

Migración de Winisis a Koha

Experiencia de la Biblioteca de la Facultad de Ingeniería, UBA

Por Claudia Casarín, Marcos Benincasa y Pablo Bianchi¹

Resumen

En el presente artículo se busca fundamentar la elección del formato MARC 21 y la implementación del Sistema Integrado de Gestión de Bibliotecas Koha, como herramientas de trabajo producto de la migración de registros bibliográficos en formato BIBUN alojados en CDS/Winisis, al tiempo que se describe el proceso de migración completo como mecanismo de extracción de datos de diversas fuentes, transformación y carga de los mismos en el nuevo sistema. A su vez se describen los recursos necesarios para la migración y los trabajos posteriores recomendados, como los próximos pasos a seguir en la implementación del sistema.

Introducción. Estado inicial

Producto de la evaluación integral de la Biblioteca Ing. E. Butty, de la Facultad de Ingeniería, Universidad de Buenos Aires realizada por Romanos de Tiratel, S. y Colombo F. (2004) donde se priorizó la calidad de las colecciones, los servicios y accesos a la misma, entre otros factores críticos de mejora, surgió paralelamente el proyecto de migración de datos y tecnologías existentes, que por encontrarse en alto grado de obsolescencia, hizo imperiosa la necesidad de planificar una alternativa que ofreciera soluciones integrales al registro y descripción de la *core collection* (al menos en una etapa inicial) facilitando su posterior mantenimiento y gestión, acoplando las funciones actuales y futuras de la biblioteca a través de un Sistema Integrado de Gestión.

Sobre la base de un control de inventario exhaustivo y expurgo de la colección principal, que se llevó a cabo los primeros años, pensamos tiempo después en dar un salto de calidad en la descripción de contenidos, con el propósito estratégico de integrar el acervo con otros catálogos colectivos a futuro, que no se limitaran a la participación local, sino también a la cooperación e integración regional e internacional, mejorando la accesibilidad, visibilidad y auspiciando el trabajo colaborativo cada vez más necesario en un entorno global. Así, los registros actuales, estructurados bajo el formato BIBUN por Resolución del Consejo Superior de la Universidad de Buenos Aires (1993), fueron transcritos al

¹ Dir. Biblioteca Central Facultad de Ingeniería; Responsable de Procesos e Informático de biblioteca respectivamente.

formato estándar internacional MARC 21, para luego ser integrados a registros de nuevos materiales y las colecciones antiguas existentes, cuyo registro y descripción aún se conservan en los ficheros manuales.

¿Por qué MARC 21?

Formato Internacional más utilizado

El *MAchine Readable Cataloging*, nacido en los EEUU, se ha expandido mundialmente a lo largo de más de cuatro décadas. Fue en 1967 cuando se lanzó por primera vez MARC, conocido como MARC II. Durante los años siguientes, el interés internacional en el sistema fue creciendo con fuerza lo que condujo al interés por crear una norma, o estándar sólido. El *American National Standards Institute* creó en 1971 MARC II como Standard o Norma ANSI Z39.2, que fue seguida de inmediato por la *International Standards Organization*, quien aprobó la Norma ISO 2709 en 1973. En aquel momento no existía una idea clara del tipo de intercambio internacional al que la existencia de estas normas podría dar lugar y parece razonable suponer que, debido a ello, muchos países intentaron generar sus propios formatos. La creación, desarrollo y mantenimiento del formato, por parte de la Library of Congress en cooperación con bibliotecas de Canadá y luego con la integración de bibliotecas europeas y de Australia, dió origen a distintas versiones de MARC, con adaptaciones a las necesidades locales, como IBERMARC, UNIMARC, USMARC o el mismo CAN/MARC usado hasta 1998 por Canadá². Todas éstas iniciativas confluyeron en el estándar unificado MARC 21. Si bien la finalidad de este trabajo no es ahondar sobre el estándar, vale recordar que éste tiene en realidad cinco expresiones o variantes, desarrollados para fines específicos, a saber:

- **Formato para datos bibliográficos:** codifica los datos para describir, recuperar y controlar los diferentes tipos de materiales bibliográficos; por ejemplo: libros, publicaciones seriadas, recursos electrónicos, mapas, música, materiales visuales y materiales mixtos.
- **Formato para datos de autoridad:** codifica la información contenida en registros de autoridad de nombres, materias y series.

