

Administración de Contenido y XML en las colecciones digitales

Alma Beatriz Rivera Aguilera

Resumen

Este trabajo es una reflexión sobre el concepto de administración de contenido en el contexto de las colecciones digitales y el uso que de los llamados sistemas de administración de contenido (CMS por sus siglas en inglés) puede darse en el manejo de las acervos electrónicos propios de cada institución. El documento reporta el modelo de sistema de administración de contenidos desarrollado por Bob Boiko y el papel del lenguaje de marcado XML en este tipo sistema. Se concluye que es necesaria la participación propositiva y se hace un llamado a no solo consumir conceptos y tecnologías sino a conocerlas a fondo, cuestionarlas y aprovecharlas con propuestas específicas; ya que los profesionales del manejo de la información poseen experiencias y conocimiento suficiente para generar aportes valiosos Finalmente se indica que un buen sistema de administración de contenidos puede ser la base de un buen sistema de administración del conocimiento.

Introducción

Este trabajo reporta sobre el concepto de administración de contenido en el contexto de las colecciones digitales y el uso que de los llamados sistemas de administración de contenido puede darse en el manejo de los acervos y servicios electrónicos propias, cada vez más comunes, en bibliotecas y archivos. Se describe el modelo de sistema de administración de contenidos desarrollado por Bob Boiko y el papel del lenguaje de marcado XML en este tipo sistema. Se reflexiona sobre la función del bibliotecario ante el dinamismo de las tecnologías, los recursos escasos y las exigencias del enfoque multidisciplinar; lo cuales enfrentan a los profesionales al reto de capacitarse constantemente y trabajar en equipo. Se concluye que es

necesaria la participación propositiva y se hace un llamado a no solo consumir conceptos y tecnologías; ya que hace falta una discusión entre profesionales de nuestra área y de otras áreas para concretar proyectos conjuntos de forma exitosa. Finalmente se señala que un buen sistema de administración de contenidos puede ser la base de un buen sistema de administración del conocimiento.

II. Colecciones y Servicios digitales en el mundo y en México.

Antes de entrar en materia técnica vale la pena contextualizar el porqué de las colecciones y servicios digitales. Tal y como las colecciones en papel la generación y administración de acervos electrónicos tiene objetivos sociales entre los que se destacan los de tipo educativo y cultural. En el texto compilado por Dowler (1997) “Gateways to Knowledge” (Portales de conocimiento) los autores reflexionan y sobre el impacto de las tecnologías en la enseñanza, el aprendizaje y la investigación; así como el papel de las bibliotecas como portales en este escenario. En el ámbito mexicano tenemos el material de Malo Alvarez y Fortes Besprovani¹ de 1999 en el cual los autores proponen atender a las necesidades de la sociedad mexicana en la era de la información a través de una red de bibliotecas digitales. Nuestras recatadas instituciones se han visto en los últimos 20 años en la necesidad soltarse el pelo y navegar la cresta de la ola tecnológica, cabe mencionar que no se ha salido mal parado de tal aventura. El bibliotecario se ha enfrentado a que tanto los procesos técnicos como los servicios a usuarios en el ámbito digital conllevan una serie de nuevas tareas que van desde la generación o selección de los archivos digitales, su preparación, preservación, catalogación, marcado, publicación, etc. Aunque ciertamente cada proyecto es diferente y existe un sin número de situaciones, son recomendables los textos de Sitts(2000) y Kenney (1999) como excelentes guías para las tareas técnicas de la generación de colecciones digitales y la administración

¹ Aunque las propuesta de estos autores solicitada por la Academia Mexicana de Ciencias hace ya 4 años no se han concretado el material sigue siendo muy valioso por su enfoque y síntesis acertada. Entre las razones que se considera incidieron en la falta de concreción de este tipo de proyectos puede vislumbrarse la falta de voluntad política y la cultura misma de los desarrollos tecnológicos que es de tipo más bien independiente y no fácilmente programable.

de este tipo de proyectos. No quisiera dejar de mencionar el reciente libro del australiano Witten (2003), de enfoque computacional el cual incluye software gratuito, que da una visión alternativa y no comercial, diferente a la norteamericana sobre como construir bibliotecas digitales².

