

DATOS Y METADATOS: LA NORMALIZACIÓN DINÁMICA DE LOS ELEMENTOS Y DE LOS PROCESOS CONSTITUYENTES DE UNA BIBLIOTECA VIRTUAL

Hernando de Larramendi, Tachi (tachi@digibis.com)
Domínguez Muriel, Jesús L. (jesus.dominguez@digibis.com)
Viedma Pélaez, Andrés (andres.viedma@digibis.com)
Hernández Carrascal, Francisca (francisca.hernandez@digibis.com)
Digibis, producciones digitales
Agenjo Bullón, Xavier
xavier.agenjo@larramendi.es
Fundación Ignacio Larramendi

Palabras clave: Bibliotecas digitales; Bibliotecas virtuales; Digitalización; Metadatos; METS; PREMIS; OAIS; JHOVE; METS/ALTO; MIX; textMD

Keywords: Digital Libraries; Virtual Libraries; Digitization; Metadata; METS; PREMIS; OAIS; JHOVE; METS/ALTO; MIX; textMD

Resumen

Se analizan los procesos que tienen lugar para la creación de los objetos digitales fruto tanto del escaneo de los materiales originales como de la descripción bibliográfica y la asignación de metadatos, en especial METS y PREMIS. Se revisa la normativa internacional para crear repositorios OAI, servidores SRU y sobre todo repositorios de preservación a largo plazo. Se examina la información que éstos requieren y la estructura de la información que debe ser almacenada en ellos conforme a la normativa actual y, en particular, al conjunto de esquemas de metadatos necesarios para ello. Se concluye que es un proceso sumamente eficiente el que transforma y codifica según los metadatos de preservación la información generada durante el proceso de digitalización, la descripción conforme a la normativa bibliográfica y el registro de procesos de creación de los objetos digitales, en especial METS/ALTO, textMD y MIX, que conformarán un repositorio de preservación.

Abstract

The different processes needed to create digital objects are analyzed in this paper. The scanning of original materials as well as the bibliographical description and metadata assignment are considered, in particular METS and PREMIS. The international standards to create OAI repositories, SRU servers, and especially long term preservation repositories are reviewed. The information and its structure requires to be storage in digital long term preservation repositories it is studied and following the current standards and mainly the metadata schemas sets necessary for that, especially METS/ALTO, textMD and MIX. At least the paper concludes that it will be a very efficient process if it can transform, codified and validated throughout JHOVE, the preservation metadata generated during the digitization process, the bibliographical description and the creation of the digital objects that will build up the OAIS long term preservation repositories.

Introducción

La definición de biblioteca virtual parece estar encaminada fundamentalmente a aquellas bibliotecas que son fruto de procesos de digitalización y no de las que se constituyen ya desde un primer momento con objetos nacidos digitalmente (*born digital*). Esta división entre bibliotecas virtuales, fruto de la digitalización, y bibliotecas digitales que son resultado de la creación digital de materiales bibliográficos o documentarios, dista mucho de estar ampliamente reconocida y, desde luego, pueden existir, partiendo de esta doble definición, bibliotecas híbridas en las que coexistan los dos tipos de objetos digitales, diferenciados por su origen. En cualquier caso, esta diferenciación es la que utilizaremos como premisa para el presente trabajo.

El proceso más habitual para configurar una biblioteca virtual es la digitalización de materiales bibliográficos según determinados criterios. Frecuentemente, el criterio fundamental es la importancia (por la razón que sea) de la pieza a digitalizar y pocas veces el estado de conservación de la misma o el peligro de desaparición que la hacen susceptible de una vida precaria y breve. Es claro que nos referimos a un caso específico de biblioteca virtual puesto que el criterio de conservación no suele ser utilizado para el proceso de constitución de una colección, sea digital o no¹.

En efecto, el caso de la digitalización de materiales bibliográficos cuya preservación y conservación es muy deseable e incluso imprescindible, puesto que es previsible su desaparición a corto plazo, constituye uno de los grandes desafíos de la profesión bibliotecaria (y archivística e incluso museológica). Nunca se repetirá lo suficiente que todos aquellos materiales bibliográficos, archivísticos y museológicos: grabados, estampas e incluso dibujos que se han impreso, estampado o trazado sobre papel, por cualquier procedimiento, son con toda seguridad materiales abocados a su desaparición más pronto que tarde, como resultado de los procesos de fabricación del papel ya a partir del segundo tercio del siglo XIX y cuyo pH ácido lo hace desintegrarse espontáneamente². A ello hay que añadir la frecuencia de su utilización como es el caso de materiales como la prensa o las novelas de quiosco, confeccionadas para su uso masivo y en modo alguno con vistas a una larga duración.

