Technical evaluation of the mEducator 3.0 linked data-based environment for sharing medical educational resources

Published version deposited in CURVE July 2013

Original citation & hyperlink:
http://ceur-ws.org/Vol-840/
http://linkedlearning2012.wordpress.com/

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

CURVE is the Institutional Repository for Coventry University
http://curve.coventry.ac.uk/open
Technical Evaluation of The mEducator 3.0 Linked Data-based Environment for Sharing Medical Educational Resources

Maurice Hendrix, Aristidis Protopsaltis, Ian Dunwell, Sara de Freitas, Panagiotis Petridis, Sylvester Arnab
Serious Games Institute, Coventry University
Coventry University Techno Park
Cheetah Road, Coventry CV1 2TL, UK
+44 (0)24 7615 8253
{MHendrix, AProtopsaltis, IDunwell, SFreitas, PPetridis, SArnab}@cad.coventry.ac.uk
Nikolas Dovrolis, Eleni Kaldoudi
School of Medicine, Democritus University of Thrace, Greece
+30 2551030329
{ndovrolis, kaldoudi}@alex.duth.gr

Davide Taibi
Institute for Educational Technologies
National Research Council of Italy
Via Ugo La Malfa 153
90146 Palermo, Italy
+39 (0)91 6809-216
davide.taibi@itd.cnr.it

Evangelia Mitsopoulou
St George’s University of London
Center for Medical and Healthcare Education
Cranmer Terrace, SW17 0RE London, United Kingdom
+ 44 (0)20 8725 1538
emitsopo@sgul.ac.uk

Stefan Dietze
L3S Research Center
Appelstr. 9a
30167 Hannover, Germany
+49 (0)511 762-17713
dietze@l3s.de

Dimitris Spachos, Panagiotis Bamidis,
Lab of Medical Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
+30 2310 999310
dimitris.spachos@gmail.com, bamidis@med.auth.gr

ABSTRACT
mEducator 3.0 is a content sharing approach for medical education, based on Linked Data principles. Through standardization, it enables sharing and discovery of medical information. Overall the mEducator project seeks to address the following two different approaches, mEducator 2.0, based on web 2.0 and ad-hoc Application Programmers Interfaces (APIs), and mEducator 3.0, which builds upon a collection of Semantic Web Services that federate existing sources of medical and Technology Enhanced Learning (TEL) data. The semantic mEducator 3.0 approach has a number of different instantiations, allowing flexibility and choice. At present these comprise of a standalone social web-based instantiation (MetaMorphosis+) and instantiations integrated with Drupal, Moodle and OpenLabyrinth systems. This paper presents the evaluation results of the mEducator 3.0 Linked Data based environment for sharing medical educational resources and focuses on metadata enrichment, conformance to the requirements and technical performance (of the MetaMorphosis+ and Drupal instantiations).

Categories and Subject Descriptors
C.4 [Performance of Systems]: Performance attributes.

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.

Linked Learning 2012, 17 April 2012, Lyon, France.

General Terms
Performance, Experimentation, Standardization, Verification.

Keywords
Performance, linked data, medical education, metadata, sharing.

1. MEDUCATOR 3.0

A significant volume of medical content supporting or delivering medical education is available within academic institutions; similarly, there is also an abundance of medical information on the web. However, without support for standards in content structure, semantics, repurposing, and reuse, this information cannot be readily linked to e-learning systems. As a European-funded Best Practice Network (BPN), the mEducator project has a dual aim at hand, first to examine existing standards and solutions and second to review how medical data can be more easily embedded, shared and reused into e-learning systems such as Learning Content Management Systems (LCMS), Content Management Systems (CMS) and repositories. The lack of universally used standardised content sharing mechanisms, identified by Bamidis at al. [1], can be singled out as one of the main challenges, which need to be addressed by mEducator. Two pilot solutions have been implemented within mEducator to evaluate two different approaches overcoming this challenge. The different approaches explored are a web 2.0 based approach, in which the interoperability is based on creating ad-hoc APIs and a semantic web-based approach. In both approaches a common standardized format has been used for describing the shared medical resources. This format is guided and formulated by
mEducator’s metadata description scheme, which is one of the fundamental outputs of the project. The semantic web-based approach forms the basis of mEducator 3.0, which implements Open Linked Education, using a set of Semantic Web Services and the provision of a number of instantiations. A standalone web instantiation and web interfaces which contain modules/extensions for the Drupal, Moodle and OpenLabyrinth systems have been implemented. This makes mEducator 3.0 more accessible to new and current users of these systems, and overcomes the lack of standardized sharing mechanisms.

