

Migración a Marc 21 del catálogo Roble de la Universidad Nacional de La Plata

por Gustavo Archuby

(Analista de Sistemas, Administrador de la Red Internet/Intranet de la Facultad de Humanidades y Ciencias de la Educación, UNLP)

Mariana Pichinini

Bibliotecaria Documentalista, Responsable Área de Procesos Técnicos, Biblioteca de la Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata

Resumen

Se presenta un informe de la conversión realizada del catálogo de libros ROBLE de la Universidad Nacional de La Plata, creado en forma cooperativa con los aportes de las bibliotecas de la Universidad y elaborado en formato propio (BIBLO); al formato MARC21. Se explica el análisis de campos realizado, las decisiones adoptadas en función del software, la estructura del catálogo y de los registros, y los productos obtenidos. Se discuten las herramientas utilizadas y los subproductos resultantes de la conversión. Se presentan los objetivos propuestos y pasos a seguir en la siguiente etapa, que incluyen el desarrollo de una interfaz para acceso online que permita exportación a XML-MARC y el desarrollo de una aplicación para upload y validación de los registros de las diferentes bibliotecas al catálogo MARC.

Introducción

La conversión del catálogo ROBLE de libros (124356 registros), desarrollado con un formato propio (BIBLO), al formato MARC21, forma parte del proyecto de conversión y actualización y se realizó durante marzo, abril y mayo de 2003.

Según la metodología establecida, en primer lugar se analizaron y compararon los campos de BIBLO con MARC21 y se estudiaron las posibles conversiones a realizar.

En segundo lugar, se estableció una tabla de equivalencias entre ambos formatos a partir de la cual se desarrollaría el software de conversión, y se adoptó una política de delimitación de niveles en cada registro importado teniendo en cuenta los campos presentes.

Finalmente, y una vez realizada la conversión, se desarrolló un Online Public Access Catalog (OPAC), con una interfaz de búsqueda mejorada y visualización de registros en formato MARC tagged.

Análisis del formato MARC21 y comparación de campos con BIBLO

Se tomó como referencia el formato MARC21, Concise Format for Bibliographic Data, editado por Library of Congress (<http://www.loc.gov/marc/bibliographic/>). Se verificó con Bibliographical Formats and Standards - Third Edition editado por OCLC (<http://www.ocic.org/bibformats/>) y

también se tuvo en cuenta el National Level Record - Bibliographic Mínima/ Level que requiere la Library of Congress para Estados Unidos.

Descripción Tag	Campos convertidos
001 Número de control	
003 Identificador de número de control	Sí
005 Fecha y hora de la última modificación	Sí
008 Campo de longitud fija- Información general	Sí
020 ISBN	Sí
040 Origen de la catalogación	Sí
041 Código de idioma	Sí
044 Código de país productor	Sí
080 Número de CDU	Sí
100 Entrada principal - Nombre personal	Sí
110 Entrada principal - Autor corporativo	Sí
111 Entrada principal - Congresos, conferencias	Si
130 Entrada principal - Título uniforme	Sí
240 Título uniforme	Sí
243 Título uniforme colectivo	Sí
245 Título propiamente dicho	Sí
246 Variaciones en el título	Sí
250 Mención de edición.	Si
260 Publicación, distribución, etc.	Sí
300 Descripción física	Sí
440 Título de la serie y entrada secundaria	Sí
490 Título de la serie sin entrada secundaria	
500 Nota general -	Sí
501 Nota de Con	
502 Nota de tesis	Sí
504 Nota de bibliografía	Sí
505 Nota de contenido	Si
546 Nota de idioma	
586 Nota sobre premios	
590 Notas locales	Sí
600 Entrada secundaria de materias: Nombres personales	
610 Entrada secundaria de materias: Nombres corporativos	
611 Entrada secundaria de materias: Congresos, reuniones, etc.	
630 Entrada secundaria de materias: Título uniforme	
650 Entrada secundaria de materias: Términos temáticos	
651 Entrada secundaria de materias: Nombres geográficos	
653 Entrada secundaria de materias: Término de indización no Controlado	Sí
700 Entrada secundaria - Nombres personales	Sí
710 Entrada secundaria - Nombres corporativos	Sí
711 Entrada secundaria - Congresos, etc.	
740 Entrada secundaria - Título analítico	
800 Entrada secundaria de serie - Autor personal	

