
7. Caso de estudio: La Biblioteca Colecciones Mexicanas y la Biblioteca Universia de Recursos de Aprendizaje

La implementación de un repositorio con el uso de las especificaciones de IMS queda fuera del alcance de este trabajo. Sin embargo, se expone un caso de estudio para comprobar que el uso de sistemas estandarizados facilita y hace posible la reutilización de los recursos y también de los metadatos.

Un hecho cada vez más frecuente es que las bibliotecas tradicionales y las bibliotecas digitales van adoptado sistemas con tecnologías abiertas (Tenant, 2002), así como sistemas de catalogación que son compatibles con otros y que fácilmente pueden transformarse para poder intercambiarse o vaciarse como registros de otras colecciones (DLC, 2003).

Aprovechando estas condiciones, se han tomando los registros de los metadatos de dos colecciones de recursos digitales. Una de éstas es una biblioteca digital, basada en Dublin Core, la otra es una biblioteca o repositorio de recursos de aprendizaje con metadatos en IMS, entre las que se buscó lograr la reutilización de los metadatos y de los contenidos. Para la exposición del caso que se hace en este capítulo, en el apartado 7.1 se hace el planteamiento de las características generales del caso, después, en el apartado 7.2, se describen los pasos que se han seguido para el proceso de intercambio de metadatos y al final, en el apartado 7.3, se presentan los resultados obtenidos.

7.1. Planteamiento del caso

El trabajo realizado fue el intercambio de datos de una colección o biblioteca digital (Colecciones Mexicanas) hacia una biblioteca de objetos de aprendizaje (Biblioteca Universia de Recursos de Aprendizaje). Se extrajeron las fichas de los metadatos descriptivos del primero y se incluyeron en la base de datos del segundo. Lo que se buscó fue la reutilización de los metadatos de la biblioteca digital para que, a través del mapeo de un estándar al otro y del respectivo vaciado de datos, se tuviera acceso a los recursos como

OI o como OA. Para ubicar el contexto de trabajo, a continuación se describe cada biblioteca y se define el objetivo del caso.

7.1.1. Descripción de las bibliotecas

Colecciones Mexicanas (CM, <http://www.coleccionesmexicanas.unam.mx/>). 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 subcolecciones diferentes (Figura 17):

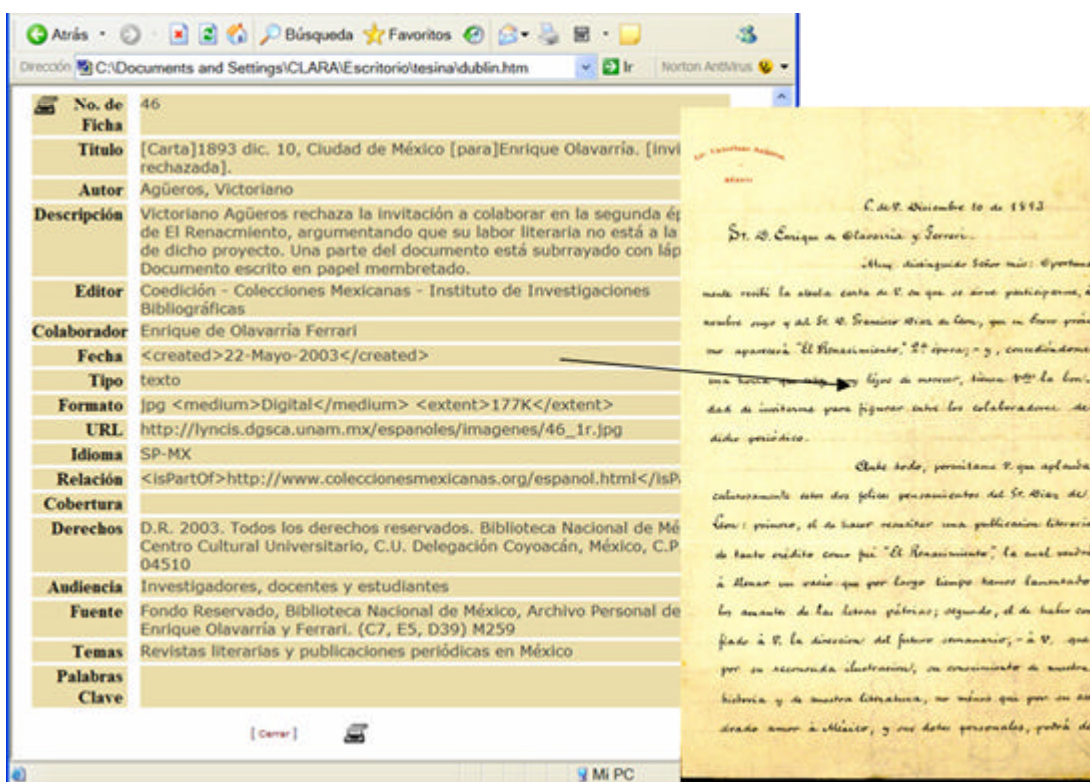
1. El Archivo Franciscano
2. Literatura Mexicana del Siglo XIX, y
3. Primeros Españoles en México