² Véase para mayor información el sitio dedicado al estándar por los canadienses en <http://www.marc21.ca/>

- **Formato para datos de existencias:** codifica información específica, como la signatura topográfica, cantidad de ejemplares y/o volúmenes, instituciones que posee un ítem, etc.
- **Formato para datos de clasificación y el formato para datos de la comunidad.**

En el presente documento se hará referencia en general al *formato para datos bibliográficos*, con algunas menciones al de *autoridad*, siendo estos dos los más utilizados e implementados en las bibliotecas.

Soporte de la Library of Congress. LOC

Los formatos MARC 21 y su documentación se han ido transformando junto con la tecnología automatizada para bibliotecas, en la medida que han ido cambiando las necesidades de la catalogación y de las mismas bibliotecas. Como se dijo, la Biblioteca del Congreso de los EEUU desarrolló originalmente el Formato MARC 21, además de mantener y publicar su documentación, aunque esta no realiza cambios o revisiones en forma unilateral. Todas las actualizaciones y modificaciones al estándar, por pequeñas que fueran, son sometidas a discusión por grupos de trabajo de diferentes instituciones. Existen principalmente dos grupos, el MARBI (Comité para la Información Bibliográfica Legible por Máquina) y el Comité Consultivo MARC, quienes son responsables de analizar y revisar la documentación del Formato MARC 21. El Comité Consultivo MARC está compuesto por representantes de las bibliotecas nacionales, los organismos bibliográficos, grupos de distribuidores, y otras asociaciones bibliotecarias y académicas. El MARBI pertenece a la Asociación Norteamericana de Bibliotecas (ALA), compuesta por tres representantes de cada una de las divisiones funcionales de la ALA: ALCTS (función de servicios técnicos); LITA (automatización); y RUSA (referencia). La ALA procura garantizar que todo tipo de experiencia y pericia estén representadas en MARBI. MARBI lleva a cabo reuniones en conjunto con el Comité Consultivo MARC en conferencias anuales y semianuales de la ALA. La actividad primordial del Comité consiste en la revisión de los documentos en discusión y las propuestas expuestas por, o a través de la Biblioteca del Congreso, que tengan como propósito efectuar cambios o adiciones a los formatos MARC, o bien, sobre el desarrollo de nuevos soportes del Formato MARC en áreas emergentes.

¿Por qué MARC 21 en la Biblioteca de Ingeniería de la UBA?

Sin duda, abordar la tarea de migración de los datos de un sistema a otro, cambiando la concepción de trabajo en un proceso crítico como lo es la catalogación, fue una decisión importante que marcará el rumbo de la biblioteca en los años subsiguientes y que implica una evaluación profunda de los costos y beneficios asociados al cambio.

Costo-beneficio

En este aspecto, el costo implica utilizar recursos (detallados más adelante) humanos y técnicos, además del recurso más costoso: tiempo. Desde esta perspectiva, hemos tomado prestado de la economía algunos conceptos como el de costo de oportunidad, entendiendo a éste como la mejor opción que dejamos de lado al tomar una decisión. En virtud de ello pensamos que la elección de MARC ha sido una decisión postergada no sólo por nuestra institución durante varios años, siendo ahora la alternativa a dejar de lado, el formato BIBUN, cuya estructura de descripción de datos sin mantenimiento ni actualizaciones, nos ata al uso exclusivamente local aislándonos del mundo y del trabajo colaborativo. En otras palabras, nuestro costo de oportunidad hoy es relativamente bajo, si lo comparamos con el costo pagado desde que en 1984 se adoptó el Formato FOCAD en nuestro país y en 1993, se adopta su hermano menor, el formato BIBUN, por Resolución del Consejo Superior de la Universidad de Buenos Aires, para todas sus bibliotecas y centros de investigación. En el trabajo de Herrera (2005) se expone claramente este aspecto en la problemática de los catálogos a nivel nacional, llegando a la conclusión de que en un escenario de recursos escasos como el de las bibliotecas de nuestro país, adoptar el formato MARC en 1980 tenía costos variables y fijos muy similares, y la decisión de no hacerlo, costos de oportunidad muy altos. Herrera dice *“En 1980 salvo MicroCDS/Isis, no existía una herramienta adecuada para catalogar en MARC. Debido a que los recursos humanos formados en sistemas y programación no interactuaron (y aún se mantienen distantes) en los ámbitos bibliotecarios, quizá se puede adjudicar a esta situación, la falta de desarrollo de los sistemas bibliotecarios y, en definitiva, la adopción de MARC”* (2005, p.110)

