

LA VISIBILIDAD Y LOS RECURSOS EDUCATIVOS:
TÉCNICAS DE
RECOLECCIÓN DE
METADATOS MEDIANTE
OAI-PMH Y LOM

Las reflexiones acerca de los nuevos cauces por los que habrá de transcurrir la información digital están ligadas al concepto de *Web semántica* o *Web 2.0*. Existe ya un importante aparato teórico que sustenta la opinión fundada de que este tipo de páginas son el futuro de la Red. Sin embargo, es necesario dar el paso hacia el plano práctico. En el presente artículo, se expone un modelo posible para la creación de una Web semántica de documentación y recursos educativos.



Domenico Fetti:
Retrato de un profesor

En los últimos años se está produciendo un crecimiento exponencial de la aparición en la red de recursos educativos digitales. Si esto ha sido así hasta fechas bien recientes, la aparición de la denominada Web 2.0¹ no hará sino incrementar esa tendencia. Es cierto que se ha dicho que la Web 2.0 es más un caso de lo que los americanos llaman *hispter*. . . es decir publicidad sobre un concepto, que no es en realidad tangible, pero en primer lugar, y como se verá más adelante, ello no es completamente cierto y, por otra parte, es indudable que la idea en si misma constituye un importantísimo catalizador y un espaldarazo para subir a la red todo tipo de recursos y entre ellos los recursos educativos.

Pocas veces habrá sido más acertada la revista *Time* que cuando señaló como *Man of the year* de 2006 a... *You*, indicando así ese protagonismo social que gracias a la red de redes, a la Web, al correo electrónico, a Internet, y en suma a procesos derivados de las 'tagging technologies' como los blogs, los podcasts, el **social bookmarking**, las **wikis**, **RSS** (y otras formas

¹ Un explicación bastante detallada de los orígenes, significado y perspectivas de la Web se encuentra como es lógico en la Wikipedia, que es justamente uno de los elementos esenciales de la Web 2.0, http://en.wikipedia.org/wiki/Web_2



Biblioteca Nacional de España

de publicación y redistribución de contenidos), el **social software**, las Web **APIs**, las **Web standards** y los servicios web, el usuario final, cada uno de los individuos que se conectan a la red, *You*, contribuye de forma decisiva a incrementar las redes sociales y el contenido de la información y el conocimiento en la red. Pocas veces digo, aunque esta elección se deriva de aquella otra ocasión del año 1982 cuando un PC fue escogido como el *hombre del año*. Si se piensa que apenas han transcurrido 25 años entre una fecha y otra, nos daremos cuenta de la enorme velocidad que ha adquirido este fenómeno, que es mucho más que tecnológico, y que se ha convertido en social. En todos los ámbitos se está dando esta gran proliferación de recursos digitales y también en el sector educativo. Ante este fenómeno no es posible quedar indiferente o inactivo, pero mucho menos ponerle puertas al campo. Quizá puedan establecerse dos estrategias bien diferenciadas aunque quizá en un momento dado acaben convergiendo en la previsible quizá irónicamente Web 3.0

Me refiero por un lado a la *recolección de metadatos*² y por otro lado a la *Web semántica*. En el recientemente celebrado I Simposio Internacional de Documentación Educativa (Palma de Mallorca, 14-16 de febrero de 2007)

2 Sobre este tema y desde un punto de vista bibliotecario me permito aconsejar la lectura de *Recursos digitales: un reto para las bibliotecas nacionales* aparecido en *Las bibliotecas nacionales del siglo XXI*: Valencia, 18-21 de mayo de 2005 / *Jornadas sobre Bibliotecas Nacionales*; coordinación, Everilda Ferriols, Miguel C. Muñoz y Romá Seguí. - Valencia : Biblioteca Valenciana, 2006. - 278 p. ; 25 cm. - (Col·lecció Professional). - ISBN 84-482-4469-9 y accesible en la red desde hace ya tiempo en la dirección: <http://bv.gva.es/documentos/Ponencias/Agenjo.pdf>