Figura 17. Organización por subcolecciones de la Biblioteca Colecciones Mexicanas

CM al ser una biblioteca digital contiene registros de metadatos que apuntan a OI, que en este caso, en su mayoría corresponden a documentos manuscritos digitalizados, algunos

embebidos en archivos HTML y algunos otros como archivos PDF (*Portable Document File*). Cada registro corresponde a una ficha descriptiva en Dublin Core que contiene la referencia hacia al documento digitalizado en formato JPG (Joint Photographic expert Group) (Figura 18). Los metadatos y los OI se alojan en el mismo servidor.



Ficha de metadatos en Dublin Core

Objeto de Información (Imagen JPG)

Figura 18. Ficha de metadatos que hace referencia al recurso

Biblioteca Universia de Recursos de Aprendizaje (BURA, <http://biblioteca.universia.net/>). Es un proyecto del portal Universia. BURA, es un repositorio de recursos de aprendizaje, contiene metadatos que apuntan a OA (Figura 19). BURA contaba ya con 10.234 registros, también de documentos antiguos principalmente del ámbito literario. En esta biblioteca se utiliza IMS LRM (Thropp & Mckell, 2001a) como estándar de metadatos y los registros se encuentran en XML.

BURA es un repositorio principalmente de metadatos que hacen referencia a los objetos alojados en otros sitios web, aunque también parte de los objetos a los que apuntan sus metadatos se encuentran alojados en el mismo sitio.

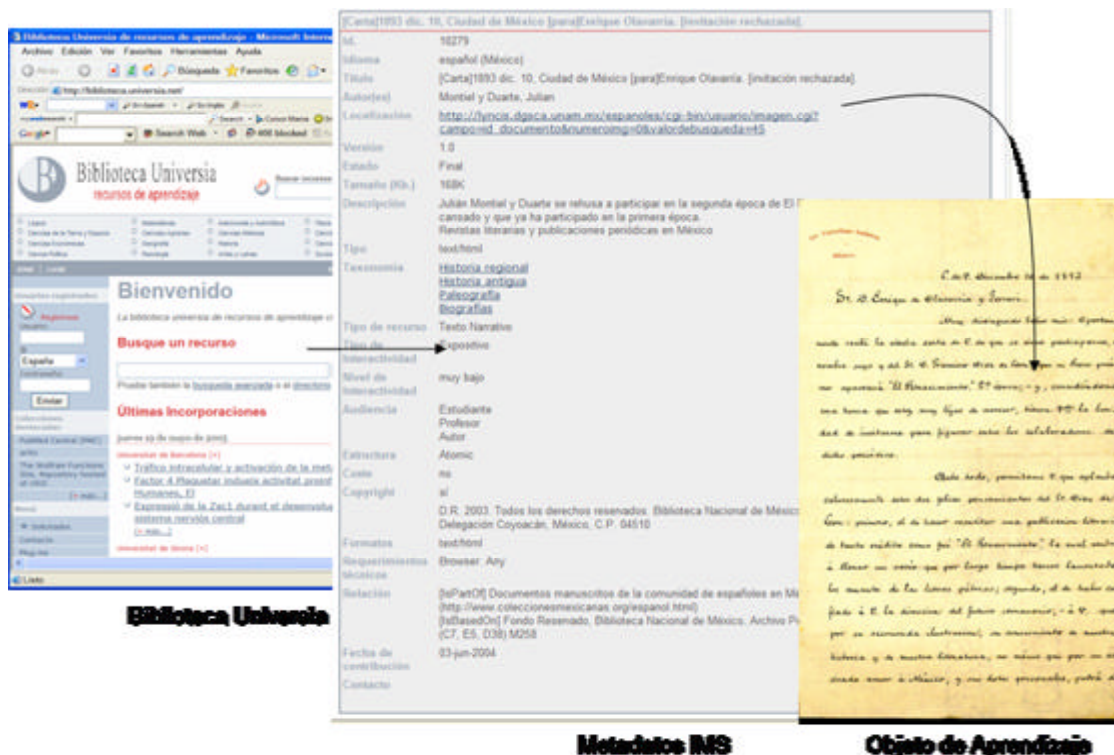


Figura 19. Biblioteca Universia de Recursos de Aprendizaje

7.1.2. Objetivo

El objetivo de este caso fue comprobar que la utilización de sistemas estándares facilita y hace posible la reutilización, tanto de recursos como de metadatos de colecciones de recursos digitales de distinta naturaleza

