available datasets (bioportal ontologies, mEducator controlled vocabularies) and reusing schemas (FOAF, Dublin Core). In mEducator's case, this is done via the `owl:sameAs` [33] property that is used along with `Subject`, `Discipline`, `DisciplineSpeciality`. For example, the `Subject` property is used to provide keywords that describe the medical resource. These keywords are identified by the user but are defined in an external source. Using the `owl:sameAs` property links to other URIs are established where the definition of the same keyword exists in other taxonomies, enabling other resources to be discovered. In this way mEducator's RDF Schema applies the fourth design principle of Linked Data, linking to other, related URIs to improve the chances of discovering other related information on the Web.

### 3.3 The mEducator API

The mEducator instances described above are stored in an RDF repository that provides a Representational State Transfer (REST) API, enabling access to the repository by third parties with no knowledge of the underlying schema used to store the data. During the inserting phase the mEducator REST API validates the RDF instances to be inserted against the mEducator RDF schema, and creates a universal unique identifier (UUID) for the mEducator resources inserted in the store according to the mEducator HTTP URI<sup>1</sup>.

The mEducator educational resource store provides simplified search mechanisms in order to retrieve information related to the resources in RDF format. The API aims to simplify the search for resources by using properties and features defined in the

---

<sup>1</sup> <http://purl.org/meducator/resources/UUID>

mEducator RDF schema. The general purpose SPARQL interface provided by the Sesame triple store is replaced with a simpler endpoint appropriately designed to handle mEducator resources. In order to create links with other resources on the Linked Data cloud the mEducator resources have been connected to the BioPortal<sup>2</sup> vocabularies with regard to disciplines and subjects. A detailed description of these search methods will be discussed in Section 3.4.

```

<rdf:Description rdf:about="http://purl.org/meducator/resources/45b5ab54-4a09-4100-88bb-93057f256da2">
  <rdf:type rdf:resource="http://www.purl.org/meducator/ns/Resource" />
  <rdfs:seeAlso rdf:resource="http://metamorphosis.med.duth.gr/uid#8332" />
  <mdc:identifier>http://www.med.helsinki.fi/tuke/meducator/Kansio_1004/Patient%201004.pdf</mdc:identifier>
  <mdc:title><![CDATA[ ECG Patient case 1004 limb and chest leads & thrombolysis ]]></mdc:title>
  <mdc:rights rdf:resource="http://purl.org/meducator/licenses/Attribution-Non-Commercial-Share-Alike" />
  <mdc:creator>
    <foaf:Person>
      <foaf:name>Timo Kuusi</foaf:name>
    </foaf:Person>
  </mdc:creator>
  <mdc:metadataCreator>
    <foaf:Person>
      <foaf:name>Jarkko Mylläri</foaf:name>
      <mdc:profileURI>http://metamorphosis.med.duth.gr/pg/profile/Jarkko?view=foaf </mdc:profileURI>
    </foaf:Person>
  </mdc:metadataCreator>
  <mdc:created>2010-04-19</mdc:created>
  <mdc:metadataCreated>2010-04-19</mdc:metadataCreated>
  <mdc:subject>
    <mdc:Subject rdf:about="http://meducator.open.ac.uk/ontology/SNOMEDCT/51308006">
      <rdfs:seeAlso>http://purl.bioontology.org/ontology/SNOMEDCT/51308001</rdfs:seeAlso>
      <rdfs:label>Thrombolysis</rdfs:label>
      <mdc:externalSource>SNOMED Clinical Terms</mdc:externalSource>
    </mdc:Subject>
  </mdc:subject>
  <mdc:resourceType rdf:resource="http://purl.org/meducator/resourceType/IMAGE"/>
  <mdc:discipline>
    <mdc:Discipline rdf:about="http://meducator.open.ac.uk/ontology/SNOMEDCT/51308001">
      <rdfs:seeAlso>http://purl.bioontology.org/ontology/SNOMEDCT/51308001</rdfs:seeAlso>
      <rdfs:label>Cardiology</rdfs:label>
      <mdc:externalSource>SNOMED Clinical Terms</mdc:externalSource>
    </mdc:Discipline>
  </mdc:discipline>
  <mdc:educationalContext>Teaching of ECG</mdc:educationalContext>
  <mdc:educationalObjectives>Reading and interpreting ECG</mdc:educationalObjectives>
</rdf:Description>

```

**Fig. 2.** Example of a mEducator resource description

Figure 2 shows an example of a mEducator resource, in which:

- the mEducator namespace (<http://www.purl.org/meducator/ns/>) is used to identify the type of the resource;
- subject and discipline are connected to the Bioportal ontology;
- educational context and objective are properties of the resource as defined in the mEducator schema;
- digital rights of the resource is connected to the Creative Commons;
- the FOAF ontology is used to reference creator and metadata creator.

<sup>2</sup> <http://bioportal.bioontology.org/>

### 3.4 The mEducator repository architecture

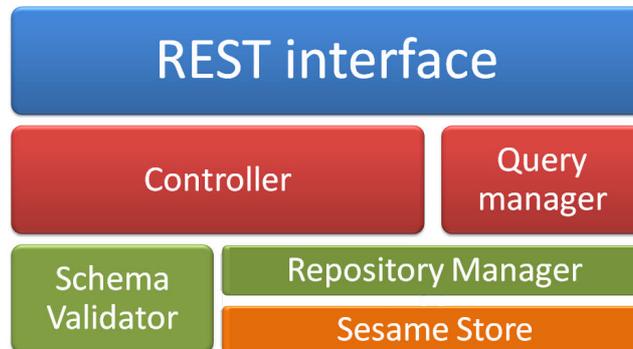
The mEducator repository stores medical learning resources in RDF format using a SESAME RDF triple store. In the project, a REST API has been created with the aim of making resources compliant with the four Linked Data principles and simplifying the repository interoperability with external environments (data stores, software applications, web agents, and so on). At present a social collaborative educational network [34], *Metamorphosis* based on the social environment Elgg, uses this API to provide a web user interface to the mEducator resources. However, as mentioned above, the API is available to any 3<sup>rd</sup> party applications to, for instance, query and retrieve resource metadata. Basically the API is an interface that external software agents (clients, web services, and so on) could use to handle mEducator resources. Moreover, future developments envisage the access to the API functionalities considering different types of authentication levels.

It is important to note that a SESAME triple store already offers a complete set of REST methods to work with the resources and the entities belonging to the store. However, these methods are designed to work with general RDF resources; for this reason it was necessary to develop functionalities specifically designed to handle and process mEducator RDF schema compliant data.

Taking into consideration the characteristics of mEducator resources, the API implements methods to insert the resources into the store, to retrieve the resources using different search criteria, and to update the description of a resource and to delete resources. The REST protocol uses the HTTP methods POST, GET, PUT and DELETE to implement the typical database CRUD (create, retrieve, update, delete) operations. The mEducator resources REST API is written in Java, specifically the project uses Jersey, the reference implementation of the JSR-311 [35], to create a RESTful endpoint for the mEducator repository.

Considering the software architecture the main part of the API consists of:

- Rest Interface: this module implements the endpoint for the services defined in the mEducator project. It manages the HTTP requests, defining the map between http URL and the Java method that implements the required behaviour.
- Controller: this module addresses the requests coming from the Rest Interface to the right handler. In particular, during the inserting phase, it uses the Schema Validator to check the validity of a new resource with respect to the mEducator schema, and for the searching functionalities it delegates the Query Manager module to address a query to the Sesame store.
- Schema Validator: this module uses the Jena Validity Report to perform validation of the RDF file representing a resource against the mEducator RDF schema. If the model is not valid, the resource is not inserted into the store and the conflicts that have caused the invalid condition is reported.
- Repository Manager: this module manages the connection with the Sesame store.
- Query Manager: this part of the architecture is responsible for managing the search requests and obtaining the results from the Sesame store.



**Fig. 3.** Main architectural components of the mEducator resources REST API.

The following query types have been developed in this version of the API:

1. the `searchByPropertyValue`, this kind of search uses the couple property-value to perform a search on the resources stored in the Sesame store.
2. the `searchByKeyword` this query creates an RDF graph of the resources with a specified keyword in one of its values.
3. the `searchByID` this query uses the DESCRIBE statement to return an RDF graph related to the resource with a specified UUID
4. the `searchBySeeAlso` this query is used to search the repository for the resources with a specified `seeAlso` property.

In retrieving the resources from the SESAME store, particular attention was given to managing the enclosures and blank nodes of the resources. In SPARQL, the DESCRIBE queries return all the RDF triples of a resource but not the RDF triples of the blank nodes related to it. A similar problem arises from deleting resources using the DELETE statement, even though in this case only the main resource is deleted and not all the related blank nodes.

## 4 Conclusions

The work presented in this paper has been designed to facilitate the connection of the mEducator repository to the Linked Data cloud. [36] reports the basic steps needed to publish data as Linked Data on the Web; it is necessary to assign a URI to the objects described by the data and to create links to other data on the Web, as well as providing metadata about the published data.

All the work that has been done so far is in line with the approaches being taken by big initiatives such as the Dublin Core Metadata Initiative. Taking this approach ensures alignment with current versions of controlled vocabularies, allowing for a rich choice of terminology (from external sources) to be made available as suggestions both during the search and the metadata filling process. The investigation of existing e-learning standards and their adoption, along with the reuse of existing RDF vocabularies within the schema, ensured that the work of mEducator is as interoperable as possible. This interoperability allows for connections to be made with other communities using strict Healthcare LOM or other schemas.

The scheme will initially be used by mEducator partners to describe their medical educational resources in two sharing mechanisms developed as part of the mEducator Best Practice Network and recommendations, and best practices will be produced. At present, the mEducator RDF store REST API assigns a valid URI to the new resources inserted into the repository, and metadata about resources are also provided using search methods. Interlinking of mEducator resources to the Linked Data cloud is currently being supported by enriching data sets with references to biomedical data available via the BioPortal vocabularies. That way, for instance, free text descriptions of subjects, keywords or disciplines are (semi-)automatically extended with references to related terms in established vocabularies.

While recent work has established the infrastructure to expose medical educational resources in mEducator, ongoing and future work is dealing with populating the data store, enhancing the enrichment process and integrating the data/store via its API into third-party educational applications. The intention is also to permit third parties to use the schema for their own resource description, while the submission of the schema to standardisation groups would also be of importance.

You can find further information about the schema, best practice guidelines, related tools and the created API at the mEducator project website <http://www.meducator.net>.

## Acknowledgements

This work was supported by the project mEducator (Multi-type Content Sharing and Repurposing in Medical Education), funded by the eContentplus Programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable (Grant ECP 2008 EDU 418006)

## References

1. mEducator | Multi-Type Content Repurposing and Sharing in Medical Education <http://www.meducator.net/>
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American*, 284(5), 34-43, (2001).
3. Gasevic, D., Jovanovic, J., Devedzic, V.: Ontology-based Annotation of Learning Object Content, *Interactive Learning Environments*, 15(1), 1-26, (2007).
4. Gašević, D., Jovanović, J., Devedžić, V., Bošković, M.: Ontologies for Reusing Learning Object Content, In Proc. of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), pp.944-945, (2005).
5. Brase, J., Nejd, W.: Ontologies and Metadata for eLearning, In S. Staab & R. Studer (Eds.) *Handbook on Ontologies*, Springer-Verlag, pp. 555-574, (2004).
6. Nilsson, M., Palmér, M. and Naeve, A.: Semantic Web metadata for e-learning - some architectural guidelines, Proc. of the 11th World Wide Web Conference (WWW2002), Hawaii, USA, May 7-11, (2002).

7. Dietze, S., Gugliotta, A., Domingue, J.: A Semantic Web Service oriented Framework for adaptive Learning Environments. European Semantic Web Conference (ESWC), Innsbruck, Austria, (2007).
8. López Cobo, J.M., Sicilia, M.A., Arroyo, S.: Making Learning Objects useful with Semantic Web Services. In *Proceedings of the e-Challenges Conference*, (2006).
9. Dunkel, J., Bruns, R., Ossowski, S.: Semantic E-Learning Agents - Supporting E-Learning by Semantic Web and agent technologies, in: Seruca, I.; Cordeiro, J.; Hammoudi, S.; Filipe, J. (Eds.), *Enterprise Information Systems VI*, Springer Verlag, pp. 237-244, (2006).
10. Simic, G., Gasevic, D., Devedzic, V.: Semantic Web and intelligent learning management systems. *Proceedings of the 2nd International Workshop on Applications of Semantic Web Technologies for ELearning*. Macéió-Alagoas, Brazil, (2004).
11. Jovanović, J, Gašević, D, Torniai, C, Bateman, S., Hatala, M. : The Social Semantic Web in Intelligent Learning Environments: state of the art and future challenges, *Interactive Learning Environments*, 17(4), 273-309, (2009)
12. Starr, J., & Gastl, A.: IsCitedBy: A Metadata Scheme for DataCite D-Lib Magazine, 17, <http://www.dlib.org/dlib/january11/starr/01starr.html>
13. REHASH Project Final Report, Repurposing, <http://www.elu.sgu.ac.uk/rehash/index.htm>
14. eViP Electronic Virtual Patients <http://www.virtualpatients.eu/about/about-virtual-patients/faqs/>
15. IEEE Learning Technology Standards Committee (LTSC) — IEEE Learning Technologies Standards Committee 2011, <http://www.ieeeltsc.org:8080/Plone>
16. Healthcare LOM Overview | MedBiquitous Consortium 2011, [http://www.medbiq.org/working\\_groups/learning\\_objects/Healthcare\\_LOM\\_Overview.html](http://www.medbiq.org/working_groups/learning_objects/Healthcare_LOM_Overview.html)
17. Konstantinidis S, Kaldoudi E, Bamidis P, Enabling Content Sharing in Contemporary Medical Education: A Review of Technical Standards, *The Journal on Information Technology in Healthcare* 2009; 7(6): 363–375
18. RDF - Semantic Web Standards 2011, <http://www.w3.org/RDF/>
19. IEEE Learning Object Metadata RDF binding LOM, <http://kmr.nada.kth.se/static/ims/md-lomrdf.html>
20. Dublin Core Metadata Element Set, Version 1.1 2011 , <http://dublincore.org/documents/dces/>
21. Langford, D.: Internet ethics . Macmillan, Basingstoke (2000) , <http://dublincore.org/documents/dcmes-xml/>
22. SKOS Simple Knowledge Organization System - Home Page 2011 , <http://www.w3.org/2004/02/skos/>
23. The Friend of a Friend (FOAF) Project | FOAF Project 2011, <http://www.foaf-project.org/>
24. Dublin Core Metadata Element Set <http://dublincore.org/documents/dces/>
25. DCMI Metadata Terms <http://dublincore.org/documents/dcmi-terms/>
26. mEducator Namespace, <http://www.meducator.net/mdc/schema.rdf>
27. The Semantic Web: Scientific American 2011 Bizer 2009
28. RDF/XML Syntax Specification (Revised) 2011, <http://www.w3.org/TR/REC-rdf-syntax/>
29. Reitz, J. M.: Dictionary for library and information science . Libraries Unlimited, Westport, Conn. (2004), [http://edutechwiki.unige.ch/en/Controlled\\_vocabulary](http://edutechwiki.unige.ch/en/Controlled_vocabulary)
30. Yu, H.Q., Dietze, S. , Li, N., Pedrinaci, C., Taibi, D., Dovrolis, N., Stefanut, T., Kaldoudi, E., Domingue, J.: A Linked Data-driven & Service-oriented Architecture for Sharing Educational Resources; *Linked Learning 2011, 1st International Workshop on eLearning Approaches for the Linked Data Age*, Heraklion, Greece (2011)
31. Linked Data - Design Issues 2011, <http://www.w3.org/DesignIssues/LinkedData.html>
32. PURL Home Page 2011, <http://purl.oclc.org/docs/index.html>
33. OWL Web Ontology Language Reference 2011, <http://www.w3.org/TR/owl-ref/>

- 34.E. Kaldoudi, N. Dovrolis, S. Konstantinidis, P. Bamidis, "Social Networking for Learning Object Repurposing in Medical Education", *The Journal on Information Technology in Healthcare*, vol. 7(4), pp. 233–243, 2009
- 35.JSR-311: JAX-RS: The Java API for RESTful Web Services.
- 36.Bizer, C., Heath, T., Berners-Lee, T., : Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3), 1-22, (2009).
- 37.Dietze, S., Yu, H.Q., Pedrinaci, C., Liu, D. and Domingue, J. (2011) SmartLink: a Web-based editor and search environment for Linked Services, 8th Extended Semantic Web Conference (ESWC), Heraklion, Greece

# Educational Semantic Wikis in the Linked Data Age: the case of MSc Web Science Program at Aristotle University of Thessaloniki

Charalampos Bratsas<sup>1,2</sup>, Anastasia Dimou<sup>1</sup>, Georgios Alexiadis<sup>1</sup>, Despoina-Eleona Chrysou<sup>1</sup>, Konstantinos Kavargyris<sup>1</sup>, Ioannis Parapontis<sup>1</sup>, Panagiotis Bamidis<sup>1,2</sup>, Ioannis Antoniou<sup>1</sup>

<sup>1</sup>Web Science, Mathematical Department, Aristotle University of Thessaloniki, Greece

<sup>2</sup>Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Greece  
cbratsas@auth.gr, andimou@math.auth.gr, galexiad@math.auth.gr, chrysoue@math.auth.gr,  
kostaskav@math.auth.gr, paraponi@math.auth.gr, bamidis@med.auth.gr,  
iantonio@math.auth.gr

**Abstract.** Wikis are nowadays a mature technology and further well established as successful eLearning approaches that promote collaboration, fulfill the requirements of new trends in education and follow the theory of constructivism. Semantic Wikis on the other hand, are not yet thoroughly explored, but differentiate by offering an increased overall added value to the educational procedure and the course management. Their recent integration with the Linked Data cloud exhibits a potential to exceed their usual contribution and to render them into powerful eLearning tools as they expand their potentialities to the newly created educational LOD. Web Science Semantic Wiki constitutes a prime attempt to evaluate this potential and the benefits that Semantic Web and linked data bring in the field of education.

**Keywords:** *Wiki Engines, Semantic Wiki, e-learning, Linked Data*

## 1 Introduction

Semantic Web is the current state-of-the-art, whose unprecedented intension is to unlock the value of content; even though knowledge modeling is significantly difficult due to its complexity. Nowadays, this is mainly accomplished using the two prominent semantic technologies of ontologies and Linked Data [1]. Although their diffusion is not wide, the education field is well passed [2]. There are several educational projects which have already started to use Linked Data. Among them are meducator [3]<sup>1</sup>, LUCERO JISC Project<sup>2</sup>, dotAC<sup>3</sup>, RKB Explorer<sup>4</sup> and SemTech<sup>5</sup>.

---

<sup>1</sup> <http://www.meducator.net/>

<sup>2</sup> <http://lucero.open.ac.uk/>

<sup>3</sup> <http://www.dotac.info/>

<sup>4</sup> <http://www.rkbexplorer.com/explorer/>

<sup>5</sup> <http://www.semtech.ecs.soton.ac.uk/>

This paper attempts to highlight in a brief but informative way the integration of Advanced Semantic Wikis with the Semantic Web and Linked Data in the form of Semantic Learning Management Systems (SLMS). Web Science Semantic Wiki (WSSW) is used as a reference for this purpose and in a sense emphasizes on how Semantic Wikis might render to be a rich alternative eLearning approach in the Linked Data Age.

## 2 Web Science Semantic Wiki Presentation

WSSW was initially implemented as a project within the framework of Master Program “Web Science” of Aristotle University of Thessaloniki, Greece. It has nowadays evolved into a SLMS to support the adoption of Linked Data in the field of educational procedure following closely the tense towards Linked Data in Education as it is determined by the University of Southampton and the Open University in UK.

WSSW manages to conjunct successfully the two predominant semantic technologies. On the one hand, it uses the classical ontological structure to model its knowledge base and on the other hand it renders it available to be queried by the Semantic Web through its Triple Store connector. The purpose was to outreach the self-contained perspective of Semantic Wikis and try to expose its content to the Linked Data Cloud as a first step before the overall integration with it.

### 2.1 The platform

SMW+<sup>6</sup>-community option- was the platform’s core software base together with a set of accompanying extensions to the Semantic Mediawiki<sup>7</sup> powered by MediaWiki. As far as the extensions were concerned, the Halo Extension<sup>8</sup> was specifically selected to facilitate WSSW’s use for the more inexperienced users. It provided the WSSW with a semantic annotation bar and an autocompletion feature but its true value emerged in conjunction with the rest of the available extensions. Another essential extension used was the Semantic Gardening Extension<sup>9</sup>. It was installed to upload and maintain the WSSW ontologies and to continuously check for their consistency and referential integrity. It ensured that any anomalies, pages without annotations, and undefined entities would be detected and finally exported the ontology as it was modified after the users’ intervention in order to be reused. The extension of Semantic Form<sup>10</sup> was also broadly used to allow users to enter their data semantically annotated without the need to learn the Semantic Mediawiki’s syntax. Instead, the Semantic annotations were added indirectly through the simultaneous use of the corresponding templates. TreeView, (Treeview<sup>5</sup><sup>11</sup> and Javascript dtree) was yet another extension used that

---

<sup>6</sup> <http://wiki.ontoprise.de/>

<sup>7</sup> [http://semantic-mediawiki.org/wiki/Semantic\\_MediaWiki/](http://semantic-mediawiki.org/wiki/Semantic_MediaWiki/)

<sup>8</sup> [http://wiki.ontoprise.de/smwforum/index.php/Help:Halo\\_extension](http://wiki.ontoprise.de/smwforum/index.php/Help:Halo_extension)

<sup>9</sup> [http://smwforum.ontoprise.com/smwforum/index.php/Help:Semantic\\_Gardening\\_extension](http://smwforum.ontoprise.com/smwforum/index.php/Help:Semantic_Gardening_extension)

<sup>10</sup> [http://www.mediawiki.org/wiki/Extension:Semantic\\_Forms](http://www.mediawiki.org/wiki/Extension:Semantic_Forms)

<sup>11</sup> [http://smwforum.ontoprise.com/smwforum/index.php/Help:TreeView\\_extension](http://smwforum.ontoprise.com/smwforum/index.php/Help:TreeView_extension)

requested subtrees automatically and allowed the demonstration of WSSW's tree as an alternative way to navigate through the wiki. TripleStoreConnector<sup>12</sup> Basic Extension was installed to enhance WSSW's querying capabilities. It contained the Jena Triplestore connector and connected the the Triplestore with the SMW<sup>+</sup> and a tolerant SPARQL endpoint. It allowed queries of the semantic data directly from within the wiki or remotely via a SPARQL endpoint. This led into better search results since inverse, equal and transitive properties were reflected in the queries. A SMW User Manual Extension<sup>13</sup> enabled immediate access to help, context-sensitive selection of help articles and provided direct feedback in the online community, (submit bugs and ask questions). Semantic Results Format Extension<sup>14</sup> was installed because it bundled a number of result formats for the inline queries. Timelines, eventlines googlebars, googlepies and semantic maps and calendars were applied to present several queries in a more comprehensive and representative way.

Additionally, several other extensions were installed to support the better function of the WSSW and to succeed in better presentation of the content.

## 2.2 Structure Description

The approach followed to represent the structure of the actual master program was based on an ontology, which provided the essential building blocks to best suit the WSSW needs. It consisted of a primary, master's program ontology in combination with the imported FOAF<sup>15</sup> (Friend Of A Friend) one (Figure 1). The master's program key ontology was implemented initially on Protégé, it was then modified through the WSSW Ontology Browser and obtained its instances, dynamically, after the users' entries. On the other hand, the FOAF was imported to fulfill the specific descriptive needs of the master's manpower, (instructors and students). What follows is a brief description of the WSSW structure via presenting some of its elements.

The FOAF class "Person" is used to describe all the persons involved. The Master's ontology distinguishes them into the classes of "AcademicStaff" and "Student", and each one has its own subclasses. Apart from the class "Person", the Master's ontology also has the "Announcement" class that has as instances all the announcements of the Master and the class "Assignment" which includes all the assignments announced at the lectures. An assignment might be the same for all students or individual ones might be assigned to specific students each time.

In addition, the class "Course" includes all the information concerning each separate course and the "Degree" class is related to the class "Person" so as to define the higher educational level of a Person. The class "Lecture" describes the specific lectures of a course and the "Material" refers to all books, presentations and notes available on each lecture. The "News" class includes all news related to the Master and the class "UniversityOrganisation" is a subclass of FOAF class "Organization".

---

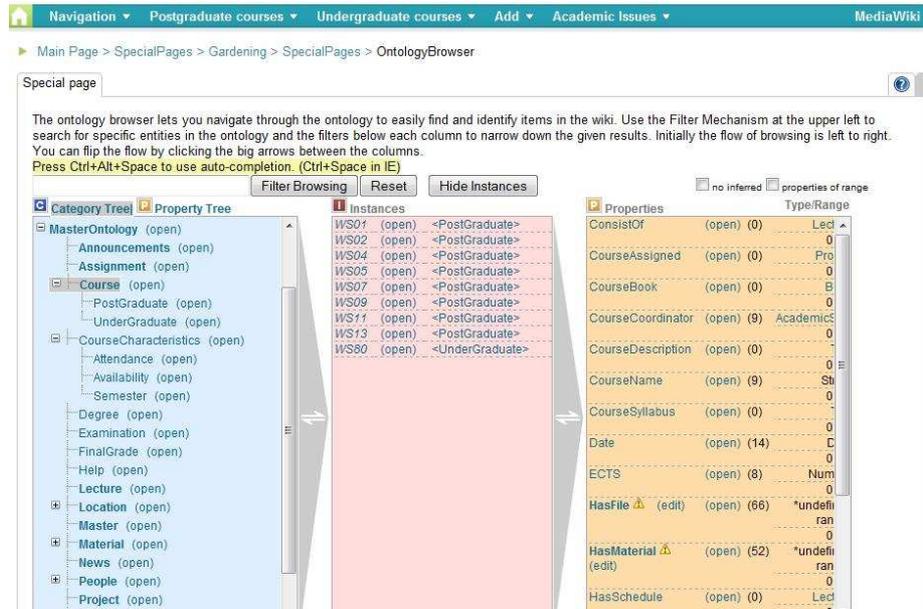
<sup>12</sup> [http://smwforum.ontoprise.com/smwforum/index.php/Help:TripleStoreConnector\\_Basic](http://smwforum.ontoprise.com/smwforum/index.php/Help:TripleStoreConnector_Basic)

<sup>13</sup> [http://smwforum.ontoprise.com/smwforum/index.php/Help:User\\_Manual\\_extension](http://smwforum.ontoprise.com/smwforum/index.php/Help:User_Manual_extension)

<sup>14</sup> [http://smwforum.ontoprise.com/smwforum/index.php/Help:Semantic\\_Result\\_Formats](http://smwforum.ontoprise.com/smwforum/index.php/Help:Semantic_Result_Formats)

<sup>15</sup> <http://xmlns.com/foaf/spec/>

Several forms were also implemented to allow easier user editing of the WSSW but to also ensure that the content would be automatically annotated and to maintain a consistent data structure. Each form creates a new instance for the class it defines. The fields that the users are asked to complete are the values of the properties of the instance. Each form is accompanied by its template which contributes further in the content presentation. Forms are used to create new and/or edit existing instances.



**Fig. 1.** The key ontology at the ontology browser

Each WSSW registered user owns a unique profile created via the “PeopleForm” and depending on the category he belongs to, he is assigned with different access rights in the available forms. For instance, only the Master’s and course Coordinators have access to the “NewsAnnouncementsForm” form and are allowed to add new entries at the associating sections. The Master’s Coordinator can also add new, or renew existing Courses via the “CourseForm” and can additionally assign a Thesis to an eligible student using the “ThesisForm”. On the other hand, the Course Coordinators may add new Lectures for their Courses, announce the Project topics and assign them to one or more Students, always through the associating forms. In general, each time a registered user attempts to enter a new entry, a table list, result of an inline query, informs him of any already existing entries and instances of that certain category and when he completes the values at the form’s fields, autocompletion feature based on concepts, facilitates his work.

A course example is presented at Figure 2. The course’s overall timetable is gathered at the course’s main page as it comes up combining its separate lectures properties. Its material is, also, collected via an inline semantic query which sum ups all separate materials added by individual lecturers. The course’s calendar follows as

a result of inline query to the lectures timetable (not illustrated at the Figure 2). Moreover, the users may browse the WSSW via the Semantic TreeView on the right of the context pane.

The screenshot displays the WSSW interface with two main tables and a Semantic TreeView on the right.

Lecture	Start Time	End Time	Lecture Title	Course/Instructor
WS11-01	26 February 2010 17:15:00	26 February 2010 21:00:00	Logic & programming in the Web	Melalides
WS11-02	5 March 2010 17:15:00	5 March 2010 21:00:00	Semantic Social Web and Web 3.0	Bratass
WS11-03	12 March 2010 17:15:00	12 March 2010 21:00:00	Ontology Development tutorial via protege	Bratass
WS11-04	19 March 2010 17:15:00	19 March 2010 21:00:00	Semantic Knowledge Representation and Reasoning	Bratass
WS11-05	26 March 2010 17:15:00	26 March 2010 21:00:00	A brief introduction to cryptography	Stamatou
WS11-06	16 April 2010 17:15:00	16 April 2010 21:00:00	Linked Data on the Web	Bratass
WS11-07	23 April 2010 17:15:00	23 April 2010 21:00:00	Ambient Intelligence, Sensor Nets	Nikolentzas
WS11-08	30 April 2010 17:15:00	30 April 2010 21:00:00	Future Internet, Internet of Things	Mylonas
WS11-09	7 May 2010 17:15:00	7 May 2010 21:00:00	Distributed processes, Cloud computing, P2P	Chatzigiannakis
WS11-10	14 May 2010 17:15:00	14 May 2010 21:00:00	Semantic Knowledge Representation and Reasoning	Bratass
WS11-11	24 June 2010 18:15:00	24 June 2010 20:00:00	Examination	
WS11-12	15 July 2010 18:15:00	15 July 2010 21:00:00	Project's Presentation	Bratass

Lecture	MainReadingsTitle	MainReadingLinks
WS11-02	SMW+ Semantic MediaWiki RDF Specification W3C RDF Homepage Halo Project	<a href="http://semanticweb.org/wiki/Project_Halo">http://semanticweb.org/wiki/Project_Halo</a> <a href="http://smwforum.ontoprise.com/smwforum/index.php/Main_Page">http://smwforum.ontoprise.com/smwforum/index.php/Main_Page</a> <a href="http://semantic-mediaWiki.org/wiki/Semantic_MediaWiki">http://semantic-mediaWiki.org/wiki/Semantic_MediaWiki</a> <a href="http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/">http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/</a> <a href="http://www.w3.org/RDF/">http://www.w3.org/RDF/</a>
WS11-03	Editing Description Logic Ontologies with the Protégé OWL Plugin Ian Horrocks (2008) Ontologies and the semantic web. Commun. ACM, Vol. 51, No. 12, pp. 59-67. A nice overview of the differences between OWL and relational databases. Papers relating to OWL, Ontologies and Protége are available in citeulike tagged as "xml summer school". Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, Chris Wroe (2004) OWL Pizzas: Practical Experience of Teaching OWL-QL. Common Errors and Common Patterns In Proc. of European Conference on Knowledge Acquisition (EKAW'04), Vol. 3257 (2004), pp. 63-81. An overview of common errors and pitfalls (with solutions) to building ontologies in OWL using pizzas as an example.	<a href="http://protege.stanford.edu/plugins/owl/publications/OL2004-protége-owl.pdf">http://protege.stanford.edu/plugins/owl/publications/OL2004-protége-owl.pdf</a> <a href="http://www.comlab.ox.ac.uk/people/ian.horrocks/Publications/download2008hor08a.pdf">http://www.comlab.ox.ac.uk/people/ian.horrocks/Publications/download2008hor08a.pdf</a> <a href="http://www.citeulike.org/gaj/xml/summer-school">http://www.citeulike.org/gaj/xml/summer-school</a> <a href="http://www.co-ode.org/resources/papers/ekaw2004.pdf">http://www.co-ode.org/resources/papers/ekaw2004.pdf</a> <a href="http://www.co-ode.org/ontologies/">http://www.co-ode.org/ontologies/</a> <a href="http://lwl.cs.manchester.ac.uk/tutorial/protege/tutorial/">http://lwl.cs.manchester.ac.uk/tutorial/protege/tutorial/</a> <a href="http://protege.stanford.edu/">http://protege.stanford.edu/</a> <a href="http://www.co-ode.org/resources/tutorials/ProtégeOWLTutorial.pdf">http://www.co-ode.org/resources/tutorials/ProtégeOWLTutorial.pdf</a>

The Semantic TreeView on the right shows a hierarchical structure of the course content, including categories like People, Academic Staff, Secretarial Personnel, Students, Assignment Owners, Graduated Students, Project Participants, Thesis Candidates, Courses, PostGraduate, UnderGraduate, and Lectures, with sub-items for each lecture (WS01 to WS09).

Fig. 2. A course as it is presented at WSSW

### 2.3 Highlights

The WSSW provides the usual benefits the most Semantic Wikis tend to exhibit and as these are summarized in scientific papers. It differentiates however in terms of content presentation and availability. Most Semantic Wikis seem to lack a much desired flexibility in the ways that content becomes available to their users. The WSSW approach caters for this need with a more thorough use of semantic annotations and attains a greater level of flexibility. It manages to offer consistent and well presented content in multiple and rich ways. WSSW users may choose one of the several different presentations that best suits their needs and experience enhanced browsing and navigation capabilities.

Apart from the inline links in regular semantic wikis, the WSSW offers alternative ways to navigate through its content. For instance, users may browse the WSSW via its compendious Factbox which sums up the semantic annotations referred at each specific page or choose a Semantic TreeView on the right of the context pane. In addition, one may use the most conventional way of the menu, which is an inline semantic query on each one. Besides regular navigation, enhanced Semantic Search Option and friendly Query Interface are provided to the users to serve their needs.

A well-adapted Semantic Calendar accompanied with Semantic Maps allows users to be informed of the timetables and the exact places associated with each lecture. This and all other intersperse information of the WSSW are readily available via inline semantic queries.

Another strong point of the WSSW is that it exploits the advantages of a Jena Triple Store Connector that besides storing, it also exposes data to the Semantic Web

and allows queries to be performed on them from other remote endpoints. This feature makes the difference comparing to the common Semantic Wikis which are independent of the rest Linked Data Cloud. The uppermost goal is the WSSW to be able, also, to incorporate data from Linked Data Cloud as well.

At the design stage of WSSW, SMW+ set of extensions was the only option for a wiki powered by MediaWiki to expose its knowledge base to the Semantic Web and to also allow queries via SPARQL. One has nowadays additional options when choosing a triple store connector/extension with each providing different functions and advantages. A comparison of their features is available at [4].

### 3 Conclusions and Future Work

Through WSSW we demonstrated that Advanced Semantic Wikis, integrated with the Semantic Web and Linked Data in the form of SLMS, exhibit true potential in becoming rich alternative eLearning approaches in the Linked Data Age. For this purpose a brief sum up of the up-to-date work on the field of education and Linked Data was presented mainly focusing on the WSSW. The used platform, its ontologies, its structure and its semantic functions were concisely reported.

As far as future milestones are concerned it is important to overcome the barrier that predefined ontologies cause and broaden the presented content by importing other sources from the Linked Data Cloud without applying any changes to our ontology. It is also of critical importance for the MSc in Web Science to incorporate an extension to this Semantic Wiki which will allow the instructors to query multiple repositories and aggregate selected resources to recommend to students and in general, to embody data from the Linked Data cloud. Another milestone is to use SKOS representations of scientific classifications and to characterize, the available learning objects of the WSSW, so as to clearly define their content and to make it easily searchable. The uppermost goal is a complete solution which will make broad use of Linked Data.

