

# Migration of E-Learning Objects from Database to IMS XML Standard

Giedrius BALBIERIS

*Computer Networking Department, Kaunas University of Technology  
Studentų 50-416, LT-3031 Kaunas, Lithuania  
e-mail: balbgied@pit.ktu.lt*

Vytautas RĖKLAITIS

*Graduate School of Information Systems  
University of Electro-Communications, Tokyo, Japan  
e-mail: vyta@is.uec.ac.jp*

**Abstract.** Since establishment of World Wide Web a number of e-learning tools and resources have been created and successfully used in every educational institution. Established standards such as IMS and SCORM currently provide means for e-learning asset portability and reuse. Most of such implementations have a database back-end. Data from such a back-end RDBMS can be exported into IMS XML and used by standard compliant e-learning platforms. After reviewing facilitating technologies and similar solutions authors state that there is no viable solution for database to IMS conversion. Next they present own DB-to-IMS XML conversion method. As a conversion example authors introduce and use own developed EduMMDB – a set of e-learning tools with database back-end. After presenting conversion tool ELSTD authors summarize gained experience and possibilities of improvement.

**Key words:** e-learning object, conversion database, IMS XML standard.

## 1. Introduction

Distance education is around for couple decades. Since the beginnings a great number of e-learning courses, training pieces as well as learning tools has been created. A great number of proprietary and own e-learning platforms had been created and deployed across education institutions. Those tools and courses were integrated into those proprietary e-learning systems and not available to anything else. These environments contain quite a lot of e-learning information which would be ready for reuse and ported to other systems only if converted to standards. Since most of such in-house developed e-learning tools have a database back-end often standardization means converting data from relational database into formats provided by learning standard. In this paper we shortly look at the standards, overview conversion technologies, most similar solutions, propose own conversion method and share experience gained from implementation and carried-out experiments.

### 1.1. *E-Learning Standards*

What e-learning standards provide, at first, is reusability and interoperability of learning resources. This is achieved via describing each thematic data chunk with standardized e-learning descriptor. When bundling up together this thematic data chunk and its' description (so called metadata) we have a learning object (LO). Provided initially we have a course as a single lump, which often is a case, then standardization is a process consisting of three tasks (Balbieris and Rėklaitis, 2002):

1. Learning object identification.
2. Data migration.
3. Metadata identification and generation.

As learning object identification and metadata generation topics are concluded in previous paper (Balbieris, 2003), this paper covers second issue – data migration. Before elaborating on own data migration method lets review what has been done in the field of learning data migration and what are the underlying technologies to achieve this task.

### 1.2. *Conversion Technologies and Solutions*

In research done following existing implementations providing XML standardization have been examined: DC-DOT (DC-DOT project, 2003), SCORPION (Subramanian and Shafer, 1998), My Meta Maker (Severiens, 2003) etc. Most of them provided metadata authoring, few provided existing HTML metadata discovery. All supported simple Dublin Core (Dublin Core project, 2003) XML dialect and only DC-DOT supported IMS (IMS project, 2003) standard specification. Nevertheless none of the tools supported learning object extraction, conversion from database. On the other hand there are few database to XML converters (Turau, 2001; XML-SPY, 2003), most of which just do a raw data dump and provide no means for matching alternatives, conditional matching etc. Whilst there is research done (Bourret, 2003a; Bourret 2003b) on advanced database to XML matching and conversion, still there is no viable e-learning data conversion solution.

DB to XML conversion, metadata identification are not new scientific problems and had been formulated in few decades since the beginning of library sciences. Technologically all means for data migration, metadata identification are in place. Still we note, that e-learning metadata discovery and data conversion is a specific process. We base this on following assumptions:

1. Variety of data source types.  
E-learning courses include all possible means to increase the learning efficiency: text (plain, rich text, hypertext, electronic documents), images and graphic illustrations, animations, sound and video fragments. As most media items are referenced by back-end database, it is possible to extend the database to XML matching by using and analysing this extra relationship information.
2. E-learning process and components specifics.  
E-learning itself is somewhat specific due to the typical phases and set of learning chunks it contains. Following good pedagogic practice suggestions (World bank,

2003) e-learning content is already structured and usually has a set of alike phases (known as Kolb cycle): course information, activities and assignments, testing etc..