SIDOC 2007 conviene mencionar el trabajo publicado como ponencia tercera y titulada *Las nuevas tecnologías en los procesos de cooperación documental. Aumento de la visibilidad en la red* presentada por Laureano Felipe Gómez Dueñas, investigador de la Universidad de Lasalle y coordinador del área de Desarrollo Tecnológico de la Biblioteca Nacional de Colombia. Sin duda, don Laureano Felipe Gómez Dueñas es conocedor de las técnicas que yo voy a mencionar más adelante y que el enumera de forma suficientemente desarrollada quizá a falta de un nivel mayor de profundidad, pero que en cualquier caso constituye una versión excelente de lo que es posible llevar a cabo y que propone efectuar en Redined³.

Sin embargo, tal vez por la bibliografía que maneja, su excelente trabajo de síntesis no parece darle una especial importancia, que a mí me parece imprescindible para el alcance de este proyecto, a los recursos didácticos. También parece que el enfoque es absolutamente teórico y se echa en falta alguna idea práctica para poner en marcha algunas de las muchas y buenas ideas que el señor Gómez Dueñas expresa en su trabajo.

Hago aquí un inciso para señalar que discurrir sobre la Web semántica no es el objetivo principal de estas páginas aunque sí considero realizarlo en otro complementario más o menos *hegelianamente* con vistas a obtener una idea sintética de actuación.

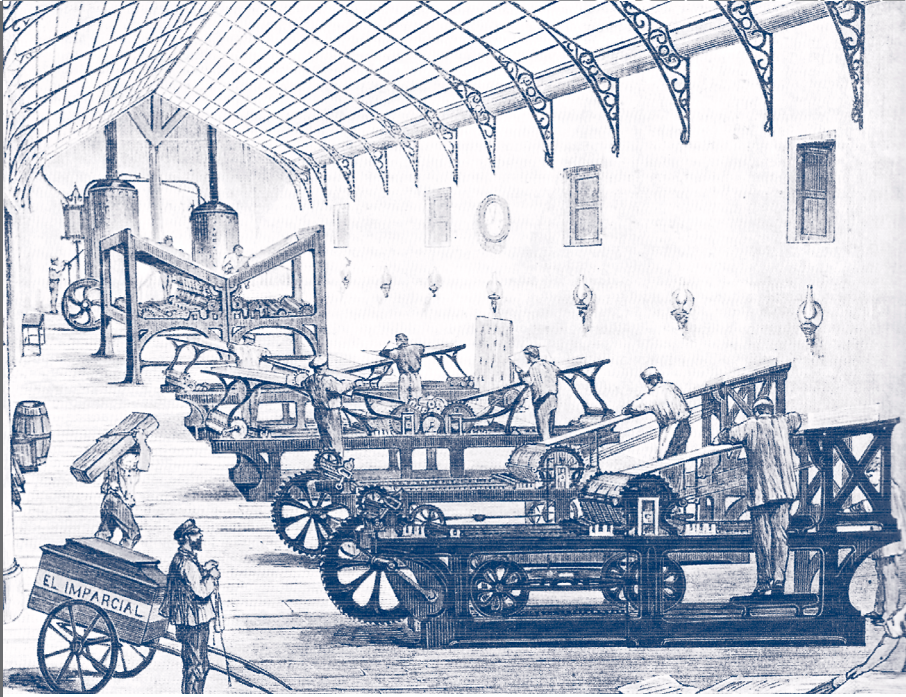
El esquema de funcionamiento de la recolección de metadatos se asienta sobre dos tipos de actores, los proveedores de contenido que son quienes crean y difunden los recursos educativos y quienes, para poder participar en el sistema, deben abrir sus recursos educativos, en la forma de un servicio Web de repositorios, a los recolectores de metadatos.

