



Temas de Biblioteconomía

**La edición-e y su impacto en las BU
Acceso y gestión de los RE:
los metadatos
y los lenguajes de marcado**

Autor: César Martín Gavilán

Fecha: 02/02/09

Edición electrónica: impacto en las bibliotecas universitarias

Es fácil entender porque al final de la era de la imprenta las bibliotecas universitarias, y de hecho todas las bibliotecas, se encuentran algo confundidas: la “tecnología” sobre la que se ha construido su misión durante los pasados quinientos años está siendo usurpada. El desarrollo de la imprenta en el siglo XV y su industrialización el siglo XIX hicieron de las bibliotecas lo que ahora son ... O, para ser más precisos, lo que eran hasta 1993, cuando empezó la web.

La mayor parte de lo que los bibliotecarios saben sobre la organización de la información no es más que una depuración y mejora del trabajo de Melvil Dewey, Charles Ammi Cutter y otros bibliotecarios pioneros del siglo XIX. Tal y como Google demuestra cada día de manera convincente, el control de autoridades y la clasificación ya no son ni de lejos las únicas o las mejores respuestas a las necesidades de información.

La edición electrónica no sólo incorpora un nuevo soporte documental a las colecciones: su marcado carácter disruptivo en el modelo de flujos de creación y acceso a la información pone en tela de juicio el futuro de la biblioteca. En este sentido, las bibliotecas universitarias deben encontrar y articular su papel en la ecología presente y futura de la información. Si no pueden o no lo hacen, los gestores de las universidades invertirán en otras prioridades y la biblioteca se atrofiará de manera lenta pero segura, convirtiéndose en un museo del libro de poco uso.

Migración de las colecciones impresas a las electrónicas

Hay tres tipos de documentos a tener en cuenta cuando consideramos la migración del formato impreso al electrónico: obras de referencia, revistas y libros. La migración está casi completa en los dos primeros casos y acaba de empezar en el tercero.

La conversión de índices y abstracts al formato electrónico empezó a mediados de los años 80 con la aparición de los CD-Roms y se completó a mediados de los años 90 cuando aparecieron las versiones web de estos productos. Las enciclopedias pasaron al formato electrónico del mismo modo y en la misma época. A continuación siguieron las obras de referencia de derecho y economía, cuyas versiones impresas requerían una árdua tarea de actualización. Estos productos han sido unos sustitutos claramente superiores a sus predecesores impresos y, en muchos casos, los productos impresos han sido abandonados. Los grandes paquetes de referencia, como el Biography Resource Center o el Literature Resource Center de Gale, se estuvieron disponibles a través de la Web a finales de los 90 y a comienzos de 2000

aparecieron un gran variedad de materiales de referencia especializados. No está muy claro que estos productos electrónicos hayan sido tratados como auténticas alternativas de sus equivalentes impresos, siendo posible todavía encontrar obras claramente obsoletas en las secciones de referencia con un uso prácticamente inexistente.

Empezando con Lexis/Nexis y el InfoTrac de IAC, las revistas a texto completo empezaron a encontrarse disponibles electrónicamente. A mediados de los 90 este contenido se trasladó a la web con índices y resúmenes. Estas revistas crecieron a medida que otros agregadores se incorporaban al mercado y muchos editores sacaban versiones web de sus títulos bien como productos gratuitos o como suplemento de la versión impresa. JSTOR añadió fondos retrospectivos a esta mezcla. A comienzos de los 2000 casi todas las publicaciones importantes de la mayoría de disciplinas se encontraban disponibles electrónicamente. En una gran parte de los casos las bibliotecas no contemplaban los productos de los agregadores como sustitutos de la versión impresa aduciendo que el cambio constante del contenido de estas colecciones las hacía poco fiables y, por lo tanto, no se podían aceptar como una alternativa. Esto ha llevado a mantener durante años la suscripción simultánea a la versión impresa. El reemplazo de la versión impresa por la electrónica se ha aceptado mejor en el caso de títulos puntuales, y especialmente cuando los presupuestos son limitados o los editores empiezan a penalizar el modelo de precio impreso + electrónico.

Aunque las bibliotecas se han movido con cautela, los usuarios se ha arriesgado más en este terreno. En la mayoría de bibliotecas el uso de las obras de referencia y revistas impresas se ha reducido de forma rápida y constante. Esto se puede saber siguiendo las estadísticas de fotocopias y la recolocación de las publicaciones en los estantes. También es probable que la facilidad de utilización y la potencia de la Web, especialmente cuando las colecciones a texto completo forman parte del producto o cuando se utilizan servicios de enlace, como SFX, aumentan el uso de las publicaciones periódicas. Debería haber ahorros importantes en la migración del soporte impreso al electrónico de un 20% a un 60% tanto en el procesamiento como en la gestión de las colecciones.

Los libros electrónicos académicos empezaron a aparecer a finales de los años 90, cuando NetLibrary introdujo las primeras colecciones. Tras un comienzo accidentado, Netlibrary y otros proveedores de libros electrónicos se establecieron en el mercado. Readex y otras han incorporado grandes colecciones retrospectivas de libros electrónicos. El Proyecto Gutenberg ha ido digitalizando y haciendo accesible de forma gratuita títulos libres de derechos desde comienzos de los 70, al igual que en España la Biblioteca Virtual Miguel de Cervantes, creada en 1999, aunque estos esfuerzos han tenido poca repercusión en las adquisiciones de las bibliotecas. A finales del 2004 Google

convulsionó el mundo bibliotecario al anunciar su asociación con cinco grandes bibliotecas de investigación (Harvard University, New York Public Library, Stanford University, University of Michigan y University of Oxford) a través del Google Print Library Project, que fue rebautizado como Google Books Library Project al año siguiente. El proyecto pretende digitalizar y hacer accesible de forma electrónica millones de volúmenes grandes bibliotecas (actualmente hay 28 bibliotecas afiliadas al proyecto, 7 de ellas de fuera de los EEUU, entre las que se encuentran 2 españolas: la Biblioteca de la Universidad Complutense de Madrid y la Biblioteca de Catalunya). Poco después Internet Archive lanzó un proyecto competidor, la Open Content Alliance, centrado en títulos libres de derechos de autor. Hasta la fecha se ha hablado mucho de los libros electrónicos, pero hay poco consenso a cerca de que los libros electrónicos sean una alternativa adecuada para los impresos. Como consecuencia ha habido pocos cambios en la práctica bibliotecaria. De todas maneras, no parece descabellado apuntar que esto podría cambiar en un futuro cercano según los datos que ahora se manejan, una vez se han dado por superados ciertos problemas tecnológicos, ergonómicos y comerciales que afectaban al libro electrónico.