### 3. Standardization technology.

E-learning standardization is possible mainly due to technological advancements in data description languages and using XML set of technologies in particular. The most flexible in expression XML has many dialects (Document Type Definition – DTD, Schema, Resource Description Framework – RDF etc). E-learning is specific as it uses DTD XML dialect and provides huge number of metadata tags for all possible learning expressions to handle. For example import of all current IMS tags in ELSTD tool produces a tree of 1545 elements. Provided there is a database schema consisting of several hundred items, the number of matching combinations becomes too high for matching and conversion by hand. Here term schema is commonly used (Rahm and Bernstein, 2001) to denote database structure as well as a XML tree structure, including each item name, type, constraints etc.

#### 1.3. Metadata Annotation and Identification Issue

The issue of metadata is the issue of next few decades. Metadata has been around for number of years already. Each word processor, spreadsheet, presentation tool has a metadata description feature. Often this is achieved via dialog 'Document properties' with metadata fields such as 'Title', 'Subject', 'Keywords' etc. Nevertheless due to extra time needed for annotating – nearly nobody uses any other metadata except file naming and location (directory) based cataloging. Standard committees provide all metadata specifications and verification tools (IMS project, 2003; SCORM project, 2003). Also there are convenient interfaces for metadata authoring such as Learn eXact Packager etc.

E-learning object and metadata identification, packaging tasks require huge workload and so far there had been no means to automate that. In fact there had been attempts for automatic metadata identification in other areas like library science etc (Subramanian and Shafer, 1998), but those particular solutions are specifically tied to subject domain and yet are not finished. The issue with learning objects is similar and even more complicated: there are many more metadata fields to use, data sources etc.

Summing up above sections we conclude that there is a huge need of dedicated tool for e-learning object and metadata identification, conversion to standards.

## 2. Proposed E-Learning Database Standardization Method

Below we propose own e-learning database to IMS XML standardization solution. Solution is based on schema matching method, described in next Section (2.1), its' implementation in ELSTD tool (Section 2.2).

### 2.1. Matching Method

The matching problem is implicitly a two-sided process: at one side there are learning chunks represented as various data formats, at another side there is a formal framework

for expressing standardized learning objects. In matching process a set of relationships between both structures are discovered and recorded as XSLT transformations. Later as expert decides that sufficient number of matchings has been achieved, generated XSLT rules are applied as unattended routine conversion of database into IMS XML standardized learning objects.

Considering the fact that formal data conveyed by database and standard schemes may be recorded in different languages, same items may be denoted with different synonyms and similar terms may have multiple occurrences, we conclude that initial information provided by database and standard schemes is not sufficient for automatic schema matching. Because of that we require to describe each matching item with a set of keywords, later on used for schema matching.

Provided that there is IMS standard descriptor tag vector and words describing each tag is in place, relationship between tags and each description word can be expressed as matrix  $S$ :

$$S_{nm} = \begin{bmatrix} s_{11} & s_{21} & s_{31} & \dots & s_{n1} \\ s_{12} & s_{22} & s_{32} & \dots & s_{n2} \\ s_{13} & s_{23} & s_{33} & \dots & s_{n3} \\ \dots & \dots & \dots & s_{ij} & \dots \\ s_{1m} & s_{2m} & s_{3m} & \dots & s_{nm} \end{bmatrix}. \quad (1)$$

Here columns  $i = 1 \dots n$  – tags from standard DTD, rows  $j = 1 \dots m$  – words describing the tags.

Also provided that there is a RDBMS database schema and words describing each of the schema items, relationship between RDBMS database schema items and each descriptor word can be expressed as matrix  $D$ :

$$D_{ps} = \begin{bmatrix} d_{11} & d_{21} & d_{31} & \dots & d_{p1} \\ d_{12} & d_{22} & d_{32} & \dots & d_{p2} \\ d_{13} & d_{23} & d_{33} & \dots & d_{p3} \\ \dots & \dots & \dots & d_{kl} & \dots \\ d_{1s} & d_{2s} & d_{3s} & \dots & d_{ps} \end{bmatrix}. \quad (2)$$

Here columns  $k = 1 \dots p$  – RDBMS database schema items, rows  $l = 1 \dots s$  – words describing each of database schema items.

