Linked education: interlinking educational resources and the web of data

Conference Item

How to cite:

Dietze, Stefan; Yu, Hong Qing; Giordano, Daniela ; Kaldoudi, Eleni ; Dovrolis, Nikolas and Taibi, Davide (2012). Linked education: interlinking educational resources and the web of data. In: The 27th ACM Symposium On Applied Computing (SAC-2012), Special Track on Semantic Web and Applications, 25 - 29 March 2012, Trento, Italy.

For guidance on citations see FAQs

© 2012 ACM
Version: Accepted Manuscript
Link(s) to article on publisher’s website:
http://www.acm.org/conferences/sac/sac2012/

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data reuse policy on reuse of materials please consult the policies page.
Linked Education: interlinking educational Resources and the Web of Data

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

Hong Qing Yu  
Knowledge Media Institute, The Open University, Milton Keynes, MK76AA, UK  
h.q.yu@open.ac.uk

Daniela Giordano  
University of Catania, Dipartimento Informatica, Catania, Italy  
dgiordan@dit.unict.it

Eleni Kaldoudi, Nikolas Dovrolis  
Democritus University of Thrace, Greece  
kaldoudi@med.duth.gr, dovroli@alex.duth.gr

Davide Taibi  
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


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, 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 Core\(^1\), IEEE Learning Object Metadata (LOM) [7] or ADL SCORM\(^2\) and query interface mechanisms such as OAI-PMH\(^3\) or SQI\(^4\) 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 [13].

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 [18], 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 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 SQI.

---

In this paper we introduce a general approach for a
challenges by following the below principles:
the EC-funded mEducator
SPARQL endpoints. In addition, current metadata sto res largely
Dublin Core) while other repositories offer JSON-ba sed feeds or
Education
landscape of Web APIs of individual repositories. F or instance,
Web-scale search of educational resources faces a h eterogeneous
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.

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.

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\(^3\) environment which has been implemented as part of
the EC-funded mEducator\(^6\) 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, the PubMed\(^7\) 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

\(\text{LD-principles are applied to model and expose metadata of both educational resources and services/APIs. In this way, not only resources are interlinked but also services and resources are exposed in a standardized and accessible way.}\)

\(\text{Heterogeneous learning repositories, i.e. their interfaces (services) are integrated on the fly by reasoning and processing of LD-based service semantics (see 1).}\)

\(\text{Metadata retrieved from heterogeneous Web repositories is automatically lifted into RDF and exposed as LD accessible via de-referencable URIs.}\)

\(\text{Automated enrichment and clustering mechanisms are exploited in order to interlink data produced by (3) with existing datasets as part of the LD cloud.}\)

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, the PubMed\(^7\) 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

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

3. OVERALL APPROACH
Our general-purpose approach aims at (i) integrating heterogeneous educational Web resources and (ii) exposing its

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

3. OVERALL APPROACH
Our general-purpose approach aims at (i) integrating heterogeneous educational Web resources and (ii) exposing its

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

\(\text{Related work in Section 2 while Section 3 illustrates our overall approach. In Section 4 we introduce an implementation dealing with integration of educational services while in Section 5 we focus on integration of educational data. Section 6 illustrates a proof-of-concept prototype application which makes use of our data and services integration approach while we discuss and conclude the paper with Section 7.}\)
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](image)

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 PubMed\(^{13}\) or OpenLearn\(^{14}\).

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)

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. RESTful 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:

1. **Educational Services Integration**

   Our implementation of the educational data and service integration layer builds on existing research \(^{19}\) 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: iServe\(^{15}\) & SmartLink\(^{16}\) \(^{3}\) are two public LD-based environments dealing with two different kinds of service annotations separately, namely functional and non-functional service annotations stored in dedicated RDF stores. In addition, SmartLink provides a Web-based interface\(^{17}\) which allows annotation of services and browsing of existing descriptions within the SmartLink\(^{18}\) and iServe repositories via a unified user interface. The proposed services integration approach consists of the following steps:

   1. Editing and publishing RDF service annotations through the SmartLink Web interface. Service consumers can use browsing and search facilities to navigate through available service descriptions.

---

\(^{13}\) http://www.pubmed.gov

\(^{14}\) http://www.open.ac.uk/openlearn

\(^{15}\) http://iserve.kmi.open.ac.uk

\(^{16}\) http://smartlink.open.ac.uk

\(^{17}\) http://smartlink.open.ac.uk/smartlink

\(^{18}\) http://thedatahub.org/dataset/smartlink
2. Services discovery: a set of RESTful APIs has been developed to let third-party applications discover and invoke suitable educational services.