Lo que no está tan claro es hasta qué punto las bibliotecas han abandonado las versiones impresas y han reinvertido aquellos recursos en otras áreas. Hay un ahorro evidente si se procesan menos ejemplares en papel, si disminuye el número de volúmenes a guardar y reencuadernar. Pero todo apunta a que pocas bibliotecas tienen estrategias claras para gestionar esta migración y cómo y cuando reaprovecharan esos recursos. Y tampoco hay demasiadas bibliotecas que tengan prisa por moverse decididamente en esta dirección.

Una fuente adicional de ahorro podría estar en la selección de materiales. Los recursos electrónicos se suelen encontrar a menudo agrupados en paquetes, a diferencia de sus equivalentes impresos. Por un lado este agrupamiento puede ser una desventaja, sobre todo cuando el paquete ha sido realizado por grandes editores comerciales y agregadores de revistas según sus propios intereses, pero por otro el ahorro de tiempo en la selección y gestión de la adquisición puede llegar a ser considerable. Por ejemplo, suscribirse a Ebrary da acceso a decenas de miles de libros electrónicos a partir de una única decisión, a diferencia de las muchas horas que un bibliotecario debería dedicar en este proceso de selección. Modelos de adquisición dirigida por el usuario, como PDA (Patron Driven Acquisition, o Adquisición impulsada por el usuario) de NetLibrary, traslada la tarea de selección a los usuarios de la biblioteca, y al menos en algunos casos puede ser notablemente más efectiva que la selección tradicional. De todas maneras, como este tipo de modelo de selección amenaza con desplazar el papel tradicional del bibliotecario, es probable que haya resistencia al cambio.

Las bibliotecas necesitan pasar con determinación de los documentos impresos a los electrónicos y, al mismo tiempo, trabajar con diligencia para recuperar el ahorro que hace posible esta migración. Si no se da ese segundo paso, se perderán gran parte de los beneficios que aporta la edición electrónica. Llevar a término las dos tareas requiere disciplina, y una evaluación rigurosa y continúa de la práctica.

Retirada de las colecciones impresas patrimoniales

Desde el momento que las bibliotecas cambian la colección de formado papel a electrónico, las colecciones patrimoniales tienen un objetivo diferente y es necesario gestionarlas de otro modo. Aunque algunos documentos en papel sigan siendo importantes, en particular las monografías en los campos de las humanidades y las ciencias sociales, en general los documentos impresos dejarán de ser la parte principal de las colecciones de trabajo. Este cambio puede producir un rendimiento significativo, especialmente en el uso del espacio. Además, serán necesarias nuevas estrategias y modelos de financiación para la preservación a largo plazo y el acceso a este material. Si no se desarrollan estrategias claras, nuestra capacidad para reutilizar el espacio será limitada. No se puede olvidar que más pronto que tarde los gestores académicos se darán cuenta que resulta inaceptable utilizar los mejores lugares de una universidad para almacenar libros y revistas poco utilizados, y actuarán en consecuencia.

Por suerte, la estructura subyacente sobre la que esta estrategia se puede construir está bien establecida. Grandes bibliotecas de investigación y algunos consorcios han construido almacenes externos de alta densidad y tienen experiencia en la gestión de las colecciones de estos espacios y en facilitar el acceso. Un ejemplo cercano es el GEPA (Garantia d'Espai per a la Perpetuïtat de l'Accés) del CBUC. Podrían existir propuestas de gestión regional o incluso nacional, aunque ninguna opción parecerá buena a los bibliotecarios de mentalidad tradicional que no quieran enemistarse con el profesorado que valora sobre todo la cercanía de "sus" libros. Una excepción sencilla a esto podría ser la colección de publicaciones de JSTOR. Muchas bibliotecas podrían descartar estos volúmenes. Después de todo, este fue el objetivo de JSTOR desde el comienzo. En el supuesto de que la comunidad bibliotecaria pudiera establecer estrategias nacionales o regionales de almacenamiento y preservación a largo plazo de las colecciones en papel, las bibliotecas podrían retirar o descartar tranquilamente sus colecciones patrimoniales en papel, especialmente las que ya se encuentran disponibles en soporte digital, y con el tiempo reorientar el uso del espacio.

Reorganización de la biblioteca como espacio de aprendizaje informal

Hasta hace muy poco el espacio de estudio en la mayor parte de las bibliotecas era una mezcla de cabinas, mesas y algún puesto de lectura más cómodo para

usuarios individuales. A mitad de los años 90 se desplegó un número elevado de ordenadores, aunque muy a menudo estos ordenadores estaban configurados y gestionados como si fueran ordenadores públicos que casualmente habían sido instalados en la biblioteca. En los últimos años se ha realizado un importante esfuerzo en muchas bibliotecas para repensar y reorganizar los espacios de estudio y crear lo que se denomina “Information or Academic Commons” (Bienes comunes informacionales o académicos). Los primeros proyectos de “commons” generalmente se centraban en traer tecnología a las bibliotecas y a menudo colaboraban los servicios de informática de la universidad. Proyectos más recientes se han centrado en desarrollar espacios que faciliten el trabajo en grupo e implican la asociación con otros servicios y colectivos universitarios implicados en lograr el éxito académico del estudiante (Learning centres o CRAIS). Los centros multimedia y las salas de ensayo no son inusuales, como tampoco lo es la colaboración con servicios que proporcionan a los profesores apoyo tecnológico y pedagógico. Siguiendo el modelo de la cadena americana Barnes and Noble, o de las europeas FNAC o Feltrinelli, los bares se están convirtiendo en puntos de acceso a la información habituales. El objetivo es crear espacios confortables, animados y activos dónde los estudiantes puedan interactuar entre ellos, con información y tecnología, y dónde se pueda encontrar ayuda para la utilización de los recursos de la biblioteca y de la tecnología. Cada vez más estos espacios se conciben tanto como lugares donde crear conocimiento como lugares donde acceder al conocimiento. Al mismo tiempo, las bibliotecas se enfrentan al desorden provocado por los teléfonos móviles y los ordenadores portátiles, que se pueden utilizar a cualquier lugar de los edificios gracias a las redes inalámbricas. Lo que está claro es que los espacios abiertos y relativamente homogéneos que habían sido el modelo de construcción de las bibliotecas desde los años 50 ya no funciona. Lo que se necesita es una nueva mezcla de diferentes tipos de espacios y de entornos de trabajo capaces de acoger usos y ambientes diferentes. Será necesario compartir el espacio de la biblioteca con otros socios y es probable que la distinción entre el espacio de la biblioteca y el de otros espacios más informales de la universidad se difumine. Con la retirada de colecciones en papel será posible reorganizar el espacio.

