

Development of learning objects repositories by metadata reutilization from a digital collection: from Dublin Core to IMS

Clara López Guzmán¹, Francisco García Peñalvo²

¹ Dirección General de Servicios de Cómputo Académico
Universidad Nacional Autónoma de México, Mexico

² Facultad de Informática y Automática, Universidad de Salamanca, Spain

^{1,2} Instituto Universitario de Ciencias de la Educación, Universidad de Salamanca, Spain
clara@servidor.unam.mx, fgarcia@usal.es

Abstract. Reutilization, in e-learning context, leads immediately to the concept of learning object and it is always associated with the objects reutilization among platforms or among educational systems, nevertheless digital objects allows the reutilization both of the resources and its metadata, even with systems not directly linked with education but useful as resources to support learning. Due to languages and databases with open technologies and the use of metadata standards this interchange is possible. In this document is explained the experience of share a collection of digital objects that uses Dublin Core metadata standard with a library of learning resources that uses IMS metadata standard.

Formación de repositorios de objetos de aprendizaje a través de la reutilización de los metadatos de una colección digital: de Dublin Core a IMS

Clara López Guzmán¹, Francisco García Peñalvo²

¹Dirección General de Servicios de Cómputo Académico, Universidad Nacional Autónoma de México

²Facultad de Informática y Automática, Universidad de Salamanca

^{1,2}Instituto Universitario de Ciencias de la Educación, Universidad de Salamanca

¹clara@servidor.unam.mx, ²fgarcia@usal.es

Resumen: En el contexto *e-learning* hablar de reutilización lleva de inmediato al concepto de objeto de aprendizaje y se asocia siempre a la reutilización de objetos entre plataformas o entre sistemas educativos, sin embargo las bondades de los objetos digitales permiten la reutilización tanto de los recursos como de sus metadatos, incluso con sistemas no directamente vinculados con la educación pero que sirven como recursos de apoyo para la enseñanza. Gracias a lenguajes y bases de datos con tecnología abierta y al uso de estándares de metadatos este intercambio es posible. En este documento se expone la experiencia del vaciado de datos de una colección de objetos digitales que hace uso del estándar de metadatos Dublin Core hacia una biblioteca de recursos de aprendizaje que utiliza el estándar de metadatos IMS.

1. Introducción

El máximo aprovechamiento de la gran cantidad de recursos disponibles en Internet se logrará cuando la ubicuidad sea su principal característica, cuando los sistemas se intercomunicuen sin importar su campo de aplicación y cuando las tecnologías sean transparentes para los usuarios. Hacia esto apunta lo que se conoce como Web Semántica¹ con la que se pretende formar una infraestructura común y de cooperación que permita compartir y reutilizar los datos a través de aplicaciones, empresas y comunidades. A pesar de estas iniciativas tan amplias, actualmente, se distingue que cada sector trabaja en el desarrollo de sus propios estándares para la interoperabilidad de sus sistemas. Este es el caso del sector bibliotecario, que promueve el uso de publicaciones electrónicas y desarrolla sus bibliotecas digitales y, por otra parte, el sector educativo que desarrolla objetos de aprendizaje y forma repositorios de ellos, ambos en búsqueda de facilitar a sus agremiados el acceso, uso y reutilización de los recursos disponibles. Sin embargo hay una línea muy delgada entre ambos tipos de repositorios y la falta de recursos digitales que apoyen y complementen los sistemas *e-learning* los está llevando a comunicarse y compartir recursos.

¹ <http://www.w3c.org/2001/sw/>

Con el interés de compartir recursos y para su reutilización en el ámbito educativo ha surgido el concepto de “objeto de aprendizaje”, aplicado a materiales digitales creados como pequeñas piezas de contenido o de información [1], con la finalidad de maximizar el número de situaciones educativas en que las que el recursos pueda ser utilizado.