La ventaja que posee el uso de software para la recolección de metadatos (metadata harvesting) es que su diseño, programación, implementación y puesta en producción es relativamente simple. Puede hablarse de una horquilla de precios que oscila entre los 60.000 y 90.000 euros, sin hacer referencia naturalmente ni al equipo físico ni a las comunicacio-

nes, ni al apoyo probablemente de un técnico de sistemas y dos bibliotecarios o 3 documentalistas para mantener actualizada la base de datos. No es ese, por lo tanto, el mayor coste que ha de abordarse para un proceso de recolección de metadatos sino uno de muy distinto precio, incluso inferior para cada una de las instalaciones, que naturalmente ha de multiplicarse por el número de proveedores de información recolectable. El esquema de funcionamiento de la recolección de metadatos se asienta sobre dos tipos de actores, los proveedores de contenido que son quienes crean y difunden los recursos educativos y quienes, para poder participar en el sistema, deben abrir sus recursos educativos, en la forma de un servicio Web de repositorios, a los recolectores de metadatos. Éstos son el otro pilar que se encarga de recolectar los metadatos de los recursos educativos servidos por los proveedores de datos. Habitualmente los recolectores disponen de un servicio web de búsqueda, consulta y visualización de la información almacenada. Estos servicios de búsqueda pueden estar dotados además de interfaces específicamente diseñadas para una comunidad, proporcionando puntos de búsqueda apropiados a la recuperación de recursos educativos.

Con ello me estoy refiriendo, lógicamente, al modelo de recolección de metadatos definido por Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) que requiere la implantación de repositorios OAI

³ <http://www.redined.mec.es/>



Imprenta

(proveedores de datos) que utilicen, además del mínimo establecido por el protocolo *Dublin Core* sin cualificar, DCMI 15836:2003, los esquemas de metadatos LOM *Learning Object Metadata*. En efecto, mientras que *Dublin Core* es una descripción entre bibliográfica y documental muy sencilla con solo 15 elementos a cumplimentar, los LOM tienen una estructura más compleja. Una aplicación mínima del protocolo OAI requeriría tan sólo la utilización de *Dublin Core* y los repositorios así definidos podrían ser recolectados por cualquier software que cumpla el mismo protocolo, es decir, funcionarían correctamente en un entorno generalista. Sin embargo, la implantación de este modelo para la comunidad educativa, con sus requisitos específicos, necesitaría de una estructura de metadatos más precisa, tal y como viene definida por el esquema IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata, que pudiera canalizar la reutilización de objetos de aprendizaje, la localización de la existencia de estos objetos o recursos y facilitar la interoperabilidad entre los sistemas de gestión de aprendizaje. Además, pasar de la estructura LOM a *Dublin Core* es un proceso que puede parametrizarse por medio de XML, incluyendo las tablas de identificadores y descriptores que puedan utilizarse.

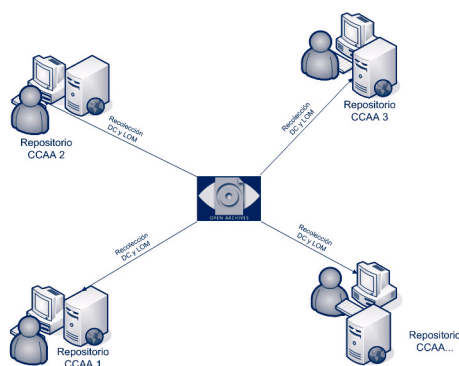
Como ya se ha dicho, son muchísimos los recursos didácticos que se van creando, pero también es evidente que resultan difíciles no ya de recuperar, sino de localizar, de conocer su mera existencia. Por lo tanto, aquí se propone una iniciativa clásica basada en el denominado modelo *seed money* para la creación de estas estructuras de información (LOM) como puntos de acceso a los recursos educativos y repositorios OAI donde puedan en primer lugar cargarse y en segundo lugar recuperarse desde uno (o varios) recolectores OAI-PMH. Puede verse el Recolector y Directorio de Colecciones y Recursos Digitales del Ministerio de Cultura, puesto en producción hace justamente un año cuando se escriben estas páginas, en la dirección. <http://roai.mcu.es>

Atendiendo los aspectos prácticos, se deberían incentivar mediante una política de subvención, que ya se ha visto que no han de ser excesivamente cuantiosas dado el coste de un repositorio entre los 60.000 y los 18.000

