MATTER: Metadata Migration and Annotation Tool of Teacher Education Resources

María Jesús Rodríguez-Triana, Guillermo Vega-Gorgojo,
Juan Ignacio Asensio-Pérez, Alejandra Martínez-Monés,
and Yannis Dimitriadis

GSIC-EMIC, University of Valladolid, Spain
{chus@gsic, guiveg@tel, juaase@tel, amartine@infor, yannis@tel}.uva.es

Abstract. Despite ongoing research on learning repositories, share and reuse of Teacher Education (TE) resources remains scarce. One of the reasons is that TE communities use their own in-house resources and repositories in a rather isolated way, thus limiting resource exchange in different contexts. Further, current learning metadata specifications such as Dublin Core (DC) and LOM do not address the description of pedagogical characteristics of resources demanded by TE practitioners. To overcome these limitations, the Share.TEC project aims to provide a federated TE metadata repository based on the Common Metadata Model (CMM) as the shared reference metadata model. Key for the success of the Share.TEC project is the migration of existing TE metadata records to the federated repository, performing the necessary format conversions to CMM. Besides, practitioners should be able to make TE-based annotations of resources. These two functionalities are covered by the Metadata Migration and Annotation Tool of Teacher Education Resources (MATTER) that is the subject of this paper. Although MATTER is purposed for the Share.TEC project, it can be potentially used to support the federation of other learning repositories.

1 Introduction

With the advent of e-learning programmes, there have been big efforts on producing digital resources, or so-called learning objects [1]. Digital resources typically refer to specific chunks of educational content that can exist and interoperate at different levels of granularity, ranging from a simple image to a whole lesson in computer science. Since producing high-quality digital contents can be a time-consuming and expensive process, sharing of resources should be eased as much as possible allowing the reusability of resources in different learning contexts [2]. Thus, there is ongoing research in metadata standards in order to enable cataloging, searching and reuse of educational resources. The Dublin Core (DC) [3] and IEEE Learning Object Metadata (LOM) [4] are two popular metadata standards that are commonly employed to annotate resources [5]. Learning object repositories such as iLumina [6] can then be used to search, browse, exchange and retrieve digital resources.

One specific domain of application of digital resources and repositories is the Teacher Education (TE) field [7]. Promoting TE is paramount to accomplish with Lisbon Strategy [8] objectives towards the building of an European knowledge society. However, sharing and reusing TE resources remains an elusive goal. One major limitation is that TE communities tend to be isolated, using their

own in-house resources and repositories, thus precluding resource exchange in other contexts. Moreover, TE practitioners have specific requirements for searching appropriate resources referred to the pedagogical characterization of digital content [9]. Despite this, current metadata learning standards do not define the necessary elements to describe this information. As a result, existing repositories do not allow TE practitioners to formulate their intended searches.

In order to overcome the aforementioned limitations, the EU-supported Share.TEC project [10] aims to propose a federated TE metadata repository dedicated to fostering the sharing of digital resources in the TE field. This system should aggregate metadata describing TE resources located in external repositories, while providing advanced search and brokerage functionalities for the retrieval of relevant digital content within the TE community. At the core of this system is the Teacher Education Ontology (TEO) [11] that has been recently released, defining a conceptualization of the TE domain. In addition, the Common Metadata Model (CMM) [12] has been defined for the annotation of resources with relevant TE metadata in compliance with TEO vocabularies.

However, a major issue of the Share.TEC system is the integration of existing thirdparty TE repositories in order to expose their resources to the TE community in an homogenous way. This integration is specially challenging since third-party TE repositories may employ disparate metadata formats such as DC, LOM or even proprietary ones. Further, available resources should include TEO-based annotations to allow practitioners to perform searches using specific TE terminologies. One feasible approach adopted by the Share.TEC project for this issue could be the translation of existing third-party metadata records to the CMM format. This way, obtained CMM metadata records could be stored in the Share.TEC system, making these resources available for the TE community. In addition, practitioners could provide TEO-based annotations in order to enable relevant searches of TE resources.

This paper presents the application that has been devised to fulfil the aforementioned functionalities: the MATTER tool. More specifically, it can be used to automatically harvest the metadata records of a repository, converting DC and LOM formats to CMM-compliant records. Besides, this application can be used interactively to include additional metadata according to the CMM and the TEO ontology, thus enriching the description of resources with pedagogical information that provides additional guidance for the search and selection of TE resources to use. The reminder of the paper is structured as follows: Section 2 introduces a general review of metadata models and repositories in the domain of Teacher Education. Section 3 is devoted to explain the design and implementation of the MATTER tool. Finally conclusions and future work are summarized in Section 4.

