

Linked Education: interlinking educational Resources and the Web of Data

Stefan Dietze¹, Hong Quing Yu², Daniela Giordano³, Eleni Kaldoudi⁴, Nikolas Dovrolis⁴, Davide Taibi⁵

¹ L3S Research Center, Leibniz University, Hanover, Germany
dietze@l3s.de

² Knowledge Media Institute, The Open University, MK7 6AA, Milton Keynes, UK
h.q.yu@open.ac.uk

³ University of Catania, Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Viale A. Doria 6, 95125, Catania, Italy
dgiordan@diit.unict.it

⁴ Democritus University of Thrace, Greece
kaldoudi@med.duth.gr, ndovroli@alex.duth.gr

⁵ Italian National Research Council, Institute for Educational Technologies, Italy
davide.taibi@itd.cnr.it

ABSTRACT

Research on interoperability of technology-enhanced learning (TEL) repositories throughout the last decade has led to a fragmented landscape of competing approaches, such as metadata schemas and interface mechanisms. However, so far Web-scale integration of resources is not facilitated, mainly due to the lack of take-up of shared principles, datasets and schemas. On the other hand, the Linked Data approach has emerged as the de-facto standard for sharing data on the Web and offers a large potential to solve interoperability issues in the field of TEL. In this paper, we describe a general approach to exploit the wealth of already existing TEL data on the Web by allowing its exposure as Linked Data and by taking into account automated enrichment and interlinking techniques to provide rich and well-interlinked data for the educational domain. This approach has been implemented in the context of the mEducator project where data from a number of open TEL data repositories has been integrated, exposed and enriched by following Linked Data principles.

Categories and Subject Descriptors

H.3.5 [On-line Information Services]: *data sharing, web-based services*. E.1 [Data]: *distributed data structures*. K.3 [Computers and Education].

General Terms

Management, Design, Experimentation, Standardization.

Keywords

Linked Data, Education, Semantic Web, SOA, eLearning, Technology-enhanced Learning, TEL, Web 2.0.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAC'12, March 25-29, 2012, Riva del Garda, Italy.

Copyright 2012 ACM 978-1-4503-0857-1/12/03...\$10.00.

1. INTRODUCTION

Throughout the last decade, research in the field of technology-enhanced learning (TEL) has focused fundamentally on enabling interoperability and reuse of learning resources and data.

That has led to a fragmented landscape of competing metadata schemas, such as, Dublin Core¹, IEEE Learning Object Metadata (LOM) [8] or ADL SCORM² and query interface mechanisms such as OAI-PMH³ or SQI⁴ which are exploited by educational resource repository providers to support interoperability. To this end, although a vast amount of educational content and data is shared on the Web in an open way, the integration process is still costly as different learning repositories are isolated from each other and based on different implementation standards [14].

In the past years, TEL research has widely attempted to exploit Semantic Web [1] technologies in order to solve interoperability issues. However, while the Linked Data (LD) [2] approach has widely established itself as the de-facto standard for sharing data on the Semantic Web, it is still not widely adopted by the TEL community. This is despite the fact that it approach provides a set of well-established principles and (W3C) standards (RDF, SPARQL [19], use of URIs) aiming at Web-scale data interoperability which have produced an ever growing amount of data sets and schemas. Thus, the Linked Data approach offers a strong potential to substantially alleviate the challenges addressed above.

While there is already a large amount of educational data available on the Web via proprietary and/or competing schemas and interface mechanisms, the main challenge for the TEL field is to (a) start adopting LD principles and vocabularies while (b) leveraging on existing educational data available on the Web via non-LD compliant means. Following such an approach, four

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

² <http://www.adlnet.org>

³ <http://www.openarchives.org/pmh/>

⁴ <http://www.cen-Itso.net/main.aspx?put=859>

major research challenges need to be taken into consideration to ensure Web-scale interoperability:

- a) *Integrating distributed data from heterogeneous educational repositories*: educational data and content is usually exposed by heterogeneous services/APIs such as OAI-PMH or SQL. Therefore, interoperability is limited and Web-scale sharing of resources is not widely supported yet [13].
- b) *Dealing with continuous change*: in highly distributed Web-based environments, frequent changes occur to available Web APIs. That is, services as well as repositories are usually added, modified or removed regularly.
- c) *Metadata mediation and transformation*: educational resources and the services exposing those resources are usually described by using distinct, often XML-based schemas. Therefore, schema/data transformation (into RDF) and mapping are important requirements in order to leverage on already existing TEL data.
- d) *Enrichment and interlinking of unstructured metadata*: existing educational resource metadata is usually provided based on informal and poorly structured data and use of controlled vocabularies is limited and fragmented. Therefore, to allow machine-processing and Web-scale interoperability, educational metadata needs to be enriched, that is transformed into structured and formal descriptions by linking it with widely established LD vocabularies and datasets on the Web.

