

## LA BIBLIOTECA DIGITAL REVISITADA

Luis Ángel García Melero  
[luisangelgarciamelero@gmail.com](mailto:luisangelgarciamelero@gmail.com)

### RESUMEN

Este artículo pretende exponer algunas normas e iniciativas sobre bibliotecas digitales redactadas o actualizadas desde 2001, año en el que Ernesto García Camarero y el autor del artículo publicaron el libro *La biblioteca digital*. Se analizan brevemente las últimas tendencias en la edición digital y de la industria de la digitalización y se repasan los conceptos de colección digital, biblioteca digital y repositorio. Se exponen las ideas básicas del movimiento acceso libre y se describe el protocolo OAI-PMH al que se considera el mecanismo más adecuado para la creación de proyectos cooperativos de bibliotecas digitales y para facilitar la recuperación de los objetos electrónicos. Tras describir los dos principales recolectores españoles (*Recolecta* e *Hispana*), se exponen algunos estándares para la descripción de objetos (Dublin Core, MARCXML, MODS, MADS), la organización de metadatos descriptivos, administrativos y estructurales y la transferencia de información (METS), la conservación de los documentos digitales (PREMIS) y para la consulta y recuperación de la información (SRU).

PALABRAS CLAVE: Biblioteca digital; Repositorio; Edición electrónica; Normas.

### INTRODUCCIÓN

Hace ocho años que Ernesto García Camarero y yo escribimos el libro *La biblioteca digital*<sup>1</sup> En la introducción le calificábamos de obra *de alborada*, de amanecer de un nuevo tiempo en el contexto de la edición digital y de la biblioteconomía. También decíamos que evitábamos dar preponderancia a las tecnologías para evitar una rápida obsolescencia de sus contenidos. La suerte de aquél nuestro hijo no ha sido del todo mala: se han vendido ejemplares y ha sido citado, principalmente por autores extranjeros. Después, he escrito en foros profesionales<sup>2</sup> sobre la conservación del patrimonio digital, he impartido algún curso sobre las bibliotecas digitales<sup>3</sup> y he

<sup>1</sup> García Camarero, Ernesto *La biblioteca digital* / Ernesto García Camarero, Luis Ángel García Melero. – Madrid: Arco/Libros, 2001. – 380 p. – (Instrumenta bibliológica). – ISBN 84-7635-486-X

<sup>2</sup> *La conservación del patrimonio digital*. – En: Acta de la IX Reunión de Trabajo de la Asociación Española de Bibliografía: Madrid, 5 y 6 de febrero de 2004. – Todavía en prensa, aunque, mucho me temo, que todos los textos de esta reunión quedarán inéditos.

*Algunas iniciativas relacionadas con la recopilación y conservación del patrimonio digital*. – En: Boletín de la Confederación de Asociaciones de Archiveros, Bibliotecarios, Museólogos y Documentalistas, v. 54 (2004), nº 1-2, enero-junio, p. 429-456

*La recopilación y conservación del patrimonio digital*. – En: Jornadas sobre bibliotecas nacionales: las bibliotecas nacionales del siglo XXI, Valencia, 18-21- de mayo de 2005 / coordinación, Everilda Ferriols, Miguel C. Muñoz y Romà Seguí. – Valencia: Generalitat Valenciana, Conselleria de Cultura, Educació i Esport ; Biblioteca Valenciana, 2006. – P. 115-147

*El archivo de documentos personales digitales*. – En: [dosdoce.com](http://dosdoce.com), diciembre 2005

<sup>3</sup> Curso de bibliotecas digitales organizado por el Instituto de Administración Pública Adolfo Posada, celebrado en Oviedo los días 6 y 7 de noviembre de 2008

propuesto a mis superiores jerárquicos proyectos que no han tenido éxito debido a los medios necesarios y a mi planteamiento en un contexto demasiado administrativo, carente de sensibilidad suficiente por la cultura y la ciencia en su acepción más amplia<sup>4</sup>. Con ello, quiero decir que he procurado mantenerme al día en este periodo de tiempo leyendo bibliografía profesional e informándome de los proyectos internacionales y nacionales que iban surgiendo, aunque he de reconocer que no ha resultado fácil debido a la superabundancia, casi inflación, de literatura.

Cuando la presidenta de ANABAD, D<sup>a</sup> Pilar Gallego, me escribió un correo electrónico solicitando mi colaboración en un número monográfico del *Boletín de la ANABAD*, pensé que podría resultar de interés escribir sobre los cambios habidos en los años transcurridos desde la edición del libro *La biblioteca digital*.

## TENDENCIAS EN LA EDICIÓN

Desde entonces se han producido o consolidado algunas tendencias que ya apuntábamos. Sin duda alguna, la primera es que la edición creada en formato digital ha aumentado de forma exponencial. Prosigue el debate sobre la pervivencia del libro impreso. Pienso que todavía continuará durante bastante tiempo, entre otros motivos, porque la evolución tecnológica no es sincrónica en todos los países, ni siquiera dentro de uno sólo, a pesar de los avances habidos en los programas desarrollados para que los miembros de la sociedad sepan manejar las tecnologías de la información. El continuo abaratamiento de los sistemas informáticos, su mayor capacidad y las investigaciones para conseguir que un solo dispositivo aúne todas las facilidades (comunicación oral y escrita, acceso a los recursos disponibles en las redes de telecomunicaciones, captura de imágenes fijas y en movimiento, de reproducción de sonido, etc.), cuyo principal exponente son los teléfonos móviles de última generación, tampoco están consiguiendo esta sincronía. Resulta lógico si lo meditamos un poco, pues, aun dentro de una misma nación desarrollada, conviven diferentes generaciones con distintas educaciones, formaciones, concepciones culturales e ideológicas y con capacidades adquisitivas dispares. Conste que estas palabras no suponen una crítica negativa o una visión derrotista. La sociedad siempre ha sido, es y será así y en este hecho radica el origen de muchos conflictos generacionales, sociales e internacionales.

Volviendo a la edición digital, el debate sobre la supervivencia de lo analógico ha ocasionado que los editores publiquen versiones impresas y numéricas de un mismo texto, permitiendo la descarga sin coste o con determinadas condiciones de la versión en línea. El motivo es que aún no ven del todo claro cuál será el futuro, a la vez que siguen con atención los progresos tecnológicos y los estudios de mercado de los e-books, como el *Amazon Kindle 2*, *iLiad Reader*, *Hanlin eReader V3*, *Sony Reader PRS 700*, *CyBook 3* o *Papyrus*, entre otros,

---

<sup>4</sup> En el 2005, cuando trabajaba en la Sección de Publicaciones del Ministerio de Administraciones Públicas (MAP), propuse que se creara una biblioteca digital en la que se almacenaran, conservaran y se pusieran a disposición de la ciudadanía las publicaciones en línea editadas por la Administración General del Estado (AGE) que escaparan a los controles de los números internacionales normalizados (ISBN e ISSN) y al depósito legal. A pesar de la consultoría contratada posteriormente por el Ministerio de la Presidencia con la empresa Digibis, el proyecto no ha llegado a nacer. La causa de postergarlo fue el elevado nivel de inversión de recursos económicos, humanos y tecnológicos. Por esta razón se han perdido y se está perdiendo una parte del patrimonio digital español creado por la administración central.

Junto a este hecho cada vez avanzan más la autoedición, el acceso libre y gratuito a contenidos que, como sucede con un número cada vez mayor de revistas científicas, están avalados por consejos de redacción que supervisan su calidad. A ello ha contribuido la expansión del software libre y de conceptos como el *copy left*, las licencias *Creative Commons* y de documentación libre de GNU. Uno y otros han dado lugar a nuevos conflictos en los países desarrollados como la pugna del software libre y de los programas gratuitos contra las aplicaciones informáticas propiedad de las empresas que las han desarrollado y los derechos de autor contra el dominio público. En el fondo, ambas son manifestaciones de una misma postura ideológica.

Las administraciones públicas no son ajenas a estos cambios en el sector de la edición. Por una parte, tienen que defender el sector editorial y la propiedad intelectual mediante leyes propuestas por ellas, debatidas y aprobadas por los parlamentos. Por otra promueven la edición electrónica y la reducción de las tiradas de los impresos bajo el pretexto del ahorro presupuestario, pero, al parecer, sin analizar las posibles repercusiones en los sectores de las artes gráficas y de la distribución, cada vez más abocados a una reconversión integral. Aumentan las publicaciones oficiales en línea que escapan a otras disposiciones, también promulgadas a su iniciativa, como son los números internacionales normalizados y el depósito legal. Así pues, las distintas administraciones públicas tienen que hacer verdaderos equilibrios para atender los intereses de las industrias culturales, que mueven un importante volumen de dinero y de puestos de trabajo, y las demandas de una parte creciente de la sociedad que pide libre acceso y circulación del conocimiento y entretenimiento por medio de las tecnologías de la información y comunicaciones y que está ocasionando una revolución igual o mayor a la que se originó con la invención de la imprenta.