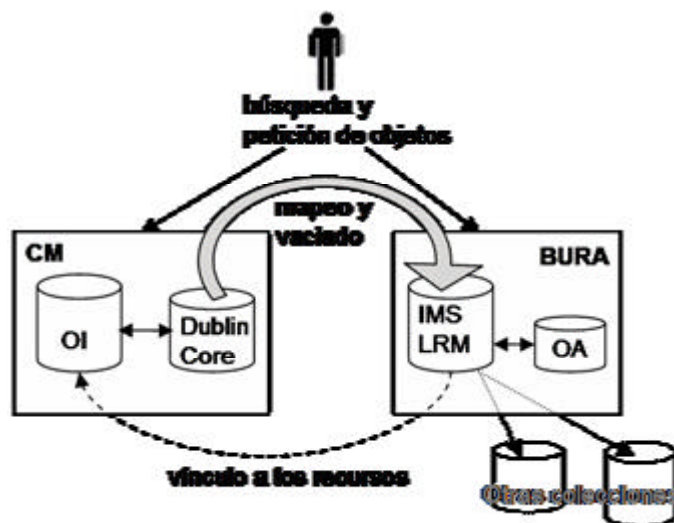


Figura 20. Reutilización de los metadatos y de los objetos de una Biblioteca Digital a través de IMS LRM

Para llegar a este objetivo, se tomaron dos bibliotecas digitales (CM y BURA) y se buscó el intercambio de metadatos entre ellas. En la Figura 20 se ejemplifica el proceso del vaciado de metadatos de CM a BURA. Después de este proceso de vaciado, fue posible el acceso a los objetos desde dos sistemas distintos que hacen uso del mismo recurso, pero vinculado desde repositorios en distintos contextos y con esquemas de metadatos distintos como son Dublin Core e IMS LRM.

7.2. Proceso

Aunque las tecnologías de ambos proyectos son diferentes, el haber utilizado sistemas de metadatos estándares y tecnologías abiertas permitió el intercambio, siguiendo el siguiente proceso:

- 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.

En los siguientes subapartados se describe con detalle cada uno de estos procesos.

7.2.1. Análisis del modelo IMS de BURA

El primer análisis se hizo sobre el modelo utilizado en BURA, para la definición de su esquema de metadatos. Se encontró que utiliza el grupo completo de metadatos de IMS RLM, es decir, las nueve categorías que se revisaron en el subapartado 4.2.3, con todos los elementos que cada una contiene.

La implementación del modelo de metadatos está basada en la especificación de IMS RLM a través de XML *Schema* (Trhopp & Mckell, 2001b). La plantilla que se utiliza es la versión 1.2 (disponible en http://www.imsglobal.org/xsd/imsmd_v1p2p2.xsd).

Se llevo a cabo la revisión de los requisitos de conformidad que la especificación indica (ver apartado 6.1), a fin de asegurar la interoperabilidad del repositorio, en ésta y en futuras aplicaciones. Sobre los metadatos se verificó su concordancia con LOM, tanto en la definición de etiquetas como en la jerarquía de los grupos, asimismo se identificó que no se utilizan extensiones, por lo tanto no era necesario revisar si habían elementos que

estuvieran en conflicto semántico con los elementos de LOM. Por otra parte, sobre el repositorio se verificó que estaba preparado para procesar todos los elementos LOM y que no hace ningún cambio en los datos para sus retransmisiones. Así que, después de estas revisiones se concluyó que BURA cumplía con los requisitos de conformidad de IMS LRM.

7.2.2. *Análisis de los metadatos en CM*

Colecciones Mexicanas utiliza una base de datos en PostgreSQL²⁶ en la que se definió la estructura de los metadatos basándose en Dublin Core *qualified*, es decir, el que hace uso de los refinamientos de los elementos principales para hacerlos más detallados (ver Anexo A).

Además de los quince campos originales que Dublin Core propone, se agregaron tres elementos más: URL, palabras clave (*keywords*) y audiencia (*audience*), quedando 18 elementos como metadatos descriptores principales de cada objeto y, como algunos de ellos tenían refinamientos (*date, format y relation*), se tenían en total 20. Por tanto, comparando los metadatos de CM y de BURA, se encontró que había 20 elementos (listados en la Tabla 10) que debían ser integrados a un sistema de metadatos que contiene más de 70 elementos (los correspondientes de IMS en BURA).

Elementos Dublin Core	Refinamientos	Agregados
title	Date.created	URL
creator	Format.medium	Keywords
subject	Format.extent	Audience
description	Relation.ispartof	
Publisher		
Contributor		
Type		
Format		
Identifier		
Source		
Language		
Coverage		
Rights		

Tabla 10. Elementos de CM

Dado que los metadatos estaban programados como base de datos, se detectó la necesidad

²⁶ PostgreSQL es un administrador de bases de datos relacional, de código abierto, <http://www.postgresql.org/>.

de transformarlos en un formato que permitiera convertirlos fácilmente a XML, por lo que de la base de datos se extrajeron archivos en texto plano. Para minimizar la posibilidad de errores en el vaciado y para hacer las pruebas más fáciles, se decidió hacer la extracción de los datos en tres archivos, uno por subcolección, a los que se identificó como:

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

Como ejemplo de estos archivos, en la Tabla 11 se presenta los datos que corresponden al registro número cuatro de la subcolección *E*, extraído en formato .txt. El contenido de cada metadato está definido por un caracter separador (|), se mantiene un orden y una secuencia acorde a la estructura de los metadatos. La misma estructura de dicho registro lo tienen el resto de los registros extraídos por lo que se facilitan su transformación posterior como archivos XML.