In this paper we introduce a general approach for a *Linked Education*⁵ environment which has been implemented as part of the EC-funded mEducator⁶ project and aims at tackling the above challenges by following the below principles:

- (1) LD-principles are applied to model and expose metadata of both educational resources and services/APIs. In this way, not only resources are interlinked but also services and resources are exposed in a standardized and accessible way.
- (2) Heterogeneous learning repositories, i.e. their interfaces (services) are integrated on the fly by reasoning and processing of LD-based service semantics (see 1).
- (3) Metadata retrieved from heterogeneous Web repositories is automatically lifted into RDF and exposed as LD accessible via de-referencable URIs.
- (4) Automated enrichment and clustering mechanisms are exploited in order to interlink data produced by (3) with existing datasets as part of the LD cloud.

We discuss related work in Section 2 while Section 3 illustrates our overall approach. In Section 4 we introduce an implementation dealing with integration of educational services while in Section 5 we focus on integration of educational data. Section 6 illustrates a proof-of-concept prototype application which makes use of our data and services integration approach while we discuss and conclude the paper with Section 7.

2. RELATED WORK

Web-scale search of educational resources faces a heterogeneous landscape of Web APIs of individual repositories. For instance,

⁵ <http://linkededucation.org>: an open platform to share results focused on educational LD. Long-term goal is to establish links and unified APIs and endpoints to educational datasets.

⁶ <http://www.meducator.net>

the PubMed⁷ repository provides an OAI-PMH-based service where response messages are based on XML in OAI-DC (OAI Dublin Core) while other repositories offer JSON-based feeds or SPARQL endpoints. In addition, current metadata stores largely use XML and relational databases and often consist of poorly structured text lacking formal semantics.

Services and APIs appear and are removed from the Web frequently and might change behavior and interfaces according to new requirements. Therefore, it is crucial to aim at resolving the underlying heterogeneity. Facilitating easy-to-use service representations based on standard service vocabularies (e.g. SAWSDL [15] and WSMO-Lite [10]) is an important requirement to allow service providers and consumers to interact. Here, we are applying LD technologies to both (a) educational service and APIs and (b) educational data in order to facilitate data as well as services interoperability. The LD principles have led to an ever increasing amount of LD-compliant schemas and data-sets⁸ as well as general-purpose tools and APIs.

Efforts were made already to improve interoperability in the field of education, e.g., by exploiting semantic technologies. For instance, an IEEE LOM-RDF binding⁹ was drafted, but was (a) discontinued and (b) only focused on the binding aspect rather than further working towards a fully LD-compliant approach. A peer-to-peer (P2P) architecture (LOP2P) for sharing educational resources among different learning institutions is proposed in [14]. A similar P2P architecture was also proposed in the EduLearn project [13] and [6]. Meanwhile, Simple Query Interface (SQI) is introduced in [18] 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. These approaches share a number of disadvantages. For instance, instead of accepting the heterogeneous landscape of the Web, they impose either a common schema or interface approach on the underlying stores.

The work described in [16] utilizes Semantic Web as well as service technologies to enable adaptation to different learning contexts through a matching mechanism between a specific context and available learning data. However, this work neither considers approaches for automatic service discovery nor is it based on common standards. Also, mediation between different metadata standards is not supported. [5] follows a similar approach but is fundamentally based on a single shared ontology. These issues apply as well to “Smart Spaces” [17] for learning. [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 [7] allows a mediation between different services based on a “connector service”.

From a more pragmatic angle, educational institutions started to expose their data based on Linked Data principles, such as The Open University (UK)¹⁰, the National Research Council (CNR,

⁷ <http://www.ncbi.nlm.nih.gov/pubmed/>

⁸ <http://richard.cyaniak.de/2007/10/lod/>

⁹ <http://dublincore.org/educationwiki/DCMIIEEELTSTTaskforce/RDFPAR>

¹⁰ <http://data.open.ac.uk/>

Italy)¹¹ or Southampton University (UK)¹². However, while this is a crucial step towards well-interlinked educational Web data, it is important to note that these efforts mainly focus on exposing data of individual institutions while interlinking with 3rd party data is not yet within the primary scope.

3. OVERALL APPROACH

Our general-purpose approach aims at (i) integrating heterogeneous educational Web resources and (ii) exposing its metadata as well-structured and interlinked Linked Data. The proposed architecture includes three layers: *Educational (Web) data and service layer*, *Educational data and service integration layer* and *Educational application and presentation layer* that are shown in Figure 1.