### 4 References

1. Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data – the story so far. International Journal on Semantic Web and Information Systems (IJSWIS), 2009, pp.1-22
2. Bamidis P.D., Kaldoudi E., Pattichis C., "mEducator: A BEST PRACTICE NETWORK FOR REPURPOSING AND SHARING MEDICAL EDUCATIONAL MULTI-TYPE CONTENT", In proceedings of PRO-VE 2009, Springer Verlag 2009, IFIP Advances in Information and Communication Technology 307, pp. 769-776
3. Bratsas C., Kapsas G., Konstantinidis S., Koutsouridis G., Bamidis P.D., "A Semantic Wiki within Moodle for Greek Medical Education", in Pocc of 22nd IEEE International Symposium on Computer-Based Medical Systems, New Mexico, USA, 2009, pp.1-4s.
4. Comparison of Semantic MediaWiki triplestore connectors, [http://www.mediawiki.org/wiki/Extension:SparqlExtension#Comparison\\_of\\_Semantic\\_MediaWiki\\_triplestore\\_connectors](http://www.mediawiki.org/wiki/Extension:SparqlExtension#Comparison_of_Semantic_MediaWiki_triplestore_connectors)

# A Linked Data-driven & Service-oriented Architecture for Sharing Educational Resources

Hong Qing Yu<sup>1</sup>, Stefan Dietze<sup>1</sup>, Ning Li<sup>1</sup>, Carlos Pedrinaci<sup>1</sup>, Davide Taibi<sup>2</sup>, Nikolas Dovrolis<sup>3</sup>,  
Teodor Stefanut<sup>4</sup>, Eleni Kaldoudi<sup>3</sup> and John Domingue<sup>1</sup>

<sup>1</sup> Knowledge Media Institute, The Open University, MK7 6AA, Milton Keynes, UK

<sup>2</sup> Italian National Research Council, Institute for Educational Technologies, Italy

<sup>3</sup> School of Medicine, Democritus University of Thrace, Greece

<sup>4</sup> Technical University of Cluj-Napoca, Romania

**Abstract.** The two fundamental aims of managing educational resources are to enable resources to be reusable and interoperable and to enable Web-scale sharing of resources across learning communities. Currently, a variety of approaches have been proposed to expose and manage educational resources and their metadata on the Web. These are usually based on heterogeneous metadata standards and schemas, such as IEEE LOM or ADL SCORM, and diverse repository interfaces such as OAI-PMH or SQI. Also, there is still a lack of usage of controlled vocabularies and available data sets that could replace the widespread use of unstructured text for describing resources. On the other hand, the Linked Data approach has proven that it offers a set of successful principles that have the potential to alleviate the aforementioned issues. In this paper, we introduce an architecture and prototype which is fundamentally based on (a) Linked Data principles and (b) Service-orientation to resolve the integration issues for sharing educational resources.

**Keywords:** Linked Services, Semantic Web, Linked Data, SOA, eLearning, Technology-enhanced Learning.

## 1 Introduction

Recently, learning resource repositories have been widely used in higher education to enhance the learning process for both students and course creators or teachers. One important feature of Web-based learning repositories is to make educational resources reusable and shared. Although most of the learning repositories provide open access to their educational resource repositories, the integration process is still costly as different learning repositories are isolated from each other and based on different implementation standards [18]. Therefore, an intuition approach is to federate different learning repositories from different institutes, which can foster the aims of reusing and sharing educational resources without costly duplicating them into local learning repositories. Following such an approach, four major research challenges need to be taken into consideration to ensure Web-scale interoperability:

1. *Retrieving distributed data from heterogeneous Learning repositories:* distributed heterogeneous data and services are neither widely reused nor integrated into learning application environments sufficiently. Standardized methodologies to solve heterogeneities between terminologies used by distinct data or service providers are not available. Therefore, interoperability and scalability of current E-Learning applications is limited [17].
2. *Dynamic metadata mediation:* since learning contents and their metadata are stored in heterogeneous repositories, the metadata and the services consuming the content are usually described by using distinct schemas such as Dublin Core<sup>1</sup> and IEEE Learning Object Metadata (LOM) [11], and a variety of vocabularies accordingly. Therefore, to achieve interoperability, distinct schemas need to be mapped and aligned.
3. *Extensible learning environment:* in highly distributed Web-based environments, frequent changes occur to available interfaces. That is, services as well as repositories are usually added, modified or removed regularly based on the new requests and new considerations.
4. *Enriching existing educational resource metadata:* many educational resource metadata stored in different Web repositories are non-aligned, incomplete, and poorly structured, i.e. free text is still widely used for describing educational resources. Therefore, to allow machine-processing and Web-scale processing, distributed educational metadata needs to be enriched, that is transformed into structured and formal descriptions. This may lead to wider interoperability by linking to existing vocabularies, schemas and taxonomies.

The most recent development of the Semantic Web [1] is Linked Data (LD) [2] which has successfully established a set of principles to expose data and metadata on the Web and has led to the widespread availability and use of schemas, vocabularies and data sets spanning across all application domains. In that, the Linked Data approach offers opportunities to substantially alleviate the challenges addressed above.

In this paper we introduce a Linked Data-driven & Service-oriented architecture to address the discussed issues above. The four major contributions are:

1. Linked Data-principles are applied to model and expose metadata of both educational resources and educational services and APIs. In this way, not only resources are connected but also services' description and resources are exposed in a standardized and accessible way.
2. Existing heterogeneous and distributed learning repositories, i.e. their Web interfaces (services) are integrated on the fly by reasoning and processing of Linked Data-based service semantics (see step 1).
3. Metadata retrieved from heterogeneous Web repositories, for instance IEEE LOM resource metadata, is automatically lifted into RDF and exposed as Linked Data, exposed and accessible based on Linked Data principles.

---

<sup>1</sup> <http://dublincore.org/documents/dces/>

A set of RESTful APIs is developed on top of the integration framework to allow third party applications to consume and interact with the data exposed by our approach.

The remaining sections of the paper cover: Section 2 discusses the background and motivation of our work. Section 3 illustrates the overview of our proposed architecture. Section 4 describes the different parts of the architecture while Section 5 introduces an educational application from the biomedical field. Section 6 discusses the related work and Section 7 finally concludes the paper.

## 2 Background and Motivation

One use case motivating a general architecture to address above research problems is a federated search across educational data crossing distributed Web repositories. More specifically, the user would search educational resources based on, for instance, a set of keywords, content types, educational subjects or preferred language. That however requires to: (1) integrate suitable resource stores via their APIs and Web services that can deliver correct educational resources; (2) dynamically invoke the suitable services and mediate the results into a single standard and meaningful response; (3) transform and automatically expose it as structured well-interlinked Linked Data.

One way to enable educational resources inside individual learning repositories to be accessed and integrated through distributed environments is to develop Web services and open APIs on top of educational repositories. However, this naturally introduces a heterogeneous API landscape, as it is characteristic for the current Web. For example, one Web service interface from the PubMed<sup>2</sup> repository provides an OAI-PMH-based REST-ful service where response messages are based on XML. Other Web services offering educational Linked Data from The Open University<sup>3</sup> provide a SPARQL [25] end point where response messages are RDF-based. Furthermore, different services may also operate on a certain language based on the target user groups. Meanwhile, there are multiple types of educational resource metadata schema such as IEEE LOM and Dublin Core. Therefore, dynamic mediation between different metadata schemas and API interfaces is required. However, the current metadata standards are mainly stored and exposed based on XML and relational databases and consist largely of poorly structured text keywords, lacking formal semantics and interlinking with established vocabularies to allow more advanced discovery and inference mechanisms. That leads also to highly ambiguous descriptions, which, considering the medical domain, for instance, can prove as crucial factors for the value and quality of resource descriptions.

Although some efforts have been made on providing an IEEE LOM-RDF binding<sup>4</sup>, this early work was (a) discontinued a few years ago and (b) only focused on the binding aspect rather than further working towards a Linked Data-principles compliant approach, e.g. by reusing elements of established Linked Data schemas or linking metadata to established Linked Data vocabularies.

---

<sup>2</sup> <http://www.ncbi.nlm.nih.gov/pubmed/>

<sup>3</sup> <http://data.open.ac.uk/>

<sup>4</sup> <http://dublincore.org/educationwiki/DCMIIEEEELTSTaskforce/RDFPAR>

While our work considers the integration of arbitrary data on the Web, we use the mEducator educational resource RDF description schema (see mEducator project<sup>5</sup>) as a schema according to which we lift existing (non-RDF) metadata for further processing and exposure as fully interlinked Linked Data. The mEducator RDF schema<sup>6</sup> is fully compliant with Linked Data principles.

The other important factor of the challenge is the distributed and Service-oriented environment. The services operating on educational repositories are very dynamic, in that they might change behaviors and interfaces according to new requirements. Also, new services (and stores) regularly join the environment while others might disappear. Therefore, the extendable environment feature is a very important requirement mainly considering supporting the modification, adding or removal of services without any impact to upper layers (e.g. existing educational applications and user interfaces). Therefore, facilitating easy-to-use service representation based on standard service vocabularies (e.g. SAWSDL [15] and WSMO-Lite [13]) is an important requirement to allow third party service providers to independently provide meaningful service descriptions and interlink their services with the environment.

In this paper, we are building on top of two fundamental technologies which facilitate data interoperability (a) Semantic Web technologies to facilitate data interoperability and (b) Services-orientation to allow interoperability at the repository level.

The four principles of LD are [2].

1. Use URIs to identify things.
2. Use HTTP URIs so that these things can be referred to and looked up ("dereference") by people and user agents.
3. Provide useful information about the "thing" when its URI is dereferenced, using standard formats such as RDF/XML.
4. Include links to other, related URIs in the exposed data to improve discovery of other related information on the Web.

The above principles have proved largely successful throughout the past years, leading to an ever increasing amount of LD-compliant schemas and data-sets<sup>7</sup> as well as general-purpose tools and APIs allowing to generate and interact with LD.

Linked services adopt the LD principles to support publishing of service annotations as LD, expressed in terms of a simple generic service conceptual model that can be consumed by both humans and machines. While there is a growing set of tools supporting the development and processing of Linked Services, the iServe platform<sup>8</sup> is one of the implementations to publish Linked Services [16]. In order to cater for interoperability, iServe uses what can essentially be considered the maximum common denominator between existing Semantic Web Services (SWS)

---

<sup>5</sup> <http://www.meducator.net/>

<sup>6</sup> <http://purl.org/meducator/ns/>

<sup>7</sup> <http://richard.cyganiak.de/2007/10/lod/>

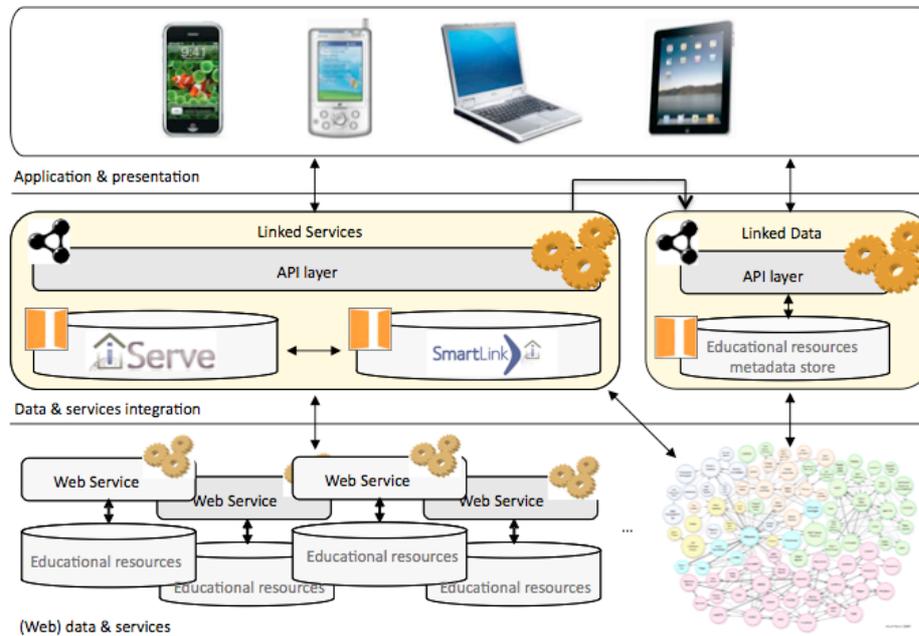
<sup>8</sup> <http://iserve.kmi.open.ac.uk/>

formalisms that we refer to as the Minimal Service Model (MSM). Services are defined as having a number of operations. Each operation has an Input and Output *MessageContent* and Faults. *MessageContent* is defined to may have *MessageParts*. Additionally, iServe adopts the SAWSDL, WSMO-Lite and hRESTS [14] vocabularies. The SAWSDL vocabulary is used to capture *modelReference*, *liftingSchemaMapping* and *loweringSchemaMapping* that are important for lifting and integrating services' outputs. WSMO-Lite extends SAWSDL with a model to specify the semantics of the particular service annotations. The hRESTS vocabulary allows modeling of additional information necessary for Web APIs.

### 3 Overall Architecture

In this section, we provide an overview of a general-purpose framework which aims at (i) integrating heterogeneous educational resources and (ii) exposing its metadata as well-structured and interlinked Linked Data. Our overall proposed architecture includes three layers: *(Web) data and service layer*, *Data and service integration layer* and *Application and presentation layer* that are shown in Figure 1.

- *The (Web) data and service layer* consists of available educational resource metadata, Web services and data sources such as the ones part of the Linked Data cloud.
- *The data and service integration layer* is based on the Linked Services approach, exploiting Linked Data based service annotation, the APIs to broker services and an RDF repository for exposing enriched educational resource data. The iServe [16] & SmartLink [6] repositories store two different kinds of service annotations separately, namely functional and non-functional service annotations.
- *The application and presentation layer* uses the APIs provided by the data & services integration layer to interact with underlying data & services and provides an interface to end-users.



**Fig. 1.** Architecture overview.

The proposed architecture supports a particular methodology as below:

1. *Query across distributed educational repositories:* The Linked Services environment allows distributed queries across distributed and heterogeneous educational repositories (via their services/APIs) on the fly while query results will be lifted into RDF automatically. The Linked Services component facilitates dynamic discovery of suitable stores/services (for instance services matching a particular user language or subject) and execution by adhering to particular execution constraints. New repository services can be added by simply adding semantic annotations of the services via our Linked Services annotation environments. The service discovery engine searches the semantically suitable services by matching service annotations with service request parameters that are passed through the application layer. Suitable services for a given query are invoked sequentially. Finally, different service outputs are lifted into the RDF mEducator RDF schema. That way, heterogeneous service responses are consolidated into a single schema that simplifies further processing by the application layer.
2. *Storage of selected metadata in RDF store:* Exposing of retrieved results as Linked Data/RDF in an RDF store. The RDF repository supports two main purposes: (1) allows content providers who do not have a metadata publishing platform to publish their educational resource metadata through the application layer; (2) allows metadata sets from the distributed learning

repositories (Step 1) to be enriched and exposed in a Linked Data-compliant way. The RDF store is implemented based on a Sesame RDF store with a BigOWLim and mEducator RDF schema compliant repository and a dedicated REST API. Each resource entity owns a unique URI identifier that can be dereferenced.

3. *Data enrichment and interlinking*: Metadata can be enriched based on LD datasets, e.g. particular ones from the biomedical field. That is of particular importance for metadata properties such as keywords, discipline or subject where established vocabularies exist on the Web, and in particular the LD cloud. Enrichment takes advantage of available APIs such as the ones provided by *Bioportal*<sup>9</sup>, which allow access to a vast number of established taxonomies and vocabularies, such as *SNOMED*<sup>10</sup>, *MESH*<sup>11</sup> or *Galen*<sup>12</sup>. That way, unstructured free text, for instance the keyword “Thrombolysis”, is enriched with unique URIs of structured LD entities - such as <http://purl.bioontology.org/ontology/SNOMEDCT/89551006> which refers to a related concept within the *SNOMED Clinical Terms* ontology – which allow not only further reasoning on related concepts but also enables users to query for resources by using well-defined concepts and terms as opposed to ambiguous free text.

## 4 Data and Services integration: the Linked Services Approach

Our current implementation of the data and service integration layer builds on existing Semantic Web Services research, namely the Linked Services approach. As discussed in Section 2, the Linked Services approach uses a rather lightweight service annotation schema and applies Linked Data principles to the services domain. Based on RDF models describing core elements of the service (e.g., operations, input, output) services are discovered and executed based on a given set of service consumer constraints.

To this end, we use the iServe open platform for publishing semantic annotations of services based on a direct application of LD principles [16]. As introduced previously, the MSM – iServe’s underlying RDF schema - focuses on functional properties while the SmartLink service annotation model focuses on non-functional properties (NfP) as extension to the MSM. SmartLink NfP data are stored in a dedicated RDF store that is synchronized with the iServe repository at runtime.

The whole lifecycle of the Linked Services approach includes 3 major steps:

1. Editing and publishing service annotations through the SmartLink service annotation online interface. SmartLink allows service providers to use a Web form to easily annotate service properties step-by-step following the extended MSM schema. Meanwhile, service requesters can also look up services

---

<sup>9</sup> [http://www.bioontology.org/wiki/index.php/BioPortal\\_REST\\_services](http://www.bioontology.org/wiki/index.php/BioPortal_REST_services)

<sup>10</sup> <http://www.ihtsdo.org/snomed-ct/>

<sup>11</sup> <http://www.nlm.nih.gov/mesh/>

<sup>12</sup> <http://www.co-ode.org/galen/>

manually apart from using our automatic API-based service discovery and invocation approach.

2. Services discovery based on service request: a set of RESTful APIs has been developed to let third-party applications send service requests distinguished by different parameter values to discover semantically suitable educational resource services.
3. Services invocation and lifting: finally, the identified services are invoked and heterogeneous service responses are lifted to comply with the mEducator Learning Content RDF description schema.

The rest of this section will introduce above steps in details.

#### **4.1. SmartLink: Linked Services editor and search environment**

In order to provide a Linked Services editor that allows the annotation of Web services and Web APIs without any pre-existing documentation, a services annotation and search tool is developed, *SmartLink*<sup>13</sup> (SeMantic Annotation enviRonmenT for Linked services). SmartLink allows annotation of Web services and Web APIs based on the MSM from scratch, that is, without any pre-existing services documentation such as WSDL or HTML files, as assumed by existing annotation tools (Section 1). SmartLink operates on top of LD stores such as iServe and is an open environment accessible to users simply via OpenID<sup>14</sup> authentication.

SmartLink exploits an extension of the MSM schema including a number of additional non-functional properties. These non-functional properties cover, for instance, contact person, developer name, Quality of Service (QoS), development status and service license. MSM-schema properties are directly stored in iServe, while additional properties are captured in a complementary RDF store based on OpenRDF Sesame<sup>15</sup>. The following Figure 2 depicts the overall architecture of the SmartLink environment.

---

<sup>13</sup> <http://smartlink.open.ac.uk> & <http://kmi.open.ac.uk/technologies/name/smartlink>

<sup>14</sup> <http://openid.net/>

<sup>15</sup> <http://www.openrdf.org/>

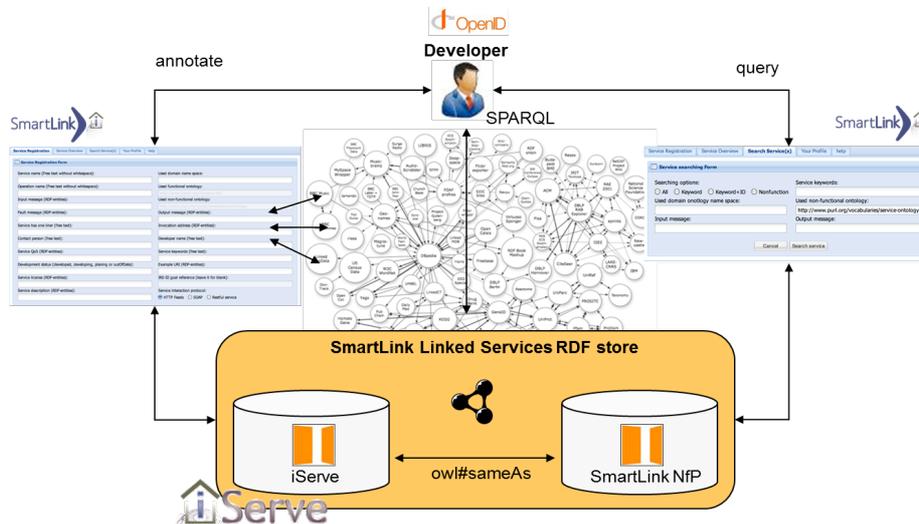


Fig. 2. SmartLink process and architecture.

Being a Linked Data-compliant environment, one of the core features of MSM is the capability to associate service descriptions with so-called *model references* that refer to RDF descriptions in external vocabularies defining the semantics of the service or its parts. That way, for instance, a particular *service response message* can be associated with an external RDF description which details and further describes the nature of the response. However, while this feature is useful and even necessary in order to provide meaningful service models, finding appropriate model references across the entire Web of data is a challenging task. Therefore, SmartLink uses established Linked Data APIs – currently the WATSON<sup>16</sup> API – to identify and recommend suitable model references to the user.

#### 4.2. A RESTful API for Linked Services discovery and execution

A RESTful API was developed and designed to allow third party applications to interact with our RDF annotations of educational services and APIs, for instance, to discover and execute services.

The service discovery function of our developed REST-ful API takes three service requirement parameters of category, subject and language, which can each be referenced by RDF entities defined in Link Data cloud, such as <http://www.daml.org/2003/09/factbook/languages#English>. One example vocabulary for service categories is the *Service-finder ontology*<sup>17</sup>. The subject finally specifies the educational domain that the service operation deals with. Different vocabularies can be used to define subjects, for instance the Open Learn classification vocabulary<sup>18</sup>

<sup>16</sup> <http://watson.kmi.open.ac.uk/>

<sup>17</sup> <http://www.service-finder.eu/ontologies/ServiceCategories>

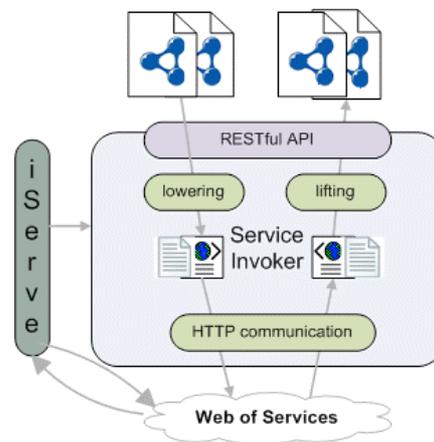
<sup>18</sup> <http://meducator.open.ac.uk/ontologies/open-learn-classification.rdf>

describes *OpenLearn*<sup>19</sup> learning subject categories. The language specifies the language requirement for service operation output, and in that, allows the discovery of services which offer access to educational data in a specific user language. All the requirement properties are optional (but at least one property should be specified). By taking these service request parameters into account, the discovery API circulates a SPARQL query to the iServe and Smartlink service annotation repositories for gaining information about suitable services. The discover response message is a RDF based output that contains service name, service description, subject, endpoint and output language. The Listing 1 is one example of the service discovery RDF for a service request.

```
<rdf:Description rdf:about="http://www.meducator.net/services#UH-Helsinki-OAI-DC-metadata">
  <hasAddress rdf:resource="http://helda.helsinki.fi/dspace-oai/request?verb=ListRecords&metadataPrefix=oai_dc"/>
  <liftingSchemaMapping rdf:resource="http://meducator.open.ac.uk/xsparql/uhlifting"/>
  <hasDeveloper>University of Helsinki</hasDeveloper>
  -<hasOneLiner>
    Learning resources metadata from University of Helsinki, Metadata in OAI-Dublin Core XML. Feed compliant with OAI-PMH.
  </hasOneLiner>
  <subject rdf:resource="http://meducator.open.ac.uk/ontologies/open-learn-classification#education"/>
  <metadataLanguage rdf:resource="http://www.daml.org/2003/09/factbook/languages#English"/>
</rdf:Description>
```

**Listing 1** Service description excerpt.

A particular invocation method is provided by the API that supports the lifting of service responses – whatever response message format (e.g., XML or JSON) – into RDF compliant with the mEducator RDF schema. The service invocation API method takes service invocation parameters provided by the application layer and returns a mEducator RDF schema compliant RDF output.



**Fig. 3** Service invocation and lifting process.

Service invocation and lifting: based on service input and output annotations and corresponding lifting/lowering schema description, an RDF service invocation message is generated by dynamically matching parameters (specified by client users)

<sup>19</sup> <http://www.open.ac.uk/openlearn>

to service semantic annotations. As shown in Figure 3, the RDF input message will go through the “lowering” process of the invocation API to the actual input format of the service. A service is then invoked with the lowered input. Meanwhile, each individual service may give different output result such as XML, RDF or JSON. In order to provide service response messages compliant with the mEducator RDF schema, native output from service invocations will be transformed via the “lifting” step to RDF compliant with the mEducator RDF schema. Not only does the RDF output enable data mashup at semantic level, it also could be used, directly or in combination with other RDF data, for the invocation of further services through the invocation API.

## **5 Metamorphosis+: An Application Layer Combining Social Computing with Semantic Data and Services Integration**

The data and services integration architecture presented in previous sections is fully exploited in Metamorphosis+, which merges the paradigms of semantic and social web to produce an environment for sharing educational resources in health sciences. MetaMorphosis+ (or MM+) realizes the integration of a novel social environment as the application layer and user interface with the semantic data and services linking architecture presented here.

At the application layer, MetaMorphosis+ can be viewed as two distinctive and interacting networks. The first one is a network of persons, including authors, potential authors and final users of learning objects (students, teachers or others, e.g. educational managers, etc). The second is a network of educational resources. The network of persons is functioning in a way similar to other social networks. Persons can interact with each other via their personal blogs, declare friends and create their own interest groups. At a different level, educational resources themselves create an equivalent social network. Educational resources in MetaMorphosis+ can be resources residing in a Learning Management System (LMS), in another educational repository, or merely available on the Web. Resources (like humans) are represented in MetaMorphosis+ by their profile. Educational resources as social objects can exhibit different aspects of ‘object sociality’ [12]: (a) the obvious connections via common tags; (b) connections based on collective usage and other related interaction of human users; (c) social connections based on the inheritance as realized via educational content repurposing; (d) semantic connections realized via semantic annotations and linking of educational resources. This later social dimension is fully realized via semantic data and service linking as described in the previous sections.

The application layer, via the Resource Profile View allows viewing and management and annotation (Figure 4) of the educational resource metadata as they are lifted from the RDF triple store. Additionally, a new resource and its metadata can be declared via the application layer, with the respective updating of the triple store.

**Edit profile**

educational resource

General Description Classification Educational Repurposing Vocabularies Companion

Resource Title

Identifiers

[add]

-- identifier type --

-If the resource is protected by a Creative Commons licence, please choose which one. (If your resource is not protected by any IPR licence, please get one by visiting the Creative Commons website, and then state which licence you acquired below).

If the resource is protected by any other IPR licence please state which one:

Quality stamp

[add]

Resource Language

-- select language --

Language of metadata

-- select language --

Save

**Fig. 4.** Resource annotation in Metamorphosis+

The application layer fully exploits the search and retrieval APIs searching for relevant resources two distinct mechanisms: (a) either search in the RDF store; or (b) on the distributed learning repositories semantically linked via the data and services integration layer. The search can be focused on general keywords or keywords found inside specific profile fields of each resource (Figure 5)

**Search for educational resources**

New search

Save search parameters under title: Save

Basic search

Keyword(s):

Search

Advanced search

Specific search

**Fig. 5.** Search interface of MetaMorphosis+. The user can do the simple search by putting keywords or do advanced search by specifying the interested metadata fields.

## 6 Related Work

A peer-to-peer architecture (LOP2P) for sharing educational resources among different learning Institutions is proposed in [18]. LOP2P aims at helping different educational institutions to create course material by using shared educational resource repositories. The major advantage of the peer-to-peer approach is that the learning repositories can be easily integrated after adding the LOP2P plugin and mediation layer to each different repository. A similar peer-to-peer architecture has also been proposed in the EduLearn project [17] and [8]. Meanwhile, Simple Query Interface (SQI) is introduced in [23] designed to query different learning repositories using a comment query language. However, query format and result format have to be agreed among different repository providers before using the query functionalities, which means that a wrapper service is required to ensure compliancy of all involved repositories with the agreed format. These approaches are sharing three disadvantages: (1) instead of accepting the heterogeneous landscape of the Web, all approaches impose either a common schema or interface approach on the underlying stores. Therefore, the heterogeneity issue hasn't been solved. (2) The sharing functionalities are limited by using a defined mediation layer and the mediation is based on syntactic matching, which is not an efficient mechanism to deal with an open and distributed environment.

The work described in [20] and [21] utilizes Semantic Web as well as Web service technologies to enable adaptation to different learning contexts by introducing a matching mechanism to map between a specific context and available learning data. However, this work neither considers approaches for automatic service discovery nor it is based on common standards. Hence, the reuse and automatic allocation of a variety of services or the mediation between different metadata standards is not supported. These issues apply to the idea of "Smart Spaces" [22] for learning as well. The work in [3] follows the idea of using a dedicated personalization Web service that makes use of semantic learning object descriptions to identify and provide appropriate learning content. Neither is the integration of several distributed learning services within the scope of this research, nor is the allocation of services at runtime. Further related research on [9] and [10] allows a mediation between different services based on a so-called "connector service".

## 7 Conclusion and Future Work

Integrating educational resources becomes more and more important since plenty of educational resources are published to be available online while a range of interfaces and description approaches are being used. In this paper, we introduced a Linked Data-driven & Service-oriented architecture to resolve the integration issues by addressing interoperability issues in a distributed and heterogeneous environment via dynamic metadata mediation towards an extensible and fully Linked Data-compliant resource description layer. The Metamorphosis+ application, an open environment for (biomedical) education has been developed based on our proposed architecture.

While the presented work is ongoing research and tackles a number of distinct challenges such as metadata interoperability, services discovery or data mediation, there are plenty of opportunities for future works. In the short term and most importantly, these cover: (1) investigating ways to enable efficient, accurate and dynamic enrichment of educational data, what involves research fields such as text mining, entity recognition and ontology mapping; (2) extending the framework with additional open repositories and data stores to further showcase and evaluate our services integration approach; (3) integrating the APIs of the “data & services integration layer” with additional third party applications to further evaluate the performance and scalability of the architecture. In particular, (1) also needs to cover the consideration of security and authentication aspects while context-awareness is an important aspect to be taken into account with respect to (2). Additionally, many aspects of our proposed framework are domain-independent and it is intended to deploy similar approaches in different areas, such as eScience.

## References

1. Berners-Lee, T., Hendler, J., and Lassila, O. (May 17, 2001). The Semantic Web Scientific American Magazine. March 29, 2009.
2. Bizer, C., T. Heath, et al. (2009). Linked data - The Story So Far. Special Issue on Linked data, International Journal on Semantic Web and Information Systems (IJSWIS).
3. Baldoni, M., Baroglio, C., Brunkhorst, I., Henze, N., Marengo, E. and Patti, V. (2006). A Personalization Service for Curriculum Planning. In proceedings of *14th Workshop on Adaptivity and User Modeling in Interactive Systems*, Hildesheim.
4. Davies, J., Domingue, J., Pedrinaci, C., Fensel, D., Gonzalez-Cabero, R., Potter, M., Richardson, M., and Stincic, S. (2009). Towards the open service web. *BT Technology Journal*, 26(2).
5. Dimitrov, M., Simov, A., Konstantinov, M., Momtchev, V. WSMO Studio – a Semantic Web Services Modelling Environment for WSMO (System Description). In: Proceedings of the 4th European Semantic Web Conference (ESWC). Number 4519 in LNCS, Innsbruck, Austria (2007) 749–758.
6. Dietze, S., Yu, H.Q., Pedrinaci, C., Liu, D. and Domingue, J. (2011) SmartLink: a Web-based editor and search environment for Linked Services, 8th Extended Semantic Web Conference (ESWC), Heraklion, Greece
7. Dietze, S., Gugliotta, A. and Domingue, J. Supporting Interoperability and Context-Awareness in E-Learning through Situation-driven Learning Processes, Special Issue on Web-based Learning of International Journal of Distance Education Technologies (JDET), IGI Global, 2008.
8. Hadzic, M., D’Souza, R., Hadzic, F., Dillon, T. Thinking PubMed: an Innovative System for Mental Health Domain. Computer-Based Medical Systems, IEEE Symposium on, pp. 330-335, 2008 21st IEEE International Symposium on Computer-Based Medical Systems, 2008
9. Henze, N. (2006). Personalized E-Learning in the Semantic Web. Extended version of 4. *International Journal of Emerging Technologies in Learning (iJET)*, 1(1).
10. Henze, N., Dolog, P., and Nejdil, W. (2004). Reasoning and Ontologies for Personalized E-Learning. *Educational Technology & Society*, 7(4).

11. IEEE, IEEE Standard for Learning Object Metadata, *IEEE Std 1484.12.1-2002*, vol., no., pp.i-32, 2002, doi: 10.1109/IEEESTD.2002.94128, URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1032843&isnumber=22180>
12. Kaldoudi E, Dovrolis N, Giordano D, Dietze S., Educational Resources as Social Objects in Semantic Social Networks, Proceedings of the Linked Learning 2011: 1<sup>st</sup> International Workshop on eLearning Approaches for the Linked Data Age, Herakleio, May 2011
13. Kopecky, J., Vitvar, T., and Gomadam, K. 2008. MicroWSMO. Deliverable, Conceptual Models for Services Working Group, URL: [http://cms-wg.sti2.org/TR/d12/v0.1/20090310/d12v01\\_20090310.pdf](http://cms-wg.sti2.org/TR/d12/v0.1/20090310/d12v01_20090310.pdf).
14. Maleshkova, M., Pedrinaci, C., and Domingue, J. (2009). Supporting the creation of semantic restful service descriptions. In Workshop: Service Matchmaking and Resource Retrieval in the Semantic Web (SMR2) at 8th International Semantic Web Conference.
15. Maleshkova, M., Kopecky, J., and Pedrinaci, C. (2009). Adapting SAWSDL for semantic annotations of restful services. In Workshop: Beyond SAWSDL at OnTheMove Federated Conferences & Workshops.
16. Pedrinaci, C., Liu, D., Maleshkova, M., Lambert, D., Kopecky, J., and Domingue, J. (2010) iServe: a Linked Services Publishing Platform, Workshop: Ontology Repositories and Editors for the Semantic Web at 7th Extended Semantic Web Conference.
17. Prakash, L. S., Saini, D. K., and Kutti. N. S., 2009. Integrating *EduLearn* learning content management system (LCMS) with cooperating learning object repositories (LORs) in a peer to peer (P2P) architectural framework. *SIGSOFT Softw. Eng. Notes* 34, 3 (May 2009), 1-7. DOI=10.1145/1527202.1527212 <http://doi.acm.org/10.1145/1527202.1527212>
18. Rafael de Santiago, Andre L.A. Raabe, "Architecture for Learning Objects Sharing among Learning Institutions-LOP2P," IEEE Transactions on Learning Technologies, pp. 91-95, April-June, 2010.
19. Sheth, A. P., Gomadam, K., and Ranabahu, A. (2008). Semantics enhanced services: Meteor-s, SAWSDL and SA-REST. *IEEE Data Eng. Bul* 1, 31(3):8-12.
20. Schmidt, A., Winterhalter, C. User Context Aware Delivery of E-Learning Material: Approach and Architecture, *Journal of Universal Computer Science (JUCS)*. 10(1), 2004.
21. Schmidt, A. "Bridging the Gap Between E-Learning and Knowledge Management with Context-Aware Corporate Learning (Extended Version)", In proceedings of *Professional Knowledge Management (WM 2005)*, Springer.
22. Simon, B., Dolog., P., Miklós, Z., Olmedilla, D. and Sintek, M. Conceptualising Smart Spaces for Learning. *Journal of Interactive Media in Education*. 2004(9), from <http://www.jime.open.ac.uk/2004/9>, 2004.
23. Simon, B., Massart, D., Assche, F., Ternier, S., Duval, E., Brantner, S., Olmedilla, D., and Miklos. Z. A simple query interface for interoperable learning repositories. In Proceedings of the 1st Workshop On Interoperability of Web-Based Educational Systems, pages 11-18, 2005.
24. Ternier, S., Duval E., and Vandepitte, P. LOMster: peer-to-peer learning object metadata, *Proc. of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2002, ISBN 1-880094-45-2.
25. World Wide Web Consortium, W3C Recommendation, SPARQL query language for RDF, 2008, (<http://www.w3.org/TR/rdf-sparql-query/>).

# Zenaminer: driving the SCORM standard towards the Web of Data

Ernesto Mudu<sup>1</sup>, Luca Schiatti<sup>1</sup>, Giuseppe Rizzo<sup>1,2</sup>, and Antonio Servetti<sup>1</sup>

<sup>1</sup> Dipartimento di Automatica e Informatica, Politecnico di Torino  
Corso Duca degli Abruzzi, 24, 10129 Torino, Italy  
ernesto.mudu@studenti.polito.it, luca.schiatti@studenti.polito.it,  
giuseppe.rizzo@polito.it, antonio.servetti@polito.it

<sup>2</sup> Department of Multimedia Communication, EURECOM  
2229, Route des Crêtes, 06560, Sophia Antipolis, France  
giuseppe.rizzo@eurecom.fr

**Abstract.** In this paper we present Zenaminer, a software architecture for linking the SCORM standard to the Linked Open Data cloud. The main idea is to overcome some limitations of the SCORM Content Packaging standard, the most used standard for sharing e-learning contents, in order to migrate this material into the Web of Data. We analyze the design of a RESTful Web Service for SCORM where each single SCO (Sharable Content Object) is exposed on the web as raw data that can be created, read, updated, and deleted through the HTTP protocol. In addition, borrowing the idea from the concept of web mashup, we propose to build the Learning Management System interface as a mashup that uses Zenaminer web API to access SCO data and that customizes the presentation and interaction with the help of additional CSS and Javascript. To achieve this goal, for the creation of the HTML material, we suggest to use a “light formalism” defined by the W3C with the name Slidy so that the presentation remains separated from the content. Finally, Zenaminer allow users to enrich the SCORM resources with comments. Comments can be contents themselves that extend and improve the original material of the teacher. Unstructured comments are automatically annotated with Spotlight linking them to the LOD cloud through DBpedia.