The rest of this paper briefly introduces the architecture of the mEducator 3.0 platform, section 3 evaluates the metadata enrichment process, section 4 the conformance of the mEducator 3.0 platform, section 3 evaluates the metadata description framework, and finally section 5 evaluates the technical performance of the mEducator 3.0 through a practical test.

2. BACKGROUND

2.1 Architecture

The architecture of mEducator 3.0 has been described in detail by Yu et al. in [2] and Dietze et al in [3] originating from earlier work by Dietze et al. [4]. It consists of the following three layers.

- The (Web) data and service layer is made up of available Learning Resource metadata (LRM), Web services and data sources, such as the Linked Data cloud [2].
- The data and service integration layer is based on the Linked Services [5] vision and uses the iServe [5] and SmartLink [6] repositories. Each repository has Application Programmers Interfaces (APIs), which makes it easy to programmatically manipulate the data in the repositories. Following Linked Data principles, the data is described in the Resource Description Format (RDF) [7]. The repositories store functional and non-functional annotations separately.
- The application and representation layer forms the interface with which end-users interact. Different web interfaces/instantiations have been developed as implementation of the application and representation layer. Namely, a standalone web interface/instantiation (MetaMorphosis+), and three web interfaces which consist of extensions of the Drupal, Moodle and OpenLabyrinth environments\(^1\). All these use the APIs of the data and service layers, showing the ease with which the services can be integrated into existing systems.

2.2 Metadata Schema

In order to be able to share, exchange, and search medical educational resources across different institutions, there is a need for a standardized format. The mEducator metadata schema introduced by Mitsopoulou et al. in [8] provides such a format. Whilst the standardization process of this schema is an ongoing process, it is used as a basis for the mEducator project and, hence mEducator 3.0 is constructed around it. The metadata schema can be in various forms, such as text, images but also Serious Games [11]. In order to be able to implement the vision of Open Linked Education, the metadata scheme needs to be compliant with the principles of Linked Data [5]; therefore, the metadata are described in accordance with the Resource Description Framework (RDF) [7], using the RDF/XML serialization instead of the standard XML binding of IEEE LOM. The serialization as RDF builds on the RDF serialization of IEEE LOM proposed by Nilsson [12], and the Dublin Core standard [13][14]. The standardized metadata, following the Linked Data principles can be used in a wide variety of different systems. E.g. Serious Games can pull in relevant learning materials [15].

The schema is explained in more detail in [8]. In the schema each resource has a number of properties, most of which are optional. Only the following fields are mandatory, but authors are encouraged to provide as much information as possible:

- The title of the resource.
- The unique identifier (i.e. URL, URN, OkkamID etc.).
- The IPR license for (re)use.
- The language of its content.
- The language of its metadata.
- A short description about its content.
- Some keywords related to it.
- Its metadata creator.
- The date of the creation of the metadata.

The multiplicity of properties varies. Each resource has only one title, identifier, IPR license, creator etc, but it can for example have multiple keywords, educational outcomes, etc. Most properties, such as description, technical description, take textual values in different languages. Other properties make use of controlled vocabularies either defined within mEducator i.e. repurposing type, educational outcomes, or of medical taxonomies, controlled vocabularies (e.g. MESH [19], SNOMED [20]). The properties can be grouped in categories such as educational, repurposing or general, according to the nature of their content.

2.3 Metadata Enrichment

Currently many educational resources exist in a range of different repositories distributed via the web. These are unaligned and often incomplete or process limited structure [2], or example unstructured plain text is still widely used. However these resources can be transformed into a more structured and aligned form, making them machine processable. For example in the case of mEducator 3.0, the word “thrombolysis” contained in the mdc:title property can be enriched with the unique URI from DBpedia: \(http://dbpedia.org/resource/Thrombolysis\) which refers to the corresponding DBpedia resource. Or to \(http://www.co-ode.org/ontologies/galen#Thrombolysis\) which references to a matching concept within the GALEN ontology. This process is called enrichment and in the case of mEducator 3.0 this is achieved by linking to existing vocabularies and resources with known schemas and taxonomies [21].

