

## La biblioteca virtual: función y planteamiento

Xavier Agenjo Bullón  
*Fundación Ignacio Larramendi*

Francisca Hernández Carrascal  
*Digibis, Producciones Digitales*

### Resumen

Se plantean los cimientos teóricos del trabajo y se indica que al igual que ocurrió con el Catálogo Colectivo de Patrimonio Bibliográfico no sólo basta con un buen planteamiento, sino que es imprescindible el esfuerzo sostenido a lo largo de los años. Se enfoca la biblioteca virtual desde el punto de vista del usuario y del lector y se pone de manifiesto cómo sus intereses van siempre encaminados a conseguir el documento primario; por lo tanto el objetivo del bibliotecario debe ser siempre ahorrar tiempo al lector (a conseguir el libro que desea). Se pone de manifiesto cómo en una biblioteca virtual es posible acceder desde el mismo OPAC<sup>1</sup> a múltiples fuentes de información digital y cómo desde los buscadores de propósito general se puede acceder a las bibliotecas virtuales. Se insiste en que la creación de metadatos permite que los objetos digitales estén disponibles, accesibles y visibles en conjuntos de información cada día mayores gracias a la recolección de metadatos que utilizan los agregadores. Se hace la referencia a la digitalización de imágenes como primer paso para constituir bibliotecas virtuales y cómo avanzar más allá mediante el reconocimiento óptico de caracteres. Se hace especial hincapié en los nuevos dispositivos de lectura móviles y cómo es necesario enfrentarse con los problemas de los derechos de autor desde un punto de vista técnico además de jurídico y de la necesidad de identificar unívocamente cada objeto digital. Se hace también referencia a las nuevas reglas de descripción de recursos y de acceso a los mismos, absolutamente imprescindibles para el entorno digital y cómo es necesario que los metadatos puedan transmitirse utilizando a su vez otros metadatos que contengan toda la información sobre el objeto digital y el objeto digital mismo. Se insiste en la necesidad de preservar los recursos digitales y de elaborar políticas de preservación digital al revés de como se ha hecho con los libros impresos y, por último, se establece la siguiente meta de las bibliotecas virtuales consistente en elaborar ontologías para la Web semántica mediante el uso de entidades y relaciones diseñadas específicamente para ese fin.

### Descriptor:

Bibliotecas virtuales; Bibliotecas Digitales; Digitalización, Reconocimiento óptico de caracteres, SRU, OpenSearch, OAI-PMH, METS, METSRights, Libros electrónicos, DOI, OAIS, PREMIS, RDA, Ontologías, RDF, RDFa.

---

<sup>1</sup> El trabajo de Francisco García y Purificación Moscoso, El portal bibliotecario y el OPAC: Ayuda en línea específica y compartida. Boletín de la ANABAD. LVIII (2008) nº 2 135-148, proporciona un análisis actualizado de este aspecto.

## 1. Introducción

Con el título y subtítulo de este artículo queremos rendir homenaje a Mercedes Dexeus que no sólo dio el impulso inicial al Catálogo Colectivo del Patrimonio Bibliográfico en la forma en que hoy lo conocemos<sup>2</sup>, sino también por la magnífica realidad que constituye este proyecto, veinticuatro años más tarde de su iniciación. Lo que ha distinguido durante este cuarto de siglo a este proyecto, con independencia de sus características técnicas, ha sido el trabajo continuado e ininterrumpido a lo largo de los años. Queremos así señalar cómo, independientemente de las potenciales oportunidades de las bibliotecas virtuales y de la avanzada tecnología que se aplique en ellas, sólo el esfuerzo continuado permitirá construir unas herramientas verdaderamente eficaces.

Debemos de hacer un excursión aquí para decantarnos por la expresión “Biblioteca Virtual” en lugar de “Biblioteca Digital”, pues aunque sean muchas y buenas razones las que avalen la segunda denominación, creemos que en la actualidad todas las bibliotecas son digitales y sólo serían verdaderamente virtuales aquellas que conforman conjuntos bibliográficos que no tienen que tener necesariamente una existencia real ni en el tiempo ni en el espacio, sino que se conforman dinámicamente a instancias del usuario<sup>3</sup>.

El recorrido de la biblioteca clásica a la biblioteca virtual ha sido mostrado por nosotros, o al menos lo hemos intentado, en dos publicaciones, *Del Catálogo Colectivo a la biblioteca virtual*<sup>4</sup> y *De las Bibliotecas Virtuales a las ontologías y Web Semántica*<sup>5</sup>. Por

---

<sup>2</sup> Mercedes Dexeus Mallol. “El Catálogo Colectivo : función y planteamiento”. En: Homenaje a Justo García Morales: miscelánea de estudios con motivo de su jubilación. Madrid : Asociación Española de Archiveros, Bibliotecarios, Museólogos y Documentalistas, 1987. pp. 123-140.

<sup>3</sup> Hay también otras razones (pero *a sensu contrario*) que suministran Luis Ángel García Melero y Ernesto García Camarero en su conocida y recomendable publicación “La Biblioteca Digital publicada hace casi 10 años”. García Camarero, Ernesto y Luis Ángel García Melero. “La biblioteca digital. Madrid”, Arco-Libros, 2001.

Debe actualizarse este texto con el artículo que se publica en este mismo número monográfico del Boletín de la ANABAD dedicado a las bibliotecas digitales y que hemos podido consultar como pre-print: García Melero, Luis Ángel. “La biblioteca digital revisitada”. Agosto de 2009. En prensa. <http://eprints.rclis.org/16945/> [Consultado el 28 de febrero de 2010]

<sup>4</sup> Agenjo Bullón, Xavier, y Francisca Hernández Carrascal. “Del catálogo colectivo a la biblioteca virtual: la Biblioteca Virtual de Patrimonio Bibliográfico.” En: I Jornadas sobre Patrimonio Bibliográfico en Castilla-La Mancha: actas: 12, 13 y 14 de noviembre, Alcázar de Toledo. Toledo: Junta de Comunidades de Castilla-La Mancha, 2003. 391-418.

<sup>5</sup> Agenjo, Xavier y Francisca Hernández. “De las bibliotecas virtuales a las ontologías y web semántica.” En: La biblioteca pública: compromiso de futuro: actas / II Congreso Nacional de Bibliotecas Públicas, Salamanca, 17, 18 y 19 de noviembre de 2004. Madrid: Ministerio de Cultura, Subdirección General de Información y Publicaciones, 2004. 354-364. <http://travesia.mcu.es/portallnb/jspui/handle/10421/788> [Consultado el 28 de febrero de 2010]

Anterior a esta artículo puede consultarse además Hernández, Francisca, y Carlos Wert. “XML, ¿una infraestructura para la biblioteca digital?: el proyecto COVAX.” En: La biblioteca pública: portal de la sociedad de la información / I Congreso Nacional de Bibliotecas Públicas, Valencia, 29, 30 y 31 de

último, no dejará de tener su interés confrontar las ideas ahí expresadas con la realidad perceptible *hic et nunc* en los artículos *Perspectivas tecnológicas en el libro y las bibliotecas del futuro*<sup>6</sup> y el *Acceso a la información digital* de María Luisa Martínez-Conde<sup>7</sup>.

## 2. Save the time of the user: el OPAC

Las premisas teóricas que se desprenden de estas publicaciones (a las que sería conveniente añadir el trabajo que se cita<sup>8</sup>) podrían compendiarse de forma brevísima en la cuarta de las cinco leyes de uno de los teóricos de la biblioteconomía de todos los tiempos Shiyali Ramamrita Ranganathan (Sirkazi, Tamil Nadu, 9 de agosto de 1892-Bangalore, 27 de septiembre de 1972) '*Save the time of the user*'. En efecto, el usuario, el lector, por utilizar la terminología clásica, desea acceder a una información o a un documento concreto, o a otros que ni siquiera conoce, pero que responden verdaderamente a su necesidad, mediante ese conjunto de estructuras y funciones que existen desde hace milenios y que reciben el honroso nombre de bibliotecas. Ya que el usuario puede estar situado en cualquier parte del mundo y, gracias a la red, en este caso, para actualizar conceptos, gracias a la nube, la biblioteca recibe el nombre de biblioteca virtual porque consigue que esa información esté disponible para el lector en cualquier sitio y en cualquier momento.

La biblioteca puede, además, salvar las fronteras del tiempo y del espacio mediante su virtualidad, pero también, y esto es aún más importante, las barreras económicas, dado el carácter esencialmente democrático<sup>9</sup> de esta institución.

Si era común en los tratados de biblioteconomía clásicos, el de Manuel Carrión es sin duda el más conocido, seguir el proceso técnico para describir las funcionalidades que debe satisfacer una biblioteca, la biblioteca virtual da la vuelta completamente a todo ese proceso. Lo que el lector encuentra, y no debe encontrar más que eso, es la información, cuya procedencia o método de elaboración no debe ser para él, al menos en principio, ningún tipo de obstáculo, problema o inquietud. El usuario accede, pues, a la información a través de una búsqueda, que como veremos no será necesario siquiera

---

octubre de 2002.- Madrid: Ministerio de Cultura, Subdirección General de Información y Publicación, 2002. 75-80.

<http://travesia.mcu.es/portalnb/jspui/handle/10421/1163> [Consultado el 28 de febrero de 2010]

<sup>6</sup> Agenjo Bullón, Xavier. "Perspectivas tecnológicas en el libro y las bibliotecas del futuro." *El libro, las bibliotecas y los archivos en España: a comienzos del Tercer Milenio*. Madrid: Sociedad Estatal España Nuevo Milenio, 2002. 177-191.

<sup>7</sup> Martínez-Conde, María Luisa. "El acceso a la información digital". En: *El libro, las bibliotecas y los archivos en España a comienzos del tercer milenio*. Madrid : Sociedad Estatal España Nuevo Milenio, 2002. 131-139.

<sup>8</sup> Agenjo Bullón, Xavier. "Directrices para la elección y/o desarrollo de un sistema integrado de gestión de bibliotecas." En: *Tratado básico de biblioteconomía*. Madrid: Universidad Complutense, 1995. 85-103.

<sup>9</sup> Agenjo Bullón, Xavier. "Las bibliotecas (informatizadas) y la democracia." *Educación y Biblioteca*, n° 77 (1997): 60-62.

que se efectúe en la propia biblioteca virtual, sino a través de cualquiera de los grandes buscadores que en el siglo XXI han popularizado y democratizado hasta unos extremos inimaginables antes de la invención de Tim Berners-Lee, hace sólo veinte años, el acceso a la información.

Pero la biblioteca virtual no solo proporciona, bien a través de su sistema de consulta, bien mediante la utilización de un buscador de propósito general, o bien, incluso, mediante herramientas características de la denominada Web 2.0 esa información, sino que proporciona el material bibliográfico mismo que el lector busca, o hacia dónde se le ha redirigido tras las oportunas navegaciones.

En efecto, el lector podrá enviar a su buzón de correo electrónico, bajar al disco duro de su ordenador, direccionar a la impresora el documento mismo que buscaba. O incluso, podrá almacenarlo dentro de un dispositivo de lectura, como puede ser un ciberlibro o un teléfono móvil o cualquier otro dispositivo móvil de estas características, para consultarlo o leerlo cuando le sea más útil o placentero.

**FUNDACIÓN IGNACIO LARRAMENDI**

INICIO  
LA FUNDACIÓN  
IGNACIO HERNANDO DE LARRAMENDI  
BIBLIOTECAS VIRTUALES FHL  

- Textos introducción
- Características de esta edición

 BIBLIOTECA VIRTUAL IGNACIO LARRAMENDI DE POLÍGRAFOS  

- Preámbulo
- Búsqueda polígrafos
- Búsqueda por obras
- Colección de polígrafos españoles
- Colección de polígrafos hispanoamericanos
- Colección de polígrafos portugueses
- Colección de polígrafos brasileños
- Autores de los estudios críticos FHL

 BIBLIOTECA VIRTUAL DE PENSADORES TRADICIONALISTAS  
 BIBLIOTECA VIRTUAL DE COMENTARISTAS DE ARISTÓTELES  
 BIBLIOTECA VIRTUAL DE DOCUMENTACIÓN HISTÓRICA EN CATEDRALES  
 BIBLIOTECA VIRTUAL MENÉNDEZ PELAYO  
 CARLISMO  
 ARCHIVO FAMILIAR MONTIANO

Aggregator RSS Repository OAI SRU Server: SRU Buscar polígrafos → DIGIBIB

Buscar en todas las Bibliotecas Virtuales FHL

Cualquier campo:

Título:

Autor:  [seleccionar]

Editor:

Lugar de edición:

Materia:  [seleccionar]

Lengua:

Año de publicación:  entre  y

Buscar Limpiar

**Ayuda**

Si quiere buscar una frase, enciérrala entre comillas. Por ejemplo: "Miguel de Cervantes".

En cada campo puede introducir una o más palabras; el operador que enlaza los términos es Y.

Los caracteres \* y ? buscan términos que comparten una misma raíz o palabras de ortografía dudosa. Por ejemplo: "ilustr\*", recuperará las palabras ilustrador, ilustraciones, ilustrada, ilustrativo, etc.; y "primiti?a" recuperará los términos primitiva y primitiua. No deben usarse estos caracteres como inicio de un término de búsqueda.

BIBLIOTECA VIRTUAL IGNACIO LARRAMENDI

DE ARISTÓTELES  
COMENTARISTAS DE ARISTÓTELES

DOCUMENTACIÓN HISTÓRICA EN CATEDRALES

**Figura 1. Pantalla de búsqueda de las Bibliotecas Virtuales FHL**  
<http://www.larramendi.es/i18n/consulta/busqueda.cmd>

Para ello, el diseño de la interfaz de consulta debe cumplir determinados requisitos y aportar una serie de funcionalidades<sup>10</sup>. En primer lugar debe ser una consulta simple y sencilla, que no necesite del conocimiento de un iniciado para su mejor manejo y deberá disponer, por supuesto, de unas condiciones de accesibilidad que afortunadamente están perfectamente descritas por el W3C<sup>11</sup>. Naturalmente si se trata de un usuario experimentado podrá utilizar unas funcionalidades más amplias como podrían ser la delimitación de la información que se desea obtener mediante el procedimiento de restringirla a determinados campos de la información, incluso yendo un poco más allá, podrá beneficiarse del enorme potencial que proporcionan los métodos de análisis de la información y la estructuración de la misma que suponen las bibliotecas virtuales semánticas<sup>12</sup>.

Los distintos procedimientos que permiten que el usuario tenga acceso a la información bibliográfica mediante las funcionalidades y recursos que le suministra la biblioteca virtual pueden resumirse en dos. Uno, la visibilidad, de la que trataremos más adelante, con los *sitemaps*, el uso de metadatos *Dublin Core* y el protocolo OAI-PMH; y dos, el catálogo que ya podemos denominar tradicional en línea, es decir el OPAC. El OPAC ha sufrido importantísimas transformaciones en los últimos años y existe un fuerte debate sobre lo que efectivamente debe mostrar. En relación con este aspecto es iluminador el informe realizado por la OCLC en *2009 Online catalogs : what users and librarians want : an OCLC report*<sup>13</sup>.

---

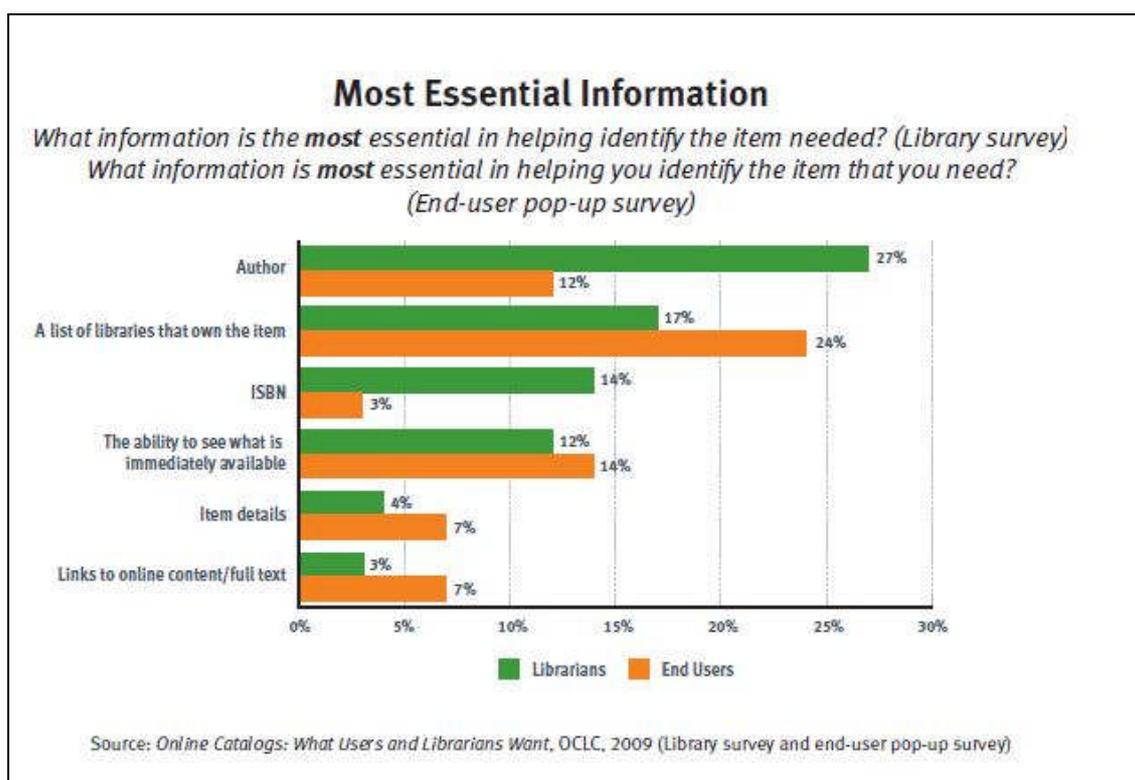
<sup>10</sup> Entre las muchas fuentes existentes destaca por su calidad y por lo ajustado a los problemas actuales de las bibliotecas virtuales y digitales la serie *Usability Inspection of Selected Digital Libraries (Part 1 of 5)* de Lorraine Paterson, analista del proyecto JISC [Joint Information Systems Committee] *Usability and Contemporary User Experience in Digital Libraries*. <http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2/ux2.aspx>, y que mantiene un *blog* denominado *Researching Usability* <http://lorrainepaterson.wordpress.com/> [Consultado el 22 de febrero de 2010]

Para todas estas cuestiones desde un análisis clásico debe verse el libro Fernández Molina, Juan. *Los catálogos de acceso público en línea : el futuro de la recuperación de información bibliográfica*. [Málaga]: Asociación Andaluza de Bibliotecarios, 1998. Es recomendable leer el prólogo, pp. 9-11.

<sup>11</sup> *Web Accesibility Initiative*. <http://www.w3.org/WAI>. [Consultado el 15 de febrero de 2010].

<sup>12</sup> Kruk, Sebastian y W. D. McDaniel. *Semantic digital libraries*. Berlin: Springer, 2009.

<sup>13</sup> Calhoun, Karen, y OCLC. *Online catalogs : what users and librarians want : an OCLC report*. Dublin Ohio: OCLC, 2009. <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf> [Consultado el 15 de febrero de 2010]



**Figura 2. Una de las gráficas más reveladoras del informe de OCLC sobre lo que quieren los usuarios y lo que quieren los bibliotecarios**

Un aspecto fundamental que debe preverse en cualquier OPAC es lo relacionado tanto con la accesibilidad como con la usabilidad. Ambas garantizan que todo tipo de usuarios pueda acceder a la información sea cual sea su capacidad de manejar el OPAC. Desde sus orígenes el *World Wide Web Consortium* (W3C)<sup>14</sup>, a través de la *Web Accessibility Initiative* (WAI)<sup>15</sup>, ha establecido unas pautas para alcanzar diferentes niveles de accesibilidad. La norma UNE 139803:2004<sup>16</sup> recoge las *Web Content Accessibility Guidelines* (WCAG) 1.0<sup>17</sup> que han quedado atrasadas desde la publicación de la recomendación WCAG 2.0 en diciembre de 2008. La *Web Accessibility Initiative* recomienda el uso de WCAG 2.0<sup>18</sup> y la actualización de los sistemas compatibles con la 1.0.

Por otro lado, en los últimos tiempos se han desarrollado un número considerable de aplicaciones Web que permiten tratar la información de una manera más eficaz, pero que tenían la desventaja de no cumplir los requisitos de accesibilidad mencionados. El

<sup>14</sup> <http://www.w3.org/> [Consultado el 15 de febrero de 2010].

<sup>15</sup> <http://www.w3.org/WAI/> [Consultado el 15 de febrero de 2010].

<sup>16</sup> [http://www.inteco.es/Accesibilidad/Normativa\\_1/Descarga/DescargaUNE\\_139803](http://www.inteco.es/Accesibilidad/Normativa_1/Descarga/DescargaUNE_139803) [Consultado el 15 de febrero de 2010].

<sup>17</sup> <http://www.w3.org/WAI/intro/wcag.php> [Consultado el 15 de febrero de 2010].

<sup>18</sup> <http://www.w3.org/TR/WCAG20/> [Consultado el 15 de febrero de 2010].

caso más notorio era el de las aplicaciones basadas en AJAX (abreviatura de *Asynchronous JavaScript and XML*). Esta situación está ya solventada y las normas de accesibilidad han variado notablemente desde el establecimiento de WAI-ARIA, the *Accessible Rich Internet Applications Suite*<sup>19</sup> destinadas a aplicaciones y contenidos Web dinámicos desarrollados con AJAX, DHTML, *JavaScript* y otras tecnologías similares.

En España<sup>20</sup> la Ley 56/2007, de 28 de diciembre, de Medidas de Impulso de la Sociedad de la Información establece (art. 18) que *"a partir del 31 de diciembre de 2008, las páginas de Internet de las Administraciones Públicas satisfarán, como mínimo, el nivel medio de los criterios de accesibilidad al contenido generalmente reconocidos. Excepcionalmente, esta obligación no será aplicable cuando una funcionalidad o servicio no disponga de una solución tecnológica que permita su accesibilidad."* Además extiende (art. 19) esta exigencia a las *"páginas de Internet cuyo diseño o mantenimiento financien total o parcialmente como las páginas de Internet de entidades y empresas que se encarguen de gestionar servicios públicos apliquen los criterios de accesibilidad antes mencionados. En particular, será obligatorio lo expresado en este apartado para las páginas de Internet y sus contenidos de los Centros públicos educativos, de formación y universitarios, así como, de los Centros privados que obtengan financiación pública. Las páginas de Internet de las Administraciones Públicas deberán ofrecer al usuario información sobre su nivel de accesibilidad y facilitar un sistema de contacto para que puedan transmitir las dificultades de acceso al contenido de las páginas de Internet o formular cualquier queja, consulta o sugerencia de mejora."*

Además de lo mencionado, con fecha de 23 de julio de 2009 se publicó el último borrador de las *User Agent Accessibility Guidelines (UAAG) 2.0*<sup>21</sup> de la *W3C Web Accessibility Initiative (WAI)*. Este documento proporciona directrices para reducir las barreras de accesibilidad en el diseño de los agentes de usuario (navegadores y otro software para recuperar y representar contenido en la Web) y promover no sólo la accesibilidad de las interfaces de usuario, sino también la conexión con otras aplicaciones, especialmente las de ayuda.

El Inteco (*Instituto Nacional de Tecnologías de la Comunicación*) mantiene por su parte unas manejables *Guías Prácticas de Comprobación de la Accesibilidad Web*<sup>22</sup> que se ha actualizado el 12 de febrero de 2010 con un nuevo capítulo dedicado a un aspecto esencial en el campo de las bibliotecas digitales *Comprobación de la Accesibilidad: Legibilidad y Metainformación*.

### 3. Hasta el infinito y más allá: las metabúsquedas

También desde el OPAC es posible acceder en la actualidad, o debería serlo, especialmente en una biblioteca virtual, a un gran volumen de recursos en línea sean estos accesibles mediante metabuscadores o bien mediante otros protocolos ofrecidos al usuario. La necesidad de expandir la búsqueda y recuperación de información a

<sup>19</sup> <http://www.w3.org/WAI/intro/aria.php> [Consultado el 17 de febrero de 2010]

<sup>20</sup> La Universidad de Alicante mantiene la página *Accesibilidad Web* que recopila diferentes recursos sobre este tema, incluyendo la legislación española y europea.  
<http://accesibilidadweb.dlsi.ua.es/?menu=quees> [Consultado el 17 de febrero de 2010]

<sup>21</sup> <http://www.w3.org/TR/UAAG20/> [Consultado el 17 de febrero de 2010]

<sup>22</sup> [http://www.inteco.es/Accesibilidad/Formacion\\_6/Manuales\\_y\\_Guias/guias\\_comprobacion](http://www.inteco.es/Accesibilidad/Formacion_6/Manuales_y_Guias/guias_comprobacion) [Consultado el 17 de febrero de 2010]

múltiples bases de datos, según diferentes protocolos e instaladas sobre diferente software ha conducido a la creación de sistemas de metabúsquedas, búsquedas federadas o búsquedas distribuidas<sup>23</sup>.

El ampliamente conocido y utilizado protocolo Z39.50 está siendo superado por diferentes servicios de búsqueda Web que suplen la principal deficiencia de esta norma y que no es otra que utiliza el *Internet Protocol*, pero es necesario desarrollar pasarelas para su funcionamiento en la Web. Así, SRU (*Search/Retrieval via URL*)<sup>24</sup>, que ha venido a cubrir la funcionalidad de Z39.50, es un protocolo que establece la sintaxis de una búsqueda (expresada en *Common Query Language*) y los mecanismos de petición y respuesta entre un cliente y un servidor en un entorno Web. *OpenSearch*, desarrollado originalmente por Amazon, especifica la forma de interrogar a un servicio de búsqueda Web, la forma de construcción de una búsqueda y el envío de resultados en la forma de RSS o Atom. Por medio de APIs abiertas o *plugins*, los servidores *OpenSearch* (p.e., *Nature*, *WorldCat*, *Wikipedia*, etc.) proporcionan los resultados de una búsqueda, mientras que los clientes, agregadores y navegadores<sup>25</sup>, utilizan las descripciones del servicio de búsqueda *OpenSearch* para incluir un determinado sitio en su lista de motores de búsqueda y así lo hacen *Firefox*<sup>26</sup>, *Internet Explorer*<sup>27</sup> o *Chrome*<sup>28</sup>.

En este sentido conviene estar atento a algunas iniciativas como las de OASIS o NISO. OASIS (*Organization for the Advancement of Structured Information Standards*)<sup>29</sup> es un consorcio que impulsa el desarrollo, la convergencia y la adopción de estándares abiertos para la sociedad de la información, formado por más de 5.000 participantes, en representación de más de 600 organizaciones y miembros individuales en 100 países, y que colabora con diferentes instituciones de normalización como ISO. OASIS desarrolla su trabajo a través de diferentes comités y grupos de trabajo, entre los cuales es especialmente interesante el *OASIS Search Web Services Technical Committee*<sup>30</sup>. Hasta la fecha este comité técnico se ha ocupado de la integración de diversos enfoques de servicios de búsqueda y recuperación en la Web en un modelo unificado partiendo de

<sup>23</sup> McCallum, Sally. "A look at new information retrieval protocols: SRU, OpenSearch/a9, CQL, and XQuery". In *World Library and Information Congress: 72<sup>nd</sup> IFLA General Conference and Council*. IFLA, 2006.

<http://archive.ifla.org/IV/ifla72/papers/102-McCallum-en.pdf> [Consultado el 17 de febrero de 2010]

<sup>24</sup> <http://www.loc.gov/standards/sru/specs/search-retrieve.html> [Consultado el 17 de febrero de 2010]

<sup>25</sup> [http://www.opensearch.org/Community/OpenSearch\\_search\\_engine\\_directories](http://www.opensearch.org/Community/OpenSearch_search_engine_directories) [Consultado el 17 de febrero de 2010]

<sup>26</sup> [https://developer.mozilla.org/en/Creating\\_OpenSearch\\_plugins\\_for\\_Firefox](https://developer.mozilla.org/en/Creating_OpenSearch_plugins_for_Firefox) [Consultado el 17 de febrero de 2010]

<sup>27</sup> <http://www.microsoft.com/windows/ie/searchguide/en-en/default.msp> [Consultado el 17 de febrero de 2010]

<sup>28</sup> <http://www.google.com/chrome/intl/en/webmasters-faq.html> [Consultado el 17 de febrero de 2010]

<sup>29</sup> <http://www.oasis-open.org/home/index.php#googtrans/auto/es> [Consultado el 15 de febrero de 2010].

<sup>30</sup> [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=search-ws](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=search-ws) [Consultado el 15 de febrero de 2010].

las especificaciones SRU y *OpenSearch*. Este modelo unificado recibe el nombre de *Abstract Protocol Definition*<sup>31</sup> que establece el marco para el desarrollo de aplicaciones concretas y el modelo de funcionamiento de un servicio Web de búsqueda y recuperación a través de los ‘*dynamic bindings*’ (que podría traducirse como vínculos o agregaciones dinámicas) que son descripciones estructuradas de un servicio que, al ser leídas por una máquina, permitirán la configuración dinámica de la misma para adaptarse al protocolo descrito. El resultado del trabajo del OASIS *Search Web Services* TC se ha concretado en la definición de vínculos dinámicos para *OpenSearch* y SRU/CQL 1.2 y en la elaboración de la versión 2.0 de SRU y CQL, ambas publicadas como borradores con fecha 22 de julio de 2009<sup>32</sup>. Por su parte, NISO ha creado el grupo de trabajo *Metasearch Initiative*<sup>33</sup> cuyo objetivo es potenciar el desarrollo de normas, elaborar guías de buenas prácticas y herramientas que incrementen la eficacia de los metabuscadores para los proveedores de servicios, los proveedores de contenidos y los usuarios finales.

The screenshot shows the SRU search page for the Fundación Ignacio Larramendi. At the top, there is a language selection dropdown and a Google Translate logo. The main header includes the foundation's name and a search bar. A navigation menu on the left lists various virtual libraries. The main content area is titled 'SRU Servidor SRU' and contains text explaining the SRU protocol, its use of CQL, and provides the URL for the server: <http://www.larramendi.es/i18n/sru/sru.cmd>.

**Figura 3. Página de búsqueda SRU de las Bibliotecas Virtuales FHL**  
<http://www.larramendi.es/i18n/sru/sru.cmd>

Son también muy importantes los programas que permiten que una búsqueda realizada en un recurso de información pueda transferirse a otros directamente como es el caso de *OpenUrl*<sup>34</sup>. El planteamiento de *OpenURL* viene a cubrir la necesidad de identificar por

<sup>31</sup> Denenberg, Ray. "Search Web Services - The OASIS SWS Technical Committee Work." D-Lib Magazine 15, 1/2 (2009).  
<http://www.dlib.org/dlib/january09/denenberg/01denenberg.html> [Consultado el 15 de febrero de 2010].

<sup>32</sup> <http://www.loc.gov/standards/sru/oasis.html> [Consultado el 17 de febrero de 2010]

<sup>33</sup> <http://www.niso.org/workrooms/mi> [Consultado el 17 de febrero de 2010]

<sup>34</sup> ANSI/NISO Z39.88 - The OpenURL Framework for Context-Sensitive Services

medio de una URL un mismo recurso en diferentes servicios Web (*'context-sensitive services'*). Así, se utiliza habitualmente para transmitir citas y registros bibliográficos de un recurso que mediante un servidor de enlaces traslada la URL con los metadatos del registro y del servicio requerido (texto completo, repositorios, índices, etc.) al proveedor de estos servicios. En enero de 2010 el *United Kingdom Serials Group* y NISO anunciaron el primer informe elaborado por el KBART (*Knowledge Bases and Related Tools*) Working Group, el *KBART Recommended Practice (NISO RP-9-2010)*<sup>35</sup>, que forma parte de una iniciativa conjunta de ambas organizaciones para examinar los problemas relacionados con la implantación de *OpenURL*. La OCLC mantiene un registro de servidores de enlaces, *OCLC OpenURL Resolver Registry*, dónde se puede dar de alta un servidor y consultar los existentes, aunque no todos sean públicos<sup>36</sup>.