## LA CONVERSIÓN DE LA CULTURA ANALÓGICA

Estas mismas tecnologías se han manifestado como unas herramientas formidables para la democratización de la cultura al tiempo que permiten la conservación de los soportes originales que contienen los principales hitos de la humanidad. Antes sólo unos pocos podían leer e investigar directamente sobre piezas únicas<sup>5</sup> como los manuscritos de la Biblia, los Beatos medievales, las obras de Shakespeare, Cervantes, Goethe..., los cuadros de Botticelli, Velázquez, los grabados de Goya, las partituras de Mozart, Beethoven, Chopin, etc., los documentos de los archivos históricos. Ahora, si el examen inmediato de los originales no resulta imprescindible para el estudio, multitud de personas pueden leerlos o contemplarlos en la pantalla de un ordenador conectado a Internet rompiendo las barreras del espacio y del tiempo.

Ha surgido una auténtica industria de la digitalización cuyo objetivo es convertir las manifestaciones científicas y culturales pintadas, escritas o impresas de forma analógica a una secuencia de dígitos, de ceros y unos, a una codificación digital que representa, en muchos casos, con igual o más calidad (permiten suprimir algunas “manchas”) al original. Es una industria transversal, que implica a muchos subsectores (el fotográfico, el informático, el documental, el telemático y el editorial, al menos), y, por supuesto, a los centros que conservan las obras (archivos, bibliotecas, museos, filmotecas...)

---

<sup>5</sup> García Camarero, Ernesto *Redes y dueños del conocimiento = Networks and owners of knowledge*. En: *Banquete Nodos y redes*. – Madrid: Turner 2009. - p. 119-130 También disponible en la URL <http://www.elgranerocomun.net/Redes-y-Duenos-del-Conocimiento.html> [Consulta realizada el 30 de julio de 2009]

La digitalización ha generado nuevos intereses. Por una parte, las bibliotecas (instituciones en las que me centraré) han visto en la digitalización una nueva forma de hacer frente al cambio que se estaba produciendo en sus servicios a raíz de la expansión de Internet; por otra, empresas como Google<sup>6</sup>, han hallado un nuevo mercado o una manera de originar más expectativa distribuyendo y facilitando el acceso a los materiales digitalizados. Las posibilidades de este nuevo negocio han suscitado asuntos relacionados con los derechos de autor e, incluso, se han iniciado demandas judiciales para no perder beneficios económicos. Las bibliotecas, como la Biblioteca Nacional de España, se han percatado de que pueden lograr algún ingreso mediante la impresión bajo demanda de sus libros escaneados por una empresa, con la excusa del coste que puede suponer al usuario final la obtención de una copia impresa en su sistema informático.

Es cierto que la digitalización de los bienes muebles culturales ha contribuido a la conservación de los originales y a una mayor difusión de los mismos de forma que cualquier individuo pueda consultarlos, pero también es innegable que está surgiendo una nueva industria con sus aspectos positivos y negativos.

## LAS COLECCIONES DIGITALES

En 2001, año de edición de *La biblioteca digital*, aparte del Archivo de Indias, modélico ya entonces para los archivos históricos, citábamos 12 proyectos de colecciones y bibliotecas digitales españolas. Sin duda alguna habría muchos más, pero nos centramos en los que considerábamos más destacados. En julio de 2009, *Hispana*, el recolector del Ministerio de Cultura, reseña 376<sup>7</sup> colecciones; el número de recursos recopilados por *Recolecta*<sup>8</sup>, el recolector de la ciencia abierta española, iniciativa de la Red de Bibliotecas Universitarias Españolas (REBIUN) y la Fundación para la Ciencia y la Tecnología (FECYT), ascendía a 177 recursos recolectados (revistas y depósitos digitales) en esa misma fecha.

El reciente informe titulado *Situación de los repositorios en España*<sup>9</sup> especifica que en España existen 44 repositorios<sup>10</sup> según el directorio *BuscaRepositorios*<sup>11</sup>, que se han ido

<sup>6</sup> Recuérdese que la iniciativa de Google de digitalizar colecciones casi enteras de libros existentes en bibliotecas norteamericanas, provocó la alarma del presidente francés Jacques Chirac. Éste propuso que la Unión Europea crease una gran biblioteca digital europea para difundir la identidad cultural de los Estados miembros y para hacer frente a los intereses comerciales e intelectuales de Google. Surgió así *The European Library* (TEL) y sus subproductos *Europeana* y *Europeana local*. Sin embargo, estos proyectos no han impedido que las bibliotecas europeas firmaran acuerdos con esta empresa para digitalizar sus fondos bibliográficos, distribuirlos y comercializarlos total o parcialmente. En España, al menos, la Biblioteca de la Universidad Complutense de Madrid, la Biblioteca de Catalunya y la Biblioteca Nacional de España han alcanzado algún tipo de acuerdo con Google.

<sup>7</sup> <http://roai.mcu.es/es/inicio/inicio.cmd> [Consulta efectuada el 31 de julio de 2009]

<sup>8</sup> <http://www.recolecta.net/buscador/index.jsp> [Consulta efectuada el 31 de julio de 2009]

<sup>9</sup> *Situación de los repositorios institucionales en España* : informe 2009 / Remedios Melero, [et al.]. – [S.l.] : Acceso Abierto, 2009. – 54 p. : gráf. Disponible en la dirección electrónica [http://www.accesoabierto.net/sites/default/files/Informe2009-Repositorios\\_0.pdf](http://www.accesoabierto.net/sites/default/files/Informe2009-Repositorios_0.pdf) [Consulta efectuada el 3 de agosto de 2009]

<sup>10</sup> *Directory of Open Access Repositories (OPENDOAR)* URL: <http://www.opendoar.org/find.php> [Consulta realizada el 3 de agosto de 2009] reseña uno más: 45 repositorios españoles en total. *Registry of Open Access Repositories (ROAR)* URL <http://roar.eprints.org/index.php> [Consulta efectuada el 3 de agosto de 2009] describe 55.

desarrollando desde el año 2001 en el que se creó el primero de nuestro país, el TDX, dedicado a las tesis doctorales. Este informe, redactado a partir de las respuestas facilitadas a una encuesta distribuida entre 104 instituciones, concluye, entre otros aspectos, que los principales tipos de objetos digitales que albergan son artículos y tesis doctorales a texto completo disponibles en acceso abierto. Las áreas del conocimiento sobre la que versan son, por este orden, humanidades y ciencias sociales, ciencias de la salud, ingeniería, ciencias naturales y artes plásticas. El archivo lo efectúa de forma delegada una biblioteca o servicio especializado cuando se trata de material digitalizado y, en segundo lugar, los autores de las tesis doctorales y de materiales docentes. Las plataformas informáticas empleadas para su constitución son de software libre<sup>12</sup> y el sistema de metadatos más empleado es *Dublin Core* con calificadores. Las cifras expuestas hablan por sí solas del avance habido en estos 8 años.

En los párrafos anteriores han aparecido términos (colección digital, biblioteca digital, repositorio, archivo abierto) que pueden confundir a alguien no muy versado en esta materia. En las siguientes líneas procuraré definir y clarificar estos conceptos, aunque estas definiciones son particulares y están sujetas a debate y revisión.

Una colección digital es un conjunto de documentos electrónicos cuyas descripciones bibliográficas residen en las bases de datos que albergan los registros de los restantes materiales impresos que constituyen el fondo de una biblioteca. Los objetos digitales están almacenados en un ordenador servidor y están enlazados con las descripciones bibliográficas mediante hipervínculos que facilitan la visualización, la lectura en pantalla o descarga de documentos electrónicos en un puesto de trabajo informático. La colección digital está relacionada con el concepto de biblioteca híbrida, que son las que todavía predominan.

Una biblioteca digital es una institución que selecciona, adquiere, cataloga y custodia documentos textuales, icónicos, gráficos o sonoros, creados digitalmente o convertidos a formato numérico y los pone a acceso público a través de las redes de telecomunicaciones. Lo sustancial, pues, es que su “fondo” está constituido exclusivamente por documentos electrónicos en línea, con las ventajas e inconvenientes que los caracterizan, y que está albergado en un sistema informático.