---

\(^1\) http://www.meducator3.net/metamorphosis; http://www.meducator3.net/drupal; http://www.meducator3.net/moodle; http://www.meducator3.net/openlabyrinth
3. METADATA ENRICHMENT EVALUATION

One of the most important considerations when forming a Linked Data environment is the metadata enrichment process as described in section 2.3. Enrichment is achieved by linking data to existing repositories in known formats and achieved machine processability. This section evaluates the metadata enrichment process under mEducator 3.0 as a measure of how successful the Linked Data principles have been applied within the mEducator 3.0 environment. The metadata enrichment procedure in mEducator 3.0 is predominantly undertaken using the DBPedia Spotlight\(^2\) tool. DBPedia is itself an example of data enrichment, using as its source the publically-editable Wikipedia, adding semantic structure. DBpedia Spotlight is a tool for annotating references of DBpedia resources in text, providing a solution for linking unstructured information sources to the Linked Open Data cloud through DBpedia. In mEducator 3.0, the free text contained in several metadata properties (description, title and educational outcomes), has been enriched with concepts expressed in DBpedia using DBpedia Spotlight. Such enrichment allows 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.

Another consideration when assessing the efficacy of the enrichment process in mEducator 3.0 is the number of resources enriched and the number of links with DBpedia concepts. The mEducator RDF triple store contains a total of 375 distinct educational resources, of which 297 (79.2\%) have been enriched through linking to DBpedia. The number of individual enrichments in the data store is 1352, involving a total of 508 distinct terms from DBpedia. The mean number of enrichments per enriched resource is 4.5 (min=1, max=42). Error! Reference source not found. lists the number of enrichments obtained for each property used to trigger the enrichment of the resources. The most used properties for enrichment are mdc:description (54\%), mdc:title (24\%) and mdc:educationalOutcomes.

4. FUNCTIONAL EVALUATION

This section evaluates mEducator 3.0 against the functional requirements specified by the mEducator project. The refined requirements focus on the differences between the two solutions in mEducator, hence focusing mainly on a developer-centric description of functionalities: i.e. sharing and searching medical educational resources across a multiplicity of distributed and heterogenic repositories. The evaluation uses a grid-based approach, which shows to what extent each of the specified requirements are met. Evaluation is performed by accessing the different instantiations and checking the mentioned functionalities are present.

4.1 Requirements overview, classification

A full account of user requirements for a content sharing platform for medical education can be found in mEducator technical report D1.2\(^3\), further refined in D5.1\(^1\). A short summary of the identified requirements is listed below.

1. Publishing content on mEducator, primarily:
   a. Giving the URI of the resource (i.e. the content item)
   b. Describing the content item by predefined metadata

2. Repurposed content
   a. Extra information content re-used
   b. Purpose. The differences from the original.
   c. Author and new IPR.

3. Content organization

   Content should be categorized according to medical taxonomies for the specific medical domain.

4. Content search

   General search for related content across all metadata and browsing by type, prior usage, rating, taxonomies used.

5. Feedback on content

   The environment should support comments/reviews & display this information to potential users. Content usage should be reported and potential users should be notified.

6. Content quality issues

   An authoritative quality stamp or content reviews by known professionals in the field should be possible.

7. Content IPR issues

   Content providers can choose and apply predefined Intellectual Property Rights (IPR) licenses to their content.

8. Language issues (multilingualism)

   The instantiations and communication facilities, content, and metadata are made available in multiple languages.

9. Content sharing model

   Measurement and reporting of user activity for rewarding participation-based credits.

10. Social networking and other web 2.0 functionalities

    Social networking support for the user and content items.

11. Personalization/customization

    Users can personalize their learning experience. The resources presented are adjusted to an author/consumer profile. Content should be proposed based on the current content and past activities of the user. Users can create, manage and share ones’ own collections and searches.

12. Links to prerequisites for content usage (optional)

    Prerequisites such as installed software (players, readers) can be indicated.