En el mundo se han dado dos corrientes sobre el tema de bibliotecas digitales: enfoque computacional y el bibliotecológico, ambos han tenido grandes logros y, aunque a veces separados en no pocas ocasiones han colaborado exitosamente (Borgman, 2000). Para conocer detalles del escenario mexicano en fuentes impresas se puede revisar entre otros a Sánchez (2002) con enfoque computacional quien hace un registro de los proyectos más importantes, si bien no todos los existentes. Ya en 1997 Feria, con un enfoque bibliotecológico, refería importantes avances en los servicios y colecciones electrónicas generadas en diversidad de proyectos por la Universidad de Colima. En los últimos 5 años se han presentado por una gran variedad de instituciones bibliotecarias y archivos decenas de ponencias sobre el desarrollo de colecciones y servicios digitales. Existen 2 foros dedicados al tema de las tecnología de información: El antiguo Coloquio de Automatización de Bibliotecas hoy Interfaces de la Universidad de Colima enfocado en el quehacer bibliotecológico, aunque con gran presencia de cómputo y pedagogía y el evento del Grupo Amigos en la Universidad de las Américas en Cholula Puebla, cuyo primer encuentro bienal fue en 2000 sobre las bibliotecas digitales y el segundo en 2002 sobre administración del conocimiento. El encuentro de Cholula está más enfocado a los aspectos computacionales, sin embargo tanto en ponentes como en asistentes hay presencia bibliotecaria. En muchas de las presentaciones se ha tenido la oportunidad de conocer proyectos de gran calidad algunos con participación multidisciplinaria, en esta línea un ejemplo interesante es el proyecto Hermes de la UNAM.

² Este enfoque de alguna manera coincide con el de los proyectos del Tec. De Monterrey y la UDLA que se han venido desarrollando con fondos de CONACYT por personal de cómputo de ambas instituciones, y que ha generado algunos software y diseños de acceso libre.

El gobierno, por su parte, ha tomado alguna iniciativa de apoyo a este tema con el grupo de CONACYT dedicado a las bibliotecas digitales, el cual está liderado por el enfoque informático.

Es un hecho que las colecciones digitales son cada día más comunes en los archivos y bibliotecas, y que la generación y el manejo de dichas colecciones requieren de la intervención de profesionales de múltiples disciplinas: computólogos, editores, bibliotecarios, diseñadores gráficos, diseñadores instruccionales, entre otros.

III.Administración de Contenido

El los últimos años los bibliotecarios hemos ampliado nuestro vocabularios profesional con términos como administración del conocimiento, minería de datos, bibliotecas digitales, metadatos, Dublin Core, XML, lógica difusa, etc. Estos conceptos provienen de diversas disciplinas y en no pocas ocasiones han surgido de cruces disciplinares, se lista aquí sin pretender ser exhaustivos algunos conceptos y su correspondiente origen multidisciplinar:

- 1.Computación: Biblioteca Digital, Lenguaje Natural, Inteligencia Artificial, Minería de Datos, Metadatos, RDF, Bases de Datos.
- 2.Bibliotecología/Cómputo: Registro bibliográfico, formato MARC, Dublin Core
- 3.Edición/Computación: Lenguajes de Marcado, Publicación Electrónica
- 4.Administración/Computación: Administración del Conocimiento y a partir del comercio electrónico aplicaciones concretas de agentes de inteligencia artificial, XML, bases de datos
- 5.Diseño gráfico aspectos de publicación en web y arquitectura de información.
- 6.Diseño instruccional (Pedagogía/Cómputo): Educación a Distancia, Estilos de Aprendizaje, Objetos de Aprendizaje.
- 7.Filosofía/Matemática/Computación: Lógica booleana, lógica difusa.

El concepto que aquí nos interesa es el de Administración del Contenido (Content Management) , originado en el área del comercio

electrónico y presentado por Bob Boiko en 2001, quien para abordarlo como punto de partida define (cap. 1)³:

1. Datos: Pequeños elementos de información que no permiten interpretación. Las computadoras fueron creadas para procesar datos.
2. Información: Todas las formas comunes en las cuáles se registra una comunicación (textos, música, imágenes, video y archivos como hojas electrónicas, presentaciones y cualquier otra forma de archivo propietario), repositorios de mensajes.
3. Contenido: Información con contexto y significado humano, es información enriquecida que se “envuelve” en datos simples. Es un compromiso de colaboración entre la utilidad de los datos y la riqueza de la información.
4. Metadatos: lo datos que “envuelven” la información y representan una versión simplificada del contexto y del contenido de la información
5. Sistema de Administración de Contenido: Un CMS administra el contenido a través de los datos que lo representan (metadatos) y pretende funcionalidad y control sobre la creación y distribución de la información