Reorientación de las herramientas, recursos y conocimientos bibliotecarios

Cuando pensamos en el futuro de los servicios bibliotecarios, es útil tener en cuenta las College Students’ Perceptions of Libraries and Information Resources de la OCLC, que afirma:

- El uso de los recursos generales de Internet está muy extendido entre los estudiantes universitarios.
- Utilizan habitualmente motores de búsqueda, correo electrónico y mensajería instantánea para conseguir e intercambiar información.

- La biblioteca no es ni la primera ni la única opción para estos buscadores de información.
- Los motores de busca son el lugar favorito para empezar una búsqueda y los encuestados indican que Google es el que han utilizado más recientemente para empezar las búsquedas.
- Entre los estudiantes que han empezado a buscar utilizando un motor de búsqueda, el 48 % acabaron en la Web de una biblioteca.
- Un 41% siguieron utilizando la Web de la biblioteca, pero sólo un 10% afirmaron que el lugar Web de la biblioteca satisfacía sus necesidades de información.
- Un 27% dijeron que también tuvieron que utilizar otros recursos.

El resultado de este estudio confirma que las bibliotecas no están consideradas como la opción prioritaria para acceder a los recursos electrónicos, incluso entre los estudiantes universitarios que tienen un alto grado de conocimiento de estos recursos. Los estudiantes viven en la Web, empiezan y a menudo acaban sus investigaciones con Google, y mayoritariamente utilizan la biblioteca como un lugar para estudiar. Es una cruda realidad aceptada por los bibliotecarios, pero a todos nos gusta pensar que los estudiantes universitarios son diferentes. Esto quizás sea verdad ahora, pero puede cambiar rápidamente.

Lo más importante de este estudio y otros similares es que muestran con claridad que si la biblioteca escoge permanecer aislada, será dejada de lado. Las fuentes de información alternativas quizás no son tan exhaustivas o autoritativas como las que se encuentran o están suscritas por la biblioteca, es cierto, pero son suficientemente buenas y se adaptan fácilmente y sin costuras a los intereses de los estudiantes y cada vez más a los del profesorado. El principal espacio digital dónde los estudiantes harán su trabajo académico será el sistema de gestión de cursos del campus o “aula virtual”. Esto puede funcionar en el caso de los sistemas centrados en la institución, pero dada la importancia de la colaboración interinstitucional entre los investigadores, es posible que los sistemas disciplinarios nacionales o internacionales sean más efectivos. Tanto los estudiantes como los profesores utilizarán los motores de busca generales como sus principales instrumentos de descubrimiento. Los recursos y herramientas bibliotecarias, el conocimiento ha de estar allí dónde se encuentran los usuarios.

Una idea debería quedar clara: si no es posible acceder a la biblioteca a través de Google, el usuario no llegará. Las bibliotecas han de aplicar estrategias de enlace para que esto sea fácil y simple. Un ejemplo a seguir podría ser el de WorldCat de OCLC, cuyos registros aparecen indexados en los grandes motores de búsqueda y son recuperables. También debería ser transparente. Sería necesario desplegar una doble estrategia:

1. Las bibliotecas han de introducir sus recursos y conocimientos técnicos dentro de los recursos y herramientas que los estudiantes y profesores utilizan en su vida diaria, evitando “inventar” nuevos sistemas salvo que no hubiera otra alternativa. El Open WorldCat de OCLC, con el enlace al proyecto Google Book es un buen ejemplo de cuál es la opción adecuada. Una completa integración de los recursos bibliotecarios en los sistemas de gestión de cursos (los sistemas de “aula virtual”) debería ser una prioridad para la biblioteca.

2. Las bibliotecas deberán reorientar la atención presencial para especializarla y dar respuesta a problemas más complejos y difíciles. El objetivo no sería sustituir la atención presencial y las relaciones que se establecen a través de ella, sino encontrar la forma de mejorarla y construir relaciones más fuertes. Los mostradores de información tradicionales, aunque se hayan ampliado con chat y correo electrónico, probablemente no son la mejor estrategia. Es posible que sea necesario elaborar una mezcla de tutoriales, instrumentos de aprendizaje y clases presenciales.

Al mismo tiempo que disminuye la importancia de los métodos basados en las herramientas de muchas de las actividades bibliotecarias, han entrado al terreno de las bibliotecas aspectos como la evaluación de la autoridad de los recursos, la integridad académica y la propiedad intelectual. Hay oportunidades claras de situar bibliotecarios en los centros para la enseñanza y el aprendizaje (CRAI) e involucrarlos formalmente en los programas de investigación de los estudiantes. Haría falta plantearse la posibilidad de tener blogs dirigidos a cursos individuales o departamentos, y que la biblioteca tuviera presencia en espacios sociales como MySpace o Facebook.