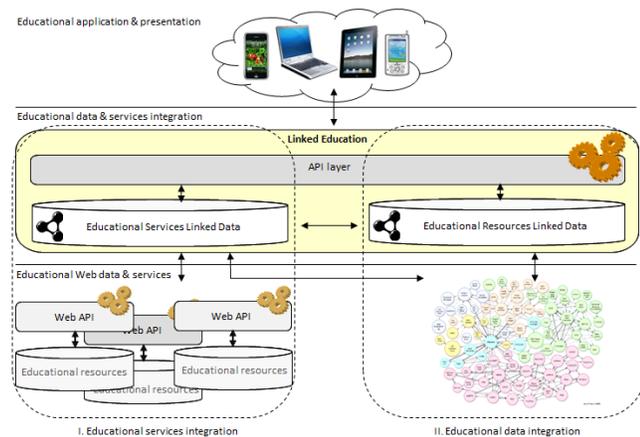


Figure 1. Approach overview

- The *Educational (Web) data and service layer* consists of available educational Web services and data, such as metadata of existing educational objects provided by open public educational repositories, such as PubMed¹³ or OpenLearn¹⁴.
- The *Educational data and service integration layer* is fundamentally based on exploiting LD principles to annotate and interlink educational services and data.
- The *Educational application and presentation layer* uses the APIs provided by the educational data & services integration layer to interact with underlying data & services and provides an interface to end-users.

The proposed approach supports a particular methodology consisting of two fundamental steps which are both facilitated by Linked Data technologies:

- Step I. Educational Services Integration
(facilitated by *Educational Services Linked Data* on the left)
- Step II. Educational Data Integration
(facilitated by *Educational Resources Linked Data* on the right)

¹¹ <http://data.cnr.it>

¹² <http://data.southampton.ac.uk/>

¹³ <http://www.pubmed.gov>

¹⁴ <http://www.open.ac.uk/openlearn>

Integration of educational data and content needs to consider two challenges: integration at the repository-level facilitated by repository-specific APIs and integration at the (meta)data-level. *Step I* aims at integrating educational services and APIs in order to facilitate repository-level integration. To this end, it is concerned with resolving heterogeneities between individual API standards (e.g. SOAP-based services vs. REST-ful approaches) and distinct response message formats and structures (such as JSON, XML or RDF-based ones). In order to enable integration of such heterogeneous APIs, we exploit LD principles to annotate individual APIs in terms of their interfaces, capabilities and non-functional properties (*Educational Services Linked Data*). That allows to automatically discover and execute APIs for a given educational purpose while resolving heterogeneities between individual API responses (Section 4). All educational data retrieved in *Step I* is transformed from their native formats into RDF.

Step II deals with the actual integration of heterogeneous educational data (as retrieved by Step I) by exposing retrieved educational (RDF) data as well-interlinked LD. As starting point, all generated RDF is stored in a dedicated, public RDF store (*Educational Resources Linked Data*) which supports two main purposes: exposing existing educational (non-RDF) data in a LD-compliant way and allowing content/data providers to publish new educational resource metadata. To enrich and interlink the educational data, two approaches are being followed:

- II.a Automated interlinking of datasets
II.b Automated clustering and classification.

While exposing educational data as RDF is one substantial requirement to follow LD principles, mere transformation of data does not improve its quality. Therefore, we enrich descriptions by automated data enrichment techniques to establish links with established vocabularies available on the LD cloud. Enrichment takes advantage of available APIs which allow access to a vast number of established taxonomies and vocabularies. That way, unstructured free text is enriched with unique URIs of structured LD entities to allow not only further reasoning on related concepts but also enables users to query for resources by using well-defined concepts and terms. In addition, automated clustering and classification mechanisms are exploited in order to enable data and resource classification across previously disconnected repositories.