4
88
[Carta]1875 abr. 25, Ciudad de México [para]Enrique Olavarría. [encargo de Enrique de Olavarría y promoción de algunas obras musicales mexicanas en España].
Morales, Melesio
Melesio Morales expresa el dolor que le causó la muerte de su suegra. Está enterado de que la enfermedad ocular de Olavarría ha cedido. Ha cumplido el encargo de éste: ver muestrario de imprenta, boletines y tipos. Agradece que dé a conocer a México en el extranjero; pide haga conocer sus obras musicales en Madrid. Olavarría se encuentra en Madrid. Documento escrito en papel esquila.
Coedición - Colecciones Mexicanas - Instituto de Investigaciones Bibliográficas
Enrique de Olavarría Ferrari <created>22-05-2003</created> texto jpg
<medium>Digital</medium><extent>111K</extent>
http://lyncis.dgsca.unam.mx/espanoles/imagenes/
SP-MX <isPartOf>http://www.coleccionesmexicanas.org/espanol.html</isPartOf>
Archivo Personal de Enrique Olavarría y Ferrari. Archivos y manuscritos del siglo XIX y XX. México
D.R. 2003. Todos los derechos reservados. Biblioteca Nacional de México. Centro Cultural Universitario, C.U. Delegación Coyoacán, México, C.P. 04510
Investigadores, docentes y estudiantes
Fondo Reservado, Biblioteca Nacional de México, Archivo Personal de Enrique Olavarría y Ferrari. (C6, E5, D6) M88

Tabla 11. Registro extraído en texto plano de la base de datos de CM

7.2.3. *Análisis de los datos de CM vs. BURA*

Una vez realizado el análisis de los metadatos de cada biblioteca, se procedió entonces a hacer una revisión comparativa de éstos, a fin de poder realizar el vaciado de los datos de CM hacia el esquema utilizado por BURA. Aunque en la guía de implementación de IMS se propone una correspondencia para el mapeo entre los elementos de LOM (que son la base de IMS) y de Dublin Core, en este caso no se pudo adoptar en su totalidad ya que la

correspondencia está pensada para exportar datos de LOM hacia Dublin Core.

Elemento en BURA	Elemento de CM
General <general> Identifier Title Catalogentry Language Description Keyword Coverage Structure Aggregation Level	- title identifier - description keywords + subject coverage - -
Life Cycle <lifecycle> Version Status Contribute	- - creator + publisher + date.created
Meta-Metadata <metametadata> Identifier Contribute Metadata Schema Language	- - - - -
Technical <technical> Format Size Location Requirement Installation Remarks Other Platform Requirements	format format.extent URL - - - -
Educational <educational> Interactivity Type Learning Resource Type Interactivity Level Semantic Density Intended End User Role Context Typical Age Range Difficulty Typical Learning Time Description Language	- - - - audience - - - - - -
Rights <rights> Cost Copyright and Other Restrictions Description	- - rights
Relation <relation> Kind Resource	- source
Annotation <annotation> Entity Date Description	- - -
Classification <classification> Purpose Taxon Path Description Keyword	- - - -

Tabla 12. Correspondencia de elementos entre BURA y CM

Es decir, la especificación contempla el intercambio de un sistema de descripción muy específico hacia un sistema muy general, por lo que no hay riesgo de inconsistencia o pérdida de datos.

En este caso, se tiene el problema inverso: el vaciado parte del sistema más general hacia uno más detallado (de Dublin Core a LOM). Por otra parte, el mapeo de la especificación está limitado a DC *unqualified* (sin refinamientos), así que las reglas del mapeo de la especificación tampoco cubre todos los metadatos que se utilizaron en CM, faltan los refinamientos y los tres elementos que se agregaron. Por ello, además de identificar el mapeo que correspondiera a los elementos de cada esquema, fue necesario hacer ajustes para aquéllos que se detectó que podían corresponder con algún otro elemento, por el tipo de contenido. Quedando como correspondencia entre elementos la que se muestra en la Tabla 12.

Después de hacer la correspondencia y el acomodo de los datos elementos CM, se observó que sólo fue posible reutilizar 17 de los 20 elementos disponibles. El elemento `format.medium` (formato del medio), cuyo valor era “digital”, no tuvo correspondencia con ningún elemento. El elemento `language`, aunque pudieran haber tenido una correspondencia con `general.language`, se decidió no utilizarlos ya que su contenido no cumplía con la normalización de datos de BURA, por ejemplo, `language` en CM contenía “sp-mx” y en BURA se tenía “es-mx”. Por último, `type` tampoco se utilizó, aunque pudo ser compatible con `educational.learninresourcetype`, pero tampoco cumplía con la normalización de los datos de BURA.