De la adquisición de la colección a la gestión de contenidos

La transición de la información en formato impreso al formato electrónico está clara y el impacto es evidente. Pero hay una segunda transición, el impacto de la cual no se ha reconocido del todo: la transición del contenido adquirido al contenido en acceso abierto. Esta segunda transición tendrá más influencia si cabe que la primera en lo que serán y harán las bibliotecas en el futuro, pero todavía no ha sido considerada ni analizada en profundidad. El número de publicaciones de acceso abierto ha crecido de forma constante. También hay un conjunto creciente de datos que indican que los autores incrementan el impacto de sus artículos cuando estos se encuentran disponibles en una publicación de acceso abierto. También se ha argumentado que el acceso abierto acelera la velocidad a la que avanza la ciencia. Pero entre la disputa de los defensores de el acceso abierto y la retórica a menudo engañosa de los editores comerciales que tratan de proteger su mercado, es fácil olvidarse de la transformación fundamental que está teniendo lugar.

Lo que es más importante aquí es que el avance del movimiento de la literatura de acceso abierto (digital, en línea, gratuita y libre de la mayoría de restricciones de copyright y licencias) repercute directamente en muchas cosas que hacen las bibliotecas, incluyendo muchos proyectos de biblioteca digital. Desde el punto de vista de los estudiantes y el profesorado el crecimiento del acceso abierto significa que hay más material científico de calidad accesible de forma gratuita (y que se encuentra fácilmente con Google y Google Scholar). Esto los libera de depender de la biblioteca universitaria como la única fuente de materiales científicos. Con el tiempo esto querrá decir que la colección de material adquirido de la biblioteca, tanto en formato electrónico como impreso, será menos importante o por lo menos su valor más relativo. La buena noticia es que cuando esto ocurra las bibliotecas tendrán que comprar menos. Y todavía mejor será que esto pasará en primer lugar en el ámbito de las revistas de ciencia y tecnología, donde el coste de los materiales ha aumentado a una velocidad de vértigo durante varias décadas. La otra cara de la moneda es que hasta ahora lo que las bibliotecas han hecho es hacer accesibles colecciones adquiridas, y si este cometido declina también lo harán los bibliotecarios. Si quieren continuar existiendo será esencial para las bibliotecas articular un cambio en la función de sus colecciones. Para conseguir esto resulta importante recordar lo que han hecho siempre la mayoría de las bibliotecas académicas:

1. Adquirir colecciones por apoyar a su comunidad u organización.
2. Custodiar colecciones especiales de ejemplares únicos o valiosos para la sociedad.

En el pasado dominaba el primer papel. En el futuro, el segundo será más importante. En el pasado, las colecciones que se gestionaban eran básicamente manuscritos y libros raros. En el futuro la mayor parte de lo que se gestione será digital. Una parte serán versiones digitales de colecciones especiales tradicionales, pero de manera creciente serán documentos nacidos digitales y productos digitales de las tareas de investigación. Gestionar las primeras se entiende de una manera razonable, gestionar las segundas será un reto, especialmente cuando grandes conjuntos de datos se hagan frecuentes con colecciones de datos ubicadas y a tiempo real en muchas áreas de la ciencia y las ciencias sociales. Mientras que en estos momentos entre el 80 y 90% de los recursos destinados a la colección de una típica biblioteca académica se dedican a la compra de fondo (lo que incluye no sólo el coste de los documentos impresos o electrónicos, sino también el coste de seleccionarlos, procesarlos y gestionarlos), y el 10-20 % restante se destina a las colecciones especiales tradicionales, en los próximos 20 años menos del 50% de las inversiones relacionadas con la colección se dedicarán a la adquisición de fondo y por encima del 50% irá a parar a “gestionar” el contenido digital. Las inversiones en las colecciones especiales tradicionales

continuarán aproximadamente al mismo nivel. Es importante reconocer el carácter inevitable de esta transición y aceptarla.

Hay una serie de retos que es necesario tener en cuenta a la hora de afrontar esta evolución:

Preservación: En primer lugar, las bibliotecas deben desarrollar las habilidades e infraestructuras para gestionar colecciones de contenido con las cuales no se está familiarizado. También hay que desarrollar tecnologías y estrategias para la preservación de información digital a largo plazo. Ahora se sabe cómo preservar bits durante décadas, pero hay que ser capaces de hacerlo durante siglos. Entre las habilidades importantes que se piden estará la habilidad de ayudar y apoyar al profesorado en la creación y recogida de este contenido. Este será un nuevo papel muy importante para los bibliotecarios especializados.

Reconocimiento: Un segundo reto será conseguir que el nivel de apoyo de esta actividad sea equiparable al nivel de apoyo que actualmente se da en el caso del fondo comprado. El valor de los fondos comprados está claro, especialmente para los usuarios. La gestión del contenido tiene un valor similar, pero no está siempre tan claro. Dado que el acceso abierto proporciona información sin ningún coste al usuario, ¿por qué ha de invertir una universidad para convertirse en proveedor de información? Muchas universidades descubrirán rápidamente el valor de gestionar, porque proporciona a los investigadores un mecanismo por compartir sus resultados y el proceso da reconocimiento y prestigio tanto a los investigadores como a la propia universidad.

Estrategia: Un tercer reto será equilibrar los beneficios del programa de gestión entre las diferentes unidades académicas del campus. Las bibliotecas gestionarán cosas diferentes para los historiadores que para los biólogos, pero han de estar haciendo algo para todo el mundo. El reto final será mantener una disciplina haciendo el cambio. No es posible construir un programa de gestión a menos que redistribuyamos recursos que ahora se utilizan para comprar fondo. Es de suponer que los editores lo pondrán todo lo difícil que puedan, y que muchos bibliotecarios opondrán resistencia al cambio porque es una amenaza a su papel tradicional. Hace falta una estrategia explícita aprobada por el campus porque un desarrollo sin rumbo e incremental no tendrá éxito. Este tipo de estrategia será difícil, porque realmente exigirá cancelar y dejar de adquirir documentos publicados a partir de la suposición de que el contenido estará disponible en formato de acceso abierto. Desde una perspectiva a largo plazo esto podría ser cierto, pero partiendo de la base del día a día, documento a documento, es poco probable que sea perfecto o a menudo ni tan solo aproximadamente sustituible. La forma más fácil de funcionar podría ser limitar el presupuesto

del contenido adquirido y pedir a los seleccionadores (tanto bibliotecarios como profesorado) que se ciñeran al presupuesto. También es probable que durante el periodo de transición haga falta un sistema de obtención de documentos (SOD) potente y reforzado y/o un mecanismo para la adquisición rápida de títulos concretos.