En este contexto, pensamos que la adopción de MARC 21 no debe tomarse como una implementación tardía. Siguiendo con la corriente de pensamiento económico ortodoxo, lo ya hecho debe tomarse simplemente como un costo hundido; esto es, un costo en el que se ha incurrido y ya

no puede recuperarse³, y a partir de allí avanzar hacia la modernización e integración de los catálogos, lo que resulta más eficiente y efectivo desde el estándar internacional MARC 21.

Más allá de las consideraciones previas, hay razones propias de nuestra biblioteca que ameritaron tomar la decisión de migrar, o dicho de otro modo, facilitaron y disminuyeron los costos de migración. Estas razones se fundan en el hecho de que la base de datos bibliográfica migrada es relativamente pequeña y con tipología documental homogénea, lo que permite trabajar con un marco bien definido, cuyo alcance en parte está dado por las características propias del acervo de la biblioteca.

Una última consideración que auspició esta decisión tiene que ver con la articulación de los trabajos que se vienen efectuando, como se señaló en la introducción, sobre la colección principal y las necesidades inmediatas de trabajar colaborativamente en la catalogación retrospectiva, lo que ofrece un valor agregado a los trabajos realizados de reinventariado y futuros controles de inventario.

En síntesis, la elección de MARC 21 tiene razones intrínsecas al formato y también a la situación interna de la biblioteca como a la relación de ésta con su entorno, constituyéndose así en una decisión estratégica con vista a un mejor posicionamiento actual y futuro.

¿Por qué Koha?

La elección de Koha como Sistema Integrado de Gestión para la Biblioteca, se basa en numerosos aspectos de diversa índole, aunque pueden simplificarse en dos dimensiones claramente identificables. Por un lado, las características que responden a las necesidades propias de la biblioteca, y por otra parte, las cualidades propias del sistema que evidencian robustez y sostenibilidad como proyecto a largo plazo. Entre los aspectos más destacables que encontramos están los siguientes:

Filosofía open source

Cuando se hace referencia al *open source* o software de código abierto, implica algo más que una definición antitética al software propietario, denominación con la que se conoce a los sistemas y desarrollos cerrados o licenciados. El código abierto es un movimiento, una filosofía que representa formas de trabajar distintas con repercusiones sociales, tecnológicas y por supuesto, económicas. Busca dar total libertad de utilización del software a los usuarios, lo que conlleva a los desarrolladores

³ También conocido como costo caído o costo histórico, y que por supuesto para nuestro argumento no debe tomarse necesariamente como un costo monetario.

o creadores a poner en total disposición el código fuente, que además de ser copiado, es por lo general modificado y redistribuido sin restricciones.

En los entornos tradicionalmente cerrados del software privativo, el crecimiento está supeditado a factores endógenos del propio ámbito de desarrollo, en cambio en la filosofía open source, el rasgo distintivo de apertura requiere de nuevos espacios públicos virtuales y colaborativos, que dan lugar a las llamadas “comunidades”. Una comunidad, está formada por un conjunto de personas que con intereses comunes se incorporan o conforman voluntariamente un grupo con independencia de su perfil social, económico, cultural y geográfico, que inicialmente es pequeño, para luego ir incrementando sus miembros hasta crear en ocasiones una comunidad organizada. Como señalan Cobo, A., Gómez, P., Pérez, D., y Rocha, R. (2005) este fenómeno social de “voluntariado” en torno al desarrollo del software, comprende el motor social de todo el movimiento, que lógicamente es favorecido por la aparición, crecimiento y extensión de la World Wide Web, y de las redes de modo abarcativo. El uso intensivo de Internet es socio fundacional del open source, en tanto que sin una red libre sería impracticable el desarrollo de esta tendencia, y a su vez si ésta el crecimiento de la Red tampoco sería el que conocemos.