Entre las funciones características de las bibliotecas virtuales aplicables a los OPAC están sin duda la exportación de datos para los gestores de referencias bibliográficas. No suele hacerse demasiado hincapié en que para que estos gestores sean verdaderamente eficaces, la biblioteca virtual, al igual que genera *Dublin Core* (ISO 15836:2009), debe generar COinS (*ContextObjects in Spans*<sup>37</sup>). Los metadatos COinS, que forma parte de la especificación *OpenURL*, son descripciones bibliográficas incluidas en el código HTML de un sitio Web, que los clientes, habitualmente gestores de referencias bibliográficas como *CiteULike*, *Zotero*, *Mendeley* y otros pueden leer e interpretar para descargar las descripciones<sup>38</sup>. Este método de embeber metadatos bibliográficos en el código HTML de un sitio Web puede hacerse con otras estructuras como RIS<sup>39</sup>, MODS, *Dublin Core*, etc., y como veremos más adelante es uno de los métodos que está permitiendo el enorme crecimiento de datos para la Web semántica.

---

[http://www.niso.org/kst/reports/standards?step=2&project\\_key=d5320409c5160be4697dc046613f71b9a773cd9e](http://www.niso.org/kst/reports/standards?step=2&project_key=d5320409c5160be4697dc046613f71b9a773cd9e) [Consultado el 17 de febrero de 2010]

<sup>35</sup> [http://www.niso.org/apps/group\\_public/documents.php?wg\\_abbrev=kbart](http://www.niso.org/apps/group_public/documents.php?wg_abbrev=kbart)  
[Consultado el 15 de febrero de 2010].

<sup>36</sup> <http://www.oclc.org/productworks/urlresolver.htm> [Consultado el 15 de febrero de 2010].

<sup>37</sup> <http://ocoins.info/>  
[Consultado el 15 de febrero de 2010].

<sup>38</sup> En el momento que se redacta esta nota [18 de febrero de 2010] no está operativo *COinS* ni para el OPAC de la Biblioteca Nacional ni para el de REBIUN, ni para las Bibliotecas Públicas, ni el Catálogo Colectivo de Patrimonio Bibliográfico, aunque sí para la Biblioteca Virtual de Patrimonio Bibliográfico.

<sup>39</sup> Este formato toma el nombre de la empresa *Research Information Systems Incorporated* que lo desarrolló como formato de intercambio entre sistemas de citas bibliográficas.

**OpenURL COinS: A Convention to Embed Bibliographic Metadata in HTML**

*stable version 1.0*

**Abstract**

COinS (ContextObjects in Spans) is a simple, ad hoc community specification for publishing OpenURL references in HTML.

**Contents**

- Main Page
  1. [Introduction](#)
  2. [Specification : OpenURL ContextObject in SPAN \(COinS\)- Embedding Citation Metadata in HTML](#)
  3. [Discussion : How to use COinS in HTML](#)
  4. [Details](#)
    1. [Empty SPANs](#)
    2. [Why "Z3988"?](#)
    3. [What is a ContextObject?](#)
    4. [Choosing the type of ContextObject for Compatibility](#)
    5. [XHTML](#)
    6. [why the span element?](#)
    7. [why class and title attributes?](#)
  5. [Implementations](#)
    1. [Embedding Sites](#)
    2. [COinS Processors](#)

**Figura 4. Página de COinS**

No es este el momento de explicar aquí el funcionamiento de este tipo de herramientas, pero sí de citar los magníficos tutoriales preparados por Julio Alonso Arévalo<sup>40</sup> denominado *Los Gestores de Referencias Sociales: índices de popularidad y descubrimiento científico*.

Más adelante entraremos en las consideraciones de la Web 2.0<sup>41</sup> que evidentemente pueden aplicarse tanto en el OPAC como en cualquier otra parte de la biblioteca virtual. Sin embargo, y por estar relacionado con la presentación de la primera interfaz de la aplicación, es decir el OPAC, y ser al igual que los ya citados *Dublin Core* o *COinS*, metadatos generados por éste, sí conviene citar los RSS (*Really Simple Syndication*), gracias a los cuales es posible obtener una información constante, a modo de suscripción, de las variaciones que experimenta la biblioteca virtual. Estas pueden referirse a nuevos registros en la base de datos, a nuevas funcionalidades de la biblioteca virtual o, en general, a cualquier novedad que ésta experimente. Conviene mencionar que parece disminuir en los últimos 24 meses el uso específico de estas herramientas, favorecido por técnicas de *microblogging* como *Facebook* o *Twitter*, cada vez más frecuentes en la página de inicio de las bibliotecas, máxime si a estas

<sup>40</sup> <http://comunidad20.sedic.es/?p=273> [Consultado el 15 de febrero de 2010].

<sup>41</sup> Resulta excelente el trabajo realizado por Didac Margaix y publicado por la APEI *Informe APEI sobre web social* que describe la situación a finales de 2008. <http://informeapeiwebsocial.dmaweb.info/> [Consultado el 15 de febrero de 2010]

Debe actualizarse con *A guide to using Web 2.0 in Libraries* publicado por *cilips (Chartered Institute of Library and Information Professionals in Scotland)* <http://www.slainte.org.uk/files/pdf/web2/Web2GuidelinesFinal.pdf> [Consultado el 15 de febrero de 2010]

Y su *Slainte2.0* <http://www.slainte.org.uk/slainte2/index.html>, permanentemente actualizado

circunstancias se le une el hecho de funcionalidades como 'Create a feed', incorporada a Google Reader a principios de 2010 que permite generar suscripciones a novedades, incluso en aquellos sitios que no disponen de RSS o Atom.



Figura 5. Página de la Fundación Ignacio Larramendi en Facebook

Facebook tiene un número más amplio de seguidores y usuarios, superando al parecer los 350 millones, que Twitter y según algunos observadores y analistas de las redes sociales parece que el primero se está decantando cada vez más hacia un público generalista, mientras que el segundo, con la sexta parte de usuarios aproximadamente, se utiliza más como agencia de noticias de multitud de instituciones y especialistas en las distintas ramas del saber. En todo caso, se trata de aspectos de la Web que se mueven rapidísimamente, hace 18 meses *Second Life* o *MySpace* parecían más utilizadas de lo que son ahora.



Figura 6. Página de DIGIBIS, Producciones Digitales en Twitter

#### 4. Nunca mucho trigo hizo mal año: Los catálogos colectivos virtuales

Otra forma de dar respuesta a la necesidad de buscar recursos de información en múltiples bases de datos es la que ofrece el protocolo *Open Archives Initiative Protocol for Metadata Harvesting*. Si bien no se trata, ni mucho menos, de un protocolo de búsqueda al modo de los expuestos en el apartado anterior, los sistemas construidos en torno a este protocolo producen como resultado la formación de 'catálogos colectivos virtuales', lo que tiene mucho de pleonasma.

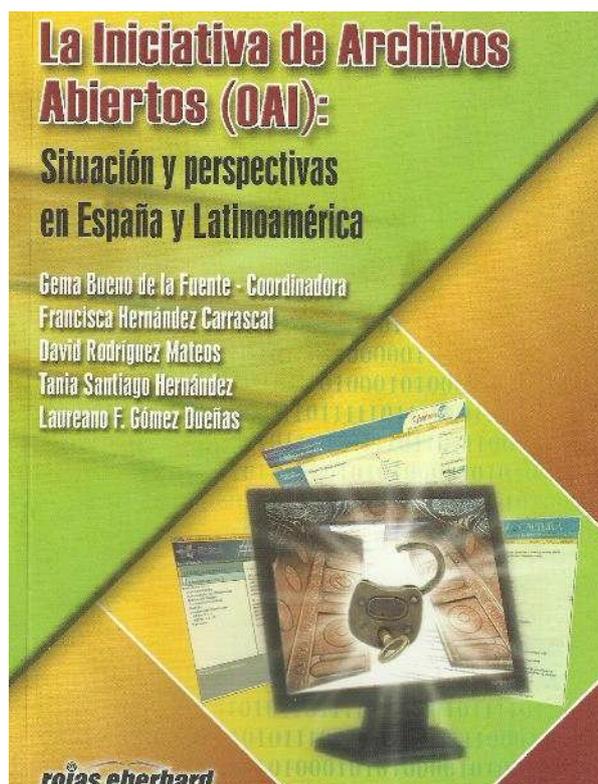


Figura 7. Cubierta de La Iniciativa de Archivos Abiertos (OAI) editado por Rojas Eberhard

Ya hemos descrito en otra parte<sup>42</sup>, y existe una amplia bibliografía sobre ello, la arquitectura del protocolo OAI. Se trata de un sistema en el que una de las partes actúa como recolector<sup>43</sup> de los metadatos y configura un repositorio común con una interfaz de búsqueda y recuperación de información; mientras que los repositorios individuales actúan como bases de datos abiertas a los recolectores. La comunicación entre el recolector y los repositorios individuales se realiza a través de peticiones OAI-PMH que

<sup>42</sup> Bueno de la Fuente, Gema, y Francisca Hernández Carrascal, 'El protocolo OAI-PMH', en *La Iniciativa de Archivos Abiertos (OAI): situación y perspectivas en España y Latinoamérica*. Bogotá: Rojas Eberhard, 2008. 33-58.

<sup>43</sup> Para la definición de este término y otros véase Barrueco Cruz, José Manuel, y Imma Subirats Coll, 'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo'. Swets Blackwell, 2003.

<http://eprints.rclis.org/177/2/cardedeu.pdf> [Consultado el 22 Febrero 2010].

se transmiten a través de HTTP y se devuelven como conjuntos de bytes en XML<sup>44</sup>. El objetivo de las peticiones OAI es que el recolector (o proveedor de servicios) obtenga los metadatos que ofrecen los repositorios (proveedor de datos) individuales. El resultado es que el recolector puede ofrecer los metadatos recolectados a través de una interfaz de búsqueda interrogable por cualquier usuario y habitualmente con una funcionalidad similar a la de cualquier OPAC. Es decir, al usuario final le permite acceder desde un único punto a todos los metadatos recolectados sin tener que saber o conocer la ubicación de cada uno de los repositorios que son el origen de la información. Siguiendo el modelo de los catálogos colectivos, es una forma de acceder e interrogar a múltiples bases de datos sin ser en sí un protocolo de búsqueda.

OAI-PMH representa uno de los mayores aciertos en el acceso a bases de datos bibliográficas (y otros datos) en los últimos 10 años. Gracias a su sencillez y facilidad de aplicación, a la que no es ajeno el hecho de que el esquema básico para los metadatos sea Dublin Core (ISO 15836:2009<sup>45</sup>) se ha extendido enormemente, empujando, y recibiendo a su vez el empuje, de movimientos como *Open Access (OA)* y dando lugar a 'repositorios abiertos', en el sentido de los principios de difusión de contenidos que animan este movimiento, pero también en el sentido de ser susceptibles de ser recolectados. Seguramente quienes más se han beneficiado de la rápida expansión del protocolo y de las aplicaciones basadas en él han sido los denominados repositorios institucionales y las bibliotecas digitales o virtuales, términos ambos difícilmente distinguibles si no es por referencia a la política de incremento de la colección digital que subyace en cada uno de ellos y la relación entre los autores de las publicaciones y el propio repositorio. Sin embargo, consideramos que todos los temas que se tratan en este artículo son aplicables tanto a bibliotecas virtuales como a repositorios institucionales y que la diferenciación es más una cuestión organizativa y administrativa que práctica o tecnológica y, mucho menos, normativa. Sí es de señalar cómo el retraso en la aplicación de la normativa de facto OAI por parte de gran parte de las empresas de software para bibliotecas ha provocado que los repositorios institucionales se hayan constituido en muchos casos en una biblioteca digital B, paralela al sistema de gestión bibliotecario utilizado para los materiales bibliográficos. Puede que este sea el motivo por el cual esté tan extendida la idea de que un repositorio institucional no tiene nada

<sup>44</sup> <http://www.openarchives.org/pmh/> [Consultado el 15 de febrero de 2010]

<sup>45</sup> ISO 15836:2009 Information and documentation -- The Dublin Core metadata element set. Hay que decir que ya que la normativa sobre este tema es considerable resulta muy útil el procedimiento de sindicación de la página de la norma dentro del sitio web de ISO [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=52142](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=52142) [Consultado el 20 de febrero de 2010] <http://www.iso.org/iso/rss.xml?csnumber=52142&rss=detail> [Consultado el 20 de febrero de 2010]

Además, la página de cada norma siempre hace referencia a otras normativas afines. Por ejemplo, la página de la norma ISO para Dublin Core informa sobre:

- ISO 15511:2009 Information and documentation -- International standard identifier for libraries and related organizations (ISIL);
- ISO 8459:2009 Information and documentation -- Bibliographic data element directory for use in data exchange and enquiry
- ISO 28500:2009 Information and documentation -- WARC file format

que ver con una biblioteca virtual y viceversa y la poca permeabilidad que hay entre ambas estructuras siendo en todo tan similares<sup>46</sup>.

El gran éxito de los sistemas OAI-PMH puede verse en las cifras que arrojan algunas aplicaciones. Según *Wayback Machine*<sup>47</sup>, en junio de 2003 *OAIster*, creado en 2002<sup>48</sup> como un proyecto de la *University of Michigan Digital Library Production Services* contaba con 1,089,937 registros de 142 instituciones. En 2009 OCLC anunció que pasaba a alojar y administrar *OAIster* en incluyó los registros recolectados a través de OAI en *WorldCat*. Las cifras a principios de 2010 son de más de 23 millones de registros procedentes de 1.100 repositorios. El hecho de que una biblioteca virtual disponga del protocolo OAI le da por sí mismo la oportunidad de ser visible a través del mayor catálogo del mundo o de ser buscable a través de los grandes buscadores como *Google* o *Yahoo* que indizan directamente *OAIster*, salvando así la opacidad de muchas bases de datos. Claro está que para ello es preciso dar de alta un repositorio en diferentes registros, *OAI Registered Data Providers*<sup>49</sup> o en los proveedores de servicios como *OAIster*, *Registry of Open Acces Repositories (ROAR)*<sup>50</sup>, *Recolecta: Recolector de Ciencia Abierta*<sup>51</sup> o *Hispana, Directorio y Recolector de recursos digitales*<sup>52</sup>. Aunque parezca obvio no debe serlo tanto porque no es infrecuente tener que rastrear la red para conocer la dirección de un determinado repositorio y en muchas ocasiones es imposible.

---

<sup>46</sup> Relacionado con estos aspectos resulta muy completa la Memoria descriptiva del proyecto: Repositorio UJI: Espacio digital para la investigación. (Universitat Jaume I) <http://www.repositori.uji.es/jspui/bitstream/10234/7376/1/MemDescrip09.doc> [Consultado el 20 de febrero de 2010]

<sup>47</sup> <http://web.archive.org/web/20030217011435/www.oaister.org/o/oaister/> [Consultado el 20 de febrero de 2010]

<sup>48</sup> Puede verse la historia de *OAIster* en <http://www.lib.umich.edu/digital-library-production-service-dlps/oaister/> [Consultado el 20 de febrero de 2010]

<sup>49</sup> <http://www.openarchives.org/Register/ValidateSite> [Consultado el 20 de febrero de 2010]

<sup>50</sup> <http://roar.eprints.org/> [Consultado el 20 de febrero de 2010]

<sup>51</sup> <http://www.recolecta.net/buscador/> [Consultado el 20 de febrero de 2010]

<sup>52</sup> <http://hispana.mcu.es/es/estaticos/contenido.cmd?pagina=estaticos/presentacion>. Hispana ofrece, a fecha 20 de febrero de 2010, 1.152.234 registros de 121 repositorios (institucionales y bibliotecas digitales o virtuales). Hay que recordar que su andadura se inició en 2006 con 121.470 registros recolectados en 25 de repositorios. [Consultado el 20 de febrero de 2010]

The screenshot shows the OAlster database page on the OCLC website. The page is titled "The OAlster® database" and features a navigation menu on the left with options like "At a glance", "OAlster metadata contributors", "How to contribute", "OAlster history", "Accessing OAlster", and "Support, training and documentation". The main content area is divided into sections: "AT A GLANCE" (listing availability through WorldCat.org, record types, and contributor count), "NEWS" (with links to recent updates), and "Access to OAlster" (describing search options). A "Subscribe to OCLC updates and offers" button is also present.

Figura 8. Acceso a OAlster en OCLC

El protocolo OAI admite que una determinada comunidad establezca además del esquema de metadatos básico, *Dublin Core* (oai\_dc), otros esquemas como MARC 21, EAD, SWAP, etc. Seguramente el caso más singular de acuerdo en una comunidad lo representa *Europeana*. Uno de los métodos de incremento de registros en *Europeana* se realiza a través del protocolo OAI, sin embargo, organizativamente esto representa un gran problema puesto que se trata de administrar la recolección directa de miles de repositorios. Por este motivo *Europeana* concretó durante 2009 una nueva organización basada en la recolección de agregadores nacionales o temáticos que se encargan de la recolección de los repositorios de un país y a su vez son recolectados por *Europeana* y planteó un nuevo esquema de metadatos que venía a intentar solventar una serie de deficiencias que presenta actualmente su sitio Web (europeana.eu). Estas deficiencias son, básicamente, la dificultad de mostrar miniaturas de los contenidos digitales en las páginas de resultados y en la navegación, la imposibilidad de recolectar objetos digitales para acceder directamente a ellos y la falta de contenidos procedentes de archivos europeos. En agosto de 2009 *Europeana* publicó una serie de documentos que reflejaban la nueva estructura de elementos requeridos a los participantes en el proyecto para la nueva versión del sitio Web y que giran en torno a la versión 3.2.2 del esquema *Europeana Semantic Elements (ESE)*<sup>53</sup>. Así, los proveedores de contenidos deben cumplir, para facilitar la recolección de metadatos y de objetos digitales, la ampliación

<sup>53</sup> [https://146.48.82.158/c/document\\_library/get\\_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602](https://146.48.82.158/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602) [Consultado el 20 de febrero de 2010]

El 16 de febrero de 2010 *Europeana* difundió la versión 4.1.1 de Definition of the *Europeana Data Model elements* que reutiliza otras ontologías y modelos como *Identity of Resources on the Web* (IRW), *OAI Object Reuse and Exchange* (ORE), *Simple Knowledge Organization System* (SKOS), *Dublin Core* (DC) y *Friend of a Friend* (FOAF).

del esquema *Dublin Core*, recomendando la utilización de 22 nuevos elementos para una mayor precisión a los datos, y la introducción de 12 elementos específicos de Europeana, cuyo objetivo es una mejor presentación de los datos en el sitio Web.



**Figura 9. ESE consta de 15 elementos Dublin Core básico, 22 elementos Dublin Core Cualificados y 12 elementos propios de Europeana**

Otra muestra de la modificación del protocolo está también representada por una iniciativa europea *DRIVER (Digital Repository Infrastructure Vision for European Research)*<sup>54</sup> que tiene como objetivo establecer una infraestructura cohesionada y paneuropea de repositorios digitales y ofrecer servicios y funcionalidades para investigadores y para el público en general. A través de *DRIVER* se recolectan creaciones científicas, informes científicos y técnicos, informes de trabajo, pre-prints, artículos y datos para lo cual los repositorios deben cumplir unas directrices de aplicación que se superponen al protocolo OAI-PMH para la adecuación de éstos a la metodología y normativa específica de la red de repositorios institucionales europeos. *DRIVER* es a su vez uno de los agregadores especializados de Europeana a la que aporta las publicaciones científicas.

Más allá del acierto en lo que constituye una publicación científica para *DRIVER*, sobre lo cual habría bastante que discutir ya que se asimila en la práctica a las publicaciones generadas por una institución académica o científica, lo cierto es que el esquema organizativo de repositorios europeos ha mostrado ya alguno de sus problemas más inmediatos. El primero, muy recomendado por Europeana, es el registro de los derechos de autor de los recursos y el segundo la utilización de sistemas de identificación unívoca de los recursos por medio de sistemas como DOI<sup>55</sup>. Es fácil darse cuenta de que si los

<sup>54</sup> <http://www.driver-repository.eu/Driver-About/About-DRIVER.html> [Consultado el 20 de febrero de 2010]

<sup>55</sup> A DOI está dedicado todo el epígrafe 9

metadatos de un recurso son recolectados por diferentes proveedores de servicios, unos temáticos o especializados como *DRIVER* y otros como los agregadores nacionales, la duplicación de descripciones será inmediata y el método más sencillo y efectivo de deshacer esta duplicidad es la utilización de un sistema de identificación digital del recurso.

La novedad más importante dentro de OAI ha sido la especificación del complemento OAI-ORE (*Open Archives Initiative Object Reuse and Exchange*)<sup>56</sup>. Esta norma proporciona las bases para la creación de aplicaciones y servicios de visualización, preservación, transferencia, etc., y para la mejora del acceso a las agregaciones que los usuarios utilizan en habitualmente en su interacción con la web, incluyendo documentos Web, documentos multiformato en repositorios institucionales, datos de investigación y colecciones de fotos y música en línea.

De forma simultánea al mundo de las bibliotecas, la aplicación de OAI se está extendiendo a los archivos y museos y se está trabajando en dos aspectos muy interesantes como *COBOAT*<sup>57</sup> y el *OAICatMuseum*<sup>58</sup> financiado por la *Andrew Mellon Foundation*

Por último y para finalizar este capítulo queremos dejar constancia del último proyecto que está desarrollando Herber van Sompel<sup>59</sup> *The Memento Project*<sup>60</sup> que permitiría navegar por la Web temporalmente y acceder a sitios concretos en el estado en que se encontraban en un momento determinado del tiempo.

## 5. Imágenes (y otros objetos multimedia)

A pesar de que no existe una única normativa que regule los procesos de digitalización, sí existen recomendaciones y guías establecidas por diferentes proyectos que sirven para que cada biblioteca virtual desarrolle su propio entorno normativo en función de la calidad de las reproducciones, sus diferentes usos (tanto por las personas como por aplicaciones y máquinas) y tipos de acceso, el sistema de información en que se integren para su gestión, incluyendo los metadatos necesarios para ello, y los requisitos de preservación digital. Estas recomendaciones sirven no sólo para proyectos de digitalización, sino que también son aplicables a las características de los objetos que nacen digitalmente, lo que frecuentemente se olvida como si estos documentos no fueran a verse afectados por los mismos problemas de preservación que los digitalizados.

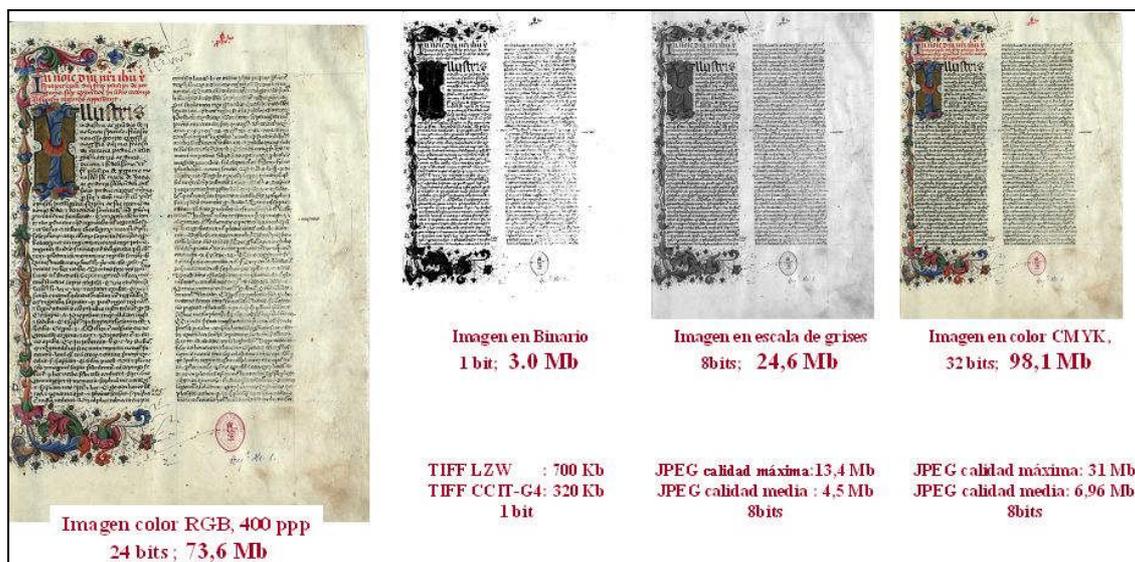
<sup>56</sup> <http://www.openarchives.org/ore/> [Consultado el 20 de febrero de 2010]

<sup>57</sup> <http://www.oclc.org/research/activities/coboat/default.htm> [Consultado el 15 de febrero de 2010]

<sup>58</sup> <http://www.oclc.org/research/activities/oaicatumuseum/default.htm> [Consultado el 15 de febrero de 2010]

<sup>59</sup> *Research Library Prototyping Team Leader* en *Los Alamos National Laboratory*, considerado el padre del protocolo OAI-PMH y participante principal en la especificación de SFX, OpenUrl y OAI-ORE.

<sup>60</sup> <http://www.mementoweb.org/> [Consultado el 20 de febrero de 2010]



**Figura 10. Profundidad de codificación de puntos**

Cada biblioteca virtual debe establecer los requisitos técnicos de digitalización en cuanto a la resolución, esquema de bits utilizado para representar un punto (blanco y negro, escala de grises o color), la compresión de los ficheros obtenidos en el proceso de escaneado, los formatos de ficheros, y debe adecuar estos requisitos a las características de los originales. Todo ello redundará en la calidad y fiabilidad de la biblioteca virtual de cara a sus usuarios, así como en su sostenibilidad. Una fuente de información de gran calidad y extensión para el desarrollo de los requisitos de un proyecto de digitalización es *Llevando la teoría a la práctica: Tutorial de digitalización*<sup>61</sup> de imágenes de la biblioteca de la *Cornell University*. En ella se puede encontrar una explicación detallada de todos los conceptos que se han mencionado, así como su influencia en la construcción de una colección digital de calidad, y una recopilación de fuentes de información complementarias, entre las que se cuentan los requisitos de algunos de los grandes proyectos de digitalización<sup>62</sup>. Otra obra de referencia básica para estas cuestiones es la 3ª edición (2007) de la *NISO Recommended Practice, A Framework of Guidance for Building Good Digital Collections*<sup>63</sup> destinada a servir de guía a instituciones de memoria para la planificación de proyectos de creación de colecciones digitales (obsérvese que no se trata sólo de proyectos de digitalización) y a entidades que financian, subvencionan o promueven el desarrollo de colecciones digitales de calidad.

Queremos también señalar en este apartado que a menudo existe una gran disparidad en los formatos de los ficheros (TIFF, JPEG, JPEG 2000, PDF, etc.) de objetos digitales que se utilizan en los diferentes proyectos, con diferentes fines, lo cual dificulta la

<sup>61</sup> <http://www.library.cornell.edu/preservation/tutorial/> [Consultado el 21 de febrero de 2010]

<sup>62</sup> *Representative Institutional Requirements for Access*  
<http://www.library.cornell.edu/preservation/tutorial-spanish/presentation/table7-2.html> [Consultado el 21 de febrero de 2010]

<sup>63</sup> <http://framework.niso.org/> [Consultado el 21 de febrero de 2010]

selección de los más convenientes o la detección de efectos colaterales a los formatos elegidos lo cual tiene una gran influencia en la continuidad de los proyectos de digitalización. La *Library of Congress* mantiene un sitio Web de referencia en el que analiza una gran cantidad de formatos (imágenes, texto, audio, video, etc.)<sup>64</sup>, sus ventajas, desventajas, uso más conveniente y adopción en los programa de esta biblioteca. estos cambian con la aparición de nuevos formatos y normas incidiendo en la selección del formato más adecuado.

En España debe seguirse el trabajo del *Grupo de Trabajo de Colecciones Digitales*<sup>65</sup>, formado por representantes de la Subdirección General de Coordinación Bibliotecaria, de la Biblioteca Nacional y de las Comunidades Autónomas y coordinado por María Luisa Martínez-Conde. Este grupo de trabajo ha traducido las *Directrices para proyectos de digitalización: de colecciones y fondos de dominio público en particular para aquellos custodiados en bibliotecas y archivos*<sup>66</sup>, resultado del trabajo de un grupo de expertos de IFLA e ICA (*International Council on Archives*) invitados por la UNESCO.

## **6. Mil palabras valen más que una imagen: el texto completo y el reconocimiento óptico de caracteres**

Uno de los inventos tecnológicos más importantes es el del reconocimiento óptico de caracteres. Aunque no existe acuerdo sobre quién creó este invento, parece que Raymond Kurzweil reúne bastantes posibilidades para ser considerado el padre de la criatura. Lo que más nos interesa es que el sistema que él creó en 1974, y que se denominaba "*Omni-font*" y evolucionó comercialmente hasta convertirse en *Nuance Textbridge*, ha experimentado una enorme mejora, tanto en el grado de reconocimiento de prácticamente todo tipo de letrerías o tipos como de la optimización de los procesos, que alcanzan tasas de reconocimiento muy altas. Una ratio razonable de producción con 3 estaciones de trabajo estaría en torno a estas cifras: 6,5 segundos por página, lo que supone una media de 13.807 páginas/día, 575 páginas/hora, u 8.963.675 caracteres/hora en condiciones óptimas; y 10.508 páginas/día, 437,8 páginas/hora, o 6.826.035 caracteres/hora en situaciones reales, es decir considerando puntos de parada.

Desde un primer momento se detectó como problema específico los juegos de caracteres que se reconocían. Eran tiempos muy anteriores a ASCII e incluso a EBCDIC, el juego

---

<sup>64</sup> Sustainability of Digital Formats Planning for Library of Congress Collections <http://www.digitalpreservation.gov/formats/> [Consultado el 21 de febrero de 2010]  
Puede completarse con DROID (Digital Record Object Identification), aplicación de código abierto para la identificación automática de los formatos de ficheros. <http://droid.sourceforge.net/> [Consultado el 21 de febrero de 2010]

<sup>65</sup> [http://www.mcu.es/bibliotecas/MC/ConsejoCB/GruposTrabajo/GT\\_Colecciones.html](http://www.mcu.es/bibliotecas/MC/ConsejoCB/GruposTrabajo/GT_Colecciones.html) [Consultado el 21 de febrero de 2010]

<sup>66</sup> Directrices para proyectos de digitalización: de colecciones y fondos de dominio público en particular para aquellos custodiados en bibliotecas y archivos / estas directrices son el resultado del trabajo de un grupo de expertos de IFLA e ICA (International Council on Archives) invitados por la UNESCO; traducidas por el grupo de trabajo de colecciones digitales de las Comunidades Autónomas y el Ministerio de Cultura. – Madrid: Ministerio de Cultura, Secretaría General Técnica, D.L. 2005. – 183 p.; 21 cm – ISBN 84-8181-271-4

de caracteres que poseía específicamente IBM. Naturalmente se estaba muy lejos de la norma ISO/IEC 10646:2003 *Information technology -- Universal Multiple-Octet Coded Character Set (UCS)*<sup>67</sup> para todos los procesos relacionados con textos escritos en todas las lenguas del mundo. Unicode es un subconjunto de la citada norma que mantiene *The Unicode Consortium*<sup>68</sup> cuya última versión es la 5.2.0<sup>69</sup>. Sin embargo, a pesar de esta normativa existen considerables problemas en el intercambio de información relacionados con el juego de caracteres y si, además, se considera que dichos problemas pueden aparecer en aquellos términos específicamente susceptibles de ser considerados como elementos de campos de búsqueda privilegiados, como el autor, el título, el problema puede llegar a ser definitivo e irresoluble. Un caso específico de este problema, que no es raro en bibliotecas virtuales, es la representación de las diferentes transliteraciones que puede presentar un nombre propio, pongamos por caso, por lo que los registros de autoridad deberían considerar todas las variantes posibles.

Este efecto de los errores en las transliteraciones puede observarse en los grandes catálogos. Ya se ha indicado en el Epígrafe 4 que la creación de un repositorio OAI puede permitir la recolección de los metadatos de la biblioteca virtual por un recolector tan potente como *OAIster* que por su intermediación nos lleva a un catálogo del volumen que presenta *WorldCat*<sup>70</sup>, en el que en octubre de 2009 se integró la base de datos *OAIster*. También de gran tamaño es el conjunto de catálogos a los que se accede mediante el *Karlsruher Virtueller Katalog KVK*<sup>71</sup>, en el que con muchísima frecuencia el usuario podrá percibir cómo en las respuestas aparecen palabras incorrectamente transcritas, de lo que fácilmente podrá deducir que su respuesta se verá limitada por el hecho de que en determinados casos el término que buscaba el lector se ha reconocido incorrectamente.