Era claro que el vaciado era pobre por la cantidad de elementos que quedaban vacíos, pero también era evidente que muchos de los elementos de BURA podían llenarse de forma automatizada por los datos generales de ambas bibliotecas y de cada subcolección. Así que se procedió a revisar los elementos vacíos y a tratar de asignarles datos de forma intuitiva (por datos conocidos) y /o a través de los vocabularios controlados que BURA ya tenía establecido por haberse apegado a IMS. Así que, por grupos, se realizó el análisis y se obtuvieron reglas generales que afectaban a todos los elementos y reglas por grupos, que se describen a continuación.

7.2.4. Reglas generales

La etiqueta <identifier> IMS la reserva para una estandarización futura, así que se ha quedado vacía 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 se proponen en el Modelo de Información de Metadatos IMS. En el Anexo C se presenta un extracto de la guía que propone el Modelo para el llenado de cada elemento y a la que puede referirse para identificar los vocabularios y recomendaciones que se han asumido para las siguientes reglas por grupos.

7.2.5. Reglas por grupos

Para el grupo *general* (Tabla 13) se encontró contenido para todos los elementos:

- En los casos <title>, <entry>, <keyword> y <coverage>, éstos tuvieron correspondencia con los campos utilizados en CM. Para <keyword> también se agregó el contenido de <subject>, ya que éste no pudo corresponder con <classification.subject>, dado que BURA utiliza la clasificación UNESCO²⁷ y no había coincidencia con los temas de <subject> de CM, así que para aprovechar el contenido de este campo se unió al de palabras clave.
- <catalog> tomó valor según el nombre de cada subcolección, es decir para los registros correspondientes al archivo F, se asignó el valor “Colecciones Mexicanas - Archivo Franciscano”, de igual forma se hizo para los otros dos archivos. Aquí puede observarse la conveniencia de hacer archivos separados para cada subcolección, ya que, como este caso, habrá otros que requieren datos distintos para un mismo elemento. Así el vaciado de datos podrá hacerse de acuerdo a las particularidades de cada subcolección.
- A <structure> se le asignó el valor “atomic”, opción posible del vocabulario

²⁷ La clasificación UNESCO, es un sistema de clasificación del conocimiento para la ordenación de recursos por temas, a través de un sistema de codificación numérica. El nivel temático mayor se codifica con dos dígitos y se denominan *campos*, los *campos* contienen varias *disciplinas*, codificadas con 4 dígitos y éstas tienen *subdisciplinas* que se codifican con 6 dígitos. Por ejemplo, 58 corresponde a *pedagogía*, por lo que 5801 a *teoría y métodos Educativos* y 580101 corresponde a *medios audiovisuales*. Puede consultarse la lista completa en <http://etpx22.bs.ehu.es/varios/unesco.htm>.

para éste elemento (consultar Anexo C), ya que se consideró que los objetos que se manejan en CM tienen una estructura atómica (mínima).

- Para <aggregationlevel> se dieron valores distintos para cada subcolección, ya que éstas tienen contenidos distintos. Los valores se asignaron según los niveles de agregación que el estándar define, por lo que a F y E se les dio un nivel de agregación 3 que corresponde a documentos HTML unidos por un índice y a L se le asignó el nivel de agregación 1, por ser éstos documentos PDF que se consideran con un nivel de agregación mínimo.

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{identifier}
1.4	language	Es-mx
1.5	description	CM{description}
1.6	keyword	CM{keywords}+CM{subject}
1.7	coverage	CM{coverage}
1.8	structure	atomic
1.9	aggregationlevel	Si F: 3; Si L: 1; Si E: 3

Tabla 13. Grupo de metadatos *general*

El elemento <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.

Para el grupo *lifecycle* (Tabla 14): el elemento <version> quedó vacío porque CM no maneja número de versión; a <status> se le asignó el valor “final” (de vocabulario), ya que los objetos se encuentran en estado final.

2	lifecycle	
2.1	version	
2.2	status	final
2.3	contribute	
2.3.1	role	Autor editor content provider
2.3.2	entity	CM{creator} CM{publisher} Si F: Biblioteca Nacional de México, Si L: DGSCA, Si E: Espamexis
2.3.3	date	CM{date <created>}

Tabla 14. Grupo de metadatos *lifecycle*

Al elemento <role>, subelemento de <contribute>, se le asignaron tres roles de colaboradores: “autor”, “editor” y “content provider”, valores del vocabulario, y cuyas entidades se reflejaron en <entity> importando los valores de <creator> y

<publisher> para los dos primeros roles. Para el tercer rol se asignó un valor distinto para cada subcolección (Ver Tabla 14) que se extrajo de los valores de CM, así como <date> que sólo utilizó la fecha de creación del resto de fechas que manejaba CM.