En resumen

Es posible decir que la edición electrónica representa cambios sustanciales para la biblioteca universitaria, tanto en la manera de concebir la formación de la colección, la manera de pensar sobre el espacio o la forma en que los bibliotecarios utilizan sus conocimientos técnicos, y en todos los casos una redistribución importante de los recursos presupuestarios y de personal para afrontar las nuevas necesidades.

La frontera que separa la biblioteca del resto de la universidad y el entorno de información exterior se está desdibujando. La biblioteca se vuelve un lugar menos diferenciado, menos exclusivo. Con todo, y a pesar de que algunos de los retos a los que se enfrenta la biblioteca universitaria son complejos, la misión de la biblioteca no se cuestiona: debe seguir haciendo accesible el conocimiento a los usuarios y colaborando con la misión de la organización que la acoge.

Acceso y gestión de los recursos electrónicos: los metadatos y los lenguajes de marcado

Sin duda, el desarrollo del formato MARC (Machine Readable Cataloging) durante los años 60 fue un avance revolucionario para la biblioteconomía moderna: estableció las bases que permitieron trasladar las bibliotecas a la era de los ordenadores y proporcionó una estructura común para registrar y transferir datos bibliográficos entre ordenadores. Juntamente con unas reglas de catalogación (principalmente las AACR), el formato MARC ha permitido y permite a las bibliotecas compartir la catalogación a gran escala, y de esta manera conseguir aumentar de forma considerable la eficacia de las tareas de catalogación, así como establecer la creación de grandes bases de datos de bibliotecas centralizadas, como la gestionada por la OCLC, que actualmente es el recurso bibliográfico más importante del mundo.

Pero desde los inicios del MARC el entorno tecnológico ha cambiado radicalmente. Cuando se creó el formato MARC la memoria de los ordenadores era muy cara, tan cara que cada carácter era un tesoro. Muy pocas personas tenían acceso a un ordenador, ni en el trabajo ni mucho menos en casa.

Internet no era más que una idea y el XML (Extensible Markup Language) aún se encontraba a décadas de existir.

En la actualidad hay que ser conscientes de que los registros bibliográficos no son patrimonio exclusivo de los sistemas de catalogación de bibliotecas, sino que se están utilizando en una amplia variedad de sistemas automatizados en el entorno de las bibliotecas, como por ejemplo los sistemas de préstamo interbibliotecario, los repositorios, en los directorios de recursos en línea de revistas electrónicas y bases de datos, etc. Y en muchas ocasiones, el formato MARC no encaja bien con estos sistemas, y la falta de una infraestructura rica de metadatos hace que las bibliotecas tengan que inventar soluciones que pueden hacer imposible la construcción de sistemas integrados de gestión de metadatos.

Los documentos, metadatos y TICs implicados en el nuevo panorama bibliotecario electrónico, digitalizado, son heterogéneos por naturaleza. Es por tanto imprescindible desarrollar mecanismos de interacción entre componentes diversos que puedan trabajar y entenderse de forma transparente, sin afectar al usuario. Es necesaria una interoperatividad a todos los niveles (funcional, sintáctica y semántica), entendiendo interoperatividad como la capacidad de dos o más entidades diferentes (sistemas o productos) para trabajar conjuntamente y sin un esfuerzo especial con el propósito de ofrecer al usuario los servicios esperados

La situación actual exige tener sistemas automatizados, políticas y procesos que permitan a las bibliotecas crear metadatos bibliográficos, asimilar metadatos bibliográficos de otros, mejorarlos, producirlos tanto de forma simple como compleja y hacer todo eso con facilidad y efectividad. También es necesario una infraestructura de metadatos bibliográficos que acepte cualquier metadato que se le presente y que permita producir registros simples cuando sea necesario, o registros complejos cuando se le pida hacerlo.

En conclusión, el interés por la interoperatividad se basa en

- Las bibliotecas son cada vez más digitales y tienen problemas para compartir e integrar datos
- Los servicios se dan en un entorno de red heterogéneo con múltiples tecnologías, aplicaciones, formatos de datos, etc.
- Es preciso “ocultar”, hacer invisible esa heterogeneidad para dedicarnos a ofrecer servicios de forma eficaz y de calidad
- No se trata de una cuestión supérflua: afecta directamente al acceso a la información
- Mejorar la interoperatividad optimiza el valor económico de la información: aumenta su consumo, la visibilidad y permite economías de escala.

Entre los múltiples aspectos que ayudan a la interoperatividad de los sistemas bibliotecarios, destacan los metadatos y los lenguajes de marcado, puesto que en muchos casos hacen posible la recuperación de la información.

Definiciones de metadatos

Los metadatos adquieren relevancia a partir del triunfo de la tecnología internet y de la web. Se carece de un consenso conceptual sobre el término “metadato”, por lo que se aborda desde diferentes puntos de vista, como por ejemplo:

- metalingüístico: “ Información sobre la información”
- etimológico: “Datos {junto a, con, entre} los datos”
- referencial: “Datos asociados a los objetos documentales que exigen a sus usuarios potenciales (personas o programas) de tener un conocimiento previo de su existencia”.

Los metadatos son un signo, en el sentido de “algo que se pone en el lugar de otra cosa o por alguna otra cosa” (Umberto Eco). Esa capacidad de sustituir a la cosa, al objeto documental es lo que posibilita que pueda representarla, revelando su estructura y valor semántico. De esta forma diremos que los metadatos sirven para identificar, describir, localizar, recuperar, organizar y preservar información y conocimiento, por parte de las personas y las máquinas. Ejemplos de metadatos son las fichas de catálogos, registros de bases de datos, etiquetas meta de páginas web. Entre los beneficios asociados a la utilización de metadatos está la de agilizar la recuperación de información, pues se indiza la representación del objeto documental y no el objeto documental mismo, o la de permitir que se comparta e integre información heterogénea (en soportes, ámbitos y disciplinas).

Resumiendo, podemos decir que los metadatos son un conjunto de información sobre objetos web, comprensible por máquinas. El objetivo de los metadatos es el acceder a recursos electrónicos con mayor facilidad gracias a sus elementos identificadores del recurso. Los metadatos más importantes y relevantes son los metadatos Dublin Core. MARC también constituye un sistema de metadatos para el entorno bibliotecario, pero fue concebido con propósitos distintos a la descripción de recursos electrónicos en la web.