El mayor problema que presenta el OCR es que, aunque en muchos casos se logra una elevada tasa de precisión<sup>72</sup> en el reconocimiento de los caracteres, ésta raramente llega a ser del 100 % y con frecuencia es necesario utilizar conjuntos de metadatos para clarificar y perfeccionar las búsquedas. Una solución puede ser, desde luego, crear los metadatos descriptivos del texto o bien hacer la búsqueda sobre el texto reconocido ópticamente, aunque probablemente sea la combinación de ambas lo que puede asegurar la mayor fiabilidad y precisión en las respuestas. Es preciso tener en cuenta que las bibliotecas virtuales se están moviendo en la actualidad en escenarios de búsqueda como prácticamente no pudieron ser jamás soñados.

<sup>67</sup> [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=39921](http://www.iso.org/iso/catalogue_detail.htm?csnumber=39921) [Consultado el 15 de febrero de 2010].

<sup>68</sup> <http://www.unicode.org> [Consultado el 15 de febrero de 2010].

<sup>69</sup> <http://www.unicode.org/versions/Unicode5.2.0/> [Consultado el 15 de febrero de 2010].

<sup>70</sup> 72.000 bibliotecas, 174.618.797 registros bibliográficos y 1.544.467.117 registros de fondos y localizaciones a fecha de 19 de febrero de 2010

<sup>71</sup> [http://www.ubka.uni-karlsruhe.de/kvk/kvk/kvk\\_span.html](http://www.ubka.uni-karlsruhe.de/kvk/kvk/kvk_span.html) [Consultado el 15 de febrero de 2010].

<sup>72</sup> Como ejemplo, la nueva interfaz de *Gallica* ofrece para cada publicación el porcentaje de reconocimiento óptico de caracteres, avisando con ello a los usuarios de su influencia en las búsquedas. <http://gallica.bnf.fr/ark:/12148/bpt6k70861t.texte.r=les+fleurs+du+mal.f2.langEN> [Consultado el 19 de febrero de 2010]

Por otra parte, no se debe obviar uno de los mayores problemas que tiene el OCR, cuando se pretende aplicar a cualquier tipo de página. En efecto, y ya desde un primer momento (la Biblia de Gutenberg está impresa a dos columnas) los caracteres que conforman las palabras, que componen las frases y que articulan los textos están segmentados por todo tipo de soluciones tipográficas, ilustraciones, versos, ladillos, que naturalmente el proceso de OCR ha de solventar tratando de una forma segmentada esa estructura compleja que el lector fácilmente puede interpretar, pero que para un proceso típicamente secuencial como el del OCR, presenta considerables dificultades.

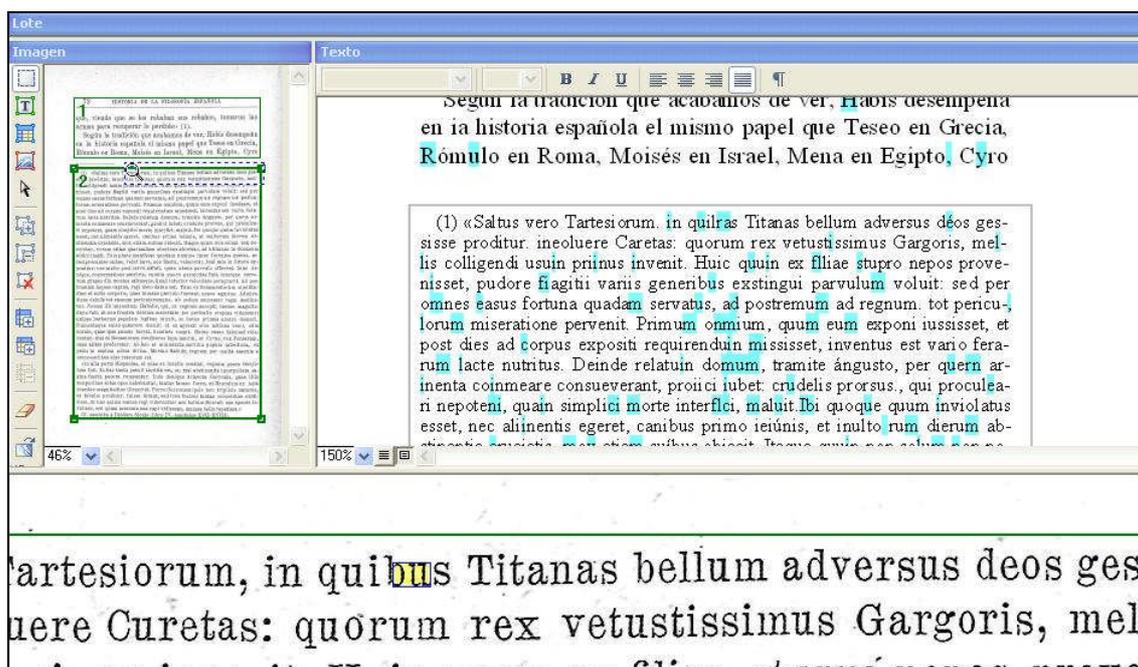


Figura 11. Detección de posibles errores en un programa de reconocimiento óptico de caracteres

Es evidente que a la vista de estos problemas y aún siendo extraordinaria la enorme facilidad y el gigantesco progreso que ha supuesto el OCR para el procesamiento informático de textos, es necesario añadir determinadas técnicas de mejora que van desde el uso de diccionarios especializados, la aplicación de determinados algoritmos que devuelvan la tasa de frecuencia de un error o incluso sistemas avanzados de tratamiento y proceso de la información que a modo de un sistema experto proponen una interpretación de la lectura de determinado carácter que validada por quién lleva a cabo la operación, retroalimenta todo el sistema. En cualquier caso, son muchos los procedimientos que existen para mejorar el proceso de OCR y sus resultados y para conocer a fondo la situación conviene seguir el proyecto IMPACT<sup>73</sup>, promovido por la Unión Europea dentro del VII Programa Marco, dónde se presenta un abanico de procedimientos y métodos para optimizar este proceso.

También es muy interesante cómo el fenómeno de las folksonomías y la anotación social se ha podido aprovechar para lo que podríamos denominar depuración

<sup>73</sup> <http://www.impact-project.eu/home/> [Consultado el 15 de febrero de 2010].

cooperativa, en la cual los propios lectores de los textos corrigen los errores evidentes e incluso proponen lecturas cuando los errores no son tan evidentes. Estos procedimientos se han aplicado en muchos proyectos y sin duda destaca entre ellos el que se lleva a cabo en *Australian Newspapers Digitization Programme*<sup>74</sup> de la *National Library of Australia* dónde la depuración de la información por parte de los lectores ha cambiado radicalmente las expectativas sobre el uso de herramientas sociales de este tipo y que ha tenido una influencia impensable, según reconocen los responsables del mismo, en los costes humanos y económicos del proyecto.

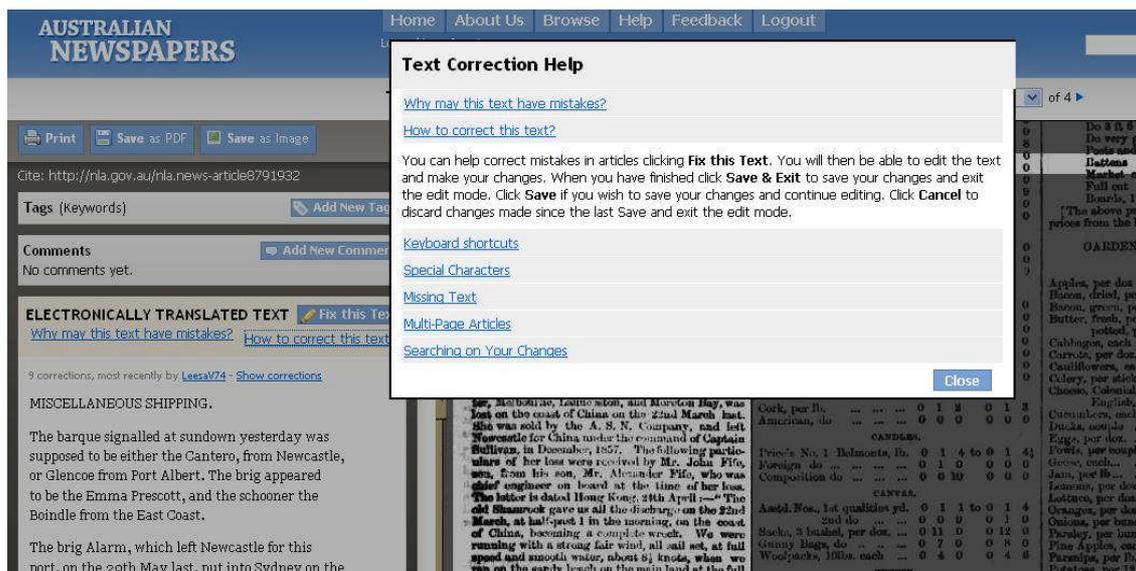


Figura 12. Interfaz de corrección social del proyecto *Australian Newspapers Digitization Programme*

Un problema que siempre ha preocupado en los proyectos de OCR ha sido el de, simultáneamente, reconocer los distintos cuerpos de las diferentes tipografías que pueden coincidir en una página y ajustar la caja (lo que por desgracia se viene a llamar ‘*layout*’ en español) al mismo tiempo que se conserva la estructura original. Por otro lado, uno de los complementos más importantes en el uso de OCR es la posibilidad de estructurar los textos obtenidos en ficheros XML<sup>75</sup> conforme a esquemas específicos, lo

<sup>74</sup> [http://www.nla.gov.au/ndp/project\\_details/](http://www.nla.gov.au/ndp/project_details/) [Consultado el 15 de febrero de 2010].

<sup>75</sup> Hernández Carrascal, Francisca, y Xavier Agenjo Bullón. “Normativa XML para bibliotecas.” *Boletín de la ANABAD* Ejemplar dedicado a Recopilación de normas internacionales en Bibliotecas / coord. por Pilar Domínguez Sánchez. (2002)Tomo 52, nº 4 : 97-122.

Se observará que tanto este artículo como el resto de los que aparecieron en ese número del *Boletín de la Anabad* dedicado a *las Normas internacionales en bibliotecas. Análisis, estado de la cuestión y perspectiva* y coordinado por Pilar Domínguez, recoge un amplio conjunto de normas a las que hay que añadir algunas de las que se mencionan a lo largo de este trabajo. Pero es esencial recalcar que buena parte de las normas que se aplican a las bibliotecas virtuales, si no todas, son las mismas, oportunamente actualizadas para ajustarse a una realidad cambiante. Se quiere con ello dar un mentís a la opinión bastante extendida de la proliferación de nueva normativa. Esto no es cierto, por el contrario existe una permanencia de las mismas y una constante actualización. Justamente lo que ha conseguido la comunidad bibliotecaria es actualizar el conjunto normativo (desde la Z39.2) lo que sin duda ha permitido su extensión a casi todo el mundo, lo que no ocurre en casi ningún otro terreno. Así puede verse cómo un ejemplo básico para multitud de aplicaciones como es el Sistema Métrico Decimal no está extendido en

cual hace que el texto reconocido y codificado de forma normalizada, pueda ser tratado por muy diferentes procedimientos. La solución más extendida para ambas cuestiones es ALTO (*Analyzed Layout and Text Object (ALTO) XML Schema*)<sup>76</sup> que se desarrolló originalmente para el proyecto *METAE* y que ha obtenido un reconocimiento como norma de facto al estar mantenida por la *Network Development and MARC Standards Office* de la *Library of Congress* desde el 10 de agosto de 2009<sup>77</sup>. En el comité editorial de esta norma que se encuentra en estos momentos en la versión 2.0<sup>78</sup> figuran representantes de los mayores proyectos de digitalización del mundo, *CCS Content Conversion Specialists GmbH*, como creadores del esquema, y bibliotecas tan señaladas como la *British Library*, la *National Library of Finland*, la *University of Kentucky*, la *Library of Congress*, la *OCLC* y la *Koninklijke Bibliotheek*. Por si esto fuera poco, ALTO, que en último término, es un esquema XML en el cual cada palabra o carácter está perfectamente ubicado en su página mediante el establecimiento de sus correspondientes coordenadas, puede encapsularse, como otros esquemas, dentro de *METS*<sup>79</sup> y así transmitirse de la misma manera en que *METS* transmite, por ejemplo, ficheros de imágenes.

---

todo el mundo. De hecho, si se compara la normativa citada en los artículos de las notas 5 y 6, o incluso con el capítulo citado en la nota 7, con la que aparece en este artículo se verá hasta qué punto son idénticas.

<sup>76</sup> <http://www.loc.gov/standards/alto/>. Véase también <http://www.content-conversion.com/alto/> [Consultado el 15 de febrero de 2010].

<sup>77</sup> <http://www.loc.gov/standards/alto/news.php#lc-maintenance> [Consultado el 15 de febrero de 2010].

<sup>78</sup> <http://www.loc.gov/standards/alto/alto-v2.0.xsd> [Consultado el 15 de febrero de 2010].

<sup>79</sup> <http://www.loc.gov/standards/alto/techcenter/use-with-mets.php> [Consultado el 15 de febrero de 2010].

```

</a:Styles>
<a:Layout>
  <a:Page ID="PAG0001" HEIGHT="3174" WIDTH="2221" PHYSICAL_IMG_NR="1" PROCESSING="OCRPROC001">
    <a:PrintSpace HEIGHT="3174" WIDTH="2221" HPOS="0" VPOS="0">
      <a:TextBlock ID="PAG001BLK0001" HEIGHT="197" WIDTH="1303" HPOS="223" VPOS="64" STYLEREF="PARSTYLE0001">
        <a:TextLine ID="PAG001LIN0001" WIDTH="1302" HEIGHT="196" HPOS="223" VPOS="64">
          <a:String CONTENT="LA" HEIGHT="180" WIDTH="285" HPOS="223" VPOS="80" WC="1" STYLEREF="TEXTSTYLE0001"/>
          <a:SP WIDTH="127" HPOS="635" VPOS="72"/>
          <a:String CONTENT="AUTONO" HEIGHT="184" WIDTH="890" HPOS="635" VPOS="64" WC="1" STYLEREF="TEXTSTYLE0001"/>
        </a:TextLine>
      </a:TextBlock>
      <a:TextBlock ID="PAG001BLK0002" HEIGHT="53" WIDTH="1052" HPOS="570" VPOS="290" STYLEREF="PARSTYLE0002">
        <a:TextLine ID="PAG001LIN0002" WIDTH="1051" HEIGHT="52" HPOS="570" VPOS="290">
          <a:String CONTENT="DIARIO" HEIGHT="42" WIDTH="117" HPOS="570" VPOS="300" WC="1" STYLEREF="TEXTSTYLE0002"/>
          <a:SP WIDTH="34" HPOS="721" VPOS="300"/>
          <a:String CONTENT="REPUBLICANO" HEIGHT="42" WIDTH="246" HPOS="721" VPOS="297" WC="1" STYLEREF="TEXTSTYLE0002"/>
          <a:SP WIDTH="36" HPOS="1003" VPOS="297"/>
          <a:String CONTENT="DEFENSOR" HEIGHT="42" WIDTH="180" HPOS="1003" VPOS="294" WC="1" STYLEREF="TEXTSTYLE0002"/>
          <a:SP WIDTH="37" HPOS="1220" VPOS="296"/>
          <a:String CONTENT="DEL" HEIGHT="40" WIDTH="62" HPOS="1220" VPOS="296" WC="1" STYLEREF="TEXTSTYLE0002"/>
          <a:SP WIDTH="35" HPOS="1317" VPOS="297"/>
          <a:String CONTENT="PARTIDO" HEIGHT="41" WIDTH="150" HPOS="1317" VPOS="295" WC="1" STYLEREF="TEXTSTYLE0002"/>
          <a:SP WIDTH="37" HPOS="1504" VPOS="290"/>
          <a:String CONTENT="ÚNICO," HEIGHT="48" WIDTH="117" HPOS="1504" VPOS="290" WC="1" STYLEREF="TEXTSTYLE0002"/>
        </a:TextLine>
      </a:TextBlock>
      <a:TextBlock ID="PAG001BLK0003" HEIGHT="35" WIDTH="578" HPOS="795" VPOS="398" STYLEREF="PARSTYLE0003">
        <a:TextLine ID="PAG001LIN0003" WIDTH="577" HEIGHT="34" HPOS="795" VPOS="398">
          <a:String CONTENT="DE" HEIGHT="30" WIDTH="61" HPOS="795" VPOS="402" WC="1" STYLEREF="TEXTSTYLE0003"/>
          <a:SP WIDTH="22" HPOS="878" VPOS="404"/>
          <a:String CONTENT="AVISOS" HEIGHT="31" WIDTH="178" HPOS="878" VPOS="401" WC="1" STYLEREF="TEXTSTYLE0003"/>
          <a:SP WIDTH="28" HPOS="1084" VPOS="400"/>
          <a:String CONTENT="Y" HEIGHT="29" WIDTH="29" HPOS="1084" VPOS="400" WC="1" STYLEREF="TEXTSTYLE0003"/>
          <a:SP WIDTH="29" HPOS="1142" VPOS="398"/>
          <a:String CONTENT="NOTICIAS" HEIGHT="33" WIDTH="230" HPOS="1142" VPOS="398" WC="1" STYLEREF="TEXTSTYLE0003"/>
        </a:TextLine>
      </a:TextBlock>
    </a:Page>
  </a:Layout>
</a:Document>

```

Figura 13. Ejemplo de marcado de un fichero ALTO

La distribución de objetos digitales a texto completo en las bibliotecas virtuales puede hacer uso de otros formatos como PDF, normalmente ofrecidos como ficheros de imagen con el texto obtenido del proceso de OCR oculto. Esto permite su indización y, por tanto, la búsqueda y recuperación a texto completo al igual que ocurre con los ficheros ALTO. Para la preservación a largo plazo de documentos electrónicos se recomienda el uso de PDF/A como define la norma *ISO 19005-1:2005 Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1)*<sup>80</sup>.

Tanto por razones de preservación a largo plazo de los documentos digitales en formato texto, como por una buena gestión de las bibliotecas virtuales es necesario consignar en los metadatos administrativos las características técnicas de los ficheros de texto, lo cual debe hacerse por medio del esquema *textMD*. El esquema *textMD*<sup>81</sup> permite detallar las propiedades de los ficheros texto como la codificación (calidad, plataformas, software, etc.), los caracteres (juego de caracteres, cuerpo de los mismos, códigos de control, etc.), el idioma, la letrería, los lenguajes de marcado utilizados y cualquier requisito técnico necesario para la correcta impresión, visualización y secuenciación de páginas. En definitiva, todos los datos que registran cómo se ha elaborado un documento textual y sus características para garantizar su uso en el tiempo.

Los esquemas mencionados hasta ahora hacen referencia o bien a la disposición del texto en una página referenciándolos por su relación con las coordenadas de la imagen

<sup>80</sup> [http://www.iso.org/iso/catalogue\\_detail?csnumber=38920](http://www.iso.org/iso/catalogue_detail?csnumber=38920)[Consultado el 15 de febrero de 2010].

<sup>81</sup> <http://www.loc.gov/standards/textMD/>[Consultado el 15 de febrero de 2010].

(ALTO) o bien a las propiedades técnicas de los ficheros texto. Sin duda, es posible una mayor explotación de los textos por medio de la utilización de marcas lógicas. Como marcas lógicas entendemos un conjunto de marcas que permiten reflejar la estructura de un texto y sus divisiones (páginas, títulos, epígrafes, secciones, párrafos, notas, etc.), el tipo de textos (prosa, teatro, verso, cartas), el registro de variantes lingüísticas de una palabra, la normalización de términos (nombres de persona, de entidades, geográficos) o el enlace de términos y textos con otros documentos o datos externos al contenido marcado, pero estrechamente relacionado con él. Este conjunto de marcas, junto con recomendaciones para su aplicación es el objeto de la *Text Encoding Initiative*<sup>82</sup> que desde 1994 ha venido actualizando las *TEI Guidelines* para el marcado de textos digitales y que constituye un compendio, junto con el esquema TEI, de edición digital.

Aunque probablemente demasiado complejo para la edición manual de un libro electrónico a partir de un objeto digital contenido en una biblioteca virtual, es previsible que se desarrolle un subconjunto de etiquetas TEI (al modo de *TEILite*) dedicado específicamente a ello. De hecho, en las últimas intervenciones que pueden leerse en la lista TEI-L hay muchas en esta línea que expresan el manifiesto interés de quienes han editado determinados textos complejos en generar ficheros ePub. Otra muestra de ello es el *TEI Special Interest Group on Scholarly Publishing*<sup>83</sup>, entre cuyos objetivos se hallan los de establecer recomendaciones para generar diferentes formatos de ficheros a partir de TEI como HTML, PDF y formatos de libros electrónicos como ePub. Igualmente, ya se han desarrollado algunos conversores para pasar de TEI a ePub (p.e. *tei2epub*<sup>84</sup>) aunque dada la flexibilidad de la norma no serán exhaustivos. Sin duda, la amplísima experiencia de este colectivo de editores digitales va a ser muy útil cuando la edición de libros electrónicos requiera de una mayor sutileza, del mismo modo que se ya se ha utilizado este conocimiento para las ediciones científicas (de hecho, existe ya desde hace tiempo un conversor de TEI para LaTeX, muy usado en ciencias exactas e ingenierías).

## 7. Pro captu lectoris habent sua fata libelli: los ciberlibros<sup>85</sup> o los eBooks<sup>86</sup>

Una nueva funcionalidad que va a tener una decisiva importancia es la posibilidad de descargar desde una biblioteca virtual un libro electrónico. Los bibliotecarios estamos acostumbrados a generar subproductos de las bases de datos bibliográficas desde su

<sup>82</sup> <http://www.tei-c.org/index.xml> [Consultado el 15 de febrero de 2010].

<sup>83</sup> <http://www.tei-c.org/Activities/SIG/Publishing/> [Consultado el 15 de febrero de 2010].

<sup>84</sup> *epub-tools*: Collection of open source (BSD) tools for generating and managing ePub documents <http://code.google.com/p/epub-tools/> [Consultado el 19 de febrero de 2010].

<sup>85</sup> Véase Fundación del Español Urgente (Fundéu BBVA) <http://www.fundeu.es/Recomendaciones.aspx?frmOpcion=RECOMENDACION&frmFontSize=2&frmIdRecomendacion=519> [Consultado el 19 de febrero de 2010]

<sup>86</sup> Es muy asequible la entrada en el blog *Anatomía de la edición: estudios sobre el sector editorial* titulado *12 hitos de la edición digital*. <http://www.anatomiadelaedicion.com/2010/02/12-hitos-edicion-digita/> [Consultado el 15 de febrero de 2010]

nacimiento. Por paradójico que pudiera resultar, el primer subproducto que se obtuvo fue el de la ficha manual. Poco a poco ese subproducto se fue modificando y, aunque sin abandonar la ficha tradicional, fue posible obtener, mediante diversos procedimientos, distintos tipos de información codificada. Así se empezaron a generar cintas de fotocomposición para la edición de bibliografías y catálogos; cintas conforme a ISO 2709 para favorecer la catalogación cooperativa (más tarde mediante servidores Z39.50 y ahora, afortunadamente, mediante SRU); metadatos Dublin Core con el objetivo, entre otros, de generar repositorios OAI-PMH; registros METS para intercambiar información codificada sobre todos los aspectos de los objetos digitales y la ubicación de los mismos; documentos primarios de artículos de revistas, capítulos de libros y otros materiales bibliográficos, en modo imagen, en PDF, para direccionarlo a una impresora o volcarlo en un ordenador; y ahora, con los dispositivos móviles, para descargar el documento primario en un lector de libros electrónicos.

Aunque el eBook, ciberlibro, libro electrónico o libro digital está en pleno desarrollo tecnológico, en el cambio de año de 2009 a 2010 comenzaron a aparecer estadísticas y análisis del mercado editorial en el que se daba cuenta de que por primera vez los libros electrónicos habían superado, para determinados libros, las ventas de los libros en papel.

No es propósito de este artículo analizar los problemas que pueden existir y que, con seguridad, se van a dar entre los libros electrónicos producidos como resultado de los proyectos de digitalización y el sector editorial. El conflicto va a ser presumiblemente largo y complejo como ya se ha dado, y aún no se ha resuelto, entre *Google Books* y el mundo editorial. Suponer que otros proyectos de bibliotecas digitales, por partir de colecciones públicas y ser editados por entidades sin ánimo de lucro como son las bibliotecas no va a suponer también un problema es pecar de ingenuidad. Por citar dos ejemplos claros, si existe la posibilidad de descargar *Les fleurs du mal* de *Gallica*, la biblioteca digital de la *Bibliothèque Nationale de France*, o las obras de Charles Dickens de la *British Library*, gratuitamente y en cualquiera de los formatos más estandarizados y a dispositivos claramente accesibles en el mercado, es imposible no esperar una reacción por parte del sector editorial. En el caso de España otros procedimientos tradicionales de acceder al contenido de los libros como son la fotocopia e incluso, por asombroso y antisocial que nos pueda parecer, el préstamo, están siendo objeto de todo tipo de cánones y gabelas.

Por lo tanto, y en mucho mayor medida, ocurrirá lo mismo con los libros electrónicos. Sin embargo, difícilmente las bibliotecas pueden renunciar a este proceso de creación de objetos digitales complejos a partir de los libros conservados en sus estanterías por ser consustancial a los objetivos biblioteconómicos facilitar al lector el acceso a la información ("*Save the time of the user*"). Y, además, esto no debe olvidarse, porque es imprescindible salvar al menos el contenido de la práctica totalidad<sup>87</sup> de los libros y

---

<sup>87</sup> Lógicamente, están exentos de este peligro los libros editados en ejemplares para bibliófilos en papel verjurado y aquellos, poquísimos (sobre todo en España), que siguen la normativa para papel permanente ISO 9706:1994 "Información y documentación - Papel para documentos - Requerimientos para permanencia", a pesar de su enorme utilidad. Si las bibliotecas españolas requirieran a los editores españoles la impresión de textos en papel permanente como hacen las bibliotecas norteamericanas, sin duda se avanzaría en esta línea.

[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=17562](http://www.iso.org/iso/catalogue_detail.htm?csnumber=17562) [Consultado el 14 de febrero de 2010]

otros materiales bibliográficos impresos a partir de la segunda mitad del siglo XIX y hasta nuestros días que se van deteriorando como consecuencia del carácter ácido del soporte papel en el que están impresos<sup>88</sup>.

Por otra parte, las bibliotecas virtuales son una representación de una determinada cultura lo que está constituyendo un campo de batalla entre gigantes de la informática como *Google*<sup>89</sup>, gigantes de la distribución de libros como *Amazon*, y gigantes de la preservación de la cultura nacional como *Gallica*<sup>90</sup> o la *British Library*. Se trata de un mundo tremendamente cambiante, en el cual los nuevos dispositivos de lectura (ciberlibro, teléfonos móviles, *tablets*, etc.) están provocando un cambio de paradigma en la edición de libros y textos que está sacudiendo el mundo editorial, la distribución de libros, las bibliotecas, incluidas las bibliotecas digitales y virtuales, y las grandes empresas de Internet como *Google*, *Apple*, *Yahoo* o *Microsoft*. La muestra de los cambios que se están avecinando es que ninguna de las grandes empresas de Internet es ajena a ella. A las propuestas como *Google Books*, que se ha convertido por arte de la tecnología y de una aplicación muy laxa de la gestión de derechos de autor en el primer editor del mundo con casi 10 millones de títulos, se han opuesto en muchos países algunas iniciativas que intentan controlar la edición y difusión en Internet de la cultura escrita de un país. Ahí están las protestas de la China, de la India, de Francia, y de la Unión Europea, como entidad administrativa, y la práctica totalidad de los editores del mundo, en contra de un monopolio *de facto* ejercido por *Google* sobre la cultura de la humanidad. Por contra, Europeana con casi 6 millones de objetos digitales, *Gallica*, o la *British Library*<sup>91</sup> (en convenio con *Microsoft*) no quieren que sea un único proveedor quien decida qué y cómo editar la cultura de un país, con el peligrosísimo olvido, por cierto, de la preservación a largo plazo.

La presencia anglosajona en *Google Books* es abrumadora, lo que equivale también a decir que la presencia de la cultura hispana es ridícula. Si sólo existe como referencia para el acceso a la cultura hispana lo que *Google Books* pueda ofrecer, será evidente que las instituciones españolas, si no hacen otra cosa, ceden su capacidad para generar

---

<sup>88</sup> Agenjo Bullón, Xavier. "La cuarta salida de El monje digital y sus problemas hemerográficos : una recapitulación." Boletín de la ANABAD. Tomo 54, nº 4 (2004): 119-138.

<sup>89</sup> Aunque se cita permanentemente a Google, no siempre se tienen en cuenta los análisis sobre algo que es mucho más que un programa de búsqueda en la Web. Para este propósito nos interesan tres títulos:

Jeanneney, Jean-Noël. *Quand Google défie l'Europe. Plaidoyer pour un sursaut*. Paris, Mille et une nuits, 2006 [Hay traducción española publicada por la Universidad de Valencia en 2007]

Jarvis, Jeff. *What would Google do?*. New York, HarperCollins, 2009

Vise, David A. *The Google story*. Updated edition for Google's 10<sup>th</sup> Birthday. London, Pan Books, 2008.

<sup>90</sup> Es muy interesante el Rapport sur la numérisation du patrimoine écrit de Marc Tessier preparado para el Ministre de la Culture et de la Communication de Francia, el 12 de enero de 2010. Conviene contrastar este informe con el libro de Jeanneney que, no se olvide, era Director de la Bibliothèque Nationale de la France cuando lo escribió.

<http://www.culture.gouv.fr/mcc/content/download/3520/23115/file/Rapport%20sur%20la%20numerisation%20du%20patrimoine%20ecrit.pdf> [Consultado el 14 de febrero de 2010]

<sup>91</sup> <http://www.bl.uk/news/2005/pressrelease20051104.html> [Consultado el 15 de febrero de 2010].

contenidos digitales en español y difundir la cultura hispánica. Otros ejemplos como la *World Digital Library*<sup>92</sup>, *Internet Archive*<sup>93</sup> o el *proyecto Gutenberg*<sup>94</sup>, presentan los mismos desequilibrios en cuanto a la presencia de la cultura hispana en sus bases de datos, lo que no es de extrañar porque en estas iniciativas no existe una presencia ni española ni iberoamericana. Esto es lo que está ocurriendo en la actualidad, en un futuro próximo el problema se agravará por la falta de control de los procesos de edición, preservación digital a largo plazo, investigación lingüística o desarrollos informáticos lingüísticos adaptados al español y realizados por españoles o hispanos (segunda lengua del mundo por número de habitantes). En este panorama seguiremos dependiendo de los análisis, los desarrollos y los planes de empresas totalmente ajenas a la cultura española, tanto en su conocimiento como en su gestión.

Pero, dónde hay que hacer especial hincapié es en el campo de trabajo representado por las áreas mencionadas y por la edición digital para la Web, para dispositivos móviles, para libros electrónicos, para nuevos modos de edición, etc., que pueden suponer un cambio en la cultura española justamente a través del medio más extendido y que más está cambiando: Internet. El fenómeno del ciberlibro es tan imparable como lo fue la música en soporte digital y transmisible por Internet en la primera década de este siglo. Si se deja este terreno libre, otros lo harán.