De esta forma, invirtiendo una cantidad de dinero no excesivamente grande podrán multiplicarse por toda España y a través de las CCAA un gran número de repositorios, como mínimo uno por Comunidad, donde se editarían e integrarían los recursos educativos compuestos por el propio recurso y sus correspondientes metadatos.

euros, para que aquellas instituciones que deseen contar con uno se comprometan a poner a disposición de la red virtual los correspondientes equipos y comunicaciones, y, sobre todo, y esta es la dovela en la que se apoya todo el arco, en la incorporación de metadatos LOM a los recursos didácticos que se vayan creando o se hayan creado ya. De esta forma, invirtiendo una cantidad de dinero no excesivamente gran-

Esquema de funcionamiento del modelo



de podrán multiplicarse por toda España y a través de las CCAA un gran número de repositorios, como mínimo uno por Comunidad, donde se editarían e integrarían los recursos educativos compuestos por el propio recurso y sus correspondientes metadatos. Ni que decir tiene que, una vez instalado, el coste de mantenimiento del sistema es mucho menor con un incremento notabilísimo de la visibilidad, y de la productividad, de los recursos educativos.

En España existe una importante e interesante actividad en ese sentido, pero que se mueve más en el campo del desarrollo teórico que en el de la aplicación práctica, siendo ésta relativamente sencilla desde el punto de vista informático y documental y no excesivamente costosa desde el punto de vista presupuestario. Probablemente el coste mayor y tal vez el mayor esfuerzo descansa en la logística de la operación que pudiera ser lo que faltase en el excelente trabajo publicado en la REDOC 29, 4 (septiembre-diciembre 2006) pp. 555-571 por Miguel Ángel Marzal García-Quismondo, Javier Calzada-Prado y Aurora Cuevas Cerveró que con el título *Desarrollo de un esquema de metadatos para la descripción de recursos educativos: el perfil de aplicación MIMETA* han definido un perfil fruto de los trabajos de los proyectos DOTEINE e IACORIE que han desembocado, como digo, en el perfil aplicación MIMETA.

El trabajo citado, fuertemente académico, repasa de forma generalmente correcta la bibliografía existente y los proyectos realizados hasta el momento, no sólo para la creación de iniciativas similares a la que aquí se viene proponiendo y muy en particular por el *IEEE Standard for Learning*

Object Metadata, tras tratar previamente del *Dublin Core*. Llama un tanto la atención que, a pesar de los proyectos que citan no mencionen la relación de OAI con LOM y sí el borrador de la norma ISO / IEC 19778, en fase de preparación aparecido en diciembre de 2005 y que al parecer trata de combinar, cito textualmente "la expresividad de LOM con la simplicidad de Dublin Core". Este borrador de norma está constituido por 52 elementos agrupados en 6 categorías. La relación directa de OAI con LOM puede verse en las *Guidelines for encoding identifiers in Dublin Core and IEEE LOM metadata* que se encuentra accesible en la dirección: <http://www.ukoln.ac.uk/metadata/dcmi-ieee/identifiers>

Yo soy de la opinión de que en lugar de crear una norma mixta sería más útil mantener el *Dublin Core* -sin cualificar- como elemento básico de la recuperación lo que permite ya de entrada recuperar todos los materiales bibliográficos, y el LOM como un esquema añadido tal y como recomiendan las directrices para la creación de repositorios OAI que, además del repetidamente citado *Dublin Core*, incorporan otras como la EAD, MARC21, E-Prints, (también interesantísimo para estos propósitos), RSLP *collection description schema*, la UDDI/WSDL y *The branding schema*. Esta información está muy fácilmente accesible en el *Sitio* principal de OAI en la dirección:

<http://www.openarchives.org/OAI/2.0/guidelines-repository.htm>

El trabajo citado recoge un número considerable e implantaciones de estos estándares y también perfiles de implantación empleados en Bibliotecas Digitales educativas, de los cuales el más conocido para los lectores de este trabajo será sin duda el denominado perfil ARIADNE www.ariadne-eu.org, cuyo proyecto finalizó el año 2000, pero que, a través de una Fundación que lleva su nombre, se encarga de mantener en funcionamiento el *Search and Index Learning Object* (SILO).