Una vez revisados los conceptos básicos Boiko (2001b) señala 2 elementos fundamentales de los objetos digitales, que al comprenderlos pueden ayudarnos a construir lo mejor posible un CMS: formato y estructura

Formato, es como se codifica la información para que una computadora pueda leerlo, en general lo que leen los equipos son 0's y 1's, es decir código binario, el cual puede recibirse con diferentes composiciones o formatos de archivo, los cuáles aunque son básicamente código binario no siempre son fáciles de mover de un formato de archivo a otro. Por ejemplo en imágenes para publicación en Web tenemos al JPEG y al GIF como estándares de código binario sin embargo el publicación impresa son más bien el EPS y el TIFF

³ Boiko no pretende que estas definiciones sean totalmente formales sino más bien ilustrativas para entender su propuesta de Sistemas de Administración de Contenido o CMS. Sin embargo corresponden fundamentalmente con las definiciones de datos e información que utiliza Svenonius (2000, cap1), conocida autora sobre los conceptos fundamentales de la catalogación.

(Boiko, 2001b, p. 13). Cada estándar tiene su manera de representar el código binario, por ejemplo los gráficos vectoriales se almacenan como ecuaciones con valores que al ser calculadas permiten que se vean las formas de las imágenes. En el caso del texto el estándar puede ser ASCII o UNICODE, y sobre ello un arreglo que lo convierte en .DOC o HTML.

Un CMS debe pues (Boiko 2001b, p, 20) aceptar y crear los formatos binarios necesarios y aplicarlos de forma consistente. Con respecto a los externos que recibe reconocer, analizar y sintetizar variedad de formatos en un sistema coherente.

Estructura, es como se organiza la información al vaciarla en un objeto digital (Boiko, 2001b, p. 21). Puede categorizarse la estructura por tipo de la siguiente forma (pp. 26-28):

1. *Estructura por División*: esta en relación a la división del contenido en piezas utilizables, pieza puede referirse aquí desde una palabra hasta el URL de un sitio web
 - a. Segmentos⁴ [de una colección digital]: artículos, folletos, cartas, mails, imágenes, etc.
 - b. Elementos dentro de un segmento: título, resumen, cuerpo, párrafo, texto en negrita, nota al pie, menú de opciones, etc.
2. *Estructura de Acceso*: la necesaria para acceder el contenido
 - a. Jerarquías, Tablas de contenido
 - b. Índices
 - c. Referencias cruzadas o ligas
 - d. Secuencias de vista o “browsing”
3. *Estructura Administrativas*: Atributos que permiten encontrar y administrar el componente de contenido. Autor, fecha de creación, número de versión, estado de revisión.
4. *Estructura inclusiva*: Qué componentes incluyen otros. Referencias a imágenes por ejemplo.

Boiko (2001b, p. 29) utiliza el concepto de Arquitecto de Contenido o Metator (metautor?) como el individuo que divide información homogénea y la marca para accederla y administrarla; este

⁴ El término segmento no me parece muy afortunado pues puede dar lugar a confusión, sin embargo es el que usa Boiko, sugiero que se tome como segmento o pieza de una colección digital.

profesional debe crear jerarquías, índices, estructuras de referencias cruzadas y secuencias.

Los anteriores conceptos en relación a estructura como identificadora de elementos que facilitan el acceso a los contenidos son tremendamente familiares para nosotros los bibliotecarios y podemos ver claramente el valor de nuestra experiencia en su desarrollo. El desarrollo de un CMS conllevan, sin embargo, tareas que van más allá de nuestra disciplina; por ejemplo, la implementación de este tipo de sistemas implica un alto componente de programación; lo cual nos enfrenta una vez más al trabajo colaborativo.

IV. Administración de contenido y bibliotecas.

Como ya se dijo, el concepto de administración de contenido nace en el mundo del comercio electrónico, pero se puede ver claramente que es útil para bibliotecas y archivos. Wu y Liu (2001, p 434) revisan el tema en relación al futuro de las bibliotecas académicas y proponen que dichos sistemas sean capaces de generar un buen mantenimiento sobre las colecciones y portales digitales a través de las siguientes características:

1. Administración con herramientas que por medio de un interfase simple y flexible permitan el control manual del contenido
 2. Autoría a través de conversión de diferentes plataformas y formularios para la creación sencilla.
 3. Publicación en servidores locales y no locales en múltiples idiomas y plataformas
 4. Seguridad para evitar alteraciones no autorizadas de contenido o ataques de hackers
 5. Escalabilidad para asegurar el crecimiento y expansión de software y hardware
 6. Flujo de trabajo de fácil adopción que permita incorporar tareas múltiples como anuncios, estadísticas, alertas, nuevos contenidos.
- Wu y Liu (2001), explicitan los elementos de la propuesta de Boiko de un CMS que se centra en las actividades de coleccionar, administrar y publicar.