Aunque habían sido ya numerosos los proyectos dados a conocer, la edición de libros digitales tomó una plena realidad cuando el mayor distribuidor de libros del mundo, *Amazon*, empezó a comercializar un aparato lector denominado *Kindle*. No nos interesa aquí extendernos sobre las funcionalidades de estos dispositivos, su alcance y sus límites, aunque los autores de este trabajo prevén una rapidísima difusión de los lectores de libros electrónicos (*eReaders*), sobre todo entre la población más joven. El formato que ha venido utilizando *Kindle* desde un primer momento tiene una estructura cerrada, se denomina MOBI<sup>95</sup> y parece destinado a tener un considerable éxito ya que *Amazon* es con diferencia el mayor distribuidor mundial de libros, dentro de lo que podría denominarse B2C. En realidad, *Amazon* usa el formato AZW, prácticamente igual a MOBI. Un aspecto esencial para el predominio de determinado formato es disponer de una gran colección de información que utilice esta estructura y es previsible que el formato MOBI se afirme con proyectos como el anunciado en enero de 2010 por la *British Library* de poner 65.000 libros<sup>96</sup> a disposición de los usuarios de forma gratuita y utilizando justamente la estructura MOBI.

---

<sup>92</sup> <http://www.wdl.org/es/> [Consultado el 14 de febrero de 2010]

<sup>93</sup> <http://www.archive.org/index.php> [Consultado el 14 de febrero de 2010]

<sup>94</sup> [http://www.gutenberg.org/wiki/Main\\_Page](http://www.gutenberg.org/wiki/Main_Page) [Consultado el 14 de febrero de 2010]

<sup>95</sup> Una excelente descripción y análisis de los formatos para e-books puede encontrarse en la entrada Wikipedia contributors, "Comparison of e-book formats", Wikipedia, The Free Encyclopedia, 11 February 2010, 06:01 UTC, [http://en.wikipedia.org/w/index.php?title=Comparison\\_of\\_e-book\\_formats&oldid=343300175](http://en.wikipedia.org/w/index.php?title=Comparison_of_e-book_formats&oldid=343300175) [Consultado el 14 de febrero de 2010].

<sup>96</sup> Una información detallada puede leerse en Brindley, Lynne. "A new global readership for forgotten literary gems." *The Times* (UK) 7 Feb 2010.

No se nos acusará de cinismo si afirmamos que *Microsoft*, la empresa que ha colaborado con la *British Library* para este logro, no es que esté interesada especialmente en favorecer a *Amazon*, sino más bien en oponerse al proyecto *Google Editions* que se lanzará en la primera mitad de 2010 como librería virtual con más de 500.000 títulos y que utilizará los formatos ePub y PDF, mucho más abiertos que el anterior. Una solución salomónica es la adoptada por *Gallica*, por *Internet Archive* y otros que permite la descarga en ambos formatos, tanto MOBI como ePub, a los que añade HTML y PDF.

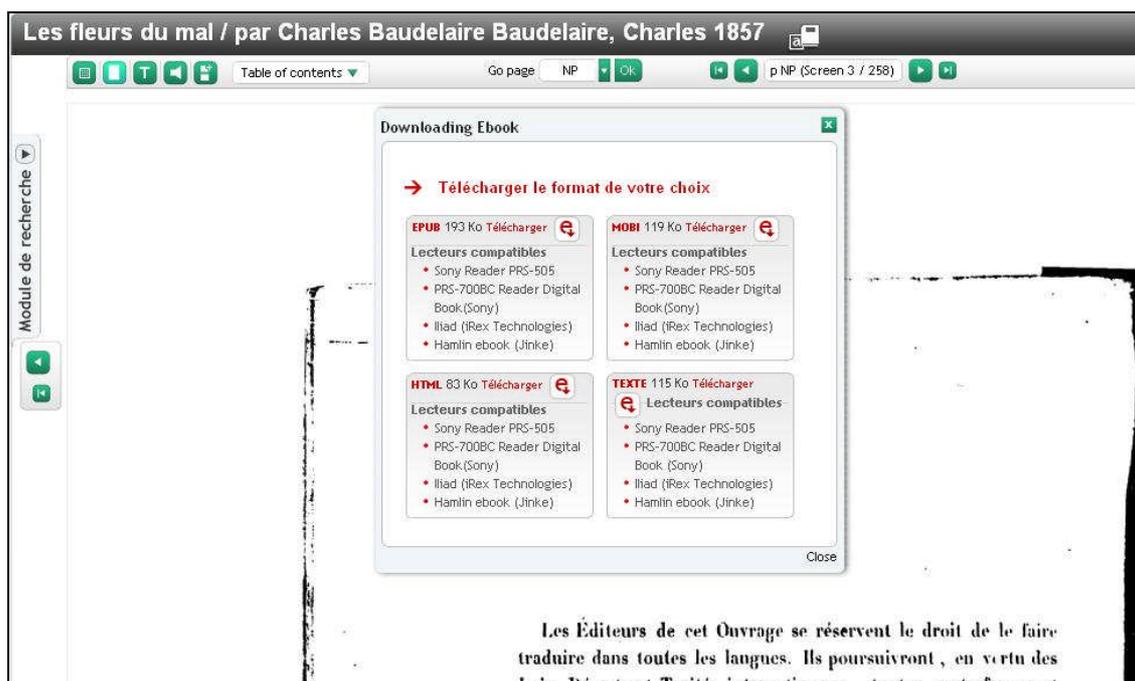


Figura 14. Ejemplo de descarga de libros electrónicos de Gallica

Para analizar algo más la estructura de *ePub* conviene recordar que se trata de una iniciativa emprendida por la *International Digital Publishing Forum (IDPF)* y que consiste en un archivo ZIP que contiene 3 ficheros XML, *Open Publication Structure (OPS)*, *Open Packaging Format (OPF)* y *Open Container Format (OCF)*. Es muy importante señalar que no sólo Google utiliza o va a utilizar *ePub* para la descarga de sus libros, sino que también una entidad tan potente como Internet Archive a través de *Bookserver*<sup>97</sup>, utiliza también el formato *ePub* en la mayoría de los casos, aunque también utiliza MOBI, PDF, TXT, etc. Por poner un ejemplo claro, la práctica totalidad de las obras de Galdós están accesibles en *Bookserver* en formato MOBI y *ePub*.

[http://entertainment.timesonline.co.uk/tol/arts\\_and\\_entertainment/books/article7017994.ece](http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/books/article7017994.ece) [Consultado el 14 de febrero de 2010]

<sup>97</sup> <http://www.archive.org/bookserver> [Consultado el 14 de febrero de 2010]

Sin duda alguna, la revolución que está suponiendo ya la nueva realidad del libro electrónico<sup>98</sup> modificará de forma sustancial la estructura de la biblioteca, así como sus funciones, si bien y como se ha dicho anteriormente no supondrá en absoluto la desaparición de la biblioteca como elemento intermediario, primero por su capacidad de orientar al lector, en este caso al futuro descargador de libros electrónicos, sobre las fuentes más interesantes y de mejor calidad –los libros de Galdós estarán disponibles en más de una biblioteca virtual. El problema se dará en aquellos casos en los que el acceso a la información digital en forma de libro digital sea de pago. Habrá, pues, que establecer ese puente democrático citado y asumir parcialmente el coste por procedimientos que no nos atrevemos a aventurar, pero que probablemente supondrán el pago de determinada tarifa como en la actualidad ocurre con la obtención de fotocopias, el canon digital, o el propio préstamo de libros. Será muy conveniente que todo ello sea objeto de un pacto y no de una imposición, de una gabela, cobrada además por particulares, en claro retroceso de la modernidad que, entre otras cosas supuso la eliminación del cobro delegado de impuestos.

The screenshot shows the Bookserver interface for the book 'Doña Perfecta (1903)'. The page is divided into several sections:

- View the book:** Contains a thumbnail of the book cover and a list of download options: Read Online (~330 pg), PDF (Google.com) (11 MB), EPUB (beta) (~330 pg), Kindle (beta) (~330 pg), Daisy (beta) (~330 pg), and Metadata (29 kB). It also includes 'All Files: HTTP' and a 'Help reading texts' link.
- Resources:** Includes a 'Bookmark' link.
- Doña Perfecta (1903) details:**
  - Author: Pérez Galdós, Benito, 1843-1920; Lewis, Edwin Seelye, 1868-1935
  - Publisher: New York : American Book Company
  - Year: 1878
  - Possible copyright status: NOT\_IN\_COPYRIGHT
  - Language: Spanish; English
  - Digitizing sponsor: Google
  - Book from the collections of: Oxford University
  - Collection: europeanlibraries
  - A link to an Open Library page: This book has an [editable web page on Open Library](#).
- Description:** Book digitized by Google from the library of Oxford University and uploaded to the Internet Archive by user tpb.
- Reviews:** A section with the prompt 'Be the first to write a review' and a note 'Downloaded 13 times'.
- Selected metadata:**
  - Identifier: doaperfecta01galdgoog
  - Scanner: google
  - Mediatype: texts
  - Identifier-access: <http://www.archive.org/details/doaperfecta01galdgoog>
  - Oclc-id: 2616979
  - Scandate: 20070223
  - Copyright-region: US
  - Identifier-ark: <ark:/13960/t8jd517m>
  - Imagecount: 300
  - Source: <http://books.google.com/books?id=yLUGAAAQAAJ&oe=UTF-8>
  - Ocr: ABBYY FineReader 8.0
  - Locn: 03026551

**Figura 15. Página de "Doña Perfecta" en Bookserver, desde donde se puede descargar en una variedad de formatos y características técnicas**

Como se ha comentado en varias ocasiones la disponibilidad de determinada solución informática no soluciona por sí misma lo que podríamos llamar en términos quizá excesivamente economistas un modelo de negocio. Por otra parte, es probable que intervengan en este modelo otros actores todavía no previstos como podrían ser los operadores telefónicos y, por supuesto, la acción impositiva del Estado, rememorando

<sup>98</sup> El 18 de febrero de 2010 se difundió el Informe sobre el Libro Digital realizado por la Federación del Gremio de Editores de España junto con la Fundación Germán Sánchez Ruipérez sobre la base de una encuesta realizada en noviembre de 2009 a numerosas editoriales españolas.

<http://www.dilve.es/dilve/getArchivoDocumentacion.do?iddocumento=921> [Consultado el 19 de febrero de 2010]

cierta célebre frase de Faraday cuando mostraba a un diputado del Parlamento inglés un generador de energía eléctrica mediante electroimanes.

En cierto sentido, la biblioteca vuelve a tomar uno de los papeles que le fueron propios en la Edad Media cuando los estudiosos se dirigían a las bibliotecas, frecuentemente monacales, para intentar conseguir una copia de un manuscrito. En efecto, no es muy diferente en cuanto a la función, copiar un manuscrito a partir de un original y descargar de la red una copia para utilizarla posteriormente en el PC o en un dispositivo móvil, como el libro electrónico. Con cierta intervención del Estado, no cabe duda de que el mercado, que en este caso es completamente global, puesto que el lector puede descargarse un libro desde cualquier sitio, en cualquier momento e incluso más, sin saber verdaderamente dónde está situado el servidor del que obtiene la información, que estará probablemente situada en la nube, irá ajustando la oferta y la demanda a partir de variables como precio y calidad.

The screenshot shows the website of Fundación Ignacio Larramendi. The main header features the foundation's logo and name. Below the header, there is a navigation bar with links for RSS, OAI Repository, SRU, and a search bar. The main content area displays a bibliographic record for 'Testamento otorgado en Santander un mes antes de morir (2009) - Menéndez Pelayo, Marcelino.' The record includes a thumbnail of the book cover, a list of export formats (MARCXML, MODS, BIBTEX, JISC, METS), and a table of metadata. The metadata table is as follows:

Sección:	Obras polígrafos
Número de control:	FIL20090008588
Autor:	Menéndez Pelayo, Marcelino. Testamento. 1912. Español (1856-1912)
Título:	Testamento otorgado en Santander un mes antes de morir / Marcelino Menéndez Pelayo
Área de datos:	Recurso en línea (1 fichero)
Publicación:	Madrid : Fundación Ignacio Larramendi, 2009 (Madrid : DIGIBÍS, S.L., 2009)
Notas:	Original: Menéndez Pelayo, Marcelino (1856-1912). Testamento otorgado en Santander un mes antes de morir. - Santander : Sociedad Menéndez Pelayo, 2000. - 1 carpeta (1h. pleg. [8] p.) : 30 cm. Tipo de archivo: Texto

**Figura 16. Registro bibliográfico con exportación de metadatos en diferentes esquemas y descarga en formatos para libros electrónicos (ePub y MOBI)**

Sin duda, existirán límites a esta utopía de biblioteca universal, o Aleph, marcados no sólo por los costes que pueda llevar aparejado este proceso y que en la actualidad se encuentran, evidentemente, en plena fase de ajuste, sino también los propios límites físicos que habrá de arrostrar en un momento dado la red. Hay que tener en cuenta que, por ejemplo, no se está tan lejos del límite empírico que señaló la Ley de Moore (y que probablemente tendría lugar hacia 2019), si bien como prevé Kurzweil, que ya ha aparecido aquí como descubridor del reconocimiento óptico de caracteres, hay alguna posibilidad de superar dicho límite y, sobre todo, y es lo que más nos interesa aquí, recordar que no es únicamente la capacidad y velocidad de los procesadores el punto en el que se basa la mayor velocidad de la red. De hecho, Kurzweil pronostica que a partir de 2020 un nuevo tipo de tecnología basada en la computación óptica o cuántica tomará

el relevo. Los recientes experimentos del español Ignacio Cirac no son tan optimistas. En cualquier caso el físico austríaco Jacob D. Bekenstein postuló que la capacidad total de información del universo<sup>99</sup>, basada en el número de partículas elementales que este contiene, tendría unos límites definidos<sup>100</sup>.

En cualquier caso, si bien no de una forma tan rápida como la sustitución de las grabaciones analógicas por las digitales y su distribución fulgurante en la red, es posible que veamos en muy poco tiempo disminuir considerablemente el número de libros impresos, aunque no se debe perder de vista que, aunque la imprenta manual y el papel continuo sustituyeron a la imprenta manual y al papel verjurado, se han seguido haciendo hasta nuestros días ediciones de bibliófilo en condiciones análogas a las que tuvieron los libros impresos hasta el primer tercio del siglo XIX y que, además, en el momento del auge de las grabaciones sonoras digitales se ha detectado un pequeño movimiento de vuelta a la estampación de discos en vinilo.

Por último, y para cerrar este apartado dedicado a los libros digitales y dentro de los subproductos de las bibliotecas virtuales, no estará de más recordar que mantiene cierta fuerza, incluso acentuada, la edición bajo demanda que en un momento dado se dió prácticamente por desaparecida. O mejor dicho, la impresión bajo demanda para satisfacer determinados gustos lectores en el soporte tradicional. Es posible, incluso, que las bibliotecas virtuales incluyan esta funcionalidad y junto a los sistemas de descarga de la información para libros digitales permitan encargar la impresión, en realidad un conjunto de fotografías, de una edición bajo demanda de libros raros de carácter facsimilar, o bien ediciones<sup>101</sup> en las que la legibilidad se ha mejorado, aumentando los márgenes, subiendo los tipos, etc.

## 8. Genus irritabile vatum: El desafío de los derechos de autor

Sin duda uno de los aspectos más vidriosos de la época actual es el relacionado con los derechos de autor y la Web. En modo alguno se pretende en este artículo dar ningún tipo de solución a cuestión tan espinosa, si bien los autores quieren imaginar que la solución o soluciones que se adopten sean grupo de un consenso o pacto entre todas las partes implicadas, en el que todos obtengan algún beneficio y todos obtengan alguna pérdida. Todos estos aspectos ligados a la teoría de la decisión son extraordinariamente complejos y como ya se demostró hace más de 60 cuando Kenneth Arrow<sup>102</sup> enunció su

---

<sup>99</sup> Bekenstein, Jacob D, y Marcelo Schiffer. "Quantum Limitations on the Storage and Transmission of Information." quant-ph/0311050 (2003). <http://arxiv.org/abs/quant-ph/0311050>. [Consultado el 14 de febrero de 2010].

<sup>100</sup> Pero incluso esos límites se están atacando ya. Puede verse un artículo tan reciente como *Going Beyond Moore's Law by Using the Third Dimension* publicado el 18 de enero de 2010. <http://www.physorg.com/news183024614.html> [Consultado el 14 de febrero de 2010]

<sup>101</sup> Aunque son muchos los ejemplos que podrían proporcionarse, conviene mencionar por la riqueza de su catálogo y la versatilidad de su fondo, el proyecto *Bibliolife*. <http://bibliolife.com/> [Consultado el 14 de febrero de 2010]

<sup>102</sup> Arrow, Kenneth Joseph. *Elección social y valores individuales*. Ministerio de Economía y Hacienda. ISBN 978-84-7196-111-2. La extensión de alguna de sus ideas sin duda aparece (así como un uso

famoso teorema, es una cuestión tan irresoluble como las planteadas por Heisenberg o Gödel en sus respectivos principios y teoremas.

En cualquier caso, no puede negarse que los derechos de autor han ido íntimamente vinculados a los medios de producción y difícilmente podrán aplicarse los mismos principios a la copia de un original mediante la imprenta manual, la imprenta mecánica<sup>103</sup> y la copia digital, especialmente si la copia digital tiene lugar en la red con lo que tiende a ser ilimitada y su coste tiende a cero<sup>104</sup>.

A partir de las premisas señaladas en las notas anteriores vamos a elegir para este artículo el modelo que supone METSRights<sup>105</sup> tanto por su calidad como por la integración con el resto de tratamientos automatizados de la información que se presentan aquí.

No debe olvidarse, sin embargo, que la actual polémica tanto en los medios de comunicación como, sobre todo, en las disputas legales mantenidas por el proyecto *Google Books* en tribunales de todo el mundo está dando origen a una jurisprudencia específica sobre este problema. Desde una perspectiva europea es muy importante citar el último documento presentado por el Comité de Alto Nivel de la Europea sobre *The Digital Library* en noviembre de 2009.

---

moderado del teorema de Minimax de Von Neumann) en la *teoría de la justicia* de John Rawls y en su concepto básico de *well-ordered society*.

<sup>103</sup> Existen varios estudios de conjunto sobre la propiedad intelectual y los más interesantes para nosotros los que tienen un punto de vista bibliográfico y bibliotecario Reyes Gómez, Fermín. *El libro en España y América : legislación y censura, siglos XV-XVIII*. Madrid: Editorial Arco/Libros, 2000.

En cualquier caso nos acogemos a la declaración de IFLA a este respecto ‘The IFLA Position on Copyright in the Digital Environment’ <http://www.ifla.org/en/publications/the-ifla-position-on-copyright-in-the-digital-environment> [Consultado 18 Febrero 2010] y a la sección creada específicamente sobre ella, el *Committee on Free Access to Information and Freedom of Expression (FAIFE)* <http://www.ifla.org/en/faife> Resulta fundamental consultar el reciente “Limitations and Exceptions to Copyright and Neighbouring Rights in the Digital Environment: An International Library Perspective.” y ‘Limitations and Exceptions to Copyright and Neighbouring Rights in the Digital Environment: An International Library Perspective’ <http://www.ifla.org/en/publications/limitations-and-exceptions-to-copyright-and-neighbouring-rights-in-the-digital-environm> [Consultado 18 Febrero 2010].

También es una referencia el Manifiesto por el Dominio Público originado en el marco del proyecto europeo *Communia (The European Thematic Network of the Digital Public Domine)* <http://www.publicdomainmanifesto.org/node/8> [Consultado 18 Febrero 2010].

<sup>104</sup> El JSCA promueve una encuesta de carácter internacional sobre el crítico problema de las denominadas obras huérfanas que es necesario consultar y seguir con atención (y si es posible participar en él) <http://sca.jiscinvolve.org/2010/02/10/spanish-version-contribute-to-international-survey-on-orphan-works/> Esta ahora accesible en español. [Consultado el 14 de febrero de 2010]

<sup>105</sup> <http://www.loc.gov/standards/mets/news080503.html>. El esquema puede encontrarse en <http://cosimo.stanford.edu/sdr/metsrights.xsd> [Consultado el 14 de febrero de 2010]

Para la elaboración del *METSRights Schema*, publicado por la *Library of Congress*<sup>106</sup> el 13 de septiembre de 2006, se tuvieron en cuenta otras iniciativas sobre la expresión de los derechos de autor como *Open Digital Rights Language* (ODRL), el *XrML REL* (*eXtensible Rights Markup Language*) o la *Digital Library Federation's Electronic Resource Management Initiative* (DLF ERMI). El *METS Editorial Board*, tras los oportunos análisis comparativos, propuso que el esquema *METSRights* se centrara en los recursos digitales controlados o propiedad de un repositorio digital; en la declaración de los derechos de autor y derechos conexos de los recursos digitales y en la identificación de los derechohabientes más que intentar expresar todos los derechos posibles para su uso en sistemas de *Digital Rights Management*, y en simplificar la declaración lo más posible dado el rápido cambio de escenario de los DRM.

Así, *METSRights* tiene 3 elementos principales: los tipos de derechos (derechos de autor, con licencia, dominio público, contractuales y otros) y la declaración asociada a cada tipo de derecho; la identificación de los derechohabientes e información de contacto; y el contexto o los permisos y tipo de usuarios asociados a un objeto digital, así como las restricciones a estos permisos.

Aunque como ya se ha dicho *METSRights* adolece de una cierta antigüedad su análisis es particularmente rico y, sobre todo, se integra perfectamente con el resto de los metadatos de un objeto digital cuyos derechos de autor se hacen patentes a través de esa declaración. Un análisis más actual y en el que también están presentes estos elementos puede encontrarse en Québec, *Laboratoire d'étude sur les politiques publiques et la mondialisation*, ENAP, 2010 14 p. (*Rapport évolutif. Analyse des impacts de la mondialisation sur la culture au Québec; Rapport 8*)<sup>107</sup>. Pero no debe de perderse de vista que el contexto europeo y por lo tanto el texto de referencia, si bien no enfoca la situación desde un punto de vista tecnológico es el informe final del *High Level Expert Group on Digital Libraries, Digital Libraries, Recommendations and Challenges for the Future*<sup>108</sup>. Llama la atención que prácticamente la mitad del informe final está relacionado con los aspectos de la propiedad intelectual en las bibliotecas digitales, si bien contiene otras recomendaciones de gran valor. Por lo tanto, y en el entorno europeo resulta muy significativo las iniciativas por la Comisión Europea en el VII Programa Marco y en el contexto del programa ICT-PSP [*Information and Communication Technologies Policy Support Programme*]<sup>109</sup> (anterior *e-Content*).

---

<sup>106</sup> Esta información puede ampliarse con el resumen ejecutivo Coyle, Karen: "Rights Expression Languages: A Report for the Library of Congress". February 2004  
<http://www.loc.gov/standards/relreport.pdf> [Consultado 18 Febrero 2010]

<sup>107</sup> [http://www.leppm.enap.ca/leppm/docs/Rapports\\_culture/Rapport\\_8\\_culture.pdf](http://www.leppm.enap.ca/leppm/docs/Rapports_culture/Rapport_8_culture.pdf)

<sup>108</sup> High Level Expert Group on Digital Libraries. "Digital Libraries: Recommendations and Challenges for the Future : Final Report". December 2009  
[http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/hleg/reports/hlg\\_final\\_report09.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/hlg_final_report09.pdf)  
[Consultado 18 Febrero 2010]

<sup>109</sup> [http://ec.europa.eu/information\\_society/activities/ict\\_psp/about/index\\_en.htm](http://ec.europa.eu/information_society/activities/ict_psp/about/index_en.htm) [Consultado el 14 de febrero de 2010]

Desde luego una biblioteca virtual debe estar constituida no sólo por un conjunto bien ordenado de objetos digitales que cuenten cada uno de ellos con una declaración y una documentación lo más exhaustiva posible sobre los derechos, tanto si los tiene como si son de dominio público, o si se encuentran en esa situación de indefinición que ha venido a calificarse como obra huérfana. Pero no sólo la biblioteca virtual debe tener perfectamente explícitamente los derechos de todos y cada uno de los objetos digitales, siguiendo el esquema METSRights u otro, sino también con un subsistema de gestión de los mismos que esté operativo ya en la interfaz de acceso público u OPAC o en los distintos módulos y procedimientos de descargas e impresión de la información en todas las modalidades. No debe pasarse por alto el comentario de Andrés Boix Palop titulado *La adaptación del concepto normativo de libro a la pluralidad de soportes en la Ley 10/2007, de 22 junio, de la lectura, del libro y de las bibliotecas*<sup>110</sup>. El comentario del Profesor Andrés Boix Palop es sumamente exhaustivo, pero podría resumirse en el epígrafe de la p. 88 y en el que señala "*Apuntes sobre la deficiente atención en la Ley 10/2007 a la efectiva inclusión del libro electrónico en sus previsiones*"<sup>111</sup>. En ese sentido, conviene recordar que el DOI dispone de un subconjunto de metadatos destinados así mismo a codificar los derechos de autor que debería estar así mismo interconectada y retroalimentar todo el proceso de intercambio de información y objetos digitales.<sup>112</sup>

## 9. Del ISBN al DOI

Nadie duda que el ISBN<sup>113</sup>, no se olvide que es una norma<sup>114</sup> ISO, ha tenido una enorme significación para el comercio internacional de libros e incluso, aunque menor, para identificar unívocamente a los mismos. Menor porque no siempre se modifica el ISBN cuando se trata de una nueva edición que, por otra parte, a veces intencionadamente o no, se confunde con una reimpresión, es decir que no proporciona una identificación perfecta. En cualquier caso, se trata de una norma muy vigente como pone de manifiesto su reciente actualización a 13 caracteres y, sin duda, ha tenido una importancia enorme el hecho de que resultara tan útil para, mediante los algoritmos diseñados por la EAN

<sup>110</sup> Muñoz Machado, Santiago, *Comentarios a la Ley de la lectura, del libro y de las bibliotecas : (Ley 10 2007, de 22 de junio)*. Madrid: Iustel, 2008, pp. 47-94

<sup>111</sup> Como se sabe el reciente Anteproyecto de Ley de Economía Sostenible incluye una disposición adicional, la que ha generado gran polémica y que tampoco prevé la inclusión del libro electrónico en el modelo de economía, sostenible o no, que se ha diseñado. En concreto, la Disposición final primera dedicada a la "*Modificación de la Ley 34/2002, de 11 de julio, de Servicios de la Sociedad de la Información, el Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba Texto Refundido de la Ley de Propiedad Intelectual y la Ley 29/1998, de 13 de julio, reguladora de la jurisdicción contencioso-administrativa, para la protección de la propiedad intelectual en el ámbito de la Sociedad de la Información.*"

<http://www.economiasostenible.gob.es/ley-de-economia-sostenible/> [Consultado el 14 de febrero de 2010]

<sup>112</sup> <http://www.acescritores.com//> [Consultado el 14 de febrero de 2010]

<sup>113</sup> E-Books and ISBNs: a position paper and action points from the International ISBN Agency. [http://www.bisg.org/docs/isbn\\_agency.pdf](http://www.bisg.org/docs/isbn_agency.pdf) [Consultado el 14 de febrero de 2010]

<sup>114</sup> ISO 2108:2005 Information and documentation -- International standard book number (ISBN)

(*European Article Number*), transformar esta numeración en un código de barras. No se debe descartar, sino todo lo contrario, el que en un futuro próximo esa misma información se traslade al interior de un chip con los datos catalográficos que poseen las bases de datos ISBN sobre un número determinado. Si todo ello es utilísimo para editores y libreros y también para bibliotecarios, lo es mucho menos para los lectores, como claramente muestra la ilustración de la figura 2.

Seguramente ocurrirá lo mismo con el DOI (*Digital Object Identifier*), resultará utilísimo para editores y bibliotecarios, aunque probablemente al lector le resulte indiferente, si bien en el caso de que sea posible descargarlo para una cita bibliográfica, -tal vez mediante *COinS*- acabe siendo muy utilizado aunque de forma transparente para los lectores. Como ya se ha dicho antes, las bibliotecas virtuales se incorporan de forma creciente al mundo editorial, aunque no sea su intención primera, en el mismo momento en el que hacen públicos no sólo los registros secundarios, sino también los primarios, es decir, objetos digitales conformados por la descripción bibliográfica, los ficheros y los textos, y otras informaciones relevantes y que puede ser descargadas en un *ePub*. Parece claro, por lo tanto, que los objetos digitales deben estar claramente identificados y es seguro que las importantes distinciones que establecen las FRBR y, por consiguiente, las RDA sobre obras, expresiones, manifestaciones e ítem<sup>115</sup>, ayudarán considerablemente a ello, dada su naturaleza multiforme.

The screenshot shows the homepage of the DOI System. The header includes the DOI logo and the text 'The DOI System' and 'Developed by The International DOI Foundation (IDF)'. There are links for 'Search Guidelines', 'Contact Us', and 'Members On'. The main content area has a 'Welcome to the DOI System' section with text explaining the system's purpose and management. There are also sections for 'In the News' and 'Resources'.

Figura 17. Sitio de la Fundación Internacional DOI

El DOI lo mantiene la *International DOI Foundation* (IDF) que puede suministrar los prefijos DOI, pero que recomienda que se adquieran directamente a través de agencias registradoras<sup>116</sup>, de las que existen ya 6<sup>117</sup>. Según DOI, el objeto registrado puede ser

<sup>115</sup> Es muy interesante observar, aunque se trate de una versión beta, del nuevo catálogo colectivo suizo *Metakatalog der Schweizer Hochschulbibliotheken und der Schweizerischen Nationalbibliothek*. <http://www.swissbib.ch/TouchPoint/start.do> [Consultado el 17 de febrero de 2010]

<sup>116</sup> [http://www.doi.org/registration\\_agencies.html](http://www.doi.org/registration_agencies.html) [Consultado el 17 de febrero de 2010]

<sup>117</sup> [http://www.doi.org/registration\\_agencies.html](http://www.doi.org/registration_agencies.html) [Consultado el 17 de febrero de 2010]

físico o digital<sup>118</sup> y con cualquier nivel de granularidad, es decir que pueden obtenerse DOIs para artículos, capítulos de monografías, fotografías, etc. Desde un punto de vista bibliotecario es capital conocer que en fecha tan reciente como el 1 de diciembre de 2009 se creó la organización *DataCite.org*<sup>119</sup> que aunque no está operativa va a ser la agencia registradora que más se ajuste a las necesidades de las bibliotecas virtuales. Los miembros fundadores de este organismo son la *British Library*, el *Technical Information Center of Denmark* (DTIC), la *TU Delft Library*, el *National Research Council's Canada Institute for Scientific and Technical Information* (NRC-CISTI), la *California Digital Library (University of California Curation Center)*, la *Purdue University (USA)* y la *German National Library of Science and Technology* (TIB). Si se tiene en cuenta que los miembros fundadores son importantes bibliotecas digitales, es muy posible que de su enseñanza se desprendan muchos avances, aunque se trate de colecciones digitales incluidas en el mundo de la investigación científica y técnica. Los objetivos de esta iniciativa son el establecimiento de un fácil acceso a los datos procedentes de la investigación científica en Internet, el incremento de la aceptación de los datos de investigación como contribuciones legítimas y citables del registro científico y el apoyo al archivo de datos que permitan resultados verificables y reutilizables para futuros estudios. *DataCite* promoverá el intercambio de datos, el aumento de acceso y la mejor protección de las inversiones en investigación.

Otra agencia registradora europea de gran interés es *mEDRA (Multilingual European Registration Agency of DOI)* que ha creado un identificador persistente normalizado para cualquier tipo de propiedad intelectual en la red. Tiene su origen en la empresa creada en 1994 denominada *mEDRA Srl (Società a responsabilità limitata)* a instancia de *L'Associazione Italiana Editori*. Interesa destacar que entre los socios de *mEDRA* se encuentra *Editrain*<sup>120</sup>, bien conocida en el mundo editorial español. Si bien *mEDRA* parece una agencia registradora más específica para el mundo europeo, es posible que se encuentre dirigida específicamente para el mundo editorial, mientras que *DataCite* tiene una orientación más clara hacia el mundo de las bibliotecas virtuales. Un aspecto importante es mencionar que DOI utiliza unos metadatos específicos, *MPEG21 Rights Data Dictionary*, para la gestión de la propiedad intelectual y sería muy interesante establecer una transferencia de ellos a los metadatos *METSRights*. Por otra parte, conviene tener presente que *mEDRA* envía sus metadatos siguiendo el modelo de *ONIX*<sup>121</sup>, aunque todavía no desarrollado en España sí está presente en un proyecto tan relevante para el mundo editorial como puede ser *DILVE (Dilve - Distribuidor de información del libro español en venta)*<sup>122</sup>.