Matrix  $D$  columns are RDBMS database schema items, matrix  $S$  columns – standard descriptor XML tags. Both matrices have describing words in rows. As shown in algorithm below the matching happens by matching finding nearest neighbor for column  $D_{k*}$ ,  $k = 1 \dots p$  in set of columns  $S_{i*}$ ,  $i = 1 \dots n$ :

1. /\* DB2XML - match described DB and XML schemes. \*/
2. function DB2XML
3.      $\forall$  dataDescriptor  $\in D_{k*}$
4.         sumMax = -1

```

5.      sumInd = -1
6.       $\forall$  stdDescriptor  $\in S_{i*}$ 
7.          sum = dataDescriptor  $\times$  stdDescriptorT
8.          if sum > sumMax
9.              sumMax = sum
10.             sumInd = i
11.          end if
12.      end # stdDescriptor
13.      saveMatch(k, i)
14.  end # dataDescriptor
15. end sub

```

In algorithm shown above each data source item  $k$  is matched to the closest standard tag  $i$ . The nearest distance is found based on description similarity (descriptor vector product). Note: before making product in line 7 both vectors need to be normalized into vector product required dimensions. Descriptors may contain special symbols to denote special functions such as:

- 1) implicit matching rule, first symbol “!”,
- 2) data class rule, first symbol “:”,
- 3) conditional rule, first symbol “?”,
- 4) variable definition, first symbol “\$”.

Schema matching is an interactive process. Recurrently used it settles down standard descriptor thus requiring just database schema annotation. Standardization process is semi-automatic: after correct description, transformation rules are automatically generated and unassisted data conversion takes place.

### 3. Experiment Implementation

#### 3.1. Standardization Tool ELSTD

There is a great variety of data sources in today's e-learning computing as well as unknown information formats of the future. Because of that for conversion implementation a more generic approach had to be taken. This was achieved via using plug-in technologies for each and every data source possible. Provided the flexibility and richness of expression in XML set of technologies and the fact that standards themselves are also expressed as XML dialect, it is very natural for generalization also choose an XML-native technology. Thus for ELSTD implementation XSLT (standard XML transformation language) was chosen. By selecting XSLT to express matching rules following preconditions are taken over:

1. Data source primary matching into XML subset.
2. Manifold processing as well as less efficient transformation load.

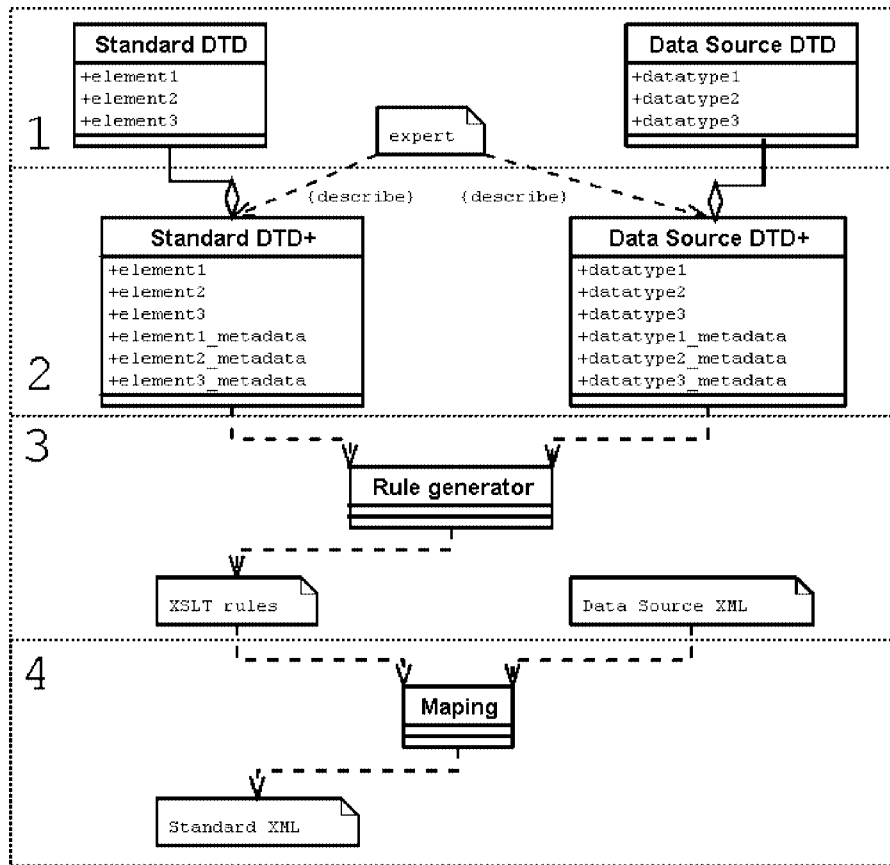


Fig. 1. Semi-automatic standardization process.

Accepting these preconditions we propose following semi-automatic standardization architecture (Fig. 1).

Semi-automatic standardization process consists of 4 stages:

1. Unification of data from various data sources.  
In this stage all data sources are converted into XML subset for later uniform processing. All documents are converted into HTML, databases exported into XML using existing database XML export engines (Turau, 2001; Bourret, 2003a).
2. Standard and data source enrichment.  
This stage is needed to cover all schemes into same coordinate term system. This includes translation of data schema names into same language, annotation of data schema elements with synonyms, marking for specific processing such as conditional matching etc.
3. Transformation rule creation.  
Based on extended data schema description best match from both schemes is achieved and XSLT transformation rules are produced. These rules are taken from

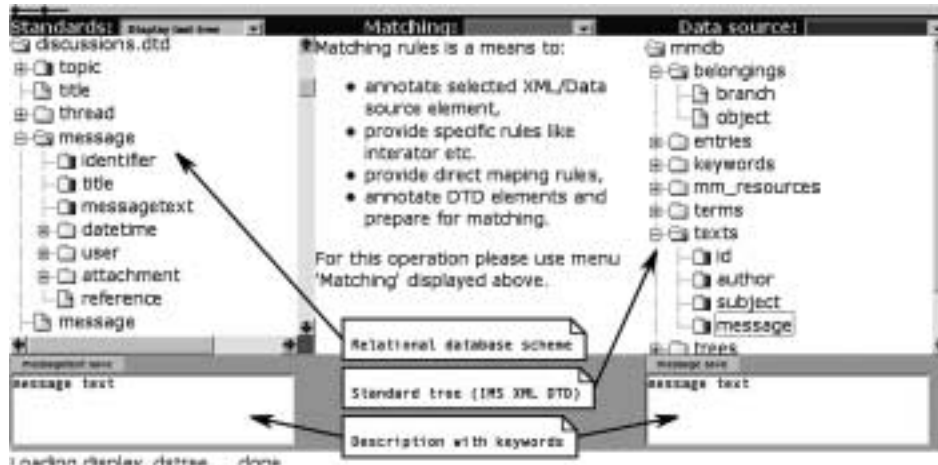


Fig. 2. Standard and data source item annotation.

predefined matching pattern library. This library as well as the matching learning chunk logic is the central place for future extensions, customization and standard substitution.

#### 4. Unified data-to-standard conversion.

Final conversion in stage four is the most computing intense part and involves application of created XSLT transformations for converting data sources and producing IMS standard based XML.

Developed in Perl ELSTD tool is aiming at easy installation and deployment on any default campus server. The tool provides a Web based migration process. As we see in the picture (Fig. 2) ELSTD provides 3 vertical columns: left and right for displaying and operation on standard and source data schema trees and middle one is for presenting matching rules.

In first step data source is provided as database access information, – in RDBMS case, or as a course zip file, – in HTML case. After analysis and conversion to internal data schemes both standard and data source are presented to the expert for assessment and enrichment (Fig. 2)<sup>1</sup>. Expert may add or exclude some attributes from both trees and add synonyms, translate words into a common language etc. After the enrichment matching process takes place and generated matching rules are saved as XSLT templates. Steps 2 and 3 are repeated until satisfactory rules production is achieved. Data conversion (Step 4) is the most labor intensive process as it involves processing huge amount of data and lots of transactions etc. After conversion resulting learning objects are accessible as a downloadable zip package and ready for integration into any IMS compliant system. This zip archive conforms to IMS Content Packaging specification (IMS project, 2003) and contains global metadata descriptor, exported learning objects with each individual metadata descriptor.