Un repositorio (término procedente del latín *repositorium* cuya traducción al castellano sería almacén o alacena) es un lugar donde se guarda algo, según la vigésimo segunda edición del Diccionario de la Lengua Española. Si extendemos este significado a las tecnologías de la información, *un repositorio, depósito o archivo es un sitio centralizado donde se almacena y mantiene información digital, habitualmente bases de datos o archivos informáticos*<sup>13</sup>. En el ámbito de la biblioteconomía digital un repositorio es una plataforma informática (equipos físicos y conjunto de programas) dedicada a albergar, conservar y difundir documentos digitales producidos por los miembros de una institución académica (universidad y/o centro de investigación) o que

---

<sup>11</sup> *BuscaRepositorios* Dirección electrónica: <http://www.accesoabierto.net/repositorios/> [Consulta realizada el 3 de agosto de 2009]

<sup>12</sup> Los programas más empleados son DSpace seguido a gran distancia por GNU Eprints, Digbib, Fedora, Greenstone, Etd-db, OJS y CDS. A ellos hay que añadir D-NET desarrollado en el marco de DRIVER y que próximamente se implementará en *Recolecta*

<sup>13</sup> Wikipedia Dirección electrónica: <http://es.wikipedia.org/wiki/Repositorio> [Consulta efectuada el 4 de agosto de 2009]

trabajan en un ámbito del conocimiento<sup>14</sup>. Un repositorio tiene una serie de características que se exponen en una política. En ella se contemplan aspectos tales como el registro de las personas que pueden realizar el archivo de los objetos digitales (el autor, una biblioteca o una tercera persona en nombre del creador), el almacenamiento de los materiales, los tipos y formatos de documentos que se pueden depositar, los metadatos o descripciones de los mismos, las normas que se deben utilizar, la accesibilidad de las publicaciones, la sujeción a la legislación de propiedad intelectual, la conservación de los objetos, etc.<sup>15</sup>

## EL MOVIMIENTOS *OPEN ACCESS*<sup>16</sup>

Los repositorios están relacionados con el denominado movimiento *Open Access* o acceso abierto. Éste es otro de los riesgos a los que se enfrenta la edición convencional. Surgió precisamente debido a los crecientes costes de la publicación, las limitaciones de los derechos de autor y a la necesidad de agilizar la comunicación de los conocimientos científicos y técnicos empleando las tecnologías de la información y las telecomunicaciones. Este movimiento se materializa en la publicación en línea de revistas, actas de congresos, libros y en la constitución de repositorios de dominio público, en los que los contenidos se pueden descargar de forma gratuita. Es también una forma de *amortizar la inversión de investigación de un país al poner a disposición de los usuarios el libre acceso a la documentación derivada de la misma*, como se expresa en la presentación del sitio web Acceso Abierto.

Aunque algunos teóricos de este pensamiento fijan sus orígenes en la década de los años 60 y 90 del pasado siglo, coincidiendo con la aparición del hipertexto y de la WWW, su desarrollo no comienza hasta el 2001 y las posteriores declaraciones de la comunidad científica internacional y de las denominadas instituciones de la memoria (archivos, bibliotecas y museos)<sup>17</sup>.

Los repositorios también han interesado a la Comisión Europea y a la legislación española en materia de ciencia y tecnología.

---

<sup>14</sup> Clifford A. Lynch, en su artículo *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*, publicado en ARL, nº 226 (February 2003), p. 1-7, define un repositorio institucional como *un conjunto de servicios que ofrece la Universidad a los miembros de su comunidad para la dirección y distribución de materiales digitales creados por la institución y los miembros de esa comunidad. Es esencial un compromiso organizativo para la administración de estos materiales digitales, incluyendo la preservación a largo plazo cuando sea necesario, así como la organización y acceso o distribución.*

<sup>15</sup> Para iniciarse en esta materia se recomienda la lectura de las siguientes publicaciones: *Cómo crear un Repositorio Institucional : Manual LEADIRS II* / por Mary R. Barton y Margaret M. Waters. – 2004-2005. Disponible en la URL <http://www.recolecta.net/buscador/documentos/mit.pdf> [Consulta realizada el 4 de agosto de 2009]; *Informe APEI sobre acceso abierto* / por Julio Alonso, Imma Subirats y M<sup>a</sup> Luis Martínez Conde. – 2008. – URL: <http://eprints.rclis.org/15107/1/informeapeiaccesoabierto.pdf> [Consulta efectuada el 4 de agosto de 2009] y *La Iniciativa de Archivos Abiertos (OAI) : situación y perspectivas en España y Latinoamérica* / Gema Bueno de la Fuente [et al.]. – Bogotá : Rojas Eberhard, 2007. – 350 p.

<sup>16</sup> Entre los sitios web dedicados a los repositorios y al acceso abierto a la literatura científica, sugerimos la consulta del sitio web español denominado Acceso abierto, URL <http://www.accesoabierto.net/>

<sup>17</sup> La IFLA publicó en 2004 *The IFLA Statement on Access to Scholarly Literature and Research Documentation* URL: <http://ifla.queenslibrary.org/V/cdoc/open-access04.html> y, en mayo de 2009, ha editado la declaración titulada *Enhancing the Debate on Open Access: A joint statement by the International Federation of Library Associations and Institutions and the International Publishers Association* URL: [http://www.ifla.org/files/hq/documents/enhancing-the-debate-on-open-access\\_final-20090505.pdf](http://www.ifla.org/files/hq/documents/enhancing-the-debate-on-open-access_final-20090505.pdf) [Consultas realizadas el 4 de agosto de 2009]

El Séptimo Programa Marco de la Comisión Europea (FP7), que abarca el periodo 2007-2013, ha puesto en marcha un proyecto piloto que facilitará el acceso libre a los resultados de las investigaciones financiadas por la Unión Europea editados en publicaciones con comités de revisión. La salvedad, que ha ocasionado diversas críticas, es el establecimiento de un periodo de retención de 6 a 12 meses, dependiendo del área de conocimiento. Se supone que esta dilación permitirá a los editores científicos obtener un rendimiento por el capital invertido en su publicación. Transcurrido este periodo, los beneficiarios de las subvenciones comunitarias están obligados a facilitar el libre acceso a los manuscritos finales o a los artículos aparecidos en revistas científicas y técnicas incluyéndolos en un repositorio. Esperemos que sea una medida coyuntural, hasta que se alcance un acuerdo definitivo con el sector editorial que se encuentra en una tesitura similar a la que debieron experimentar los copistas de manuscritos con la progresiva expansión de la imprenta. Lo que importa es que con éste y con el proyecto DRIVER<sup>18</sup>, también financiado por la Comisión Europea, el acceso libre a los contenidos científicos cada vez se va abriendo más paso en la comunidad académica internacional.

El Ministerio de Ciencia y Tecnología del Gobierno español ha elaborado un Borrador de Anteproyecto de Ley de la ciencia y la tecnología<sup>19</sup> cuyo artículo 33 está dedicado a la *Publicación en acceso abierto*. En él se especifica que se *impulsara el desarrollo de repositorios, propios o compartidos, de acceso abierto a las publicaciones de su personal de investigación*. También establece que los investigadores cuya actividad haya sido financiada con cargo a los Presupuestos Generales del Estado, publicarán una versión digital de los resultados y conclusiones que se difundirá a través de repositorios de acceso abierto. Esperemos que este borrador, tras superar los procedimientos legislativos pertinentes, sea aprobado por el Congreso y el Senado y que, al menos, se mantengan los conceptos recogidos en el citado artículo.

## **UN SISTEMA MUY NUEVO Y, AL MISMO TIEMPO, MUY VIEJO**

Desde la antigüedad los bibliotecarios estamos acostumbrados a organizar la información contenida en los distintos soportes que han ido apareciendo a lo largo de la historia: tablillas cuneiformes, papiros, manuscritos en pergamino o papel, impresos, microfilmes, microfichas, discos de vinilo, casetes, discos compactos, vídeos, etc. Un objeto digital en línea no deja de ser un soporte que puede residir en nuestro sistema informático local o en uno remoto.

Los bibliotecarios, o como se nos quiera denominar, también sabemos que una biblioteca, o como se la desee designar, es un sistema de comunicación (también el cuerpo humano lo es), que pone en contacto la edición mundial en cualquier medio con una comunidad de usuarios que precisa formarse, informarse o distraerse.

---

<sup>18</sup> DRIVER (*Digital Repository Infrastructure Vision for European Research*) tiene como finalidad la creación de una infraestructura de conocimiento en el área de la investigación europea que facilite la distribución de contenidos científicos mediante una red virtual de repositorios institucionales de los estados miembros de la Unión Europea. Para ello ha publicado unas directrices a cuya normativa y metodología se deben adecuar los distintos repositorios. Se puede obtener más información en el sitio web de DRIVER cuya URL es <http://www.driver-repository.eu/>

<sup>19</sup> Disponible en la URL [https://lcyt.fecyt.es/wp-content/uploads/2009/02/borrador-cero-alcyt-11\\_02\\_09.pdf](https://lcyt.fecyt.es/wp-content/uploads/2009/02/borrador-cero-alcyt-11_02_09.pdf) [Consultado el 4 de agosto 2009]

Los bibliotecarios somos conscientes de que para llevar a cabo esta misión, se necesitan una serie de políticas (de adquisiciones, de catalogación y clasificación, de normas de acceso, de servicios bibliotecarios, de conservación...), unas normas que aseguren la uniformidad y calidad de nuestros productos (en la actualidad, bases de datos), de recursos humanos, técnicos, económicos y de un local donde podamos desarrollar nuestra actividad y que nos albergue a nosotros, al mobiliario preciso y al fondo o colección de documentos organizados.