**Keywords:** SCORM, e-learning, LOD, DBpedia, web service, REST, learning object

## 1 Introduction

The evolution of digital communications and the more and more relevant presence of the Web in our society lead to significant changes also in the e-learning scenario. A new term is being used to define this change, “e-learning 2.0” [6]. In relation to the same shift happened in the Web, among the various innovations that characterize the shift from the 1.0 to the 2.0 system, in this work we focus on the new paradigm defined as “Web of Data”. That is, the old Web of mostly

human readable documents has to change in a Web of micro contents that can be easily processed by machines [1] and by the users. The big advantage is in the ability to manipulate every single part of a documents on-the-fly with the possibility to aggregate and remix the data in order to generate new documents of enhanced value. Micro contents are the building blocks on which Linked Data can operate, they represent contents that convey one idea and that can be accessed through a single URL in the space of the Web. Examples of micro contents can be the arrival and departure times for an airplane flight, the abstract from a long publication, or other similar. As a consequence of this transformation, the Web is changing from a read-only platform to a read-write platform where users, starting from data available on-line, can also share, remix and create new and original material [12]. Web users are becoming both consumers and producers, i.e., prosumers.

These ideas are not new at all in the e-learning scenario. Here, teaching material is defined in the form of micro contents named Sharable Content Objects (SCO). SCOs form the basis of usability, interoperability, and adaptation. They are the building blocks that can be independently produced, stored, indexed, composed and evaluated. SCOs can be considered a compromise between raw data (e.g., image, video, audio, text) and a whole lesson: they have a context, but small, so they can flexibly arranged to form new lessons or presented in a different shape changing the appearance of their elements (e.g., logos, titles, etc.). For example, in this work we have defined a SCO as a group of slides on the same topic inside a lesson. Thus, the production of teaching material implies the production of a number of SCOs that can be collected to create an archive of sharable micro contents. Nevertheless, no services have been defined to migrate this information into the Web of Data. SCOs are shared as documents, ZIP archives, that a user need to download and extract before being able to use them in a new project.

The paradigm of the "Web as a Platform" [10] requires instead these contents to be exposed on the Web through an Application Programming Interface (API) that allows a set of operations on them, i.e., create, read, update, delete. From a Linked Data point of view, a Learning Object Repository (LOR) can be seen as a Web Service (WS) where, briefly, each LO is a resource (and so identified by URI) that can be processed using the HTTP methods (GET, PUT, POST, and DELETE) of the Representational State Transfer (REST) [7]. In this paper, we propose a novel architecture to expose on the Web, as a Linked Data resource, the teaching material of a Shareable Content Object Reference Model (SCORM) packet, the de-facto standard in the production of e-learning content. Our objective is twofold, the design of a new Web interface for SCORMs and the migration of LOs into the Linked Data Cloud where the "consumer" can infer, connect and aggregate data from different repositories in order to create new value added resources. The added value is not in the data itself, but in how contents are combined in new ways, how they are presented with new interfaces, in how they are enriched far beyond they original content defined by the teacher.

The remainder of this paper is organized as follows. A review of the current state of the art is presented in Section 2. The key ideas of our approach, with some details on the issues of exporting SCORMs by means of a REST web service, are described in Section 3. Then, in Section 4, we discuss the concept of separation of concerns, which is at the basis of our tool. In Section 5 we describe the architecture of the web service and, finally, in Section 6 the use case used to test our work. Conclusions follow in Section 7 where we stress on the benefits introduced by the tool into the e-learning community.

## 2 Related Work

E-learning systems has assumed an important role in the Web community for the important value of the contents which they shared. In this context, contents are named sharable content objects (SCO) and they hold information about lectures and metadata useful to describe better the inferred contents. Based on the assumption that the Web provides the best opportunity to maximize both access to and the reuse of learning contents, the overall objective of these systems are to make sharable content objects usable, reusable and interoperable. In order to address this, our goals are to define and to implement a model to export teaching materials adhering the Web of Data.

Many efforts has been spent to make SCORM environment interoperable<sup>3</sup> within the Web cloud. Many of them focused to make interoperable LMSs in a cloud infrastructure [5] e [4] other, instead, to perform interoperability sharing learning objects. Although the idea of making a network of services which may handle teaching materials is interesting for what concerns managing the author information, security and network reliability, on the other hand it requires other formalism to make interoperable the atomic information. In addition, it is far from the idea of making the Web of Data linkable, which is one of the best practice of the Linked Data.

Vossen and Westerkamp [14] introduced the need to share SCO by means of service. They proposed a service oriented architecture (SOA), but any specific architecture or implementation was described. In [3], Redondo et al. discussed typical problems of SCORM-related standards and proposed a web service-oriented approach as a solution. They introduced the concept of Everything as a Service (Eaas) and they mapped each single SCO as a service. A WSDL exposed information about services available to gather information from the each SCO. Gonzalez et al [8] extended Moodle<sup>4</sup> to the SOA paradigm. Modules and courses were available from the Web and sharable from a different kind of applications, making it a portable tool. To sum up, these approaches try to expose features of these LMSs, using the service-oriented architecture. Our work, instead, exposes data from each single SCO in a raw way, in order to make this data linkable and navigable from the Web applications.

---

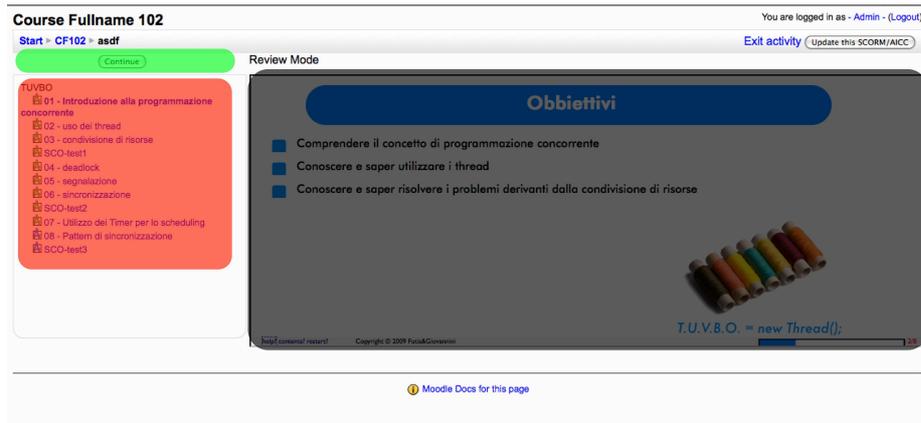
<sup>3</sup> According to the Open Knowledge Initiative (OKI) definition, interoperability is about making the integration as simple and cost effective as technologically possible.

<sup>4</sup> <http://moodle.org/>

### 3 A web service for SCORM

SCORM, i.e., the Sharable Content Object Reference Model, is a reference model for the creation of web-based e-learning material with the objective to ensure interoperability, reusability, and accessibility. The SCORM content packaging section specifies how the course material should be packaged, usually in a ZIP file, and described. The course is defined as a collection of Sharable Content Objects (SCOs) that can be associated to a lesson or a part of it. A Content Object must be a web-deliverable learning unit, that is usually designed as an HTML page with CSS and Javascript so that it can be launched in a web browser.

SCORM packages are managed by a Learning Management System (LMS) that, among other important functions such as administration, tracking, and reporting, provides also a mean to display this information to the users. In order to correctly deliver the contents, the LMS parses the package metadata to understand the course structure, known as the “activity tree”, and to know how to launch each SCO. Since most of times the material takes the form of slides, “traditional” LMS use a fixed HTML frameset structure, as shown in Fig. 3, that includes a list of the available lessons, a set of navigation buttons, and a panel where the slides are shown.



**Fig. 1.** Screenshot of the web interface of the Moodle Learning Management System.

This kind of interface presents some limitations:

- The SCO designer can completely define the content and presentation of each single SCO, but it can not control at all the interface used to present and navigate it (that is part of the LMS functions);
- The SCO designer’s work area is limited by the dimension of the HTML frame defined by the LMS, thus it is difficult to integrate different media and materials, i.e., a video of the teacher, slideshows, comments;

- SCOs are included in the LMS with the graphical style defined by their own SCO designer, style that can be unhomogeneous if an LMS integrates different SCO sources. If content and presentation were independent, it would be instead easy to redefine each SCO in order to present an homogeneous presentation;
- A fixed interface and the impossibility to separate presentation from content make hard to adapt the interface to different displays, i.e., smartphones, tablets, netbooks, etc.

To overcome these limitations, we present *Zenaminer*, a web service to export SCORM packages through a RESTful web API, that is a web service implemented using HTTP and the principles of REST (REpresentational State Transfer). An architecture where a SCORM package is not shared as a single object (the ZIP packet), but as a collection of smaller objects, the SCOs, each of them accessible independently from the others. It is the implementation of the new paradigm of the “Web of Data”, SCOs are publicly available as read/write raw data that can be retrieved or updated using a REST interface as described in the next sections.

The key factors that enable *Zenaminer* to provide such a functionality are two:

- the knowledge of the SCORM content packaging standard, to import the ZIP files;
- the knowledge of the content “internals”, i.e., how the slide information is defined, to separate the content from the presentation.

*Zenaminer*, for the creation of HTML documents, suggests the introduction of a “light formalism” proposed by the W3C with the name Slidy [11]. Slidy defines simple keywords for the class attribute of HTML tags that can be used to identify HTML elements as slides, titles, sidebar, incremental lists, etc. – each of them with a particular presentation style defined by CSS or behavior implemented in Javascript.

In short, SCORM contents are available as resources of the web service given a specific URI. The SCO designer can freely define the interface as he likes it, then contents will be dynamically retrieved from *Zenaminer* using AJAX (Asynchronous Javascript and XML) calls. As illustrated in Section 5, web API give access to each single slide of the course as well as to the table of contents for the navigation. In addition *Zenaminer* extends the SCORM metadata to include also references to additional information such as the video recording of a lesson, synchronization information between video and slides, etc.

Taking into consideration another characteristic of what is defined as e-learning 2.0, that is collaborative learning, *Zenaminer* enables students to enrich the SCORM resources with comments on the course, the lesson, or the slide. Comments can be contents themselves that extend and improve the original material provided by the teacher. Furthermore, unstructured comments are au-

tomatically annotated with DBpedia Spotlight<sup>5</sup> linking them to the Linked Open Data cloud through DBpedia [1][2].

## 4 Content presentation

In order to structure the presentation of contents we adopted the Slidy formalism, which uses HTML for the description of its items. The choice to adopt HTML becomes important in our approach because it opens to all devices able to connect and visualize Web pages and, more important, it intrinsically performs the separation of contents from the presentation (also called view). According to that, we may define data able to exchange information and how this data may be used to build the presentation: this is obtained with the use of the Cascading Style Sheet (CSS) file. By means of it, a view maker may define models to render specific items or group of them. These models are named class and are referred to an HTML page. The interaction between HTML page and CSS is performed by means of selectors, which are able to point to specific items or class of an HTML page. In this context, the Javascript, a “scripting” language, performs the possibility to select an item or a set of them and to change the behaviour of the view, previously defined. In addition, it works as a tool to enrich dynamically the presentation, e.g. making table of contents (TOC), suggesting the value of an acronyms, changing font or window dimensions. Then, it performs an important role to dynamically access structured data without any layout details, or raw data, which are coming from different archives spread within the Web cloud. The technology that allows to gather data from external sources is the Asynchronous Javascript and XML (AJAX).

Our work exploits the Separation of Concerns (SoC) principle, introduced above, in the context of a Learning management system. The difference between data and view allows to navigate through resources, making customizable views which better respond to the need of who provide contents (author) and who use, reuse and redistribute them (user). Data is presented without layout details, raw, but it is rich of semantics: data is stored with the information about inference to others and are exposed to users by means of REST APIs. Our approach exploits the MVC (Model-View-Controller) paradigm, in which the view is the set of presentation rules for the data, the Model is the amount of data raw available for a generic topic, while Controller is the set of methods which are needed to create the communication channel from the Model and the View. Our data Model is deployed in the system which it holds SCORM packages, while the View is created on all user machines whenever a user require it. The Controller, instead, is located on the user machine where the view is created and on the LMS where SCORMs are available.

To provide the maximum interoperability and reusability, contents indexed and managed by our LMS are slides, lectures and courses. All of these are Web resources available by URIs time-invariant. This feature is important because we

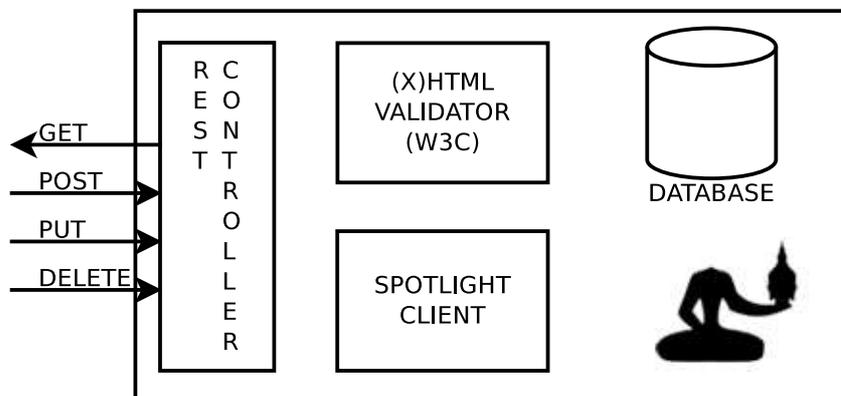
---

<sup>5</sup> <http://dbpedia.org/spotlight>

need to perform durability and reusability to each Web resource. Slide contents are described by means of HTML Slidy formalism; each slide is composed of title, pictures, vertical and horizontal scroll bars, interactive items (buttons, check box, etc.). Lectures are organized as follows: slides and video fragments. A fragment is a part of video which is related to a particular lecture. To map lecture and fragments, we introduced a synchronization file. A video is available and accessible by means of a URI with the above requirements. Course, lecture and slide exploit the SoC principle, separating the view from the data model. The video, instead, holds the information about its presentation and it is shared without changing.

## 5 Architecture

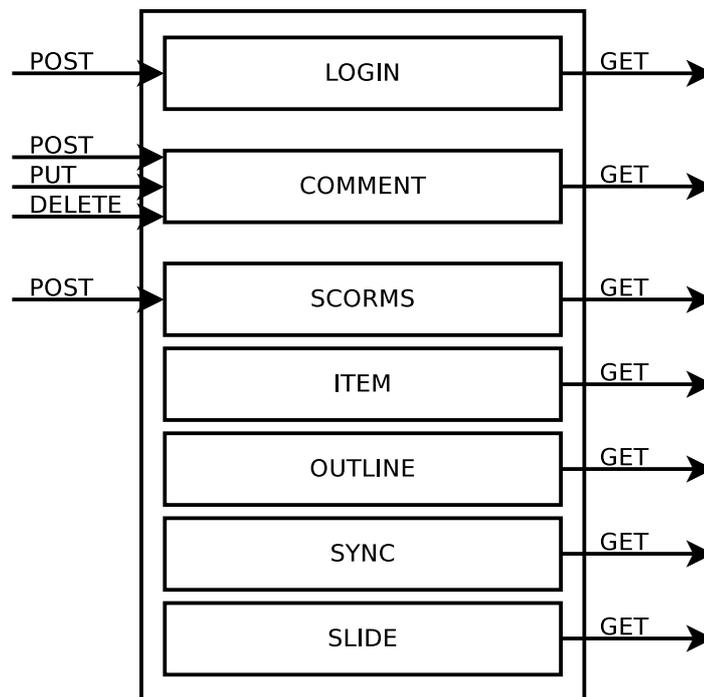
The architecture is structured in four building blocks that are summarized in Figure 2:



**Fig. 2.** Zenaminer architecture. The REST controller is the interface between Zenaminer and the Internet; the W3C validator is used to check that imported (X)HTML files are well-formed; the Spotlight client provides automatic annotation; the database stores imported SCORMs together with comments and annotations.

1. The REST controller is the interface between Zenaminer and the Internet. Its features are grouped in three sets (as shown in Figure 3): user management and authentication (login), get and post of comments (comment) and management of SCORM packages (scorms, item, outline, sync and slide). Each of these features is implemented using the REST architecture, thus resources are available through URIs in the Internet. Access to REST calls that provide access to (GET) or modify (POST, PUT, DELETE) existing contents can be limited using the existing facility for user management and authentication.

2. A local validator is used to verify that (X)HTML files are valid and well-formed before they are parsed. When a new SCORM package is imported each HTML file that contains slides is validated using the remote interface offered by the official W3C validator<sup>6</sup>.
3. When a comment (an enrichment) is received it is passed to the Spotlight client. The Spotlight client sends a request to DBpedia Spotlight service and receives an annotated version of the comment.
4. The SQL Database is used to store imported SCORM packages together with comments and annotations.



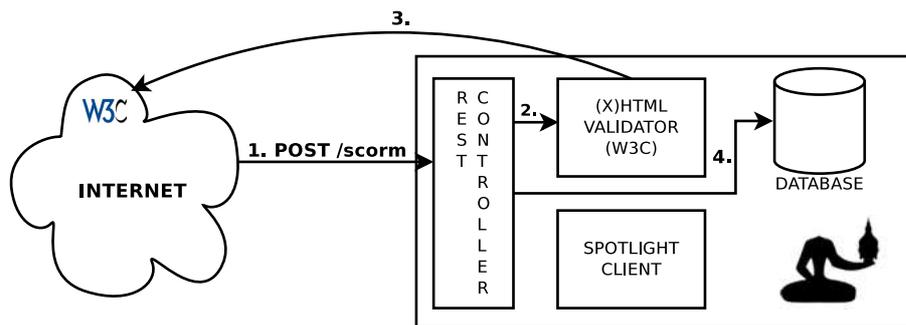
**Fig. 3.** Detail of the REST controller, its three main features are: user management, comment management and SCORM management.

Figure 4 shows the workflow for the upload of a SCORM package into Zenaminer, it is performed in the four steps below:

1. the client sends a **POST** request to `/scorm` page, including in the request body the SCORM package;

<sup>6</sup> <http://validator.w3.org>

2. the REST controller receives the package, unpacks and analyzes the files containing slides and sends them to the W3C validator;
3. the W3C validator validates individual files and reports the result to the REST controller;
4. if all files have been validated correctly the package is imported into Zenaminer. Otherwise if files are not validated Zenaminer sends to the client an error message containing errors generated by the W3C validator.



**Fig. 4.** Workflow describing the upload of a SCORM package in Zenaminer: 1. a SCORM package is received from a client; 2. (X)HTML files are forwarded to the validator; 3. (X)HTML files are validated by sending requests to the official W3C validator; 4. if all documents are valid, the SCORM package is stored to the database.

Zenaminer gives the possibility to its users to enrich the content of a given content (e.g. a slide) by adding a textual comment to it. Such comments are named enrichments because they are meant to add more information to the content of a SCO, for example an enrichment to a slide could be a proof to a theorem, a correction to its content, an in-depth explanation to a concept, etc.. In order to avoid that content of an enrichment is buried in the database we added automatic annotation in order to enable machines to access to the content, enhancing search and the possibility to display complementary information.

Figure 5 shows the actions performed by Zenaminer when a comment is received:

1. the client sends a POST request to /comment page, the message body contains the text of the comment;  
An example of enrichment is:  
President Obama called Wednesday on Congress to extend a tax break for students included in last year's economic stimulus package, arguing that the policy provides more generous assistance.
2. the text is forwarded to the Spotlight client;

- the Spotlight client sends a request to DBpedia Spotlight and obtains as response an automatically annotated version of the comment. As an additional parameter to the annotate request we could pass a minimum confidence and support values that will impact on precision and recall of the annotation itself;

The request URI to annotate the enrichment in the previous example is:

```
GET http://spotlight.dbpedia.org/rest/annotate?
text=President%20Obama%20called%20Wednesday%20on%20Congress%20to
%20extend%20a%20tax%20break%20for%20students%20included%20in%20last
%20year%27s%20economic%20stimulus%20package,%20arguing%20that
%20the%20policy%20provides%20more%20generous%20assistance.
&confidence=0.4
&support=20
```

The response (in JSON format below) contains a reference to the annotated term in the enrichment (@surfaceForm) and the DBpedia URI relative to the annotated content (@URI):

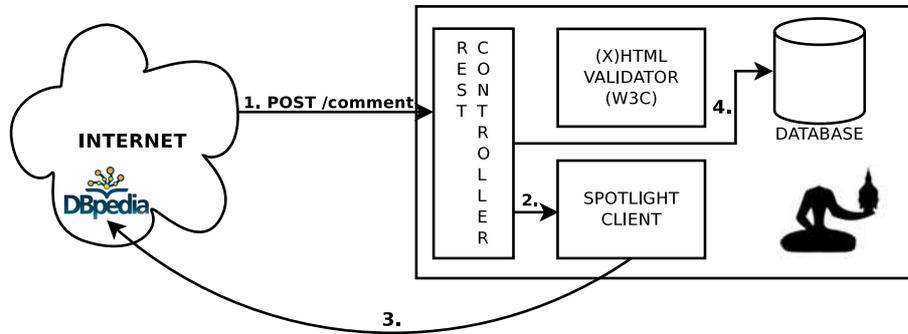
```
{
"@URI": "http://dbpedia.org/resource/United_States_Congress",
"@support": "9321",
"@types": "Organisation,Legislature",
"@surfaceForm": "Congress",
"@offset": "44",
"@similarityScore": "0.14802740514278412",
"@percentageOfSecondRank": "0.6257434730652487"
}
```

- finally, the enrichment and its annotations are stored into the database.

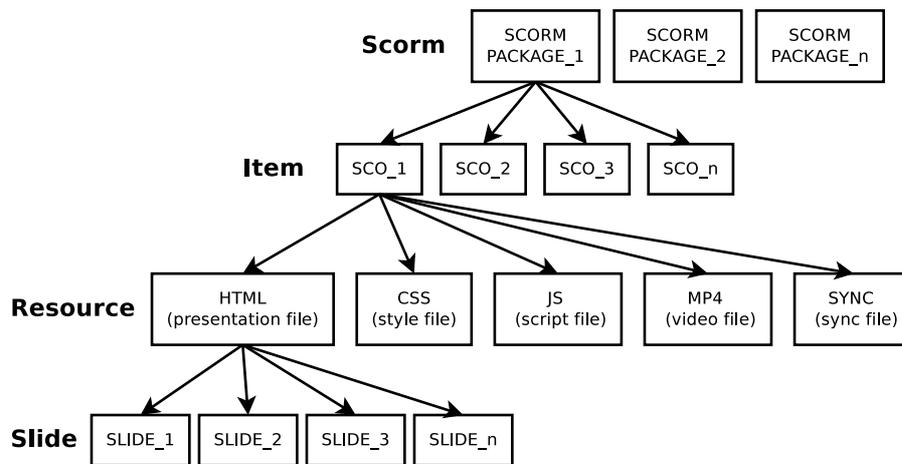
A SCORM package includes a XML file named “manifest”, it describes a course and is used to list all lectures in it, all resources (files) associated to each lecture and some sequencing rules used to define how the user accesses to SCOs. Figure 6 shows in a tree structure how the manifest file is represented into the relational database. In addition, a SCORM package lists several SCOs, each SCO includes an HTML file containing slides, a CSS file to describe the style, a Javascript file to describe the behavior. Additional files can be included in a SCO: videos, subtitles or files describing synchronization between a video and the slides. Files linked into the manifest are then packaged together into a ZIP file. Currently Zenaminer does not support sequencing rules, thus management of such rules is up to the SCO designer.

## 6 Use case

The validation phase of the Web-service was conducted during the course “Multimedia Environments” (Academic Year 2010/2011) for the Master of Science in Cinema and Media Engineering at the Politecnico di Torino. Zenaminer was thought as a service in order to design e-learning projects (SCORM Packages)

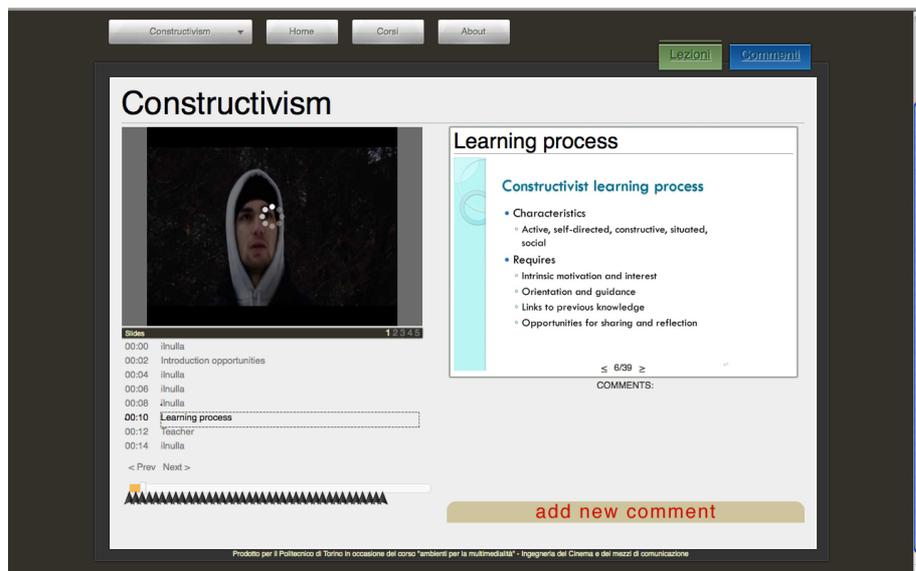


**Fig. 5.** Workflow describing the reception of a comment in Zenaminer: 1. a comment is received from a client; 2. the content of the comment is forwarded to the Spotlight client; 3. the Spotlight clients contacts DBpedia Spotlight to obtain an annotated version of the comment; 4. the comment together with annotations are stored to the database.



**Fig. 6.** A tree representation of a SCORM package in Zenaminer. A SCORM package is a set of SCOs, each SCO represents a lesson in a course, each SCO contains several files like HTML, CSS, JS or videos. HTML files containing a presentation are parsed and slides are extracted.

for the course “Multimedia Environments”. The objective for the students was to create a SCORM package defining both content and presentation. Students were divided in 20 teams, each team defined a personal learning environment building different interfaces (using the SoC concept). Students acted as SCO designers, the projects were the use case for Zenaminer and we verified the potentiality of the separation of content from presentation. The demo is available at <http://eridano.polito.it:8080>. Figure 7 and 8 show two different views of the same raw data done by two different groups of students. Such interfaces are able to show same contents in two different ways. The controller of each interface collects the list of SCORM packages stored into Zenaminer. Depending on the SCORM package selected by the user, the controller gathers the entire set of lectures for that package and displays the related Table of Contents (ToC). The user is, then, free to navigate the lectures (SCOs) following the ToC, when a lecture is selected the controller requests all files associated to that lecture and the view is updated accordingly. For example, in both figures the selected SCO had a set of slides and a video associated to it. Thanks to the sync file, the view is able to synchronize the video with the slides.



**Fig. 7.** One of the user interfaces designed by students for the course Multimedia Environments.

The Web Service was developed with the framework Pylons 1.0<sup>7</sup>. The source code<sup>8</sup> is released under the GNU GPLv3. In order to better balance and dis-

<sup>7</sup> <http://pylonshq.com>

<sup>8</sup> <https://sourceforge.net/projects/zenaminer/>

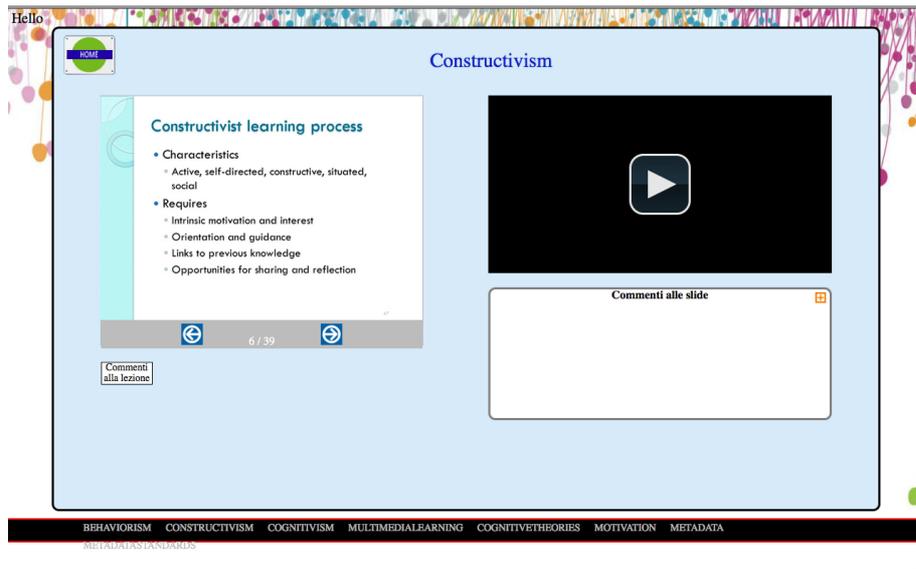


Fig. 8. Another example of user interfaces designed by students for the course Multimedia Environments.

tribute the load we decided to use Apache 2.2 HTTP Server<sup>9</sup> as interface between the requests towards the Web Service. As relational database we used PostgreSQL 9<sup>10</sup>; it stores data extracted from the SCORM by the web-service. Raw data stored in the database are available through REST calls and serialized into JSON format. The W3C Markup Validation Service is used to validate the HTML of the lectures that must be conform to W3C standards. Finally we used Spotlight to automatically annotate the enrichments of learners.

## 7 Conclusions

In this paper we proposed and argued a Web Service Architecture for e-learning. Zenaminer goes beyond the limitations of the SCORM and allows a complete separation of content and presentation. Content creators are free to define learning environments without constraints of traditional LMSs. Zenaminer does not limit the definition of the interfaces, the graphic design and the learner interaction with a SCO. In the field of e-learning the presentation (the interface) does not only involve the aesthetic appeal and the usability but also affects the learning process. In fact the interface design [13] and the adaptation of interfaces to the learner [9] are two key factors for e-learning success.

<sup>9</sup> <http://httpd.apache.org>

<sup>10</sup> <http://www.postgresql.org/>

The “Separation of Concerns” (SoC) allows SCO designers to improve e-learning acceptance. For example a SCO designer could create different environments for the same course according to the learners’ computer literacy or could improve the interoperability of SCOs designing different interfaces for different devices (e.g. Smart Phones). Not secondary is the possibility to integrate and manage multimedia contents as separate entities from the learning content. In traditional LMSs a multimedia content must be integrated in a SCO while, in Zenaminer, content-types other than text (e.g. videos) can be managed and integrated regardless of SCOs. This feature does not compromise the SCORM compatibility with traditional LMSs.

Finally Zenaminer allows collaborative learning. Learners have an active role in the definition of contents thanks the possibility to enrich lectures with personal contributions. Content creators are free to define different models of collaborative learning; they can implement a Wiki model or they can choose a model with less collaboration. In addition to collaborative learning it is possible to integrate contents thanks to instruments offered by the Semantic Web. The Automatic annotation allows the integration between of content with hypertext link navigation in order to satisfy more effectively the information needs of the learner. Furthermore the SCO designer could decide to allow learners to disambiguate or integrate automatic annotations. The use of ontology-based annotation pushes Zenaminer to be a Linked Data LMS. The architecture proposed in this paper allows Instructional designers to better implement learning theories (e.g. Cognitivism, Constructivism) and teaching strategies. For example they could choose to minimize the content of a course and focus on collaborative learning; this approach satisfies the key aspect of Constructivism that considers fundamental the collaboration between learners [15]. Moreover this work represents a step towards the publishing of LMS related data to the Web of Data and new future works are planned. First of all, to be totally compliant to the Linked Data paradigm we want to investigate about the conversion of LMS related data to RDF, and to expose them using a SPARQL endpoint. Then we want to investigate about the use of disambiguation process through the DBpedia Spotlight. The idea behind is offer to users the possibility to better understand ambiguous concepts by means of the semantic disambiguator. This tool may highlight ambiguous concepts in a SCO and may suggest multiple references per each ambiguous concept, then user may choose the most interesting.

## Acknowledgments

Thanks to all students of the course “Multimedia Environments” academic year 2010/2011 of the Politecnico di Torino. They were developers and analysts of our Web Service with a vibrant cooperation spirit, which was fundamental in achieving this important goal. A special thanks to SCO designers Jacopo Berta, Federica Tina Bossa (who developed the SCO view shown in Figure 7) and Lucia Marengo, Alice Ferrari (who developed the SCO view shown in Figure 8).

## References

1. Berners-Lee, T.: Linked data. *International Journal on Semantic Web and Information Systems* 4(2) (2006)
2. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - a crystallization point for the web of data. *Web Semantics Science Services and Agents on the World Wide Web* 7(3), 154–165 (2009)
3. Díaz Redondo, R., Fernández Vilas, A., Pazos Arias, J.: Educateca: A web 2.0 approach to e-learning with SCORM. In: Cellary, W., Estevez, E. (eds.) *Software Services for e-World, IFIP Advances in Information and Communication Technology*, vol. 341, pp. 118–126. Springer Boston (2010)
4. Dong, B., Zheng, Q., Qiao, M., Shu, J., Yang, J.: BlueSky cloud framework: An e-learning framework embracing cloud computing. *Lecture Notes in Computer Science Cloud Computing* 5931, 577–582 (2009)
5. Dong, B., Zheng, Q., Yang, J., Li, H., Qiao, M.: An e-learning ecosystem based on cloud computing infrastructure. In: *Proceedings of the 2009 Ninth IEEE International Conference on Advanced Learning Technologies*. pp. 125–127. ICALT '09, IEEE Computer Society, Washington, DC, USA (2009)
6. Downes, S.: E-learning 2.0. *eLearn 2005* (October 2005)
7. Fielding, R.T., Taylor, R.N.: Principled design of the modern web architecture. *ACM Transactions on Internet Technology* 2(2), 115–150 (2002)
8. Gonzalez, M., Penalvo, F., Guerrero, M., Forment, M.: Adapting lms architecture to the soa: An architectural approach. In: *Proc. of the Fourth International Conference on Internet and Web Applications and Services, ICIW '09*. pp. 322–327 (May 2009)
9. Mödritscher, F., Barrios, V.M.G., Gütl, C.: Enhancement of SCORM to support adaptive e-learning within the scope of the research project AdeLE. In: Nall, J., Robson, R. (eds.) *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2004*. pp. 2499–2505. AACE, Washington, DC, USA (2004)
10. Murugesan, S.: Understanding web 2.0. *It Professional* 9(4), 34–41 (2007)
11. Raggett, D.: Slidy - a web based alternative to Microsoft PowerPoint. W3C (2006)
12. Richardson, L., Ruby, S.: *RESTful Web Services*. O'Reilly Media, Inc. (May 2007)
13. Selim, H.M.: Critical success factors for e-learning acceptance: Confirmatory factor models. *Computers & Education* 49(2), 396–413 (September 2007)
14. Vossen, G., Westerkamp, P.: Why service-orientation could make e-learning standards obsolete. *International Journal of Technology Enhanced Learning* 1(1), 85–97 (2008)
15. Wilson, B.G.: *Constructivist Learning Environments: Case Studies in Instructional Design*. Educational Technology Publications, illustrated edn. (1996)

# Developing controlled vocabularies for educational resources sharing: a case study

Daniela Giordano<sup>1</sup>, Isaak Kavasidis<sup>1</sup>, Concetto Spampinato<sup>1</sup>,  
Panagiotis Bamidis<sup>2</sup>

<sup>1</sup> Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Università di Catania,  
Viale A. Doria 6, 95125 Catania, Italy

<sup>2</sup> Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki,  
PO Box 323, 54124 Thessaloniki, Greece  
{dgiordan, ikavasidis, cspampin}@dieei.ing.unict.it  
{bamidis}@med.auth.gr

**Abstract.** We present a novel controlled vocabulary for the classification of Educational Resource Type and Media type that has been developed within the mEducator Best Practice Network, and discuss the rationale behind its design and its SKOS implementation. Several available controlled vocabularies were analyzed, but none was found entirely satisfactory for the project's purposes. Although the driving motivation of this effort was to account for and deal with multi-type content variety that is especially characteristic of the medical field, the result of this process and the proposed organization generalise fairly well to broader educational/learning contexts. We discuss the proposed vocabularies pointing out the multifaceted role they may fulfil in the linked learning era.