810	Entrada secundaria de serie - Autor corporativo	
811	Entrada secundaria de serie - Reunión	
830	Entrada secundaria de serie - Título uniforme	
900	Entrada digitalizador	Sí
910	Signatura topográfica	Sí
920	Inventario	Sí
930	Grupo temático	

Las entradas secundarias de materia no se pudieron convertir, ya que los formatos utilizados en el catálogo no discriminan los temas siguiendo la estructura de MARC, por lo cual se colocaron todos en un mismo campo (653 Término de indización no controlado).

El resto de los campos MARC que no se convirtieron corresponden a datos no registrados en ROBLE y, por lo tanto, no forman parte del procedimiento de conversión. En la medida en que se incluyan en futuras actualizaciones, se analizará el campo al cual se deben agregar.

3. Decisiones adoptadas en función de la estructura del catálogo ROBLE

3.1. Supresión de indicadores

Debido a que no se podían establecer automáticamente los indicadores para todos los campos y teniendo en cuenta que los indicadores de los campos más importantes (grupos 100, 200 y 700) no se podían inferir a partir de la estructura de los datos, se decidió la supresión de los mismos en todos los registros.

3.2. Creación de niveles en los registros

Los registros provenientes de las diferentes bibliotecas no son uniformes, ya que hay diferencias entre los formatos que no permiten su convertibilidad homogénea. Esto podría resultar en una falta de uniformidad del catálogo en general. Para evitar esta posibilidad, se establecieron varios niveles de registros dentro del catálogo, que se definieron en función de la cantidad de campos MARC presentes y la importancia del bloque de MARC existente.

Nivel 3: Grupo mínimo

Bloques obligatorios: 200 y 900

Bloques opcionales: 000, 100, 700 Título y datos de ubicación y ejemplar

Nivel 2: Grupo medio

Bloques obligatorios: 200, 300 y 900

Bloques opcionales: 000, 100, 400, 500 y 700

Descripción bibliográfica completa sin descripción temática

Nivel 1: Grupo máximo

Bloques obligatorios: 200, 300, 600 y 900 Bloques opcionales: 000, 100, 400, 500 y 700

Descripción bibliográfica y temática

Cabe aclarar que los bloques 000, 100, 400, 500 y 700 pueden estar presentes o no, ya que un documento puede no contener los datos de dichos bloques. Por ejemplo, puede haber un documento que no posea autor. Estos bloques, por lo tanto, no se pueden incluir en la

obligatoriedad, aún cuando su presencia esté incluida en los niveles.

3.3. Corrección de registros

Un registro cata lo gráfico mínimo tiene que tener título y ubicación. Se pudieron detectar algunos registros sin campo título que pertenecían a congresos, a los que se agregó automáticamente la leyenda [trabajos, actas, ponencias, resúmenes, etc.". También se detectaron registros sin ubicación en los que se determinó a partir del contexto y se agregó automáticamente.

3.4 Eliminación de campos mal cargados

La existencia de campos mal cargados requería un trabajo manual de edición para su reparación e impedía la conversión automática. Estos datos fueron eliminados, ante la imposibilidad de determinar el dato correcto.

4. Chequeo de consistencia de los datos

Los chequeos que permiten detectar fallos en los registros, como campos con errores de carga, inexistentes, etc., que deberían corregirse/eliminarse para evitar afectar la conversión, se realizaron básicamente con CISIS utilities y WXIS

Estos son:

Existencia, correcta longitud y cantidad de ocurrencias de los campos en la base Chequeo de valores tabulados (como códigos de países, idiomas y bibliotecas)

5. Procedimientos de conversión

5.1. Conversión de idiomas y países en código 150 a código MARC

Se realizó un gizmo de conversión de lenguas y países, mediante una tabla de traducción ISO-MARC para lenguas (isoMarcLen) y países (isoMarcPai). El gizmo¹ dejó en la base de datos de salida los valores del campo país convertidos a formato MARC.