Pueden cambiar los medios utilizados (el cálamo fue sustituido por la pluma estilográfica y ésta, por la máquina de escribir que, a su vez, fue remplazada por los ordenadores; al catálogo en forma de libro le siguió el catálogo en fichas y, a éstas, los registros albergados en un disco duro), adecuarse las políticas y las normas o cambiar la terminología, pero, al final, si lo analizamos bien, permanece la función esencial de la biblioteca y de los bibliotecarios: atender las necesidades informativas, formativas o recreativas de nuestros usuarios, ya sea un niño, estudiante, profesor, investigador o un alto cargo de una empresa pública o privada mediante, la organización de una colección de objetos que contienen la información precisa en cada caso. Heráclito y Parménides tenían razón los dos: en la vida, todo fluye, pero hay algo que permanece constante a lo largo del tiempo, que es lo que identifica a un ser, aunque vaya cambiando.

La anterior digresión viene a cuento de que lo expuesto en los párrafos anteriores sucede también con la biblioteca digital y los repositorios, aunque la propia naturaleza de los objetos digitales (obsolescencia de las tecnologías y de los formatos y “deterioro” de la información) impone ciertos condicionantes. A esa conclusión se llega al leer la bibliografía especializada y la norma internacional ISO 14721<sup>20</sup>. Esta especificación define lo que es un OAIS (Sistema Abierto de Archivo Información)<sup>21</sup>. En ella se describe un modelo genérico que sirve de referencia en el diseño y puesta en marcha de archivos, repositorios o bibliotecas digitales orientados a la preservación de documentos electrónicos. También es un patrón de referencia que permite evaluar los sistemas dedicados a la conservación digital.

El estándar ISO 14721 establece las responsabilidades, agentes, funciones, tipos de información, estrategias de preservación y los niveles de cooperación de los sistemas abiertos de información de archivo.

Las responsabilidades de un OAIS son el establecimiento de criterios de selección, la comprobación y aceptación de la información procedente de las entidades que la producen, si cumplen las reglas establecidas para su incorporación al sistema, la garantía de que la información conservada se pone a disposición de los usuarios de forma comprensible y la aplicación de políticas y procedimientos que aseguren la preservación de la información y la difusión de copias autenticadas.

Los agentes que intervienen son los productores de la información (las personas o clientes que proporcionan los documentos a conservar), el personal del archivo encargado de realizar las funciones, los procesos y servicios y los usuarios que acceden al sistema para consultar los objetos conservados.

---

<sup>20</sup> *ISO 14721-2003 Open archival information system* : Reference model. – Genève: International Standard Organization, 2003

<sup>21</sup> *Organización formada por personas y sistemas que han aceptado la responsabilidad de preservar información y hacerla disponible para una determinada comunidad de usuarios.*

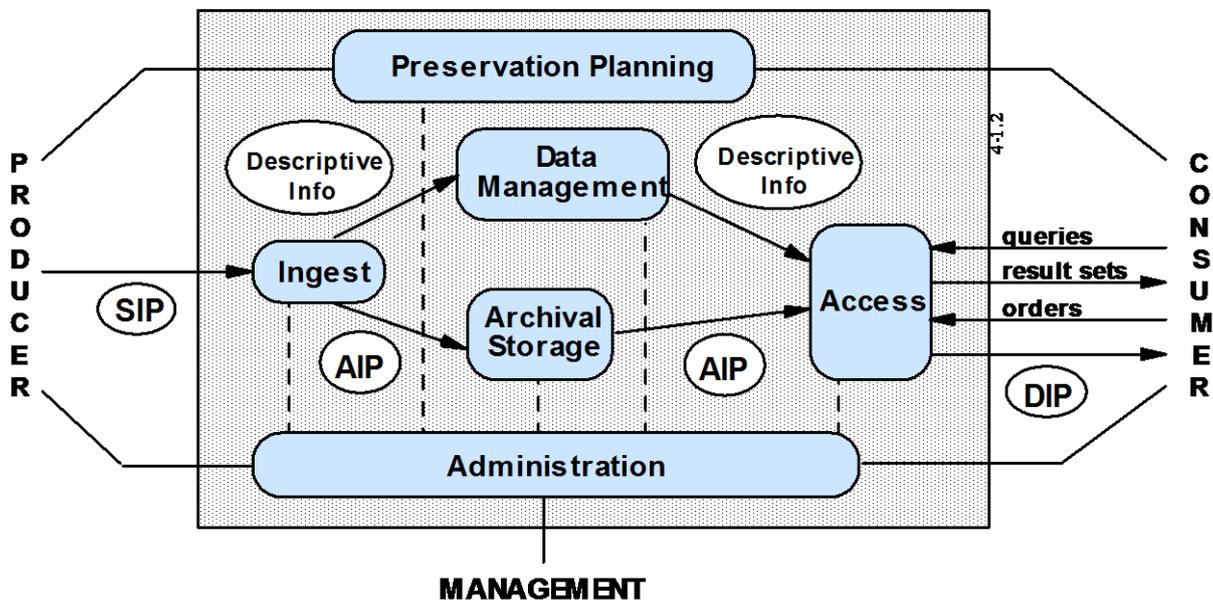
Las funciones que se llevan a cabo en el OAIS son la incorporación o ingreso (*Ingest*) de los paquetes de información, formados por el documento digital y sus metadatos, la gestión de los datos (*Data Management*), el almacenamiento del objeto con fines de conservación (*Archival Storage*), la recuperación de la información y el acceso a los objetos (*Access*), la planificación de la conservación (*Preservation Planning*) y la administración del sistema (*Administration*).

Además de la información descriptiva (*Descriptive Info*), se distinguen tres paquetes de información que circulan en las funciones básicas (*Ingest*, *Archival Storage* y *Access*) de un archivo abierto: información de entrega (*Submission Information Package* o *SIP*), información de archivo (*Archival Information Package* o *AIP*) y el paquete de información de difusión (*Dissemination Information Package* o *DIP*). El *SIP* es el remitido por los productores y su forma y contenido deben ser establecidos entre aquellos y el OAIS; el *AIP* es el resultante de la incorporación del objeto al archivo completado con las informaciones propias para la gestión de la información y el *DIP* es el que se remite y muestra al usuario de manera que pueda distinguir la información solicitada de forma comprensible.

ISO 14721 también especifica la estructura que deben tener los paquetes de información y que consta de cuatro elementos: Información del contenido (*Content Information* o *CI*), Información de preservación / descripción (*Preservation Description Information* o *PDI*), Información de empaquetado e Información descriptiva. La *CI* contiene en uno o varios archivos informáticos el objeto a conservar así como la información complementaria que los usuarios necesitan para comprender el origen, historia, significado, etc. del mismo. La *PDI* recoge metadatos necesarios como los identificadores internos o externos, las relaciones del objeto con su entorno y con otros objetos, información de procedencia (creación del objeto, acciones realizadas para preservarlo, cambios de custodia, etc.) e indicación de integridad (datos que permiten comprobar la autenticidad e integridad del objeto, como las marcas digitales). La información de empaquetado incluye datos que agrupan los detalles contenidos en *CI* y *PDI*. Por último, la información descriptiva son metadatos que facilitan la identificación, búsqueda y recuperación del objeto.

El Modelo de referencia de un OAIS contempla distintas estrategias para preservar los materiales digitales como la transformación del formato del objeto, el rejuvenecimiento y preservación de los soportes, la emulación de los programas informáticos para acceder a los datos y la conservación del código fuente de las aplicaciones informáticas que ofrecen acceso a los mismos.

Por último, el estándar que nos ocupa también se plantea los requisitos para que se puedan consultar y acceder a los datos existentes en distintos archivos abiertos o repositorios de objetos digitales. Así mismo, establece cuatro tipos de sistemas de conservación de documentos electrónicos de acuerdo con las posibilidades de colaboración de un OAIS con otro: independencia (no existe ningún tipo de interacción con otros OAIS); cooperación (un OAIS actúa como usuario de otro); recursos compartidos (existen acuerdos entre OAIS para ello) y federados (varios OAIS sirven a una comunidad global a la que se ofrecen herramientas de búsqueda comunes. Se añadirá a este modelo un catálogo común que permita buscar los objetos gestionados por distintos OAIS).



Esquema funcional de un OAIS<sup>22</sup>

## LA FORMA MÁS RACIONAL DE CONSTITUIR BIBLIOTECAS DIGITALES

El movimiento de acceso abierto y la norma ISO 14721 exponen conceptos, respectivamente, para la captura, conservación y difusión gratuita de la bibliografía científica y para diseñar procedimientos y estructuras organizativas de bibliotecas digitales, pero se precisa de una herramienta que la lleve a cabo y, sobre todo, que racionalice los esfuerzos desarrollados en la creación de repositorios de objetos electrónicos para evitar duplicaciones innecesarias. Esa herramienta es el protocolo OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*)<sup>23</sup> cuya versión 2.0 se puso en funcionamiento en junio de 2002, aunque los orígenes de la investigación se remontan a la década de 1990.