Un componente importante para la reutilización de un objeto es que esté adecuadamente descrito, a través de sus correspondientes metadatos [2]. Si los metadatos utilizados se apegan a los estándares propuestos por los grupos importantes y se basan en tecnologías abiertas también la reutilización de éstos toma un papel de gran valor para compartir información entre repositorios. La consistencia semántica en la descripción de los recursos y la normalización de los datos que cada elemento contenga, también son factores clave para la posible reutilización a través de sistemas.

El propósito de este trabajo es exponer la experiencia de cómo se han reutilizado los metadatos de la biblioteca digital Colecciones Mexicanas para alimentar la Biblioteca Universia de Recursos de Aprendizaje, para ello, primero se introduce a los conceptos referentes a bibliotecas digitales y repositorios de objetos de aprendizaje, para después exponer el caso práctico de cómo se llevo a cabo el vaciado de datos entre ambos sistemas.

2. Colecciones organizadas de objetos digitales

El Web puede ser considerado como una gran colección de objetos digitales. Se pueden encontrar recursos de diferentes formatos y para aplicaciones de todo tipo y sobre cualquier tema. Sin embargo, el problema es que esta colección no tiene un orden, la búsqueda es compleja entre tanta información dispar y la permanencia de la información es impredecible. Para que los recursos de información disponibles en el Web sean verdaderamente explotables se han dado iniciativas y tecnologías para la organización de los recursos. Específicamente para el sector educativo se habla de estos sistemas de organización como bibliotecas digitales y en los últimos años se habla de las bibliotecas o repositorios de objetos de aprendizaje.

2.1 Bibliotecas digitales

Al agrupar o concentrar recursos digitales de información se obtienen colecciones digitales, de las cuales encontramos un sinnúmero en Internet de muy diversos tipos y temas. Cuando estas colecciones se organizan con un sistema descriptivo a través de metadatos (catalogación) y se les asocian algunas facilidades para la búsqueda y uso de la información (servicios), se conocen entonces como bibliotecas digitales [3]. Las bibliotecas digitales basan el contenido de sus repositorios en objetos de información [4], que trascienden al concepto de documento de las bibliotecas tradicionales, refiriéndose a todo tipo de objeto que provea información, como imágenes, videos, animaciones y multimedia.

2.2 Metadatos

Un registro de metadatos consiste en un conjunto de atributos o elementos necesarios para describir el recurso en cuestión. El concepto de metadato existe antes de la aparición de Internet, pero en los últimos años se ha popularizado mucho dada la necesidad que hay de organizar la información en el Web y para la estandarización con vistas a la interoperabilidad de los sistemas de información.

Entre los estándares de catalogación o de metadatos más comúnmente utilizados en las bibliotecas digitales se encuentra la Iniciativa de Metadatos Dublin Core (DCMI), desarrollada para la descripción de un amplio universo de recursos en red ya que su aplicación es de carácter muy general. La propuesta más sencilla de esta Iniciativa, que se conoce como *DCMI Simple*, está conformado por 15 elementos²:

título	editor	fuentes
creador	fecha	idioma
materia	tipo de recurso	relación
descripción	formato	cobertura
colaborador	identificador	derechos

Para una especificación semántica más precisa algunos elementos hacen uso de *refinamientos* (llamados también *calificativos*) a fin de facilitar la búsqueda y recuperación de los recursos. Cuando se hace uso de refinamientos y/o del elemento adicional *audiencia* se conoce entonces como *DCMI Qualified* [5].

Gracias a su simplicidad, DCMI es compatible con otros sistemas de catalogación y puede ser fácilmente implementado en bases de datos o en ficheros XML³ (*eXtended Markup Language*).

2.3 Repositorios de objetos de aprendizaje

Los objetos de aprendizaje son agrupados y almacenados en repositorios de objetos de aprendizaje. De éstos, se identifican dos tipos [6]: los que contienen objetos de aprendizaje y sus metadatos, en éstos los objetos y sus descriptores se encuentran dentro de un mismo sistema e incluso dentro de un mismo servidor; los que contienen sólo los metadatos, en este caso el repositorio contiene sólo los descriptores y se accede al objeto a través de una referencia a su ubicación física que se encuentra en otro sistema o repositorio de objetos.