Koha, sistema integrado de gestión de bibliotecas (ILS), fue creado por *Katipo Communications* para la *Horowhenua Library Trust* en Nueva Zelanda. Nace luego de dos años de desarrollo en 1999, para entrar en funcionamiento en el año 2000. De la lengua Maorí, Koha significa “obsequio” o “donación”, y en concordancia con ello, se presenta como un sistema de código abierto cuya existencia, crecimiento y difusión ha sido posible gracias a la filosofía open source.

Vasta comunidad

Las ventajas de una gran comunidad de usuarios, desarrolladores, colaboradores, profesionales y entusiastas de todo el mundo, hacen propicio el crecimiento sostenido del sistema y su actualización permanente, generando nuevas funciones y mejoras, a la vez que dan soporte a cientos de bibliotecas, mediante subcomunidades, wikis, foros, etc.

Una comunidad amplia y creciente, no sólo ofrece soluciones en el presente, sino que otorga fiabilidad mediante ajustes colaborativos y fundamentalmente aporta sustentabilidad al proyecto. Con más de diez años de vida Koha, se vislumbra como un sistema joven con mucho potencial de

crecimiento para afrontar los desafíos futuros que las bibliotecas, los bibliotecarios y la comunidad de usuarios demanden.

Estabilidad

Al momento de escribir el presente artículo contamos con muchísimos más antecedentes de implementaciones exitosas (véase párrafo siguiente *uso en el mundo*), lo que ofrece sobradas garantías de software estable. Sin embargo, algún tiempo atrás cuando decidimos abordar la posibilidad de adoptar Koha, las experiencias eran menos conocidas en nuestro país y las evaluaciones debieron centrarse más en la herramienta en sí, que en los antecedentes cercanos. De esta manera, además de ensayar cada una de las versiones que salieron, se ha podido corroborar el sólido funcionamiento de la base de datos relacional MySQL, en conjunto con la eficiente interfaz Web, programada con el potente lenguaje Perl, que como todo sistema que corre sobre Linux, garantiza economía de recursos de hardware y seguridad informática.

Modularidad

Integrando todas las funciones de una Biblioteca, Koha destina sendos módulos que pueden implementarse gradual, parcial o totalmente. Incluye los más usuales como Circulación o gestión automatizada del préstamo, Catalogación con soporte MARC 21 bibliográfico y de autoridades, gestión de compras, PP, gestión de usuarios y credenciales, OPAC con funcionalidades de la Web 2.0 y Web Social.

Además permite configurar servicios de alerta, informes personalizables, gestor de noticias, planificador de tareas y calendario, administración de perfiles de trabajo, y numerosas herramientas complementarias, con mucha flexibilidad y customerización.

API estándar

La ventaja de incluir numerosas interfaces de aplicaciones, comúnmente utilizadas en la Web y particularmente utilizadas en las bibliotecas, como el protocolo z39.50, o SRU inclinó aún más la balanza en favor de koh. API (del inglés *Application Programming Interface*), es el conjunto de funciones y procedimientos (o métodos, en la programación orientada a objetos) que ofrece cierta

biblioteca⁴ para ser utilizado por otro software como una capa de abstracción. Koha integra mediante estas funciones con servicios de Open URL, LDAP, SIP2, Open Search de Amazon, LibraryThing, permitiendo también otros servicios como OAI/PMH, de gran utilidad en las iniciativas de Archivos Abiertos. Estas características, aseguran la portabilidad de los datos al tiempo que ofrecen un gran abanico de posibilidades para la interoperabilidad de los sistemas.

Uso en el mundo

Koha es uno de los SIGB libres más usados en el mundo, quizá el más usado. Es difícil saber con precisión la cantidad de bibliotecas que lo utilizan, pero podemos estimar un mínimo de acuerdo a quienes se registraron por voluntad propia en el wiki de Koha (<http://wiki.koha-community.org/>) y de lib-web-cats, de Library Technology (<http://www.librarytechnology.org/libwebcats/>), tendencia que por otra parte ya se vislumbraba un par de años atrás. Del primero se obtiene la siguiente tabla, que suma más de 800 bibliotecas:

África	32
Norteamérica	78
Centroamérica y caribe	4
Sudamérica	76
Asia del este	14
Sudeste asiático	4
Asia del sur	117
Sudeste asiático	334
Europa	128
Europa	51

Antes de decidir la elección de este SIGB se hizo una extensa búsqueda. Se barajaron nombres como Emilida, Openbiblio, PMB, Evergreen, entre otros, pero ya en aquel momento Koha se distinguía por sobre el resto.