**Keywords:** Resource Type, Media Type, controlled vocabularies, SKOS, Linked Data, metadata reuse

## 1 Introduction

Thesauri and taxonomies (which extend simple lists of terms by expressing also the relationship between terms) are currently acknowledged as an essential tool in any successful knowledge management effort [1]. In a taxonomy terms are arranged or linked in a tree, so that narrower, more specific “children” terms fall under broader, more generic “parent” terms; in thesaurus relationships of affinity, synonyms and relatedness among terms can be expressed. Controlled vocabularies, more simply, consist of a flat list of terms used for indexing or categorizing; their goal is to ensure consistency in indexing, tagging, or categorizing and to guide the user to the desired information. In this respect, controlled vocabularies can be used to assist the users either in filling metadata for resource description, and conversely, can be proposed in the user search interfaces to assist during the search process. This latter practice somehow addresses a recent criticism of taxonomies/controlled vocabularies, i.e. that often they are too far from the user language, and therefore are unlikely to be used spontaneously for searching. Hence the support for effective resources retrieval is questioned. On the other hand, “folksonomies” emerging from social tagging of

resources do provide a collection of terms closer to the end-user language. Taxonomies and folksonomies are not mutually exclusive or incompatible, and there is a fast growing body of literature pointing out the benefit of their integration [2], [3].

In systems that resort to user generated metadata, as content sharing solutions often do, there is a tension between the design choice of enforcing controlled vocabularies versus allowing free tagging and filling of the metadata fields. This tension originates from considering that in the indexing phase, when a resource is published, and should be described as effectively as possible, the user, who is not a professional indexer, should be aware of the vocabularies, and know how to use them. Often this is not the case, and even when the user is aware of the existence of a classification system, its actual usage, especially in the case of vocabularies with a very large number of terms (e.g., MeSH) will depend on how easily it can be browsed and understood. On the other hand, filling with unconstrained vocabularies metadata fields, often results in poor quality because of semantic ambiguity in the name of the fields [4].

From a Linked Data (LD) perspective [5] some of these issues can be solved or ameliorated at a fundamental level. In fact, the LD paradigm places emphasis on the reuse of available metadata fields even from different schemas, which inherently favours the reuse of the most useful and meaningful fields; whereas concerning the actual choice of controlled terms, there is the freedom to choose any system that suits the need of the community, as long as it is treated consistently with the DL principles and links are created with other data sets dealing with similar terminologies. An even better practice is to create through RDF links explicit mappings, by resorting to RDF or OWL properties between similar terms in different classification systems (e.g., through the use of OWL's "same as", or "relates to" or SKOS's "closeMatch", "exactMatch", etc.). One clear advantage of this approach is the enabling of automatic metadata enrichment by resorting to the LD Cloud (e.g., [6], [7]): this practice may result into more robust search and retrieval systems also by assisting the users with dynamical suggestion of terms during metadata publishing or during the searches.

The goal of this paper is twofold. First, we present a controlled vocabulary concerning Educational Resource type and Media type that has been developed in the context of the mEducator project, funded within the eContentPlus Programme and concerned with best practices in innovative technical solutions for educational content sharing and repurposing. This vocabulary has a specific *raison d'être* in the context of a very wide and open network of stakeholders broadly concerned with medical education and also with the practical issue of repurposing educational resources that are often overspecialised and very expensive to develop. Then we discuss generalisation issues and, based on the experience accrued within the project we derive some implications in the context of the Linked learning movement.

## **2 Background: multi-type content sharing and repurposing**

The mEducator project ("Multi-type Content Sharing and Repurposing in Medical Education, [www.meducator.net](http://www.meducator.net)) is a Best Practice Network funded by the European Programme eContentPlus and is concerned with the evaluation of two innovative ways to share medical educational contents across European institutions and across

the community of educators (i.e., one solution utilizing semantic web services technology and one solution based on mashup technology). Among the project's goals is the provision of recommendations to relevant standardization bodies (e.g. to Medbiquitous, that currently employs the HLOM, the Healthcare extension of the IEEE LOM, [8]) regarding suitable, additional extensions to the metadata schema to effectively support both the sharing and the repurposing process. In this regard, an action of the project has been concerned with the actual design and development of these extensions. After several rounds of polling the partners about inclusion of the available fields from the HLOM, and critical re-examination of the HLOM from the Technical Reference Group of the mEducator Consortium, an RDF-based meducator schema was developed [9], reusing existing metadata fields and including new meducator fields dealing specifically with the repurposing history of a resource and other pedagogical aspects of the resource, e.g. medical educational outcomes.

One line of investigation of mEducator that is at the basis of the work presented in this paper concerns the identification of educational content types that are specific to medical education and to test the descriptive adequacy of the currently available e-learning metadata standards against these content types. For example, a Virtual Patient (VP) is a specific medical content type that employs a problem-based learning approach and allows the learners to follow either linear or branched paths to diagnose the patient and suggest treatments [10]. Similarly, interactive medical images (e.g., X-rays) where the learners can draw tracings to identify anatomical structures or pathologies/abnormalities and obtain visual feedback are a novel resource type especially suitable for the development and refinement of visual perception skills [11]. Initially, the distinction between content type and resource type was blurred, and the only agreement within the Consortium on the semantics of the field was that it should try to capture the nature of the educational resource without any reference to topics or disciplines. Thus, during the first iteration of testing and validation of the proposed meducator metadata schema, the field "content type" was included, and it was left as free text to fill (i.e., the initial decision was not to use any controlled vocabulary). The rationale of this decision was to gather data about the variability of the type of educational items that would be contributed by the content providers of the Consortium, and use this information to inform subsequent decisions about the meducator schema and the potential need to resort to controlled vocabularies.

### **3 Methodology**

#### **3.1 Deriving the requirements**

The first version of the schema, where content type was a free text field, was used by the Consortium Partners to describe the educational resources that each one had agreed to contribute to the project. This was done by resorting to MetaMorphosis, a social network based on ELGG platform [12], where the users from the participating institutions could register and contribute their resources by filling the metadata forms. The data from 350 forms contributed by about 100 users from the various institutions were analysed by means of content analysis.

From the analysis of how the users filled the field it was apparent a conflation between format (e.g. PPT presentation), educational content type (e.g., a tutorial, or a case study) and tool/interaction type. For example, we had users who referred to "webTraces" to indicate that their contents were images annotated with graphical feedback, since such content had been developed within the LCMS "WebTrace"[11]. This conflation of several aspects into the same field, pointed to the need to clearly differentiating "media/format" aspects from "content type", and to attempt a differentiation between the pedagogical role of the resource and the rhetoric of presentation/interaction.

Some semantic overlap with notion of "subject/topic" was noted in the way users filled the form, although this was not entirely unexpected. In addition, the analysis of the content to be shared revealed that, consistently with the aim of the mEducator project, the type of materials to be shared amongst were broader in scope, and would include design documents such as curricula, or elaborations over official documents (typically, clinical guidelines) to be considered jointly (i.e. as companion resources); other cases would include referral to general resources available on the web, where up to date materials would be issued regularly on a given topic or discipline.

The broad typology of sharing needs from the mEducator community target users was an additional reason to support the choice of resorting to a controlled vocabulary.

The general requirements for the controlled vocabulary were:

- Maximise reuse from existing vocabularies, also in the light to ensure machine processability and compatibility with existing standards.
- Cater for the need of sharing materials that might not be "strictly" educational. This implies complementing a focus on the instructional/pedagogical aspect of the resource (e.g. a tutorial) with a focus on the informational model (e.g. a "news", regardless of whether it is being delivered through TV, Internet, or Newspaper). A clear separation of the informational model from the media of delivery would facilitate repurposing across different media.
- In the characterization of the media, novel types emerging from the Web 2.0 paradigm should be considered (e.g., Wikis, blogs, and the like).
- Take into account the variety of pedagogical approaches that are in place in the community and are reflected in community specific "content types". This requirement might be referred to as pedagogical flexibility [13].
- Completeness, with respect to the specific "content types" in use in the community of medical educators.
- Support efficiency in the retrieval, but also in the annotation process, i.e., the vocabulary should be as lean as possible and understandable at a glance.

### **3.2 Related work**

A widely shared criticism of current e-learning standards is that they fail to address adequately the "learning" [14], and, in general, miss crucial information required for efficient searching and automated processing [13]. From the perspective of the e-

learning community, two works are especially relevant. In [13] the LOM "learning-resource type" field is criticised on the ground that the controlled values mix instructional (e.g., *exercise, simulation, experiment*) and format information (e.g., *diagram, figure, slide, table*) that need to be separated. This is exactly the situation in which we incurred by leaving the field free to be filled by the users. It should be noted that the same problem occurs in the HLOM extension, where both instructional (e.g., *tutorial, Virtual Patient, simulation*) vocabulary values and format values (e.g., *narrative text*) have been added. The solution proposed in [12] is an ontology, with root class Instructional Object. The philosophy of this work is to subsume under the subclass "Concept" the specific domain of interest (definition, facts, laws and processes) and to complement "Concept" with "Satellite" elements, a subclass that captures the instructional information associated to any concept of the domain (e.g., *real world problem, exercise, exploration, example non-example, explanation, remark, evidence, etc.*). Format and media aspects are not addressed in this ontology.

On a similar note, but with a different goal, i.e., supporting automated annotation, [15] proposes the ALOCOM ontology to capture both the content structure and the content type of a Learning Object (LO). The need to express the Structure of the LO, stems from having to deal effectively with the different granularity levels of a LO, which is seen as consisting of various "content fragment" (non further decomposable media); "content object" (an aggregation of fragments with added navigational features), and "learning object" (an aggregation of content object on a learning objective). The proposed classes for learning object are: *lesson, course, chapter, test and tutorial*. From the content type perspective, the pedagogical roles are captured in the "supporting" subclass of content object, and at the moment include: *example, exercise, reference, description, illustration, question, and answer*. This is somehow similar to the "satellite" subclass in [13].

There exist several other Resource Types and Media Type vocabularies. Some have been developed within broader metadata standard specifications, e.g., Dublin Core [16], some have been specifically developed by educational communities, such as the RDN/LTSN (Resource Discovery Network/Learning and Teaching Support Network) [17]; a comprehensive listing of the variety of vocabularies developed in the attempt to capture pedagogical aspects is provided in [18], where some general requirements that should inform their development are also provided.

Concerning the media, in addition to the MPEG-7 ontology [19] there are specialisations in narrower domains, e.g., the RDA/ONIX framework for resource categorization in the media publishing domain [20]; whereas another set of relevant terminology is available through the categories "publication format" and "publication component" widely accepted medical thesauri MeSH [21]. Another relevant work is the SIOC ontology [22], providing terms especially relevant to characterize emergent social media.

### 3.3 Design considerations

The methodology used to craft the taxonomies or the controlled vocabularies for mEducator was to analyze first existing vocabularies, taxonomies and ontologies, to evaluate their fit for reuse. Selection of the most appropriate terms was then

performed based on coverage of the educator requirements, and then adding the missing concepts emerged from the first field testing of the mEducator schema.

Critical analysis of the available resources, however, pointed out some shortcomings with respect to mEducator needs. A common problem was lacking expressivity in the description of learning/teaching activities implied or supported by the resource, other ones were too biased towards lecturing-based pedagogy or tutorial expository modes; others were too much oriented toward media fragments (e.g., [19]). Others would be too focused on various granularity levels in content organization (e.g. [15]) whereas others used the more practical approach of condensing in one term more than one aggregation level (e.g., "course/module/unit" in the RDN/LTSN vocabulary [17]). This latter approach was adopted as a design criteria, since it was deemed a good practice, alternative to pursuing generalisation into more abstract, catch all terms, that often tend to be semantically ambiguous. The expected benefit is lessening the cognitive load on the user without sacrificing precision of retrieval, at least as far as concerning identification of the nature of the resource.

Analysis of the identified sources mentioned in section 3.2 resulted in the selection of 45 concepts obtained by picking up from the sources, merging with mEducator community specific terms, and in some cases rewording some concepts. Nearly none of them were mutually exclusive; thus, from the user annotation perspective, the classification grid must enable multiple selection to characterise reasonably the resource to be shared.

In general, the term selection and refinement process was carried out taking into consideration that in the end, the resources will be indexed by the end-users and not by professional cataloguers, therefore attempting to obtain manageable lists (e.g., a too long list should be categorised in a way meaningful to the end-user), with terms readily visualised and familiar (ideally, self-explaining).

## **4 The mEducator Resource Type Taxonomy**

Resource Type in the context of the mEducator project is defined as:

*A classification of the Resource based on its informational qualities, taking into account the nature of the information provided, how it is organized, presented or collected (e.g., a tutorial, a textbook, a simulation) and the nature of the interactions that are expected to take place between the users and the content (e.g., reading, practicing, experimenting, assessing, etc.). Resource type includes all the artifacts that traditionally have a role in the context of educational practice, and any document, product or tool that has a role during the practice of healthcare professions and that can be either the focus or a supporting resource for the design of an educational experience (e.g., a laboratory test, or a clinical practice guideline).*

On purpose, aspects strictly related to media, format and implementations were not considered, to give primacy to the nature of the information provided by the resource. The 45 concepts were classified according to the following three broader categories:

### 1) Educational practice artefact

**Definition:** any document, product or tool designed and developed *specifically* for use in any phase of the teaching/learning process, including the supporting institutional and administrative activities, or generated as a result of the educational activity.

**Comment:** this category includes all the artefacts that have a role strictly within the context of educational practice, and as such, would be of no particular value in the context of practicing a healthcare profession (e.g., a Game, a Study Guide, a reading list)

### 2) Professional practice artefact

**Definition:** any document, product or tool that is used during the practice of healthcare professions.

**Comment:** The emphasis is on the artefacts that have their “raison d’être” in the professional context (e.g. a clinical record or diagnostic test) but that can be effectively introduced in the educational activities to design realistic, authentic learning experiences. Resources of this type might not be readily available in digital format; therefore the capability to identify and retrieve them would facilitate their repurposing to customize learning/teaching activities.

### 3) Reference Material

**Definition:** any collection of information that is not routinely used in professional/or educational practice, but it is used only for reference purpose.

**Comment:** for this category a complete mapping to a selection of MeSH terms was possible, and the Mesh definitions were reused.

To facilitate reuse, the vocabulary has been further organised into "Core Resource Type " and "Medical Resource Type" vocabulary, this latter being an extension of the Core Resource Type. Table 1 lists the terms under each category, and definitions for each term are provided in Appendix. It must be noted that a Resource can be indexed according to multiple terms from the same or distinct category; for example, a “course/module/unit” might contain a “simulation”, “problem/exercise with feedback” and “reading list”, and therefore should be indexed with these four terms.

#### 4.1 Comments on the vocabulary values: towards generalisation

It is interesting to note that a good proportion of the items that have been identified are quite general and would apply to any teaching/learning setting, regardless of the specific domain, although this analysis has not been done in the light of completeness with respect to other domains. It is easy to see that for many items collected under the Medical resource type, equivalent ones can be found in other domains.

In the educational practice artefacts, a Virtual Patient could be seen as a type of Simulation, and Clinical Case Study as a domain-based specialisation of Case Study; making these specialisations explicit as medical type extensions increases precision.

"Teaching file" was an interesting case of semantic differences across communities, since medical groups such as radiologists refer to a well established practice of creating files with Images to be discussed/diagnosed by the students, whereas other communities (paediatrics, general educators) used the term to refer to a set of organized resources meant for the teacher. Thus the approach was to keep the term as used by radiologists and include it in Medical Resource Type, and introduce the notion of Resource Pack that generalizes the second sense of "Teaching file" (see definition in Appendix).

**Table 1.** The mEducator Resource Type taxonomy. In appendix 1 definitions are listed.

<b>mEducator Resource Type Taxonomy</b>			
	<b>Educational Practice Artefact</b>	<b>Professional Practice artefact</b>	<b>Reference material</b>
<b>Core Resource Type</b>	Course/Module/Unit Lecture (recording) Lecture Slide/Presentation Lecture Notes/Handout Textbook/Chapter Tutorial Reading List Worked example Demonstration Simulation Case Study Practical Problem/Exercise Problem/exercise (with Feedback) Game/Serious Game Assessment Item/Instrument Student generated content Resource Pack Study Guide Syllabus/Programme/Curriculum Educational Policy	Professional Practice Guideline Scientific Journal Article Diagnostic algorithm Table Dataset Database Software	Atlas Bibliography Dictionary Encyclopaedia Handbook Index Legislation Laboratory manuals Terminology
<b>Medical Resource Type</b>	<b>Virtual Patient</b> <b>Teaching File</b> <b>ClinicalCaseStudy</b>	<b>Clinical Practice Guideline</b> <b>Diagnostic/Laboratory test</b> <b>Diagnostic/Laboratory test (annotated)</b> <b>Medical/Diagnostic algorithm</b> <b>Medical Form</b> <b>Clinical record</b> <b>Patient Education Handout</b>	<b>Anatomical Atlas</b> <b>Pharmacopoeias</b>

Concerning the "Professional Practice" collection of terms, it can be noted that, many specific terms can have a counterpart in other, highly specialized professional domains (e.g., law, economics, psychology, etc.) that have professional codes, ethics, practice guidelines, data collection tools, software to support the job activity. For example, a "Diagnostic algorithm" or a "Diagnostic/laboratory test" makes sense also in most engineering/technical disciplines, and, as such, could be part of other domain-specific extensions. Similar considerations apply for the Reference materials. As mentioned before, the reason why one would index and share these types of resources, that are somehow "raw materials" (not re-elaborated for educational purposes) is that they can be well suited to be the building blocks of constructivist and "authentic" learning experiences. Also, it might be argued, they can indirectly foster the process of community building (that is most often mediated by sharing tools and tricks of the trade) that can make a system for sharing resources successful or not.

## 5 The mEducator Media Type Taxonomy

The property "Media type" of the mEducator learning resource refers to the media type of the resource, or the media types embedded in it, if this is a complex one. It is possible to have multiple values for this field, e.g., an image sequence accompanied by an audio clip, or a blog post embedding a video clip. Media type may refer also to the packaging standards that are used to interoperate the resource across different learning platforms (e.g., SCORM package, MVP-Medbiquitous Virtual Patient) but leaves out details such as file digital format and size.

In order to facilitate identification of "fragments" for repurposing, and of classic e-learning "packages" for interoperability, Media Type has been organized in four categories: Primary Media, MultiMedia, Web/Social Media and Media Package. The Web/Social Media category reuses terms from the SIOC ontology [22].

Table 2 lists the terms selected for each category, whereas Table 3 provides the definitions for Primary Media and Multi Media

**Table 2.** The mEducator Media Type taxonomy.

<b>mEducator Media Type Taxonomy</b>			
<b>Primary Media</b>	<b>MultiMedia</b>	<b>Web/Social Media</b>	<b>Media Package</b>
Text	Interactive Multimedia	Site	SCORM
Image	Immersive Environment	Forum	IMS
Sketch/Graphical Annotation		Blog	DICOM
Animation		Post	MVP (VirtualPatient)
Audio		Podcast	
Video		Webinar	
3D Model		Wiki	

**Table 3.** mEducator Resource Type Taxonomy: definitions for Primary Media and MultiMedia

<b>Definitions for mEducator Media Type Taxonomy</b>	
<b>Primary Media Sub-headings</b>	
Text	A resource consisting primarily of words for reading. Example: an e-book without interactive features or multimedia elements. Comment: same as <a href="http://purl.org/dc/dcmitype/Text">http://purl.org/dc/dcmitype/Text</a>
Image	A static visual representation other than text Same as: <a href="http://purl.org/dc/dcmitype/StillImage">http://purl.org/dc/dcmitype/StillImage</a>
Sketch/Graphical Annotation	An image or text augmented by layer information. May combine attaching text labels to graphical elements is a natural visual notation that appears in many kinds of hand drawn diagrams, such as those appearing in user manuals, to indicate part-whole relationships. Narrower than Image.
Animation	A series of visual representations imparting an impression of motion when shown in succession. Include visual output from a simulation. Note: specializes <a href="http://purl.org/dc/dcmitype/MovingImage">http://purl.org/dc/dcmitype/MovingImage</a>
Audio/Sound	A resource primarily intended to be heard. Same as <a href="http://purl.org/dc/dcmitype/Sound">http://purl.org/dc/dcmitype/Sound</a>
Video	A camera-based recording of visual and audible components. Audio might not be included. Note: specialises <a href="http://purl.org/dc/dcmitype/MovingImage">http://purl.org/dc/dcmitype/MovingImage</a>
3D Model	A three-dimensional digital representation of an object.
<b>MultiMedia - Subheadings</b>	
Interactive Multimedia Resource	A resource that aggregates any combination of text, audio, still images, animation, video, and requires interaction from the user to be understood, executed, or experienced. Examples include AJAX Web pages, Applets, e-books with multimedia elements.
Immersive Environment/Virtual Reality	An artificial, interactive, computer-created scene or "world" within which a user can immerse themselves. May resort to gestural controls, motion tracking, and computer vision respond to the user's actions and movements. Examples include Second Life, or a virtual reality simulation of surgical procedures. Narrower than Interactive Resource

## 6 Implementation

IMS Vocabulary Definition Exchange (VDEX) [23] is a standard currently being used for the interoperability of eLearning vocabularies; however, since mEducator places emphasis on compatibility with Semantic Web application, and one of the solutions that is going to be tested is based on Semantic Web Services and Linked Data, SKOS [24] was chosen to represent mEducator vocabularies/taxonomies, also in accordance with the JISC recommendations on terminology services and technologies [25]. SKOS (Simple Knowledge Organization System) is the W3C specification for representing thesauri, classifications, subject headings, taxonomies, and folksonomies, and fully supports multilingualism.

The vocabularies namespaces are:

- <http://purl.org/educator/resourceType/>
- <http://purl.org/educator/mediaType/>

whereas the RDF files are available at:

- <http://www.educator.net/mdc/resourceType.rdf>
- <http://www.educator.net/mdc/mediaType.rdf>

respectively.

The schema used for attaching Resource and Media types to the educational resources, is the mEducator schema that can be found at [www.purl.org/educator/ns](http://www.purl.org/educator/ns). In Figure 1 an excerpt of an instance of a resource that uses the implemented vocabularies is shown.

*Resource "Biomolecular Structure Function analysis practicals" type is "Practical", "Simulation", "Software", "Resource Pack"*

```
<mdc:resourceType rdf:resource="http://purl.org/educator/resourceType#practical"/>
<mdc:resourceType rdf:resource="http://purl.org/educator/resourceType#simulation"/>
<mdc:resourceType rdf:resource="http://purl.org/educator/resourceType#software"/>
<mdc:resourceType
rdf:resource="http://purl.org/educator/resourceType#resourcePack"/>
```

*Resource "Biomolecular Structure Function analysis practicals" media is "animation", "3D model":*

```
<mdc:mediaType rdf:resource="http://purl.org/educator/mediaType#3DModel"/>
<mdc:mediaType rdf:resource="http://purl.org/educator/mediaType#animation"/>
```

**Fig. 1.** An example of the instance section characterizing the learning resource "Biomolecular Structure Function analysis practicals". **mdc** is the alias of the mEducator schema.

## 7 Discussion

Resource Type and Media Type have been designed to work jointly to profile the resources. Another option could have been to include in the medical resource type extensions some other specific cases where the Resource Type implies the Media, as is the case for VPs (to a certain extent), or for traceable medical images (that imply both images and graphical annotation). This might lead to some redundancy, but the trade-off could be achieving more immediate recognition of a concept from the users. This possibility will be re-evaluated after completing the testing of the vocabulary in its current form with the end-users.

Resource Type and Media Type are not the only (mandatory) controlled vocabularies that were adopted in the mEducator schema. A richer profiling of the learning resource is achieved, with respect to controlled vocabularies by properties regarding the IPR licensing schema (property reuse and values from Dublin Core), educational level of the intended audience and learning outcomes taxonomy. This latter one was developed anew and given its specificity for the medical domain its description is outside the scope of the paper. However, it should be noted that this learning outcomes taxonomy plays an important role, similar to the characterization

of the content with respect to the domain (e.g., fact, procedure, law, etc.) included in the ontologies in [13] and [15].

The development of the proposed vocabularies can be considered from two angles: 1) an exploration of what constitutes a content type from an educational perspective, and 2) an attempt to a characterization that is meaningful for the community of medical educators and that is useful in terms of assisting the retrieving of content with some repurposing goal in mind. To this aim it should be remarked that the level of granularity that has been adopted has been very much affected by the notion of *repurposing* that is key to mEducator. In fact, repurposing is a middle ground between the notion of reusing a learning object in its entirety, and reusing some of its media fragments or content fragments [15], this latter requiring a much more fine-grained approach to annotation. On the other hand, *repurposing* presupposes that a fairly complex and elaborated resource keeps its general connotation (e.g., a virtual patient), but is repurposed to take into account different contexts, e.g., different cultures, different languages, different measurement systems, or to cover yet an additional learning outcome. In this sense it seems reasonable to annotate at a level that does not explicitly take into account content structure, and yet may allow specifying the nature of the components included in the resource.

Concerning the practical ways to support the adoption and the correct use of controlled vocabularies (to reap their benefits), issues more directly related to the users must be taken into account. One lessons from the development of the vocabulary activity was that any classification system, no matter how much "self-explaining" its categories and terms are, becomes increasingly meaningful to the user engaged in the indexing process as long as he or she gains a full view of the facets used in the classification. Interestingly, this leads us into the realm of user-interface design, where the recommendation is to show contextually all the top and first level categories from which the user can select their indexes. This requirement is even more important when the coexistence of the folksonomic approach is allowed (i.e., a user who does not recognize a suitable controlled term to describe the resource can enter a new term), as is the case in our project. For example, if the user thinks that resource content type is a WIKI s/he should be immediately aware that that option has been contemplated in the Media Type section.

## 8 Concluding remarks

We are currently involved in two lines of activity. The first one is concerned with evaluating and validating the proposed vocabularies, as implemented in the two mEducator solutions to share medical content. The proposed vocabulary will be then presented to Medbiquitous ([www.medbiq.org](http://www.medbiq.org)), the international consortium dedicated to develop information technology standards for healthcare education and competence assessment, as part of the goals of mEducator.

The second line of action regards the formalisation of the mappings of the chosen terms towards DC, HLOM, MeSH and other sources used in developing the vocabulary. This will be done taking into account also the output of a recent JISC project on the Vocabulary Mapping Framework [17], where a mapping of

vocabularies from source standards is provided, with the aim to support the building of transformations between any of them. It is expected that through this activity and through the RDF linking of related vocabularies, a valuable resource to assist in the tailoring of vocabularies for specific communities will be made available.

From a research perspective, we plan to investigate how the network of mappings that becomes realised by connecting vocabularies can complement the information that can be extracted by mining the free text fields of the meducator schema, as proposed in [27], to further facilitate the resource retrieval process.

## Acknowledgements

This work was supported by the project mEducator (Multi-type Content Sharing and Repurposing in Medical Education), funded by the eContentplus Programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable (Grant ECP 2008 EDU 418006)."

## References

1. Lambe, P.: *Organising knowledge: taxonomies, knowledge and organisational effectiveness*. Chandos Publishing, Oxford (2007)
2. Kiu C., Tsui, E.: TaxoFolk: a hybrid taxonomy-folksonomy classification for enhanced knowledge navigation. *Knowledge Management Research & Practice* 8, 24--32 (2010)
3. Bamidis P.D., Kaldoudi E., Pattichis C.: From Taxonomies to Folksonomies: a roadmap from formal to informal modeling of medical concepts and objects". In: 9th IEEE International Conference on Information Technology and Applications in Biomedicine, Larnaca, Cyprus, 5-7 November (2009)
4. Heath B.P., McArthur, D.J., et al.: Metadata lessons from the iLumina digital library. *Commun. ACM* 48, 7, 68--74 (2005)
5. Bizer, C., Heath, T. et al.: Linked data - The Story So Far. Special Issue on Linked data, *International Journal on Semantic Web and Information Systems (IJSWIS)* (2009).
6. Mannens, E., Troncy, R. et al.: Automatic metadata enrichment in news production. In: 10th Workshop on Image Analysis for Multimedia Interactive Services, pp. 61--64, IEEE (2009)
7. Dietze, S., Benn, N. Yu, H. et al.: Comprehensive service semantics and light-weight Linked Services: towards an integrated approach". In: 8th International Semantic Web Conference (ISWC 2010), Shanghai, China, November (2010)
8. [http://www.medbiq.org/std\\_specs/standards/index.html#HCLOM](http://www.medbiq.org/std_specs/standards/index.html#HCLOM)
9. Mitsopolou, E., Taibi, D., Giordano, D. Dietze, S., Yu, H.Q., Bamidis, P., Charalampos, B., Woodham, L: Connecting medical educational resources to the Linked Data cloud: the mEducator RDF Schema. In: 1st Linked Learning Workshop, ESWC, Heraklion, Crete, May 29, (2011)
10. Kaldoudi, E., Bamidis, P., Papaioakeim, M., Vargemezis, V.: Problem-Based Learning via Web 2.0 Technologies. In: 21st IEEE International Symposium on Computer-Based Medical Systems, pp. 391--396, IEEE (2008)
11. Giordano, D., Leonardi R.: Web-trace and the learning of visual discrimination skills. In: 1st International Workshop on Pen-based Learning Technologies, Catania, Italy. 24-25 May, pp. 80-85.. IEEE CPS (2007)

12. Kaldoudi, E., Dovrolis, N., Konstantinidis, S., Bamidis, P.D.: Social networking for learning object repurposing in medical education”, J. Inform Techn Healthcare, 7, 4, 233--243 (2009)
13. Ullrich, C.: The learning-resource-type is dead, long live the learning- resource-type!," Learning Objects and Learning Designs, 1, 1, 7--15 (2005)
14. Jonassen, D., Churchill, D.: Is there a learning orientation in learning objects?" International Journal on E-learning 3(2): 32-41 (2004).
15. Jovanovic, J., Gasevic, D., Devedic, V.: Ontology-based Automatic Annotation of Learning Content. International Journal on Semantic Web and Information Systems, 2, 2, 91--119 (2006)
16. <http://dublincore.org/documents/dcmi-type-vocabulary/>
17. Barker P., Cross, P., Fernandez, A. et al. : RDN/LTSN resource type vocabulary- v.1.0. Available at: <http://www.intute.ac.uk/publications/rdn-ltsn/types/>
18. Currier S., MacNeill S., et al.: Vocabularies for describing pedagogical approach in e-learning: a scoping study. In: DC-2006 International Conference on Dublin Core and Metadata Applications: metadata for knowledge and learning.
19. MPEG-7 Ontology, <http://www.w3.org/2005/Incubator/mmsem/XGR-mpeg7/>
20. Dunsire, G.: Distinguishing Content from Carrier: The RDA/ONIX Framework for Resource Categorization. D-Lib Magazine 13, 1/2 (2007)
21. MeSH, <http://www.ncbi.nlm.nih.gov/mesh>
22. The SIOC ontology. <http://sioc-project.org/>
23. [http://www.msglobal.org/vdex/vdexv1p0/imsvdex\\_bestv1p0.html](http://www.msglobal.org/vdex/vdexv1p0/imsvdex_bestv1p0.html)
24. SKOS, Simple Knowledge Organization System, [www.w3.org/2004/02/skos](http://www.w3.org/2004/02/skos)
25. Tudhope D., Koch, T., Heery R.: Terminology Services and Technology - JISC state of the art review, <http://www.ukoln.ac.uk/terminology/JISC-review2006.html> (2006)
26. The vocabulary mapping project, <http://cdlr.strath.ac.uk/VMF/index.htm>
27. Giordano, D., Faro, A., et al.: Feeding back learning resources repurposing patterns into the “information loop”: opportunities and challenges. In: 9th IEEE International Conference on Information Technology and Applications in Biomedicine, Larnaca, Cyprus, 5-7 November (2009)

## Appendix: Definitions for Educational practice artifacts

<b>Definitions for Educational Practice Artifacts</b>	
Course/Module/Unit	A sequence of activities designed to advance student skills, knowledge, and attitudes in a particular discipline and to help students meet requirements as prescribed in a curriculum. Example: a course delivered through a LCMS (e.g., Moodle)
Lecture (Recording)	The video or audio recording, or the transcript of the exposition of a given subject delivered before an audience, for instructional purposes. Example: a speaker presentation at a conference
Lecture Slide/Presentation	Materials that are projected to support the delivery of lectures or presentations in a module or a course. Example: an annotated PPT presentation, an image, picture or diagram.
Lecture Notes/Handout	Notes containing topical information prepared as study/reference material for lecture(s) or a course. Example: a document distributed to the participants in a seminar
Textbook/Chapter	Book or chapter of a book explicitly meant for the study of a specific subject, characterised by the systematic presentation of knowledge about the subject.
Tutorial	A resource that provides guided, practical information about a subject

	Example: on-line primer to a software for statistical analysis
Reading List	A list of recommended or required sources which provide additional information on the subject being studied
Worked Example	A written problem or exercise designed to illustrate step-by-step how to perform a task or how to solve a problem.
Demonstration	A video or audio recording, or a transcript, demonstrating a skill or a procedure in practice.
Simulation	A representation of a process, activity or situation. Designed to support problem-based learning or exploratory learning, or the comprehension of the dynamics of a bio-physical or other complex process. Example: a simulated patient interview, a software simulating a spring oscillation at the varying of the spring parameters and the force applied to the spring.
Virtual Patient	An interactive computer simulation that allow the learner to take the role of a health care professional and develop clinical skills such as making diagnoses and therapeutic decisions.
<b>Teaching file/Clinical Case Study</b>	The presentation of a clinical case through diagnostic imaging modalities. The images contained in the file are typically supplemented with the following information; case title, history/presentation, findings, diagnosis, discussion.
Case Study (Non Clinical)	A detailed account or a process or activity, prepared to assist the student in determining what factors led to its success or failure. Example: a detailed analysis of an institutional change/innovation, a description of a critical incident.
Practical	An activity to develop the practical skills of a subject, or an examination of such practical skills. Example: an activity to be carried out in the laboratory.
Problem/exercise (no feedback)	A task, or other effort to be performed by the student to develop, maintain, or increase skill or cognitive abilities. May include Projects and Fieldwork
Problem/exercise (with feedback)	A task, or other effort to be performed by the student to develop, maintain, or increase skill, including problem-solving, and for which feedback is readily available to the student.
Game/Serious Game	A structured activity, usually undertaken for enjoyment, used to teach about a subject or to develop specific skills while playing.
Assessment Item/Instrument	An item, activity, system or instrument designed to measure student learning. May include question/answers pairs or practical skill demonstration. Example: Self-assessment questions, Multiple choices, Objective Structured Clinical Examination (OSCE)
Student generated content	Any content generated by a student of a group of students as a result of coursework. Can be used to exemplify good and bad practices, or shared for reference. Examples: entries in a Wiki, final coursework, a discussion thread in a Forum
Resource Pack	A complete package for a course, module or unit including learning resources of several types and supporting documentations (e.g., teacher/learning instructions). Sometimes referred to as "teaching file" in educational contexts other than medicine.
Study Guide	A written guide created to provide direction, and point out critical information to the students. It may include techniques for problem solving and hints to manage the study process (e.g. topic sequencing and timing, learning and testing strategies)
Syllabus/Programme/Curriculum	A document describing the contents of units, courses, and courses of study. May include information about access requirements, delivery, activities and assessment modalities.
Educational Policy	A document containing statements about how an educational system should operate.

# Applying and Extending Semantic Wikis for Semantic Web Courses

Lloyd Rutledge<sup>1</sup> and Rineke Oostenrijk<sup>2</sup>

<sup>1</sup> Faculty of Informatics, Open Universiteit  
Amsterdam, The Netherlands

Lloyd.Rutledge@ou.nl

<sup>2</sup> Netherlands Environmental Assessment Agency  
Bilthoven, The Netherlands

**Abstract.** This work describes the application of semantic wikis in distant learning for Semantic Web courses. The resulting system focuses its application of existing and new wiki technology in making a wiki-based interface that demonstrates Semantic Web features. A new layer of wiki technology, called “OWL Wiki Forms” is introduced for this Semantic Web functionality in the wiki interface. This new functionality includes a form-based interface for editing Semantic Web ontologies. The wiki then includes appropriate data from these ontologies to extend existing wiki RDF export. It also includes ontology-driven creation of data entry and browsing interfaces for the wiki itself. As a wiki, the system provides the student an educational tool that students can use anywhere while still sharing access with the instructor and, optionally, other students.

**Keywords:** wikis, semantic wikis, distance learning, linked data

## 1 Introduction

A wiki helps multiple people at a distance to cooperate in writing and sharing documents. The Semantic Web enables people of multiple institutions to create data that machines on the web can exchange and share. A semantic wiki combines these types of systems: it lets multiple people at multiple places cooperate in making documents and data, along with the programs that retrieve, organize and present this data. Distance learning is another form of distant collaboration, in which not just documents and data are exchanged, but also skills and knowledge, and the means of conveying them. The ability of semantic wikis to enable distance and collaborative development of Semantic Web systems has special ramifications for distant learning about the Semantic Web.