---

<sup>118</sup> Ávila Álvarez, Antonio María. "El libro electrónico". En: Boletín económico de ICE, Información Comercial Española. Nº 2978, 2009, pags. 13-22.

[http://www.revistasice.com/cmsrevistasICE/pdfs/BICE\\_2978\\_13-22\\_0265C30C37E5ED1B176B22EC57E8E818.pdf](http://www.revistasice.com/cmsrevistasICE/pdfs/BICE_2978_13-22_0265C30C37E5ED1B176B22EC57E8E818.pdf) [Consultado el 17 de febrero de 2010]

<sup>119</sup> <http://www.datacite.org/> [Consultado el 17 de febrero de 2010]

<sup>120</sup> <http://www.editrain.com/content/seccionweb1.php> [Consultado el 17 de febrero de 2010]

<sup>121</sup> <http://www.editeur.org/8/ONIX/> [Consultado el 17 de febrero de 2010]

<sup>122</sup> [http://www.dilve.es/dilve/dilveweb/index\\_dilve.jsp](http://www.dilve.es/dilve/dilveweb/index_dilve.jsp) [Consultado el 17 de febrero de 2010]

Hay que destacar, por último que DOI es, en el momento en que se escriben estas páginas, un *Draft International Standard 26324 Information and Documentation – Digital Object Identifier System*. Probablemente cuando se publiquen estas páginas será ya una norma ISO, por cierto relacionada con UNICODE, un muy conocido y utilizadísimo subconjunto de la norma ISO/IEC 10646:2003 *Information technology -- Universal Multiple-Octet Coded Character Set (UCS)*.

## 10. Las RDA: un giro copernicano

Cierto es que para acomodarse a esta nueva situación el documento secundario o referencia bibliográfica, que suele ser el primer punto de acceso del lector a la información, ha variado sustancialmente. La publicación, que coincidirá seguramente con las de estas páginas, de las RDA (*Resource Description and Access*), previstas para junio de 2010, facilitan justamente un concepto fundamental, desde el punto de vista informático como es el acceso al documento<sup>123</sup>. Para que estas nuevas reglas de descripción y acceso a la información hayan podido establecerse con rigor ha sido imprescindible un trabajo importante de análisis previo que se plasmó ya en 1998 en un documento de una importancia trascendental en la historia de las bibliotecas, los denominados FRBR y sus normas hermanas de las que vamos a hablar en este apartado<sup>124</sup>. En efecto, este modelo basado en el análisis entidad-relación tiene muy presente lo que desde el punto de vista de la programación informática podríamos llamar programación orientada a objetos, siendo un objeto el conjunto de acciones que pueden experimentar en el tiempo los materiales bibliográficos.

Hasta 1998 el modelo conceptual para la descripción bibliográfica ha estado basado en los denominados *Principios de París* de 1961<sup>125</sup>, en la *International Standard*

---

<sup>123</sup> Haciendo una abstracción verdaderamente descomunal podría decirse que la signatura topográfica de un ejemplar constituía ya un primer indicio de cual era el tipo de acceso al documento bibliográfico y que incluso el tipo de signatura, para quien anduviera familiarizado con ellas, podría dar una indicación de cuáles eran los privilegios y restricciones que se atribuían a ese ejemplar concreto.

<sup>124</sup> IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records. Feb. 2009. Including amendments and corrections to date [http://www.ifla.org/files/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf) [Consultado el 28 de febrero de 2010]

Traducción española: Requisitos funcionales de los registros bibliográficos : informe final / Grupo de estudio de la IFLA sobre los Requisitos funcionales de los registros bibliográficos ; traducción de Xavier Agenjo y María Luisa Martínez-Conde. — [Madrid] : Ministerio de Cultura, 2004. — ISBN 84-8181-213-7. <http://travesia.mcu.es/portalanb/jspui/bitstream/10421/422/1/frbr.pdf> (En Travesía, Ministerio de Cultura de España) <http://www.ifla.org/files/cataloguing/frbr/frbr-es.pdf> (Servidor de IFLA) [Consultado el 28 de febrero de 2010]

<sup>125</sup> International Conference on Cataloguing Principles (Paris : 1961). Report. London : International Federation of Library Associations, 1963. Also available in: Library Resources and Technical Services, 6 162-167 (1962); and Statement of principles adopted at the International Conference on Cataloguing Principles, Paris, October, 1961. (London : IFLA Committee on Cataloguing, 1971) Puede consultarse la traducción española realizada por Elena Escolano en:

*Bibliographic Description* y en los códigos nacionales<sup>126</sup> revisados a partir de esta norma, entre los que las *Anglo-American Cataloguing Rules, Second edition* (AACR2)<sup>127</sup>, publicada en 1978, tuvo una amplia e intensa influencia. Desde 1990<sup>128</sup>, fecha del *Stockholm Seminar on Cataloguing* organizado por el programa de IFLA UBCIM, se hizo patente la necesidad de modificar las prácticas catalográficas para adaptarlas a los cambios producidos tanto en las bibliotecas como en la producción bibliográfica y atender así a la descripción y acceso de los nuevos materiales electrónicos, garantizando al mismo tiempo el control bibliográfico universal de los materiales bibliográficos tradicionales que también habían sufrido un incremento explosivo, y a los avances informáticos.

#### **a) Functional Requirements for Bibliographic Records (FRBR)**

En 1991 se formó el *Study Group on Functional Requirements for Bibliographic Records* que dio lugar en 1998 al modelo conceptual de los *Functional Requirements of Bibliographic Records*<sup>129</sup> que desde entonces se ha convertido en una de las aportaciones más influyentes de la biblioteconomía.

---

[http://www.bne.es/es/Servicios/NormasEstandares/Docs/Paris\\_1961.pdf](http://www.bne.es/es/Servicios/NormasEstandares/Docs/Paris_1961.pdf) [Consultado el 28 de febrero de 2010]

<sup>126</sup> Code Comparisons to Paris Principles. In: "3rd IFLA Meeting of Experts on an International Cataloguing Code (IME ICC)". [www.loc.gov/loc/ifla/imeicc/pdf/codecompare\\_Cairo.pdf](http://www.loc.gov/loc/ifla/imeicc/pdf/codecompare_Cairo.pdf)  
Barbara Tillet. Results of the code comparisons: a summary. In: "3rd IFLA Meeting of Experts on an International Cataloguing Code (IME ICC)". [Consultado el 28 de febrero de 2010]

<sup>127</sup> Joint Steering Committee for Development of RDA. *A Brief History of AACR*. <http://www.rda-jsc.org/history.html> [Consultado el 28 de febrero de 2010]

<sup>128</sup> Barbara B. Tillett. IFLA Study on the Functional Requirements of Bibliographic Records: Theoretical and Practical Foundations. In: 60th IFLA General Conference - Conference Proceedings - August 21-27, 1994  
<http://ifla.queenslibrary.org/IV/ifla60/60-tilb.htm> [Consultado el 28 de febrero de 2010]

<sup>129</sup> IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records : Final Report*. Approved by the Standing Committee of the IFLA Section on Cataloguing September 1997. As amended and corrected through February 2009 September 1997.  
[http://www.ifla.org/files/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf) [Consultado el 28 de febrero de 2010]

The screenshot shows the IFLA website page for the Functional Requirements for Bibliographic Records (FRBR) group. The page has a red navigation bar with links for 'About IFLA', 'Membership', 'Activities and Groups', 'News', 'Publications', 'Annual Conference', and 'Corporate Partners'. The main content area is titled 'Functional Requirements for Bibliographic Records' and includes a search bar, a table of contents, and a list of resources. The table of contents includes 'About this group', 'Events', 'Publications', 'FRBR Bibliography', 'FRBR Discussion List', 'Frequently Asked Questions', 'Related Efforts', 'Review Group Reports', 'Review Group Roster', 'Reviewing Process', 'Working Group on Aggregates', 'Working Group on FRBR/CRM Dialogue', and 'Workshops and Meetings'. The 'Publications' section lists 'Functional Requirements: the FRBR Family of Models' and 'Current text'.

Figura 18. Página oficial en IFLA de los FRBR, a través de la cual se puede acceder al informe, a sus traducciones y a un amplio conjunto de recursos relacionados<sup>130</sup>

En el modelo pre-FRBR, y más concretamente en el modelo ISBD, una descripción bibliográfica se corresponde básicamente con una determinada edición de una obra, compuesta por sus elementos distintivos, reflejados en las áreas de título y mención de responsabilidad, mención de edición, área de edición, producción y distribución y área de notas. Un ejemplo de ediciones serían las de 1998 y de 2009 de los *Functional Requirements for Bibliographic Records: Final Report*. Según la lógica del modelo FRBR, una 'obra' 'se realiza a través de' una 'expresión', que se concreta en una 'manifestación', representada físicamente en cada uno de los ejemplares o 'item'. Para las FRBR el término 'obra' es un contenido intelectual específico y no la concreción de éste en un libro, documento o recurso. Una obra se 'expresa', valga la redundancia, a través de un medio de expresión y utiliza un lenguaje específico. Los conceptos 'obra' y 'expresión' resultan ser los más 'novedosos' a la concepción catalográfica tradicional, no tanto porque no fueran analizados o definidos teóricamente<sup>131</sup> con anterioridad, sino por estar escondidos detrás de un buen número de reglas, normas de puntuación y notas. El hecho de figurar explícitamente como entidades, con una serie de atributos específicos que es preciso registrar, contribuye a comprender y retener el sentido del concepto y, sobre todo, a que sea un conjunto de informaciones que deben proporcionarse siempre, y no sólo en determinados casos<sup>132</sup>.

<sup>130</sup> <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records> [Consultado el 28 de febrero de 2010]

<sup>131</sup> William Denton. FRBR and the History of Cataloging. In: *Arlene G. Taylor. Understanding FRBR: What it is and how it will affect our retrieval tools*. (Westport, Conn: Libraries Unlimited, 2007). <http://pi.library.yorku.ca/dspace/handle/10315/1250>

<sup>132</sup> Como ejemplo de lo dicho, el concepto de entrada autor-título, que puede asimilarse según qué casos a la definición FRBR de 'obra' y 'expresión' se prescribía en las Reglas de Catalogación para la

Seguramente, la implantación de este modelo traerá complicaciones a los catalogadores, acostumbrados a no reflejar más que manifestaciones, pero es de crucial importancia para las bibliotecas virtuales, entre otras cuestiones, para poder ofrecer a los usuarios un registro completo de las múltiples relaciones que pueden establecerse entre los originales y las copias digitales, o entre las distintas versiones electrónicas de un objeto digitalizado o nacido digitalmente. Así, según el modelo FRBR pueden consignarse como distintas manifestaciones las versiones de un original en formato imagen (p.e., JPEG), en formato texto (codificado en los mencionados más arriba TEI o en ALTO), en formato de libro electrónico (p.e., ePub). Los FRBR consideran cada una de estas copias o versiones como nueva manifestación, lo que puede apreciarse por algunos de los atributos para describir manifestaciones como son el soporte físico, los requisitos del sistema, las características del archivo, el modo de acceso y la dirección de acceso. Por supuesto, el original es una manifestación más y entre todas ellas deben establecerse las correspondientes relaciones.

Este ha sido el modelo utilizado por la *Fundación Ignacio Larramendi* para las *Bibliotecas Virtuales FHL*<sup>133</sup> y especialmente para la *Biblioteca Virtual Ignacio Larramendi de Polígrafos*<sup>134</sup>, adelantándose con ello a la próxima publicación de las RDA que, como veremos más adelante, utilizan como modelo base los FRBR.

The screenshot shows the website interface for Fundación Ignacio Larramendi. On the left is a navigation menu with categories like 'LA FUNDACIÓN', 'BIBLIOTECAS VIRTUALES FHL', and 'BIBLIOTECA VIRTUAL IGNACIO LARRAMENDI DE POLÍGRAFOS'. The main content area displays MARC 21 fields for a record. A red box highlights the following fields:

046	1875 2 1926
100	1  \$a Bonilla y San Martín, Adolfo \$d 1875-1926
370	1  \$a Madrid \$c España \$f Comunidad de Madrid
370	1  \$a Madrid \$c España \$f Comunidad de Madrid
372	1  \$a Catedrático de filosofía \$a Historiador
375	1  \$a Hombre
377	1  \$a spa
856	1  \$y Proyecto Filosofía en español \$q HTML \$u http://www.filosofia.org/ave/001/a058.htm

A red callout box on the right points to these fields with the text: "Campos y subcampos RDA".

Figura 19. Campos y subcampos MARC conforme al metamodelo RDA

elaboración de entradas secundarias, pero no para las entradas principales. Un análisis precursor de los FRBR fue el realizado para el desarrollo del sistema de gestión bibliográfico de la Biblioteca Nacional, ARIADNA, en el que las entradas de autor-título se realizaban en todos los casos, dando una gran coherencia al catálogo.

<sup>133</sup> <http://www.larramendi.es/i18n/estaticos/contenido.cmd?pagina=estaticos/BVFHL> [Consultado el 28 de febrero de 2010]

<sup>134</sup> <http://www.larramendi.es/i18n/estaticos/contenido.cmd?pagina=estaticos/bibliotecaIL> [Consultado el 28 de febrero de 2010]

Extraña que IFLA esté manteniendo ambos modelos para los registros bibliográficos. Por un lado difunde los FRBR<sup>135</sup> y publica, en 2009, la *Statement on International Cataloguing Principles (ICP)*<sup>136</sup>, mientras que, por otro, continúa revisando y editando las ISBD como código general para la elaboración de reglas de catalogación y como modelo para la descripción bibliográfica de las agencias bibliográficas nacionales. La última *‘edición consolidada’*<sup>137</sup> se propone como una revisión y reunificación de todas las ISBD publicadas, así como una adaptación a los FRBR. Sin embargo, esta adaptación es meramente un lavado de cara terminológico puesto su objetivo declarado continúa siendo la descripción de manifestaciones.

### **b) Functional Requirements for Authority Data (FRAD)**

En la construcción de bibliotecas virtuales debe tenerse muy en cuenta el uso de herramientas como los ficheros de autoridad por las mismas razones que éstos se utilizan en los catálogos bibliográficos. A menudo se olvida que una biblioteca virtual o un repositorio institucional no es sino una extensión de las funcionalidades de un catálogo bibliográfico y hasta se puede apreciar que hay quienes consideran una ventaja el hecho de no utilizar herramientas propias de la gestión de bibliotecas tradicionales, lo que lleva a ‘reinventar la rueda’ con una cierta asiduidad, normalmente cuando se descubren las ventajas para el tratamiento de la información de las viejas técnicas bibliográficas. En el análisis *Repositorios institucionales universitarios*<sup>138</sup> realizado por José Manuel Barrueco y Cristina García Testal sobre una recopilación de datos de 2008 se señala *“La escasa importancia que parece dársele a proporcionar metadatos suficientes... Un repositorio no es un fin en sí mismo. Su utilidad no debe acabar cuando sus contenidos son encontrados a través de su interfaz web o de un buscador general tipo Google, sino que debe ofrecer las bases para proporcionar servicios de valor añadido a su comunidad de usuarios. Esto sólo se conseguirá proporcionando metadatos completos y de calidad”*.

Y concluyen que *“Sería recomendable que Rebiun tomara la iniciativa en la normalización de aspectos clave como clasificaciones temáticas o vocabularios*

<sup>135</sup> <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records> [Consultado el 28 de febrero de 2010]

<sup>136</sup> [http://www.ifla.org/files/cataloguing/icp/icp\\_2009-en.pdf](http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf). El objetivo de los Principios de Catalogación de 2009 es guiar el desarrollo de códigos de catalogación para que tengan en cuenta las entidades, atributos y relaciones definidas en FRBR, FRAD y FRSAD.

La versión española puede consultarse en:

<http://www.bne.es/es/Servicios/NormasEstandares/Docs/principioscatalogacion2009.pdf> [Consultado el 28 de febrero de 2010]

<sup>137</sup> International Standard Bibliographic Description (ISBD). Preliminary consolidated ed. (2007). [http://www.ifla.org/files/cataloguing/isbd/isbd-cons\\_2007-en.pdf](http://www.ifla.org/files/cataloguing/isbd/isbd-cons_2007-en.pdf)

La versión española puede consultarse en:

<http://www.bne.es/es/Servicios/NormasEstandares/Docs/ISBDconsolidada.pdf> [Consultado el 28 de febrero de 2010]

<sup>138</sup> Barrueco Cruz, José Manuel y Cristina García Testal. “Repositorios institucionales universitarios: evolución y perspectivas”. En: “Interinformación: XI Jornadas Españolas de Documentación : 20, 21 y 22 de mayo de 2009”, Auditorio Palacio de Congresos de Zaragoza, 2009. pags. 99-108

[http://www.fesabid.org/zaragoza2009/Libro\\_Actas\\_Fesabid\\_2009.pdf](http://www.fesabid.org/zaragoza2009/Libro_Actas_Fesabid_2009.pdf) [Consultado el 28 de febrero de 2010]

controlados recomendados a través de la creación de una guía de buenas prácticas de aplicación en todas las universidades”.

The screenshot shows the IFLA website page for Functional Requirements for Authority Data (FRAD). The page layout includes a navigation menu on the left with categories such as 'About IFLA', 'Activities and Groups', 'Publications', and 'Corporate Partners'. The main content area is titled 'Functional Requirements for Authority Data' and includes a search bar, a list of related resources, and a section for translations. The page is designed with a green and white color scheme.

**Figura 20. Página oficial en IFLA de los FRAD, a través de la cual se puede acceder al informe, a sus traducciones y a un amplio conjunto de recursos relacionados**

El *IFLA Working Group on Functional Requirements and Numbering of Authority Records* (FRANAR) se creó en 1999 para extender el modelo FRBR a los registros de autoridad. El resultado de su trabajo, después de diferentes revisiones, se ha publicado en junio de 2009<sup>139</sup>. En él se señala que una de las funciones más importantes de los ficheros de autoridad es, desde el punto de vista del usuario final, el apoyo a las búsquedas bibliográficas en forma de notas y referencias o como soporte para la

<sup>139</sup> Functional Requirements for Authority Data - A Conceptual Model. Edited by Glenn E. Patton. (München: K.G. Saur, 2009) <http://www.ifla.org/publications/functional-requirements-for-authority-data> [Consultado el 28 de febrero de 2010]

Existe una traducción española de una versión previa del informe final que puede resultar útil para un acercamiento a esta norma. En cualquier caso, convendría actualizar la traducción para que sea la de informe final. Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas Grupo de Trabajo sobre Requisitos Funcionales y Numeración de Registros de Autoridad (FRANAR). Requisitos funcionales para datos de autoridad: un modelo conceptual. Arroyo Fernández, Domingo, trad. ; Rodríguez Gordo, Pablo, trad. Madrid, Ministerio de Cultura, 2007. <http://hdl.handle.net/10421/421> [Consultado el 28 de febrero de 2010]

Se acaba de publicar la versión española definitiva elaborada por el equipo de trabajo que dirige Elena Escolano en la Biblioteca Nacional [http://www.bne.es/opencms/es/Servicios/NormasEstandares/Docs/FRAD\\_espaxol.pdf](http://www.bne.es/opencms/es/Servicios/NormasEstandares/Docs/FRAD_espaxol.pdf) [Consultado el 5 de marzo de 2010]

creación de sistemas de búsqueda multilingües. Otras funciones de singular interés son el uso cooperativo de los ficheros de autoridad en el entorno de la Web, tanto a nivel nacional como internacional, con ejemplos como el *Virtual International Authority File*<sup>140</sup> (VIAF) que implica la fusión de ficheros de autoridad procedentes de diferentes fuentes e idiomas, o la apertura de estos ficheros para su reutilización por diferentes aplicaciones según la filosofía de *Linked Open Data* que veremos más adelante.

**Figura 21. Resultados de la búsqueda sobre Averroës en la página de VIAF [con un ligero retoque gráfico]**

Los FRAD detallan las entidades y atributos necesarios para el control de autoridades, pero introducen como novedad algunos atributos específicamente destinados a mejorar la contextualización de los registros de autoridad. Así, los atributos como género, lugar de nacimiento y muerte, país, campo de actividad, profesión y ocupación tienen como objetivo aumentar la información que se proporciona y permitir la construcción de navegaciones a otras entradas relacionadas. De este modo, los ficheros de autoridad pasan a ser no sólo un apoyo a la búsqueda bibliográfica sino recursos informativos en sí mismos para usuarios finales<sup>141</sup>. Evidentemente, no sólo los usuarios finales se

<sup>140</sup> <http://www.oclc.org/research/activities/viaf/default.htm>. Se puede acceder a la consulta de VIAF en <http://viaf.org/> [Consultado el 28 de febrero de 2010]

Es muy interesante y atractiva la nueva interfase a la que se ha actualizado recientemente el VIAF. Conviene recordar que VIAF utiliza el protocolo SRU (<http://code.google.com/p/oclsrw/>) para que distintos clientes puedan descargarse los registros de autoridades multilingües.

<sup>141</sup> Esta ha sido la filosofía seguida por la *Biblioteca Virtual Ignacio Larramendi de Polígrafos* que ha introducido estos atributos, recogidos en la *Update 10* de octubre de 2009 del *MARC 21 Format for Authority Records*, en su fichero de autoridades, que como se ha indicado recoge ya determinadas estipulaciones de las RDA.

beneficiarán de este nuevo modelo, sino también otras aplicaciones relacionadas con la Web Semántica (i.e., ontologías) que necesitan con urgencia descripciones primordiales de calidad sobre personas, instituciones, familias, lugares, etc., momento en el cual se mira hacia las bibliotecas, sorprendidos, de que esa información exista y en tal cantidad. Además, a partir del atributo de idioma de una entidad (no confundir con el idioma de catalogación), su escritura y transliteración, es posible desarrollar ficheros de autoridad multilingües y multiescritura y, por tanto, el uso de puntos de acceso multilingües en bases de datos bibliográficas o en bibliotecas virtuales.

Figura 22. Página oficial en IFLA de los FRSAD

### c) Functional Requirements for Subject Authority Data (FRSAD)

Para completar el análisis de los registros de autoridad con todo lo relacionado con las materias se creó en 2005 el *Working Group on Functional Requirements for Subject Authority Records* (FRSAR) que elaboró los *Functional Requirements for Subject Authority Data* (FRSAD): *a conceptual model -- Draft Report*, en fase de revisión<sup>142</sup>. En el modelo FRSAD una *obra* tiene como materia un *tema* cuya denominación es un *nomen*. Este es uno de los cambios que introduce más complejidad en el sistema, pero que facilitará el tratamiento de la información a diferentes niveles o con diferentes fines, como es la separación de un concepto del nombre utilizado para designar ese concepto.

El impacto<sup>143</sup> de FRBR, FRAD y FRSAD como modelo del universo bibliográfico ha sido enorme, lo cual puede apreciarse en que es la espina dorsal de las nuevas reglas de

<sup>142</sup> <http://www.ifla.org/node/1297> [Consultado el 28 de febrero de 2010]

<sup>143</sup> FRBR Review Group, IFLA Cataloguing Section. FRBR bibliography (2006).

[http://archive.ifla.org/VII/s13/wgfrbr/FRBR\\_bibliography.pdf](http://archive.ifla.org/VII/s13/wgfrbr/FRBR_bibliography.pdf)

Puede completarse con The FRBR Blog <http://www.frbr.org/> [Consultado el 28 de febrero de 2010]

catalogación *Resource Description and Access* y en la consiguiente adaptación del formato MARC 21. Es evidente que esto significa un cambio radical en el modelo descriptivo no sólo de los catálogos bibliográficos, sino también de las bibliotecas virtuales, lo que hace ineludible dedicarle espacio suficiente en este artículo.

#### **d) Resource Description and Access (RDA)**

Las AACR2 han sido, sin duda, el código de catalogación más influyente que ha existido porque, combinado con el formato MARC y las ISBD, han tenido un alcance mundial. En 1997 se convocó en Toronto, Canadá, una reunión para iniciar la revisión de las AACR2 y para responder al nuevo panorama de la todavía muy reciente Web. Poco a poco fueron desarrollándose los diferentes borradores de las entonces denominadas AACR3 y fue quedando claro que el volumen de los cambios era de tal calado que no podía constreñirse al marco de una nueva edición actualizada de las AACR. En 2005 *el Joint Steering Committee for the Revision of AACR* (JSC) anunció un nuevo rumbo para la revisión de las reglas y se abandonó la idea de su revisión optando por la elaboración de una nueva normativa<sup>144</sup> que se adecuara convenientemente a un entorno digital, que proporcionara directrices tanto para la descripción como para el acceso a los recursos, que considerara todo tipo de recursos, analógicos y digitales (y especialmente éstos), y que permitiera la creación de registros que pudieran utilizarse en todo tipo de entornos (Internet, en la Web, en OPACs, etc.).

En resumen, las nuevas reglas de catalogación pasaron a denominarse RDA (*Resource Description and Access*) y adoptaron como modelo subyacente el de los FRBR y FRAD lo que supuso la reconceptualización tanto de las descripciones bibliográficas como de los registros de autoridad. Tras el consiguiente periodo de elaboración de la norma, de los vocabularios complementarios y del material anejo y de su revisión, el borrador final<sup>145</sup> se difundió en 2008, anunciándose la publicación definitiva para 2009. Este plazo ha sido retrasado y está prevista la publicación de las RDA<sup>146</sup> definitivas para junio de 2010.

---

<sup>144</sup> <http://www.rda-jsc.org/rda.html#background> [Consultado el 28 de febrero de 2010]

<sup>145</sup> <http://www.rda-jsc.org/rdafulldraft.html> [Consultado el 28 de febrero de 2010]

<sup>146</sup> <http://www.rdaonline.org/> [Consultado el 28 de febrero de 2010]

	<p>Joint Steering Committee for Development of RDA</p>
Overview	<p>The Joint Steering Committee for Development of RDA (JSC) is the new name of the Joint Steering Committee for Revision of AACR. The JSC was responsible for maintaining the Anglo-American Cataloguing Rules, and is now working on a new code, "RDA: Resource Description and Access" scheduled to be released in June 2010. In order to enable the JSC and constituencies to concentrate on this work, the last update to AACR2 was in 2005.</p>
News and Announcements	
RDA	<p>The <i>Anglo-American Cataloguing Rules (AACR)</i> are "designed for use in the construction of catalogues and other lists in general libraries of all sizes. ... The rules cover the description of, and the provision of access points for, all library materials commonly collected the present time."<sup>147</sup> [Rule 0.1]</p>
Submitting Proposals to Revise AACR	<p>The current text is the <b>Second Edition, 2002 Revision</b> (with 2003, 2004, and 2005 updates) which incorporates all changes approved the JSC through February 2005. The rules are published by:</p> <ul style="list-style-type: none"> <li>▪ The American Library Association</li> <li>▪ The Canadian Library Association</li> <li>▪ CILIP: Chartered Institute of Library and Information Professionals</li> </ul>
Strategic Plan	<p>For information on purchasing AACR, see the <a href="#">AACR2 Web site</a>.</p>
JSC Members	<p>AACR is published in English and has been translated into other languages; see "<a href="#">Translations of AACR</a>."</p>
Annual Reports	<p>For information on the history of AACR see "<a href="#">A Brief History of AACR</a>."</p>
Working Documents	
Historic	

**Figura 23. Página del Comité Conjunto Permanente para el Desarrollo de las RDA<sup>147</sup>**

Lógicamente, las RDA han desatado el proceso de adaptación del formato MARC 21 a la nueva realidad catalográfica. Los procedimientos ya conocidos de actualización<sup>148</sup> de este formato a través de MARBI<sup>149</sup>, propuestas, aprobaciones y aplicación, ya han dado lugar a la *Update 10*<sup>150</sup>, publicada en octubre de 2009 y hay que analizar los muy importantes cambios<sup>151</sup> de la *Update 11* difundida a principios de marzo de 2010<sup>152</sup>. Las modificaciones de la *Update 10* suponen en conjunto la creación de nuevos campos en los formatos de autoridades y bibliográfico, lo cual no modifica su estructura. Sin embargo, la *Update 11* supone un considerable cambio pues contiene la categorización de tipos de registros para identificarlos como obras, expresiones o manifestaciones o el establecimiento de las relaciones previstas por las RDA entre registros de distinto tipo y la codificación de las relaciones (p.e., relaciones entre obras, entre expresiones) mismas supondrá un enorme cambio en las prácticas catalográficas y su comprensión. Otro gran cambio anunciado es la posibilidad de vincular registros internamente en una aplicación a través de los números de identificación registrados en el subcampo \$0, pero también a través de URIs a fuentes de información externas (p.e., a codificaciones en RDF). Ya se comentará más adelante, pero esto es de enorme significación para el desarrollo de aplicaciones semánticas.

<sup>147</sup> <http://www.rda-jsc.org/index.html> [Consultado el 28 de febrero de 2010]

<sup>148</sup> <http://www.loc.gov/marc/development.html> [Consultado el 28 de febrero de 2010]

<sup>149</sup> <http://www.loc.gov/marc/marbi/marbi.html> [Consultado el 28 de febrero de 2010]

<sup>150</sup> <http://www.loc.gov/marc/bibliographic/ecbdhome.html> [Consultado el 28 de febrero de 2010]

<sup>151</sup> Véase el documento RDA in MARC de enero de 2010 firmado por la propia Sally McCallum <http://www.loc.gov/marc/RDAinMARC29.html> [Consultado el 26 de febrero de 2010]

<sup>152</sup> Véanse los cambios para el formato para registros bibliográficos en <http://www.loc.gov/marc/bibliographic/bdapndxg.html>; para registros de autoridad en <http://www.loc.gov/marc/authority/adapndxf.html>; y para registros de fondos en <http://www.loc.gov/marc/holdings/hdapndxf.html>

**R | D | A**  
RESOURCE DESCRIPTION & ACCESS

**RDA Toolkit**   **FAQ**   **RDA Background**   **Press Releases**   **JSC**   **AACR2**

**RDA Toolkit WEBINAR**

**RDA Toolkit**  
Coming June 2010

The resource to help you navigate from AACR2 to RDA—the new, unified standard for resource description and access, designed for the digital world and an expanding universe of metadata users. RDA Toolkit highlights:

- RDA instructions that are searchable and browsable
- AACR2 Rule Number Search of RDA instructions
- Workflows, mappings: tools to customize the RDA instruction set to support organizational training and processes.
- Two views of RDA content—by table of contents and by RDA element set
- What you need to evaluate and implement RDA; to make cataloging decisions based on principles; to increase efficiency; to facilitate collaboration; and to help position the community for the future by making bibliographic data accessible on the Web.
- Full text of AACR2 with links to RDA.

The content of RDA has been developed in a collaborative process led by the Joint Steering Committee (JSC). The project is overseen by the Committee of Principals representing American Library Association, Canadian Library Association, CILIP, Library of Congress, Library and Archives Canada, British Library, and National Library of Australia. The RDA Toolkit is published by the American Library Association, the Canadian Library Association, and CILIP (through its publishing imprint Facet Publishing).

**RDA Toolkit Pricing**

**Licensing fees for the RDA Toolkit (US Pricing)**

**Base price:**

Annual licensing fee for one user at a time, unlimited number of registered users: \$325

**Figura 24. Página oficial de las RDA con las novedades hasta su publicación. Aquí se muestra el RDA Toolkit para el aprendizaje de la nueva normativa<sup>153</sup>**

Ni que decir tiene que las RDA ya han causado una enorme polémica con división de opiniones entre los que piensan que el modelo tradicional de las AACR2 estaba suficientemente contrastado y sólo necesitaba de una evolución y aquellos que consideran que las RDA no han terminado de dar el paso hacia un nuevo modelo descriptivo como requiere la actual Web, incluida la Web Semántica, y que el lastre de tener que ser compatible con los miles de millones de registros producidos según las AACR2 únicamente contribuirá a retrasar una alteración radical del modelo. Como ejemplo, para esta última corriente de opinión, representada por Karen Coyle, continuar con la transcripción de la portada es no sólo innecesario, sino que choca con el modo de expresar datos de la web semántica. Por desgracia, y para mantener la continuidad con el modelo catalográfico anterior, las RDA no sólo deben ser compatibles con AACR2, sino también con la puntuación ISBD, lo cual no es buena noticia para todos aquellos que se ven en la obligación de especificar la puntuación de los campos y subcampos. Esto nos lleva a una segunda cuestión que es cómo va a desarrollarse la adaptación del software, comercial o libre, de gestión bibliotecario o de generación de repositorios a la nueva normativa, especialmente desde los puntos de vista de las tareas del catalogador y de los usuarios.