Aun a riesgo de ser muy sintético, podemos definir este protocolo como un programa informático que, con una frecuencia determinada, establece una conexión telemática con una o varias bases de datos accesibles en la web, busca en ellas los registros existentes que responden a unos parámetros (fechas de creación o modificación, asuntos de los que tratan, etc.), captura los metadatos o descripciones que satisfacen estos criterios y las descarga en una base de datos centralizada o catálogo colectivo sin ninguna mediación humana. Los objetos digitales, es decir, los documentos, permanecen en la base de datos consultada enlazados con la base de datos central mediante una dirección electrónica contenida en el registro. De esta forma un usuario, al buscar en el sistema de

<sup>22</sup> Ilustración tomada del libro *Reference Model for an Open Archival Information System (OAIS) : Recommendation for Space Data System Standards*. – Washington : Consultative Committee for Space Data System, 2002. – 148 p. que sirvió de base para la elaboración de la norma ISO 14721 2003

<sup>23</sup> *Open Archives Initiative* URL <http://www.openarchives.org/> Consulta realizada el 10 de agosto de 2009

recuperación de la información del catálogo colectivo las publicaciones existentes sobre un determinado argumento de búsqueda, puede visualizarlas o descargarlas pulsando sobre el vínculo.

En la estructura OAI-PMH hay dos clases de participantes o agentes: los proveedores de datos y los proveedores de servicios. Los primeros son sistemas informáticos, que almacenan objetos digitales en bases de datos, archivos abiertos o repositorios y los registros que los describen. Estos sistemas soportan el protocolo y se exponen al programa recolector. Los proveedores de servicios son también sistemas informáticos que disponen de un programa de aplicación denominado recolector (*harvester*, en inglés) que recopila los registros de los proveedores de datos y los almacena en una base de datos central. También cuentan con un motor de búsqueda para efectuar consultas y con un interfaz de usuario para la realización de las búsquedas y la prestación de otros servicios. Los agregadores son otro tipo de participantes que se encargan de capturar las descripciones de distintos proveedores y los proporciona a otros proveedores de servicios completándolos con algún valor añadido.

Este protocolo se basa en una serie de normas internacionales de hecho y de derecho como son el protocolo http empleado para servir peticiones a proveedores de datos mediante los métodos GET y POST; XML<sup>24</sup> para codificar los registros de metadatos y las respuestas de los proveedores de datos; *Dublin Core*<sup>25</sup>, formato de metadatos básico que deben soportar los proveedores de datos e ISO 8601<sup>26</sup> para las fechas expresadas en el formato UTC (Coordinated Universal Time).

OAI-PMH también especifica cómo deben ser los identificadores únicos de los ítems, la estructura de los registros y las seis operaciones o verbos (*Identify*, *ListMetadataFormats*, *ListSets*, *ListIdentifiers*, *ListRecords* y *GetRecord*) que el programa de recolección puede solicitar al proveedor de datos. En todos los casos, las respuestas que el proveedor de datos envía al programa de recolección serán documentos XML.

El protocolo OAI-PMH se ha extendido internacionalmente y sus especificaciones se contemplan en los proyectos de la Comisión Europea dedicados a la creación de una red de repositorios de la bibliografía científica comunitaria. Es una tecnología, concebida inicialmente para la búsqueda y recuperación de *e-prints* producidos por la comunidad académica de uno o varios ámbitos del conocimiento, que se ha extendido a todas las instituciones de la memoria, pues facilita la constitución de una red de repositorios temáticos o institucionales y de bibliotecas digitales que convergen en una base de datos central que posibilita la recuperación de la información y de los documentos donde quiera que se encuentren. Ayuda, pues, a racionalizar los recursos humanos, técnicos y económicos. En mi opinión, todavía adolecen de cierta insuficiencia en la recuperación de la información por materias, lo que se debe a la pluralidad de asuntos y a la diversidad de tipos de centros que participan. Poco a poco, mediante ontologías o mediante macro tesauros se podrán ir solventando estas disfunciones.

---

<sup>24</sup> Extensible Markup Language (XML) 1.0. -5th ed.- W3C Recommendation 26 November 2008. – URL: <http://www.w3.org/TR/xml/> [Consulta efectuada el 12 de agosto de 2009]

<sup>25</sup> ISO 15836 2009 *Information and documentation : The Dublin Core metadata element set*. – 2<sup>nd</sup>. Ed. - Genève : The International Standard Organization, 2009. – 5 p.

<sup>26</sup> ISO 8601 2004 : *Data elements and interchange formats : Information interchange: Representation of dates and times*. – 3<sup>rd</sup>. ed. – Genève : The International Standard Organization, 2004. – 33 p.

## LOS RECOLECTORES ESPAÑOLES

En España disponemos de dos grandes recolectores de objetos digitales a los que nos hemos referido anteriormente: *Recolecta* e *Hispana*. El primero está más orientado a la literatura científica y académica expresada, principal pero no exclusivamente, en *e-prints*, artículos de revistas y tesis editadas en acceso abierto resultantes de proyectos financiados con fondos públicos<sup>27</sup>. Además, según, se expone en la página de acogida de su sitio web, tiene como objetivos 1) el impulso, apoyo y coordinación del desarrollo cohesionado de la red interoperable de repositorios digitales para el acceso abierto, difusión y preservación de los resultados de la investigación científica en España; 2) el desarrollo de servicios y funcionalidades de valor añadido sobre los resultados de esta investigación para los investigadores y el público en general y 3) la coordinación de la agenda nacional con el desarrollo internacional de la infraestructura global de repositorios digitales científicos como parte integrante del Espacio Europeo de Investigación.

*Hispana*<sup>28</sup>, uno de los resultados del Grupo de Trabajo de Colecciones Digitales del Consejo de Cooperación Bibliotecaria del Ministerio de Cultura, es un directorio y recolector de recursos digitales. Es también el punto de acceso común a los objetos digitales de bibliotecas y archivos españoles. Constituye, pues, un agregador de contenidos de las bases de datos de recursos digitales conforme a la Iniciativa de Archivos Abiertos (OAI) promovida por la Unión Europea.

Sus participantes son todas las instituciones españolas de las distintas administraciones públicas y de las fundaciones y entidades sin ánimo de lucro que disponen de colecciones digitales conforme a dicha iniciativa.

El directorio reseña los proyectos de digitalización que se están llevando a cabo en España y facilita el acceso a las colecciones, cuya cifra asciende a 376<sup>29</sup>. De esta forma se puede cumplir con el objetivo de permitir la coordinación de los proyectos que se están realizando para evitar la digitalización dos o más veces de la misma obra.

El recolector permite la consulta, recuperación de la información, visualización de las referencias en múltiples formatos y de los objetos digitales<sup>30</sup> albergados en 102 repositorios que satisfacen las especificaciones OAI y que aparecen marcados con el icono OAI.

Su amplitud es mayor que la de *Recolecta*, pues comprende más tipos de documentos, incluidos materiales de algunos archivos e impresos digitalizados. La temática es también más amplia, pues recopila los metadatos de los objetos de los repositorios de *Recolecta* más los de otras áreas del conocimiento producidas o conservadas en diversas clases de bibliotecas y entidades culturales. No solo cumple con todos los requisitos de OAI y con las Recomendaciones de la Comisión Europea sobre digitalización y accesibilidad en línea del material cultural, si no también con casi todos los estándares

<sup>27</sup> *Recolecta* contiene 570.241 objetos digitales procedentes de unos 177 recursos.

<sup>28</sup> *Hispana (Biblioteca digital)* En: [http://es.wikipedia.org/wiki/Hispana\\_\(biblioteca\\_digital\)](http://es.wikipedia.org/wiki/Hispana_(biblioteca_digital)) [Consulta realizada el 14 de agosto de 2009].

<sup>29</sup> Dato facilitado por la Coordinación de Hispana el 14 de agosto de 2009

<sup>30</sup> Según la Coordinación de la iniciativa Hispana, el número de objetos ascendía a 974.370 a 14 de agosto de 2009, si bien esta cifra se incrementa a diario por lo que debe ser considerada coyuntural.

internacionales, por no decir todos, relacionados con la digitalización y las bibliotecas digitales. Próximamente se va a poner en funcionamiento una nueva versión que incluirá las últimas especificaciones del proyecto comunitario DRIVER.

Sin embargo, tanto en *Recolecta* como en *Hispana* se echa en falta la disponibilidad de publicaciones en línea de acceso gratuito editadas por fundaciones no lucrativas con interés científico. Éste, junto con los documentos oficiales de las distintas administraciones públicas españolas disponibles en Internet, es un campo de actuación que queda por abarcar. Tal vez procediera que los Ministerios correspondientes (Ciencia e Innovación y Cultura) informaran a estas dos clases de instituciones de la existencia del libre acceso, de la OAI y les prestase el apoyo necesario para que creasen repositorios conformes a OAI-PMH y facilitasen la recolección de los metadatos de sus publicaciones. Sin duda, redundaría en una mayor visibilidad de sus productos culturales y en una utilización más intensiva por los usuarios, que a menudo no conocen su existencia.