The Master’s course *Capita Selecta Thema - Semantic Web* of the Informatics faculty at the Open Universiteit applied semantic wikis as distance education tool. Now the software and the didactic technique developed for and used in this course are being adapted and extended for an upcoming Bachelor’s course on the Semantic Web. As with most Open Universiteit courses, these Semantic Web

courses apply distance learning techniques. Students in both courses develop, as part of their graded assignments, Semantic Web systems on semantic wikis. This work explains and demonstrates this application of semantic wikis as course material in distance learning about the Semantic Web.

This paper starts with a discussion of related work in semantic wikis. We then describe Semantic Web educational activities at the Open Universiteit and our use of semantic wiki's in them. The last section presents OWL Wiki Forms, an extension of existing semantic wiki's developed for Semantic Web courses.

## 2 Related Work

A variety of related research and systems apply to our wiki-based approach for developing courseware for distant learning about the Semantic Web. This section starts by presenting a broad selection of concepts and tools that apply to Semantic Web courses. We then present the central tool we apply: Semantic MediaWiki. This section concludes with the wiki interface tool we use most: Semantic Forms.

### 2.1 Foundational Semantic Web tools

Our application of semantic wikis as Semantic Web courseware is influenced and supported by Semantic Web tools beyond wikis. This section starts by presenting Protégé, an ontology editor that many Semantic Web courses use. We then present a progression of different systems that lead to semantic wikis.

Protégé-OWL (or just *Protégé*) is a GUI-editor for developing ontologies, queries and rules in Semantic Web formats [7]. Protégé supports development and visualization of RDF, RDFS, OWL and SWRL. It is easily installed and free of charge. Protégé is frequently used in courses about the Semantic Web. Protégé is a good implementation of many technical Semantic Web components. However, it functions mostly as a stand-alone back-end developer's interface, lacking as such a direct end-user interface.

*Semantic browsers* such as Tabulator [3] form one type of easily accessible end-user interface to Semantic Web data. While they offer search and analysis interfaces in a variety of ways, they usually have a view for a single resource, shown as a table of the resource's properties with their assigned values. In earlier work, we presented a semantic browser whose focus was generating outline structure, whose links than lead to such tabular annotation displays for given resources [12].

*Wikis* provide an important and large-scale online source for highly collaborative authoring of information. Initiatives such as Wikiwijs [13] provide platforms on which teachers share education material with each other. Wikis are also helpful for distance learning because their interfaces make facilitate online collaboration.

The best-known wiki is *Wikipedia*<sup>1</sup>. Wikipedia's *infoboxes* provide annotations of pages using a simple editing format, which results in tabular displays of facts for given Wikipedia pages. Infobox displays resemble semantic browsing in their structured presentation of properties and linked values. An important difference, however, is that each page's infobox is human-edited separately, whereas semantic browsers automatically generate their displays.

The *DBpedia* project brings Wikipedia's infoboxes into the Semantic Web by converting them to RDF and providing Semantic Web-based querying and browsing interfaces to the infobox data [1]. DBpedia data is then integrated into the Linked Open Data dataset [4]. All this makes the infobox format on Wikipedia a large and important source of annotations on the Semantic Web. The most important technology for wikis is *MediaWiki*<sup>2</sup>, which is used for, among other things, Wikipedia.

## 2.2 Semantic MediaWiki

*Semantic wikis* are wikis that add crowd-editing of data to that of documents. With semantic wikis, people can add, manage, distribute and present data with wiki interfaces as they do with documents, but so that machines process the data as well. The most important software for semantic wikis is *Semantic MediaWiki (SMW)* [9], an extension of MediaWiki. This tool supports making systems that follow the vision and high-level functionality of the Semantic Web. However, the individual Semantic Web components only become explicitly visible in SMW's RDF export feature.

Semantic MediaWiki, like MediaWiki, is free of charge and open source, and therefore not only easily acquired, but also easy to build further as assignments for student projects. There are at least 216 active, public Semantic MediaWiki installations online with several types of applications, for fields such as companies, culture, training and management, and 33 specifically for education [14]. The industrial use of SMW for data management is growing of, as shown by the development and adoption of industry-oriented Semantic MediaWiki extensions such as SMW+<sup>3</sup>.

Table 1 shows how various Semantic Web constructs are mimicked by or directly implemented in features from the incremental layers of SMW technologies. Semantic MediaWiki starts by recognizing that MediaWiki itself has features similar to Semantic Web components, as indicated in the "MediaWiki" column of Table 1. The next column show features of the SMW extension that build on top of MediaWiki to add Semantic Web features. The final column shows the indirect Semantic Web support provided by Semantic Forms, which is an extension on top of Semantic MediaWiki.

First, SMW treats each MediaWiki link like a triples joining the current page with the link end via an unspecified property. The foundation of Semantic

---

<sup>1</sup> <http://www.wikipedia.org>

<sup>2</sup> <http://www.mediawiki.org/>

<sup>3</sup> <http://www.ontoprise.de/en/solutions/smw-plus/>

**Table 1.** Ontological features implemented in progressive layers of Semantic MediaWiki technologies

Feature		MediaWiki [9]*	SMW [9]*	Semantic Forms [8]
RDF	triples	[[xxx]] link (untyped), infoboxes [1]	[[ <i>pred</i> : <i>obj</i> ]] queried and exported as triple	infobox assigns SMW annotation
	type	category	export	infobox assigns category
	Property		property	infobox/form row
RDFS	datatype		SMW <i>datatype</i> : affects query, display, sorting and export	specialized input field
	Resource	page	export as URI from wiki	each page annotated by infobox/form
	Class	category	exported	usually own infobox/form
	domain			property assignment to infobox/form
	range			autocompletion on category "Has default form" property
	subClassOf	categories in categories	in queries and export	
	subPropertyOf		SMW subproperty: queries and export	
OWL	inverseOf		invert-property template [15]	
	sameAs, equivalent-Class/Property	#REDIRECT	in query processing [15] export uses SMW's <i>equivalent URI</i> [16] and <i>ontology import</i> [15]	

*\*all content of this column cites this reference unless stated otherwise*

MediaWiki is its typed links, which extend MediaWiki link syntax by adding specified properties. This handling as triples becomes literal in the RDF export feature.

In addition to this Semantic Web-based foundation of typed links, SMW has many smaller, albeit sometimes more indirect, equivalences of Semantic Web functionality. SMW offers a variety of datatypes. These influence how their data is displayed and sorted, as well as explicit coding in RDF export. MediaWiki's special page namespace for categories offers a rough equivalent of Semantic Web classes in internal behavior. SMW extends this by class-like handling of categories in queries, and by explicitly making categories be classes in RDF export. Semantic MediaWiki adds a wiki page namespace for properties, whose mem-

bers then function as equivalents of Semantic Web properties, and are exported explicitly as such.

Categories of categories in MediaWiki have subclass-like inheritance in internal Semantic MediaWiki queries. SMW has its own subproperties, with similar internal inheritance, and also with explicit RDF export. The `invert-property` template distributed with Semantic MediaWiki provides interface behavior similar to that indicated by `owl:inverseOf`. MediaWiki's `#REDIRECT` is treated by SMW queries in some ways as `owl:sameAs`. Most important for the ability to link internal data to the Linked Data cloud are Semantic MediaWiki's import ontology and equivalent URI functions, which link given internal pages to previously existing external Web resources.

The Semantic MediaWiki Linked Data extension (*SMW-LDE*) is an extension on top of SMW+ that gives a wiki improved access to external linked data [2]. This improved access includes unification of data from multiple external sources in internal querying and in other aspects of the wiki interface. SMW-LDE illustrates the growing role of semantic wikis for linked data. While such external data integration is an important part of Semantic Web systems in general, our work here focuses more on ontology modeling on the wiki and the exporting and integration of data from these ontologies made on the wiki.

### 2.3 Semantic Forms

*Semantic Forms* (here just “*Forms*”) is an extension on top of Semantic MediaWiki that facilitates the entry of data and automates its presentation for browsing [8]. Data entry facilitation comes as a form-based user interface for adding annotations to wiki pages. Semantic Forms then generates presentations of this form-entered data for each wiki page in tables similar to Wikipedia infoboxes, providing automatically rendered semantic browsing of data in the wiki. The rightmost column of Table 1 shows how Semantic Forms implements certain Semantic Web system features. Here we make our own observations and comparisons, since Forms does not explicitly target Semantic Web functionality but instead implements more general data management.

The *infoboxes* in Semantic Forms are MediaWiki templates. MediaWiki *templates* serve as functions that accept parameters in generating wiki page code. Each parameter in a Forms infobox template corresponds with a SMW property. A call to an infobox template from a wiki page generates SMW code that assigns the parameter value to the given property for that page.

Forms offers special form-based interfaces to let the user enter the specifications that drive the generation both of these templates for infobox displays and of the forms that populate them. Through these special forms, the user indicates which properties appear in each infobox/form pair, and optionally which category the infobox puts its pages into. Semantic Forms applies the SMW datatype of each property to determine interface details of how its forms accept data input for that property. By specifying in this manner the properties asked for in each form and the appearance of each type of infobox, the user sets up the equivalent of a data model for the wiki.

Some Semantic Form features reflect certain aspects of the properties `rdfs:domain` and `rdfs:range`. That a property is assigned to a form and an infobox suggests that the form's class is included in its domain. Similarly, Forms's "autocompletion on category" parameter of form value input field specifications suggests a `rdfs:range` for the property. In addition, some `rdfs:range` interface behavior shows in the assignment of default forms to pages linked from a given property that have not yet been created.

These features of MediaWiki, Semantic MediaWiki and Semantic Forms offer a semantic browsing environment with forms for entering data. Installations of these tools let users create and export RDF code for integration with Semantic Web sets such as the Linked Data Cloud. In addition, administrators can fine-tune their installations so that the user interface reflects additional constructs from the Semantic Web. In the next section, we discuss our application of these tools to Semantic Web educational activities. The section after that presents the OWL Wiki Forms extension of these tools, for use in upcoming courses.

## 2.4 Ontowiki

There are other semantic wikis beyond Semantic MediaWiki. One that is particularly relevant to this work and its venue is OntoWiki [6]. While, the collection of semantic features OntoWiki offers is similar to that of Semantic MediaWiki and its extensions, it is not built on top of MediaWiki. Advantages that OntoWiki has SMW include smoother integration with existing Linked Data sources and direct internal support of OWL constructs. For our work here, we apply Semantic MediaWiki and its extensions because they comprise a larger development endeavor with a wider variety of data processing features.

## 3 Applying Semantic Wikis to Semantic Web Courses

The semantic wiki technologies from the previous section apply to making a variety of Semantic Web systems. This section describes our application of these technologies to administer classes and student projects about the Semantic Web.

### 3.1 Bachelor's Thesis Project

The *Semantic Friendly Forms* Bachelor's thesis project at the Open Universiteit built an extension of Semantic Forms [5]. The primary functionality it adds is a form-based interface for modifying pre-existing code for input forms and data presentation tables. With just Semantic Forms, on the other hand, the form pages for page forms and infoboxes only created them. Any subsequent changes were edited manually.

Semantic Friendly Forms also provides forms for assigning certain RDF, RDFS and OWL properties to any property the semantic wiki uses. These Semantic Web standard properties are `rdf:type`, `rdfs:label`, `rdfs:comment`,

`rdfs:domain`, `rdfs:range` and `owl:isInverseOf`. The form also lets users indicate that the current property is an `owl:SymmetricProperty`. In addition, Semantic Friendly Forms processes `rdfs:domain` annotations to determine which properties to include in autogeneration of forms and infoboxes.

Semantic Friendly Forms is implemented as PHP encoding of MediaWiki “Special pages”, just as all of Forms and much of SMW is.

### 3.2 Master’s Course

The Open Universiteit recently gave a one-time Master’s “Capita Selecta Thema” course about the Semantic Web [11]. This course used Semantic MediaWiki, along with Semantic Forms and Semantic Friendly Forms, as the core with which students share their data and interfaces with the teacher, and with each other [11]. It was also used for the general course administration and as a platform for large student projects. Validation and some higher-level Semantic Web processing was carried out by external tools, including Protégé.

The challenge for each student was to make a system with the functionality and interface of the Semantic Web. Each student developed an ontology in OWL for their project on the semantic wiki. These ontologies then became populated and presented with Semantic Forms-based interfaces that the students made for them, in part with Semantic Friendly Forms. Each project also produced additional interfaces for its data in the form of SMW query result tables and code for showing them with the various presentations formats SMW offers. Finally, the students set up their projects so that the RDF export feature of Semantic MediaWiki export each as an RDF file that confirmed to their OWL-defined ontology.

The shared use of the course wiki allowed students to share their work in progress with the teacher. The teacher could access, use, and even edit the students systems at any time from any web browser. In this Master’s course, all students worked on different project with different topics on a common wiki. Students could thus see each other’s projects and code. In addition, student project sometimes used components of other student projects, as was encouraged.

Another communal online aspect of the course was its series of online sessions. We used the online lecture tool Elluminate<sup>4</sup> for a series of lectures. In these sessions, both teachers and students gave presentations. In addition, students were able to use Elluminate’s Application Share tool to present and give demos of their projects on the common semantic wiki. Fig. 1 has a reproduction of an online session of this course, with demonstrator of a student’s project on the wiki. Other forms of communication in the course are email, a discussion group, individual student-teach online sessions with demos and, of course, the wiki itself.

The definition of ontologies was a core part of each project. The students defined these in OWL as separate OWL files. They then used Semantic Forms and Semantic Friendly Forms to build the user interfaces that provided entry,

---

<sup>4</sup> <http://www.illuminate.com/>

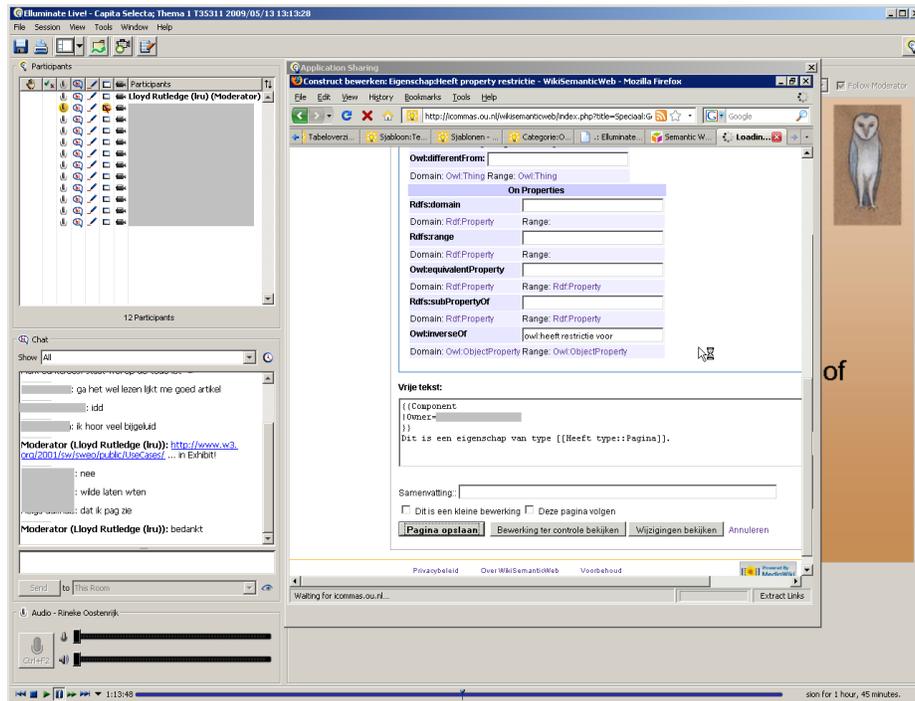


Fig. 1. Screen display of an online class with a semantic wiki

browsing and RDF exporting of data conforming to these ontologies. The ability of Semantic Friendly Forms to assist in the modification of existing infobox templates and forms was exploited by the students in building and implementing their ontologies. However, the defining of the data interface and of the OWL code it reflects remained technically separate tasks.

### 3.3 Master's Course Project on OWL

One of the student projects from this Master's course was about OWL itself [10]. This project provided form interfaces for annotating wiki pages with OWL properties and classes. It thus enables users to enter OWL-defined ontologies as SMW pages and Forms infoboxes. The project also processed some of these OWL constructs for OWL-related inferencing inside the wiki. As with Semantic Friendly Forms, this project offers a form page for defining properties. It also has form pages for annotating classes, restrictions and OWL properties on any resource. These four forms provide annotation with the majority of RDF, RDFS and OWL constructs. This project provided the entry of OWL-defined ontologies inside the wiki instead of as external files. However, these internally defined ontologies still had no automated influence on the creation of the entry forms and infoboxes for data populating these ontologies.

Like all students in the course, this student had no administrator access to the computer hosting the wiki while developing this project. Thus the project involved no PHP code. Instead, all functionality was encoded as SMW page and template code. While programming this functionality in PHP instead could improve performance, this project shows how much the wiki interface alone can define such interaction and processing of a semantic wiki installation.

### 3.4 Bachelor's Course

The Open Universiteit is developing a new bachelors course about the Semantic Web <sup>5</sup>. The techniques applied in this course are for the most part adaptations and extensions of those for the Master's course. One technical extension is of the semantic wiki course tools, which we package in the tool OWL Wiki Forms, described in the next section. This extension adds internal wiki interface behavior that more directly reflects the Semantic Web constructs used.

The Bachelor's course follows much of the approach of the Master's course. This includes communication via e-mail, discussion groups, wikis, individual on-line meetings and schedule online lectures. As with the Master's course, the core of the Bachelor's course is that each student builds a Semantic Web system on a semantic wiki. A key difference is that each student has the same assignments, and each student implements these on their own wiki, which the teacher can also access. In addition, all students can access a common course wiki for general course information as well as manuals and examples of the semantic wiki technology the course applies.

While the Master's course projects were general systems with high-level features, the Bachelor's course assignments are smaller in scale and focused more on understanding and applying individual components of Semantic Web technologies. This motivates one requirement in developing OWL Wiki Forms: that its interface directly reflects more individual Semantic Web technology components.

As with the Master's course, the Bachelor's courses system development is split up into assignments that represent different levels of Semantic Web technologies. These are: RDF and data, SPARQL and data presentation, RDFS, OWL and rules. When each assignment is given a passing grade, the starting point of the next assignment, which is in part the teacher's answer to the previous assignment, is loaded onto the student's wiki. In this way, all students start each assignment with the same foundation.

## 4 Extending SMW for Semantic Web Courses

This section presents the software that we developed for use in courses about the Semantic Web. The features and techniques from the software developed for the Semantic Friendly Forms project and the Master's class OWL project are unified and applied to the development of new software, *OWL Wiki Forms*

---

<sup>5</sup> <http://www.studieaanbod.ou.nl/T64211.htm>

**Table 2.** Ontological features implemented in OWL Wiki Forms

Feature		Ontology editing	Browsing	Data Input
RDF	Property	inherited by object and datatype properties		
RDFS	datatype	pull-down menu on property form	associated SMW datatype sets display	determines Forms input type
	Resource	own form		
	Class	own form	determines ... infobox form	
	domain	property form value fields	determines row in ... infobox form	
	range	autocomplete on class	new page links to form	value autocompletion
	subPropertyOf	property form value fields		
OWL	inverseOf	value fields	shows in both	
	equivalent-Property	autocomplete on property		
	ObjectProperty	own form		
	DatatypeProperty	own form		
	Symmetric-Property	checkbox in property form	infoboxes	
	Transitive-Property		<i>shows in chain of infoboxes</i>	
	(Inverse-)Functional-Property			
	Restriction	own form		

*italics = still under development*

(*OWF*). OWL Wiki Forms focuses on direct support for standard Semantic Web constructs in the wiki interface itself. Table 2 shows how OWF provides support for various Semantic Web components. It identifies four areas of support: *ontology editing*, *querying*, *browsing* and *data input*. Ontology editing consists of form-based interfaces for defining ontologies inside the wiki. Semantic Media-Wiki queries should return not only the triples asserted in the wiki page code and infobox parameters but also the triples inferred by processing the relevant RDF, RDFS and OWL code.

OWL Wiki Forms also generates data display infobox based on the ontologies entered, which provide central browsing interface for the wiki. The tool similarly generates data input forms. Each of the following subsections discusses one of these semantic wiki interface components.

#### 4.1 Ontology editing

The editing of ontologies is an important part of a complete Semantic Web system. One aim of this work is thus to include ontology editing features in OWL Wiki Forms. Its primary interface for letting user's edit ontologies is encoded

using Semantic Forms. This gives the user a form-based interface for entering annotations using classes and properties from RDF, RDFS and OWL.

Our adaptation of Semantic Forms infobox templates provides the page display of this ontological data. It also ensures the corresponding annotations with categories and properties that we define for standard Semantic Web constructs by using Semantic MediaWiki. Our code for these constructs ensures that SMW's RDF export of any wiki content data includes the RDF, RDFS and OWL triples relevant for the ontologies around that data.

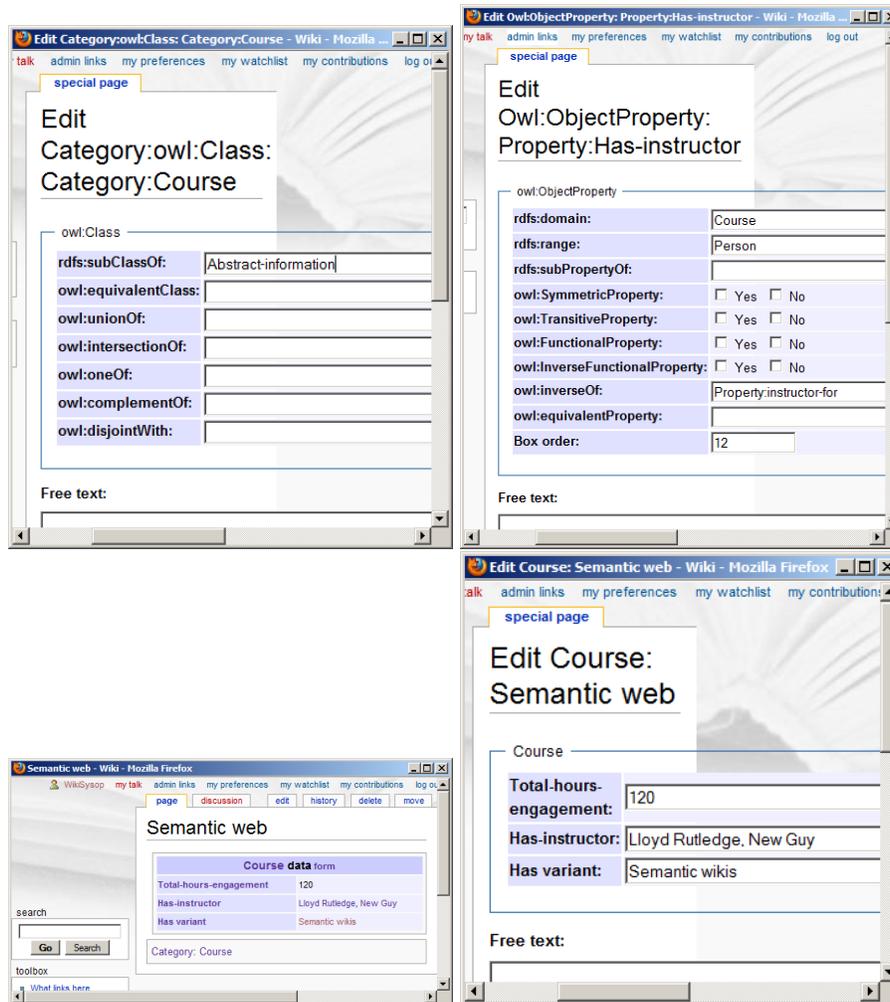
OWL Wiki Forms provides roughly the same forms for entering ontologies that the Masters course OWL project did: forms for general resources, classes, properties and restrictions. These correspond with, respectively, the Semantic Web standard classes `rdfs:Resource`, `rdfs:Class`, `owl:DatatypeProperty`, `owl:ObjectProperty` and `owl:Restriction`. Each of these forms allows the entry of value assignments for properties from RDF, RDFS and OWL. Assignments using Semantic MediaWiki's equivalent URI and ontology import facilities cause all Semantic Web definitions entered with these ontologies forms to appear appropriately in RDF exports from the wiki. Fig. 2 shows screen displays for forms for classes and object properties.

Each of the five main OWF forms for editing ontologies provides for the entry of values for standard properties. Their entry here results in not only explicit RDF export of what is entered but also corresponding internal wiki behavior for data display and entry. The property forms have value entry fields for various standard properties. Two of these are `rdfs:domain` and `rdfs:range`. The values entered for these two particular properties have later influence on the entry and presentation of data in OWL Wiki Forms, as described later. For object property forms, the `rdfs:range` field accepts a class name. For datatype property forms, it has a pull-down menu for selecting the `rdfs:Datatype` of the property.

The object property form also has three standard properties that refer to other properties. These are `rdfs:subPropertyOf`, `owl:inverseOf` and `owl:equivalentProperty`. The form assists ontologists in entering values for these properties by providing autocompletion for existing property names.

Some OWL property classes have no additional OWL properties of their own. These include `owl:SymmetricProperty`, `owl:TransitiveProperty`, `owl:FunctionalProperty` and `owl:InverseFunctionalProperty`. Protégé uses checkboxes rather than class assignments to indicate that properties are members of these classes. Similarly, Semantic Friendly Forms uses a checkbox to indicate that a property is a `owl:SymmetricProperty` [5]. OWL Wiki Forms takes the same approach, using the checkbox feature in Semantic Forms. When such a box is checked, it generates code that assigns the given property to the corresponding class assignment. All four of these checkboxes are on object property forms. A checkbox for `owl:FunctionalProperty` is also on datatype property forms.

By providing relevant internal ontological structure with each RDF export, OWL Wiki Forms facilitates the integration of wiki-generated triples into broader Semantic Webs such as the Linked Data Cloud. For Semantic Web courses, such exports let students see their wiki-edited ontologies in other Semantic Web tools



**Fig. 2.** Screen displays of the OWL Wiki Forms screen displays for a class form, an object property form, a page annotation form and an infobox

such as Protégé, and thus exercise with the integrability of the data their wiki implementations create. The ontologies entered via OWF also drive the inferencing, presentation and data input interfaces described later in this section. One limitation is that all of this processing currently only applies to ontological information entered via the wiki. Potential future work includes processing external ontological data for inferencing, browsing and data input assistance.

## 4.2 Querying and Browsing

Semantic MediaWiki has its own query language for internal data. One goal of OWL Wiki Forms is to have results from these queries include the inferencing that is defined by Semantic Web constructs entered in the wiki. Our initial means of programming this is as queries in infobox templates that find property assignments from that should infer properties of the current page, and then transform these query results into SMW code for the corresponding OWL assignments.

One important type of query result to include these inferences in is the property value display in infoboxes. In Wikipedia, Semantic Forms and OWL Wiki Forms, infobox templates provide a consistent table-based data report for the pages in a given category. Fig. 2 shows an infobox generated by OWL Wiki Forms. In Semantic Forms, a user determines the infobox rows and the properties they display. In OWF, on the other hand, infobox template code is generated automatically based on relevant ontologies in the system. OWL Wiki Forms determines each infobox's rows and how they appear by processing Semantic Web ontology constructs.

OWF also extends the processing of generation of specific Semantic Web constructs in generating infoboxes. The most important aspect of the infoboxes, the assigning of the rows themselves, is generated by OWL Wiki Forms by processing the `rdfs:domain` property. The processing by Semantic Friendly Forms of property domains [5] is applied here: if a property has a class as an `rdfs:domain`, then the property gets a row in the infobox for that class.

OWL Wiki Forms's generation of infoboxes applies the `rdfs:range` property by having links to nonexistent pages go to forms for the corresponding classes, so that the pages are registered with this class when the user later adds them. To do so, it generates code using Semantic Forms's "Has default form" property. Here, not only links from infoboxes trigger such forms, so do links anywhere in the wiki that lead to new pages for which range-based inferencing assigns them to a class.

OWF infoboxes show for each property not just the direct assignments from that page's form, as Semantic Forms does, but also the inferred values of the property. For example, infobox displays of values of symmetric properties show all values of this property for the current page, including names of pages with annotations with the same property pointing to the current page. Properties with an `owl:inverseOf` have each page in its triples show in the infobox of the other, albeit, unlike with `owl:SymmetricProperty` under different property names. Similarly, we plan development for `owl:TransitiveProperty` by having a value for such property rows in all infoboxes along the chain of pages link by the property.

## 4.3 Data Input

OWL Wiki Forms aims to process ontologies in order to guide the interface in assisting both system designers and data enterers. As with infoboxes, the `rdfs:domain` property determines the rows in a class's input form. The

`rdfs:range` property, on the other hand, plays a different role here. Fig. 2 shows a screen display for a page’s data input form, as generated by OWF processing of these and other properties.

One means of assisting data enterers is autocompletion on property value fields. To provide this, OWF processes the `rdfs:range` properties of each property in order to feed autocompletion in entering its values. Here, as the user types, autocompletion offers a pull-down menu of the wiki pagenames that are members of the classes in the property’s range.

OWL Wiki Forms achieves this by generating corresponding Semantic Forms “`autocomplete on category`” parameter code. On the Semantic Web level, this gives the autocompletion the names of pages representing resources within the given classes. The user can still enter new names outside this list. As described earlier, such a new name will then appear in the resulting infobox as a link to a form for one of the range classes that generates a new page of that name.

## 5 Conclusion

The Informatics faculty of the Open Universiteit offers Semantic Web courses that apply semantic wikis. These wikis support the didactic approach of these courses, in which knowledge of the Semantic Web and its components and technologies is tested by means of system development assignments. The functionality of this semantic wiki supports writing and sharing course material, the implementation of student programming assignments, communication between the teacher and each student, and between the students. Since distance learning applies here, the communicative aspects of the wiki are particularly useful. The technical contributions of this work include new support for Semantic Web components directly within a semantic wiki interface. These functional extensions of current semantic wikis are programmed in the software package OWL Wiki Forms. More information about OWL Wiki Forms is available at its website at <http://icommas.ou.nl/lru/OWLwikiForms/>.

## 6 Acknowledgements

Funding for this work comes in part from the Open Universiteit IPO online begeleiding (online teaching) project and from the OUNL’s Informatics department. Students of the OUNL course Capita selecta thema - Semantic Web provided feedback from their use of implementation in the execution of their class projects. The Semantic Friendly Forms implementation discussed in this work was part of an OUNL Bachelor’s degree team project. The students from this project, Henri Gerrits and Rens te Kloese, also gave comments beyond the scope of their project that were helpful for this work. Open Universiteit colleagues Stef Joosten and Laura Benvenuti gave insightful discussions about this work. Harold Pootjes and Leo Wiegerink collaborated in developing the Bachelor’s Semantic Web course, during which they helped with many of the concepts presented here.

## References

1. S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, and Z. Ives. DBpedia: A nucleus for a web of open data. In *In 6th Intl Semantic Web Conference, Busan, Korea*, pages 11–15. Springer, 2007.
2. C. Becker, C. Bizer, M. Erdmann, and M. Greaves. Extending SMW+ with a linked data integration framework. In *9th International Semantic Web Conference (ISWC2010)*, November 2010.
3. T. Berners-Lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer, and D. Sheets. Tabulator: Exploring and analyzing linked data on the semantic web. In *Proceedings of the 3rd International Semantic Web User Interaction*, 2006.
4. C. Bizer, R. Cyganiak, and T. Heath. How to publish linked data on the web. <http://www4.wiwi.fu-berlin.de/bizer/pub/LinkedDataTutorial/>, 2007. Revised 2008. Accessed 04/03/2011.
5. H. Gerrits and A. L. te Kloese. Semantic Friendly Forms. [http://icommas.ou.nl/lru/OWLwikiForms/index.php?title=Semantic\\_Friendly\\_Forms](http://icommas.ou.nl/lru/OWLwikiForms/index.php?title=Semantic_Friendly_Forms), 2010. Accessed 04/03/2011.
6. N. Heino, S. Dietzold, M. Martin, and S. Auer. Developing semantic web applications with the ontowiki framework. In T. Pellegrini, S. Auer, K. Tochtermann, and S. Schaffert, editors, *Networked Knowledge - Networked Media*, volume 221 of *Studies in Computational Intelligence*, pages 61–77. Springer, Berlin / Heidelberg, 2009.
7. H. Knublauch, M. A. Musen, and A. L. Rector. Editing description logic ontologies with the Protégé OWL plugin. *CEUR*, 104, 2004.
8. Y. Koren. Semantic Forms. MediaWiki extension, [http://www.mediawiki.org/wiki/Extension:Semantic\\_Forms](http://www.mediawiki.org/wiki/Extension:Semantic_Forms), 2011. Accessed 04/03/2011.
9. M. Krötzsch, D. Vrandečić, M. Völkel, H. Haller, and R. Studer. Semantic wikipedia. *Journal of Web Semantics*, 5:251–261, September 2007.
10. R. Oostenrijk. Implementeren en toepassen van Web Ontology Language (OWL) in Semantic MediaWiki (SMW). Master’s course student project, Informatics Faculty, Open Universiteit, 2009.
11. L. Rutledge. Semantic wiki’s voor onderwijs - linking data, kennis en leren. In *Nationale Informatica Onderwijs Congres (NIOC 2009)*, Utrecht, The Netherlands, 2009.
12. L. Rutledge, J. van Ossenbruggen, and L. Hardman. Making RDF presentable: integrated global and local Semantic Web browsing. In *Proceedings of the 14th international conference on World Wide Web, WWW ’05*, pages 199–206, New York, NY, USA, 2005. ACM.
13. R. Schuwer and F. de Vries. Structuring OER using learning trajectories. In *Open Ed Conference 2010*, Barcelona, Spain, 2010.
14. Semantic MediaWiki Community Wiki. Browse data page for public active sites. [http://smw.referata.com/w/index.php?title=Special:BrowseData&\\_cat=Sites&Public=1&Status=Active](http://smw.referata.com/w/index.php?title=Special:BrowseData&_cat=Sites&Public=1&Status=Active), 2011. Accessed 04/03/2011.
15. Semantic MediaWiki project. Semantic MediaWiki. <http://semantic-mediawiki.org/>, 2011. Accessed 04/03/2011.
16. D. Vrandečić and M. Krötzsch. Reusing ontological background knowledge in semantic wikis. In M. Völkel, S. Schaffert, and S. Decker, editors, *Proceedings of the First Workshop on Semantic Wikis - From Wikis to Semantics*, Budva, Montenegro, Juni 2006.

# Navigating learning resources through linked data: a preliminary report on the re-design of Organic.Edunet

Miguel-Angel Sicilia<sup>1</sup>, Hannes Ebner<sup>2</sup>, Salvador Sánchez-Alonso<sup>1</sup>, Fernando Álvarez, Alberto Abián, Elena García-Barriocanal

<sup>1</sup> Information Engineering Research Unit, Computer Science Department,  
University of Alcalá, Polytechnic Building, Ctra. De Barcelona km. 33.6  
28871 Alcalá de Henares (Madrid), Spain

{msicilia, salvador.sanchez, fernandol.alvarez, alberto.abian, elena.garciab}@uah.es

<sup>2</sup> School of Computer Science and Communication, Royal Institute of Technology,  
Lindstedtsvägen 3, 10044 Stockholm, Sweden  
hebner@kth.se

**Abstract.** Learning objects repositories have grown and matured in the last years, being currently a cornerstone for open education. Several current systems are offering metadata openly through mainstream harvesting protocols or providing standardized query interfaces. Also, the use of standardized vocabularies or ontologies is becoming more common to provide a degree of semantic interoperability. However, learning object metadata is typically not linked across repositories, and it is not providing a way to navigate by using other sources of data available on the Web. The linked open data (LOD) approach provides the framework for the evolution of learning object repositories into a more flexible system of sharing learning resource metadata. This paper describes how linked data has been integrated in the design and re-design of the export mechanisms of Organic.Edunet, a federation of learning repositories in the domain of organic agriculture that uses an RDF store and several ontologies to browse and search resources. The paper focuses on how the existing search and semantic browsing mechanisms can benefit from the use of LOD across repositories.

**Keywords:** learning objects, repositories, linked data, ontologies, Organic.Edunet

## 1 Introduction

Learning object repositories have grown and matured in the last years, being considered currently an essential component for e-learning (Ochoa and Duval, 2009). Each of these repositories provide some particular interaction style for searching, browsing and navigating across learning resources, typically including text and metadata search along with some form of topical navigation. Some repositories provide also capabilities that extend its reach to resources or metadata distributed

across different repositories (Klemke et al., 2010). This is currently achieved mainly by two mechanisms: harvesting and distributed search. In federated systems using harvesting, the metadata from a number of repositories is collected into a central location and regularly updated. Protocols like OAI-PMH are typically used for that process (van de Sompel et al., 2004). In distributed search, a query is distributed to several repositories and the query results are put together and eventually ranked according to some criteria. The Simple Query Interface (SQI) is an example of a protocol enabling that kind of arrangement (Simon et al., 2005).