5.2. Conversión del formato 8/810 a MARC21

Una vez realizados todos los cambios previos y eliminados aquellos registros que no estaban en condiciones de ser transformados, se ejecutó el procedimiento de transformación. Al ejecutar este proceso utilizando CISIS (específicamente MX), con el parámetro PROC² se crea una base en formato MARC21 a partir de la base en formato BIBLO.

5.3. Eliminación de autores duplicados

En el formato MARC, los autores citados como parte de una nota de contenido se ingresan también como entrada secundaria de autor.

Un mismo autor es citado en varias partes de la nota de contenido, por ejemplo, como autor de diferentes capítulos de una misma obra, por lo que al momento de realizarse la

¹ Proceso que cambia los valores existentes por otros definidos en una tabla de conversión.

² Procedimiento que permite modificar el contenido y la estructura de los registros de una base Isis.

conversión, estos autores se agregan en el bloque 700 como autores secundarios y aparecen, por lo tanto, en forma duplicada, tantas veces como aparezcan en la nota de contenido.

Se desarrolló una script de WXIS que permite eliminarlos.

6. Utilitarios desarrollados

PROC (toMARC21.prc)

Procedimiento en lenguaje de formateo CISIS para la transformación de registros BIBLO a registros MARC21.

Es el corazón del proceso de conversión. El código está documentado identificando cada campo a convertir.

Ejemplo de llamada al proc:
mx dbname proc=@toMarc21.prc create=dbMarc —all now

AUTORES DUPLICADOS (duplicados.xis)

Este desarrollo identifica los registros que tienen esta situación, y elimina los valores duplicados del campo.

duplicados.xis es un script wxis que recibe como parámetros la base de datos, el tag del campo del cual se van a eliminar las ocurrencias duplicadas y como opcionales a partir de qué número de registro y sobre cuántos registros se realizará la tarea.

Ejemplo de llamada al scrípt:
wxis IsisScript=duplicados.xis db=dbname tag=700 from=500 count=6000

ASIGNACION DE NIVELES (niveles.xis)

Para la asignación de niveles, fue necesario desarrollar una herramienta que catalogara los registros de acuerdo con el nivel al que pertenecen e informara la cantidad total de registros de cada nivel.

Se realizó una aplicación de orden general, que permitiera, entre otras cosas, obtener los números de los registros que cumplen determinada condición o sólo el total, y cuáles son los registros que no la cumplen.

La especificación final es una aplicación que recibe como parámetros el nombre de la base de datos y una condición determinada en formato Isis y devuelve, de acuerdo con una selección previa, la cantidad total de registros dentro del rango que cumplen con la condición o, en caso contrario, los números de esos registros. Como parámetro opcional, se puede especificar, en el caso de la salida por MFNs, imprimir además otros datos mediante un formato y también invertir la condición, de manera tal que la aplicación devuelva los registros que no la cumplen.

Ejemplo de llamada al script:

wxis IsisScript=niveles.xis db=dbname

condicion=nivell.txt salida=MFNS

Cabe aclarar que, por una cuestión de comodidad, la condición se especifica en un archivo de texto externo.

7. Consideraciones finales

En esta primera etapa del proyecto se pudieron alcanzar todos los objetivos propuestos y los resultados obtenidos fueron satisfactorios.

El catálogo colectivo de libros de la Universidad Nacional de La Plata se convirtió al formato MARC21 y, como productos de la conversión, se obtuvo el detalle de la metodología para futuras conversiones, un procedimiento específico y un conjunto de aplicaciones.

Adicionalmente, se obtuvieron subproductos para la limpieza de registros, el establecimiento de niveles y la eliminación de ocurrencias duplicadas de un campo, entre otros.

En una etapa posterior, se propone una metodología de actualización de los datos ingresados por cada biblioteca y el desarrollo de una interfaz de salida en lenguaje XML.