Bibliotecas nacionales

Como se mencionó con anterioridad, la implementación de las Bibliotecas Nacionales de Colombia (<http://catalogornbp.bibliotecanacional.gov.co/>), Filipinas (<http://koha.nlp.gov.ph/>), Malawi (<http://www.catalogue.mw/>),

⁴ Aquí el término biblioteca refiere a las llamadas “librerías” de los diferentes lenguajes de programación, que contienen un conjunto de funciones específicas para dichos lenguajes.

Venezuela (<http://sisbiv.bnv.gob.ve/>), Armenia (<http://tert.nla.am/>), opera de alguna manera como “cita de autoridad”, garantía de fiabilidad, dando además una idea de la escalabilidad del sistema, considerando que en el presente caso el punto de partida es una biblioteca universitaria.

Aspectos de las fuentes a migrar

EL proceso de migración, se basa principalmente como se dijo antes, en la traducción de registros BIBUN a MARC 21. Sin embargo y ante las dificultades que esto supone, tal traducción no es posible sin un previo trabajo de organización de los datos, limpieza, y traducción en nueva información. La búsqueda de patrones o constantes, es parte del proceso que conlleva a una traducción más precisa y eficiente entre los distintos esquemas, y posibilita la extracción de manera adecuada. El proceso realizado, se conoce como un mecanismo de ETL, *Extracción Transformación y Carga* (en inglés *Extract, Transform and Load*), en el que es necesario además de la validación de datos, un *crosswalk schema*, o mapeo, donde se presentan las equivalencias entre los datos de origen y destino. Dicho de otro modo, el proceso de transformación requiere que se indique para nuestro caso, cuál es el equivalente de determinado campo o subcampo BIBUN en MARC 21.

Pero la complejidad no radica sólo en este aspecto, considerando que además las equivalencias entre formatos no son tan lineales como uno desearía, hay que agregar al proceso de extracción realizado, otras fuentes de datos, entre las que utilizamos MS-Excel y las Google Spreadsheets de Drive sobre las que hay registrados procesos y actividades en relación a los registros y que deseamos transferir al nuevo sistema. De esta manera el proceso completo de ETL, da como resultado un archivo que contiene un lote de registros MARC 21 auténticos, con el agregado de numerosa información local adicional, que al importar a nuestro ILS agrega valor, fundamentalmente sobre los registros de existencias (*holding data*), o como en Koha se suele denominar, registros de items.

A continuación se describen brevemente las fuentes de datos sobre las que se extrajo la información:

Bibliográfica: Sistema BD: Winisis v1.5. Formato: BIBUN. Cantidad de Registros bib.: 11.000. Incluye: Documentos impresos. Bibliográfica: monografías, series, tesis, congresos. No contempla: Documentos en otros soportes diferentes de papel. Analíticas.	Hojas de cálculo: Software: MS Excel. Estructura: Sobre textos fuera de circulación. Ejemplares en reparación. Ejemplares dados de baja. Cantidad de registros: 1500. Contempla dos tablas con la información de estructura.
--	--

Otras fuentes: Google Spreadsheets
Estructura: Lista de términos autorizados (descriptores) y datos fuente.
Cantidad de registros: 9000.
Contempla una tabla con los tesauros, términos utilizados y relaciones básicas.

Recursos necesarios

Una variedad de recursos fueron necesarios para sustentar el proyecto, sin los cuales difícilmente se podría haber avanzado. El recurso principal fue el humano, aunque sin el apoyo de las autoridades y la posibilidad de contar con las herramientas informáticas necesarias, no podríamos haber salido de una fase inicial o instancia de prueba.

Humanos

Fue imprescindible contar con una serie de características como equipo, en su conjunto. Como en todo proyecto, priorizamos la conformación de un equipo, entre bibliotecarios e informático, donde las principales habilidades o competencias del conjunto fueran:

- Manejo fluido de la PC, sistemas operativos MS-Windows y Linux.
- Conocimiento de formatos involucrados, BIBUN y MARC 21, y estándares RCA2.
- Profundo conocimiento de la base a migrar, en todos sus detalles; interpretaciones locales del formato de origen (BIBUN) y de las convenciones de carga establecidas a lo largo del tiempo.
- Afinidad con el software libre, no sólo técnica sino filosóficamente.
- Conocimiento de Koha: Los parámetros a configurar, principales y secundarios.
- Manejo de bases de datos relacionales, lenguaje Perl, HTML, expresiones regulares (*regexs*) y otras herramientas de procesamiento de texto plano.
- Lectura y escritura fluida en inglés, para leer documentación y preguntar en listas de correo, IRC, foros, etc., donde muchas veces no hay información en castellano.
- Actitud proactiva e iniciativa para la resolución de problemas emergentes.
- Trabajo en equipo: Fue esencial generar una sinergia, consecuencia de escuchar mutuamente la conveniencia o no de las decisiones, desde el aspecto bibliotecológico, documental e informático.

Pre-migración⁵

Antes de realizar la migración en sí sabíamos que debíamos analizar lo mejor posible los registros de la base a migrar. Para esto, se utilizaron una serie de comandos de consola de Linux junto con pequeños scripts (personalizados para cada caso) para obtener información relevante, que permita encontrar errores tipográficos y corroborar que las hipótesis de cada caso se cumplieran en todos los campos de la base. Algunos casos fueron:

- ☐ Conteo de ocurrencias de uso de cada campo y subcampo de toda la base.
- ☐ Listado de veces que se repitió un contenido de campo/subcampo. Esto fue particularmente útil para encontrar errores en campos en los cuales los valores posibles forman parte de un conjunto finito, como ser *código de idioma* o *código de país*.
- ☐ ISBN e ISSN válidos para todo caso.
- ☐ Búsqueda de errores triviales como presencia de subcampo en algunas ocurrencias y en otras no.

En el proceso se hicieron sucesivos reportes, se realizaron las correcciones que éste planteaba. Luego se volvía a realizar un reporte, y así, hasta no encontrar errores que los script no puedan corregir automáticamente⁶. En adición al análisis exhaustivo de la base, se debió verificar que cada registro tenga algunos datos esenciales, como el número identificador de registro bibliográfico, número identificador de ejemplar (inventario del ítem, *barcode*) y tipo de ejemplar (*item type*): El tipo de ejemplar imprescindible para luego poder establecer reglas de circulación.

Script inicial

Una serie de tareas automáticas fueron agrupadas en un script inicial. Este extrae del archivo maestro de Winisis los registros para pasarlos a TXT, luego realiza una serie de correcciones, devolviendo un reporte de lo hecho en los casos más relevantes. Dentro de las correcciones realizadas podemos nombrar el cambio de codificación de caracteres, y la eliminación de errores automáticos como ser caracteres inválidos, campos y subcampos vacíos y fines de línea espúreos. Finalmente generando un archivo ISO-2709 pero respetando la estructura de un registro MARC. Esto

⁵ Si bien la palabra *migración* no es la más indicada (quizá lo sea *traducción*) se utilizará porque es la más ampliamente difundida para denominar este proceso.

⁶ Término propio de la jerga de las ciencias de la computación, referido a algo automático, pero con un aparente elemento de magia; referido a un proceso más o menos complejo, oculto de la vista del usuario, resultando en algo que "*simplemente funciona*".

se hizo así para utilizar la infinidad de herramientas que existen partiendo de esa base, en especial las bibliotecas⁷ que ofrece Perl para la manipulación de registros, lo que permite hacer prácticamente cualquier cosa.

Migración

Tabla de migración

La herramienta elemental para la transformación de los datos, está dada por el cruce de información entre los formatos, el crosswalk o data mapping, es un trabajo exclusivo del equipo. Para esto se utilizó una tabla compartida en línea, usando una planilla de cálculo en *Google Spreadsheet*. En esta se comenzó por poner el mapeo de datos, pero dada la complejidad del trabajo esta tabla, terminaron siendo 13 tablas, alcanzando un tamaño de 119 filas y 24 columnas, lo que da idea de la dimensión del trabajo.

Como es de esperar, en muchos casos el mapeo es simple: Lo que en un caso se ingresaba en un lugar, en MARC 21 debe estar en otro. Pero muchos otros casos implican procedimientos más o menos complejos, donde el movimiento depende del estado de otros subcampos, o requieren tratamientos especiales antes de ser copiados. Para esto se utilizó la tabla principal, pero también permanente diálogo e incluso diagramas y cuadros, para evitar que se “escapen” casos posibles.