While harvesting and distributed search have proved effective for aggregating repositories for the purpose of search, they do not allow actual browsing and navigation across different systems. In the case of systems based on harvesting, the navigation is still local to the centralized metadata store, and thus restricted to the repositories that are inside the federation. In the case of distributed search, browsing is also limited as only metadata records of query results are provided, and not links to categorizations of resources or to other resources not in the same repository, so cross-repository navigation is not allowed. The common problem with these systems is that they restrict navigation either by using a concrete request/response format or by restricting to the metadata already harvested in a central location, which is in most cases not providing navigation to other resources in its original systems. A distributed browsing and navigation system across repositories would require a way to expose metadata or query results that provides links to permanently identifiable resources along with links to elements of terminologies or taxonomies that are able to mediate navigation to other systems.

The collection of practices and recommendations for exposing linked open data (Bizer, Heath and Berners-Lee, 2009) are providing the required infrastructure for allowing that distributed navigation possible by exposing resource information in a common RDF format and enabling navigation through linked open data (LOD) to other systems. However, the deployment of the infrastructure for realizing that concept requires a re-conceptualization of the way repositories are commonly offering metadata and search facilities, along with the development of specialized browsers that are capable of traversing the Web of linked data. This paper provides the ideas and initial design elements for the implementation of such approach, using the Organic.Edunet repository<sup>1</sup> as an example. The research objective of the present work is laying out the foundations and requirements for exposing LOD in learning object repositories that enable cross-repository search and navigation. In consequence, the scope of the proposals presented focuses on how to give support to these scenarios and it is not addressing other infrastructural aspects (e.g. bootstrapping or the practicalities of migrating/extending current repositories).

The rest of this paper is structured as follows. Section 2 briefly describes the requirements from the data provider and consumer perspectives. Then, Section 3 describes the main decisions in the way IEEE LOM metadata is exported as linked data. Section 4 describes how the availability of generalized linked data following the conventions exposed could be exploited for browsing across repositories. Finally, conclusions and outlook are provided in Section 5.

---

<sup>1</sup> <http://portal.organic-edunet.eu/>

## 2 Requirements

Data provider requirements are no essentially different from those of exposing linked data in other domains in which some metadata schemas are already in place. The most widespread specific standard for providing descriptive metadata of learning objects is IEEE LOM (IEEE, 2002). There are a number of systems providing ontologies or RDF mappings of IEEE LOM, however they are not standardized. Nilsson et al. (2008) published a mapping of IEEE LOM into the Dublin Core Abstract Model that is discussed within the joint DCMI/IEEE LTSC Taskforce<sup>2</sup>. This mapping was also used within Organic.Edunet to annotate resources using the repository tool Confolio (Ebner, 2009). The mapping sets the basis for exposing both Dublin Core and IEEE LOM metadata by using a shared format.

Further, there is a need to consider the particular nature of learning objects and the fact that a LOM metadata record is describing two very distinct types of information: (a) *factual* information on the resource, e.g. identifier, language, technical description, and (b) *contextual* information in which the resource is described for a particular educational context, i.e. difficulty is relying on some particular target user population, and it is possible that another metadata record for the same resource is having a different value as it is considering a different target. Such a contextualisation is supported by the Confolio tool and its backend (Ebner, 2008). One common use case is to harvest metadata from generic repositories and to provide context by adding educational metadata. Separate named graphs with relations between them are used to express metadata (in our case educational metadata) and meta-metadata such as provenance information and access control.

The general availability of learning object metadata as LOD across repositories would enable a number of applications that are beyond the reach of current systems. From the viewpoint of filtering and selection, it would be possible to build systems that mediate search through terminologies or ontologies representing educational systems and competencies required for particular profiles. Particularly, it would be possible to navigate through these representations to browse and search, complementing current distributed query and harvesting mechanisms. Another key application would be that of aggregating metadata descriptions for the same learning object, creating new opportunities for filtering based on quality, and having a way to measure the reusability of the resource across different educational contexts (Sicilia and García-Barriocanal, 2003).

In order to enable cross-repository search and browsing through LOD, there is a need for some conventions on exposing resource metadata and also some descriptions of the repository themselves. These are described in what follows.

## 3 Exporting IEEE LOM as linked data

Organic.Edunet features two different approaches for storing metadata: (1) the distributed repository tools using a triple-store with an abstraction of named graphs

---

<sup>2</sup> <http://dublincore.org/educationwiki/DCMIIEEEELTSCTaskforce>

and an implementation of the DCMI/IEEE draft using DCAM (Ebner, 2009), and (2) on the federated portal side, an OWL-based repository based on HP Jena with a relational datastore backend, using an OWL representation of IEEE LOM combined with several ontologies, including one with an specific terminology for organic agriculture and agroecology (Sánchez-Alonso, 2009). The repository tools within the federation expose the metadata following the linked data approach and according to the abovementioned DCMI/IEEE draft. In a different way, the process of exporting linked data through the portal required the construction of a new module that uses the existing SPARQL endpoint to translate the native RDFOWL representation to the RDF export described below, combined with a module creating additional RDF links whenever viable in an automated way.

Identifiers should be representing at least two types of entities: the object themselves (i.e. the Web contents) and the metadata records, as the same objects may be described several times in different repositories (Downes, 2004). In our case, all the resources are external to the portal and identified by URIs, so we are exposing only the second kind of objects. The resource identifiers themselves in IEEE LOM are specified in element `General.Identifier`. In our case, the URI is the object of all the resource statements about the resource, with the exception of the `Meta-metadata` category, that will have as object the document requested. In consequence, a fragment expressed in Turtle syntax for the description of an example resource would be as follows (namespace declarations are omitted for brevity):

```
# Meta-metadata information
<>
  lom-meta:schema lom-meta:LOMv1.0
  lom-meta:language <http://dbpedia.org/page/English_language>;
  ...
# The description
<http://confolio.vm.grnet.gr/scam/6/resource/271>
  lom-gen:language <http://dbpedia.org/page/English_language>;
  lom-gen:keyword [
    a rdf:Alt;
    rdf:_1 "energy consumption"@en ;
    rdf:_2 "energiforbruk"@no ;
  ];
  lom-gen:structure lom-gen:atomic;
  lom-gen:coverage <http://dbpedia.org/page/Austria>;
  lom-edu:interactivityType lom-edu:expositive;
  lom-life:editor [
    lom-life:contributor [
      foaf:name "Petros Lamas"
      foaf:workplaceHomepage <http://www.ea.gr>
    ]
  ]
  lom-life:validator [
  ...
```

There are several cases in which links to other linked data sites can be provided, including language of the metadata record and of the resource, which can be mapped to DBPedia resources. The lifecycle of the resources in Organic.Edunet requires metadata on the content provider (editor) and on the party that underwent quality assessment (validator). Both are using internally vCards, however in the exporting through linked data, these are converted into FOAF sentences linking to organizations where available. Another interesting case is `lom-gen:coverage` in

which spatial coverages are linked to the DBPedia, and can also be linked to other LD datasets representing geopolitical information. In general, the vocabulary values used (e.g. lom-edu:expositive) are represented using the (namespace, value) construct proposed as a general mechanism in Nilsson et al. (2003).

Linking to ontology/terminology elements is achieved by exposing the Organic.Edunet ontology as linked data by using the following convention: `<http://www.organic-edunet.eu/ont/element>`. Where *term* is one of the terms, instances or properties in the ontology. The RDF description enables the navigation through the is-a hierarchy but also laterally by following predicates. For example:

```
<http://www.organic-edunet.eu/ont/biological-fertilizer>
  oe:is-made-from
    <http://www.organic-edunet.eu/ont/plant-origin-processed-product>;
```

A special case is that of element `Classification` which is used for various purposes in Organic.Edunet. IEEE LOM provides a vocabulary of purposes for this field, but these have been extended in the project with additional ones that were selected to provide a better description of classifications. The following fragment shows an example of IEEE LOM and extended classification purposes.

```
oe:details <http://www.organic-edunet.eu/ont/soil-fertility>
```

The `oe:detail` property is an `rdfs:subPropertyOf` of the vocabulary property `lom-cls:educationalObjective`. This is one of the IEEE LOM purpose values refined currently in the portal, a similar approach could be used for others as for example, `prerequisites`.

The abovementioned conventions are the basis for exposing metadata in LOD format, including links to arbitrary terminologies. An additional step is that of providing links to other resources. This is implemented by extending the meanings of the IEEE LOM `Relation.Kind` element (which has a value space based on Dublin Core). As there are no data for this element in Organic.Edunet, descriptions are extracted from other parts of the portal. Currently and for demonstration purposes, links based on co-occurrence of user-created tags are automatically added.

## 4 Browsing across repositories by using IEEE LOM

Architecting a browser that works over repositories publishing linked data following the abovedescribed conventions requires components to start the search process and to traverse links across metadata to reach related resources. We limit our discussion to reaching resources that are related through common reference to entities or concepts in terminologies that are also published as linked data. Organic.Edunet features a form of “berrypicking” search as described by Bates (1989). The existing implementation was based on the traversal of ontology relations, namely subclass for the main navigation and arbitrary properties for an alternative traversal. The selection one or several terms is done by navigation, iteratively and search processes are done on the central metadata repository that harvests regularly the provider systems. A REST-based interface is invoked from a Flash-based visualization, using JSON for the transfer of data for the rendering of the traversal tree. The re-design to exploit linked data consumption entails two main aspects:

1. Changing the interfaces serving terms during navigation to provide ontology terms across distributed linked terminologies. This allows the visualization to render any kind of relationship that is provided as backlinks in the description of the terms.
2. Changing the invocation from a local search to a distributed one. This can do in two flavors: (i) using text descriptions to go through conventional distributed query systems and (ii) selecting repositories or aggregators depending on the self-declaration of the terminologies used by repositories.

In both cases, repositories need to expose some form of standardized query interface as the SQI. The second approach would require knowing which terminologies are in use by each repository, so that RDF links as applicable in `Classification` are used for the selection. A kind of self-declaration from repositories may follow the linked data approach by exposing information about themselves using linked data. An example is as follows:

```
<http://www.organic-edunet.eu/lor>
  dc:requires <http://www.organic-edunet.eu/ont/>;
  dc:requires <http://aims.fao.org/aos/agrovoc/>;
  ...
```

The W3C VoID<sup>3</sup> vocabulary for expressing metadata about RDF datasets can be used for the same purpose, concretely using the `void:vocabulary` property. Also, a property similar to VoID `sparqlEndpoint` could be used to provide the SQI endpoint of the repository. While many repositories are not using any kind of controlled vocabulary there are some relevant cases starting to use them as the LRE Thesaurus<sup>4</sup>, the Organic.Edunet ontology or the different classification systems used in MACE<sup>5</sup>. In other cases, they can be generated automatically from metadata, as has been experimentally tested with AGROVOC in Organic.Edunet by using KEA<sup>6</sup> keyphrase extraction mechanisms. As an example, let's consider the current semantic navigational interface in Organic.Edunet showed in Figure 1. The OE ontology is providing mappings to terms in the AGROVOC thesaurus via `owl:sameAs`, which has very recently started to be exposed as linked data<sup>7</sup>. This enables the navigation to move to AGROVOC for the collection of query terms. Once the user has finished the query formulation, the terms selected would be sent via SQI to the underlying repository (in this case Organic.Edunet) or to any other that is exposing their description. The `requires` statements serve the role of selecting the more appropriate repositories according to terminologies used. Repositories can be discovered by crawling or by trying with the prefixes of URIs of resources included in the `relation` statements of records obtained.

The results snippets in Figure 1 would then potentially correspond to resources in several repositories without the need for central harvesting or registries of repositories. Also, relations between learning objects and to other linked data sets (e.g. as in coverage) can be exploited for query expansion.

---

<sup>3</sup> <http://www.w3.org/TR/void/>

<sup>4</sup> <http://lre.eun.org/node/87>

<sup>5</sup> <http://portal.mace-project.eu/>

<sup>6</sup> <http://www.nzdl.org/Kea/>

<sup>7</sup> <http://aims.fao.org/website/Linked-Open-Data/sub>

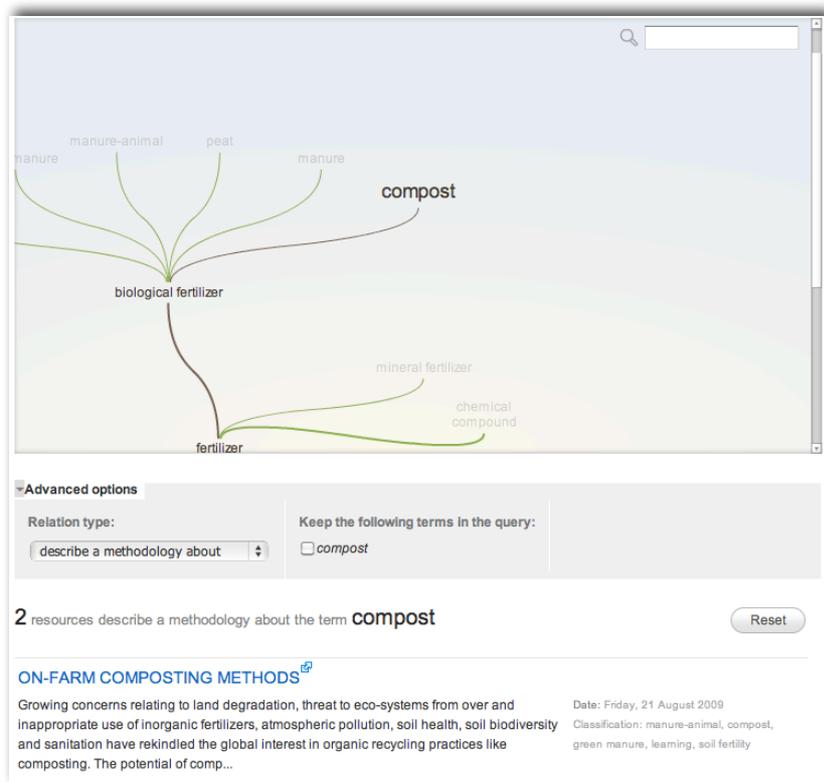


Fig. 1. The “berrypicking” interface exposed in Organic.Edunet

## 5 Conclusions and outlook

The linked data approach represents an opportunity to evolve existing learning object repository technology towards the open exposure of metadata in a form that enables novel approaches to search and navigation that are not restricted to centralized metadata stores and that enable navigating across repositories without a need for integrating them beyond the provision of links to other linked data sources.

Linked data support in the Organic.Edunet portal is currently in a prototype stage and in ongoing development, so it may still be subject to change. It has been used in a different way from the beginning within the repository tools of the providers to the portal based on Confolio, complying with DCMI/IEEE drafts for encoding the metadata. The support for distributed browsing is currently based only on pilot prototypes, as there are no other learning object repositories exposing linked data that could be used for integration testing. However, the re-design of the browsing interfaces show that the approach is feasible to implement. Future work should deal with the evaluation and refinement of the conventions used to provide semantic relations between learning resources and their metadata, moving to a formal

specification that could be used in the many repositories currently exposing IEEE LOM metadata through OAI-PMH or SQL interfaces. It should be noted that the success of the approach presented here relies on the adoption of a shared set of linked data exposure conventions for repositories, which can only come from a kind of community consensus.

## Acknowledgments

The work presented in this paper has been funded with support by the European Commission through the projects Organic.Edunet and TEL-Map.

## References

- Bates, M.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13 (October 1989): 407-424.
- Bizer, C., Heath, T. and Berners-Lee, T.: Linked data - the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3): 1-22 (2009).
- Downes, S. (2004) 'Resource profiles', *Journal of Interactive Media in Education*, Available at: <http://www-jime.open.ac.uk/2004/5>.
- Ebner, H., Palmér, M.: A Mashup-friendly Resource and Metadata Management Framework, In: Wild, Kalz, Palmér (Eds.): *Mash-Up Personal Learning Environments, Proceedings of the 1st Workshop MUPPLE, European Conference on Technology Enhanced Learning (EC-TEL)*, Maastricht, The Netherlands, 17-19 September, 2008.
- Ebner, H., Manouselis, M., Palmér, M., Enoksson, F., Palavitsinis, N., Kastrantas, K., Naeve, A.: Learning Object Annotation for Agricultural Learning Repositories, *IEEE International Conference on Advanced Learning Technologies*, Riga, Latvia, 15-17 July, 2009.
- IEEE LOM, "Draft Standard for Learning Object Metadata", IEEE Learning Technology Standards Committee, IEEE 1484.12.1-2002, 15 July 2002.
- Klemke, R., Ternier, S., Kalz, M. and Specht, M. (2010), Implementing infrastructures for managing learning objects. *British Journal of Educational Technology*, 41: 873–882
- Nilsson, M.: Draft Recommended Practice for Expressing IEEE Learning Object Metadata Instances Using the Dublin Core Abstract Model. Draft IEEE P1484.12.4<sup>tm</sup>/D1 (2008).
- Ochoa, X. and Duval, E.: Quantitative analysis of learning object repositories," *IEEE Transactions on Learning Technologies* 2(3), 226–238 (2009).
- Van de Sompel, H., Nelson, M.L., Lagoze, C. and Warner, S.: Resource harvesting within the OAI-PMH framework. *D-Lib Magazine*, 10(12), 1082-9873 (2004).
- Sánchez-Alonso, S.: Enhancing availability of learning resources on organic agriculture and agroecology. *The Electronic Library* 27(5): 792-813 (2009).
- Sicilia, M.A. & García, E. (2003). On the Concepts of Usability and Reusability of Learning Objects. *International Review of Research in Open and Distance Learning* 4(2).
- Simon, B., Massart, D., van Assche, F., Ternier, S., Duval, E., Brantner, S., Olmedilla, D., Miklos, Z.: A Simple Query Interface for Interoperable Learning Repositories. In: *Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems*, pp. 11–18 (2005)

# Educational Resources as Social Objects in Semantic Social Networks

Eleni Kaldoudi<sup>1</sup>, Nikolas Dovrolis<sup>1</sup>, Daniela Giordano<sup>2</sup> and Stefan Dietze<sup>3</sup>,

<sup>1</sup> School of Medicine, Democritus University of Thrace, Alexandroupoli, Greece  
[kaldoudi@med.duth.gr](mailto:kaldoudi@med.duth.gr), [ndovroli@alex.duth.gr](mailto:ndovroli@alex.duth.gr)

<sup>2</sup> University of Catania, Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Viale  
A. Doria 6, 95125, Catania, Italy

<sup>3</sup> The Knowledge Media Institute, The Open University, MK7 6AA, Milton Keynes, UK.  
[dgiordan@diit.unict.it](mailto:dgiordan@diit.unict.it)  
[s.dietze@open.ac.uk](mailto:s.dietze@open.ac.uk)

**Abstract.** Educational content is often shared among different educators and is enriched, adapted and in general repurposed so that it can be re-used in different contexts. This paper exploits the various aspects of object-centered sociality that the educational resource can manifest in a participatory virtual community of authors and learners. These social facets include: (a) the obvious connections via common tags; (b) connections based on collective usage and other related interaction of human users; (c) social connections based on the inheritance as realized via educational content repurposing; (d) semantic connections and similarities of accompanying metadata. The concept of exploiting the various aspects of educational resources as social objects is shown in the implementation of the MetaMorphosis+ social network for educational resource sharing in health sciences.

**Keywords:** educational resource, social networking, semantic technologies, Linked Open Data.

## 1 Introduction

During the past ten years we have experienced the realization of the so-called “information society”, characterized by mass information seeking and based on the distribution, mainly via the internet, of pre-defined and standardized digital data. Current and emerging trends in information systems and communication services mark the shift towards a “knowledge society”, that is, a society based on knowledge as a value. Thus, educational resources (in the broad sense) become a central issue in our society. This paper exploits the social aspect of educational resources within a participatory virtual community of authors and learners as realized via the MetaMorphosis+ environment for sharing educational resources in medical education.

This work is conducted partly within the mEducator project, an EU funded best practice network (under the eContentPlus2008 programme, Contract Nr: ECP

2008 EDU 418006) with the aim to implement and critically evaluate existing standards and reference models in the field of e-learning in order to enable specialized state-of-the-art medical educational content to be discovered, retrieved, shared and re-used across European higher academic institutions. The following sections discuss some of the different facets of the educational object sociality, including (a) the obvious connections via common tags; (b) connections based on collective usage and other related interaction of human users; (c) social connections based on the inheritance as realized via educational content repurposing; (d) semantic connections and similarities of accompanying metadata. Then, the concept of exploiting the various aspects of educational resources as social objects is shown in the implementation of the MetaMorphosis+ semantic social network [1].

## **2 Educational Content as a Social Object**

The social Web, or Web 2.0 [2], has become an important trend during the last few years. Among the prominent social web tools, social networking websites focus on creating online communities of individuals who publish their content and activities while exploring others content and activities, thus creating virtual on-line social groups and associations.

This communication paradigm has been taken up by the community of researchers and academics and nowadays there is a thriving number of social networks dedicated to science and professional relations. Most of these virtual communities provide a way for researchers to create and populate their online profiles and share their work with others. Recently the term object-centered sociality was introduced [3] to describe the fact that strong social relationships are built mainly when individuals are grouped together around a shared object that mediates the ties between them. Therefore, we can assume that each content item on a social network site can be a source of social connectivity, catalyzing social networking in virtual spaces. This new approach to sociality has drawn attention, and current state-of-the-art research in the area involves various ways to exploit object-oriented sociality to the benefit of the community. For example, a semantic tag model has been recently introduced for representing tag data in the form of reusable constructs at a semantic level in order to share tag metadata across heterogeneous social networking sites [4]. A similar work proposes to use existing information on Web 2.0 and convert it to RDF which can be used as a flexible model for describing and integrating data [5],[6].

In this paper we take a different novel view about merging semantic technologies with the social web in order to enhance object-centered sociality and data integration. Here the social object is the educational resource. In our work, we view and exploit their 'social aspects' of educational resources in a variety of different ways as described below.

## 2.1 Building a Social Profile for Educational Resources

The most straightforward social dimension of an educational resource as a social object in a network can be realized in the conventional way of connections among profile tags. This requires a standardized metadata set to describe concisely an educational resource and thus create its social profile. Standardizing metadata for describing digital educational resources constitutes one of the main research topics in the e-learning community. Currently, the IEEE LOM (Learning Object Metadata) XML scheme seems the most prominent standard for describing learning objects [7] as it derives from a number of related standardization initiatives. LOM defines a wide range of metadata to classify and characterize learning resources, which include: overall description (cataloguing, annotations, and associations and relationships with other learning resources), technical data (file size, format, installation/usage descriptions), educational data (educational purpose, learning objectives, classification), and management data (intellectual property rights). However, there are no generally accepted conventions for properly describing learning objectives or the learning context, and although attempts have been made (such as CLEO, and Educational Markup Languages) these only capture some of the semantics; thus more complex models are needed [8]. Thus, different communities give different meanings and use different descriptions (e.g. extensions/alterations to LOM, such as Healthcare LOM, etc), really breaking down the notion of the standard. Whereas the above difficulties relate to issues of relevance, semantics and usability, a more severe critique of learning object metadata standards [9] points out that the assumption implied in the current ways of describing an educational resource is that “information equates learning”.

Since metadata currently only describes simplified technical and structural characteristics of learning objects, In order to put emphasis on the educational nature and context of a resource (as opposed to its technical integration in the digital world) one had to resort towards metatags that describe the nature of the learning activities in relation to the knowledge states that may result, or the thinking processes stimulated by the activities entailed by the resource. This approach is taken by the mEducator consortium in their recent proposal of a metadata scheme to capture pedagogical aspects of a learning resource as well as patterns of activity (e.g. repurposing episodes, but not only). The mEducator metadata scheme [10] includes a number of fields addressing different aspects of the educational resource: (a) general fields: resource title, unique identifier, URL, URN, intellectual property rights clearance/license, quality stamp (if any); (b) fields related to a general resource description: resource authors, creation date, citation (i.e. how the resource should be formally cited), keywords, content description, technical description (including any technical requirements to access and use the resource); (c) fields related to the educational aspect of the resource: educational context (for which the resource is intended), teaching/using instructions, educational objectives, expected learning outcomes, suggested assessment methods, educational prerequisites; (d) fields related to classification/taxonomy information: resource language, type, discipline, discipline

subspecialty, educational level; and (e) fields addressing repurposing: resource parents, repurposing context, repurposing description.

These user generated description metadata, when treated as social tags, create a complex and dynamic organization of educational resources in a similar fashion as in any conventional social network, thus realizing the resources' social network.

## **2.2 Resource Sociality based on Collective Usage Interaction**

Educational resources are accessed, used, shared, repurposed, and also rated, commented upon, and can be organized in a number of user specified ways in collections. Capturing and sharing information about the attention that users spend on resources in specific contexts can provide a different aspect of sociality based on the personal views and mental models of the users. This way one can build the profile of a resource as it appears to the external user, as opposed to the profile of the resource according to the view of its creator, as depicted in the description metadata.

The perspective and attention is normally captured via recording contextual attention metadata [11]. This includes data about the users' attention and activities that relate both to semantically rich actions on and interactions with educational content items as well as data on indirect interactions amongst content items. Additionally, basic interaction metadata can also be considered which includes all other basic user-system interactions that provide some kind of basic attention information (not necessarily semantically and contextually rich).

## **2.3 Resource Family Trees based on Repurposing History and Inheritance**

The term 'repurposing' refers to changing a learning resource initially created and used for a specific educational purpose in a specific educational context in order to fit a different new educational purpose in the same or different educational context. Although not formally addressed as such, educational content repurposing is what any educator is routinely engaged in when preparing a new educational experience, including preparing the educational content itself. Customarily, when an educator sets the context and goals of a new educational experience, he/she will overview existing content and/or search for new relative content and then repurpose and re-organize content to fit the purpose of the new educational experience.

There can be a variety of situations where repurposing educational content is desired. These situations, referred to as "repurposing contexts", can be of a pedagogical nature, a technical nature or both, and include the following [12],[13]: repurposing (1) in terms of the actual content; (2) to different languages; (3) to different cultures; (4) for different pedagogical approaches; (5) for different educational levels; (6) for different disciplines or professions; (7) to different content types; (8) for different content delivery media and/or technology; (9) to educational content from an initial content type that is not intended for education; and (10) for people with special

needs.

Considerable research work has targeted the field of automatic learning resource repurposing, e.g. [14],[15],[16],[17]. Only few works have really concentrated on modelling repurposing history of the content. The problem of the granularity level at which the modifications should be described is posed in [18], where the approach is to track content changes only in structural, layout and content aspects of the content (i.e. insert, delete, replace, rearrange). A shift from tracking structural changes towards studying changes in the educational context is seen in the repurposing of a specific content type (that of virtual patients) as addressed in the eVip/ReVip projects [13]. Finally, the MURLLO project from the community of language teachers, considers repurposing by tracking the changes in the “contextual metadata” that describe the intended teaching context [19]. Interestingly, metadata modifications that account for repurposing are carried out collaboratively based on Wiki technology.

In our work, we address repurposing as a means to provide a different kind of sociality for the educational resources. Thus repurposing history and inheritance are used as basic social relationship among educational resources in order to cluster resources into families. Each repurposed resource declares its parent(s) resource(s). Following iteratively the ‘parents’ in a chain of repurposing ancestors, the entire ‘family’ tree of the particular resource can be compiled.

#### **2.4 Semantic Links between Resources by Harvesting the Linked Data Cloud**

Finally, the social dimension of educational resources can be further expanded and enriched by harvesting semantically rich information existing in the Linked Data cloud. The Linked Open Data (LOD) approach is simply about “using the Web to create typed links between data from different sources. These may be as diverse as databases maintained by two organisations in different geographical locations, or simply heterogeneous systems within one organisation that, historically, have not easily interoperated at the data level [20]. The Linking Open Data Project [21] is a community project of the World Wide Web Consortium’s Semantic Web Education and Outreach Group (W3C SWEO). The goal of the project is to extend the Web with a data commons by publishing various open data sets on the Web, and making links between data items from different data sources. Since its inception in June 2007, the size of the cloud has rapidly exploded and already includes a large variety of open data sets including several research and medical data sets. This wealth of information can be used to automatically enrich educational resources metadata with references to external vocabularies, and in particular domain specific vocabularies, thus creating a rich domain specific profile and extending the resource’s social connections to other web objects. For example, in medical education the BioPortal [22] enables searching, sharing, visualizing, and analyzing a large set of biomedical ontologies, terminologies, and annotations exposed as LOD.

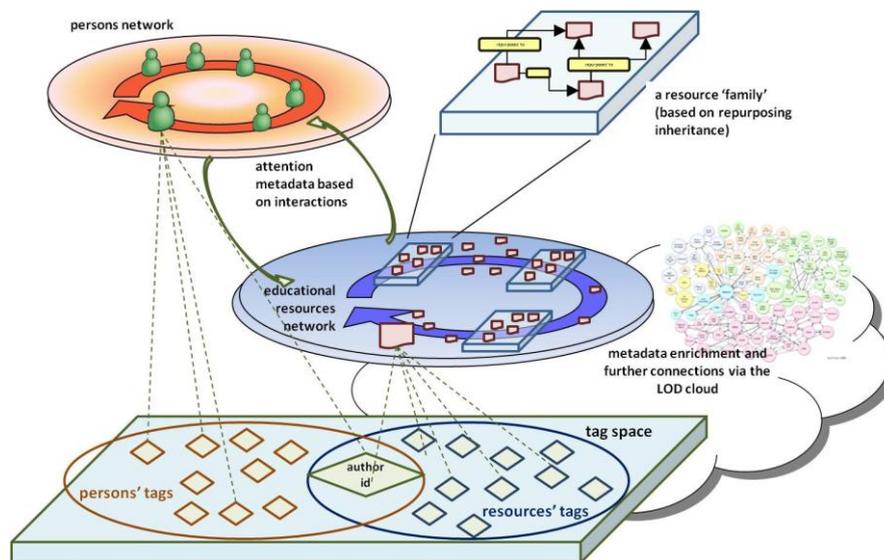
### 3 A Semantic Social Network for Educational Content Sharing

The various ways of exploiting educational resources as social objects have been implemented in the MetaMorphosis+ social network [1], which can be viewed as two distinctive and interacting networks. The first one is a network of persons, including authors, potential authors and final users of learning objects (students, or teachers or others, e.g. educational managers, etc). The second is a network of published learning resources. The network of persons is functioning in a way similar to other social networks. Persons can interact with each other via their personal blogs, declare friends and create their own interest groups. At a different level, learning resources themselves create an equivalent social network with interactions with other learning resources as well as with persons. These interactions are variable and dynamic, thus create an evolving, user centric and goal oriented organization of resources and persons, based on social dynamics.

From the point of view of the resources' social network, interactions are more complex. Each resource is described by a variety of fields that capture its basic characteristics as well as features pertaining to repurposing. This collection of fields forms the resource profile and is a virtual representation of the resource in the social network of resources. Educational resources are distributed, and they can reside anywhere on the Web (e.g. within a Learning Management System, another on-line repository, a Web page, etc), as long as their URL is known. MetaMorphosis+ only holds their metadata description and the pointer to their actual location.

The organization of educational resources is dynamically created around the four different social aspects presented in the previous section, as shown in Fig. 1. A most straightforward organization is created on the basis of user generated tags that constitute the resource profile as a set of metadata.

A second type of organization is a hierarchical one, describing the repurposing history of each resource. Each repurposed resource declares its parent(s) resource(s) and thus a dynamic organization of the resources in 'families' is generated. In MetaMorphosis+, a force-directed graph is used to depict the specific resource's family and inheritance patterns. Each node in the graph represents a resource, while the directed edges represent repurposing relationship, with the arrows pointing from the "source" objects to their "repurposed" descendants. The nodes also state the 'repurposing context', while they are active links to the resource profile where more information on the repurposing description can be obtained. For the entire resource collection, a circular directed graph representation is used which depicts all the resources with the various individual inheritance trees, usually not interconnected amongst them. A resource inheritance tree is a group of resources that have a relationship based on repurposing – this can also be viewed as resource 'family'.



**Fig. 1.** A schematic representation of the double network organization in MetaMorphosis+ showing the different social aspects of the educational resources.

The third dimension in the social organization of resources is based on a model for framing the representation and treatment of information gathered from the reuse and repurposing of learning resources from distributed repositories [23]. The model takes into account as sources of information both static user-edited or automatically generated metadata fields and the emerging, dynamic information clouds that surround a learning resource when users comment on it, tag it, or explicitly link it to other learning resources, i.e. by a combined use of strict taxonomies/controlled vocabularies with folksonomies (dynamic, user's tags facilitators/aggregators). By coordinating these separate information layers, we hope to reduce the semantic gap occurring when unanticipated contexts of use are to be described by resorting only to predefined vocabularies, and thus to improve the relevance of the retrieved resources after a query.

Finally, the social profile of the educational resources is augmented by semantic tagging. The architectural framework for semantic data and service linking and federating of disparate educational resource pools that powers the social environment is described in detail elsewhere [24]. At the lower level of this semantic technology framework, a Web data and service layer employs distributed Web services that harvest educational resource metadata from heterogeneous data sources on the Web. In the upper layer, semantic data and service integration is achieved based on the Linked Services approach and on semantic technologies such as iServe [25] and SmartLink [26]. An RDF repository exposes harvested educational resource metadata as triples.

Metadata as harvested by Web sources can also be enriched with existing LOD

vocabularies). This is of particular importance to extend rather unstructured metadata, such as keywords or free text subject and discipline descriptions with structured data based on well-known vocabularies. This is achieved by exploiting a variety of medical domain ontologies and the expanding LOD cloud to semantically annotate the existing RDF description of a resource and then expose its metadata back to the LOD cloud for further exploitation by third parties which make use of the web of LOD.

Biomedical ontologies provide essential domain knowledge to drive data integration, information retrieval, data annotation, natural-language processing and decision support. BioPortal (<http://bioportal.bioontology.org>) is an open repository of biomedical ontologies that provides access via Web services and Web browsers to ontologies developed in various formats including OWL, RDF, OBO format and Protégé frames [22]. In MetaMorphosis+ we have utilized the NCBO BioPortal's RESTful Web services programming interface to access and incorporated terms and concepts from the more than 260 ontologies provided to this day, corresponding to more than 4.5 million medical and life sciences terms. This way the MetaMorphosis+ user can annotate an educational resource with suggested standardized terms and concepts from a variety of ontologies, enriching the RDF output with dereferencable standardized terms as values for the various fields, e.g. keywords, discipline, specialty, etc. The ontologies used include amongst else for prominent medical ontologies such as SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms), ICD9/10 (International Statistical Classification Diseases and Related Health Problems), Body System (body system terms used in ICD11), MeSH (Medical Subject Headings), NCI (Meta)Thesaurus, Galen (the high level ontology for the medical domain), HL7 (the Normative RIM model v2), Biomedical Resource Ontology (BRO, a controlled terminology of resources to improve sensitivity and specificity of Web searches).

As an example, suppose a user intends to describe an educational resource by using the term/concept Telemedicine, in the list of 'Keywords' or in the 'Discipline' and 'Specialty' fields of the metadata description of the educational resource. Semantic annotation in MetaMorphosis+ can suggest a number of related standardized terms from the available ontologies. For example, the equivalent term from the NCI Thesaurus is represented by a dereferencable URI, (namely <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#Telemedicine>), thus enabling rich linking and reasoning within the semantic Web. Figure 1 shows an example of RDF triples in the mEducator scheme before and after semantic annotation in MetaMorphosis+.