Tipos de metadatos

- Según su ámbito disciplinar: MARC, DC (Dublin Core), EAD (Encoded Archival Description), GILS (Global Information Locator Service), etc.
- Según su ciclo de vida: administrativos, descriptivos, conservación, técnicos, uso
- Según su complejidad estructural: desestructurados, semiestructurados, estructurados

Esquemas de metadatos

Un esquema de metadatos es un conjunto de campos, junto con sus reglas de uso, concebidos para un propósito concreto. Son características de un esquema de metadatos una:

- Sintaxis: alude a los identificadores de contenido, su carácter obligatorio, repetible... (por ejemplo, el formato MARC 21)
- Reglas de contenido: prescriben cómo seleccionar y representar los valores de un campo (por ejemplo, las AACR2)
- Semántica: fija la codificación de los campos del esquema en la máquina, para ofrecer un formato de intercambio (por ejemplo, la ISO 2709)

Es posible decir que el esquema de metadatos MARC es un compuesto de diferentes normas, códigos, listas ... (ISO 2709, MARC21, Anglo-American Cataloguing Rules, etc.)

Visión rápida sobre los lenguajes de marcado

El lenguaje de marcado (markup language), también denominado lenguaje de anotaciones o de etiquetas, se define como un conjunto de reglas para estructurar y dar formato a un documento electrónico. Suelen utilizar etiquetas para definir el inicio y el final de un elemento: un párrafo, un título, un elemento subrayado, etc. Los lenguajes de marcas más utilizados son HTML y XML, ambos basados en el metalenguaje SGML (standard generalized markup language). Un lenguaje de marcado cumple con dos objetivos esenciales para diseñar y procesar un documento digital:

- Separa un texto en los elementos en los que se compone, como por ejemplo un párrafo, un capítulo, etc.
- Especifica las operaciones tipográficas y funciones que debe ejecutar el programa visualizador sobre dichos elementos. Las operaciones tipográficas son instrucciones de formato que se aplican a cada uno de los elementos de un documento digital, por ejemplo, visualizar un título en itálicas.

En conclusión, los lenguajes de marcado son:

- La invención técnica al reto de organizar y recuperar información electrónica.
- El sistema formal que añade datos codificados al texto para representar su significado y controlar su procesamiento.

Por último, decir que las semánticas utilizadas en los lenguajes de marcado son mucho más abiertas, versátiles y extensibles que la establecida por la ISO 2709 para los registros MARC, cuya robusta estructura de datos se define posicionalmente, lo que implica cierta limitación.

Clases de lenguajes de marcado

Existen dos clases de lenguajes de marcado, que se diferencian en teoría en cada documento pero que se pueden encontrar juntos en la práctica, aunque no sea lo habitual. No existen restricciones respecto a cuál de estas clases de lenguajes pueden utilizarse en el mismo documento

- Marcas de procesamiento: orientadas a la presentación de la información (HTML)
- Marcas descriptivas: describen las partes de texto, pero no indica cómo se deben representar, éstas pueden añadir al contenido cualquier clase de metadatos. Los lenguajes diseñados específicamente para realizar marcado descriptivo son SGML y XML.

El marcado descriptivo es muy flexible y simplifica la tarea de volver a dar formato a un texto, ya que la información de formato se encuentra separada del contenido

SGML

SGML (Standard Generalized Mark-up Language), es un lenguaje generalizado estándar para el marcado de documentos. Es un estándar internacional que permite definir lenguajes para dar formato a documentos. Por ejemplo, el HTML es un lenguaje de formato de documentos definido de acuerdo con SGML; es decir, HTML es una aplicación de SGML para dar formato a documentos de hipertexto. SGML es un lenguaje para describir otros lenguajes.

El lenguaje de marcado SGML comenzó a gestarse en 1969, cuando un equipo de investigadores de IBM desarrolló GML (Generalized Markup Language). GML se desarrolló como una vía para crear la documentación básica de la empresa en un formato electrónico transferible y fácil de gestionar, Se trata del primer lenguaje no propietario de marcado de texto (independiente del sistema en el que se crean los documentos y de la plataforma en que circulan) capaz de definir estructuras lógicas de cualquier tipo de documento, basándose en una serie de normas. GML consiguió el estatus de norma ANSI bajo la denominación de SGML (Structured Generalized Mark-up Language) en 1978. Esta norma fue adoptada como norma ISO en 1986 y forma parte de un conjunto más amplio de normas bajo la denominación genérica de "Information processing-text and office systems-standard generalized markup language". La literatura científica se refiere a este conjunto de normas genéricas y/o especificaciones subsidiarias y/o complementarias en torno a la ISO 8879:1986 como la familia de normas SGML/XML.

SGML es utilizado desde mitad de los 80 y ha permanecido bastante estable. Gran parte de su estabilidad se debe al hecho de que el lenguaje es a la vez flexible y rico en posibilidades. Sin embargo, esta estabilidad tiene un

inconveniente: el nivel de complejidad que ha producido su uso en diversos ámbitos como la World Wide Web.

SGML fue diseñado para permitir el intercambio de información entre distintas plataformas, soportes físicos, lógicos y diferentes sistemas de almacenamiento y presentación, independientemente de su grado de complejidad. SGML especifica reglas sintácticas muy generales para escribir documentos, pero ninguna etiqueta concreta y proporciona una herramienta para que cualquiera se cree sus propias etiquetas y las reglas de uso: las DTD. Las DTD (document type definition o definición de tipo de documento) de SGML define la estructura de un tipo de documento específico, abarcando varios aspectos: los elementos que pueden formar parte del tipo de documento, el nombre de los elementos y si son repetibles, el orden de los elementos, los contenidos de los elementos, qué tipos de marcado pueden ser omitidos, los atributos y sus valores por defecto y los nombres de las entidades permitidas. Por ejemplo, HTML es una DTD que fija como etiquetas permitibles <title> <head> <meta> <link> <body> <p>. En este caso, <title> y <meta> deben estar dentro de <head>, mientras que <p> sólo ocurre dentro de <body>

SGML es inherentemente jerárquico y por tanto ideal para expresar las relaciones jerárquicas de los objetos documentales. Además, permite expresar la semántica de cualquier esquema de metadatos, incluso MARC 21, por ejemplo: MARC: 245 03 \$a El nombre de la rosa /\$cpor Umberto Eco → SGML: <E245 i1=0 i2=3><E245-a> El nombre de la rosa /></E245-a><E245-c>por Umberto Eco</E245-c>. La gran desventaja es que es un lenguaje difícil para ser procesado por los programas.