Regularmente los repositorios de objetos de aprendizaje operan de forma independiente, aunque es común que los LMS (*Learning Management System*) tengan asociado un repositorio que en la mayoría de los casos es sólo para su uso dentro de la misma plataforma.

Los repositorios de objetos de aprendizaje apuntan a la utilización del sistema de metadatos específicamente orientados al campo educativo, como el propuesto por IMS que se basa en el estándar LOM (*Learning Object Metadata*) [7]. Esta propuesta contiene un grupo mínimo de elementos para la administración, ubicación y evalua-

² La descripción detallada se encuentra en <http://dublincore.org/usage/terms/dc/current-elements>

³ XML es un lenguaje de etiquetado sencillo y flexible ampliamente adoptado en la publicación electrónica y para el intercambio de datos en el Web. <http://www.w3.org/XML/>

ción de los objetos de aprendizaje, agrupados en nueve rubros que se describen en la Tabla 1, en el apartado 3.1.

3. Caso práctico

El trabajo realizado fue el intercambio de datos de una colección o biblioteca digital hacia una biblioteca de objetos de aprendizaje. Extrayendo del primero las fichas descriptivas en el estándar Dublin Core (DC) e incorporándolas en el segundo como registros IMS.

Colecciones Mexicanas (CM) es una biblioteca de documentos digitales sobre la historia de México, con documentos del siglo XVII al siglo XIX, con un total de 8876 registros, de tres colecciones diferentes: Primeros Españoles en México, Literatura Mexicana del Siglo XIX y El Archivo Franciscano. Cada registro, que corresponde a un documento digitalizado, tiene una ficha descriptiva en DC.

La Biblioteca Universia de Recursos de Aprendizaje⁴ (BURA) es un repositorio de metadatos de objetos de aprendizaje. Ésta contaba ya con 10,234 registros, también de documentos antiguos principalmente del ámbito literario. BURA utiliza IMS como estándar de metadatos y los registros se encuentran en XML.

Aunque las tecnologías de ambos proyectos era diferente, el haber utilizado sistemas de metadatos estándares permitió el intercambio, siguiendo el siguiente procedimiento:

- a) Análisis del modelo IMS de BURA
- b) Análisis de los metadatos de CM
- c) Análisis de los datos de CM vs. BURA
- d) Vaciado de datos
- e) Verificación de consistencia y compatibilidad.

3.1 Análisis del modelo IMS DE BURA

El primer análisis se hizo sobre el modelo utilizado en BURA para la estructuración de sus metadatos. Se utiliza el grupo completo de metadatos, es decir, los nueve grupos con todos los elementos que cada grupo contiene [8]. Las etiquetas principales, que definen las características de cada grupo se describen en la *Tabla 1*.

La implementación del modelo de metadatos está basada en la especificación de IMS [9] a través de XML Schema, la plantilla que se utiliza es la versión 1.2 (disponible en http://www.imsglobal.org/xsd/imsmd_v1p2p2.xsd).

⁴ <http://biblioteca.universia.net/>

Elemento	Descripción
<general>	Información general que describe el objeto de aprendizaje como un todo. Contiene 9 subelementos.
<lifecycle>	Características relacionadas con la historia y el estado presente del objeto de aprendizaje y de aquellos que han afectado a éste objeto durante su evolución. Contiene 6 subelementos.
<metametadata>	Agrupación de información sobre los mismos metadatos, no sobre el objeto de aprendizaje que se está describiendo. Contiene 10 subelementos.
<technical>	Agrupación de los requerimientos y características técnicas del objeto de aprendizaje. Contiene 11 subelementos.
<educational>	Condiciones del uso educativo del recurso. Contiene 11 subelementos.
<rights>	Condiciones de uso para la explotación del recurso. Contiene 3 subelementos.
<relation>	Define la relación del recurso descrito con otros objetos de aprendizaje. Contiene 7 subelementos.
<annotation>	Comentarios sobre el uso educativo del objeto de aprendizaje. Contiene 3 subelementos.
<classification>	Descripción temática del recurso en algún sistema de clasificación. Contiene 8 subelementos.