2 Metadata specifications and repositories for Teacher Education

This section provides an overview of metadata specifications and repositories in the elearning domain, stressing the major challenges to spread the sharing and cataloguing of educational resources among repositories. It follows a brief depiction of the TEO ontology and the CMM metadata model developed in the Share.TEC project for annotating Teacher Education resources.

2.1 Educational metadata specifications and repositories

Since producing high-quality educational resources can be a time-consuming and expensive process, it was soon recognized the necessity of sharing and reusing educational resources in different contexts. Metadata is critical to finding resources, so a number of standardization bodies have launched different metadata initiatives. In the e-learning domain, the Dublin Core Metadata Element Set (DC) [3] and IEEE Learning Object Metadata (LOM) [4] are the most popular metadata standards for resource annotation [5]. Hence LOM is specifically targeted to learning resources, while the scope of DC is broader, applying to any kind of digital resources.

The DC specification was designed to be simple, defining a set of just 15 metadata elements for describing core properties of a resource such as title, creator or date. In contrast, the LOM specification is more complex, defining a set of 76 elements grouped in nine categories referring to different metadata perspectives such as general information, technical characteristics or intellectual property rights. Both DC and LOM make use of controlled vocabularies to populate some fields, though such vocabularies are not sufficient or appropriate in many cases [13, ch. 5] [6]. As a result, organizations typically use so-called application profiles of DC or LOM defining their own vocabularies or even including new elements not covered in the standards [13, p. 142] [14].

Educational resources and their metadata are commonly stored in repositories (see some examples in [14]), allowing users to search, browse, retrieve and exchange resources. One of the main challenges related to repositories is how to provide a uniform access to different collections of resources. Indeed, there are still some problems with metadata format conversion, despite standardization efforts. For instance, there are mappings publicly available for converting metadata between LOM and DC, but data can be lost in the translation since LOM is much more extensive than DC. Further, the use of different application profiles with disparate vocabularies and metadata elements complicate the translation process [5].

In addition to differences in metadata formats, another issue for federating repositories is accessing resource metadata homogeneously. Fortunately, standardization efforts have proposed the Open Archives Initiative Protocol for Metadata Harvesting (OAIPMH) [15]. This protocol specifies an application-independent way of collecting metadata from repositories, enabling the creation of a unique set of resources coming from different collections. As a result, e-learning repositories such as ARIADNE [16] are beginning to adopt this protocol, offering OAI-PMH providers for exposing metadata records.

Repository federation also involves metadata publication in a standard way. To achieve this aim, the Simple Publishing Interface (SPI) protocol [17] has been proposed for publishing learning objects and metadata records to digital repositories. Several projects have adopted this protocol in their repositories, such as ALOE [18] or ARIADNE [16].

2.2 TEO ontology and CMM metadata model for annotating Teacher Education resources

The Teacher Education Ontology (TEO) [11] has been developed by TE researchers in the Share.TEC project in collaboration with international experts.

This ontology aims to conceptualize the relevant abstractions of the TE domain with particular emphasis on facilitating the sharing of digital resources by the TE community. Thus, TEO intends to provide an agreed and non-ambiguous vocabulary that serves as a *lingua franca* to provide meaningful descriptions of TE resources. A resource in TEO can be characterized with their pedagogical features, e.g. didactic strategy, associated to a specific knowledge area, e.g. Computer Science, and related to the competencies to be addressed, e.g. analyze. Note that these aspects were considered specially relevant for sharing resources among TE users, although cannot be described with DC or LOM.

Significantly, the Common Metadata Model (CMM) [12] has been derived from TEO in order to provide a shared reference metadata model for annotating TE resources in the Share.TEC project. A key design decision in the proposal of CMM was to adopt LOM as a basis in order to facilitate the transition of existing LOM-based repositories to the purposed CMM-based federated repository of Share.TEC. As a result, the metadata elements defined in CMM are organized in 10 categories. The first eight ones directly correspond to categories 1 to 8 of LOM, while TE specific elements (pedagogical features, knowledge areas and competencies) have been included in categories 9 and 10. Hence the allowed vocabularies for these TE elements are defined in TEO.