The current deployment of MetaMorphosis+ social network is implemented using the Elgg open source social engine (<http://elgg.com>) and is available on-line on <http://metamorphosis.med.duth.gr>. Additionally, FOAF (Friend-of-a-Friend) ontology is used for integration of human profiles, while Triplify [27] is employed to publish metadata profile fields as Linked Data. Graph representation was implemented based on the Prefuse information visualization toolkit (<http://www.prefuse.org>) and is based on the GraphML (<http://graphml.graphdrawing.org/>) standard and a FOAF/GraphML

standard interface so as to dynamically extract the data and form the graphical representation at the time of request.

```

before enrichment ↗
- <mdc:subject>
  - <mdc:Subject rdf:about="http://meducator.open.ac.uk/metamorphosis#subject1247265680">
    <rdfs:label>Thrombolysis </rdfs:label>
  </mdc:Subject>
</mdc:subject>
<mdc:description>ECG case 1004 limb and chest leads</mdc:description>
<mdc:technicalDescription>PDF</mdc:technicalDescription>
<mdc:resourceType rdf:resource="http://purl.org/meducator/mediaType#sketchGraphicalAnnotation"/>
<mdc:resourceType rdf:resource="http://purl.org/meducator/mediaType#image"/>
- <mdc:discipline>
  - <mdc:Discipline rdf:about="http://meducator.open.ac.uk/metamorphosis#discipline1351316776">
    <rdfs:label>cardiology </rdfs:label>
  </mdc:Discipline>
</mdc:discipline>

after semantic enrichment ↗
- <mdc:subject>
  - <mdc:Subject rdf:about="http://meducator.open.ac.uk/ontology/SNOMEDCT/51308006">
    - <rdfs:seeAlso>
      http://purl.bioontology.org/ontology/SNOMEDCT/51308001
    </rdfs:seeAlso>
    <rdfs:label>Thrombolysis</rdfs:label>
    <mdc:externalSource>SNOMED Clinical Terms </mdc:externalSource>
  </mdc:Subject>
</mdc:subject>
<mdc:description>ECG case 1004 limb and chest leads</mdc:description>
<mdc:technicalDescription>PDF</mdc:technicalDescription>
<mdc:resourceType rdf:resource="http://purl.org/meducator/resourceType/Sketch-graphicalannotation"/>
<mdc:resourceType rdf:resource="http://purl.org/meducator/resourceType/IMAGE"/>
- <mdc:discipline>
  - <mdc:Discipline rdf:about="http://meducator.open.ac.uk/ontology/SNOMEDCT/51308001">
    - <rdfs:seeAlso>
      http://purl.bioontology.org/ontology/SNOMEDCT/51308001
    </rdfs:seeAlso>
    <rdfs:label>Cardiology</rdfs:label>
    <mdc:externalSource>SNOMED Clinical Terms </mdc:externalSource>
  </mdc:Discipline>

```

**Fig. 2.** The above depictions of RDF/XML snippets show the differences of the RDF output before (above) and after (below) the semantic enrichment via the BioPortal. Before enrichment the values for the 'Keyword' and the 'Discipline' fields appear as strings with no linking to an ontology, while after enrichment these values are substituted with dereferencable URIs corresponding to standardized ontology terms.

The main goal of this implementation is to collect and organize pilot educational content within the mEducator project and test and re-engineer the metadata scheme for describing educational content in medical education. During the first few months of deployment the environment exhibits more than 100 registered users and more than 350 educational resources, including 80 repurposed resources. Although about half of the resources are in English language, there is a representation of more than 15 other European languages. The resources included in the environment are distributed among the various educational levels, 33% intended for undergraduate medical education, 23% intended for postgraduate/resident studies and 21% for continuing

life-long education, while 22% are intended for educating the public. The majority of resources are of conventional content types, such as lecture notes and books (34%), lecture presentations (12%) and graphs/diagrams/figures/images (16%). Clinical cases, teaching files and virtual patients are 12% of available resources, while there is a small but notable representation of Web 2.0 type resources (4%), serious games (2%), algorithms (4%), and simulators (1%). When it comes to the repurposed resources, 84% of them have only one parent, while the rest declare two parent resources. A total of 42% of the repurposed resources have declared 2 repurposing contexts and 22% more than two repurposing contexts. All repurposing contexts are well represented, with repurposing to “different technology”, and “different educational level”, being the most common.

#### **4 Discussion**

In retrospect, it is possible to identify three generations of information technology supported learning. The first generation is based on multimedia technology support, such as videos, CD-ROMs or other stand-alone educational software. The second generation employs telematic technologies and it is basically set up as teaching via the Web, where conventional educational material, and entire educational courses, is delivered via the network to remote students. The last, emerging generation, is about Web based learning, where the Internet is used as a means to create active, context based, personalized learning experiences. This last generation of e-learning shifts the emphasis from ‘teaching’ to ‘learning’ and from the notion of technology as a didactic mediator to the notion of a sociable, peer-supported, involved learner. This new learning paradigm inevitably places emphasis on educational resources and virtual communities of practice. Although a lot of effort has been put in the area of educational content development, description, and sharing, currently there is no prominent clear and standards-based solution for the seamless sharing of educational content in medicine and in general.

Current efforts mainly address the problem of educational content sharing via centralized or distributed repositories, but such approaches do not consider notions such as author/learner participation and collaboration, nor do they address issues of usage and inheritance as resources are shared, re-used and repurposed. The MetaMorphosis+ social network presented in this paper presents a novel approach for capturing the multi-faceted social aspect of educational resources, as they form the social-object basis of a human network of educators and learners. The goal is to provide a conceptually different approach to educational resource search and retrieval via ‘social’ associations amongst learning resources and their authors and ultimate consumers.

**Acknowledgments.** The authors acknowledge invaluable comments and input with the mEducator consortium partners. This work is funded in part by the mEducator project (Contract Nr: ECP 2008 EDU 418006 mEducator) under the eContentplus

programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

## References

1. MetaMorphosis+, <http://metamorphosis.med.duth.gr>
2. O'Reilly T.: What is Web 2.0: Design Patterns and Business Models for the Next Generation of software. (2005) <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
3. Engeström, J.: Why Some Social Network Services Work and Others Don't. The Case for Object-Centered Sociality. (2005). [http://www.zengestrom.com/blog/2005/04/why\\_some\\_social.html](http://www.zengestrom.com/blog/2005/04/why_some_social.html)
4. Kim, H.-L., Decker, S., Breslin, J.G.; Representing and Sharing Folksonomies with Semantics", J. Inf. Sci. 36, 57-72, (2010)
5. Bojars U., Breslin, J.G., Finn, A., Decker, S.: Using the Semantic Web for Linking and Reusing Data across Web 2.0 Communities. Web Semantics: Science, Services and Agents on the World Wide Web, 6, 21-28 (2008)
6. Bojars U., Breslin, J.G., Peristeras, V., Tummarello, G., Decker, S.: Interlinking the Social Web with Semantics. IEEE Intelligent Systems, 23, 29-40 (2008)
7. LOM working draft v4.1 (2000) <http://ltsc.ieee.org/doc/wg12/LOMv4.1.htm>
8. Rehak, D.R., Mason, R.: Keeping the Learning in Learning Objects", in Littlejohn, A. (ed): Reusing Online Resources. pp. 20-34, Routledge, London (2003)
9. Jonassen, D., Churchill, D.: Is There a Learning Orientation in Learning Objects? International Journal on E-learning, 3, 32-41 (2004)
10. mEducator Schema, <http://purl.org/meducator/ns>
11. Wolpers, M., Najjar, J., Verbert, K., & Duval, E.: Tracking Actual Usage: the Attention Metadata Approach. Educational, Technology & Society, 10, 106-121 (2007)
12. Kaldoudi, E., Dovrolis, N., Konstantinidis, S., Bamidis, P.D.: Social Networking for Learning Object Repurposing in Medical Education", J. Inform Techn Healthcare, 7, 233-243 (2009)
13. Balasubramaniam, C., Poulton, T., Huwendiek, S.: Repurposing Existing Virtual Patients; an Anglo-German Case Study. Bio-Algorithms and Med-Systems, 5, 91-98 (2009)
14. Zaka, B., Kulathuramaiye, N., Balke, W.-T., Maurer, H.: Topic-Centered Aggregation of Presentations for Learning Object Repurposing. In Proc. World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education (E-Learn), Las Vegas (2008)
15. Jovanović, J., Gašević, D., Verbert, K., Duval, E.: Ontology of Learning Object Content Structure. In Proc. 12th International Conference on Artificial Intelligence in Education, pp. 322-329. Amsterdam (2005)
16. Lenski, W., Wette-Roch, E.: The TRIAL-SOLUTION Approach to Document Re-use Principles and Realization. In Proc. of Workshop on Electronic Media in Mathematics, Coimbra (2001)
17. Najjar, J., Klerkx, J., Vuorikari, R., Duval, E.: Finding Appropriate Learning Objects: an Empirical Evaluation. in Research and Advanced Technology for Digital Libraries, LNCS, vol. 3652, pp. 323-335. Springer Berlin, Heidelberg (2005)

18. Meyer, M., Bergstraesser, S., Zimmermann, B., Rensing, C., Steinmetz, R.: Modeling Modifications of Multimedia Learning Resources Using Ontology-Based Representations. In: Advances in Multimedia Modeling, LNCS vol. 4351, pp. 34-43 (2006)
19. Wang, C., Dickens, K., Davis, H., Wills, G.: Community Tools for Repurposing Learning Objects. In: Second European Conference on Technology Enhanced Learning, pp. 378-392, Crete, Greece (2007)
20. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story so Far. International Journal on Semantic Web and Information Systems (IJSWIS), 5, 1-22 (2009)
21. LOD, <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
22. Noy N F, Shah N H, Whetzel P L, Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A. Chute, D.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic. Acids. Res. 37, W170-3 (2009)
23. Giordano, D., Faro, A., Maiorana, F., Pino, C., Spampinato, C.: Feeding Back Learning Resources Repurposing Patterns into the “Information Loop”: Opportunities and Challenge. In The Proceedings of ITAB2009: 9th International Conference on Information Technology and Applications in Biomedicine, Larnaca, Cyprus (2009)
24. Yu, H.Q., Dietze, S., Li, N., Pedrinaci, C., Taibi, D., Dovrolis, N., Stefanut T., Kaldoudi, E., Dominique, J.: A Linked Data-driven & Service-oriented Architecture for Sharing Educational Resources. In the Proceedings of the Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age, (2011) in press.
25. Pedrinaci, C., Liu, D., Maleshkova, M., Lambert, D., Kopecky, J., Domingue, J.: iServe: a Linked Services Publishing Platform. In the Workshop in Ontology Repositories and Editors for the Semantic Web at 7th Extended Semantic Web Conference (2010)
26. Dietze, S., Yu, H.Q., Pedrinaci, C., Liu, D. and Domingue, J. SmartLink: a Web-based editor and search environment for Linked Services, 8th Extended Semantic Web Conference (ESWC), Heraklion, Greece (2011)
27. Auer, S., Dietzold, S., Lehmann, J., Hellmann, S., Aumueller, D.: Triplify: Light-Weight Linked Data Publication from Relational Databases. In the Proceedings of the 18th International Conference on World Wide Web, pp. 621-630. ACM, NY (2009)

# Generating educational assessment items from Linked Open Data: the case of DBpedia

Muriel Foulonneau

Tudor Research Centre,  
29, av. John F. Kennedy  
L-1855 Luxembourg, Luxembourg  
muriel.foulonneau@tudor.lu

**Abstract.** This work uses Linked Open Data for the generation of educational assessment items. We describe the streamline to create variables and populate simple choice item models using the IMS-QTI standard. The generated items were then imported in an assessment platform. Five item models were tested. They allowed identifying the main challenges to improve the usability of Linked Data sources to support the generation of formative assessment items, in particular data quality issues and the identification of relevant sub-graphs for the generation of item variables.

**Keywords:** Linked Data, open data, DBpedia, eLearning, e-assessment, formative assessment, assessment item generation, data quality, IMS-QTI

## 1 Introduction

Assessment takes a very important role in education. Tests are created to evaluate what students have learned in the class, to assess their level at the beginning of a cycle, to enter a prestigious university, or even to obtain a degree. More and more assessment is also praised for its contribution to the learning process through formative assessment (i.e., assessment to learn, not to measure) and/or self-assessment whereby the concept of a third party controlling the acquisition of knowledge is totally taken out of the assessment process. The role of assessment in the learning process has considerably widened. The New York Times even recently published an article entitled “To Really Learn, Quit Studying and Take a Test” [1], reporting on a study by Karpicke et al. [2] which suggests that tests are actually the most efficient knowledge acquisition method.

The development of e-assessment has been hampered by a number of obstacles, in particular the time and effort necessary to create assessment items (i.e., test questions) [3]. Therefore, automatic or semi-automatic item generation has gained attention over the last years. Item generation consists in using an item model and creating automatically or semi-automatically multiple items from that model.

The Semantic Web can provide relevant resources for the generation of assessment items because it includes models of factual knowledge and structured datasets for the

generation of item model variables. Moreover, it can provide links to relevant learning resources, through the interlinking between different data sources.

Using a heterogeneous factbase for supporting the learning process however raises issues related for instance to the potential disparities of data quality. We implemented a streamline to generate simple choice items from DBpedia. Our work aims at identifying the potential difficulties and the feasibility of using Linked Open Data to generate items for low stake assessment, in this case formative assessment.

We present existing approaches to the creation of item variables, the construction of the assessment item creation streamline, and the experimentation of the process to generate five sets of items.

## 2 Existing work

Item generation consists in creating multiple instances of items based on an item model. The item model defines variables, i.e., the parts which change for each item generated. There are different approaches to the generation of variables, depending on the type of items under consideration.

In order to fill item variables for mathematics or science, the creation of computational models is the easiest solution. Other systems use natural language processing (NLP) to generate for instance vocabulary questions and cloze questions (fill in blanks) in language learning formative assessment exercises ([4], [5], [6]). Karamanis et al. [7] also extract questions from medical texts.

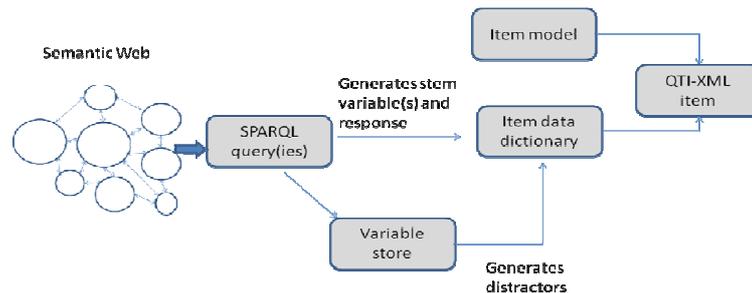
The generation of variables from structured datasets has been experimented in particular in the domain of language learning. Lin et al. [8] and Brown et al. [9] for instance generated vocabulary questions from the WordNet dataset, which is now available as RDF data on the Semantic Web. Indeed, the semantic representation of data can help extracting relevant variables. Sung et al. [10] use natural language processing to extract semantic networks from a text and then generate English comprehension items.

Linnebank et al. [11] use a domain model as the basis for the generation of entire items. This approach requires experts to elicit knowledge in specifically dedicated models. However, the knowledge happens to already exist in many data sources (e.g., scientific datasets), contributed by many different experts who would probably never gather in long modeling exercises. Those modeling exercises would have to be repeated over time, as the knowledge of different disciplines evolves. Moreover, in many domains, the classic curricula, for which models could potentially be developed and maintained by authorities, are not suitable. This is the case of professional knowledge for instance.

Given the potential complexity of the models for generating item variables, Liu [12] defines reusable components of the generation of items (including the heuristics behind the creation of math variables for instance). Our work complements this approach by including the connection to semantic datasets as sources of variables. Existing approaches to item generation usually focus on language learning [13] or mathematics and physics where variable can be created from formulae [14]. We aim to define approaches applicable in a wider range of domains (e.g., history) by reusing existing interlinked datasets.

### 3 Generating item variables from a SPARQL endpoint

An item model includes a stem, options, and potentially auxiliary information [15]. Only the stem (i.e., the question) is mandatory. Response options are provided in the case of a multiple choice item. Auxiliary information can be a multimedia resource for instance. In some cases, other parameters can be adapted, including the feedback provided to candidates after they answer the item.



**Figure 1 - Semi-automatic item generation from semantic datasets**

In order to investigate the use of Linked Data as a source of assessment items, we built a streamline to generate simple choice items from a SPARQL endpoint on the Web. The item generation process is split in different steps detailed in this section. Figure 1 shows the item model represented as an item template, the queries to extract data from the Semantic Web, the generation of a set of potential variables as a variable store, the organization of all the values of variables for each item in data dictionaries, and the creation of items in QTI-XML format from the item template and item data dictionaries. These steps are detailed in this section.

#### 3.1 Creating an IMS QTI-XML template

In order to generate items which are portable to multiple platforms, it is necessary to format them in IMS-QTI (IMS Question & Test Interoperability Specification)<sup>1</sup>. IMS-QTI is the main standard used to represent assessment items [16]. It specifies metadata (as a Learning Object Metadata profile), usage data (including psychometric indicators), as well as the structure of items, tests, and tests sections. It allows representing multimedia resources in a test. IMS-QTI has an XML serialization.

---

<sup>1</sup> <http://www.imsglobal.org/question/>

```

<choiceInteraction responseIdentifier="RESPONSE" shuffle="false" maxChoices="1">
  <prompt>What is the capital of {prompt}?</prompt>
  <simpleChoice identifier="{responseCode1}">{responseOption1}</simpleChoice>
  <simpleChoice identifier="{responseCode2}">{responseOption2}</simpleChoice>
  <simpleChoice identifier="{responseCode3}">{responseOption3}</simpleChoice>
</choiceInteraction>

```

**Figure 2 - Extract of the QTI-XML template for a simple choice item**

No language exists for assessment item templates. We therefore used the syntax of JSON templates for an XML-QTI file (Figure 2). All variables are represented with the variable name in curly brackets. Unlike RDF and XML template languages, JSON templates can define variables for an unstructured part of text in a structured document. For instance, in Figure 2, the {prompt} variable is only defined in part of the content of the <prompt> XML element. Therefore, the question itself can be stored in the item model, only the relevant part of the question is represented as a variable.

### 3.2 Collecting structured data from the Semantic Web

In order to generate values for the variables defined in the item template, data sources from the Semantic Web are used. The Semantic Web contains data formatted as RDF. Datasets can be interlinked in order to complement for instance the knowledge about a given resource. They can be accessed through browsing, through data dumps, or through a SPARQL interface made available by the data provider. For this experiment, we used the DBpedia SPARQL query interface (Figure 3). The query results only provide a variable store from which items can be generated. All the response options are then extracted from the variable store (Figure 1).

```

SELECT ?country ?capital
WHERE {
  ?c <http://dbpedia.org/property/commonName> ?country .
  ?c <http://dbpedia.org/property/capital> ?capital
}
LIMIT 30

```

**Figure 3 - SPARQL query to generate capitals in Europe**

Linked data resources are represented by URIs. However, the display of variables in an assessment item requires finding a suitable label for each concept. In the case presented on Figure 3, the ?c variable represents the resource as identified by a URI. The <http://dbpedia.org/property/commonName> property allows finding a suitable label for the country. Since the range of the <http://dbpedia.org/property/capital> property is a literal, it is not necessary to find a distinct label.

The label is however not located in the same property in all datasets and for all resources. In the example of Figure 3, we used the property `<http://dbpedia.org/property/commonName>` which provides the capital names as literals. However, other properties, such as `<foaf:name>` are used for the same purpose. In any case, the items always need to be generated from a path in a semantic graph rather than from a single triple. This makes Linked Data of particular relevance since the datasets can complete each other.

### 3.3 Generating item distractors

The SPARQL queries aim to retrieve statements from which the stem variable and the correct answer are extracted. However, a simple or multiple choice item also needs distractors. Distractors are the incorrect answers presented as options in the items. In the case of Figure 3, the query retrieves different capitals, from which the distractors are randomly selected to generate an item. For instance, the capital of Bulgaria is Sofia. Distractors can be Bucarest and Riga.

### 3.4 Creating a data dictionary from Linked Data

The application then stores all the variables for the generated items in data dictionaries. Each item is therefore represented natively with this data dictionary. We created data dictionaries as Java objects conceived for the storage of QTI data. We also recorded the data as a JSON data dictionary.

In addition to the variables, the data dictionary includes provenance information, such as the creation date and the data source.

### 3.5 Generating QTI Items

QTI-XML items are then generated from the variables stored in the data dictionary and the item model formalized as a JSON template. We replaced all the variables defined in the model by the content of the data dictionary. If the stem is a picture, this can be included in the QTI-XML structure as an external link.

## 4 The DBpedia experiment

In order to validate this process, we experimented the generation of assessment items for five single choice item models. We used DBpedia as the main source of variables. The item models illustrate the different difficulties which can be encountered and help assessing the usability of the Linked Data for the generation of item variables.

#### 4.1 The generation of variables for five item models

*Q1 - What is the capital of { Azerbaijan }?*

The first item model uses the query presented on Figure 3. This query uses the <http://dbpedia.org/property/> namespace, i.e., the Infobox dataset. This dataset however is not built on top of a consistent ontology. It rather transforms the properties used in Wikipedia infoboxes. Therefore, the quality of the data is a potential issue<sup>2</sup>.

Out of 30 value pairs generated, 3 were not generated for a country (Neuenburg am Rhein, Wain, and Offenburg). For those, the capital was represented by the same literal as the country. Two distinct capitals were found for Swaziland (Mbabane, the administrative capital and Lobamba, the royal and legislative capital). The Congo is identified as a country, whereas it has been split into two distinct countries. Its capital Leopoldville was since renamed Kinshasa. The capital of Sri Lanka is a URI, whereas the range of the capital property is usually a de facto literal. Finally the capital of Nicaragua is represented with display technical instructions “Managua right|20px”. Overall, 7 value pairs out of 30 were deemed defective.

*Q2 - Which country is represented by this flag ?*

```
SELECT ?flag ?country
WHERE {
  ?c <http://xmlns.com/foaf/0.1/depiction> ?flag .
  ?c <http://dbpedia.org/property/commonName> ?country .
  ?c <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://dbpedia.org/class/yago/EuropeanCountries>
}
LIMIT 30
```

Q2 uses the Infobox dataset to identify the label of the different countries. However, the FOAF ontology also helps identifying the flag of the country and the YAGO (Yet Another Great Ontology) [17] ontology ensures that only European countries are selected. This excludes data which do not represent countries.

Nevertheless, it is more difficult to find flags for non European countries, while ensuring that only countries are selected. Indeed, in the YAGO ontology, <http://dbpedia.org/class/yago/EuropeanCountries> is a subclass of <http://dbpedia.org/class/yago/Country108544813>. But most European countries are not retrieved when querying the dataset with <http://dbpedia.org/class/yago/Country108544813>. Indeed, the SPARQL endpoint does not provide access to inferred triples. It is necessary to perform a set of queries to retrieve relevant subclasses and use them for the generation of variables.

Out of 30 items including pictures of flags used as stimuli, 6 URIs did not resolve to a usable picture (HTTP 404 errors or encoding problem).

---

<sup>2</sup> <http://wiki.dbpedia.org/Datasets>

Q3 - Who succeeded to { Charles VII the Victorious } as ruler of France ?

```
SELECT DISTINCT ?kingHR ?successorHR
WHERE {
?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/KingsOfFrance> .
?x <http://dbpedia.org/property/name> ?kingHR .
?x <http://dbpedia.org/ontology/successor> ?z .
?z <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/KingsOfFrance> .
?z <http://dbpedia.org/property/name> ?successorHR
}
LIMIT 30
```

Q3 uses the YAGO ontology to ensure that the resource retrieved is indeed a king of France. Out of 30 results, one was incorrect (*The three Musketeers*). The query generated duplicates because of the multiple labels associated to each king. The same king was named for instance *Louis IX*, *Saint Louis*, *Saint Louis IX*. Whereas de-duplication is a straight forward process in this case, the risk of inconsistent naming patterns among options of the same item is more difficult to tackle. An item was indeed generated with the following 3 options: *Charles VII the Victorious*, *Charles 09 Of France*, *Louis VII*. They all use a different naming pattern, with or without the king's nickname and with a different numbering pattern.

Q4 - What is the capital of { Argentina }? With feedback

```
SELECT ?countryHR ?capitalHR ?pictureCollection
WHERE {
?country <http://dbpedia.org/property/commonName> ?countryHR .
?country <http://dbpedia.org/property/capital> ?capitalHR .
?country <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://dbpedia.org/class/yago/EuropeanCountries> .
?country <http://dbpedia.org/property/hasPhotoCollection> ?pictureCollection
}
LIMIT 30
```

The above question is a variation of Q1. It adds a picture collection from a distinct dataset in the response feedback. It uses the YAGO ontology to exclude countries outside Europe and resources which are not countries. A feedback section is added. When the candidate answers the item, he then receives a feedback if the platform allows it. In the feedback, additional information or formative resources can be suggested. Q4 uses the linkage of the DBpedia dataset with the Flickr wrapper dataset. However the Flickr wrapper data source was unavailable when we performed the experiment.

Q5 - Which category does { Asthma } belong to?

```
SELECT DISTINCT ?diseaseName ?category
WHERE {
?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Disease> .
?x <http://dbpedia.org/property/meshname> ?diseaseName .
?x <http://purl.org/dc/terms/subject> ?y .
?y <http://www.w3.org/2004/02/skos/core#prefLabel> ?category
}
LIMIT 30
```

Q5 aims to retrieve diseases and their categories. It uses SKOS and Dublin Core properties. The Infobox dataset is only used to find labels. Labels from the MESH vocabularies are even available. Nevertheless, the SKOS concepts are not related to a specific SKOS scheme. Categories retrieved range from *Skeletal disorders to childhood*. For instance, the correct answer to the question on *Obesity* is *childhood*.

#### 4.2 The publication of items on the TAO platform

The TAO platform<sup>3</sup> is an open source semantic platform for the creation and delivery of assessment tests and items. It has been used in multiple assessment contexts, including large scale assessment in the PIAAC and PISA surveys of the OECD, diagnostic assessment and formative assessment.

We imported QTI items generated for the different item models in the platform, in order to validate the overall Linked Data based item creation streamline. Figure 4 presents an item generated from Q1 (Figure 3) imported in the TAO platform.

**Test capital names Venezuela**

**What is the capital of Venezuela?**

- Bangui
- Caracas
- Port Moresby

Figure 4 - Item preview on the TAO platform

---

<sup>3</sup> <http://www.tao.lu>

## 5 Data analysis

The experimentation of the streamline was therefore tested with SPARQL queries which use various ontologies and which collect various types of variables. It raised two types of issues for which future work should find relevant solutions: the quality of the data and the relevance of particular statements for the creation of an assessment item.

### 5.1 Data quality challenges

In our experiment, the chance that an item will have a defective prompt or a defective correct answer is equal to the number of defective variables used for the item creation. Q1 uses the most challenging dataset in terms of data quality. 7 out of 30 questions had a defective prompt or a defective correct answer (23,33%).

The chance that an item will have defective distractors is represented by the following formula, where  $D$  is the total number of distractors,  $d(V)$  is the number of defective variables and  $V$  is the total number of variables:

$$D \times \frac{d(V)}{(V - 1)}$$

We used 2 distractors. Among the items generated from Q1, 10 items had a defective distractor (33,33%). Overall, 16 out of 30 items had neither a defective prompt nor a defective correct answer nor a defective distractor (53,33%).

As a comparison, the items generated from unstructured content (text) that are deemed usable without edit were measured between 3,5% and 5% by Mitkov et al. [18] and between 12% and 21% by Karamanis et al. [7]. The difficulty of generating items from structured sources should be lower. Although a manual selection is necessary in any case, the mechanisms we have implemented can be improved.

#### *The ontology*

Q1 used properties from the Infobox dataset, which has no proper underlying ontology. Q1 can therefore be improved by using ontologies provided by DBpedia, as demonstrated by Q2 for which no distractor issue was identified. We present Q1 and Q2 to illustrate this improvement but it should be noted that there is not always a straight equivalent to the properties extracted from the Infobox dataset.

Q5 could be improved either if the dataset would be linked to a more structured knowledge organization system (KOS) or through an algorithm which would verify the nature of the literals provided as a result of the SPARQL query.

#### *The labels*

The choice of the label for each concept to be represented in an item is a challenge when concepts are represented by multiple labels (Q4). The selection of labels and their consistency can be ensured by defining representation patterns or by using datasets with consistent labeling practices.

### *Inaccurate statements*

Most statements provided for the experiment are not inaccurate in their original context but they sometimes use properties which are not sufficiently precise for the usage envisioned (e.g., administrative capital). In other cases, the context of validity of the statement is missing (e.g., Leopoldville used to be the capital of a country called Congo). The choice of DBpedia as a starting point can increase this risk in comparison to domain specific data sources provided by scientific institutions for instance. Nevertheless, the Semantic Web raises similar quality challenges as the ones encountered in heterogeneous and distributed data sources [19]. Web 2.0 approaches, as well as the automatic reprocessing of data can help improve the usability of the Semantic Web statements. This requires setting up a traceability mechanism between the RDF paths used for the generation of items and the items generated.

### *Data linkage*

Data linkage clearly raises an issue because of the reliability of the mechanism on different data sources. Q3 provided 6 problematic URIs out of 30 (i.e., 20%). Q4 generated items for which no URI from the linked data set was resolvable since the whole Flickr wrapper data source was unavailable. This clearly makes the generated items unusable. The creation of infrastructure components such as the SPARQL Endpoint status for CKAN<sup>4</sup> registered data sets<sup>5</sup> can help provide solutions to this quality issue over the longer run.

### *Missing inferences*

Finally, the SPARQL endpoint does not provide access to inferred triples. Our streamline does not tackle transitive closures on the data consumer side (e.g., through repeated queries), as illustrated with Q3. Further consideration should be given to the provision of data including inferred statements. Alternatively, full datasets could be imported. Inferences could then be performed in order to support the item generation process.

Different strategies can therefore be implemented to cope with data quality issues we encountered. Data publishers can improve the usability of the data, for instance with the implementation of an upper ontology in DBpedia. However, other data quality issues require data consumers to improve their data collection strategy, for instance to collect as much information as possible on the context of validity of the data, whenever it is available.

## **5.2 Data selection**

The experiment also showed that the Linked Data statements should be selected. The suitability of an assessment item for a test delivered to a candidate or a group of candidates is measured in particular through such information as the item difficulty.

---

<sup>4</sup> <http://www.ckan.net>

<sup>5</sup> <http://labs.mondeca.com/sparqlEndpointsStatus/index.html>

The difficulty can be assessed through a thorough calibration process in which the item is given to beta candidates for extracting psychometric indicators. In low stake assessment, however, the evaluation of the difficulty is often manual (candidate or teacher evaluation) or implicit (the performance of previous candidates who took the same item). In the item generation models we have used, each item has a different construct (i.e., it assesses a different knowledge). In this case, the psychometric variables are more difficult to predict [20]. A particular model is necessary to assess the difficulty of items generated from Semantic Web sources. For instance, it is likely that for a European audience, the capital of the Cook Islands will raise a higher rate of failure than the capital of Belgium. There is no information in the datasets, which can support the idea of a higher or lower difficulty. Moreover, the difficulty of the item also depends on the distractors, which in this experiment were generated on a random basis from a set of equivalent instances. As the generation of items from structured Web data sources will become more elaborated, it will therefore be necessary to design a model for predicting the difficulty of generated items.

## **6 Conclusion and future work**

The present experimentation shows the process for generating assessment items and/or assessment variables from Linked Data. The performance of the system in comparison with other approaches shows its potential as a strategy for assessment item generation. It is expected that data linkage can provide relevant content for instance to propose formative resources to candidates who failed an item or to illustrate a concept with a picture published as part of a distinct dataset.

The experimentation shows the quality issues related to the generation of items based on such a resource as DBpedia. It should be noted that the measurements were made with a question which raises particular quality issues. It can be easily improved as shown with other questions. Nevertheless the Linked Data Cloud also contains datasets published by scientific institutions, which may therefore raise less data accuracy concerns. In addition, the usage model we are proposing is centered on low stake assessment, for which we believe that the time saved makes it worthwhile having to clean some of the data, while the overall process remains valuable.

Nevertheless, additional work is necessary both on the data and on the assessment items. The items created demonstrate the complexity of generating item variables for simple assessment items. We aim to investigate the creation of more complex items and the relevance of formative resources which can be included in the item as feedback. Moreover, the Semantic Web can provide knowledge models from which items could be generated. Our work is focused on semi-automatic item generation, where users create item models, while the system aims to generate the variables. Nevertheless, the generation of the items from a knowledge model as in [11] requires that more complex knowledge is encoded in the data (e.g., what happens to water when the temperature decreases). The type and nature of data published as Linked Data need therefore to be further analyzed in order to support the development of such models for the fully automated generation items based on knowledge models.

We will focus our future work on the creation of an authoring interface for item models with the use of data sources from the Semantic Web, on the assessment of

item quality, on the creation of different types of assessment items from Linked Data sources, on the traceability of items created, including the path on the Semantic Web datasets which were used to generate the item, and on the improvement of data selection from semantic datasets.

**Acknowledgments.** This work was carried out in the scope of the iCase project on computer-based assessment. It has benefited from the TAO semantic platform for e-assessment (<https://www.tao.lu/>) which is jointly developed by the Tudor Research Centre and the University of Luxembourg, with the support of the Fonds National de la Recherche in Luxembourg, the DIPF (Bildungsforschung und Bildungsinformation), the Bundesministerium für Bildung und Forschung, the Luxembourgish ministry of higher education and research, as well as OECD.

## References

1. Belluck, P. To Really Learn, Quit Studying and Take a Test. *New York Times*. January 20th, 2011
2. Karpicke, J. D., & Blunt, J. R. Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*. (2011)
3. Gilbert, L., Gale, V., Warburton, B., & Wills, G. Report on Summative E-Assessment Quality (REAQ). Joint Information Systems Committee, Southampton. (2008)
4. Aldabe, I., Lopez de Lacalle, M., Maritxalar, M., Martinez, E., Uria, L. Arikurri: an Automatic Question Generator Based on Corpora and NLP techniques, ser. Lecture Notes in computer science, vol. 4053, pp. 584–594. Springer, Heidelberg (2006)
5. Lee, J. S. Y. Automatic correction of grammatical errors in non-native English text. PhD dissertation at The Massachusetts Institute of Technology. (2009)
6. Goto, T., Kojiri, T., Watanabe, T., Iwata, T., & Yamada, T. Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 2(3), 210. (2010)
7. Karamanis, N., Ha, L. A., & Mitkov, R. Generating multiple-choice test items from medical text: a pilot study. In *Proceedings of the Fourth International Natural Language Generation Conference*, pp. 111–113. (2006)
8. Lin, Y.C., Sung, L.C., and Chen, M.C. An Automatic Multiple-Choice Question Generation Scheme for English Adjective Understanding. *Workshop on Modeling, Management and Generation of Problems/Questions in eLearning, the 15th International Conference on Computers in Education (ICCE 2007)*, pages 137-142. (2007)
9. Brown, J. C., Frishkoff, G. A., & Eskenazi, M. Automatic question generation for vocabulary assessment. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 819–826). (2005)
10. Sung, L.-C. Lin, Y.-C., Chen, M. C. The Design of Automatic Quiz Generation for Ubiquitous English E-Learning System. *Technology Enhanced Learning Conference (TELearn 2007)*, pp. 161-168, Jhongli, Taiwan. (2007)
11. Linnebank, F., Liem, J., and Bredeweg, B. Question generation and answering. *DynaLearn, EC FP7 STREP project 231526, Deliverable D3.3*. (2010)
12. Liu, B. SARAC: A Framework for Automatic Item Generation. In *2009 Ninth IEEE International Conference on Advanced Learning Technologies* (pp. 556-558).

Presented at the 2009 Ninth IEEE International Conference on Advanced Learning Technologies (ICALT), Riga, Latvia. (2009)

13. Xu, Y., Seneff, S. Speech-Based Interactive Games for Language Learning: Reading, Translation, and Question-Answering. *Computational Linguistics and Chinese Language Processing* Vol. 14, No. 2, pp. 133-160. (2009)
14. Lai, H., Alves, C., & Gierl, M. J. Using automatic item generation to address item demands for CAT. In *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. (2009)
15. Gierl, M.J., Zhou, J., Alves, C. Developing a Taxonomy of Item Model Types to Promote Assessment Engineering. *Journal of Technology, Learning, and Assessment*, 7(2). (2008)
16. Sarre, S., Foulonneau, M. Reusability in e-assessment: Towards a multifaceted approach for managing metadata of e-assessment resources. *Fifth International Conference on Internet and Web Applications and Services*. (2010)
17. Suchanek, F. M., Kasneci, G., & Weikum, G. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web* (pp. 697–706). (2007)
18. Mitkov, R., An Ha, L., & Karamanis, N. A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, 12(02), 177–194. (2006)
19. Foulonneau, Muriel, Cole, Timothy W. Strategies for reprocessing aggregated metadata. *European Conference on Digital Libraries. Lecture notes in computer science* 3652 , 290-301 (2005)
20. Bejar, I. I., Lawless, R. R., Morley, M. E., Wagner, M. E., Bennett, R. E., & Revuelta, J. A feasibility study of on-the-fly item generation in adaptive testing. *Educational Testing Service*. (2002)

# A Mobile and Adaptive Language Learning Environment based on Linked Data

Davy Van Deursen<sup>1</sup>, Igor Jacques<sup>2</sup>,  
Stefan De Wannemacker<sup>2</sup>, Steven Torrelle<sup>1</sup>, Wim Van Lancker<sup>1</sup>,  
Maribel Montero Perez<sup>2</sup>, Erik Mannens<sup>1</sup>, and Rik Van de Walle<sup>1</sup>

<sup>1</sup> Ghent University - IBBT,  
Gaston Crommenlaan 8/201, B-9050 Ledeborg-Ghent, Belgium  
`firstname.lastname@ugent.be`

<sup>2</sup> ITEC - Interdisciplinary research on Technology, Education and Communication,  
K.U. Leuven Campus Kortrijk, Etienne Sabbelaan 53, B-8500 Kortrijk, Belgium  
`firstname.lastname@kuleuven-kortrijk.be`

**Abstract.** The possibilities within e-learning environments increased dramatically the last couple of years. They are more and more deployed on the Web, allow various types of tasks and fine-grained feedback, and they can make use of audiovisual material. On the other hand, we are confronted with an increasing heterogeneity in terms of end-user devices (smartphones, tablet PCs, etc.) that are able to render advanced Web-based applications and consume multimedia content. Therefore, the major contribution of this paper is an adaptive, Web-based e-learning environment that is able to provide rich, personalized e-learning experiences to a wide range of devices. We discuss the global architecture and data models, as well as how the integration with media delivery can be realized. Further, we give a detailed description of a reasoner, which is responsible for the adaptive selection of learning items, based on the usage environment and the user profile.