4. EDUCATIONAL SERVICES INTEGRATION

Our implementation of the educational data and service integration layer builds on existing research [20] and applies LD principles to the services domain. Based on RDF descriptions describing core elements of services and APIs, these are discovered and executed in compliance with a set of service consumer constraints. We exploit two well-integrated technologies which follow LD-principles for services and API integration: iServe¹⁵ & SmartLink¹⁶ [4] are two public LD-based environments dealing with two different kinds of service annotations separately, namely functional and non-functional

¹⁵ <http://iserve.kmi.open.ac.uk/>

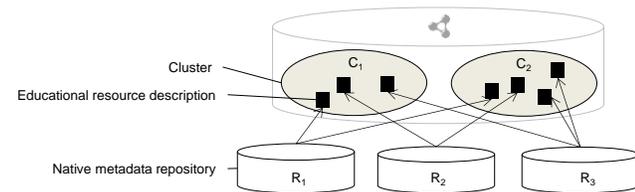
¹⁶ <http://smartlink.open.ac.uk>

We have developed enrichment mechanisms which automatically enrich poorly structured descriptions with links to related terms in well-established vocabularies. While enrichment of unstructured data poses several crucial research challenges such as entity recognition and text mining, we take advantage of available and established APIs such as the ones provided by *DBpedia Spotlight*²⁶ and *Biportal*²⁷, which partially tackle some of the related issues and allow access to a vast number of established vocabularies, such as *DBpedia*, *SNOMED*²⁸, *MESH*²⁹ or *Galen*³⁰. Enrichments allow not only further reasoning on related concepts but also enable users to query for resources by using well-defined concepts and terms as opposed to ambiguous free text.

Figure 3 depicts an example RDF resource description before and after enrichment. Please note also the enrichment with multilingual labels, as enabled by comprehensive datasets such as *DBpedia*, which contributes to resolving metadata internationalisation issues. Enrichment is implemented as automated mechanism whenever new data is pushed to the RDF store and also as semi-automated approach where users are provided with suggestions of related terms from which they can select suitable ones as part of the *MetaMorphosis+* application (see Section 6).

5.2 Interlinking educational resources across repositories

Exploratory search [11] is a useful means to navigate through large data collections [10] and is characterised as being open-ended. It involves learning and refocusing the information need while the user makes sense of the information retrieved after sequences of tentative queries that are progressively adjusted. A way to support the user in exploratory search is based on the use of a *multiple unsupervised clustering technique* [6] that uses as a basis selected subsets of the *mEducator* metadata fields (Section 4). Whereas the above mentioned terminology enrichment process tends to increase the *precision* of retrieval, and thus is key in supporting focused search, the use of clustering based on some notion of *similarity* plays an orthogonal role: it supports exploratory search by increasing the *recall* of retrieval, by suggesting possibly relevant items by utilizing clustering techniques. [6] also derives "second order" relationships among items from the analysis of the defining features of each cluster: when commonalities across clusters are found above a set threshold, association rules among the defining features can be drawn; these are used to further expand the recall, but based on new, different "facets".



²⁶ <http://dbpedia.org/spotlight>

²⁷ http://www.bioontology.org/wiki/index.php/BioPortal_REST_services

²⁸ <http://www.ihtsdo.org/snomed-ct/>

²⁹ <http://www.nlm.nih.gov/mesh/>

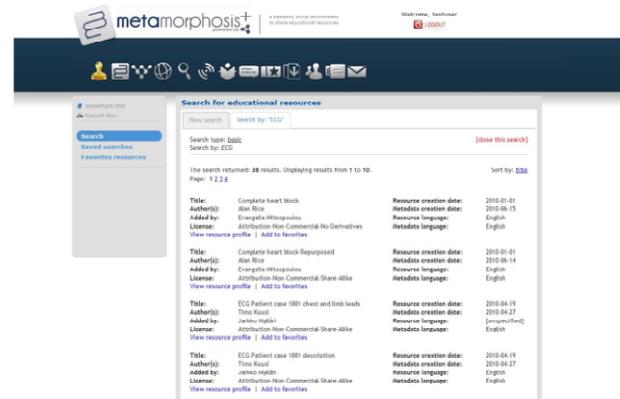
³⁰ <http://www.co-ode.org/galen/>

Figure 4. Clustering as a means to interlink educational resources across independent repositories

An interesting implication of deploying this method on the *mEducator Linked Education Resources dataset* is that it generates clusters of items which belong originally to distinct repositories (see Figure 4). In this respect, we are currently investigating appropriate metrics to analyze the provenance of the items in the clusters to derive meaningful associations at the repository level, based, for example, on subject coverage similarity, or on content type similarity. The clustering functionalities have been integrated in the RDF store described in Section 5 to allow the interlinking of resources originating from different repositories.

6. METAMORPHOSIS+: EXPLOITING LINKED EDUCATION DATASETS

The data and services integration APIs and datasets presented in the previous sections are fully integrated in *MetaMorphosis+*³¹, which merges the paradigms of semantic and social web to produce an environment for sharing linked educational resources. *MetaMorphosis+* (or *MM+*) realizes the Educational application and presentation layer (see Figure 1). To this end, *MM+* interacts with the APIs provided by the Educational services and data



integration layer (Figure 1) instead of directly retrieving and processing data from disparate sources.