Additionally, it is pursued in the Share.TEC consortium to take into account linguistic and cultural differences among users. Thus, the CMM has been extended with a series of Multicultural Metadata Models (MMM) [19] corresponding to each country. The relation between the CMM and the MMM can be seen as the collection of all the adaptions/translations of the CMM in various cultures/languages. Specifically, the MMM gains its multicultural dimension by translating the elements of sections 1 to 8 and adapting the elements and vocabularies derived from TEO.

3   MATTER: Metadata Migration and Annotation Tool of Teacher Education Resources

Once the TEO ontology and theCMMmetadata model have been implemented, the next step is the development of a federated metadata repository of TE resources annotated in conformance with CMM. An overview of the federated TE repository is shown in Figure 1. Participants in the TE federation will provide access to their repositories, offering their TE resources to the consortium. The Share.TEC system aggregates the metadata records of TE resources located in participants' repositories in a cohesive collection, allowing users to seamlessly search and retrieve TE resources.

Participants in the federation must expose their metadata records using an OAIPMH provider, thus enabling metadata harvesting to feed the Share.TEC repository. However, different metadata formats such as LOM or DC may be employed in participants' repositories, so it is necessary to perform a translation to the CMM format in order to homogenize metadata records. Further, resource metadata should include TE annotations to allow users to perform searches referred to specific aspects of the TE domain. Since these two functionalities are key for the realization of the Share.TEC vision, the authors have proposed the Metadata Migration and Annotation Tool of Teacher Education Resources (MATTER). MATTER supports the migration of metadata records between re-

positories, allowing the conversion to DC, LOM and CMM metadata formats. In addition, MATTER can be used interactively to annotate resources in the aforementioned formats.
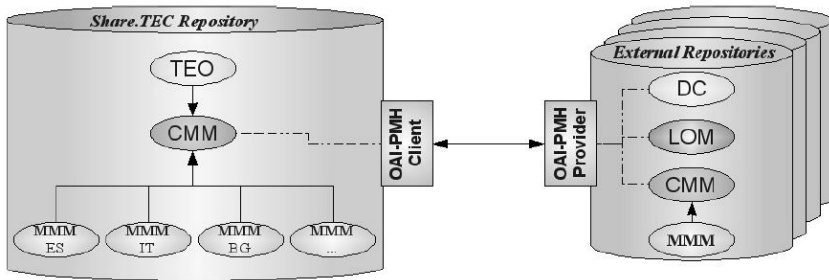

Fig. 1. Overview of the federated TE repository.

3.1 MATTER outline and logical architecture

MATTER has been devised for supporting the following three scenarios of usage. The first one refers to the migration of metadata records from a source repository to a target repository, performing the necessary format conversions. In the second scenario a TE practitioner uses MATTER for enriching an existing metadata record with additional annotations, possibly specific of the TE domain. In the last scenario, a TE practitioner employs MATTER for describing a resource from scratch. Noteworthy, MATTER is purposed to support the translation between DC, LOM and CMM metadata formats. DC and LOM are commonly employed in educational repositories, while CMM has been specifically proposed for the TE domain (see section 2).

The logical architecture of MATTER consists on a set of components shown in Figure 2. The *harvester* module is in charge of retrieving metadata records from repositories. A harvested metadata record can be automatically converted to another metadata format by the *translator*, being DC to CMM and LOM to CMM translations of special interest for the Share.TEC federated repository. Note that information can be lost in a translation to a less extensive format (e.g. LOM to DC). In addition, DC and LOM do not cover TE-specific metadata defined in CMM such as didactic strategy or collaboration level, so translated records to the CMM format will not include this information. Nevertheless, the user can supply additional metadata, if desired, interacting with the *annotator*. Finally, the produced metadata record is published in the target repository through the *publisher* component.