**Keywords:** Adaptive, Language Learning, Mobile, Web-based

## 1 Introduction

The last years, the use of e-learning environments has increased spectacularly, not only in formal educational settings, but also in working and private environments. At the same time, the possibilities within these e-learning environments increased dramatically: learning environments for instance have become easier and more pleasant to use, they allow various types of tasks and fine-grained feedback, and they can make use of audiovisual material. Moreover, while e-learning environments were traditionally offered as applications on stand-alone computers, nowadays they are more and more being rendered over the Internet. It is clear that these evolutions are related to technological evolutions, and the wide availability of fast multimedia computers and internet access.

Next to the fact that e-learning environments are more and more deployed over the Web, we are confronted with an increasing heterogeneity in terms of

end-user devices that are able to connect to the Web and consume multimedia content. Therefore, personal devices such as tablet PCs and smartphones could be used as learning devices, next to traditional desktop and laptop devices. Also, the role of personalization within e-learning environments has become more and more important. Personalization can be applied both at the learning level (i.e., adjust learning sessions according to the learner’s capabilities) and at the environmental level (i.e., adjust the rendering of the learning environment according to the characteristics of the usage environment).

The above described challenges are exactly the ones that are currently tackled in the IBBT MAPLE project (Mobile, Adaptive & Personalized Learning Experience<sup>3</sup>), which aims to make adaptive mobile e-learning possible. Therefore, in this paper, we present a Web-enabled e-learning environment that is able to offer personalized learning sessions on any device, primarily focused on language learning making optimal use of digital multimedia. In order to realize such an environment, we need the following key components:

- a *common, machine-understandable data model* that is independent of usage environments and is able to express both learning content and metadata about the learning content;
- a *logging framework* that allows to capture the behaviour and performance of the learner on a detailed level;
- a *reasoner* that is able to select learning items based on the learner’s capabilities and behaviour;
- a *media delivery platform* taking into account usage environment characteristics and restrictions.

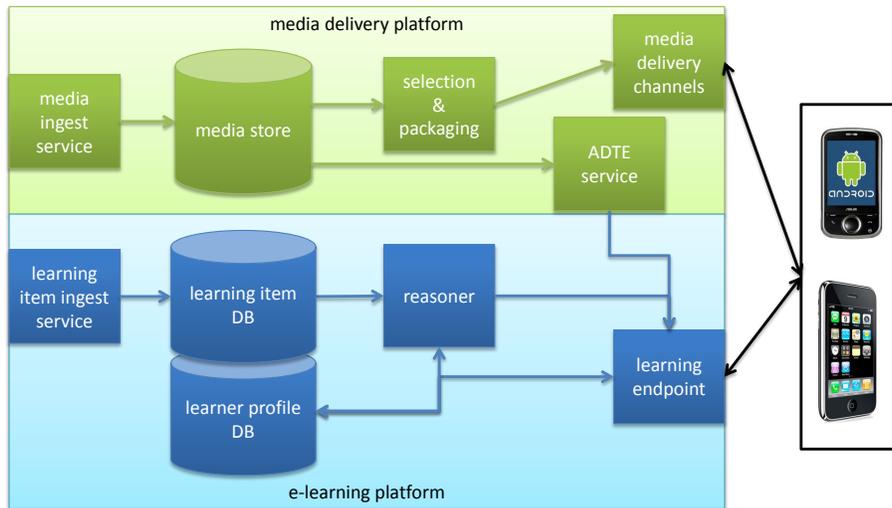
In the remainder of this paper, we provide an overview of the architecture of our adaptive e-learning platform. Further, we discuss the above described key components in more detail. Finally, we discuss related work, draw a number of conclusions, and discuss some future work.

## 2 MAPLE platform

In order to offer a highly adaptive e-learning platform that can also deal with (mobile) multimedia delivery, we designed the architecture that is depicted in Fig. 1. Two major parts can be distinguished: the e-learning platform and the media delivery platform. The e-learning platform relies on two RDF stores, i.e., a store for learning exercises and a store for learner profiles. The learning items store is filled through the learning item ingest service. More details regarding the creation of learning items and the data model according to which they are modeled are provided in Section 3. Further, the learner profile store is build up, based on the learners’ actions and preferences (see Section 3.5). The reasoner is responsible for selecting the most adequate exercise, based on the learner’s profile and environment and the available learning items. Detailed information

---

<sup>3</sup> <http://www.ibbt.be/en/projects/overview-projects/p/detail/maple-2>



**Fig. 1.** The MAPLE e-learning platform

regarding the reasoner is provided in Section 4. Finally, the learning endpoint is the communication point between learner devices and the e-learning platform.

The media delivery platform corresponds to NinSuna<sup>4</sup>, which is a metadata-driven media adaptation and delivery platform [25]. At its core, format-independent modules for temporal selection and packaging of media content are present. Almost all existing media delivery channels are supported by NinSuna: RTSP, RTMP, HTTP progressive download, and HTTP adaptive streaming. Moreover, native support for Media Fragments 1.0 [24] is provided, which enables the delivery of media fragments (i.e., temporal or track fragments) in a standardized way [15]. Finally, NinSuna comes with an Adaptation Decision-Taking Engine (ADTE), which is able to 1) detect the capabilities of the device issuing the request and 2) take a decision regarding which quality version of the requested media resource is the most adequate for the detected device. A more detailed description of the NinSuna platform can be found in [25].

The presented e-learning platform exposes its data (i.e., learning content and accompanying media resources) as linked data. More specifically, it follows the guidelines regarding the publication of linked data<sup>5</sup>: use dereferencable HTTP URIs as names for things, provide useful information using the standards (RDF, SPARQL), and include links to other URIs. Hence, within our platform, the learning items and learner profiles are available through a SPARQL endpoint, while the metadata of the media resources are published as RDF URIs. This way, services such as the reasoner and the ADTE can rely on the linked data and can start reasoning over it.

<sup>4</sup> <http://ninsuna.elis.ugent.be>

<sup>5</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

A typical e-learning scenario using this architecture is then as follows:

- (1) the learner logs in into the Web-based e-learning application using its mobile device, which contacts the learning endpoint of the e-learning platform; the end point approaches the reasoner which provides a personalized overview of the available courses;
- (2) based on the course selected by the learner, the reasoner selects an exercise from the learning item store, taking into account the learner profile and the available exercises within that course;
- (3) when the selected exercise contains media content (audio, video, or images), the ADTE of NinSuna is contacted in order to select the media resource version that fits best for the current device;
- (4) the learning endpoint renders the selected exercise in HTML and sends the response to the learner;
- (5) when the learner is solving the selected exercises, his/her answers and his/her behaviour in terms of clicks and timing is logged and sent back to the e-learning platform;
- (6) the received answers and behaviour information are used to update the learner's profile.

In the next sections, more detailed information regarding a number of components in the architecture is provided.

### 3 Data Models and Instance Generation

A number of different data models need to be developed in order to structure and define the content used on the e-learning platform. More specifically, we need the following data models:

- model for the learning items and their metadata (e.g., question, possible answers, difficulty level);
- model for the learning domain;
- model for the metadata of the media resources (e.g., bit rate);
- model for the learner profile;
- model for the logging.

In the following subsections, we provide more information regarding these different models and how they are populated. Note that all ontologies are modelled in OWL and published online.

#### 3.1 Model for learning items and their metadata

The model for learning items consists of two ontologies: one for the learning items themselves<sup>6</sup> and one for their metadata<sup>7</sup>. An example instance of a learning

<sup>6</sup> <http://multimedialab.elis.ugent.be/organon/ontologies/maple/content>

<sup>7</sup> <http://multimedialab.elis.ugent.be/organon/ontologies/maple/llomp>

**Listing 1.1.** Representing a learning item and its metadata in RDF (in Turtle).

```
1 @prefix mplc: <http://multimedialab.elis.ugent.be/organon/ontologies/  
    maple/content#>.  
  @prefix llomp: <http://multimedialab.elis.ugent.be/organon/ontologies/  
    maple/llomp#>.  
  @prefix xsd: <http://www.w3.org/2001/XMLSchema#>.  
  @prefix dc: <http://purl.org/dc/terms/> .  
5  
  <http://ninsuna.elis.ugent.be/rdf/resource/maple/blcc_47363>  
    a llomp:Exercise ;  
    dc:title "47363" ;  
    mplc:exerciseType mplc:MultipleChoice ;  
10    mplc:media <http://ninsuna.elis.ugent.be/Media/Maple/FLAA2V0#t=0,19> ;  
    mplc:task "What do Belgians eat, according to the reporter?"@en,  
        "Wat eten de Belgen volgens de reporter?"@nl ;  
    mplc:answerSpace "Les Belges mangent ..." ;  
    mplc:input [  
15      a mplc:Input ;  
      mplc:answer [  
        a mplc:Choice ;  
        mplc:isCorrect "false"^^xsd:boolean ;  
        mplc:scoreCorrect "0"^^xsd:int ;  
20        mplc:scoreFalse "0"^^xsd:int ;  
        mplc:text "des frites"@fr .  
      ] ;  
      mplc:answer [  
25        a mplc:Choice ;  
        mplc:isCorrect "true"^^xsd:boolean ;  
        mplc:scoreCorrect "1"^^xsd:int ;  
        mplc:scoreFalse "0"^^xsd:int ;  
        mplc:text "de la glace"@fr .  
      ] .  
30    ] ;  
    mplc:maxScore "1" ;  
    mplc:minScore "0" ;  
  
    llomp:educational [  
35      a llomp:Educational ;  
      llomp:difficulty llomp:medium ;  
      llomp:level llomp:A2 ;  
      llomp:learningComponent :learningComponent_44854 .  
    ] ;  
40    llomp:lifeCycle :lifeCycle_47363 .
```

item modelled according to our model is shown in Listing 1.1. We explain and illustrate both ontologies based on this example. The model is heavily based on the Learning Object Metadata (LOM, [2]). LOM specifies a conceptual data scheme and the corresponding XML-binding for metadata of learning items. We started from LOM and defined a number of extensions in order to provide improved support for learning subject, feedback and scoring, as well as better integration with media resources. Further, as mentioned before, we split our model between learning items and their metadata.

We describe not only the metadata of learning items, but also the exercises themselves. This way, they are formally represented, independent of any rendering. Moreover, they can be easily integrated with their metadata and corresponding media resources. Also, the reasoner (Section 4) will not only rely on the learning item metadata, but also on the items themselves (e.g., this type of

exercise is preferred by the learner). For the moment, six `mplc:exerciseTypes` are supported (focussed on language learning):

- Multiple Choice: given a number of answers, the learner has to choose exactly one answer;
- Multiple Response: given a number of answers, the learner has to choose one or more answers;
- Fill Gaps: given a text with some gaps, the learner needs to fill in missing text in text boxes;
- Dropdown: same as Fill Gaps, but instead of free text fields, the learner can choose between a number of predefined answers;
- Click on Text: given a text, the learner needs to click/tab on one or more words;
- Click on Zone: given an image or video, the learner needs to click/tab at one or more regions within the image or video.

Note that media elements can also occur within the first five types of exercises. For instance, a movie can be played followed by the question to solve. Only the last type (Click on Zone) uses multimedia in an interactive way as described in [19].

In Listing 1.1, a multiple choice exercise is used as example (line 9). A link to a movie fragment is provided via the `mplc:media` property (line 10), which takes as value a Media Fragment URI (see Section 3.3). The `mplc:task` description (line 11) provides the question or task in multiple languages (based on the level of the learner, the reasoner can choose if the language is presented in the native language of the learner or not). Further, the `mplc:answerSpace` (line 13) corresponds to the zone where the learner can enter its answers. Within such an answer space, `mplc:input` is provided (line 14), where each `mplc:answer` corresponds to one possible answer. In case of a multiple choice type, each answer corresponds to a `mplc:Choice`. It contains information such as ‘is this possible answer the correct one?’, ‘how much does the learner score when (s)he selects this one?’, and the possible answer itself. LOM-specific elements such as `llomp:lifeCycle` (line 40) and `llomp:educational` (line 34) are present as well.

As a part of the aforementioned LOM extensions, we added the learning component property to the educational component. Since the MAPLE project focusses on language learning, we extended this learning component property with specific support for language learning. The learning component is split up into three separate subcomponents: target language, theme and language component. The latter component can have one or more of the following subproperties:

- knowledge property: vocabulary, pronunciation, etc.;
- skill property: reading, listening, writing or speaking

We also defined a hierarchical structure for the range of the knowledge property based on which the exact knowledge URIs can be deduced. This was done in a language-independent way extendable with language-specific elements. As the

---

**Listing 1.2.** Representing a learning component in RDF (in Turtle).

---

```
1 @prefix lang: <http://kuleuven-kortrijk.be/itec/ext/ontologies/
   itec_elearning_ontology/languagecomponent/#>.
   @prefix llomp: <http://multimedialab.elis.ugent.be/organon/ontologies/
     maple/llomp#>.

   <http://ninsuna.elis.ugent.be/rdf/resource/maple/learningComponent_40001>
5   a llomp:LearningComponent ;
   llomp:theme "agriculture" ;
   llomp:targetLanguage "en-UK" ;
   llomp:languageComponent [
10     a llomp:LanguageComponent ;
       llomp:knowledge <http://kuleuven-kortrijk.be/itec/ext/ontologies/
         itec_elearning_ontology/languagecomponent/grammar/partsOfSpeech/
         substantive> ;
       llomp:skill lang:writing .
   ] .
```

---

skill and knowledge property exists next to each other, it is possible to specify the subject of an exercise very accurately. In Listing 1.2 an example instance of a learning component can be found. The exercise in this instance trains the writing skill of substantives related to agriculture.

Within the MAPLE project, we use learning items from Televic Education (TEDU)<sup>8</sup>. Currently, TEDU stores their learning items and accompanying metadata in a SQL store. Through XML feeds, the store can be accessed from outside. Hence, we implemented a converter taking as input the XML feeds and producing RDF learning items according to the above described model.

### 3.2 Model for the learning domain

The learning items are not physically arranged into courses. Which learning objects belong together is determined by the metadata, namely the learning component within the educational component of each item. The domain model consists of two type of relations: prerequisite and hierarchical relations. In the project, the domain model is supposed to be simple. It is a three level hierarchical model in which the items are first distinguished by their target language, secondly by their theme, and thirdly by their language component. Additionally, there exist prerequisite requirements between the language components, expressing one language component depends on the knowledge of another. The reasoner will take into account these prerequisites when determining what courses are available for the learner.

### 3.3 Model for media metadata

To model media resources, we rely on the W3C Media Annotations ontology [11], which is supposed to foster the interoperability among various kinds of metadata formats currently used to describe media resources on the Web. Moreover, it

---

<sup>8</sup> <http://www.televic-education.com/en/>

**Listing 1.3.** Representing a learner profile in RDF (in Turtle).

---

```
1 @prefix itec: <http://kuleuven-kortrijk.be/itec/ext/ontologies/
   itec_elearning_ontology#>.
   @prefix foaf: <http://xmlns.com/foaf/0.1/>.
   @prefix mplc: <http://multimedialab.elis.ugent.be/organon/ontologies/
     maple/content#>.

5 <http://kuleuven-kortrijk.be/itec/ext/ontologies/itec_elearning_ontology/
   maple/learners#blcc_piet_lambrecht >
   foaf:nick "piet_lambrecht" ;
   foaf:firstName "Piet" ;
   foaf:lastName "Lambrecht" ;
   itec:hasProficiency [
10     a itec:Proficiency ;
       itec:hasLearningSubject :learningComponent_47584 ;
       itec:hasScoredEvaluation [
15         a itec:ScoredEvaluation ;
           itec:score "3.2"^^xsd:float ;
           itec:scoreVariance "1.1"^^xsd:float ;
           itec:scoreScale itec:defaultEuropeanLanguageLevelScale .
       ] .
   ] ;
   itec:hasLearningGoal [
20     a itec:ScoredEvaluationLearningGoal ;
       itec:hasScoredEvaluation [
           a itec:ScoredEvaluation ;
           itec:score "4"^^xsd:float ;
           itec:scoreScale itec:defaultEuropeanLanguageLevelScale .
25     ] ;
       itec:hasLearningSubject :learningComponent_47584 .
   ] ;
   itec:preferredExerciseType mplc:DropDown .
```

---

already contains mappings to many other existing metadata formats. Further, the ontology also provides support for Media Fragment URIs.

### 3.4 Model for the learner profile

In order to steer the decision making of the reasoner, an up-to-date learner profile is required for each of the learners in the learning system. This profile holds proficiency score estimations for each of the appropriate learning subjects. Each of these values is supplemented with a reliability parameter, namely the variance of the estimator. As we focus on language learning, the proficiency scores are expressed on a continuous scale based on the discrete European Language Levels [4]. The level of A1 conforms to a score of 0, A2 to 1, B1 to 2, etc. Also, the profile keeps a list of the learning goals which were set for that learner. An example of such a learning goal could be “Achieve the B2 level for the French verb form imparfait”. The type of learning items the learner prefers can also be saved in the profile. An example instance can be found in Listing 1.3.

The properties in the model will be caught either automatically either by means of preference setting. The learner’s favourite learning item types can be edited through a preference menu and the learner’s proficiency scores will be updated by a module of the reasoner. Additionally, the ontological model sup-

**Listing 1.4.** Representing a logging abstract in RDF (in Turtle).

---

```
1 @prefix itec: <http://kuleuven-kortrijk.be/itec/ext/ontologies/
   itec_elearning_ontology#>.
   @prefix learners: <http://kuleuven-kortrijk.be/itec/instances/maple/
     learners#>.
   @prefix log: <http://kuleuven-kortrijk.be/itec/instances/maple/logging#>.
   @prefix maple: <http://ninsuna.elis.ugent.be/rdf/resource/maple/>.
5
log:learnerSession12452
  a itec:LearnerSession ;
  itec:hasSessionStart "2010-10-26T21:32:52.126"^^xsd:dateTime ;
  itec:hasSessionStop "2010-10-26T21:38:52.526"^^xsd:dateTime ;
10 itec:hasLearner learners:blcc_piet_lambrecht ;
   itec:hasSubSession [
     a itec:LearningSession ;
     itec:hasSessionStart "2010-10-26T21:32:52.229"^^xsd:dateTime ;
     itec:hasSessionStop "2010-10-26T21:38:52.501"^^xsd:dateTime ;
15 itec:hasItemObjectSession [
     a itec:ItemObjectSession ;
     itec:hasItemObject maple:blcc_47363 ;
     itec:hasSessionStart "2010-10-26T21:32:56.233"^^xsd:dateTime ;
     itec:hasSessionStop "2010-10-26T21:32:59.999"^^xsd:dateTime ;
20 itec:hasAnswerSubmittedEvent [
     itec:hasInputObject maple:inputObject_57495 ;
     itec:hasGivenAnswer maple:answer_57495 ;
     itec:dateTime "2010-10-26T21:32:59.526"^^xsd:dateTime .
   ] .
25 ] .
] .
```

---

ports properties like motivation, learning style, learner strategy, and cognitive ability's, but currently these are not used in the MAPLE e-learning platform.

### 3.5 Model for logging the learner's activity

Finally, we developed a model for describing logging information. For instance, the model is able to express information such as the start and stop of a learner session or the learner's course selection. Once the learner has chosen a course, a learning session is initiated in which the reasoner successively selects a new learning item, each time resulting in a learning item session which lasts for the time the learner interacts with the item. During such an item session a learner can give an answer, request a hint, or change his mind by changing his answer. All these interactions are logged by the system. This results in a huge amount of information which is consumed in two ways. Firstly, a part of the logging information is used at run-time by the reasoner. For instance, a score attained by the learner will affect the proficiency score of a learner's profile through the functionality of the reasoner's proficiency manager. Secondly, after runtime, the logged information will be used as input for statistical research tracing how certain interactions of the learner give information about the learning process. In Listing 1.4 an example instance can be found. The learner and learning session, and the session of the item are respectively interconnected by the `itec:hasSubSession` and the `itec:hasItemObjectSession` relation.

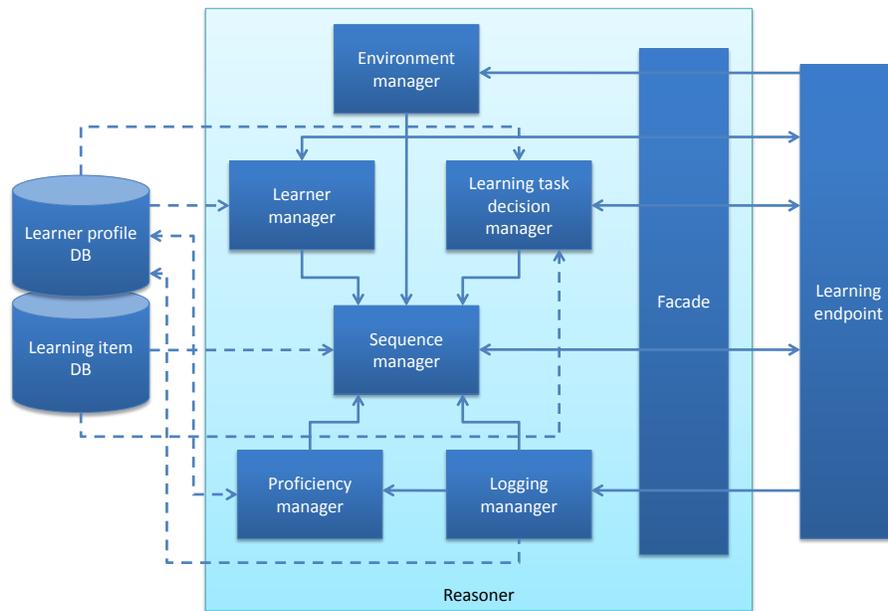
These resulting triples are partially generated in the core of the reasoner, e.g. the start and stop of the learner and the learning sessions. The low level interactions concerning one specific exercise are generated at the client and sent back to the reasoner which processes the logging and stores it in the learner profile RDF store.

## 4 Adaptive Learning Item Selection

The reasoner, introduced in Section 2, is a crucial component within the MAPLE learning system architecture as it is responsible for the adaptive learning item selection. If a learner logs in, the reasoner will first of all provide a short list of courses from which the learner can choose. As the reasoner is aware of the learning goals for each learner through the learner profile model, only courses that contribute to the not yet attained learning goals can be selected. Next, once the learner has chosen a course, the reasoner will start up a learning session and will successively decide on the exact exercise to deliver to the learner.

The reasoner takes into account the learner profile as well as some real time environmental properties. For the environmental adaptivity, both the screen capacity and connection quality of the user's device are sources of adaptivity. In case the screen size is too small, the reasoner will avoid the use of exercises with media. A slow network connection will also result in avoiding media exercises. For the learner profile adaptivity, there are two main policies which can steer the decision process. The first one is based on a theory stating that the exercise difficulty needs to be increased each time a learner has answered a series of four exercises correctly. Similarly, when four consecutive exercises are answered incorrectly, it should go down [12]. The second policy is based on a pedagogical theory which tries to keep the learner's motivation high by chasing a predefined (e.g. 70 %) correct-answer probability. This probability can be estimated based on the IRT theory ([5]) by combining the current proficiency estimation with the level and difficulty of the exercise [28, 7]. The aforementioned policies are supplemented with an event-driven feedback system. The system allows the sequencer to shift in a feedback item (instead of an exercise) to explain a learning subject once a specific and predefined condition is met. For instance, "the learner made five errors in a row against the same learning subject". This feedback item is chosen based on the learning component property which both the feedback and the exercise item have in their metadata. For both policies, also the preferred exercise types of the learner are taken into account by favouring them though not completely cold-shouldering the other exercise types.

To fulfil the aforementioned tasks, the architecture of the reasoner (shown in Fig. 2) consists of six modules, supplemented by a facade for communicating with the learning endpoint. The six reasoner modules are the Learner manager, Environment manager, Learning task decision manager, Sequence manager, Logging manager and Proficiency manager. We elucidate the functionality of these modules by means of the following example.



**Fig. 2.** The reasoner architecture

Suppose a learner’s initial profile was set by a teacher thereby providing the learning goal “Achieve the B2 level for the French verb form imparfait” and also providing an estimation for the learner’s initial level, namely A2, for “the French verb form imparfait”. When the learner logs in, the Learner manager produces a learner session. Consequently, the Learning task decision manager loads the learner’s learning goals in order to compose a three-level tree representation of all courses relevant for this learner, as explained in Section 3.2. This tree is sent to the learning endpoint which produces a representation such that the learner can navigate through the tree. Let us assume that the learner first selects ‘French’ followed by the theme ‘General’ and finally the language component ‘Imparfait’. Besides, the learner opens the preferences menu and sets the dropdown exercise type as his favourite one.

Next, the Learning task decision manager composes a learning task object which is sent to the Sequence manager. Here the learning task is sequencing the items (exercises and feedback) with the first policy of adaptivity, starting from level A2, having as a stop criterion the achievement of the level B2, and taking into account the learner’s preferred exercise types and environmental properties. Subsequently, the Sequence manager loads the sequencer necessary for the learning task. To this end, the sequencer makes use of the Environment manager, which is an access point for information on the current connection quality and the screen size of the device of the learner. At this point, the sequencer can successively decide on the id of the next item and passes its choice to the learning

endpoint, which automatically generates a visual representation and makes use of the delivery platform in case media are present.

Once the learner finishes the exercise or has read the feedback in case of a feedback item, the logging information about the interactions of the learner with the item are sent back to the Logging manager of the reasoner. The latter sends this information as a specific logging object to a couple of observer objects which all have different functionalities. For instance, there is an observer writing these logs to the learner profile RDF store. Another observer warns the sequencer when for example four exercises have been consecutively answered correctly and yet another sends the learner's score to the Proficiency manager together with the level, difficulty and the learning subject of the answered exercise. The Proficiency manager keeps the proficiency scores up to date. Prior to every decision of the sequencer, the stop criterion is tested based on a proficiency that is retrieved from the Proficiency manager. If this criterion is reached, the sequencer sequences a special concluding feedback item announcing the end of the learning session to the learner.

## 5 Related Work

The architecture of the reasoner builds further on existing proposals for generic learning system architectures such as in [20]. These architectures however have mostly been designed having an adaptive hypermedia learning system in mind. Even though most systems currently developed are based on providing learner control based on adaptive links, e.g. [3], our system is specialized in adaptive curriculum sequencing, meaning that the learning objects are sequenced in an automated way. To create an adaptive learning system the method of using ontologies has often been proposed in literature, e.g., in [23, 17, 8]. We partially rely on existing ontologies and data models, and introduced new data models such as a model for describing learning exercises and language-learning specific information. The latter were all done in collaboration with educationalists. Additionally, both the delivery platform and the reasoner take into account connection quality and screen size either to choose the right video format either to avoid sending any media to a device if they cannot be delivered in an optimal way. This way, our system implements a part of the context-awareness which has been claimed to be crucial in mobile learning [23, 27].

The ontology for the learner profile is a compact non-exhaustive synopsis of the most common learner characteristics found in literature [21, 13, 10] which can be used in steering an adaptive learning system. For the preservation of the learner's knowledge we used what is classified as an overlay model in [13]. Until now, the IEEE Learning Object Model standard LOM is considered to be the standard for many repositories storing thousands of learning objects with metadata. There have been attempts to transform the LOM metadata model into an RDF version (e.g., [18]). However, the model provided by LOM was not sufficient. Hence, we adopted part of the LOM model (by relying on previous LOM RDF efforts) and extended it with our own needs.

Our realizations in this project largely replace the functionality of the restrictive SCORM standard [1]. SCORM, an abbreviation for Sharable Content Object Reference Model, is a collection of specifications imposing a format for bundling Web-based exercises into courses, thereby imposing LOM for the metadata, as well as a data model for communicating learning scores between server and client. The standard was updated in 2004, now supporting a limited set of instructions for adaptive behavior. In practise however, the imposed syntax for adaptivity had low expressivity but remaining very complicated [14]. Although in the past SCORM had an important impact on the sharing of bundled learning courses on the web and although many tried to improve the SCORM standard [16, 22, 29], we think its starting point has become outdated. After all, we believe grouping learning objects in a container format conflicts with the principle of the Semantic Web of data in which objects are scattered over the web. Additionally, its extensibility pointed out to be low [14, 6] and the data model for exchanging learning results is limited to the exchange of a single score, thereby not fulfilling our needs of more advanced reporting of a learner's interactions with the exercises. Our formalized representation model for recording scores and interactions with exercises makes it possible to develop true interoperable exercises that are able to report learning results in a universal way. Until now, the importance for adaptive learning systems having an extendible although universally understandable learning result reporting system was largely ignored.

Gang *et al.* proposed a framework for mobile learning in [9] that approaches the challenges similarly as we did here. More specifically, a media delivery system was developed, as well as an adaptive module for learning item selection. However, they relied on MPEG-21 technology while we use the NinSuna platform, which is based on MPEG-21 principles but proven to be more efficient and generic [26]. Further, learning item selection is not based on educational properties such as skills or experience, but solely on environmental properties.

## 6 Conclusions and Future Work

In order to exploit the possibilities of Web-based e-learning environments, we proposed an e-learning architecture that is able to provide rich, personalized e-learning experiences to a wide range of devices. We discussed the various data models used within the e-learning framework. Moreover, we provided details of the reasoner, a crucial component allowing to select learning items based on the usage environment and the learner profile.

Future work consists of exploiting the possibilities of the Semantic Web even more by linking learning items to the Linked Open Data cloud. Further, data models could be optimized and linked to upcoming efforts (e.g., how to represent the life cycle of a learning item as provenance information on the Web). Also, more detailed domain models should be investigated. Regarding the reasoner, future work consists of taking into account more information obtained from the logging framework, as well as investigating how error-specific feedback could be generated (e.g., link frequently occurring errors to answers).

## Acknowledgments

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT, 50% co-funded by industrial partners), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

## References

1. Scorm 2004 4th edition version 1.1 overview, <http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Overview.aspx>
2. Standard for learning object metadata, <http://standards.ieee.org/findstds/standard/1484.12.1-2002.html>
3. Grapple, a generic responsive adaptive personalized learning environment. <http://www.grapple-project.org> (Jun 2010)
4. European language levels - self assessment grid (2011), <http://europass.cedefop.europa.eu/LanguageSelfAssessmentGrid/en>
5. Baker, F.B., Kim, S.H. (eds.): Item Response Theory: Parameter Estimation Techniques, Second Edition (Statistics: A Series of Textbooks and Monographs). CRC Press, 2 edn. (July 2004)
6. Bohl, O., Scheuhase, J., Sengler, R., Winand, U.: The sharable content object reference model (scorm) - a critical review. In: Computers in Education, 2002. Proceedings. International Conference on. pp. 950 – 951 vol.2 (2002)
7. Chen, C.M., Lee, H.M., Chen, Y.H.: Personalized e-learning system using item response theory. *Computers & Education* 44(3), 237 – 255 (2005)
8. Chi, Y.L.: Ontology-based curriculum content sequencing system with semantic rules. *Expert Syst. Appl.* 36, 7838–7847 (May 2009)
9. Gang, Z., Zongkai, Y.: Learning Resource Adaptation and Delivery Framework for Mobile Learning. In: *Frontiers in Education, 2005. FIE '05. Proceedings 35th Annual Conference* (October 2005)
10. Jia, B., Zhong, S., Wang, W., Yang, B.: The construction and evolution of learner model in adaptive learning system. *Computer Technology and Development, International Conference on* 1, 148–152 (2009)
11. Lee, W., Bürger, T., Sasaki, F., Malaisé, V., Stegmaier, F., Söderberg, J. (eds.): *Ontology for Media Resource 1.0. W3C Working Draft, World Wide Web Consortium* (June 2010)
12. Leutner, D.: Instructional design principles for adaptivity in open learning environments. *Curriculum, Plans, and Processes in Instructional Design: International Perspectives* pp. 289–307 (2004)
13. Loc Nguyen, P.D.: Learner model in adaptive learning. *World Academy of Science, Engineering and Technology* 45 (2008)
14. Mackenzie, G.: Scorm 2004 primer, a (mostly) painless introduction to scorm. Tech. rep. (2004), <http://www.pro-ductivity.com/Compliance21CFR/CTMW/scormintro.pdf>
15. Mannens, E., Van Deursen, D., Troncy, R., Pfeiffer, S., Parker, C., Lafon, Y., Jansen, J., Hausenblas, M., Van de Walle, R.: A URI-Based Approach for Addressing Fragments of Media Resources on the Web. To appear in *Multimedia Tools and Applications – Special Issue on Multimedia Data Semantics*

16. Melia, M., Barrett, R., Pahl, C.: A model-based approach to scorm sequencing. In: In Proceeding of the Sixth Annual Irish Educational Technology Users Conference (EdTech06) - Research Track. ILTA (2006)
17. Min, W.X., Wei, C., Lei, C.: Research of ontology-based adaptive learning system. In: Proc. Int. Symp. Computational Intelligence and Design ISCID '08. vol. 2, pp. 366–370 (2008)
18. Nilsson, M., Palmer, M., Brase, J.: The LOM RDF binding - principles and implementation. In: Proceedings of 3rd Annual Ariadne Conference (2003)
19. Perez, M.M., Cornillie, F., Senecaut, M.P., De Wannemacker, S., Desmet, P.: Mobile vocabulary learning: activities allowing for interaction through input enhancement. In: De Wannemacker, S., Clarebout, G., De Causmaecker, P. (eds.) *Interdisciplinary Approaches to Adaptive Learning: A Look at the Neighbours.*, pp. 15–27. Springer-Verlag (2011)
20. Rani, S., Ashok, M., Palanivel, K.: Adaptive content for personalized e-learning using web service and semantic web. pp. 1–4 (jul 2009)
21. Roll, I., Baker, R.S., Alevan, V., Koedinger, K.R.: A metacognitive act-r model of students' learning strategies in intelligent tutoring systems. In: *Intelligent Tutoring Systems.* pp. 854–856 (2004)
22. Silva, L., Mustaro, P.N., Stringhini, D., Silveira, I.F.: Using conceptual lattices to represent fine granular learning objects through scorm meta-objects. *The Electronic Journal of e-Learning* 4, 141–148 (2006)
23. Soyly, A., Vandewaetere, M., Wauters, K., Jacques, I., De Causmaecker, P., Desmet, P., Clarebout, G., Van Den Noortgate, W.: *Ontology-driven Adaptive and Pervasive Learning Environments - APLEs: An Interdisciplinary Approach.* In: De Wannemacker, S., Clarebout, G., De Causmaecker, P. (eds.) *Interdisciplinary Approaches to Adaptive Learning: A Look at the Neighbours.*, pp. 99–115. Springer-Verlag (2011)
24. Troncy, R., Mannens, E., Pfeiffer, S., Van Deursen, D. (eds.): *Media Fragments URI 1.0. W3C Working Draft, World Wide Web Consortium* (January 2011)
25. Van Deursen, D., Van Lancker, W., De Neve, W., Paridaens, T., Mannens, E., Van de Walle, R.: *NinSuna: a Fully Integrated Platform for Format-independent Multimedia Content Adaptation and Delivery based on Semantic Web Technologies.* *Multimedia Tools and Applications – Special Issue on Data Semantics for Multimedia Systems* 46(2-3), 371–398 (January 2010)
26. Van Deursen, D., Van Lancker, W., De Bruyne, S., De Neve, W., Mannens, E., Van de Walle, R.: *Format-independent and Metadata-driven Media Resource Adaptation using Semantic Web Technologies.* *Multimedia Systems* 16(2), 85–104 (2010)
27. Wang, Y.K.: *Context awareness and adaptation in mobile learning.* In: *Wireless and Mobile Technologies in Education, 2004. Proceedings. The 2nd IEEE International Workshop on.* pp. 154 – 158 (2004)
28. Wauters, K., van den Noortgate, W., Desmet, P.: *The use of irt for adaptive item selection in item-based learning environments.* In: *Proceeding of the 2009 conference on Artificial Intelligence in Education.* pp. 785–786. IOS Press, Amsterdam, The Netherlands, The Netherlands (2009)
29. Zhu, X.H., Yan, R.Q., Wu, L.P., Zhang, L.F.: *An interoperable learning sequencing model based on the flow control object.* *Computer Science and Software Engineering, International Conference on* 5, 1217–1220 (2008)