Es muy interesante estudiar el sitio Web del nuevo catálogo colectivo suizo *Metakatalog der Schweizer Hochschulbibliotheken und der Schweizerischen Nationalbibliothek*<sup>154</sup>, en el que se muestra de forma exhaustiva la visualización de los

<sup>153</sup> <http://rdaonline.org/index.html> [Consultado el 28 de febrero de 2010]

<sup>154</sup> Véase nota 112.

registros bibliográficos y de los fondos conforme a la normativa que venimos mencionando y que contribuirá a ayudar a los bibliotecarios de todo el mundo a familiarizarse con ella, aunque conviene señalar que al no estar publicadas oficialmente las RDA no puede asegurarse que la adaptación suiza se corresponda exactamente con la normativa.

Será muy interesante conocer las contribuciones y conclusiones de la sesión *Libraries and the Semantic Web*, organizada conjuntamente nada menos que por tres secciones de IFLA (*Information Technology Section, Cataloguing Section y Classification and Indexing Section*) que, entre otros aspectos, versará sobre la relación entre normas como MARC, FRBR, FRAD, etc y la Web Semántica. Todo ello demuestra hasta qué punto se trata de un proyecto transversal.

## 11.El intercambio de metadatos entre las bibliotecas virtuales

El término 'metadatos' viene utilizándose en el mundo bibliotecario desde mediados de los años 90, coincidiendo con la explosión de los recursos electrónicos en la web, y de la misma web. Durante un tiempo, este novedoso término para el mundo bibliotecario venía a ser sinónimo de descripción de recursos electrónicos, diferenciándose del proceso de catalogación de materiales bibliográficos tradicionales, lo que ha provocado una cierta desorientación. Así, los metadatos describían recursos electrónicos, mientras que los registros bibliográficos eran el resultado de la catalogación de materiales bibliográficos tangibles. La confusión llegó hasta tal punto que algunos profesionales consideraban que el formato MARC resultaba conveniente para la descripción de los materiales habituales de las bibliotecas, pero que en el caso de bibliotecas digitales era necesario disponer de metadatos en Dublin Core. No ha costado poco esfuerzo, ni tinta, deshacer esta confusión<sup>155</sup>.

Anne J. Gilliland establece una tipología, que resulta bastante clarificadora, sobre los diferentes tipos de normas relacionadas con los recursos de información y que se engloban, correcta o incorrectamente, bajo el paraguas del término 'metadatos'. En este apartado queremos hacer un resumen de los esquemas de metadatos utilizados para representar un objeto digital con fines de proceso, intercambio e información que ordenaremos según la funcionalidad que sustentan. Hemos elegido esta presentación para intentar deshacer la impresión que tienen muchos bibliotecarios de que es imposible estar al tanto de las normas que deben aplicarse, insistiendo con ello en las funcionalidades que deben proporcionarse a los usuarios de las bibliotecas virtuales y a su correcta gestión.

### a) Descripción de recursos de información

En el apartado anterior hemos hecho un breve resumen de los cambios que van a afectar al formato MARC 21 en su adaptación a las nuevas reglas de catalogación RDA. A pesar de las críticas a este formato no puede olvidarse que los catálogos bibliográficos de todo el mundo ofrecen miles de millones de registros codificados conforme a esta

---

<sup>155</sup> Sally MacCallum. Extending MARC for Bibliographic Control in the Web Environment: Challenges and Alternatives (2002).

[http://www.loc.gov/catdir/bibcontrol/mccallum\\_paper.html](http://www.loc.gov/catdir/bibcontrol/mccallum_paper.html) [Consultado el 28 de febrero de 2010]

norma que continúa siendo el formato de intercambio de información bibliográfica por excelencia entre los sistemas de gestión bibliográfica y bibliotecaria. Y, por supuesto, tiene también la ventaja de que existen multitud de conversiones que permiten que a partir de un registro bibliográfico MARC puedan obtenerse otros registros conformes a otros esquemas de metadatos, desde Dublin Core a las estructuras para citas bibliográficas. La coexistencia de diferentes formatos MARC (MARC 21, UNIMARC, etc.) y otras variantes está en el origen de la norma *MarcXchange (ISO 25577:2008 Information and documentation – MarcXchange)*<sup>156</sup>, cuyo objetivo es 'la generalización de un esquema XML para representar cualquiera de los formatos basados en ISO 2709'. Ni que decir tiene que la agencia designada para el mantenimiento de *MarcXchange* es la *Network Development and MARC Standards Office de la Library of Congress*<sup>157</sup>. La Library of Congress dispone en su sitio web de una serie de hojas de estilo para realizar transformaciones entre los diferentes esquemas de metadatos descriptivos (MARC XML, MODS, DC, ONIX), aunque fundamentalmente tienen como origen o destino el esquema MARC XML.

The screenshot shows the ISO website interface for the product ISO 25577:2008. The page is titled 'ISO 25577:2008' and 'Information and documentation – MarcXchange'. It features a navigation menu with options like 'Home', 'Products', 'Standards development', 'News and media', and 'About ISO'. A sidebar on the left contains links to 'ISO Store', 'ISO Standards', and 'By ICS'. The main content area includes a table for 'Media and price' with columns for Language, Format, and Add to basket. The table lists options for English and French in PDF and Paper formats, all priced at CHF 86,00. Below this, there is a 'General information' section with details such as 'Number of Pages: 15', 'Edition: 1 (Monolingual)', 'Status: Published', and 'Stage: 60.60 (2008-11-25)'. A right-hand sidebar titled 'These standards could also interest you' lists related standards like ISO 15511:2009, ISO 15836:2009, and ISO 8459:2009.

Figura 25. Página de la norma MarcXchange en el sitio web de ISO

<sup>156</sup> [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=43005](http://www.iso.org/iso/catalogue_detail.htm?csnumber=43005) [Consultado el 28 de febrero de 2010]

Esta norma relaciona las funciones de *MarcXchange* que puede utilizarse para representar registros MARC en XML, para la descripción bibliográfica en XML, como extensión del esquema METS (Metadata Encoding and Transmission Standard); para el intercambio de registros MARC en XML, para la transmisión de registros MARC en servicios web como SRU (*search/retrieval via URL*), como formato intermedio para la manipulación, conversión, edición y validación de datos bibliográficos y como metadatos XML embebidos en un recurso de información.

<sup>157</sup> <http://www.loc.gov/standards/iso25577> [Consultado el 28 de febrero de 2010]

### a.1 Dublin Core

Desde sus orígenes en 1995<sup>158</sup>, *Dublin Core Metadata Element Set (DCMES)* ha tenido una gran éxito de implantación. Estas especificaciones técnicas para la descripción de todo tipo de recursos han sido revisadas en diferentes ocasiones y constituyen la norma ISO 15836:2009 (y ANSI Z39.85-2007). Aunque originada en un entorno bibliotecario fuerte, con la participación de destacados miembros de la OCLC, estaba pensada para su aplicación, fundamentalmente, fuera del entorno bibliotecario, siendo lo suficientemente sencilla (15 elementos) y completa para garantizar su aplicación. La versión inicial, 1.0 de 1998 ha tenido numerosas modificaciones, hasta llegar a la actual de 2009. La norma es extremadamente básica y no incluye apenas reglas de codificación para los valores de determinados elementos, ni guía, salvo un breve comentario, para la redacción de los elementos. La rápida aceptación del conjunto de elementos Dublin Core se puso de manifiesto en su incorporación en 1998 como puntos de acceso al conjunto de atributos BIB-1 de Z39.50 con los valores 1097 a 1111.

A los 15 elementos básicos de DCMES, conocido también como *Dublin Core básico o sin cualificar*, se le añadieron en el año 2000 una serie de cualificadores y de refinamientos de codificación que ampliaban las posibilidades de descripción de los recursos y hacían la norma mucho más completa. En resumen, la especificación de todos los términos que mantiene *Dublin Core Metadata Initiative* conforman *DCMI Metadata Terms* del que DCMES (o Dublin Core sin cualificar) es un subconjunto. *DCMI Metadata Terms* es desde agosto de 2009 parte fundamental de las especificaciones de Europeana a través de *Europeana Semantic Elements (ESE) specifications v3.2.2*<sup>159</sup>. Está previsto, sin embargo completar estas especificaciones por otras de carácter más semántico: *Definition of the Europeana Data Model elements*<sup>160</sup>, muy en la línea de lo que se expresa en el apartado 12 de este artículo. En este sentido, DC ofrece la gran ventaja de que al no estar orientado a un dominio específico puede ser la estructura de datos común a la que poder convertir otras estructuras de datos procedentes de archivos, bibliotecas y museos, lo que lo hace especialmente interesante para sistemas de información creados a partir de la integración de metadatos recolectados de fuentes heterogéneas<sup>161</sup>.

<sup>158</sup> History of the Dublin Core Metadata Initiative. <http://dublincore.org/about/history/> [Consultado el 28 de febrero de 2010]

<sup>159</sup> [http://version1.europeana.eu/c/document\\_library/get\\_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602](http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602) [Consultado el 28 de febrero de 2010]

<sup>160</sup> [http://version1.europeana.eu/c/document\\_library/get\\_file?uuid=9783319c-9049-436c-bdf9-25f72e85e34c&groupId=10602](http://version1.europeana.eu/c/document_library/get_file?uuid=9783319c-9049-436c-bdf9-25f72e85e34c&groupId=10602) [Consultado el 28 de febrero de 2010]

<sup>161</sup> COVAX (Contemporary Virtual Archives in XML), financiado por el programa IST de la Unión Europea en 1999-2000, fue un proyecto pionero en el que participaron los autores del presente artículo. Para una descripción del mismo véase Hernández Carrascal, Francisca, y Carlos Wert. XML, ¿una infraestructura para la biblioteca digital?: el proyecto COVAX. En: "La biblioteca pública: portal de la sociedad de la información (presentada en el Congreso Nacional de Bibliotecas Públicas, Valencia, 29, 30 y 31 de octubre de 2002". Madrid, Ministerio de Cultura, Subdirección General de Información y Publicaciones, 2002. <http://travesia.mcu.es/portaInb/jspui/handle/10421/1163> [Consultado el 28 de febrero de 2010]

El sitio oficial de DCMI ofrece los esquemas<sup>162</sup> XML recomendados para la codificación tanto de DCMIES (DC sin cualificar) como de *DCMI Metadata Terms* (DC cualificado), así como el *DCMI RDF schema*<sup>163</sup>, y en HTML/XHTML<sup>164</sup>. Además, DC puede expresarse en el *XML Schema for Dublin Core without qualification*<sup>165</sup>, siguiendo la versión 2.0 de *The Open Archives Initiative Protocol for Metadata Harvesting*<sup>166</sup> (OAI-PMH) y en *Dublin Core Extended*<sup>167</sup>, siguiendo la norma SRU (*Search/Retrieval via URL*).

La prueba de la extensión que ha adquirido DC está en ser la estructura de datos descriptivos utilizada por la mayoría del software para repositorios digitales<sup>168</sup> y en ser el núcleo central de *Open Archives Initiative Protocol for Metadata Harvesting*. Una idea de las cifras que arrojan la combinación DC + OAI está en los más de 23 millones de objetos digitales procedentes de 1.139 repositorios que ofrece OAIster a 1 de marzo de 2010.

## a.2 MODS

En torno también al formato MARC la *Library of Congress' Network Development and MARC Standards Office* desarrolló el esquema MODS (*Metadata Object Description Schema*)<sup>169</sup> para atender la necesidad de una versión de MARC 21 en XML simplificada que fuera más amplia que los 15 elementos básicos de Dublin Core y que se adaptara a las necesidades de la comunidad bibliotecaria. Los elementos del esquema, aunque se derivan del formato MARC y conforman un subconjunto del mismo, tienen una presentación más legible (<titleinfo> en lugar de <datafield tag="245" ind1="1" ind2="0">). Su principal ventaja está en la capacidad para establecer relaciones jerárquicas entre las descripciones y proporcionar una codificación detallada de las partes de un recurso (volumen, número, capítulo, sección, párrafo, pista, etc.), pudiendo llegar hasta la codificación de secciones o párrafos dentro de un determinado texto. Esta expresión de jerarquías, que no tiene equivalente en el formato MARC 21 ha

<sup>162</sup> <http://dublincore.org/schemas/xmls/> [Consultado el 28 de febrero de 2010]

<sup>163</sup> <http://dublincore.org/schemas/rdfs/> [Consultado el 28 de febrero de 2010]

<sup>164</sup> Expressing Dublin Core metadata using HTML/XHTML meta and link elements. <http://dublincore.org/documents/dc-html/> [Consultado el 28 de febrero de 2010]

<sup>165</sup> [http://www.openarchives.org/OAI/2.0/oai\\_dc.xsd](http://www.openarchives.org/OAI/2.0/oai_dc.xsd) [Consultado el 28 de febrero de 2010]

<sup>166</sup> <http://www.openarchives.org/OAI/openarchivesprotocol.html> [Consultado el 28 de febrero de 2010]

<sup>167</sup> <http://www.loc.gov/standards/sru/resources/dcx/dcx.xsd> [Consultado el 28 de febrero de 2010]

<sup>168</sup> Repositories Support Project. Repository Software Survey, March 2009. <http://www.rsp.ac.uk/software/surveyresults> [Consultado el 28 de febrero de 2010]

<sup>169</sup> Guenther, Rebecca and Sally MacCallum. New Metadata Standards for Digital Resources: MODS and METS. "Bulletin of the American Society of Information Science and Technology". v. 29, n.2 (December/January 2003) <http://www3.interscience.wiley.com/journal/109863404/abstract?CRETRY=1&SRETRY=0> [Consultado el 28 de febrero de 2010]

provocado, al menos de momento, que sea un esquema de metadatos descriptivos ampliamente utilizado en bibliotecas y colecciones digitales. Ejemplo de ello es que es una de las extensiones de metadatos de METS más utilizada, sobre todo en proyectos que incluyen el reconocimiento óptico de caracteres y que necesitan controlar las partes de un texto. Igualmente, esta característica hace que MODS pueda ser utilizado como esquema común de transformación de otros esquemas de metadatos marcadamente jerárquicos como EAD en el entorno de proyectos que reúnen colecciones documentales archivísticas, bibliográficas y museológicas.

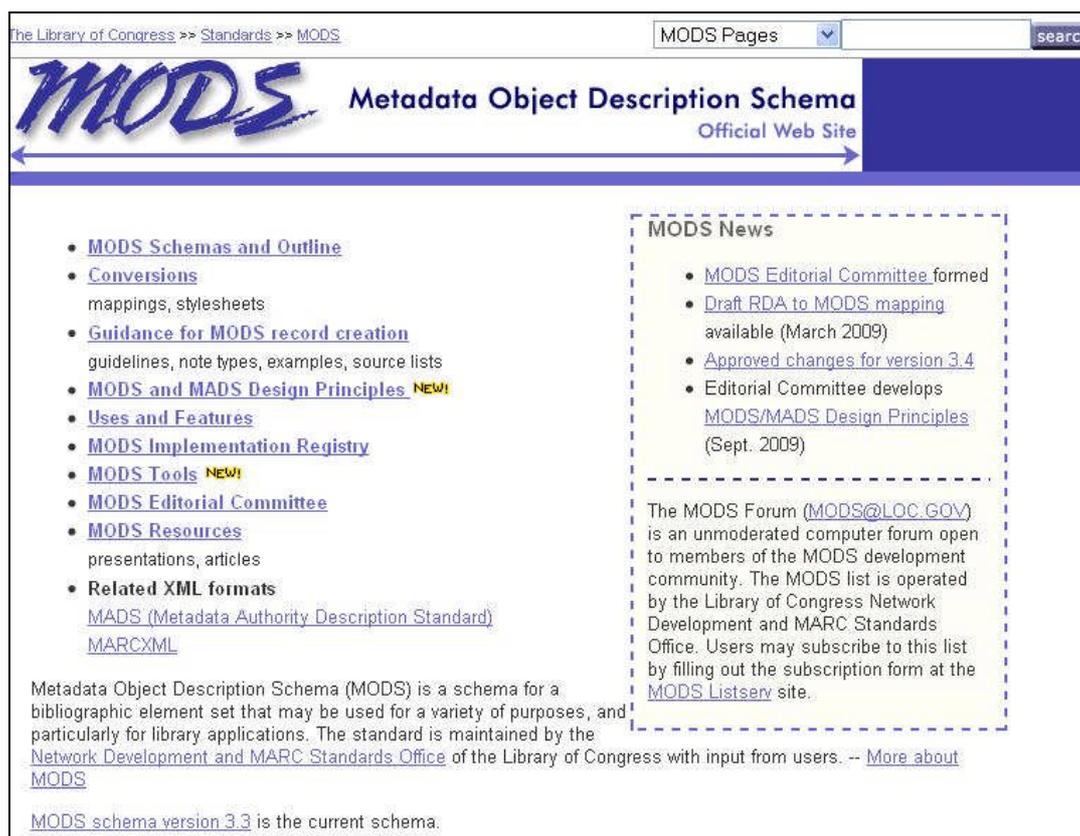


Figura 26. Página oficial de MODS

### a.3 ONIX

No debe olvidarse que el hecho de que las bibliotecas virtuales y digitales estén construidas en su mayor parte con recursos que forman parte del dominio público va a ser alterado con la irrupción masiva de publicaciones nacidas digitalmente, tanto comerciales como gratuitas, y en cualquiera de los formatos de lectura electrónica. Por ello conviene prestar atención a otro esquema de metadatos que es de gran utilidad para la transmisión de datos descriptivos y comerciales como ONIX *ONline Information Exchange*. ONIX es una familia de normas, para libros<sup>170</sup>, publicaciones periódicas<sup>171</sup> y licencias<sup>172</sup> creadas por EDItEUR<sup>173</sup> para la generación de mensajes y metadatos

<sup>170</sup> <http://www.editeur.org/93/Release-3.0-Downloads/> [Consultado el 28 de febrero de 2010]

<sup>171</sup> <http://www.editeur.org/84/Overview/> [Consultado el 28 de febrero de 2010]

<sup>172</sup> <http://www.editeur.org/85/Overview/> [Consultado el 28 de febrero de 2010]

necesarios en los procesos de comunicación entre ordenadores relacionados con la industria editorial, con la creación, distribución, licencias y derechos de propiedad intelectual de publicaciones tanto físicas como digitales. Un elemento central de esta comunicación es la descripción del producto y como tal descripción puede expresar las particularidades de las transacciones del mundo editorial, pero también puede ser reutilizado su contenido para generar descripciones bibliográficas normalizadas en otros formatos. Esta transformación y reutilización de datos descriptivos (p.e., entre ONIX y MARC<sup>174</sup>) resulta especialmente útil en los procesos de gestión de adquisiciones, y en la precatalogación y catalogación de publicaciones adquiridas. Igualmente, una biblioteca digital puede ser a su vez editor o distribuidor de sus propias publicaciones por lo que la transformación entre esquemas de metadatos descriptivos es de especial interés (véase el apartado 7 de este artículo).

Para finalizar este apartado es necesario mencionar la incorporación de los esquemas de metadatos descriptivos a las especificaciones de otras normas y protocolos, especialmente los destinados a la búsqueda y recuperación de información y a la recolección de metadatos. Así, SRU (*Search/Retrieval via URL*) incluye entre los esquemas de metadatos en los que puede transferirse<sup>175</sup> un registro bibliográfico a MODS, ONIX DTD v2.0, MARC XML, EAD, UNIMARC XML Schema<sup>176</sup> y Dublin Core Extended.

Tanto si se utilizan por sí mismos, como esquemas de metadatos descriptivos, o formando parte de otros esquemas y protocolos, es conveniente para la gestión de bibliotecas virtuales que sea posible realizar conversiones entre los diferentes esquemas de metadatos mencionados. Esto permitirá que a partir de una única descripción, por ejemplo en MARC 21, se generen otros registros en los esquemas requeridos por otras aplicaciones y protocolos. Así, la OCLC ha lanzado el *Crosswalk Web Service* (CWS)<sup>177</sup> para traducir registros de un esquema a otro. Otra iniciativa de ayuda para el desarrollo de transformaciones entre estructuras de metadatos es la *Metadata Standards Crosswalk*<sup>178</sup> de *Getty Research Institute* (*Getty's Crosswalk*), aunque solo considera los elementos que se solapan entre archivos, bibliotecas y museos.

---

<sup>173</sup> [http://www.editeur.org/2/About/#Editeur's\\_role](http://www.editeur.org/2/About/#Editeur's_role) [Consultado el 28 de febrero de 2010]

<sup>174</sup> <http://lcweb.loc.gov/marc/onix2marc.html> [Consultado el 28 de febrero de 2010]

<sup>175</sup> <http://www.loc.gov/standards/sru/resources/schemas.html> [Consultado el 1 de marzo de 2010]

<sup>176</sup> <http://www.bncf.firenze.sbn.it/progetti/unimarc/slim/documentation/unimarcslim.html> [Consultado el 1 de marzo de 2010]

<sup>177</sup> <http://www.oclc.org/research/activities/xwalk/default.htm> [Consultado el 1 de marzo de 2010]

<sup>178</sup> [http://www.getty.edu/research/conducting\\_research/standards/intrometadata/crosswalks.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html) [Consultado el 1 de marzo de 2010]

Font Size: A A A

HOME ABOUT STANDARDS ▾ COLLABORATIONS ▾ MEMBERSHIP ▾ EVENTS CONTACT

**EDItEUR**

**ONIX**

You Are Here: Home / Standards / ONIX

**ONIX**

The ONIX family includes standards for **Books**, **Serials** and **Licensing Terms** (including **RROs**).

All ONIX standards are designed to support computer-to-computer communication between parties involved in creating, distributing, licensing or otherwise making available intellectual property in published form, whether physical or digital. All are expressed in XML.

**ONIX for Books** was the first, and is the most widely-adopted, member of EDItEUR's ONIX family of standards. It was initially developed by EDItEUR jointly with **Book Industry Communication** (UK) and the **Book Industry Study Group** (US) and is now maintained under the guidance of an International Steering Committee including not only BIC and BISG but also national user groups in Australia, Belgium, Canada, Finland, France, Germany, Italy, The Republic of Korea, The Netherlands, Norway, Russia, Spain, and

STANDARDS

ONIX

ONIX for Books ▾

ONIX for Serials ▾

Licensing Terms ▾

FTP Filenaming ▾

ONIX Identifier ▾

Registration Formats ▾

E-Commerce

EDIFACT ▾

EDItEUR

Figura 27. Página oficial de ONIX

## b) Metadatos administrativos

Son todos aquellos esquemas que describen la procedencia de un objeto digital, los procesos realizados para su creación o generación, sus características técnicas, sus condiciones de acceso y derechos de propiedad intelectual, y las acciones ya realizadas o previstas relacionadas la preservación del objeto mismo. La codificación de este tipo de datos va encaminada a mantener los objetos digitales utilizables para todos los procesos y funcionalidades propios de una biblioteca virtual, para los que el acceso es el fundamental, como hemos venido insistiendo a lo largo del artículo, pero que afectan también a su gestión y administración. Todo ello significa en la práctica registrar, para saber, entre otras cosas, los formatos de los ficheros, su fecha de creación, los agentes que los han creado, el software y hardware con el que ha sido creado un fichero y cualquier otra circunstancia de este tipo. Desde luego, no son técnicas muy diferentes de las ya habituales en las bibliotecas tradicionales en los que un libro se adquiere, se registra, se cataloga e indiza, se describe físicamente, se ubica, se tejea y se revisa periódicamente su estado de conservación.

Así, como ejemplos, los ficheros de imagen resultantes del escaneado tendrán unas características técnicas que conviene codificar con el esquema MIX (*Metadata for Still Images in XML*)<sup>179</sup> versión 2.0. Se trata de datos propios de los ficheros de imagen como su formato (p.e., TIFF, JPEG) y su versión (p.e., TIFF 6.0, JPEG 2000), el esquema de bits (p.e., blanco y negro, escala de grises, color), la compresión utilizada (p.e., CCITT, LZW, JPEG 2000), el índice de compresión, los algoritmos utilizados para comprobar la

<sup>179</sup> <http://www.loc.gov/standards/mix/> [Consultado el 1 de marzo de 2010]

fijeza de los ficheros y las características de las imágenes (altura, anchura, color, perfil de color, etc).

Los ficheros texto obtenidos a partir de las imágenes por procedimientos de reconocimiento óptico de caracteres tienen también sus características técnicas que conviene registrar en el mismo momento de su creación en el esquema textMD<sup>180</sup>. Con este esquema se codificará el software y hardware utilizado para la realización del OCR y su versión, o el procesador de texto (p.e., Word, OpenOffice), el juego de caracteres (p.e., ISO 8859-1, UTF-8), la variedad de tipos y sus tamaños (p.e., Verdana 10), el lenguaje de marcado (p.e., TEI, ALTO), y los requisitos tanto de software como de hardware para su impresión y visualización. Del mismo modo que los ficheros de imagen y texto, los más habituales en bibliotecas virtuales, se pueden describir las características técnicas de ficheros de audio y video<sup>181</sup> e incluso 3D, utilizando VRML(*Virtual Reality Markup Language*)<sup>182</sup>.

Además de las características técnicas, los metadatos administrativos deben indicar todas aquellas circunstancias sobre los derechos de propiedad intelectual asociados a una copia digital. El esquema METSRights<sup>183</sup> prescribe la codificación de la declaración de derechos del objeto, la identificación del propietario de los mismos, las categorías de derechos (dominio público, copyright, licencia, contrato, etc.), y los permisos y restricciones (localizar, mostrar, copiar, duplicar, modificar, imprimir) asignados a perfiles de usuarios. Todos estos datos servirán para informar a los usuarios de las operaciones que pueden realizar o no y a la biblioteca virtual le permitirán realizar una gestión correcta de los objetos digitales en función de los derechos de autor de cada uno de ellos. No puede olvidarse que la gestión de derechos de autor constituye uno de las mayores complicaciones de una colección digital, sobre todo si, como es muy frecuente, y digno de estudio jurídico, el correcto respeto de estos derechos supone tratamientos diferenciados para las partes de un mismo objeto o bien para los distintos sujetos de derechos, como se ha visto en el epígrafe 8.

Por último, dentro de los metadatos administrativos, debemos mencionar los metadatos de preservación<sup>184</sup>. Aunque existen algunas opiniones que dicen que, estrictamente hablando, los metadatos de preservación de un objeto digital sólo afectan a un sistema de preservación digital y que serían superfluos en cualquier otro sistema, lo cierto es

---

<sup>180</sup> <http://www.loc.gov/standards/textMD/> [Consultado el 1 de marzo de 2010]

<sup>181</sup> Audio Visual Prototype Project Working Documents : Extension Schemas for the Metadata Encoding and Transmission Standard  
<http://www.loc.gov/rr/mopic/avprot/metsmenu2.html> [Consultado el 1 de marzo de 2010]

<sup>182</sup> Information technology -- Computer graphics, image processing and environmental data representation -- Extensible 3D (X3D) encodings  
[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=44681](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44681)

<sup>183</sup> <http://cosimo.stanford.edu/sdr/metsrights.xsd> [Consultado el 1 de marzo de 2010]

<sup>184</sup> Agenjo Bullón, Xavier, y Francisca Hernández Carrascal. "Exegi monumentum aere perennis (Hor. C. 3.30): preservación de materiales bibliográficos digitalizados." Boletín de la ANABAD. LVIII, 2 (2008), 101-111

que su existencia permite a cualquier biblioteca virtual hacer un seguimiento de la calidad de sus objetos digitales. El modelo *Open Archival Information System (OAIS) Reference Model*<sup>185</sup> establece la necesidad de registrar una serie de datos mínimos que permitan seguir el ciclo de vida de los objetos digitales para garantizar su accesibilidad futura, su correcta interpretación, su autenticidad e integridad, cualquiera que sea el sistema o metodología que se utilice para ello. La clave está, pues, en los datos asociados a los objetos digitales en el momento de ingreso en un sistema, procedentes de un proveedor de contenidos digitales (p.e., proceso de digitalización) o de un intercambio de información entre repositorios o bibliotecas virtuales.

En consecuencia y como fruto de un proceso de digitalización es recomendable que se registren esa serie de datos mínimos independientemente de que formalmente el ingreso de información se realice en un repositorio de preservación o no. Ha de tenerse en cuenta además que cualquier biblioteca virtual debe disponer de procedimientos de preservación básicos (p.e., copias de seguridad, actualización de versiones entre formatos, transformación de un formato a otro, comprobación de alteraciones en ficheros, etc.) que se realizan regularmente, aunque propiamente no puedan denominarse repositorios de preservación.

El esquema para este tipo de metadatos es PREMIS (*Preservation Metadata Implementation Strategies*)<sup>186</sup>, basado en el modelo de referencia OAIS que fue el adoptado por el estudio realizado en 2002 por el grupo de trabajo *Preservation Metadata Framework*<sup>187</sup>, organizado conjuntamente por la OCLC y la RLG para definir los tipos de información que se deberían asociar a un objeto digital a efectos de preservación. La versión actual es la 2.0, publicada en marzo de 2008<sup>188</sup>.

---

<sup>185</sup> ISO 14721:2003 Space data and information transfer systems -- Open archival information system -- Reference model  
[http://www.iso.org/iso/catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=24683](http://www.iso.org/iso/catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683) [Consultado el 1 de marzo de 2010]

<sup>186</sup> <http://www.loc.gov/standards/premis/> [Consultado el 1 de marzo de 2010]

La versión española de Understanding PREMIS de Priscilla Caplan, realizada por María Luisa Martínez-Conde: Entender PREMIS. Madrid, Subdirección General de Coordinación Bibliotecaria, 2009 puede consultarse en [http://www.loc.gov/standards/premis/UnderstandingPREMIS\\_espanol.pdf](http://www.loc.gov/standards/premis/UnderstandingPREMIS_espanol.pdf) o en Travesía <http://hdl.handle.net/10421/981>  
 [Consultado el 1 de marzo de 2010]

<sup>187</sup> A Metadata Framework to Support the Preservation of Digital Objects (the Framework).  
[http://www.oclc.org/research/activities/past/orprojects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf) [Consultado el 1 de marzo de 2010]

<sup>188</sup> <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf> [Consultado el 1 de marzo de 2010]

The Library of Congress > Standards > PREMIS Home

Standards Pages [v] SEARCH

**PREMIS** PRESERVATION METADATA MAINTENANCE ACTIVITY

Official Web Site

**DISCLAIMER**

▶ **PREMIS Data Dictionary for Preservation Metadata version 2.0** [PDF: 3.1MB / 217p.]

This publication includes the PREMIS Introduction, the Data Dictionary, Special Topics, Methodology and Glossary. The data dictionary and report with supporting documentation are also available as separate documents:

- > [PREMIS Data Dictionary, version 2.0](#) (PDF: 1.1MB/184pp.)  
The PREMIS Data Dictionary for core preservation metadata needed to support the long-term preservation of digital materials.
- > [PREMIS Introduction and Supporting Documentation](#) (PDF: 451K/51pp.)  
Provides information on the background, objectives, data model, implementation and other supporting documentation for the PREMIS Data Dictionary.
- > [PREMIS Data Dictionary Entity Hierarchical Listing](#)  
A hierarchical list by PREMIS semantic unit.