XML

XML (Extensible Mark-up Language) es un lenguaje universal de marcado para documentos estructurados y datos en la web, mucho más amplio, más rico y más dinámico que HTML, desarrollado por el Consorcio World Wide Web (W3C).. Pero XML es algo más que un lenguaje de marcado: XML, al igual que SGML, es un lenguaje capaz de definir lenguajes, un metalenguaje. De hecho hereda la misma estructura que SGML. Sin embargo, es mucho más sencillo y más fácil de aplicar que SGML, y está diseñado para hacer frente al gran y rápido desarrollo de las nuevas tecnologías de Internet.

En teoría HTML es un subconjunto de XML especializado en presentación de documentos para la Web, mientras que XML es un subconjunto de SGML especializado en la gestión de información para la Web. En la práctica XML contiene a HTML, pero no en su totalidad. La definición de HTML contenido totalmente dentro de XML y, por lo tanto, cumpliendo a rajatabla la especificación SGML, es XHTML (Extensible Hypertext Markup Language).

En XML no hay elementos. Cada usuario puede crear su propio lenguaje para el formato de datos y documentos según sus necesidades, siguiendo las reglas de XML a partir de unos estándares generales.

Al igual que HTML, se basa en documentos de texto plano (aquellos que están compuestos únicamente por texto sin formato, sólo caracteres) en los que se utilizan etiquetas para delimitar los elementos de un documento. Sin embargo, XML define estas etiquetas en función del tipo de datos que está describiendo y no de la apariencia final que tendrán en pantalla o en la copia impresa, además de permitir definir nuevas etiquetas y ampliar las existentes. HTML no indica lo que está representando, se preocupa más del formato (cómo se ve) que de lo que se está mostrando (qué se ve). XML describe el contenido de lo que etiqueta. Su objetivo es conseguir una página Web más semántica.

Son varios los vocabularios desarrollados en XML con el fin de ampliar sus aplicaciones. Podemos considerar fundamentales: XHTML, XSL-FO y XSLT, XLink, XPointer y Schema. Además, existen también versiones para usos específicos, como MathML (fórmulas matemáticas), SVG (gráficos vectoriales), DocBook (lenguaje para el formato de libros electrónicos), RSS (sindicación de noticias), o XBRL (partes financieros).

XML se puede emplear mediante dos herramientas: DTDs o Esquemas (que usan xml), gestiona los espacios de nombre (*namespaces*), y puede visualizarse en navegadores empleando XLS.

Ventajas

- Gestiona información jerárquica (ej: todos los documentos de una página web en un mismo registro estructurado).
- Integra información textual y multimedia (ej: un documento con texto, sonido e imagen)
- Facilita la cohabitación entre esquemas (generales y especializados) en un mismo registro
- Permite el intercambio entre esquemas (crosswalks)
- Los datos están en ascii, y no en binario propietario, lo que mejora el intercambio de información entre aplicaciones.

RDF (Resource Description Framework)

Conforme vaya creciendo la información en la Red que se marca con etiquetas XML específicas de cada actividad, más fácil será encontrar lo que se necesita. Pero esto no es más que un paso intermedio. Hace mucho que los bibliotecarios descubrieron que la forma más rápida de hallar datos no consiste en revisar la información propiamente dicha, sino en ceñirse a un conjunto de referencias mucho menor, que nos oriente hacia las fuentes útiles: es decir, un sistema de metadatos. Desde el primer momento, parte del proyecto XML

consistió en la creación de una norma gemela para crear metadatos. El RDF (Resource Description Framework) debería ser para los datos de la Red lo que los ficheros bibliográficos son para las obras de una biblioteca. Consiste en aplicar la sintaxis XML a los metadatos, y no sólo a los documentos primarios (los textos)

Ventajas:

- Permite la cohabitación entre cualquier esquema de metadatos, y por tanto, una amplia interoperatividad.
- Reduce las disparidades terminológicas o semánticas que se presentan, pues al emplear el elemento *namespace* de XML, aclara el significado de un campo de cualquier esquema de metadatos utilizado al describir un recurso de información.

Dublin Core (DC)

Dublin Core Metadata Initiative fue creada en 1995 tras un congreso en Dublin, la cual se autodescribe como una “organización dedicada al desarrollo y adopción de estándares sobre metadatos”. Su función consiste en mantener el conjunto de metadatos “Dublin Core”, el sistema de metadatos más popular en la descripción de recursos electrónicos en Internet.

Dublin Core Metadata Element Set, se publicó en 1998 y es un conjunto de elementos que se pueden aplicar en la descripción de un recurso con el objetivo de facilitar su posterior recuperación. El crecimiento exponencial de la Web llevó a la necesidad de plantear un modelo similar al del catálogo de bibliotecas, de esta forma, se popularizó el concepto de metadatos: DCMES traducido a 25 idiomas. Dublin Core es un sistema sencillo cuyo objetivo es permitir la rápida descripción de los recursos. Cualquier persona sin mucho conocimiento podría utilizar Dublin Core tras una formación simple

Se trató de definir un conjunto de metadatos o propiedades aplicables a cualquier tipo de recurso, con independencia del “sector” o “contenido” que tuviese y de esto resultaron 15 elementos, los cuales pueden ser usados libremente y que suelen ser utilizados mayormente en el contexto de las páginas web. El conjunto de elementos Dublin Core, o lo que se conoce habitualmente como el DC simple, están definidos por las norma ISO 15836:2003, recientemente actualizada (2008). Los metadatos Dublin Core pueden indicarse interna o externamente al recurso y no se obliga a codificarlos utilizando HTML, XML, RDF, ni ningún otro sistema, sino que se ofrecen alternativas para poder codificarlos en cualquiera de estos lenguajes.