3. Services invocation and lifting: identified services are invoked and heterogeneous service responses are lifted into a coherent RDF schema. Since we apply our approach to educational data harvesting and retrieval services only, all retrieved data is lifted into a shared schema for educational resources.

Being a LD-compliant environment, one of the core features of SmartLink is the ability to associate service descriptions with so-called model references that refer to RDF descriptions in external vocabularies to further define the semantics of a service. SmartLink uses established LD APIs – currently the WATSON\textsuperscript{19} API - to identify and recommend suitable model references to the user.

A dedicated REST-ful API was developed to allow third party applications to interact with our RDF service annotations, for instance, to discover and execute services and consolidate responses. For instance, a dedicated search method expects three parameters (category, subject and language) which serve as basis to identify suitable educational services and respectively, data stores. Each of these parameters can be described by dereferencable URIs of RDF entities, such as 

\[ \text{http://www.daml.org/2003/09/factbook/languages#English} \]

invocation method is provided that supports the lifting of service responses – whatever response message format (e.g., XML or JSON) – into RDF compliant with the mEducator educational resources RDF schema\textsuperscript{21}.

In order to provide service response messages compliant with the mEducator RDF schema, native responses from educational services (e.g., XML, JSON) are lifted into RDF compliant with the mEducator RDF schema (see Figure 2 for an example). The RDF output enables data mashups at the semantic level and allows to interlink results with other LD.

5. EDUCATIONAL DATA INTEGRATION

All educational metadata as retrieved in the services integration step (Section 4) is stored in a dedicated RDF store (Educational Resources Linked Data in Figure 1) containing the mEducator – Linked Educational Resources dataset\textsuperscript{24}. This store is implemented based on OWLIM\textsuperscript{25} and is compliant with the mEducator resources RDF schema. A dedicated REST API (see \textsuperscript{11}) is offered and each resource entity owns a unique, dereferencable URI, such as 

\[ \text{http://purl.org/meducator/resources/25a8c581-66d7-4186-9411-f907f783463e} \]

Figure 2. Lifting from XML-based OAI-PMH response from PubMed to corresponding RDF

A dedicated REST-ful API was developed to allow third party applications to interact with our RDF service annotations, for instance, to discover and execute services and consolidate responses. For instance, a dedicated search method expects three parameters (category, subject and language) which serve as basis to identify suitable educational services and respectively, data stores. Each of these parameters can be described by dereferencable URIs of RDF entities, such as 

\[ \text{http://www.daml.org/2003/09/factbook/languages#English} \]

to describe an expected language. An example vocabulary for service categories is the Service-finder ontology\textsuperscript{20}. The subject finally specifies the educational domain, e.g. Maths or Life Sciences that the underlying repository is targeting. Different vocabularies are used to define subjects, e.g. one classification vocabulary\textsuperscript{23} describes OpenLearn\textsuperscript{22} learning subject categories. Taking these parameters into account, the discovery API provides information about suitable services (RDF descriptions about service, subjects, endpoints and output languages). A particular invocation method is provided that supports the lifting of service responses – whatever response message format (e.g., XML or

\[ \text{http://dbpedia.org/spotlight} \]

\[ \text{http://www.open.ac.uk/openlearn} \]

\[ \text{http://purl.org/meducator/ns} \]

\[ \text{http://thedatahub.org/dataset/meducator} \]

\[ \text{http://www.ontotext.com/owlim/} \]

\[ \text{http://dbpedia.org/spotlight} \]

\[ \text{http://www.bioontology.org/wiki/index.php/BioPortal_REST_services} \]

\[ \text{http://www.ihtsdo.org/snomed-ct/} \]

\[ \text{http://purl.org/meducator/resources/25a8c581-66d7-4186-9411-f907f783463e} \]
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 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 [10] is a useful means to navigate through large data collections [9] 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 [5] 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. [5] 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".

![Figure 4. Clustering as a means to interlink educational resources across independent repositories](image)

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. MetaMorhphosis+ (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+](image)

MetaMorhphosis+ 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 MetaMorhphosis+ 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. 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

---

29 http://www.nlm.nih.gov/mesh/
30 http://www.co-ode.org/galen/
31 http://metamorphosis.med.duth.gr/
have proposed an approach which exploits LD principles to support Web-scale interoperability between educational resources. 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 point\textsuperscript{32} 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

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.