La Biblioteca del Congreso reportó recientemente (Rogers, 2002) la adquisición de un CMS que administra su sitio web y permite trabajo colaborativo en el mantenimiento y actualización de sus materiales digitales. OCLC utiliza el software Poet (Wu y Liu, 2001, p. 435) para su documentación técnica, educación a distancia, ayuda en línea y publicación en web.

V.XML y la administración de contenido.

No se explicará aquí en amplitud el concepto del XML, existen un buen número de libros y artículos que han cubierto los conceptos básicos (mis preferidos con respecto a los aspectos técnicos del XML son Harold (2001) y Morrison (2000); con respecto al uso del XML en las bibliotecas son Tennant (2001) y Kim (2000) y Miller (2000)⁵). Mencionaré, sin embargo, que el XML es un lenguaje de marcado que permite integrar a un documento electrónico tanto sus metadatos de encabezado o descriptivos como los llamados estructurales, es decir, etiquetas que describen los elementos catalográficos por un lado y la estructura del archivo u objeto digital por otro.

En algunos proyectos mexicanos de bibliotecas digitales se ha utilizado el XML en el marcado tanto de encabezado como estructural, por ejemplo en el departamento de Publicaciones Digitales de DGSCA⁶ y en la UDLA en los telegramas de Porfirio Díaz⁷.

Para Boiko (2001b, 753-756) el XML permite añadir la estructura necesaria a los objetos digitales para encontrarlos y presentarlos a los usuarios. Retomando los aspectos que este autor propone en un sistema de administración de contenido podemos ubicar en uso del XML en relación a:

⁵ En México Lafuente y Garduño (2001) publicaron un libro sobre los lenguajes de marcado donde se menciona brevemente el XML. El libro de Miller está en español y aparece, como el resto, citado en la bibliografía.

⁶ La Maestra Clara López, Jefa de Publicaciones Digitales de la DGSCA, UNAM ha presentado diversas conferencias y ponencias en donde ha explicado detalladamente el uso del XML en las colecciones digitales de dicha dirección.

⁷ Alberto Sánchez, encargado de Biblioteca Digital en la UDLA, Puebla ha reportado esta aplicación en diversos foros.

1. Colecciones: refiriéndose la creación e integración de archivos en un solo repositorio XML, para lo cual es necesario:
 - a. Herramientas de **autoría** en XML
 - b. Herramientas de **conversión** al XML
 - c. Seguimiento de las **reglas** a través de los DTD o schemas establecidos para la generación de archivos XML
 - d. **“Syndication”**⁸, es decir integración de archivos externos XML de forma automática
2. Administración: de las colecciones de objetos digitales para el acceso y distribución de archivos, lo cual podría alcanzarse con las siguientes tecnologías:
 - a. **Repositorios de objetos**, en donde cada archivo es independiente y puede ser utilizado. Es conveniente seguir estándares W3C XPath y XSL. Este es el sistema ideal de acuerdo a Boiko, pero la gente de los departamentos de informática está mas a gusto con los dos siguientes esquemas.
 - b. **Repositorios de bases de datos relacionales**. El archivo XML se guarda como un campo en una base de datos. Para manipular los archivos se necesita del herramientas como DOM (estándar de Microsoft que proporciona acceso programático a los documentos en XML o HTML)
 - c. **Sistemas de archivos**. Cada archivo se guarda en su formato original, XML, .doc
3. Publicación: entrega de la información a los usuarios para lo cual se pueden utilizar los enfoques de:
 - a. Formularios para generar versiones, utilizando estandars como ASP/COM de Microsoft o JSP/J2EE (Java) aprovechando o no el XML.

⁸ A pesar de que el Diccionario de la Universidad de Chicago traduce sydication como sindicación no me parece que sea el término más adecuado para referirse a la integración de archivos externos.

- b. DOM y XSLT (transformar archivos XML en salidas HTML).