Estos 15 elementos son:

Contenido

- Cobertura (Coverage): Cobertura del contenido del recurso. La cobertura puede ser espacial (coordenadas físicas de un lugar) o temporal (de periodo de tiempo).
- Descripción (Description): Esta etiqueta incluye una descripción o resumen del recurso.
- Tipo (Type): Tipología o categoría del recurso, por ejemplo documento técnico, diccionario, etc.
- Relación (Relation): Relación de ese recurso con otro u otros. Esto facilita en enlace entre varios recursos.
- Fuente (Source): Información sobre otro recurso del cual deriva el recurso que está siendo descrito.
- Materia (Subject): El tema o temas sobre los que trata el recurso.
- Título (Title): Señala el nombre del recurso.

Propiedad intelectual

- Colaborador (Contributor): Persona u organización que haya tenido una contribución en la creación del recurso.
- Autor (Creator): Persona física o jurídica que ha creado el documento.
- Editor (Publisher): Entidad o institución que edita dicho recurso.
- Derechos (Rights): Etiqueta que muestra los derechos de autor del recurso (copyright).

Versión / Acceso

- Fecha (Date): Fecha de creación del recurso.
- Formato (Format): Sirve para identificar el software y posiblemente, el hardware que se necesita para abrir el recurso.
- Identificador (Identifier): Etiqueta que muestra una serie de caracteres que identifican al recurso.
- Lengua (Language): Idioma en que está escrito el recurso.

Los elementos de Dublin Core son opcionales y repetibles, y el esquema, además permite emplear calificadores opcionales para cada elemento que posibilitan indicar la normativa empleada en caso de haber usado normas de descripción bibliográfica usuales. Los calificadores permiten aumentar la especificidad y precisión de los metadatos, pero pueden también introducir complejidad que disminuiría la compatibilidad con otras aplicaciones que usen Dublin Core. Dublin Core Metadata Initiative reconoce sólo dos grandes grupos de calificadores: refinación de elementos (p. ej. título alternativo) y esquema de codificación (p. ej. materia según LCSH).

Podemos decir que una entrada Dublin Core dentro de HTML tiene la siguiente sintaxis: META NAME = "DC.Creator" CONTENT = "Nombre de la entidad donde se crea">. La sigla DC, hace referencia a la utilización de Dublin Core. La etiqueta que estamos utilizando como podemos ver es la de Creator (tal y como se ve en el ejemplo). "CONTENT" es utilizado para describir el valor del elemento.

Tendencias en esquemas de metadatos: MODS

MODS (Metadata Object and Description Schema) es un esquema de metadatos descriptivo compatible con MARC y pero con sintaxis XML, que pretende superar las insuficiencias de MARC. El esquema MODS traslada a XML los metadatos MARC (hereda su semántica) pero abandona la rígida sintaxis ISO 2709. Sus etiquetas resultan más amigables: las etiquetas numéricas de MARC son sustituidas por etiquetas textuales.

Conversión o mapeo (mapping) de MARC→MODS es sencillo, pero no es reversible. Un registro MODS incluye menos información que un registro MARC, es más simple, y aunque técnicamente es posible hacer la conversión MODS →MARC el registro resultante presentará pérdidas de información importantes.

Perspectivas de uso:

- Con OAMHP
- Con Z30.50 SRW
- Para describir recursos originales.
- Con registros METS

Proyectos:

- American Memory and Global Gateway Legacy Conversion Project
- Performing Arts Encyclopedia
- Library of Congress Web Archives (LCWA) (early named MINERVA)
- Y hasta 29 proyectos más están registrados en el sitio web de la LC
- En España: Biblioteca Virtual del Patrimonio Bibliográfico

Existe un esquema de metadatos equivalente para registros de autoridad: MADS (Metadata Authority Description Schema)

Tendencias en esquemas de metadatos: METS

METS significa "Metadata Encoding and Transmisión Standard". Esquema XML que empaqueta todos los metadatos asociados con un recurso digital (descriptivos, administrativos, estructurales necesarios para recuperar, preservar y ofrecer recursos digitales). Se trata de una norma convergente e integradora, destinada a describir, transmitir y gestionar recursos digitales.

Surge en el seno de la DLF (Digital Library Federation) como respuesta al modelo de gestión de depósitos de recursos digitales OAIS. Permite una gestión integral y jerárquica de manifestaciones de una misma obra: texto, grabación sonora y vídeo de una conferencia.

Una sintaxis de transferencia METS puede trasladar contenedores de metadatos correspondientes a diversos estándares, o a ninguno, y asociarse al mismo objeto documental. Por ejemplo, un registro MODS y un registro con metadatos bibliográficos ONIX (ONline Information eXchange) procedentes de una base de datos interna de la editorial. Los campos se indexarán desde los dos registros para la búsqueda y visualización del usuario.

Partes de un registro o paquete METS:

- Cabecera: informa del doc METS mismo (identificadores, fecha de creación, actualizaciones y status)
- Metadatos descriptivos: admite esquemas DC, MARCXML y MODS, entre otros.
- Metadatos administrativos para difusión y preservación
- Sección de ficheros: identifica la totalidad de ficheros que componen el objeto
- Mapa estructural: contiene el diseño de la estructura jerárquica del doc
- Sección de comportamiento y funcionamiento: aplicaciones empleadas para visualizar el documento.

Uso: para construir colecciones digitales los usan LC, Harvard, OCLC... En España lo ha implementado el Ministerio de Cultura en la Biblioteca Virtual de Prensa Histórica y la Biblioteca Virtual del Patrimonio Bibliográfico

BIBLIOGRAFÍA

David W. Lewis. "A Strategy for Academic Libraries in the First Quarter of the 21st Century". En: *College and Research Libraries*, Vol. 68, No. 5 (Sept. 2007)

<https://idea.iupui.edu:8443/dspace/bitstream/1805/953/1/DWLewis_Strategy.pdf>

Roy Tennant. "A bibliographic metadata infrastructure for the twenty-first century". En: *Library Hi Tech*, Vol. 22, No. 2 (2004), p. 175-181

Daniel Martínez Ávila. ["Tendencias y usos de XML en Biblioteconomía y Documentación"](#)