Script de migración en sí

Una vez obtenido un archivo ISO-2709 pero con la estructura de un registro MARC (de extensión normalmente .mrc) se lo procesa registro por registro secuencialmente con un software programado en Perl. Éste representa la tabla de migración hecho software, y es el que incorpora por lo tanto todos los algoritmos necesarios para traducir el registro de BIBUN a MARC 21..

Valor agregado

Además del trabajo previo de búsqueda y corrección de errores en la base, en la misma migración se realizaron procedimientos que hacen al registro más rico y completo. Dentro de estas mejoras podemos considerar:

⁷ Se utiliza el término *biblioteca* en el sentido informático del término, como un conjunto de herramientas listas para ser usadas.

- **Lista de material fuera de circulación:** Se integró la información relativa a los ejemplares en dicho estado, presente otrora en una planilla de cálculo, en la parte relativa a la información del ejemplar, dentro del registro bibliográfico. De esta forma, toda la información está en un mismo lugar y se puede administrar todo desde Koha, pudiendo por ejemplo eliminar o no su visualización del catálogo.
- **Lista de tesauros de descriptores:** La información relativa al tesauro fuente de la que se obtuvo cada descriptor se encontraba también en un listado aparte, pudiendo también incorporarse al registro
- **Detección de idioma:** Estas incluyen, a modo de ejemplo, la posibilidad de detectar el idioma de un registro a partir del título (si por alguna razón este dato no fue ingresado). Si el dato obtenido tiene asociada con un valor de confianza que supere un umbral dado, este se incorpora el registro.
- **Compresión del registro de entrada:** Una idea novedosa fue la idea de comprimir el registro a migrar y copiarlo completo en un campo local, lo cual permite ver (eventualmente tiempo después) cómo era el registro antes de ser migrado.
- **Detección y registro de errores:** En los casos en los que no se podía permitir que el software tome alguna decisión al encontrar un error, se optó por agregar un prefijo al subcampo con un código de error. Esto nos permite migrar sin tener que resolver cada error, pudiendo buscarlo y arreglarlo luego al estar utilizando Koha.

Post migración

Una vez obtenida la base migrada sabíamos que el trabajo no terminaba ahí: Debíamos comprobar si todo había surgido como esperábamos y continuar el trabajo de implementación en Koha.

Comprobación

Aunque parezca un proceso lineal, el procedimiento se repitió algunas veces, con continuas mejoras, al encontrar errores en cada etapa. Algunos estos fueron encontrados al final de la migración, la cual fue verificada esencialmente de dos formas:

- **Manualmente:** Tomando una muestra de registros representativos sobre los que se revisa campo por campo y especialmente en los casos “difíciles” si habían sido migrados correctamente
- **Automáticamente:** Utilizando validadores de registro MARC, los cuales revisan toda la base y devuelven reportes sobre errores de distinto tipo, como ser subcampos o indicadores inválidos. Como en los demás casos, este análisis se hizo de forma iterativa, corrigiendo sucesivas veces los errores que presentaba.

Autoridades

En un principio las autoridades no se consideraron, en parte porque no tenemos registros genuinos para migrar y Koha no presentaba las facilidades para generarlas. Actualmente y concluido el proceso de migración se incluyó la generación automática de autoridades con Koha, a partir de una herramienta (Script en Perl) que trae el sistema para ejecutar por consola. Con algunos ajustes, Koha permite a partir de registros bibliográficos, generar cuasi-registros de autoridad, con campos elementales generados a partir de la base de origen, sobre los que post migración hay que completar. Esto no es propiamente un registro de autoridad completo, pero sí uno con elementos suficientes para cumplir con la función de referencia entre tablas, sobre todo para los campos 1xx, 6xx y 7xx. De este modo es posible el control de errores tipográficos no detectados anteriormente, y la consolidación de los datos, además de conservar información valiosa que ya se generaba con los términos utilizados en la indización.