Existen sistemas comerciales de Administración de Contenido que poseen alguna de las características genéricas mencionadas como Documentum, Interwoven, Poel, Sanderson CMI, Solid8 Software y muchos más (Wu y Liu, 2001, p. 434-435 y Margulius, 2002) . Sin embargo, es necesario ir considerando desde ya el mejor diseño para los sistemas locales de administración de archivos o parametrizar los comerciales en su caso, en ambos casos conviene tener conocimiento del tema. Por ejemplo, los DTD (Document Type Definition) o schemas que establecen las estructuras de los documentos marcados con XML tienen que ser definidos y aprobado por los usuarios en la mayoría de los casos.

V. Conclusiones

No se puede asegurar si este concepto será una moda pasajera o tendrá adeptos y futuro, lo que si se sabe es que la necesidad de administrar contenidos para poder ser verdaderos portales de conocimiento en apoyo a la docencia, a la investigación, al desarrollo comunitario y no de confusión sofisticada. La exigencia de orden y control en los repositorios digitales y los portales de las bibliotecas es real y por lo tanto coleccionar, administrar y publicar contenidos serán actividades en las que nos veremos y queremos estar envueltos los bibliotecarios, ya seamos de sistemas, de procesos técnicos o de consulta. En especial porque creo que es valioso lo que tenemos que aportar desde nuestra disciplina. Después de haber tenido esta primera aproximación con el concepto de administración de contenido me atrevo a proponer que deberíamos:

1. Conocer más a fondo el concepto de administración de contenido y discutirlo
2. No solo consumir conceptos y tecnologías, sino a investigarlas a fondo, cuestionarlas y aprovecharlas con propuestas específicas ya que poseemos experiencias y conocimiento para generar aportes valiosos.

3. Discutir entre profesionales de nuestra área y de otras áreas para concretar mejores proyectos relacionados con colecciones y servicios digitales.
4. Un buen sistema de administración de contenidos que pueda ser la base concreta de un buen sistema de administración del conocimiento

Bibliografía

- Boiko, B. (2001a). Understanding content management. *Bulletin of the American Society for Information Science and Technology* , octubre/noviembre, 8-13. Recuperado el 16 de enero de 2003 en Academic Search Premier.
- Boiko, B. (2001b). *Content management bible*. , New York: Hungry Minds.
- Borgman, C. L. (2000). *From Gutenberg to the global information infrastructure : access to information in the networked world*. Cambridge, Massachussets: MIT Press.
- Dowler, L. (1997). *Gateways to knowledge : the role of academic libraries in teaching, learning and research*. Cambridge, Massachussets: MIT Press.
- Harold, E.R. (2001). *XML Bible* (2a. Ed.) New York: Hungry Minds.
- Kenney, A.R. y Rieger O.Y. (1999). *Moving theory into practice : digital imaging for libraries and archives*. . Mountain View, California: Reserch Library Group.
- Kim, Hun-Hee y Choi, Chang-Seok. (2000). XML: How it will be applied to digital library systems. *The Electronic Library*. 18(3), 183-189,
- Lafuente López, R. Y Garduño Vera, R. (2001). *Lenguajes de marcado de documentos digitales de carácter bibliográfico*. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas.
- Malo Alvarez, S. y Fortes Besprovani, M. (1999). *México frente a la era de la información*. México: Academia Mexicana de Ciencias.

- Miller, D.R (2000, Verano). XML: Libraries' strategic opportunity. *Library Journal*. [Net Connect], 18-22.
- Morrison, M. et. al. (2000). *XML al descubierto*. Madrid: Prentice Hall.
- Rogers, M. (2002, Otoño). LC gets a new content management system. *Library Journal*, 6-7.
- Sitts, M.K. (2000). *Handbook for digital projects : a management tool for preservation and access*. Andover, Massachusetts: Northeast Document Conservation Center..
- Tennant, R. (2001, 15 de marzo). XML: The digital library hammer. *Library Journal*, 30-32.
- Sánchez, A. y Fernández, M.L.(2000). Bibliotecas digitales en México. *Boletín de la Sociedad Mexicana de Ciencia de la Computación*. 1(2):13-16.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, Massachusetts: MIT Press.
- Witten, I.H. y Bainbridge, D. (2003). *How to build a digital library*. Amsterdam: Morgan Kaufman.
- Wu, Y.D. y Liu, M. (2001). Content management and the future of academic libraries. *The Electronic Library*: 19(6): 432-439. Recuperado el 16 de enero de 2003 en Academic Search Premier.