13. User support

    Support to help users use the system, though a FAQ section and interaction with human users and

14. Extending the environment

---

2 http://dbpedia.org/spotlight
3 See http://www.meducator.net/?q=content/deliverables
a. Adding tools and services relevant to educational authoring,
b. Using and repurposing should be possible.

d. mEducator content repositories (optional)
   
   A content repository is available in order to host content
   from providers who do not have their content on the web.

16. User engagement and administration

   A form of user management, registration and verification, in
   line with data protection rules and regulations should exist.

Functional requirements were divided in core data and service-
related (the main focus of mEducator) and user functionalities.

4.2 Requirements discussion

An evaluation grid was used to check for presence of the required
functionalities. Most of the user requirements demonstrate at least
a partial implementation. However while the metadata schema has
fields to support multilingualism, currently the implemented
solution does not support it. Personalisation and customisation
based on a user’s profile is also not implemented and content
recommendations are not made. With regards to user support,
elements of the supporting documentation still require
construction. Revisiting the requirements specified in the previous
section, clarity in their definition and scope limited some aspects
of the evaluation. This highlights an issue that is not specific to
this project and requires considerable attention when defining any
solution in the field, as requirements and the technologies evolve.

Twelve out of 21 requirements (1, 2, 3, 4, 6, 7, 9, 10, 12, 13, 14,
and 15 from section 4.1) have been fully implemented while two
have only been partially met and the rest either have not been
implemented yet or they have been implemented with limitations.
Repurposed content and content IPR issues (D1.2-4.7) have been
implemented, the system allows selection of a predefined set of
license, most notably the creative commons license. However it is
not clear which license should be used if original and repurposed
content licenses conflict. With regards to content search, the
solution offers different types of search with different titles in the
different instantiations and streamlining the number of searches
and their naming should be considered. The solution currently
does not offer any features for collaborative editing, although
requirements for the support of Social Networking and other Web
2.0 functionalities have been met. Extending the Environment is
relatively straightforward for somebody who possesses the
required level of technical skills. Finally the instantiations do not
implement different roles with individual privileges fully, in the
case of the modules this is integrated into existing user
administration of the system.

5. PERFORMANCE EVALUATION

Non-functional requirements are an important aspect of every
system [22]. However often these requirements are never actually
formally captured, but are instead assumed. This is also the case
for mEducator 3.0. Performance is one of the main non-functional
requirements of any system. In a web-based system this mainly
means that users will get a (correct) response from the system
within an acceptable timeframe. Scalability is also very important
as it indicates how well the system copes with increasing numbers
of users. Performance is evaluated using automated software tests
using the open source JMeter\(^4\) performance measurement tool.
Scalability is evaluated using the performance evaluation data.

Testing the performance of a Web-based service or tool involves
assessing the speed that a request can be processed and serviced.
This requirement can be determined by measurement of
throughput and latency. Both can be measured by using the
timestamps at the request time and response times. At the time of
testing almost all content consisted of text and images, and did not
require especially high throughput rates, so the most important
metric for real world performances is latency, or response times.

The test was performed using the MetaMorphosis+ and Drupal
instantiations installed in different locations on different
hardware. Our test plan in JMeter, simulated accessing all major
parts of the system including login and content creation remotely
from Coventry with 1, 10, 20, 30, 40, and 50 simultaneous users.

\(^4\) http://jakarta.apache.org/jmeter/
Figure 1 demonstrates how the average response times for different pages scales up, for the MetaMorphosis + instantiation, there are 2 different groups, A and B, of which individual members roughly mirror each other in terms of scalability. We group the pages according to scalability group, shown in Table 1.