3.2 MATTER implementation

The different components that constitute the logical architecture of MATTER have already been implemented. Specifically, the harvester module implements the client side of the OAI-PMH protocol [15], enabling the collection of metadata records from repositories exposing an OAI-PMH provider. Since this protocol supports multiple metadata formats, the developed harvester can retrieve DC, LOM and CMM formatted records.
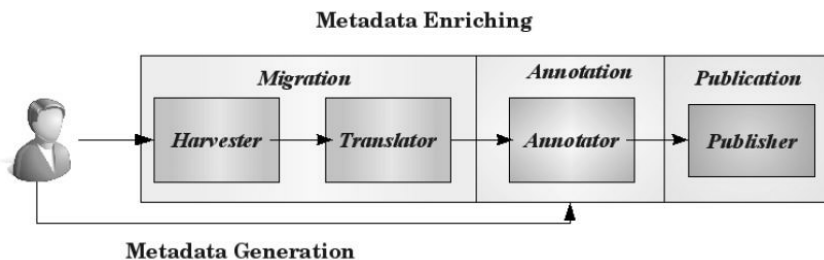
Metadata Enriching

Metadata Generation

Fig. 2. Logical architecture of MATTER.

In order to implement the translator component, different choices were considered. Since both input and output metadata records have an XML syntax, the popular Extensible Stylesheet Language Transformations (XSLT) [20] could be employed to perform the translation process. However, XSLT-based transformations generally lead to brittle implementations since XSLT code is difficult to extend or readapt; for instance, experience with XSLT for converting formats in the CAViCoLA project [21] [22] evidenced a lack of flexibility. Moreover, translations to the CMM format have to be aligned with the vocabularies defined in the TEO ontology, requiring a high degree of customization. Thus, it was decided to follow a programmatic approach for the translation process, making use of the Java Architecture for XML Binding (JAXB) API [23].

JAXB allows automatic two-way mapping between XML documents and Java objects. Given a schema definition (e.g. LOM schema), the JAXB compiler can generate a set of Java classes that allow developers to build applications that can read, manipulate and recreate XML documents. Generated Java classes can be highly customized to perform the required translations, such as making use of external vocabularies. Therefore, the implementation of the translator component of MATTER is based on JAXB, allowing any transformation from DC/LOM/CMM to DC/LOM/CMM metadata formats. Special care was taken for defining the mappings among the controlled vocabularies employed by each metadata format. In this sense, the OWL API [24] was employed in order to gather the classifications defined in the TEO ontology for the CMM format. Once the required mappings were implemented as Java classes, generation of converted metadata records as XML documents is automatically handled by JAXB.

Concerning the annotator component, it was also implemented employing JAXB as a basis. In this case, a form-based user interface was developed for obtaining the information introduced by the user in order to create or enrich a metadata record. Existing vocabularies and classifications, e.g. knowledge areas, are shown to the user so as to provide guidance in this process. Finally, generated records are serialized as XML documents and forwarded to the publisher module. This last component provides a Simple Publishing Interface (SPI) source implementation [17], enabling the publication of metadata records in repositories exposing an SPI target.

A preliminary testing of the developed prototype of MATTER has already been done. Specifically, conversions of metadata records to different formats have been successfully performed. Further, teacher educators have been able to successfully generate new CMM-compliant records, as well as enriched

existing ones retrieved from the authors' repository. These resulting files have been harvested by the Share.TEC system, ensuring that the output files follow correctly the final formats.

4 Conclusions and future work

Sharing and reusing TE resources is a long-desired goal of the TE community. Enabling the federation of existing repositories is envisioned as a required step towards this objective. With this aim, the Share.TEC project aims to provide a federated TE metadata repository based in the CMM metadata model to homogenize existing TE resources in the consortium. Further, TE practitioners demand specific search capabilities referred to teacher education characteristics, including culturally dependant aspects. To fulfill this vision, conversion of existing metadata records to CMM should be tackled, as well as enabling practitioners to annotate resources with meaningful TE characteristics.

This paper has presented the MATTER tool that has been proposed to cover the aforementioned functionalities. AlthoughMATTER is intended to be used in the Share.TEC project, it can also be employed to facilitate the federation of other learning repositories. Indeed, MATTER can be used to automatically convert DC/LOM/CMM metadata formats or as a multi-format resource annotation tool, thus it could be applied in other contexts. Moreover, MATTER has been designed to be extensible, so new metadata formats such as specific application profiles could be easily supported due to the usage of the JAXB API. Since the developed prototype has not been extensively tested, the authors plan to offer MATTER to the TE community in order to further assess the suitability of this tool, and evaluate the results and experiences from the end user perspective with the aim of improving its usability.