En el grupo *metadata* (Tabla 15) había que llenar los elementos con la información de los datos que se estaban introduciendo, así que los metadatos correspondían al catálogo de BURA y al número que el sistema le asigna (<catalog> y <entry>). Los autores de los metadatos son UNAM y UNIVERSIA, se asignó la fecha de creación correspondiente y se identificaron como metadatos del LOM versión 1.0 (<metadatascheme>), en idioma (<language>) español (Es).

3	metadata	
3.1	Identifier	
3.2	catalogentry	
3.2.1	Catalog	BURA
3.2.2	entry	{No. BURA}
3.3	contribute	
3.3.1	role	Creator
3.3.2	entity	UNAM-UNIVERSIA
3.3.3	date	01/06/2004
3.4	metadatascheme	LOMv1.0
3.5	language	Es

Tabla 15. Grupo de metadatos *metadata*

En el grupo *technical* (Tabla 16), para <format> se dio un valor a cada subcolección según el vocabulario permitido: “applicaton/pdf” para L ya que son archivos PDF, y “Text/html” para E y F que son documentos en HTML. <size> tuvo correspondencia con el elemento <format.extent> de CM. El elemento <location> pudo llenarse por correspondencia con el elemento <URL> de CM, pero sólo para una subcolección, en las otras colecciones, el URL de cada documento, se genera de forma dinámica, pero se hizo la programación correspondiente para obtenerla. A <type>, del vocabulario se le asignó “browser” ya que es el software necesario para poder visualizar los recursos, y como el tipo de *browser* puede ser cualquier, a <name> se le asignó el valor “any”. En el caso de la subcolección L se requiere de Adobe Acrobat Reader²⁸ para visualizar los PDF así que para este caso, quedó indicado en <otherplatformrequirements>. Los otros elementos quedaron vacíos al no encontrarse datos que pudieran agregarse.

²⁸ Adobe Acrobat Reader es el lector de facto para archivos PDF.

4	technical	
4.1	format	Si L: application/pdf Si E: Text/html Si F: Text/html
4.2	size	CM{format <extent>}
4.3	location	Si L: CM{URL} Si F: http://lyncis.dgsca.unam.mx/franciscanos/cgi-bin/usuario/imagen.cgi?campo=id_documento&valordebusqueda=253&numeroimg={folio} Si E: http://lyncis.dgsca.unam.mx/espanoles/cgi-bin/usuario/imagen.cgi?campo=id_documento&valordebusqueda=253&numeroimg={folio}
4.4	requirement	
4.4.1	type	Browser
4.4.2	name	Any
4.4.3	minimu mversion	
4.4.4	maximu mversion	
4.5	instalationremarks	
4.6	otherplatformrequirements	Si L: Adobe Acrobat Reader
4.7	duration	

Tabla 16. Grupo de metadatos *technical*

Los valores de los elementos de *educational* (Tabla 17) que pudieron obtenerse, fueron tomados de los vocabularios controlados, excepto <intendeduserrole> que ya se había identificado con correspondencia con <audience>. Por las características de los objetos se dedujo que el tipo de interactividad era expositivo, los tipos de recursos eran textos narrativos, el nivel de interactividad era muy bajo y se identificaron varios contextos para su uso. En el caso de las etiquetas que quedaron vacías se consideró que el asignar valor sería una decisión muy subjetiva así que se optó por no asignar dato alguno.

5	educational	
5.1	interativitytype	Expositive
5.2	learningresource type	Narrative text
5.3	interactivitylevel	Very low
5.4	semanticdensity	
5.5	intendeduserrole	CM{audience}
5.6	context	Higher Education, University Postgrade, Professional formation, Continuous Formation.
5.7	typicalagerange	
5.8	difficulty	
5.9	typicallearningtime	
5.10	description	
5.11	language	

Tabla 17. Grupo de metadatos *educational*

En el grupo de derechos de autor, *rights* (Tabla 18), el contenido para `<cost>` y `<copyrightandotherrestrictions>` se tomó de vocabularios, deduciendo que CM no tiene costo asociado al uso de sus recursos pero sí tiene restricciones para su uso, que son las especificadas en el elemento `<rights>` de CM y que correspondieron con `<description>`.