▶ **PREMIS schemas**

▶ [Understanding PREMIS](#) by Priscilla Caplan for the Library of Congress: an overview of the PREMIS Data Dictionary for Preservation Metadata (PDF: 543K/26pp.)  
Also available in:  
Spanish: [Entender PREMIS](#) (PDF: 363K/30pp.) **NEW!**

**News and articles:**

- ▶ [PREMIS in METS toolbox](#): open-source tools to support implementation of PREMIS in METS **NEW!**
- ▶ PREMIS Editorial Committee hosted [PREMIS Implementation Fair](#): Oct. 7, 2009 in San Francisco, CA
- ▶ ["Understanding PREMIS"](#) by Priscilla Caplan for the Library of Congress: an overview of the PREMIS Data Dictionary for Preservation Metadata (PDF: 543K/26pp.)  
Also available in:  
Spanish: [Entender PREMIS](#) (PDF: 363K/30pp.) **NEW!**  
Italian: [Capire PREMIS](#) (PDF: 141K/25pp.) **NEW!**
- ▶ ["Repository to Repository Transfer of Enriched Archival Information"](#) by Priscilla Caplan, DLib Magazine, November/December 2008
- ▶ ["Using METS, PREMIS and MODS for Archiving eJournals."](#) by Angela Dappert and Markus Enders, DLib Magazine, September/October 2008

Figura 28. Página oficial de PREMIS

Por último, queremos mostrar que el desarrollo de software libre específicamente dedicado a la extracción de metadatos, habitualmente metadatos técnicos, manifiesta bien a las claras la necesidad de las bibliotecas virtuales de generarlos. Aplicaciones JHOVE (*JSTOR/Harvard Object Validation Environment*)<sup>189</sup>, sirven no sólo a este fin, sino que pueden ser utilizadas para contrastar y validar la calidad de los objetos digitales.

### c) Transmisión e intercambio de objetos digitales: Metadata Encoding and Transmission Standard (METS)<sup>190</sup>

METS surge en 2002 como una iniciativa de la *Digital Library Federation*<sup>191</sup> para proporcionar los datos necesarios, en un entorno XML, para la gestión de materiales digitales, tanto para su almacenamiento como para el intercambio de información entre bases de datos, o para su difusión a los usuarios finales, dentro del modelo *Open*

<sup>189</sup> <http://hul.harvard.edu/jhove/> [Consultado el 1 de marzo de 2010]

<sup>190</sup> <http://www.loc.gov/standards/mets/> [Consultado el 1 de marzo de 2010]

<sup>191</sup> <http://www.diglib.org> [Consultado el 1 de marzo de 2010]

*Archival Information System (OAIS) Reference Model*<sup>192</sup>. Es una norma que mantiene la Library of Congress, como tantas otras, que ha distribuido el 22 de febrero de 2010 la versión 1.9<sup>193</sup>.

METS permite agrupar todos los conjuntos de metadatos de los que hemos venido hablando, los datos descriptivos, los datos administrativos, tratándolos como extensiones al esquema, así como a la estructura de los objetos digitales y su comportamiento en un sistema de información. La norma no prescribe ningún esquema de metadatos descriptivos, sino que puede ser utilizada por cualquier tipo de objeto (material bibliográfico, archivístico o museístico) y así pueden incluirse como datos descriptivos registros codificados en MARC 2709, MARC XML Schema, Dublin Core, EAD, etc. METS propone como extensiones al esquema el uso de otros metadatos como Dublin Core, MARCXML, MODS y VRA<sup>194</sup>, sin embargo cualquier comunidad puede establecer los metadatos descriptivos adecuados, lo que es conveniente que se refleje en el correspondiente perfil de METS<sup>195</sup>. Otro tanto ocurre con los datos administrativos, que puede incorporarse al propio documento METS o bien puede definirse mediante un enlace a una localización externa, tal y como se refleja en las extensiones METS.

La estructura de un objeto digital, según METS, es realizada por dos grupos de elementos básicos: el registro de ficheros y el mapa que define las relaciones entre los ficheros y las informaciones complementarias para su visualización, reproducción, etc. El mapa de la estructura de los ficheros es el corazón de los metadatos METS. En él se definen las secciones que componen el objeto (p.e., los capítulos de una obra), los títulos y numeraciones asociados a cada una de las secciones y la secuencia de ficheros, componiéndose con ello el mapa de navegación de un objeto digital.

Desde el punto de vista del usuario, el esquema de metadatos METS viene a solventar uno de los problemas más complicado de resolver, por lo prolijo, en las colecciones digitales, puesto que se trata de ofrecer copias digitales con, como mínimo, las mismas características de manejo que el objeto original (ojear un índice, ir a una determinada página, localizar un determinado capítulo o epígrafe, saber qué página se está leyendo, etc.). Y desde luego ofrece una respuesta normalizada a propuestas mucho más complejas que ésta como podrían ser la presentación relacionada de miniaturas e imágenes legibles, de copias digitales en formato texto y en formato digital, copias para visualización, preservación y reproducción, copias con distintos derechos de autor o bien copias destinadas a diferentes servicios.

---

<sup>192</sup> <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Véase también el sitio Web de Consultative Committee for Space Data Systems <http://public.ccsds.org/default.aspx> [Consultado el 1 de marzo de 2010]

<sup>193</sup> <http://www.loc.gov/standards/mets/version19/mets.xsd> [Consultado el 1 de marzo de 2010]

<sup>194</sup> Visual Resources Association <http://www.vraweb.org/> y especialmente <http://www.vraweb.org/projects/vracore4/index.html>

<sup>195</sup> <http://www.loc.gov/standards/mets/mets-profiles.html> [Consultado el 1 de marzo de 2010]

Ya hemos mencionado cómo METS se diseñó para poder ser utilizado en repositorios OAIS de preservación. En este sentido tiene muchos puntos en común con PREMIS ya que ambos, con su propia semántica, permiten la codificación de metadatos descriptivos y administrativos y relacionan los ficheros componentes de un objeto digital y su estructura. Por ello, los metadatos de un objeto digital pueden realizarse o bien usando PREMIS como esquema básico o bien utilizando PREMIS como una extensión<sup>196</sup> de METS<sup>197</sup>, propuesta ésta última más extendida y para lo que pueden utilizarse las recomendaciones *Guidelines for using PREMIS with METS for exchange*<sup>198</sup>. La versión 1.9 de METS hace una recomendación concreta para se haga de esta manera. Evidentemente, cualquier sistema que reciba un paquete de información procedente de un proceso de digitalización, por ejemplo, debe establecer sus requisitos mínimos y obligatorios, o bien haciendo referencia a alguno de los perfiles de METS existentes<sup>199</sup> o bien desarrollando sus propios perfiles. La reciente publicación *de PREMIS in METS Toolbox*<sup>200</sup> facilita considerablemente la tarea, pues se trata de herramientas de código abierto creadas por el *Florida Center for Library Automation* para la *Library of Congress* en 2009. Tal vez en un próximo futuro la normativa WARC permita un envoltorio global con características análogas a la norma ISO 25577:2008 MarcXchange.

201

---

<sup>196</sup> "Using Premis with METS" refleja una serie de perfiles que especifican el uso de los metadatos PREMIS envueltos en un esquema METS y la relación con otras extensiones de metadatos que pueden coincidir total o parcialmente con METS. <http://www.loc.gov/standards/premis/premis-mets.html>  
Pero véase más adelante la nota 77.

Resulta de gran interés el perfil *ECHO Dep Generic METS Profile for Preservation and Digital Repository Interoperability*. <http://www.loc.gov/standards/mets/profiles/00000015.html>  
[Consultado el 1 de marzo de 2010]

<sup>198</sup> <http://www.loc.gov/standards/premis/guidelines-premismets.pdf> [Consultado el 1 de marzo de 2010]

<sup>199</sup> Estos perfiles incluyen la definición de aplicación de una institución (p.e, National Library of Australia) o de un sistema concreto (p.e., Dspace).

<sup>200</sup> <http://pim.fcla.edu/> [Consultado el 1 de marzo de 2010]

<sup>201</sup> ISO 28500:2009 Information and documentation -- WARC file format  
[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=44717](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44717) [Consultado el 1 de marzo de 2010]

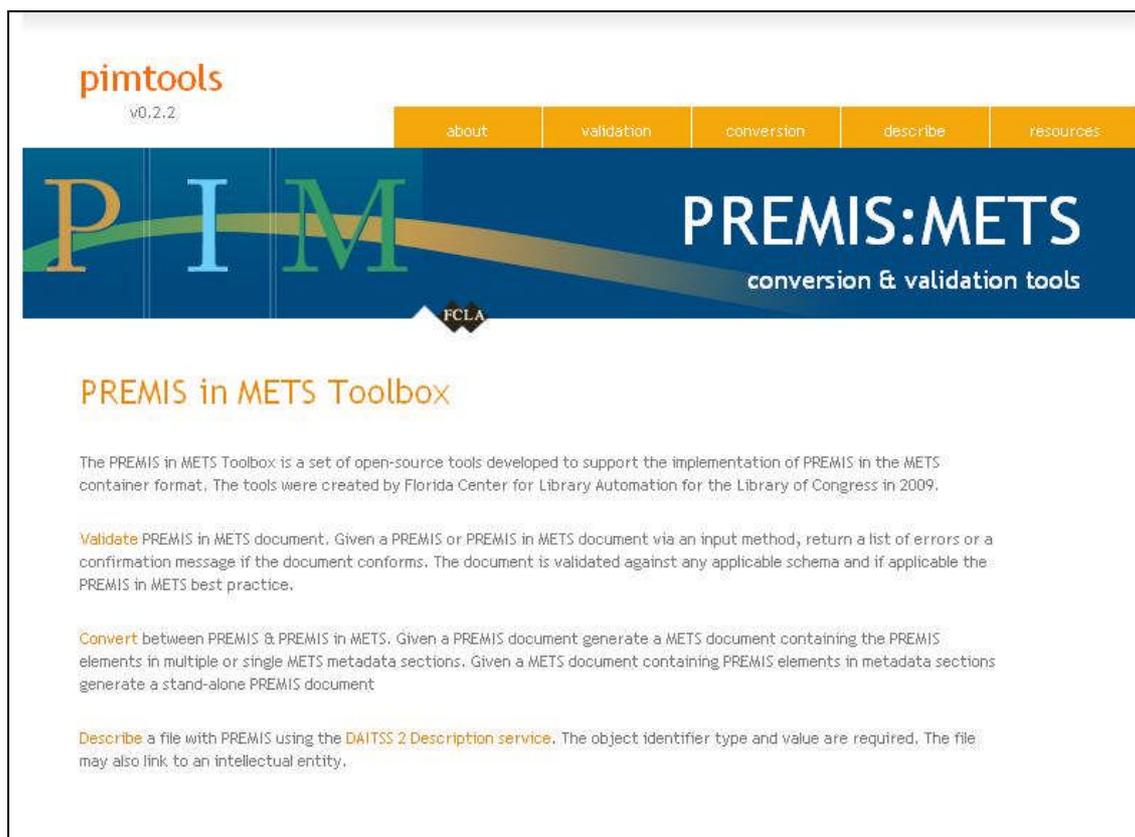


Figura 29. PREMIS in METS Toolbox

## 12. Tejiendo la Web semántica: Ontologías para bibliotecas virtuales

No trataremos de definir la web semántica<sup>202</sup>, ni las estructuras de información que la sustentan como las ontologías o los lenguajes en los que están basadas estas estructuras como RDF<sup>203</sup> o OWL<sup>204</sup>. Sin embargo, la evolución de la web semántica, entendida como la suma de datos, ontologías y aplicaciones, ha experimentado en los últimos años ciertos cambios que han tenido gran influencia en el mundo bibliotecario y que constituyen el marco para algunas experiencias reseñables que dejan ver a las claras una fuerte tendencia para reformular los datos propios de las bibliotecas y de las bibliotecas virtuales, según los requisitos de la Web semántica.

En resumen, las ontologías, como modelo y expresión de un dominio del conocimiento, pueden contribuir a mejorar la precisión de las búsquedas en la web, eliminando la ambigüedad entre términos y relacionando semánticamente unos conceptos con otros. Además, el desarrollo de programas y aplicaciones de búsqueda basadas en ontologías podría extender la recuperación de información significativa y pertinente a toda la web.

<sup>202</sup> Berners-Lee, Tim y James Hendler, Ora Lassila. The Semantic Web. "Scientific American", May 2001. <http://www.scientificamerican.com/article.cfm?id=the-semantic-web> [Consultado el 2 de marzo de 2010]

<sup>203</sup> Resource Description Framework. <http://www.w3.org/RDF/> [Consultado el 2 de marzo de 2010]

<sup>204</sup> Web Ontology Language. <http://www.w3.org/2004/OWL/> [Consultado el 2 de marzo de 2010]

Indudablemente, para que esto sea así debe existir una interconexión directa o a través de aplicaciones entre diferentes ontologías. Sin embargo, la propuesta de la web semántica para la mejora de las capacidades de búsqueda y recuperación de información en la Web no había producido resultados relevantes o con la suficiente extensión para animar a las bibliotecas a realizar el esfuerzo que conlleva la transformación de sus datos en estructuras de información semánticas. Con frecuencia, se ha señalado que muchas experiencias basadas en ontologías no han salido del marco de los proyectos de investigación o de entornos cerrados y la manifestación del poder real de la web semántica se ha ido postergando año tras año. Mientras tanto, otras propuestas no semánticas (i.e., Google) han continuado mostrando eficazmente la capacidad del tratamiento estadístico de los textos<sup>205</sup> para la búsqueda y recuperación de información, del mismo modo que ha resultado espectacular el desarrollo alternativo de los ‘*mashup web services*’ que reeditan datos procedentes de diferentes, construidos sobre APIs (*Application Programming Interface*).

The screenshot shows the Linked Data website interface. The main content area features a large, complex network diagram of interconnected circular nodes representing various datasets and services, with 'Dilpedia' at the center. The left sidebar contains a navigation menu with items like 'Home', 'Guides and Tutorials', and 'Tools'. The right sidebar includes a login section for authorized users and a syndication icon.

Figura 30. Página de Linked Data<sup>206</sup>

<sup>205</sup> Halevy, A., Norvig, P. & Pereira, F. The unreasonable effectiveness of data. "IEEE Intelligent Systems" 24, 8–12 (2009) <http://www.computer.org/portal/web/csdl/doi/10.1109/MIS.2009.36> [Consultado el 2 de marzo de 2010]

<sup>206</sup> <http://linkeddata.org/> [Consultado el 2 de marzo de 2010]

En 2006, también Tim Berners-Lee<sup>207</sup> realizó una simplificación de la visión de la web semántica proponiendo 4 reglas o pasos para la publicación de datos estructurados en RDF usando URIs<sup>208</sup> en lugar de centrarse en el desarrollo de ontologías y reglas de inferencia. Esta simplificación iba dirigida a reducir las barreras para la adopción de este sistema, del mismo modo que la Web inicial simplificó los enlaces hipertextuales entre documentos, facilitando así su extensión. Bajo estos requisitos, conocidos como *Linked Data*, el crecimiento de los datos expuestos e interrelacionados ha sido exponencial. En 2008 se inició el proyecto *Linking Open Data Community Project*<sup>209</sup> dentro del *W3C Semantic Web Education and Outreach Interest Group*<sup>210</sup> con el objetivo de abrir datos, no documentos, en RDF a la Web y estableciendo enlaces entre datos de diferentes fuentes. Abandonando, de momento, la idea inicial de la interconexión de ontologías, el entramado de datos de la Web Semántica se está realizando a través de transformación de datos en RDF y su interrelación (p.e., la persona Molière descrita en RAMEAU e interconectada con la descripción de la misma persona en el fichero de autoridades de la Biblioteca del Congreso o con la entrada de la Wikipedia). Este entramado de enlaces entre datos en RDF permitirá la navegación entre un mismo dato de una fuente a otra mediante de navegadores semánticos o de motores de búsqueda. El 2 de marzo de 2010 estaban registrados 177 ‘data sets’<sup>211</sup> que ofrecían 13.112.409.691 tripletas, con 142.605.717 enlaces entre los diferentes conjuntos de datos. Al ofrecerse de forma abierta<sup>212</sup> datos y relaciones entre datos, en lugar de documentos HTML, otras aplicaciones pueden usarlos para nuevas formas de presentación de información.

En el sitio del proyecto *Linked Open Data* se ofrece también una lista de navegadores, mezcladores (Mashups) y otras aplicaciones que permiten apreciar la dirección que está tomando la web semántica, a veces de forma poco usable, y la enorme velocidad que está imprimiendo *Linked Open Data* al desarrollo de nuevos conjuntos de datos en RDF. Muestra de ello es la publicación de datos gubernamentales<sup>213</sup> a la que se han

---

<sup>207</sup> Berners-Lee, T. (2006). *Linked Data - Design Issues*.  
<http://www.w3.org/DesignIssues/LinkedData.html>. [Consultado el 2 de marzo de 2010]

<sup>208</sup> Uniform Resource Identifier <http://tools.ietf.org/html/rfc2396> [Consultado el 2 de marzo de 2010]

<sup>209</sup> <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> [Consultado el 2 de marzo de 2010]

<sup>210</sup> <http://www.w3.org/2001/sw/sweo/> [Consultado el 2 de marzo de 2010]

<sup>211</sup> <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/LinkStatistics>  
 [Consultado el 2 de marzo de 2010]

<sup>212</sup> Véase la entrada en la Wikipedia sobre el movimiento Open Data y sus relaciones con otras iniciativas para la libre difusión de contenidos como Open Access. [http://en.wikipedia.org/wiki/Open\\_Data](http://en.wikipedia.org/wiki/Open_Data)  
 [Consultado el 2 de marzo de 2010]

<sup>213</sup> eGovernment at W3C improving access to government through better use of the Web  
<http://www.w3.org/egov/> [Consultado el 2 de marzo de 2010]

sumado de momento las administraciones de Estados Unidos<sup>214</sup>, Inglaterra<sup>215</sup> y Australia<sup>216</sup>. Recientemente el Principado de Asturias ha anunciado también la apertura de sus datos a través de la *Fundación CTIC*<sup>217</sup>. Otras experiencias de datos abiertos son las del *New York Times*<sup>218</sup> o la *CERN Library* que ofrece su catálogo bibliográfico para que pueda ser descargado por terceros<sup>219</sup>. Desde el otro lado de la conexión, proyectos como *Open Library*<sup>220</sup> hacen uso de los datos así distribuidos para incrementar tanto su contenido como el entramado de relaciones.

Entre las aplicaciones más llamativas pueden consultarse algunas de gran calidad como *BBC Programmes*<sup>221</sup> o *BBC Music Beta*<sup>222</sup>, *Faviki*<sup>223</sup> que permite anotar sitios web con etiquetas semánticas, o *DBpedia*<sup>224</sup>, que seguramente es el núcleo central de *Linked Open Data* y que es una adaptación semántica de la Wikipedia.

---

<sup>214</sup> <http://www.data.gov/> [Consultado el 2 de marzo de 2010]

<sup>215</sup> <http://data.gov.uk/> [Consultado el 2 de marzo de 2010]

<sup>216</sup> <http://data.australia.gov.au/catalogue> [Consultado el 2 de marzo de 2010]

<sup>217</sup> <http://datos.fundacionctic.org/> [Consultado el 2 de marzo de 2010]

<sup>218</sup> <http://data.nytimes.com/> [Consultado el 2 de marzo de 2010]

<sup>219</sup> De momento, el catálogo está en MARC XML y anuncian su próxima distribución en RDF <http://library.web.cern.ch/library/Library/bookdata.html> [Consultado el 2 de marzo de 2010]

<sup>220</sup> <http://openlibrary.org/> [Consultado el 2 de marzo de 2010]

<sup>221</sup> <http://www.bbc.co.uk/programmes> [Consultado el 2 de marzo de 2010]

<sup>222</sup> <http://www.bbc.co.uk/music> [Consultado el 2 de marzo de 2010]

<sup>223</sup> <http://www.faviki.com/pages/welcome/> [Consultado el 2 de marzo de 2010]

<sup>224</sup> <http://dbpedia.org/About> [Consultado el 2 de marzo de 2010]

The screenshot shows the BBC website interface with a search bar containing 'Beatles'. The page is divided into several sections:

- Coming up on TV & Radio:**
  - BBC TWO:** 'Weston' on BBC Two, 08 Mar 10 15:45. Description: Anita gets all hot under the collar with some Beatles memorabilia prompting a small, if a little off-key performance and Michael is delighted that a piece of Minton lands on his table. There's a treat in store for all watercolour lovers everywhere, and the auction throws up a few surprises.
  - BBC RADIO 2:** 'Sounds of the 60s' on BBC Radio 2, 06 Mar 10 08:05. Description: This week the A-Z of The Beatles reaches 'Wild Honey Pie'. Three-In-A-Row marks the anniversary of the Battle of the Alamo and there are tracks by Nirvana, The Amen Corner, The Flirtations, Bobby Bland, Fleetwood Mac, Gale Garnett, and Otis Redding. With...
- BBC NEWS:**
  - 'Sir Paul unveils outdoor UK shows' (NEW 15 hours ago | Wales). Description: Sir Paul, who turns 68 a couple of days before the Glasgow show, will draw on music from throughout his career including the Beatles catalogue and songs from his years with Wings. There will also more recent tracks, including music from his 2008 album Electric Arguments, which he recorded under...
  - 'Sheep leads animation shortlist' (01 Mar 10 | Arts & Culture). Description: In the new media category, the opening to the Beatles Rock Band game is shortlisted against another video game, DJ Hero for best commissioned animation. The majority of the biennial awards are judged by a jury of animation experts and are joined this year by theatre director Sir Jonathan Miller...
- BBC iPlayer:** 'Sounds of the 60s - 27/02/2010' (27 Feb 10). Description: This week the A-Z of The Beatles reaches 'Why Don't We Do It In The Road'. The featured EP is by The Rolling Stones, there's an instrumental by The Who and tracks from Ambrose Slade, Ray Charles, Roy Orbison, The Roulettes and Family Dogg. With...
- Knowledge:** 'Music - The Beatles'. Description: The Beatles The Beatles Formed 1957. Disbanded 10 April 1970. 1980: The legend of Lennon.

Figura 31. Sitio Web de BBC Music

La adaptación del mundo bibliotecario a *Linked Data* está dando lugar a la publicación de ficheros de autoridad y catálogos en RDF. Puede decirse que en los últimos años la aportación que han realizado las bibliotecas a la web semántica es sustancial lo que ha significado, sobre todo, la apertura de los ficheros de autoridad. La identificación unívoca, la desambiguación de términos, así como las relaciones entre los términos y el modo normativo de referenciarlas, tan necesarias para la web semántica, forman parte de las prácticas bibliotecarias y bibliográficas tradicionales y son el alimento básico de casi cualquier ontología, esquema o aplicación semántica. Téngase en cuenta que, independientemente de su denominación, estamos hablando de las clases 'personas', 'entidades', 'grupos', 'lugares', 'periodos cronológicos', 'periodos culturales', 'materias', etc., es decir la esencia, podríamos decir, casi para cualquier ontología. Así en 2008 la Library of Congress anunció la distribución de sus encabezamientos de materia en SKOS<sup>225</sup> y mantiene el *Library of Congress Authorities and Vocabularies service*<sup>226</sup> que ofrece los registros de autoridad según la filosofía de *Linked Data*, tanto a usuarios humanos como a máquinas y que, además, presentan la ventaja de estar vinculados con los registros de RAMEAU<sup>227</sup>. Es curioso que en el marco de la Web

<sup>225</sup> <http://www.w3.org/2001/sw/wiki/SKOS>

Véase también <http://www.w3.org/2001/sw/wiki/SKOS/Datasets> para los conjuntos de datos disponibles en SKOS (p.e., Agrovoc, Alexandria Digital Library Feature Types Thesarus, etc)

La versión en español puede consultarse en <http://skos.um.es/TR/skos-primer/>. [Consultado el 2 de marzo de 2010]

<sup>226</sup> <http://id.loc.gov/authorities/about.html> [Consultado el 2 de marzo de 2010]

<sup>227</sup> RAMEAU (*Répertoire d'autorité-matière encyclopédique et alphabétique unifié*) <http://rameau.bnf.fr/>

semántica se hayan vuelto a resucitar conceptos como *'machine readable'* que parecían obsoletos o atrasados, quizá a algunos no les parezca ya tan anticuado eso de *'Machine Readable Cataloguing'*.

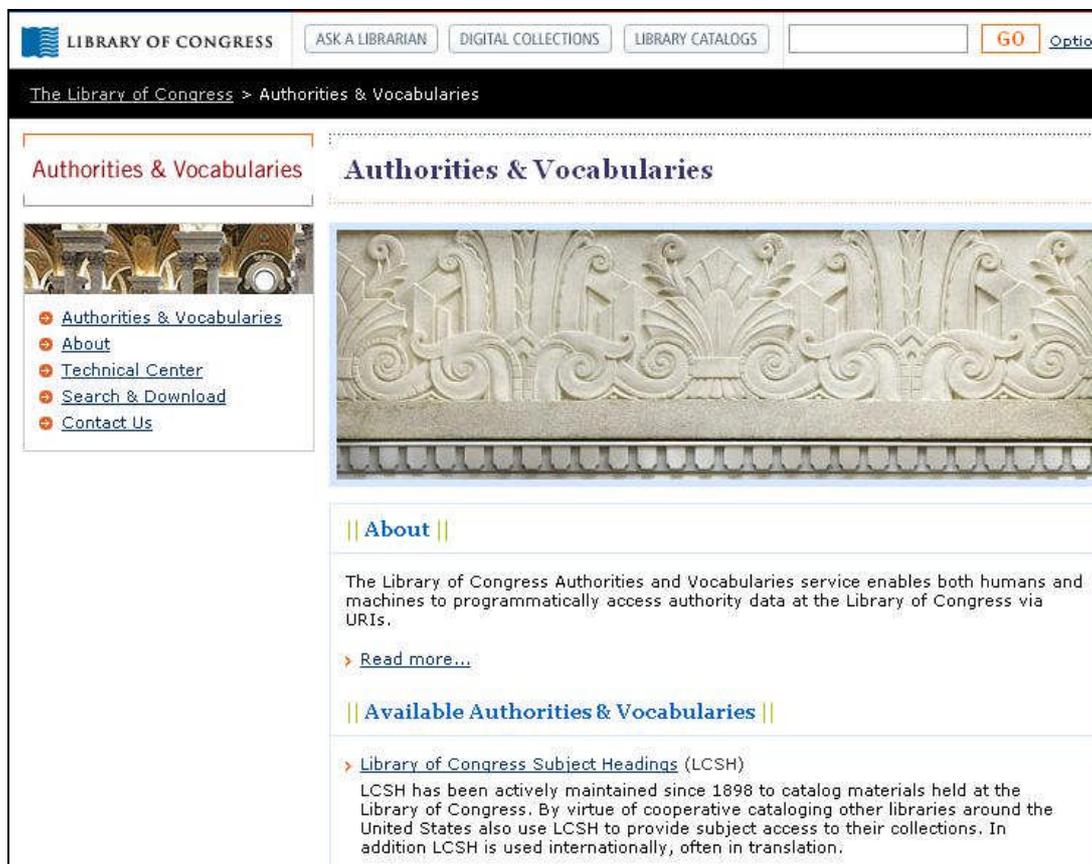


Figura 32. Autoridades de la Library of Congress en RDF

SKOS (*Simple Knowledge Organization System*) es un área de trabajo de la W3C Semantic Activity que desarrolla normas y recomendaciones para trasladar sistemas de organización del conocimiento como tesauros, clasificaciones, encabezamientos de material y taxonomías a los requisitos de la web semántica. Es también, desde el 18 de agosto de 2009, una recomendación del W3C<sup>228</sup> para representar estos sistemas de organización del conocimiento en RDF para su utilización en aplicaciones. En el mismo sentido la OCLC tiene previsto aplicar RDF a VIAF, unificando SKOS con FOAF para los registros de personas<sup>229</sup>, y a FAST (*Faceted Application of Subject Terminology*)<sup>230</sup>

también ofrece sus registros de autoridad en SKOS <http://www.es.yu.nl/STITCH/rameau/> [Consultado el 2 de marzo de 2010]

<sup>228</sup> <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> [Consultado el 2 de marzo de 2010]

<sup>229</sup> Nótese la diferencia entre 'personas' en el entorno FOAF y 'nombres de personas' en el entorno MARC. FOAF (*Friend of a friend*) es una ontología para la descripción de personas, sus actividades y relaciones con otras personas. <http://xmlns.com/foaf/spec/> [Consultado el 2 de marzo de 2010]

<sup>230</sup> <http://www.oclc.org/research/activities/fast/default.htm> [Consultado el 2 de marzo de 2010]

Otro camino para alcanzar la propuesta de *Linked Open Data* de ofrecer datos y su interrelación por medio de aplicaciones específicas es el representado por RDFa<sup>231</sup> que, dicho brevemente, permite embeber tripletas RDF en ficheros XHTML. De este modo, cualquier sitio web puede ofrecer, embebido en el código de una página, datos de autoridades o de recursos de información (digitales o no). Representa una ventaja sobre la distribución de ficheros en RDF como es la actualización de los datos y permite que los grandes motores de búsqueda como Google, Yahoo o Bing hagan uso de esta información para generar relaciones entre sitios y páginas Web. De hecho, las empresas citadas tienen servicios y herramientas semánticas, en producción o anunciados, basados en el uso de RDFa. Es muy frecuente que los datos que se ofrecen en RDF estén accesibles vía RDFa, como ocurre con los encabezamientos de materia de la Library of Congress y que puede apreciarse reclamando el código fuente de sus páginas.

Por supuesto, en el ámbito de las bibliotecas virtuales la interrelación de contenidos puede ser un avance crucial. Así, siguiendo el modelo de *Linked Open Data* o de las ontologías, debe ser posible navegar a todos los recursos digitales de un autor (por ejemplo) utilizando los registros de autoridad (multilingües y multiescritura) como entramado. Este es el modelo que ha seguido *la Ontología del Patrimonio de Cantabria*<sup>232</sup>, integrando y entrelazando datos (entre ellos ficheros de autoridad y registros bibliográficos MARC 21) de distintas procedencias y relacionados con el patrimonio cultural, y el que tiene planificado seguir la Fundación Ignacio Larramendi<sup>233</sup> para su *Biblioteca Virtual de Polígrafos* por medio de la aplicación de SKOS.

---

<sup>231</sup> <http://www.w3.org/TR/xhtml-rdfa-primer/> [Consultado el 2 de marzo de 2010]

<sup>232</sup> Proyecto financiado por la Fundación Marcelino Botín. Puede verse una descripción del mismo en Hernández Carrascal, Francisca. Intercambio de información sobre patrimonio cultural: la Ontología del Patrimonio de Cantabria., 2009 . En: “XI Jornadas de Gestión de la Información: Servicios polivalentes, confluencia entre profesionales de archivo, biblioteca y documentación”. Madrid, 2009 <http://eprints.rclis.org/17376/> [Consultado el 2 de marzo de 2010]

<sup>233</sup> De hecho, y con fecha 30 de septiembre de 2004, la empresa DIGIBÍS íntimamente vinculada a la Fundación Ignacio Larramendi, presento una solicitud de ayuda al “Plan Nacional de Investigación científica, desarrollo e Innovación tecnológica (I+d+I 2004-2007). Fomento de la Investigación técnica”, para un proyecto titulado *Ontología y Web Semántica de Polígrafos*; proyecto que fue aprobado con fecha de 11 de noviembre de 2004 para la creación de una Ontología para la elaboración de la Biblioteca Virtual Ignacio Larramendi o, por decirlo de forma más clara, para conseguir que la estructura informática de la Biblioteca Virtual Ignacio Larramendi se apoyará en cierto grado en una Ontología. [http://www.larramendi.es/i18n/img/pdf/05\\_introduccion\\_general\\_proyecto\\_b.pdf](http://www.larramendi.es/i18n/img/pdf/05_introduccion_general_proyecto_b.pdf) [Consultado el 2 de marzo de 2010]

Patrimonio de Cantabria  
Todo el patrimonio cultural y natural de Cantabria

Fundación Marcelino Botín

Buscar Buscador avanzado

Inicio Obras / Monumentos Lugares Personajes Periodos Cronológicos Instituciones Patrimonio

Inicio > ... > BÚSQUEDA > **quevedo** > Torres Quevedo, Leon...

**Torres Quevedo, Leonardo**

imprimir PDF imágenes

 Científico cántabro formado como Ingeniero de Caminos en los años 70, su labor científica iba a destacar por sus trabajos aplicados –sobre todo en el terreno de la mecánica– que le consagrarían internacionalmente como uno de los más prodigiosos inventores de la época. Su prestigio en el campo de la ciencia española quedó refrendado en el reconocimiento que le tributarían las primeras instituciones del país, como la Sociedad Española de Físicas y Química que le nombró presidente en diciembre de 1919, la Sociedad Matemática Española, que también presidió desde 1920 o la Real Academia de Ciencias Exactas, Físicas y Químicas (RACEFQ) cuya presidencia alcanzó en 1928 (habiendo ingresado un cuarto de siglo antes). A ello habría que sumar los premios y distinciones, tales como la medalla Echegaray otorgada por la RACEFQ (1916) o el premio Parville que le concedió la Academia de Ciencias de Pa ... [leer más](#)

**Datos relacionados**

Nacimiento: 1852-12-28. Mollado.