Tabla 1. Grupos principales de metadatos

3.2 Análisis de los metadatos en CM.

Colecciones Mexicanas utiliza una base de datos en *postgres*⁵ en donde se definió la estructura de los metadatos de DC con refinamientos. Además de los quince campos originales que DC propone se agregaron tres elementos más: *URL*, *palabras clave* y *audiencia*, quedando un total de 18 elementos como metadatos descriptores de cada objeto de la colección. Por tanto, comparando los metadatos de CM y de BURA, se encontró que había 18 elementos que debían ser integrados a un sistema de metadatos que contiene cerca de 80 elementos.

Además se identificó que habría que extraer la información de los metadatos de CM de la base de datos a un formato que permitiera convertirlos fácilmente a XML. Así que se propuso extraer los datos en tres archivos, uno por colección, a los que se identificó como:

- “F” para Archivo Franciscano
- “L” para Revistas Literarias del siglo XIX, y
- “E” para Primeros Españoles en México.

⁵ Un tipo de manejador de bases de datos

3.3 Análisis de los datos de CM vs. BURA

Aunque en la guía de implementación de IMS [10] ofrece una correspondencia entre los elementos de DC y de LOM, en este caso no se pudo adoptar en su totalidad ya que la correspondencia está pensada para exportar datos de LOM hacia DC, que va de un sistema de descripción muy específico hacia un sistema muy general, por lo cual no hay riesgo de inconsistencia o pérdida de datos. Este caso es el inverso, el vaciado parte del sistema más general hacia uno más detallado (de DC a LOM), por ello, además de un mapeo directo entre los elementos coincidentes, fue necesario hacer ajustes a aquellos que podían empatar y en otros caso el llenado de otros campos que no correspondían con los de DC pero que se podían deducir por las características de las propias colecciones. Adicionalmente, se detectaron algunas inconsistencias semánticas en los datos, que hacían incompatibles los datos aún cuando los elementos podrían corresponderse. Así que se consideraron parcialmente las correspondencias que propone IMS y se elaboró una guía con las reglas para el llenado de cada metadato.

Las reglas adoptadas para la correspondencia de los metadatos de DC y las deducciones hechas para cada grupo de metadatos LOM, se explican en los siguientes apartados.

3.3.1 Reglas generales

Hay etiquetas que tienen un comportamiento consistente aunque se asocien a grupos de metadatos diferentes, por lo cual se hicieron reglas generales para estas etiquetas, llegando a las siguientes consideraciones:

La etiqueta <identifíer> IMS la reserva para una estandarización futura, así que se ha quedado vacío en todos los grupos en los que aparece, al igual que los elementos que son raíz de otros que también quedan vacíos (p.e. <general> y <catalogentry>).

En los casos en que la especificación indica que el elemento utiliza un vocabulario controlado se han tomado los valores que para caso se proponen en el Modelo de Información de Metadatos IMS [8].

3.3.2 Reglas por grupos

En búsqueda de concordancia de los metadatos de CM con los de BURA, se llevó a cabo un análisis de cada grupo, obteniéndose para cada de ellos una descripción de las decisiones de llenado así como una tabla con los datos que cada elemento debe contener, como resultado del análisis y decisiones hechas. A continuación se muestra el caso del primer grupo:

Para el grupo <general> se encontró contenido para todos los elementos. <title>, <entry>, <keyword> y <coverage> tuvieron correspondencia con los campos utilizados en CM. En el caso <keyword> se incluyó tanto el contenido de <palabras clave> como de <tema> de CM ya que BURA se apega a la clasificación UNESCO para la clasificación por temas (en el grupo 9) y no hubo compatibilidad con los temas de CM, así que para conservar la información <tema> se agregó a este elemento.