6	rights	
6.1	cost	no
6.2	copyrightandotherrestrictions	yes
6.3	description	CM{rights}

Tabla 18. Grupo de metadatos *rights*

7	relation	
7.1	kind	IsBasedOn
7.2	resource	
7.2.1	identifier	
7.2.2	description	Si F: Manuscritos del siglo XVI al XIX sobre los Franciscanos en México. Si L: Revistas literarias que reflejan la sociedad mexicana del siglo XIX. Si E: Documentos manuscritos de la comunidad de españoles en México en el siglo XIX.
7.2.3	catalogentry	
7.2.3.1	catalog	CM{source}
7.2.3.2	entry	CM{source.referenciacatalográfica}

Tabla 19. Grupo de metadatos *relation*

En el grupo *relation* (Tabla 19), se quiso mantener la relación del recurso digital con el recurso en papel que se digitalizó, así que el el valor para el tipo de relación, `<kind>`, se tomó del vocabulario disponible y se asignó “IsBasedOn”. Dado que se quería tener la relación de cada subcolección con su temática, ésta se asignó en `<description>`. `<catalog>` tuvo correspondencia con `<source>` de CM y para `<entry>` se extrajo un dato particular del contenido de `<source>` que correspondía a la referencia bibliográfica de cada registro en sus bases de datos originales y que se consideró importante conservar como indentificador del recurso original (el original papel). En el grupo *annotation* (Tabla 20) no se agregó ningún dato ya que no había comentarios que agregar sobre los recursos.

8	Annotation	
8.1	person	
8.2	date	
8.3	description	

Tabla 20. Grupo de metadatos *annotation*

Se había mencionado anteriormente que BURA se apega a la norma UNESCO y que no fue posible tener una correspondencia con entre los temas de ambas bibliotecas, ya que CM tiene un sistema de clasificación especializado, en el que se eligieron temas específicos, relevantes para el tipo de documentos de las subcolecciones y que fueron definidos por especialistas en historia de México. Por ello, en el grupo *classification* (Tabla 21), <subject> no tuvo compatibilidad con los subelementos de <taxonpath> y se extrajeron los temas correspondientes a cada colección (<id>) de la norma UNESCO (<source>). El propósito fue clasificar por disciplinas (<purpose>).

9	classification			
9.1	purpose	Discipline		
9.2	taxonpath			
9.2.1	source	UNESCO		
9.2.2	taxon			
9.2.2.1	id	F: 550302 550401 550508 550621	L: 550401 620203	E: 550302 550401 550508 5501
9.2.2.2	entry	[tema correspondiente por código UNESCO]		
9.3	description	Clasificación según códigos UNESCO		
9.4	keyword			

Tabla 21. Grupo de metadatos *classification*

Una vez realizado este análisis para llegar a un proceso normalizado para el intercambio de metadatos, entonces si dió paso al proceso de vaciado de datos entre los sistemas.

7.2.6. *Vaciado de datos*

Comprendido el proceso de vaciado y establecidas las reglas a considerar en cada elemento, se extrajeron tres archivos 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 Java²⁹ llamada *CSVColmex* y se utilizaron otras clases genéricas que BURA ya tenía creadas: *CSVReader* y *CSVIndexer*.

²⁹ Una clase Java es código informático que define las variables y los métodos comunes de un grupo de objetos del mismo tipo (Sun Microsystems, 2005).