Figure 5. Resource search results in MetaMorphosis+

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. At a different level, educational resources themselves create an equivalent social network. 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)

³¹ <http://metamorphosis.med.duth.gr/>

semantic connections realized via semantic annotations and linking of educational resources. The latter social dimension is realised via semantic data and service linking as described in the previous sections.

The application layer allows viewing, management and annotation of the educational resource metadata retrieved via the APIs provided by the *Educational services and data integration* layer (Figure 1). MM+ fully exploits the search and retrieval APIs searching for relevant resources two distinct mechanisms: (a) either search in the educational resources RDF store; or (b) on the distributed learning repositories semantically linked via the data and services integration layer.

7. CONCLUSION AND FUTURE WORK

Integrating educational Web resources becomes increasingly important as plenty of metadata is published openly online. We have proposed an approach which exploits LD principles to support Web-scale interoperability between educational resources, that is, educational services as well as data. LD is adapted to describe both, services and data, allowing the integration of existing educational repositories at the service and the data level. We leverage on the wealth of existing datasets and vocabularies and allow interlinking between educational data and resources. We have introduced a set of implemented integration approaches to facilitate our vision of Linked Education, resulting RDF datasets and APIs and an application (MetaMorphosis+) which makes use of these datasets and APIs to provide an open environment for (biomedical) education. The long-term goal is to establish an unified entry point³² to well-interlinked educational datasets on the Web.

While the presented work already tackles a number of distinct TEL challenges such as metadata interoperability, services discovery or data mediation, other issues will be addressed as part of ongoing and future work. Most importantly these cover: (1) investigating additional ways to enable efficient, accurate and dynamic enrichment of educational data; (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 “Educational data & services integration layer”, with additional third party applications to further evaluate the performance and scalability of our approach.

Acknowledgements

This work is funded in part by the mEducator project (Contract Nr: ECP 2008 EDU 418006 mEducator) under the eContentPlus programme of the European Commission.

8. 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] 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
- [5] 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.
- [6] Giordano, D. Faro, A., Maiorana, F., Pino, C, Spampinato, C. (2009) Feeding back learning resources repurposing patterns into the “information loop”: opportunities and challenges. 9th Int. Conf. on Information Technology and Applications in Biomedicine, 2009. ITAB 2009. IEEE
- [7] Henze, N. (2006). Personalized E-Learning in the Semantic Web. Extended version of 4. *International Journal of Emerging Technologies in Learning (IJET)*, 1(1).
- [8] 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.
- [9] Kaldoudi E, Dovrolis N, Giordano D, Dietze S., Educational Resources as Social Objects in Semantic Social Networks, Proceedings of the Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age, Herakleio, May 2011
- [10] Kobilarov, G. and Dickinson, I. (2008). Humboldt: Exploring Linked Data, Proc. Linked Data on the Web (LDW'08), Beijing, China, April 2008
- [11] Marchionini, G. (2006) Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41-46,
- [12] Mitsopoulou, E., Taibi, D., Giordano, D., Dietze, S., Yu, H. Q., Bamidis, P., Bratsas, C., Woodham, L., Connecting Medical Educational Resources to the Linked Data Cloud: the mEducator RDF Schema, Store and API, Linked Learning 2011: the 1st International Workshop on eLearning Approaches for the Linked Data Age, CEUR-Vol 717, 2011.
- [13] 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>
- [14] 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.
- [15] Sheth, A. P., Gomadam, K., and Ranabahu, A. (2008). Semantics enhanced services: Meteor-s, SAWSDL and SA-REST. *IEEE Data Eng. Bul l.*, 31(3):8–12.

³² <http://data.linkededucation.org> aims at providing a unified entry point to educational Linked Data. Individual data sets will be interlinked based on the methods proposed in this paper.

- [16] Schmidt, A., Winterhalter, C. User Context Aware Delivery of E-Learning Material: Approach and Architecture, *Journal of Universal Computer Science (JUCS)*. 10(1), 2004.
- [17] Simon, B., Dolog., P., Miklós, Z., Olmedilla, D. and Sintek, M. Conceptualising Smart Spaces for Learning. *Journal of Interactive Media in Education*. 2004(9), 2004.
- [18] 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*, 2005.
- [19] World Wide Web Consortium, W3C Recommendation, SPARQL query language for RDF, 2008, (<http://www.w3.org/TR/rdf-sparql-query/>).
- [20] 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, in *Proceedings of Linked Learning 2011: the 1st International Workshop on eLearning Approaches for the Linked Data Age*, CEUR-Vol 717, 2011.