<sup>1</sup>Note: we used IMS extended version by WebCT to include definitions for extra tools such as bookmarks, library, image database, discussions, glossary of terms etc.

### 3.2. EduMMDB RDBMS Database as a Data Source

EduMMDB developed by authors since 1997 Copernicus project has undergone major improvements and now is in version 4. This is a multimedia catalog for sharing different learning resources, Internet links as well as in-place integrated structured discussion environment. It has been used in number of courses and now is facing the need for XML standardization.

EduMMDB is based entirely on SQL-92 compliant relational database backbone. It provides a framework for e-learning material library, discussions and cataloging of links. This is achieved by using 4 different types of objects:

1. Catalog entry (has one parent and few children objects),
2. Multimedia file – any file which is described by MIME types.
3. Link. External resource with description.
4. Discussion message. It is a piece of text with author and subject attributes.

Standardizing tree structure of EduMMDB items requires advanced data matching features. This is achieved by providing advanced RDBS-to-tree XSLT transformation patterns. Such extended approach illustrates possibilities of XSLT expression riches and provides basis for implementing custom migration plugins from other kinds of data sources (for example object databases).

Currently the biggest EduMMDB installation at Kaunas University of Technology contains 16 major categories, 360 catalogs with 1353 items. Second implementation installed and is being used at Graduate School of Information Systems, University of Electro-Communications, Tokyo, Japan. Both implementations need to be ready for integration into standard compliant learning platforms such as WebCT, Blackboard or similar.

EduMMDB is mostly RDBMS driven application within cells of database it references data in variety of data formats. Retrieval methods differ depending on data type used, because of that conditional matching rules had to be used.



Fig. 3. EduMMDB contains: folders, messages, links and files.



In the conversion experiment the most labor intensive work was in learning ELSTD tool and describing the schemes correctly. Usually description required not only additional attributes, but also defining classes, using conditional matching. In EduMMDB case one database field was denoting four types of learning objects (catalog, link, file, message). In such situation it is easier to write simple conditions and match each type separately thus also avoiding data conversion collisions.

#### 4. Summary and Future Work

In this paper we discussed problematics of data conversion from back-end relational databases into e-learning standards. Finding no alternative we proposed own conversion method and presented its' architecture. Expert involvement was inevitable thus it took few attempts till optimal easy to use file-manager like ELSTD migration tool was developed. EduMMDB – author developed virtual learning environment was used for data migration. Scheme matching model has proved itself in practice, but still there are more automation possibilities to include and those are related to:

1. analysis of hyper relations between different learning objects,
2. data mining not only data schemes and expert created keywords, but also the contents of learning objects and thus discovery of multilevel structures of learning objects.

Overall we see that it is possible to migrate learning objects from relational database into IMS XML. Although there are no premises for complete automation and there is a huge variety of learning objects, metadata thus making migration a very complicated task. Indeed there are already virtual learning environments which provide standard support (WebCT etc). Thus we strongly suggest to put a high stake for choosing virtual learning environment with standard support. Such choice will have no platform lock-in and thus will be less risky, integration-ready and in long term it will be less costly as well.

#### References

- AICC: Aviation Industry CBT Committee (2003). <http://www.aicc.org>
- Balbieris, G. (2003). *Using Naive Bayes Method for Learning Object Identification* (prepared for publishing). [http://www.metalot.com/publications/balbieris\\_naive-bayes.pdf](http://www.metalot.com/publications/balbieris_naive-bayes.pdf)
- Balbieris, G., and V. Reklaitis (1999). Multimedia data repository for integrated learning environment. The role of ODL in the Information Society. In *Proceedings of International Conference Wiss'99*. Technologija, Kaunas, pp. 64–70.
- Balbieris, G., and V. Reklaitis (2002). Reshaping e-learning content to meet the standards. *Informatics in Education*, 1, 5–16.
- Barritt, C., and D. Lewis (2000). *Reusable Learning Object Strategy: Definition, Creation Process, and Guidelines for Building*. Cisco Systems, Inc. [http://www.cisco.com/warp/public/10/wwtraining/elearning/implement/rlo\\_strategy\\_v3-1.pdf](http://www.cisco.com/warp/public/10/wwtraining/elearning/implement/rlo_strategy_v3-1.pdf)
- Bourret, R. (2003). *XML-DBMS Java Packages for Transferring Data between XML Documents and Relational Databases*. <http://www.rpbouret.com/xmldbms/readme.htm>
- Bourret, R. (2003a). *XML and Databases*. <http://www.rpbouret.com/xml/XMLAndDatabases.htm>