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Educational Resource</td>
<td>Friends files</td>
</tr>
<tr>
<td>Basic search</td>
<td>Friends bookmarks</td>
</tr>
<tr>
<td>Advanced search</td>
<td>My files</td>
</tr>
<tr>
<td>Start page</td>
<td>Scientific publications</td>
</tr>
<tr>
<td>Login</td>
<td>Exploratory search page</td>
</tr>
<tr>
<td>Delete added resource</td>
<td>Select item to connect to</td>
</tr>
<tr>
<td>Simple search</td>
<td>All files</td>
</tr>
<tr>
<td>Edit new resource</td>
<td>All Bookmarks</td>
</tr>
<tr>
<td>Write a blog post page</td>
<td>Find item</td>
</tr>
<tr>
<td>Add file</td>
<td>Collections</td>
</tr>
<tr>
<td>Your groups</td>
<td>Groups</td>
</tr>
<tr>
<td>New Group</td>
<td>Bookmarks inbox</td>
</tr>
<tr>
<td>Dashboard</td>
<td>My blogs</td>
</tr>
<tr>
<td>New Collection</td>
<td>Groups you own</td>
</tr>
<tr>
<td>My messages</td>
<td>Resource page</td>
</tr>
<tr>
<td>Sent messages</td>
<td>All site blogs</td>
</tr>
<tr>
<td>Send a message</td>
<td>Bookmarked items</td>
</tr>
<tr>
<td>Add content page</td>
<td>Blogs page</td>
</tr>
<tr>
<td>View graphical representation of Educational Resources</td>
<td>My profile</td>
</tr>
<tr>
<td>Select created item</td>
<td>Popular groups</td>
</tr>
<tr>
<td>Define as parent</td>
<td></td>
</tr>
<tr>
<td>Remove as parent</td>
<td></td>
</tr>
<tr>
<td>Search page</td>
<td></td>
</tr>
<tr>
<td>Perform search</td>
<td></td>
</tr>
<tr>
<td>Friends blogs</td>
<td></td>
</tr>
<tr>
<td>Bookmarks bookmarklet</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Server configurations.

<table>
<thead>
<tr>
<th>Drupal instantiation</th>
<th>MetaMorphosis+ instantiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS: Ubuntu Linux 10.10</td>
<td>OS: virtual windows2008r2</td>
</tr>
<tr>
<td>Server 64-bit (VM)</td>
<td>Processor: Quad Core</td>
</tr>
<tr>
<td>Processor: 1 X 2.343 GHz</td>
<td>Memory: 4GB RAM</td>
</tr>
<tr>
<td>Memory: 1884 MB RAM</td>
<td>Running a small number of</td>
</tr>
<tr>
<td>Database: MySQL 5.1.49</td>
<td>web sites/services</td>
</tr>
<tr>
<td>Server API: Apache 2.2.16</td>
<td>Connected via a 4 Gbps line</td>
</tr>
<tr>
<td>Running 56 web sites/services</td>
<td></td>
</tr>
<tr>
<td>Connected via a 1 Gbps line</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 shows a similar graph for the Drupal instantiation. In this graph, one may see that for up to 10 users, the results for the two instantiations are comparable, above 10 Drupal performs dramatically better. A good explanation for this is given by looking at the server specification (see Table 2) in terms of memory and connection speed. Interesting is that the pattern of scalability is quite different from MetaMorphosis+. With the exception of the Blogs page, add educational resource page and the start page, all pages scale in the same manner. The scalability result for the Blogs page is of orders of magnitude worse than others and needs further investigation.

6. RECOMMENDATIONS & DISCUSSION
The tests of two of the instantiations show that mEducator 3.0, meets most requirements at least to some degree.

7. CONCLUSION
This paper focused on providing a functional and performance evaluation of the mEducator 3.0 platform. mEducator 3.0 is a sharing tool for medical educational content, based on the principles of open linked data. It currently has four instantiations, to make it accessible to both novice and experienced users of learning platforms. The functional evaluation has shown that most of the sharing functionalities, which were the main focus of
mEducator 3.0, have been implemented at least to some degree. Multilingualism and personalization remain unaddressed. The MetaMorphosis+ instantiation performed better than the Drupal instantiation, but neither setup indicate readiness to handle larger (100+) user numbers. Interestingly, this also showed that the scaling of pages in the MetaMorphosis+ instantiation depends on users accessing repositories, while in the Drupal instantiation this does not seem to be the case. Assessing the difference in performance under idealized conditions would be an interesting exercise, though establishing realistic requirements for each instantiation would lead to useful practical guidelines.

8. ACKNOWLEDGMENTS
This paper has been fully supported by the mEducator project, funded by the eContentPlus programme by the European Commission (Grant ECP 2008 EDU 418006).

9. REFERENCES