Otro posible ámbito de actuación sería el sector editorial. La tecnología OAI-PMH será, sin duda, la más adecuada para una futura recopilación del depósito legal de los documentos creados digitalmente en las condiciones de acceso que especifiquen toda clase de editores. Las asociaciones de editoriales podrían constituir un repositorio proveedor de datos que agrupase los e-books o libros electrónicos publicados en España y que lo abrieran a los proveedores de servicios y a los programas recolectores de *Recolecta* y/o *Hispana*. También en este caso volvemos a lo mismo: se precisará una campaña informativa, otra de sensibilización y prestar los apoyos tecnológicos y formativos pertinentes para su implantación. Los editores podrían indicar las condiciones de acceso (gratuito u oneroso) en los metadatos administrativos, pero habría un punto único de difusión y obtención de sus documentos electrónicos. De esta manera, el usuario podría tener una visión más exhaustiva del material digital, retrospectivo y actual, existente sobre un autor o asunto.

## **OTRAS NORMAS DESARROLLADAS DESDE 2001**

Además de las ya mencionadas, se han desarrollados otros muchos estándares desde la publicación de *La biblioteca digital* en el año 2001 a los que me voy a referir en este epígrafe. No obstante, quiero advertir que esta relación es selectiva, son las que considero más relevantes con el inconveniente que siempre implica un criterio personal. Dejo, pues, abierta esta rúbrica para que se puedan agregar otras normas por estudiosos de la difusión de contenidos digitales.

En una materia que evoluciona con tanta rapidez como es la creación de repositorios de objetos electrónicos y en su conservación, se pone de manifiesto la necesidad de adoptar estándares de hecho y códigos de buenas prácticas antes de que se conviertan en normas de derechos avaladas por instituciones dedicadas a la actividad de normalización, como es International Standard Organization (ISO). El trabajo de estos organismos requiere una colaboración entre las partes implicadas, que, en ocasiones, ralentiza la redacción, adopción y edición de especificaciones técnicas. Otras veces, estas instituciones sólo confirman los estándares efectuados por centros nacionales de normalización o por iniciativas que han demostrado su utilidad y aplicación por una determinada comunidad profesional.

El lenguaje de marcas XML<sup>31</sup> es una norma anterior a estos últimos ocho años, pero que se impone en este periodo. Durante él se convierte en el sistema preferido para codificar otras especificaciones dedicadas a transcribir el texto completo de los documentos en línea, a reseñar los metadatos descriptivos, administrativos y de preservación de los objetos, al intercambio de información y a la recuperación de la información. Esto se debe, entre otras cualidades, a la capacidad de XML de independizar los datos de los fabricantes de hardware y software.<sup>32</sup>

También nos referimos en nuestro libro, a *Dublin Core*<sup>33</sup>, cuya primera edición se publicó en 1998 y que en el 2003 se consagró como norma ISO 15836. La segunda edición de este estándar apareció en febrero del presente año<sup>34</sup>. En el 2004 se dio a conocer un perfil para la aplicación de *Dublin Core* a proyectos y aplicaciones relacionadas con las bibliotecas<sup>35</sup>. En concreto, este perfil pretende 1) servir de formato de intercambio entre varios sistemas que utilizan diferentes formatos de metadatos; 2) ser empleado para la recolección de metadatos procedente de recursos propios y ajenos al dominio bibliotecario; 3) ayudar a la creación de registros catalográficos a partir de recursos procedentes de diversos sistemas; 4) transferir registros en formato MARC a otras comunidades mediante una conversión a *Dublin Core* y 5) permitir la adquisición de recursos al descubrir metadatos realizados por personas que no son bibliotecarios y que utilizan *Dublin Core*.

Este sistema para describir toda clase de objetos, diseñado para que pudiera ser utilizado por los autores de los mismos o por personal con una formación específica, ha proseguido su avance y ha logrado asentarse como estándar que se debe incluir en el protocolo OAI-PMH para reseñar los parámetros que identifican un objeto y para facilitar la recuperación de la información. Sin embargo tiene un poderoso competidor en la comunidad bibliotecaria que dispone de centenares de millones de registros bibliográficos codificados en MARC y que ha sabido reaccionar y adaptarse a las nuevas tecnologías. De hecho, a mediados de la década de los años 1990, The Library of Congress realizó el esfuerzo de adoptar el formato MARC a SGML, pero este último fue remplazado por XML. Tras la armonización de los formatos MARC utilizados por las bibliotecas nacionales de Estados Unidos, Gran Bretaña y Canadá, que dio como resultado el MARC21, se acometió su adecuación a XML.

En 2002 se publicó el MARCXML, término usado para los registros MARC codificados en XML diferenciándolos de los codificados conforme a ISO 2709. MARCXML tiene como finalidad la codificación de cualquier registro MARC (bibliográfico, autoridades, etc.) en XML usando un único esquema y tiene como objetivos: 1) Facilitar la representación de registros MARC existentes en formato XML mediante un proceso de conversión; 2) Facilitar la captura e integración de registros MARC en los procesos de recolección característicos de los proyectos OAI; 3) Permitir la descripción inicial de

<sup>31</sup> W3C REC-xml-19980210 Extensible Markup Language (XML) 1.0

<sup>32</sup> Eito Brun, Ricardo *Lenguajes de marcas para la gestión de recursos digitales* : aproximación técnica, especificaciones y referencia. – Gijón : Trea, 2008. – 469 p. : il. – (Biblioteconomía y administración cultural ; 184)

<sup>33</sup> Dublin Core Metadata Initiative <http://dublincore.org/> [Consulta realizada el 17 de agosto de 2009]

<sup>34</sup> ISO 15836 2009 *Information and documentation* : *The Dublin Core metadata element set*. – Genèva : The International Standard Organization, 2009. – 5 p.

<sup>35</sup> Library Application Profile <http://dublincore.org/documents/library-application-profile/index.shtml> [Consulta efectuada el 17 de agosto de 2009]

recursos utilizando MARC y XML y 4) Facilitar la inclusión de metadatos MARC en el texto completo de documentos XML.

En MARCXML no se plantean en ningún momento cambios en lo referente a nombres y etiquetas de campos, subcampos e indicadores. Se trata de utilizar un nuevo modelo de codificación y transferencia física del registro que no afecta a su organización lógica ni al significado de los campos que lo constituyen. Lo único que se elimina es el directorio, que deja de ser necesario pues en XML se utilizan las etiquetas de inicio y de fin de elemento.<sup>36</sup>

Para facilitar su rápida implantación y la conversión de los registros existentes estructurados en ISO 2709, la iniciativa MARCXML Framework y The Library of Congress han desarrollado una serie de utilidades y de programas como la conversión de registros MARC 2709 a MARCXML, la conversión entre registros MARCXML, Dublin Core y MODS; la presentación y visualización de registros MARCXML mediante hojas de estilo XSLT; el desarrollo de aplicaciones de conversión o transformación de registros MARCXML a otros formatos de metadatos; utilidades para la presentación y visualización de registros MARCXML en un formato fácilmente legible tipo HTML, PDF, etc. y la realización de aplicaciones para validar documentos MARCXML frente al esquema o para cualquier otro tipo de operación.<sup>37</sup>

En 2002, el mismo año en que se publicó el MARCXML, The Library of Congress Network Development and MARC Standards Office, desarrolló MODS<sup>38</sup> (*Metadata Object Description Schema*). Se trata de un esquema para un conjunto de elementos bibliográficos que puede ser utilizado para varios propósitos y, especialmente, para fines bibliotecarios. Como esquema XML, está concebido tanto para contener y reagrupar datos de registros MARC21 existentes como para crearlos. Podemos afirmar que es un sistema intermedio entre *Dublin Core* y MARC. Contiene más información que el primero y resulta más sencillo que el segundo para el usuario final. Su principal inconveniente radica en la pérdida de alguna información en la conversión bidireccional MARC – MODS – MARC. Es decir: si un registro codificado en MARC21 se convierte a MODS, no puede volver a ser convertido a MARC21 en su totalidad sin alguna pérdida de especificidad en las etiquetas o sin pérdida de datos.

---

<sup>36</sup> ISO 25577 2008 *Information and documentation : MarcXchange*. – Genève : The International Standard Organization, 2008. – 15 p.

Esta norma, equivalente a la misión de ISO 2709 pero en el entorno XML, especifica los requisitos de un formato de intercambio basado en XML de registros bibliográficos y de otros tipos de metadatos. Describe una estructura diseñada principalmente para la comunicación entre sistemas de proceso de datos y que también puede ser empleada como formato de proceso dentro de un sistema. No define la longitud del contenido de los registros ni asigna unos valores a las etiquetas, indicadores o identificadores cuyas especificaciones corresponde a un formato de implementación. ISO 25577 se puede utilizar para 1) representar un registro MARC completo o un conjunto de registros MARC en XML; 2) para la descripción del recurso original en XML; 3) como una extensión del sistema METS del que trataremos más adelante; 4) para el intercambio de registros MARC en XML; 5) la transferencia de registros MARC en servicios Web como SRU; 6) la transmisión de datos del editor; 7) como un formato temporal en todo tipo de manejo o transformación de datos (conversión, publicación, edición y validación) y 8) para metadatos en XML que pueden ser empaquetados en un recurso electrónico.