- Bourret, R. (2003b). *Data Transfer Strategies*.  
<http://www.rpbourret.com/xml/DataTransfer.htm>
- DC-DOT: Dublin Core Metadata Editor (2003). <http://www.ukoln.ac.uk/metadata/dcdot/>
- Dublin Core Project (2003). <http://dublincore.org>
- IMS project: Instructional Management Systems Global Learning Consortium (2003).  
<http://www.imsproject.org>
- Rahm, E., and P. Bernstein (2001). On matching schemas automatically. *Technical Report MSR-TR-2001-17*, Microsoft reasearch.
- Schulmeister, R. (2003). *Lernplattformen für das virtuelle Lernen. Evaluation und Didaktik*. R. Oldenbourg Verlag, München u.a.
- SCORM Project: E-Learning Standard Developed by Advanced Distributed Learning (ADL) Initiative (2003).  
<http://www.adlnet.org>
- Severiens, T. (2003). *My Meta Maker*. <http://physnet.uni-oldenburg.de/services/mmm/>
- Sosteric, M., and S. Hesemeier (2002). *When is a Learning Object not an Object: A First Step Towards a Theory of Learning Objects*. <http://www.irrodl.org/content/v3.2/soc-hes.html>
- Subramanian, S., and K. Shafer (1998). *Clustering in SCORPION Project*.  
<http://orc.rsch.oclc.org/6109/clustering.html>
- Turau, V. (2001). *DB2XML Transforming Relational Databases into XML Documents*.  
<http://www.informatik.fh-wiesbaden.de/~turau/DB2XML/index.html>
- Turau, V. (2003). *Transforming Relational Databases into XML Documents*.  
<http://www.informatik.fh-wiesbaden.de/~turau/DB2XML/index.html>
- WebCT (2003). <http://www.webct.com>
- Worldbank: *Worldbank Global Distance Education Network* (2003).  
<http://www1.worldbank.org/disted/>
- XML-SPY, (2003). *XML Spy Import/Export*. [http://new.xmlspy.com/features\\_convert.html](http://new.xmlspy.com/features_convert.html)

**G. Balbieris**, born in 1974. In 1992 graduated Biržai 2nd Secondary School as well as School of Young Programmers. In 1996 graduated Kaunas University of Technology (KUT) with bachelor diploma in informatics. In year 1998 graduated KUT with master degree in informatics. He continues his studies for PhD in the IT sciences. Last 7 years he is working in e-learning area.

**V. Rėklaitis** has been involved in networking development at Kaunas University of Technology over the last 12 years. He headed joint projects funded by EC under TEMPUS and INCO-Copernicus programmes related to both teaching of networking and developing of e-learning applications. Since April 2002 he is a professor at Graduate School of Information Systems, University of Electro-Communications, Tokyo.

## **E-mokymosi objektų konvertavimas iš reliacinės duomenų bazės į IMS XML standartą**

Giedrius BALBIERIS, Vytautas RĖKLAITIS

Nuo interneto pradžios sukurta daugybė e-mokymosi priemonių ir mokomosios medžiagos. Per pastaruosius keletą metų susiformavę e-mokymosi standartai įgalina perkelti mokomąją medžiagą, priemones iš vienu aplinkų į kitas, įvairiai komponuoti mokomuosius objektus skirtinguose kursuose. Šiame darbe autoriai analizuoja e-mokymosi objektų konversijos iš reliacinės duomenų bazės į IMS XML standartų problemą, pasiūlo savo duomenų konvertavimo metodą. Metodas patikrintas eksperimentiškai konvertuojant pačių sukurtos ir taikomos EduMMDB mokomuosius objektus. Straipsnyje pateikti esminiai struktūrų konversijos principai, išanalizuotos galimybės dar geriau automatizuoti konversijos taisyklių kūrimą.