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
-<lom xmlns="http://www.imsglobal.org/xsd/imsmd_vlp2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.imsglobal.org/xsd/imsmd_vlp2
imsmd_vlp2p2.xsd">
- <general>
- <title>
- <langstring xml:lang="es">Carta de Luis Antonio Menchaca al virrey de
Cruillas; que env a preso a la capital [M xico] a un indio esp a de la
naci n comanche: Real presidio de San Antonio de B jar, 23
agosto...</langstring>
- </title>
- <catalogentry> <catalog>coleccionesmexicanas.unam.mx</catalog>
- <entry> <langstring>500</langstring>
- <langstring>bura:coleccionesmexicanas.unam.mx:500</langstring>
- </entry>
- </catalogentry>
- <language>es_MX</language>
- <description>
- <langstring xml:lang="es">Manuscrito que contiene Carta de Luis
Antonio Menchaca al virrey de Cruillas; que env a preso a la capital
[M xico] a un indio esp a de la naci n comanche: Real presidio de San
Antonio de B jar, 23 agosto 1764.</langstring>
- </description>
- <keyword>
- <langstring xml:lang="es">Presidios Indios Diligencias
Reos</langstring>
- </keyword>
- <keyword>
- <langstring xml:lang="es">DEFENSA Y SEGURIDAD LEYES Y LEGISLACI N
PRISIONEROS PRISIONES</langstring>
- </keyword>
- <coverage>
- <langstring xml:lang="es">Manuscritos Siglo XVI - XIX</langstring>
- </coverage>
- <structure>
- <source> <langstring xml:lang="x-none">LOMv1.0</langstring>
- </source>
- <value> <langstring xml:lang="x-none">Atomic</langstring> </value>
- </structure>
- <aggregationlevel>
- <source>
- <langstring xml:lang="x-none">LOMv1.0</langstring>
- </source>
- <value>
- <langstring xml:lang="x-none">3</langstring>
- </value>
- </aggregationlevel>
- </general>
- <lifecycle>
- .
- .
- .
</lom>

```

Tabla 22. Registro de una instancia XML de IMS LRM.

El funcionamiento de las clases es el siguiente: los tres archivos extra dos de CM fueron guardados como hojas de c culo, quedando en la primera l nea los nombres de los metadatos. *CSVIndexer* lee los archivos (utilizando *CSVReader*) y para cada entrada crea

un mapa con todos los valores de las columnas, usando como claves los nombres de los metadatos. Ese mapa es pasado como parámetro de entrada al método *visit* de *CSVColmex*. *CSVColmex* contiene la lógica para la conversión de los metadatos de un registro de CM a un registro BURA. Dentro de la clase hay un método llamado *lom*, que se encarga de extraer los valores de cada registro y transformarlos a IMS 1.2, convirtiendo apropiadamente los datos y añadiendo valores de defecto a otros campos de la estructura. En el Anexo D se incluye el código de esta clase, el código de las otras clases no se incluye ya que se considera de poca relevancia para el objetivo que persigue este caso. Aplicando a cada uno de los archivos las reglas y el programa de conversión, se obtuvieron las instancias XML del Schema de IMS. En la Tabla 22 se muestra parte del código XML correspondiente a la instancia del registro número 500.

7.2.7. 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. Se comprobó que los datos eran consistentes entre ellos. Asimismo, se hizo una prueba de integrar un grupo de instancias (como la mostrada en la Tabla 22) 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.

7.3. 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 una biblioteca de objetos de aprendizaje.

Aun cuando de inicio se contaban con sólo 20 posibles datos para ser incluidos en los metadatos de LOM, finalmente fue posible el llenado de aproximadamente el 70% de los elementos, gracias a los campos que se llenaron por el uso de vocabularios de LOM o por los datos que se podían extraer de la información conocida de las bibliotecas examinadas.

El intercambio de metadatos entre los esquemas fue posible, aunque vale la pena resaltar

que no fue posible utilizar toda la información de los metadatos de CM, debido a la inconsistencia entre términos y a los diferentes sistemas de normalización que cada una manejaba. También fue notable la cantidad de análisis que tuvo que realizarse por humanos, ya que al partir de un sistema general hacia uno más específico se necesitaba inteligencia para discernir y comprobar la compatibilidad entre los datos y llegar a la mejor opción de equivalencia.

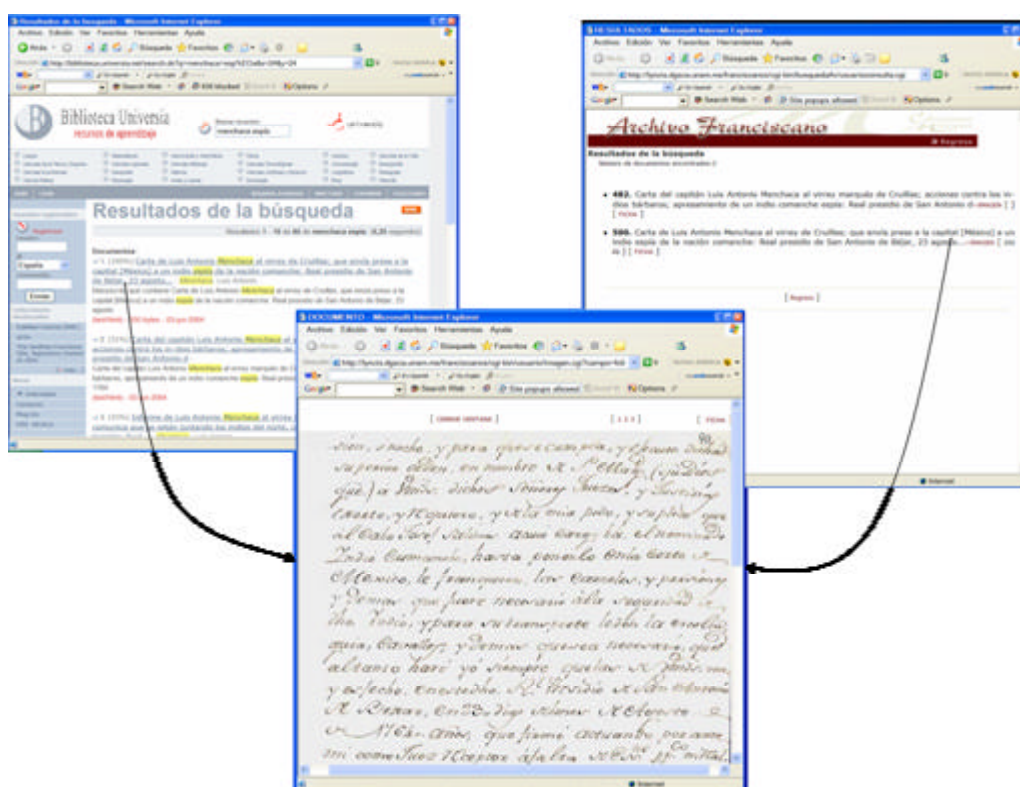


Figura 21. Acceso al objeto desde ambas Bibliotecas.

Finalmente, las instancias XML se han incorporado al catálogo de BURA. Ahora, 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 del sitio de BURA (Figura 21).