<sup>37</sup> MARCXML [www.loc.gov/standards/marcxml/](http://www.loc.gov/standards/marcxml/) [Consulta realizada el 17 de agosto de 2007] En esta dirección se puede consultar el esquema, la DTD, y distintas utilidades y programas

<sup>38</sup> *Metadata Object Description Schema (MODS)* URL: <http://www.loc.gov/standards/mods/mods-overview.html> [Consulta realizada el 18 de agosto de 2009]

En 2004 la misma Oficina de The Library of Congress desarrolló *Metadata Authority Description Schema (MADS)*<sup>39</sup>, un esquema, basado en XML, para codificar registros de autoridades de personas, entidades, congresos, temas y subdivisiones de materia. MADS sirve de complemento a MODS al vincular las autoridades que aparecen en los registros bibliográficos basados en este último con los de autoridad contenidas en registros MADS mediante el atributo genérico de MODS xlink

Aun sin tener en cuenta sus orígenes, que se remontan a 1997, antes de algunas de las normas enumeradas más arriba, promovida por la Digital Library Federation y con el respaldo técnico de The Library of Congress, que se encarga de su mantenimiento, se desarrolló *Metadata Encoding & Transmission Standard (METS)*<sup>40</sup> en el año 2001. Según se indica en la página principal del sitio web y en la Introducción y tutorial, *METS* es una norma para codificar los metadatos descriptivos, administrativos y estructurales de objetos de una biblioteca digital utilizando el lenguaje XML y para expresar las relaciones entre estos tipos de metadatos. Además ofrece un estándar útil para el intercambio de objetos digitales entre repositorios y permite asociar objetos digitales con comportamientos o servicios. Por lo tanto, podemos afirmar que esta especificación técnica fue diseñada para: 1) permitir compartir los recursos necesarios para desarrollar servicios y/o herramientas para la gestión de la información y 2) facilitar el intercambio de materiales digitales entre instituciones.

Un documento METS consta de siete secciones (Cabecera; Metadatos Descriptivos; Metadatos Administrativos; Archivo; Mapa estructural; Enlaces estructurales Y Comportamientos) cuya descripción detallada se puede encontrar en las referencias indicadas.

En el 2002 se editaron las *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos*.<sup>41</sup> Estas normas fueron elaboradas por un grupo de expertos de la International Federation of Library Associations and Institutions (IFLA) y del International Council on Archives (ICA) a petición de la UNESCO. Como especifica su título, resultan de especial interés para todos aquellos proyectos e iniciativas destinadas a la conversión a soporte digital de los documentos en papel (manuscritos, libros impresos y fotografías) de dominio público custodiados por archivos y bibliotecas. No contemplan ni los

<sup>39</sup> *Metadata Authority Description Schema (MADS)*. URL: <http://www.loc.gov/standards/mads/> [Consulta efectuada el 18 de agosto de 2009]

<sup>40</sup> Las principales fuentes de información sobre esta norma son el sitio web oficial: *Metadata Encoding & Transmission Standard (METS)* URL: <http://www.loc.gov/standards/mets/> [Consulta llevada a cabo el 19 de agosto de 2009] También se debe consultar la siguiente publicación unitaria <*METS*>: *Metadata Encoding and Transmission Standard: Primer and Reference Manual*. - Washington: Digital Library Federation, 2007. - 133 p.; gráf. Disponible en la URL: <http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf>. Una visión sintética se puede obtener a partir de la lectura de *METS: introducción y tutorial* / traducción de Ricardo Eito Brun accesible en la URL [http://www.loc.gov/standards/mets/METSOverview\\_spa.html](http://www.loc.gov/standards/mets/METSOverview_spa.html) [Consulta realizada el 18 de agosto de 2009]

<sup>41</sup> *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos* / Grupo de expertos de IFLA e ICA ; traducidas por el Grupo de Trabajo de Colecciones Digitales de las Comunidades Autónomas y el Ministerio de Cultura. - Madrid : Ministerio de Cultura, 2005. - 183 p. Los apéndices de la edición española contienen borradores de pliegos de especificaciones técnicas y una relación de proyectos de digitalización existentes en España. También disponible en la URL: [http://travesia.mcu.es/documentos/pautas\\_digitalizacion.pdf](http://travesia.mcu.es/documentos/pautas_digitalizacion.pdf) Los apéndices están actualizados a mayo de 2009.

registros sonoros ni las películas. Un rasgo pertinente de estas directrices es que se refieren a la planificación y establecimiento de proyectos de digitalización y a los elementos claves que los componen: política de selección, conversión, programas de control de calidad, gestión de la colección, presentación y mantenimiento del acceso a largo plazo. Por lo tanto, no son únicamente unas normas técnicas, si no que contemplan los aspectos organizativos. Por último, recordemos que estas normas, junto con otras más específicas (protocolo OAI-PMH), son la base de las prescripciones técnicas de las ayudas concedidas desde el 2007 por la Dirección General del Libro, Archivos y Bibliotecas del Ministerio de Cultura a proyectos de digitalización del patrimonio español.

## **NORMAS SOBRE CONSERVACIÓN DE OBJETOS DIGITALES DESARROLLADAS DESDE 2001**

Hasta ahora me he referido a normas aparecidas desde 2001 dedicadas

- al diseño conceptual y organizativo de archivos abiertos encargados de conservar y difundir objetos digitales (OAIS);
- a la cooperación entre diversos repositorios (OAI) y el procedimiento para recolectar los metadatos descriptivos de los documentos en línea (OAI-PMH):
- a la descripción de esta clase de materiales con fines de identificación y recuperación de la información (*Dublin Core*, MARCXML, MODS y MADS);
- a la estructuración de los distintos tipos de metadatos (descriptivos, administrativos y estructurales), de los ficheros que almacenan el contenido de los objetos digitales, de las aplicaciones necesarias para su visualización y gestión y a la transferencia de este conjunto de informaciones a otros repositorios con fines de cooperación y racionalización de recursos (METS) y
- a las recomendaciones para emprender proyectos de digitalización de documentos en papel existentes en los fondos de archivos y bibliotecas (Directrices elaboradas por el Grupo de expertos de la IFLA y el ICA).

Falta tratar de los estándares sobre la conservación a medio o largo plazos de los recursos electrónicos en línea, ¿De qué servirían las anteriores especificaciones si un repositorio no realizara las acciones necesarias para mantenerlos en el tiempo? Las referencias a documentos en línea incluidos en este artículo no serían accesibles a corto plazo si un archivo abierto o biblioteca digital, en el que se han depositado, no realizara estos trabajos a partir de la información contenida en sus metadatos de conservación con lo que mi colaboración sería sólo parcialmente comprendida por cualquier lector. Ésta es la misión de PREMIS<sup>42</sup>

---

<sup>42</sup> *Sitio web de Preservation Metadata Implementation Strategies (PREMIS)*: URL <http://www.loc.gov/standards/premis/> [Consulta efectuada el 19 de agosto de 2009] Desde este sitio web se puede acceder y descargar el *PREMIS Data Dictionary for Preservation Metadata*. – Version 2.0. – Washington: The Library of Congress Network Development and MARC Standard Office, 2008. - 217 p. URL: <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf> fundamentan para comprender estos metadatos. También en él se alberga la siguiente introducción cuya traducción al castellano ha sido realizada por María Luisa Martínez-Conde, de la Subdirección General de Coordinación Bibliotecaria del Ministerio de Cultura español: Caplan, Priscila *Entender PREMIS*. – Madrid: Ministerio de Cultura ; Washington : Library of Congress Network Development and MARC Standard Office, 2009. – 30 p. Direcciones electrónicas en las que se pueden acceder a la versión en castellano: [http://www.loc.gov/standards/premis/UnderstandingPREMIS\\_espanol.pdf](http://www.loc.gov/standards/premis/UnderstandingPREMIS_espanol.pdf) y <http://www.mcu.es/bibliotecas/docs/MC/PREMIS/Creditos.pdf>

PREMIS es el resultado de la labor de un grupo de trabajo con el mismo nombre establecido en junio de 2003 por la OCLC y el RLG. En él se sintetizaron las experiencias de iniciativas previas para definir metadatos de preservación. Los principales objetivos del grupo de trabajo consistieron en: 1) definir un conjunto de metadatos para la preservación, que se pudiesen implementar y utilizarse en un sistema real; 2) evaluar alternativas para su codificación, almacenamiento e intercambio; 3) preparar programas pilotos para verificar las conclusiones alcanzadas y 4) analizar las ventajas de la cooperación en la creación de este tipo de metadatos.

Los principales hitos en el desarrollo de PREMIS fueron la publicación del informe *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community* (2004) y del *Data Dictionary for Preservation Metadata*, cuya primera edición se publicó en 2005. En él se describen los metadatos esenciales para la preservación digital válidos para cualquier tipo de formato de archivo informático. También recoge terminología, ejemplos y recomendaciones.