Muerte: 1936-12-18. Madrid.

Padres: [Quevedo de la Maza, Valentina](#)  
[Torres Vildósola y Urquijo, Luis](#)

Estudios: 1871, 1876. Educación Superior: [Escuela Especial de Ingenieros de Caminos, Canales y Puertos](#), Madrid.  
1922. Doctor Honoris Causa. [Sorbona](#), París.  
Educación Secundaria: [Colegio de los Hermanos de la Doctrina Cristiana](#), París.

**Personas**

- Abascal y Carredano, José
- Abín, César
- Jenaro
- Aguilera Santiago, Ignacio
- Aquirre y Benboa, Adolfo de

+ mas

**Instituciones**

- Academia de Ciencias
- Académie des Sciences
- Laboratorio de Mecánica Aplicada
- Real Academia de Ciencias Exactas, Físicas y Natur...

**Figura 33. Leonardo Torres Quevedo en el sitio web del proyecto Ontología del Patrimonio de Cantabria**

La aplicación de las tecnologías de la web semántica en el mundo bibliotecario<sup>234</sup>, especialmente las ontologías, viene dada por diferentes aproximaciones. La primera de ellas está representada por la consideración de que los catálogos de bibliotecas, y otras bases de datos generadas en el mundo bibliotecario, son una fuente de información de primera magnitud para alimentar a la Web con datos. Vista la web semántica, o la nueva generación web como una interrelación entre datos, que supera la estructura de interrelación de documentos mayoritaria en la web actual, los catálogos de bibliotecas aportan datos sumamente estructurados, organizados, recogidos y registrados a lo largo de siglos de profesión. Las cifras de WorldCat, ya mencionadas dan una idea aproximada de la magnitud de datos representados en los catálogos bibliográficos. Si tomamos en cuenta los datos que habitualmente se codifican en un registro bibliográfico podemos apreciar que suponen un registro completo de una de las acciones humanas que más pueden contribuir al desarrollo del conocimiento, de la cultura, de las artes, de la ciencia, etc., como es la edición. Es decir, los catálogos de bibliotecas acumulan organizadamente los autores (prologuistas, editores literarios, traductores...), las obras publicadas (sus traducciones, versiones, adaptaciones...), las fechas en las que fueron publicadas, los editores e impresores que las diseñaron y produjeron, y multitud de otros datos, entre los que no son los menos importantes, los temas o materias de esas obras.

<sup>234</sup> Coyle, Karen. Understanding the Semantic Web: Bibliographic Data and Metadata. "Library Technology Reports, January 14, 2010, 5-31.

<http://alatechsource.metapress.com/content/g212v1783607/offerings/?p=2e400d0fdb3d4cc5a79498d6fa56adf4&pi=0> [Consultado el 2 de marzo de 2010]

En 2002 se publicó la recomendación *Expressing Simple Dublin Core in RDF/XML* [DCMES-XML], que ha sido revisada constantemente hasta su última versión de 2008<sup>235</sup> lo que manifiesta la preocupación de la iniciativa Dublin Core por adaptarse a las recomendaciones del W3C para la web semántica. Un fichero RDF debe estar siempre referido a un esquema (RDF Schema) que define las clases, subclases y propiedades de un determinado dominio por lo que se constituye en su modelo de referencia. En el caso de Dublin Core, el esquema RDF DCMI Metadata Terms<sup>236</sup> es el que define estos elementos, es la estructura de información que permitirá alojar instancias concretas propias de ese dominio (i.e., un registro Dublin Core describiendo un determinado recurso). RDFa será el vehículo para trasladar Dublin Core en RDF al código fuente de una página web lo que va a tener una enorme influencia en la forma en que estamos acostumbrados a recuperar y navegar por la información, y especialmente entre los recursos digitales de una biblioteca virtual.

Igualmente existen otros proyectos cuyo objetivo es la aportación de registros MARC 21 a la web semántica, transformando registros MARC 21 ISO 2709 o MARC XML Schema en RDF<sup>237</sup>. Este ha sido una de las aproximaciones realizadas por el proyecto *Ontología del Patrimonio de Cantabria*, basándose en la ontología FRBRoo<sup>238</sup>, elaborado por el *International Working Group on FRBR/CIDOC CRM Harmonisation*. Sin embargo, y a pesar de la importancia de esta transformación, sobre todo en el ámbito de *Linked Open Data*, los proyectos de este tipo presentan el problema de la falta de adecuación al nuevo modelo de referencia para la descripción bibliográfica, las RDA, o la utilización de un modelo de referencia y un entorno de aplicación, digamos, restringido como es el caso de *Bibliographic Ontology Specification*<sup>239</sup> que ha desarrollado una ontología para las citas bibliográfica. En algunos casos, a pesar de esta falta de adecuación a la nueva normativa, se han desarrollado ontologías de particular interés para las funcionalidades de una biblioteca digital como es el caso de *JeromeDL - e-Library with Semantics*<sup>240</sup>.

---

<sup>235</sup> Expressing Dublin Core metadata using the Resource Description Framework (RDF). <http://dublincore.org/documents/dc-rdf/> Contacted on September 29 2009

<sup>236</sup> <http://dublincore.org/2008/01/14/dcterms.rdf#> [Consultado el 2 de marzo de 2010]

<sup>237</sup> Styles, R., D. Ayers, y N. Shabir. Semantic MARC, MARC21 and the Semantic Web. "WWW 2008 17th International World Wide Web Conference". 2008.

<sup>238</sup> [http://cidoc.ics.forth.gr/frbr\\_inro.html](http://cidoc.ics.forth.gr/frbr_inro.html) [Consultado el 2 de marzo de 2010]

<sup>239</sup> <http://groups.google.com/group/bibliographic-ontology-specification-group?pli=1> [Consultado el 2 de marzo de 2010]

<sup>240</sup> [http://www.jeromedl.org/index.php?option=com\\_content&task=view&id=17&Itemid=38](http://www.jeromedl.org/index.php?option=com_content&task=view&id=17&Itemid=38)  
Puede verse la ontología sobre la que trabaja JeromeDL en <http://www.jeromedl.org/ontology/2.0/spec/index.html> [Consultado el 2 de marzo de 2010]

The screenshot shows the website 'The Bibliographic Ontology'. The header includes the site name and a logo. A navigation bar contains links for News, Specification, Changelog, Examples, Projects, Community, and Login. On the left, there is a 'Navigation' sidebar with a 'Recent posts' link. The main content area is titled 'Bibliographic Ontology Usecase Examples' and contains a list of examples with the following titles:

- A Complex series + proceeding + article use case
- Article
- Book
- Conference Paper
- Conference Presentation
- Contributors
- Enhanced example of the above Zotero example
- Legal Case
- Letter
- Manuscript
- Rejected article submitted to a conference
- Statue
- Thesis

**Figura 34. Ejemplos de BIBO (Bibliographic Ontology) que pueden servir para familiarizarse con una serie de casos de uso**

Aquí es donde cobran sentido las críticas a las RDA basadas en el cambio no es lo suficientemente radical para adaptarse al entorno semántico y que hubieran debido ir más allá. Como ejemplo, uno de los problemas de compatibilidad con el modelo semántico es el requisito de transcribir la portada y utilizar las fuentes de información con un orden establecido, lo que obligaría a definir clases para los datos transcritos, lo que para muchos es redundante y que puede suplirse con el acceso a la copia digital. De hecho, uno de los esfuerzos paralelos a la difusión de las RDA es su transformación en una ontología para ser usada en cualquier entorno y no solamente en el ámbito bibliográfico<sup>241</sup>, haciendo hincapié, al modo de los FRBR, en la mejora de la presentación de información a los usuarios. Hay que hacer notar que las mejoras en las búsquedas previstas por la web semántica y, en general, la mayor preocupación por mejorar la experiencia de los usuarios han llevado a una nueva generación de OPACs y al uso de facetas para facilitar la interrelación de contenidos y proponer nuevas rutas de navegación y recuperación de información. Sin embargo, las búsquedas multifacetas<sup>242</sup> no implican obligatoriamente el uso de ontologías, sino que puede construirse, y así se realiza en muchas ocasiones, a partir de los datos contenidos habitualmente en un registro bibliográfico (tipo de documento, lengua, país de origen, materia, etc.).

Por último, queremos señalar que dos de las cuestiones fundamentales a la hora de la transformación de datos bibliográficos a RDF o OWL están relacionadas con la

<sup>241</sup> Hillmann, Diane [et al.]. RDA Vocabularies: Process, Outcome, Use. "D-Lib Magazine" January/February 2010, 16, Number 1/2. <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html> [Consultado el 2 de marzo de 2010]

<sup>242</sup> Automatic construction of multifaceted browsing interfaces <http://portal.acm.org/citation.cfm?id=1099738>

precisión en la definición del dominio, es decir de la ontología, en la posibilidad o conveniencia de transformar correctamente la totalidad de los datos y, lo que no suena nada raro, en las funcionalidades que la ontología debe ofrecer a los usuarios.

### 13.El que guarda haya: aprender del papel del error y del error del papel

Una revisión muy actualizada, y con la labor de síntesis que le caracteriza, es el trabajo ya citado de Luis Ángel García Melero en el que se recogen buena parte de dos aportaciones de los autores de este artículo y que se utilizan convenientemente actualizados<sup>243</sup> para la redacción de este epígrafe.

Desde el punto de vista del usuario, del lector, que es el que se está abordando principalmente en este artículo, hay que hacer hincapié en que la preservación digital, además de su objetivo primario, que pueda utilizarse a largo plazo la información digital o digitalizada, presenta una característica que debe estar siempre presente como es la posibilidad de consultar la información preservada, hasta el punto de que la consulta es uno de los elementos básicos del modelo de preservación digital. Así, por ejemplo, y aunque el proyecto *Wayback Machine*<sup>244</sup> conserva los contenidos de las páginas web que ha ido recogiendo desde 1995, no puede consultarse. De igual modo, el nuevo proyecto, ya citado anteriormente, *The Memento Project* busca parecidos fines, pero se entiende que la búsqueda y la consulta se realizarán en el mismo modo que en la Web.

De hecho, uno de los mayores problemas que se han planteado para el megaproyecto de *Google Books* es que en él no está previsto, al menos de modo explícito, la preservación a largo plazo de los contenidos. Desde un punto de vista bibliotecario, este es un aspecto que no puede soslayarse y quizá la mejor manera de expresar el concepto que se persigue sea afirmando que preservación y acceso<sup>245</sup> son las dos caras de la misma moneda y para que ello sea una realidad hay que intentar que los procesos que conducen a la preservación también permitan el acceso, lo cual debe realizarse desde el momento de la creación del objeto digital y mediante el uso de diferentes esquemas de metadatos<sup>246</sup>.

---

<sup>243</sup> Se trata Hernando de Larramendi, Tachi [et al.] “Datos y metadatos : la normalización dinámica de los elementos y de los procesos constituyentes de una biblioteca virtual.” Interinformación: XI Jornadas Españolas de Documentación : 20, 21 y 22 de mayo de 2009, Auditorio Palacio de Congresos de Zaragoza. FESABID, 2009. 109-116.

<http://www.fesabid.org/zaragoza2009/actas-fesabid-2009/108-116.pdf> [Consultado el 23 de febrero de 2010]

Una descripción más detallada de las estructuras de preservación puede verse en: Agenjo Bullón, Xavier, y Francisca Hernández Carrascal. “Exegi monumentum aere perennis (Hor. C. 3.30): preservación de materiales bibliográficos digitalizados.” Boletín de la ANABAD LVIII, 2 (2008): 101-111.

<sup>244</sup> <http://www.archive.org/web/web.php> [Consultado el 23 de febrero de 2010]

<sup>245</sup> Un ejemplo de ello puede encontrarse en CLOCKSS (*Controlled LOCKSS*), organización sin fines de lucro formada por editores científicos y bibliotecas especializada con el objetivo de construir un archivo de preservación a largo plazo para publicaciones científicas en la Web.

<http://www.clockss.org/clockss/Home> [Consultado el 2 de marzo de 2010]

<sup>246</sup> Véase el apartado 11 de este artículo

Lo que será necesario es, en la medida de lo posible, procurar que esos procesos se retroalimenten unos a los otros, de modo que el esfuerzo en la creación de objetos digitales no redunde en una dejación de las funciones específicas de la preservación, todo lo cual es perfectamente posible si dentro del flujo de trabajo, tanto de la creación de nuevos objetos digitales, como de aquellos producidos fruto de diferentes procesos de escaneo de la información, que pudiéramos llamar analógica, se prevé la obtención de los mismos.

La preservación digital<sup>247</sup> es el conjunto de estrategias, procesos y técnicas que dan respuesta a los problemas que plantea la conservación de los materiales digitales y de los medios (hardware y software) que se emplean para su almacenamiento y consulta, y que están derivados fundamentalmente de la obsolescencia provocada por la rápida renovación tecnológica y por la inestabilidad de los soportes. Estas técnicas son muy variadas y responden a diferentes situaciones y líneas estratégicas<sup>248</sup> (copias de seguridad, copia de datos en soportes durables, migración, replicación, emulación, transferencia de datos<sup>249</sup>, etc.), aunque, en general, están destinadas a mantener los objetos digitales y sus características de acceso a largo plazo.

Para buscar una definición oficial se pueden citar las Directrices para la preservación del Patrimonio Digital de la UNESCO<sup>250</sup>:

*la preservación digital consiste en los procesos destinados a garantizar la accesibilidad permanente de los objetos digitales. Para ello, es necesario encontrar las maneras de representar lo que se había presentado originalmente a los usuarios mediante un conjunto de equipos y programas informáticos que permiten procesar los datos. Para lograrlo, es necesario que la comprensión y la gestión de los objetos digitales se realice considerándolos desde cuatro puntos de vista: como fenómenos físicos, como codificaciones lógicas, como objetos conceptuales comprensibles para el ser humano y como conjuntos de elementos esenciales que deben ser preservados para ofrecer a los futuros usuarios lo esencial del objeto.*

La preocupación internacional por la preservación digital se concretó en 2002 con la publicación del *Reference Model for an Open Archival Information System (OAIS)*<sup>251</sup>.

---

<sup>247</sup> Bollacker, Kurt D. Avoiding a Digital Dark Age : Data longevity depends on both the storage medium and the ability to decipher the information. "American Scientist". March-April 2010, Volume 98, Number 2, pp. 106. DOI: 10.1511/2010.83.106  
<http://www.americanscientist.org/issues/pub/2010/3/avoiding-a-digital-dark-age> [Consultado el 2 de marzo de 2010]

<sup>248</sup> Strodl, Stephan y Andreas Rauber. Preservation Planning in the OAIS Model. IPRES 2007.  
[http://www.digitalpreservationeuropa.eu/publications/presentations/strodl\\_final\\_3.pdf](http://www.digitalpreservationeuropa.eu/publications/presentations/strodl_final_3.pdf) [Consultado el 2 de marzo de 2010]

<sup>249</sup> Digital Content Transfer Tools (Library of Congress)  
<http://www.digitalpreservation.gov/library/challenge/data-transfer-tools.html> [Consultado el 2 de marzo de 2010]

<sup>250</sup> Directrices para la preservación del Patrimonio Digital. Preparado por la Biblioteca Nacional de Australia. División de la Sociedad de la Información. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Marzo de 2003.  
<http://unesdoc.unesco.org/images/0013/001300/130071s.pdf> [Consultado el 23 de febrero de 2010]

Se trata de una recomendación para el desarrollo de un amplio consenso sobre los requisitos de preservación que debe cumplir un archivo<sup>252</sup> para preservar a largo plazo la información digital y definir la norma *ISO Reference Model for an Open Archival Information System (ISO 14721:2003)*<sup>253</sup>. Se define en un sistema OAIS como una organización, de equipos humanos y sistemas, que tienen la responsabilidad de preservar información y hacerla disponible para una comunidad específica (*'designated community'*). Un sistema será conforme a OAIS si soporta el modelo de información descrito en la norma que, por otra parte, no especifica ningún método de implantación. No es el objetivo de esta artículo detallar el contenido de esta norma, pero sí es importante tener una noción básica del modelo informativo, que nos hará entender con mayor precisión los metadatos necesarios para la preservación.

Un sistema OAIS trabaja con paquetes de información que se distinguen según los procesos en los que participan. Así un *Submission Information Package (SIP)* es un paquete de información que se envía a un sistema OAIS por un productor. Su forma y contenido (información de contenido e información de preservación) deben ser establecidos entre el productor y el sistema OAIS. Una vez que este paquete de información haya sido incorporado al sistema se complementará con las informaciones específicas propias para la gestión de preservación con lo que se transformará en un *Archival Information Package (AIP)*. A su vez un sistema OAIS devolverá a los usuarios que lo soliciten los datos necesarios en forma de un *Dissemination Information Package (DIP)*, de forma que el usuario pueda distinguir claramente la información solicitada.

Por lo tanto, un sistema OAIS debe proporcionar una serie de funciones para ser considerado como tal, fundamentalmente las funciones de integración (*Ingest*), de gestión de datos, de archivo de preservación, de gestión de preservación, de acceso y de administración. Igualmente, forma parte del sistema una definición específica de roles y entidades en la que el *Productor* son las personas o clientes del sistema que generan o proporcionan la información que debe ser preservada, el *Archivo* es la entidad que realiza las funciones y servicios necesarios para el control de las operaciones diarias de un sistema OAIS, y, por último, el *Consumidor* (usuario) que son las personas o clientes que interactúan con los servicios de un sistema OAIS para acceder a la información preservada.

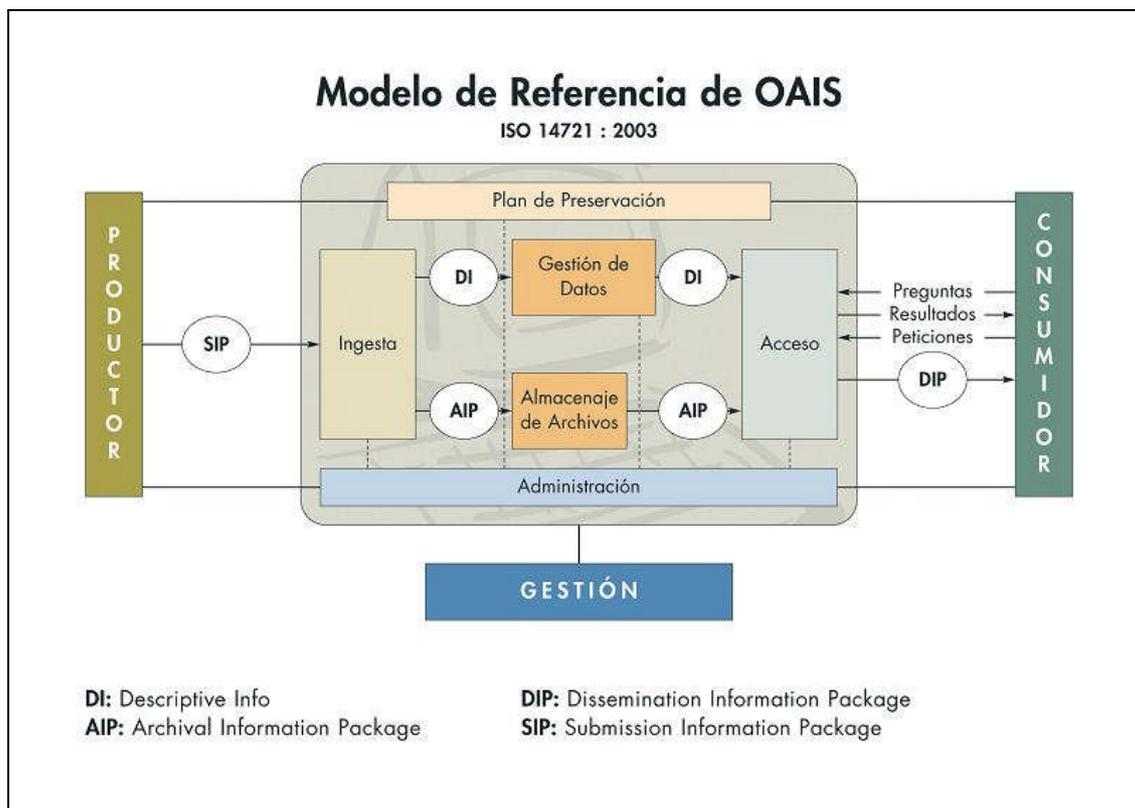
---

<sup>251</sup> Reference Model for an Open Archival Information System (OAIS) : CCSDS 650.0-B-1 : Blue Book : Recommendation for Space Data System Standards. Consultative Committee for Space Data Systems. January 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Se encuentra ya muy avanzada la revisión del Modelo de Referencia que puede consultarse en <http://public.ccsds.org/sites/cwe/riids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf> [Consultado el 2 de marzo de 2010]

<sup>252</sup> Dado el variado uso del término 'archive' en inglés ha de tenerse en cuenta que para esta norma un archivo es la organización que asume la responsabilidad de preservar una o varias colecciones digitales a largo plazo. Igualmente en esta norma el término Abierto ('Open') se refiere a que tanto la norma misma como otras normas o recomendaciones relacionadas que pudieran elaborarse se desarrollarán en foros abiertos. Por tanto, no implica ninguna connotación de restricciones o ausencia de ellas en el acceso a un archivo.

<sup>253</sup> [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=24683](http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683) [Consultado el 23 de febrero de 2010]



**Figura 35. Tomado del Modelo de Referencia para un Sistema abierto de archivo de información (OAIS): CCSDS 650.0-B-1: Blue Book, Figura 4-1**

Dada su adhesión a OAIS, el modelo de datos de PREMIS define cinco entidades especialmente importantes para las actividades de preservación digital, las entidades intelectuales, los objetos digitales, los eventos, los agentes y los derechos, y sobre ellos construye las unidades semánticas apropiadas. Las entidades intelectuales (monografías, publicaciones periódicas, partes componentes, mapas, fotografías, etc.) que son las obras<sup>254</sup> que se consideran a efectos de descripción y gestión, o en términos de FRBR las manifestaciones de una obra, que pueden tener relaciones con otras obras y con una o más representaciones digitales (p.e., las imágenes TIFF de un artículo, un fichero PDF formado por imágenes y texto oculto, o uno o varios ficheros con el texto obtenido de un proceso de OCR). Los objetos digitales son las unidades de información a las que se refieren los procesos de preservación digital y pueden ser ficheros, cadenas de bits o representaciones. Los eventos son las acciones que se realizan sobre un objeto, o por un agente, en un repositorio de preservación. Los agentes son las personas, organizaciones, sistemas, etc. relacionados con un evento en la vida de un objeto o con los derechos particulares sobre un objeto. Los derechos son declaraciones sobre permisos de preservación asociados a un objeto o agente.

<sup>254</sup> La definición de PREMIS de 'entidad intelectual' parece corresponderse con la de 'manifestación' en FRBR más que con la de 'obra' o 'expresión', a pesar de la mención al contenido intelectual. El concepto PREMIS de 'representación digital' puede corresponderse tanto con una 'manifestación' (p.e., en el caso de una obra nacida digitalmente) o con un 'item' (p.e., en el caso de la digitalización).

Cuando se produce el ingreso (*'ingest'*) de un conjunto de informaciones en cualquier repositorio, sea o no sea este de preservación, que procedan de un proceso de digitalización (un *Submission Information Package*, según OAIS) se debe disponer de los metadatos propios de preservación (según PREMIS) para poder realizar el seguimiento actual o futuro del ciclo de vida de los objetos digitales. Esto implica, como ya se ha dicho más arriba, registrar los metadatos propios de la descripción de las unidades intelectuales (descripción bibliográfica); los metadatos técnicos que reflejan las características de los ficheros; los metadatos administrativos propios de los derechos de acceso, uso y reproducción; la relación de ficheros que puede componer una representación digital; y la estructura lógica y física de sus ficheros. El registro de estos metadatos hará posible que en cualquier momento puedan volver a ser regenerados, detectando los cambios que hayan podido afectar a los objetos digitales, sobre todo aquellos que puedan verse afectados por una falta de normalización, en cualquiera de las características básicas de un objeto digital. De hecho, una de las primeras medidas de un sistema de preservación digital es la asignación de metadatos a los objetos digitales.

Al igual que el modelo OAIS, el esquema de metadatos PREMIS no define ninguna estrategia o método de implementación, ni prescribe una tipología de acciones<sup>255</sup>, ni el modo en el que deben recogerse, almacenarse o gestionarse los datos referidos a estas acciones. En cualquier caso, parece muy lógico utilizar el esquema PREMIS, dada la riqueza y precisión creciente de las distintas versiones de su diccionario de datos. En nuestra opinión, debe quedar claro que la no existencia de un repositorio de preservación normalizado no exime de la necesidad de utilizar metadatos de preservación y en concreto PREMIS. Para evaluar la conveniencia de utilizar o no determinadas estrategias *Planets Testbed*<sup>256</sup> proporciona un entorno cooperativo que combina servicios y datos y que permite experimentar esas estrategias, ofreciendo resultados que permiten su evaluación. Seguramente la cuestión fundamental dentro de la preservación digital está en la definición de las acciones de preservación más importantes y en el software que pueda desarrollar estas acciones de una forma automática. Este es uno de los objetivos de programa ICT de la Unión Europea, en torno al cual se están desarrollando algunos proyectos de gran alcance<sup>257</sup>, tanto por sus participantes como por sus objetivos. En esta línea se encuentra PLANETS (*Preservation and Long-term access through Networked Services*)<sup>258</sup>.

---

<sup>255</sup> A finales de 2009 el proyecto europeo PLANETS publicó un estado de la cuestión sobre acciones y software de preservación: "A blue print for the development of new preservation action tools" [http://www.planets-project.eu/docs/reports/PA2-D2\\_Blueprint\\_Preservation\\_Action\\_Tools.pdf](http://www.planets-project.eu/docs/reports/PA2-D2_Blueprint_Preservation_Action_Tools.pdf) [Consultado el 2 de marzo de 2010]

<sup>256</sup> <http://testbed.planets-project.eu/testbed/> [Consultado el 2 de marzo de 2010]

<sup>257</sup> El proyecto europeo PARSE.Insight (INSIGHT *into issues of Permanent Access to the Records of Science in Europe*) publicó en 2009 los resultados de una encuesta sobre la situación de la preservación digital de los datos y publicaciones de investigación. [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_SurveyReport\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf) [Consultado el 2 de marzo de 2010]

<sup>258</sup> <http://www.planets-project.eu/> [Consultado el 2 de marzo de 2010]

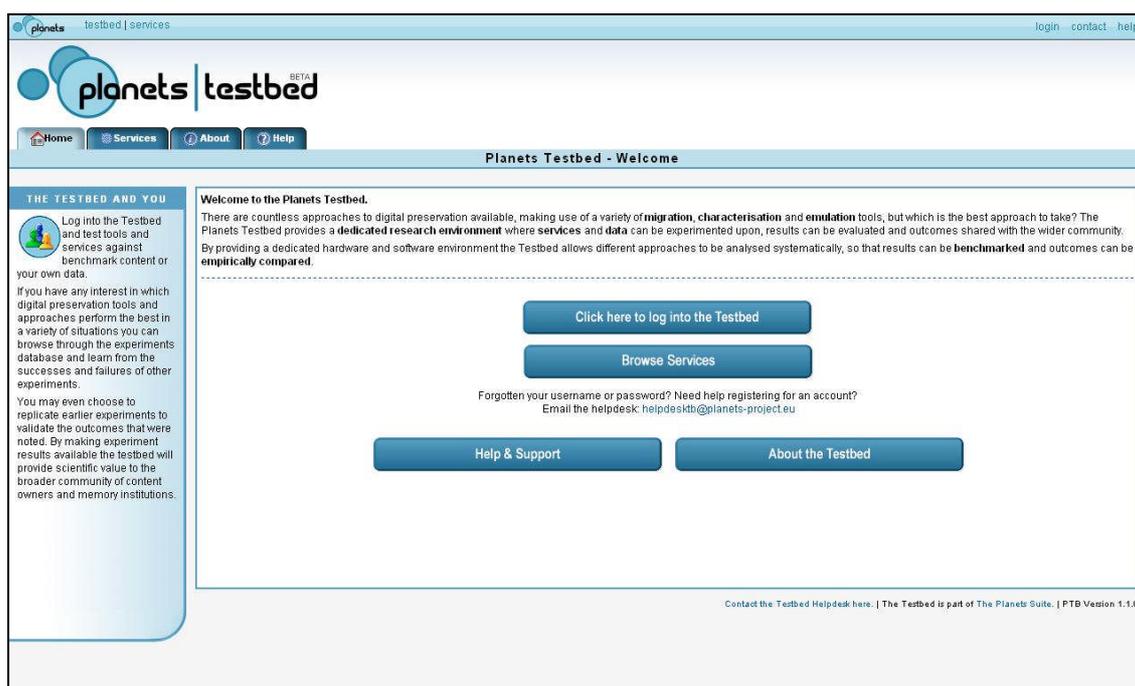


Figura 36. Sitio web de PLANETS Testbed

Todas las normas, recomendaciones y perfiles mencionados, empezando por el modelo OAIS, están sustentados en las ventajas de la normalización. Probablemente, sería muy difícil intentar llevar a cabo una preservación digital a escala global, y mucho menos cooperativa o coordinada<sup>259</sup>, si no se utilizarán procedimientos normalizados que, por un lado permitieran crear y diseñar un software específico, y por otro lado establecer acuerdos de colaboración. Cuando se piensa, por ejemplo, que la Library of Congress, probablemente la institución más potente en el ámbito bibliotecario, ha llevado a cabo acuerdos estratégicos con terceras instituciones para proceder a la preservación del patrimonio digital norteamericano, no cabe duda que esto no sería posible si no se contara con esos instrumentos de normalización. De forma complementaria, el análisis de las fuentes de información sobre preservación digital contribuirá a que las decisiones que se tomen en este aspecto estén guiadas por la normalización y la cooperación<sup>260</sup>. Parece bastante claro que una sola institución, por mucho que así esté contemplado en la legislación, pueda afrontar la organización y procedimientos necesarios para la preservación digital de un país. De aquí, evidentemente, se extrae la lección de que por muy voluntariosos que sean determinados proyectos que puedan llevarse a cabo en España para la preservación del patrimonio digital, si no participan, ni cuentan con la normalización que suponen las normas descritas, es muy posible que su resultado sea baldío.

<sup>259</sup> Skinner, K. y M. Schultz, Eds. A Guide to Distributed Digital Preservation. Atlanta, GA: Educopia Institute, 2010.  
[http://www.metaarchive.org/sites/default/files/GDDP\\_Educopia.pdf](http://www.metaarchive.org/sites/default/files/GDDP_Educopia.pdf) [Consultado el 2 de marzo de 2010]

<sup>260</sup> Una buena recopilación de fuentes de información y recursos es la que ofrece NISO a través de sus Webinars especializados.  
<http://www.niso.org/news/events/2010/preservation/resources> [Consultado el 2 de marzo de 2010]

El cada día mayor problema de la acidez del papel no hace pensar con demasiado optimismo en que se emprenderán de forma sistemática las medidas necesarias para evitar la desaparición, si no todo al menos una parte, de los recursos digitales que existen en la actualidad. Recordemos que Google, tal vez el mayor popularizador de la Web en el siglo XXI no prevé en su modelo de negocio ninguna medida de preservación a largo plazo, lo que desde el punto de vista de las bibliotecas virtuales, que parafraseando a Ortega, tienen que ser virtuales, pero que tienen que ser bibliotecas, resulta crucial. Este problema puede convertirse en otro, también sin resolver, como el de la preservación del papel, que de forma un tanto inconsciente, ni se está desacidificando, ni, lo que sería un paralelismo con la situación actual, se ha emprendido nunca una política sistemática de preservación a largo plazo, basada en la impresión sobre papel permanente.