<catalog> tomó el valor de la colección, según el caso; <structure> y <aggregationlevel> se llenaron en base al vocabulario controlado de IMS para esos elementos; <language> tenía su equivalente en CM pero en inglés (Sp-mx) así que se tomó el valor en español (Es-mx) para mantener consistencia con los datos de BURA.

En el grupo <lifecycle> el elemento <version> quedó vacío porque CM no maneja número de versión; <status> y <role> tomaron valores del vocabulario controlado y <entity> se extrajo de los valores de CM, así como <date> que sólo utilizó la fecha de creación del resto de fechas que manejaba CM.

1	general	
1.1	identifier	
1.2	title	CM{titulo}
1.3	catalogentry	
1.3.1	catalog	Si F: Colecciones Mexicanas - Archivo Franciscano Si L: Colecciones Mexicanas - Literatura Mexicana del Siglo XIX Si E: Colecciones Mexicanas - Primeros Españoles en México
1.3.2	entry	CM{folio}
1.4	language	Es-mx
1.5	description	CM{descripción}
1.6	keyword	CM{palabra clave}+CM{tema}
1.7	coverage	CM{cobertura}
1.8	structure	atomic
1.9	aggregationlevel	Si F: 3; Si L: 1; Si E: 3

Tabla 2. Grupo de metadatos *general*

3.4 Vaciado de datos

Comprendido el proceso de vaciado y establecidas las reglas a considerar en cada elemento, se extrajeron tres archivos .csv de la base de datos de CM. Un archivo por cada colección (F, L y E), a fin de hacer procesamientos independientes y tomar en cuenta las particularidades de cada grupo de datos.

Para la conversión automatizada de los registros se escribió una clase Java (todo BURA está programado en Java) que leía los archivos .csv y convertía cada fila en un archivo XML con esquema IMS. Aplicando a cada uno de los archivos las reglas y el programa de conversión, se obtuvieron las instancias XML del Schema de IMS.

3.5 Verificación de consistencia y compatibilidad

Una vez obtenidas las instancias, se compararon con los registros previamente incorporados en BURA, a fin de corroborar la consistencia de los datos resultantes de las reglas aplicadas contra los datos que el proyecto ya contenía, y se comprobó que los datos eran consistentes entre ellos. Asimismo, se hizo una prueba de integrar un grupo de instancias a BURA para identificar alguna incompatibilidad no prevista, sin embargo los ficheros se integraron sin problema, pudiéndose listar y buscar en el sistema sin ningún inconveniente.

4. Resultados

Del proceso de conversión de los ficheros se han obtenido un lote de 8876 instancias del XML Schema de IMS, que cumplen con las condiciones para ser integradas a cualquier sistema que admita el estándar de metadatos IMS, sea éste una plataforma de aprendizaje (LMS) o un repositorio de objetos de aprendizaje.

Aun cuando de inicio se contaban con sólo 18 posibles datos para ser incluidos en los metadatos de LOM, finalmente fue posible el llenado de aproximadamente el 70% (38 de 54) de los elementos, debido a los campos que se llenaron con el análisis por deducción o inferencia.

Finalmente, las instancias XML se han incorporado al catálogo de BURA. Ahora, después de la reutilización de los metadatos, puede accederse a los objetos, como objetos de información, a través del sitio de la colección en Colecciones Mexicanas o, como objetos de aprendizaje, a través de su referencia en el sitio de BURA, en donde a su vez los objetos incluidos podrán ser reutilizados (Figura 1).