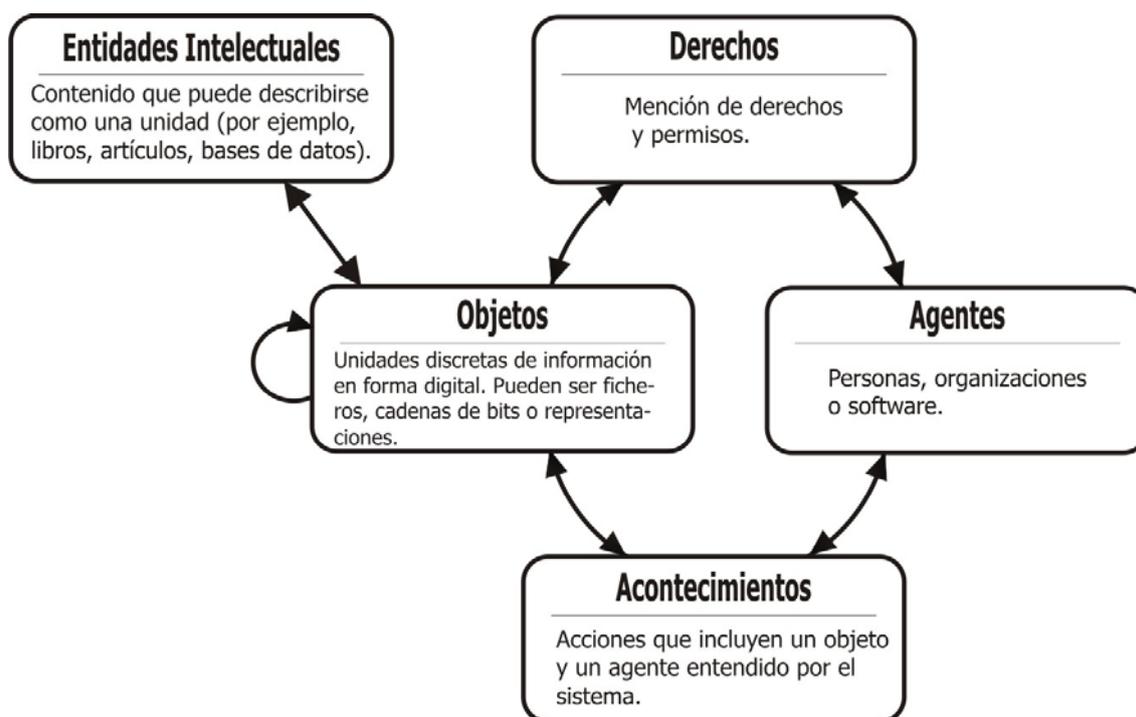
El diccionario PREMIS organiza los metadatos para la preservación en torno a cinco entidades:

1. **Entidad intelectual.** Se refiere a unos contenidos tratados como una unidad, por ejemplo una monografía.
2. **Objeto.** Corresponde a un documento en formato electrónico. Las entidades intelectuales están relacionadas con los objetos, pues son su materialización. Los objetos se clasifican en secuencias de bites (*bitstream*), archivos (*files*) procesables por un programa informático, archivos comprimidos, encriptados o incluidos dentro de otro archivo (*filestream*) y representaciones (*representation*). Estas últimas son el resultado de agrupar todos los archivos necesarios para ensamblar un documento con entidad propia y permitir una presentación completa de la entidad intelectual.
3. **Eventos.** Son las actividades que se realizan para la preservación de un objeto (registro y almacenamiento en un archivo, rejuvenecimiento, migración, etc.).
4. **Agentes.** Son las personas, entidades o aplicaciones informáticas relacionadas con un evento.
5. **Derechos.** Son los permisos para realizar ciertas acciones sobre un objeto para su preservación.

Las entidades están relacionadas entre sí. Los objetos también pueden estar conectados entre ellos. Un objeto puede ser el resultado de una transformación realizada sobre otro. Los objetos se pueden relacionar de forma estructural, como sucede en un documento html con imágenes o un sitio web compuesto por varias páginas.

Para registrar este tipo de relaciones, los objetos, eventos, agentes y derechos contarán siempre con un identificador único en el repositorio. Las relaciones se codificarán como parte de los metadatos reservados para este fin, asignándoles como valor los identificadores de las entidades relacionadas.

La siguiente figura muestra las entidades establecidas en PREMIS y las relaciones que pueden existir entre ellas.



Modelo de datos PREMIS<sup>43</sup>

En el diccionario de PREMIS no se habla de metadatos, sino de unidades semánticas. Éstas son los datos que debemos conocer para poder articular un programa de preservación; los metadatos son la forma en que los registramos.

PREMIS recoge metadatos para los Objetos, Eventos y Derechos, pero no para las entidades intelectuales ni los nombres de los agentes. Se considera que otros sistemas de metadatos descriptivos, como *Dublin Core*, MARC, MADS, etc., resultan válidos para estas entidades.

Un repositorio conforme con PREMIS debe registrar información para los metadatos considerados obligatorios en el diccionario. Se permite añadir metadatos adicionales, si no entran en conflicto con los descritos en Diccionario. No se exige ningún tipo de implementación particular para los metadatos (forma de almacenarlos, uso de XML, etc.). Sólo se formulan recomendaciones sobre la automatización de la extracción de los metadatos a partir de los objetos que se añadan al repositorio o sobre la utilización de vocabularios controlados para los valores de algunos metadatos. También se aconseja que los metadatos críticos se almacenen en sistema de bases de datos y en forma de datos asociados con los objetos digitales que describen.

Algunos de los metadatos de la entidad **Objeto** son la identificación y categoría del objeto, nivel de preservación, propiedades y características, denominación original,

<sup>43</sup> Tomado de: Caplan, Priscila *Entender PREMIS*. – Madrid: Ministerio de Cultura ; Washington : Library of Congress Network Development and MARC Standard Office, 2009. – 30 p.

almacenamiento, entorno informático, relaciones, etc.; de **Evento**, el identificador, tipo de evento, fecha y hora, información adicional, resultado de la acción, enlace con otros objetos; de la entidad; de **Agente**, la información de la persona, institución o programa informático que interviene en un evento y de la entidad **Derechos**, la información sobre un permiso asignado a uno o más objetos como los identificadores del permiso y del objeto, la referencia al documento en el que se concede la autorización y las acciones permitidas.

## **NORMAS SOBRE RECUPERACIÓN DE LA INFORMACIÓN DESARROLLADAS DESDE 2001**

También la consulta y recuperación de la información han tenido nuevos desarrollos desde el año 2001. Me estoy refiriendo a SRU (*Search/Retrieval via URL*)<sup>44</sup>, iniciativa denominada ZING (*Z39.50 International Next Generation*) hasta el año 2008. En síntesis, ZING y SRU pretenden la adaptación de la norma Z39.50 a las tecnologías web y al formato XML. Se basan en los protocolos SRU y POST y en el lenguaje de consulta CQL (*Contextual Query Language*). El primero define las operaciones que posibilitan un proceso de recuperación de la información entre aplicaciones informáticas. Las principales operaciones son *Search/Retrieve*, que precede a la ecuación de búsqueda, el número máximo de registros que se desea recuperar y el esquema XML de los registros que se reciban; *Scan* permite efectuar consultas a los ficheros índices; *Explain* establece los mensajes que se deben enviar desde el cliente al servidor para obtener información sobre las características de éste y su base de datos, así como el formato de las respuestas y *ZeeRex*, que proporciona una descripción de las bases de datos consultables.

POST es un protocolo que se emplea para el intercambio de peticiones entre ordenadores a través de la Web. Los parámetros y valores se envían como parte de la cabecera del mensaje http aunque permanecen ocultos al usuario.

CQL es el lenguaje en el que se escriben las ecuaciones de búsqueda y que se envían a un servidor a través de SRU. Acepta las construcciones características de los lenguajes de consulta de bases de datos documentales, como son las búsquedas booleanas, de proximidad, delimitadas a un campo, truncamiento de términos, etc. También admite definir distintos conjuntos de puntos de acceso y relaciones que reciben el nombre de contexto. Las relaciones permiten que indiquemos si deseamos recuperar los registros que contengan todos, algunos o exactamente los términos de la ecuación de búsqueda.

Una de las principales ventajas de CQL es que puede ser adaptado y utilizado en diversos contextos y no sólo en el entorno bibliográfico. En este último se han definido varios contextos que permiten la búsqueda en índices de campos Dublin Core, MODS (una combinación de ambos, como el contexto bib) y MARC. Otros están orientados a la consulta de tesauros en línea y a bases de datos que contienen descripciones de servicios de recuperación.

Soy consciente de que existen más normas de las que debía haber tratado, de que no he hablado suficientemente de la necesidad de la cooperación a todos los niveles para acometer el control y la accesibilidad del patrimonio digital y de que apenas me he

---

<sup>44</sup> SRU (Search/Retrieval via URL) Dirección electrónica: <http://www.loc.gov/standards/sru/> [Consulta realizada el 25 de agosto de 2009]

referido al apoyo que las instituciones públicas nacionales e internaciones, como la Unión Europea, a través de sus distintos Programas Marco, han prestado a la digitalización del patrimonio impreso y a la creación de bibliotecas digitales y repositorios institucionales, pero son temas de los que trataré en otra futura colaboración.

25 de Agosto de 2009