References

1. McGreal, R.: Learning Objects: A Practical Definition. Instruction Technology and Distance Learning 1(9) (2004)
2. Wiley, D.: Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In Wiley, D., ed.: The Instructional Use of Learning Objects. Agency for Instructional Technology (2002) 3–23
3. DCMI Usage Board: DCMI Dublin Core Metadata Element Set, Version 1.1. Recommendation, DCMI (2008) URL: http://dublincore.org/documents/dces/, last visited June 2009.
4. IEEE Learning Technology Standards Committee: IEEE standard for learning object metadata. Specification 1484.12.1-2002 (2002)
5. McClelland, M.: Metadata standards for educational resources. Computer 36(11) (2003) 107–109
6. Heath, B.P., McArthur, D.J., McClelland, M.K., Vetter, R.J.: Metadata lessons from the iLumina digital library. Communications of the ACM 48(7) (2005) 68–74
7. MERLOT-Multimedia Educational Resources for Learning and Online Teaching:

Merlot Teacher Education portal (2009) http://teachereducation.merlot.org/, last visited June 2009.

8. Kok, W.: Facing the challenge: The Lisbon strategy for growth and employment. Technical report, High Level Group (2004) URL: http://ec.europa.eu/growthandjobs/ pdf/kok report en.pdf, last visited June 2009.

9. Alvino, S., Forcheri, P., Ierardi, M.G., Sarti, L.: A general and flexible model for the pedagogical description of learning objects. In: Proceedings of the 20th World Computer Congress (WCC 2008), Milan, Italy (2008) 55–62

10. Share.TEC Consortium: Share.TEC Project Website - SHAring digital REsources in the Teaching Education Community (2008) http://www.share-tec.eu/, last visited June 2009.

11. Alvino, S., Bocconi, S., Boytchev, P., Earp, J., Sarti, L.: An ontology-based approach for sharing digital resources in Teacher Education. In: Proceedings of the Seventh International Workshop on Ontologies and Semantic Web for E-Learning (SWEL), Brighton, UK (2009)

12. Open University of Netherlands (OUNL): Common Metadata Model (CMM). Technical Report Deliverable 2.2, Share.TEC (2008)

13. Devedzic, V.: Semantic Web and Education. Springer, New York, NY, USA (2006)

14. Neven, F., Duval, E.: Reusable learning objects: a survey of LOM-based repositories. In: Proceedings of the tenth ACM international conference on Multimedia (MULTIMEDIA '02), Juan-les-Pins, France (2002) 291–294

15. Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S.: Open Archives Initiative - Protocol for Metadata Harvesting. (2008) http://www.openarchives.org/OAI/ openarchivesprotocol.html, last visited June 2009.

16. ARIADNE Foundation: ARIADNE Website - Foundation for the Knowledge Pool (2006) http://www.ariadne-eu.org/, last visited June 2009.

17. Ternier, S.: Simple Publishing Interface Specification. (2008) http://ariadne. cs.kuleuven.be/lomi/images/b/ba/CEN SPI interim report.pdf, last visited June 2009.

18. Martin Memmel: ALOE - The Social Resource and Metadata Hub (2007) http:// aloeproject.de/index.html, last visited June 2009.

19. Trinity College Dublin (TCD), ed.: Multicultural Metadata Model for Interoperability. Number Deliverable 3.1 (2009)

20. World Wide Web Consortium (W3C): XSL Transformations (XSLT). Recommendation, W3C (1999) URL: http://www.w3.org/TR/xslt, last visited June 2009.

21. Martínez-Monés, A. and Marcos-García, J. A. and Rodríguez-Triana, M. J. and Fernández-San Miguel, D.: Role-AdaptIA: A role-based adaptive tool for interaction. In: Proceedings of the Workshop on Real-Time methods, Utrecht (The Netherlands), International Conference of the Learning Sciences (ICLS) (2008)

22. Marcos-García, J. A. and Martínez-Monés, A. and Dimitriadis, Y. and Rodríguez-Triana, M. J.: Proposals for adaptive and interoperable IA systems, in . In: Proceedings of the Workshop on Interaction Analysis and Visualization for Asynchronous Communication: Analysis Methods, Tools, and Research Questions (CSCL 2009), Rhodes (Greece) (2009)

23. GlassFish Community: Java Architecture for XML Binding (JAXB) (2008) https:// jaxb.dev.java.net/, last visited June 2009.

24. University of Manchester: The OWL API Website (2009) http://owlapi.sourceforge. net/, last visited June 2009.