No cabe duda que MIMETA y el modelo IACORIE de metadatos es un proyecto ambicioso, pero cuya practicidad pueda tal vez discutirse. En todo caso está al menos parcialmente reflejado y accesible en el portal Educamadrid en su formato público de contenidos, disponible en www.educamadrid.org. En este sentido es necesario estar atentos a iniciativas del tipo de OLCOS (Open E-Learning Content Observatory Services) que promueve la creación y reutilización de Recursos Educativos Abiertos (Open Educational Resources OER), movimiento paralelo al de los Archivos Abiertos (Open Archives), iniciado en el año 2002 por la UNESCO (Forum on the Impact of Open Courseware for Higher Education in Developing Countries). Si, como ha ocurrido con las iniciativas de Archivos Abiertos, el recurso a un protocolo y unos metadatos comunes ha sido uno de los elementos de su éxito, es lógico pensar que en el entorno de los Recursos Educativos Abiertos ocurra otro tanto de lo mismo. Por ello, tengo que volver a insistir en las ventajas de la combinación de DC y LOM, y que el desarrollo de un perfil propio no me parece -al menos en principio- el mejor de los caminos.

Si, como ha ocurrido con las iniciativas de Archivos Abiertos, el recurso a un protocolo y unos metadatos comunes ha sido uno de los elementos de su éxito, es lógico pensar que en el entorno de los Recursos Educativos Abiertos ocurra otro tanto de lo mismo.

Muchos pasos han debido desandarse y muchos procesos de conversión han debido realizarse en el campo bibliotecario como consecuencia de la aplicación de formatos de codificación bibliográfica (metadatos al fin y al cabo) tan dispares como IBERMARC y CATMARC, que han dificultado el intercambio de información durante los últimos veinte años. En el campo de los archivos pueden encontrarse ejemplos de lo mismo, así mucho me temo que acabará ocurriendo lo mismo con la EAG (Guía de Archivo Codificada) de uso en España, en lugar de la Encoded Archival Description de amplio uso internacional. Yo sería partidario más bien de la adopción estricta de las dos normas y su aplicación rigurosa tal y como se acordó en México en la denominada *Resolución Manzanillo 2006* dentro de la Conferencia internacional sobre DC y aplicaciones de metadatos, 3-6 octubre 2006.

Por el contrario, la actuación que yo propongo de la construcción de un recolector de metadatos DC y de metadatos LOM (no la mezcla de DC y LOM) y repositorios y registros educativos que utilicen LOM, mediante pequeñas ayudas que permitan la inicial puesta en mantenimiento de una red educativa, me parece mucho más realista, puesto que al fin y al cabo no se trata tanto de alimentar a la *red* sino de alimentar al *propio repositorio* sin tener que estar pendientes de cuándo el recolector reunirá virtualmente los contenidos. Además el modelo propuesto, por su sujeción a ambas normativas, permitirá que los metadatos cosechados y almacenados en una base de datos sean, a su vez, recuperables por protocolos de búsqueda y recuperación de información y de localización de metadatos. Así, pueden ser aplicables normas como SRU (Search and Retrieve via URL, (<http://www.loc.gov/standards/sru>) u OpenURL (http://www.niso.org/committees/committee_ax.html) y otras técnicas como la muy eficaz para este caso de *agregadores de contenidos RSS*, todas ellas basadas en XML. Todo ello, acabará produciendo un crecimiento *iterativo* y la ordenación dinámica de un sector particularmente entusiasta y que, ateniéndose a unas reglas relativamente simples, puede potenciar enormemente su repercusión mediante la accesibilidad y visibilidad en la red, al alcance de todo el mundo, disponible en cualquier momento, y en cualquier lugar; ideal democrático de todo sistema de información verdaderamente destinado a la sociedad. ●