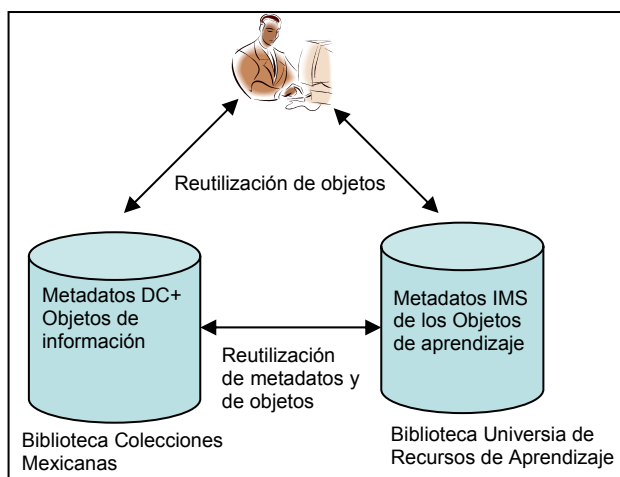


Fig. 1. Esquema conceptual de la reutilización de los recursos y de los metadatos

5. Conclusiones

El utilizar sistemas estándares, en este caso de metadatos, ha facilitado la conversión de los metadatos de un contexto sólo descriptivo y genérico, como el DCMI, hacia un sistema de metadatos de aplicación en el campo educativo, con fines descriptivos, administrativos y de evaluación, como es IMS. Aunque el proceso puede automatizarse, la intervención de personal calificado para el llenado de algunos campos sigue siendo requerido.

El que ambos sistemas hubieran sido desarrollados con tecnologías no propietarias facilitó el acceso a los códigos y a los datos, con lo cual se pudieron hacer las manipulaciones necesarias a los programas para obtener los resultados buscados en este caso.

El problema semántico es de considerable importancia en el vaciado de datos. Aunque la correspondencia entre etiquetas sea consistente, los datos contenidos no siempre guardan esa misma correspondencia, es importante entonces no perder de vista que aunque la tecnología permita el intercambio la semántica presenta casos especiales a analizar previo al ingreso de datos en un catálogo. Este problema semántico lleva a la necesidad de una investigación futura sobre el uso de ontologías en las colecciones digitales.

El impacto del idioma también es considerable ya que los vocabularios obligan a mantener datos en inglés pero la naturaleza de las colecciones tiene por su parte datos en español, así que la consistencia en el lenguaje es un problema que debe estudiarse.

Referencias

- [1] Wiley, D. A Connecting Learning Objects to Instructional Design Theory: A definition, a metaphor, an a taxonomy, Utah State University (2002).
- [2] Sicilia, M. A, García E. On the Concepts of Usability and Reusability of Learning Objects, International Review of Open and Distance Learning (Octubre, 2003), <http://www.irrodl.org/content/v4.2/sicilia-garcia.html/> (2003).
- [3] Borgman, C. L. What are digital libraries, who is building them, and why? En T. Aparac (Ed.), Digital libraries: Interdisciplinary concepts, challenges and opportunities (pp. 29). Zagreb, Benja (1999).
- [4] Leiner, B. M. The Scope of the Digital Library, DLib Working Group on Digital Library Metrics, enero 16, 1998. <http://www.dlib.org/metrics/public/papers/dig-lib-scope.html/> (1998).
- [5] Hillman, D. Using Dublin Core, Dublin Core Metadata Initiative, <http://www.dublincore.org/documents/usageguide/> (2003).
- [6] Downes S. Design and Reusability of Learning Objects in an Academic Context: A New Economy of Education?, National Research Council, Moncton, Canada, <http://www.downes.ca/files/milan.doc/> (2002).
- [7] Learning Technology Standards Commite, WG12: Learning Object Metadata, <http://ltsc.ieee.org/wg12/index.html> (2003).
- [8] IMS Global Learning Consortium, Inc. IMS Learning Resource Meta-data Information Model, Version 1.2.1 Final Specification, IMS, 28 de septiembre (2001).
- [9] IMS Global Learning Consortium, Inc. IMS Learning Resource Meta-data Specification Binding, Versión 1.2.1 Final Specification, 28 de septiembre (2001).
- [10] IMS Global Learning Consortium, Inc. IMS Learning Resource Meta-Data Best Practice and Implementation Guide, (2001).