Conclusiones. Aspectos del procedimiento

Como habíamos previsto, encontramos buena cantidad de problemas en todo el proceso, uno de ellos y quizá el más significativo fue el de exportar la base de datos bibliográfica de Winisis de forma “limpia”, sin que se filtren errores. Se dedicó mucho tiempo a analizar numerosas herramientas y metodologías de migración, encontrando fuertes debilidades a muchas de ellas, hasta dar con la adecuada. Algunos métodos de exportación generaban registros defectuosos, mientras que algunas metodologías de migración no ofrecían la versatilidad, velocidad o estabilidad necesaria para realizar un trabajo con la calidad deseada. Sin embargo aplicando soluciones alternativas, logramos sintetizar

y automatizar gran parte del proceso de migración y documentarlo para posteriores usos o mejoras, siguiendo la filosofía del software libre, para las bibliotecas deseen abordar trabajos similares.

Aporte a la comunidad

El trabajo paralelo de aprendizaje de Koha tuvo algunos frutos para con la comunidad. Creamos la comunidad *Koha-es* para hispanohablantes, que actualmente ostenta más de 400 miembros. Asimismo, actualmente la documentación de Koha la hacemos utilizando la tecnología DocBook/XML sugerida, y numerosos bugs fueron reportados y documentados con precisión para aportar aunque sea modestamente al desarrollo de este software gratuito y libre.

Próximos pasos

Aunque concluida la migración exitosamente, el camino tiene aún mucho por delante. Algunas de las tareas pendientes son:

- Implementación de módulos de Circulación, y resto del sistema: Koha tiene una enorme cantidad de funciones y la capacitación y uso de los módulos de Circulación y Reportes está entre las próximas prioridades. Para el uso del módulo de Circulación se utilizará códigos de barra en el material y tecnología RFID (similar a la SUBE) para las credenciales de usuario.
- Koha permite trabajar con un directorio centralizado de usuarios (LDAP), esto permitirá la interoperabilidad con sistemas diversos como son el sistema de gestión de alumnos SIU-Guaraní y la plataforma de e-learning libre Moodle. Incluso integración con Greenstone.

Referencias

- Automagically (s/f). En Dictionary.com. Recuperado, Septiembre 2013 de <http://dictionary.reference.com/>
- Bibliotecas usando Koha. En Library Technology (s/f). Recuperado, Septiembre de 2013 de <http://www.librarytechnology.org/map.pl?ILS=Koh>
- Brown, M. C., & Brown, Martin C. (2001). Perl : manual de referencia / M.C. Brown. Madrid, España : McGraw-Hill.
- Cobo, A., Gómez, P., Pérez, D., Rocha, R., (2005). PHP y MySQL: Tecnologías para el desarrollo de aplicaciones web. Madrid: Ediciones Díaz de Santos
- Cobo, A., Gómez, P., Pérez, D., Rocha, R., & e-libro, Corp. (2005). PHP y MySQL: Tecnologías para el desarrollo de aplicaciones web. Madrid: Ediciones Díaz de Santos.

- Dyché, J. (2000). e-Data: Turning data into information with data warehousing. Reading, Mass: Addison-Wesley
- Fernández, L. I. (2008). Nuevos aprendizajes de programación Bash en Linux. Almería: Procompal.
- Furrie, B., Rojas, L. O. G., García, B. A., Follett Software Company., & Library of Congress. (2002). Conociendo MARC bibliográfico: Catalogación legible por máquina. Bogotá, Colombia: Rojas Eberhard Editores.
- Herrera, L. A.. (2005). MARC 21 y conceptos de economía en las decisiones bibliotecarias: MARC 21 and economic concepts in the library-related decisions. Inf. cult. soc. 101-114. Recuperado, Septiembre 2013 de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17402005000200006&lng=es&nrm=iso
- Library and Archives Canada. (1997). MARC standards. Ottawa: Library and Archives Canada. Recuperado, Septiembre 2013 de <http://www.marc21.ca/index-e.html>
- Library of Congress. (1990). MARC standards. [Washington, D.C.]: Library of Congress, Network Development and MARC Standards Office. Recuperado, Septiembre 2013 de <http://purl.fdlp.gov/GPO/gpo13828>
- Mankiw, N. G. (2009). Principios de economía. Australia: Thomson.
- Romanos, S., Colombo, F. (2005). Informe diagnóstico preliminar de la Biblioteca "Ingeniero Enrique Butty" Buenos Aires: Facultad de Ingeniería. Universidad de Buenos Aires.
- Universidad de Buenos Aires (1993). Formato BIBUN. Buenos Aires: SISBI
- Universidad de Buenos Aires. Consejo Superior (1993). Resolución Nro. 3911/93.