Convendría recalcar aquí la necesidad de convencer a los editores en soporte papel de la necesidad de imprimir utilizando la norma *UNE-EN ISO 9706:1999 Información y documentación. Papel para documentos. Requisitos para la permanencia. (ISO 9706:1994)*, así como la norma *ISO 11108:1996 Information and documentation — Archival paper — Requirements for permanence and durability*. En Estados Unidos, por ejemplo, es la American Library Association una de las principales fuerzas de presión para que el sector editorial utilice la norma ANSI/NISO Z39.48-1992, norma nacional americana para la permanencia del papel de publicaciones y documentos en archivos y bibliotecas. Ambas normas, aunque no son coincidentes, utilizan el mismo símbolo de conformidad para el papel permanente, y su uso en la impresión de materiales bibliográficos está privilegiada por las bibliotecas americanas, que por su número y la fuerza de la ALA, han conseguido que sea un requisito que tienen muy presente los editores³.

Como curiosidad puede señalarse que cuando se procedió al sellado de Chernóbil, con los residuos nucleares se incorporaron las circunstancias e instrucciones para tratar aquellos restos radioactivos dentro de 1000 años, periodo en el que se supone que habrá cesado su actividad. Pues bien, las instrucciones se escribieron en... pergamino, pues a diferencia del papel o del microfilme, por no hablar ya de los soportes electrónicos, este material ha probado su capacidad para sobrevivir milenios. La posibilidad de hacerlo en piedra o en algún metal fue desestimada naturalmente por la extensión de los escritos. Evidentemente, se utilizó tinta elaborada con componentes químicos que así mismo hubieran probado su perdurabilidad.

Digitalización: datos y metadatos

Es imprescindible deshacer el equívoco que puede existir a la hora de considerar que la digitalización se limita únicamente a convertir un original en un objeto digital vicario, a través de procedimientos de captura de imagen y almacenamiento posterior en un soporte externo con estándares de compresión específicos. No, se trata de crear objetos digitales que estén formados, tanto

¹ Esto no es completamente cierto desde un punto de vista biblioteconómico ya que, en ocasiones, tanto en bibliotecas públicas como en bibliotecas universitarias es necesario emplear parte de los recursos de selección y adquisición a la renovación de materiales deteriorados por el uso frecuente. Bien lo conocen los bibliotecarios que trabajan en ese tipo de instituciones; los manuales, muy utilizados (y fotocopiados), o las colecciones de tebeos o libros ilustrados muy estropeados por lo que Mariano José de Larra llamaba con deliciosa expresión decimonónica la *natural turbulencia* de los lectores infantiles.

² Quizá por su condición de compendio y por su propuesta metodológica, incluso deontológica, puede consultarse *La cuarta salida de El monje digital y sus problemas hemerográficos: una recapitulación* / Xavier Agenjo Bullón. - En: *Boletín de la ANABAD*, ISSN 0210-4164, tomo 54, nº 4, 2004, pp. 119-138.

³ Ivar A. L. Hoel (Director de la Royal School of Librarianship and Information Science, Dinamarca y Secretario del ISO TC/46/SC10 Las normas para papel permanente esta comunicación se presentó al 64th IFLA General Conference. August 16 - August 21, 1998 (no figura en la fuente el nombre del traductor de esta comunicación al español). <http://www.ifla.org/IV/ifla64/115-114s.htm>

por los propios datos, es decir, los materiales digitalizados, como por los metadatos, es decir aquella información que permite la descripción pormenorizada de copias digitales y de sus originales, los requisitos técnicos relacionados con la captura de las imágenes o generación de textos y el almacenamiento de los mismos y sus características, y la descripción de todos y cada uno de los procesos que dan lugar a los objetos digitales definitivos. En esta comunicación nos centraremos fundamentalmente en los objetos digitales obtenidos a partir de materiales textuales o impresos, ya sea en formato imagen o en formato texto. Si ello se lleva a cabo de una forma normalizada, lo que es absolutamente imprescindible para garantizar la interoperabilidad entre sistemas, se habrá podido cumplir, con un esfuerzo muy pequeño, una buena parte de la información que se recoge en el diccionario de datos de PREMIS (PREservation Metadata: Implementation Strategies)⁴ y lo que ello implica⁵.

Metadatos técnicos para imágenes y textos

Estos metadatos se refieren a múltiples aspectos: metadatos descriptivos, metadatos administrativos, relación de ficheros, metadatos estructurales y comportamiento del objeto digital. Las condiciones inherentes a las características físicas y técnicas del objeto digital no es que se pueda, es que se debe generar de forma automática en el momento de su creación. Es en este momento cuando pueden obtenerse datos de forma fidedigna tales como la fecha de creación, el software y hardware necesario para la creación de un objeto digital, el tamaño del fichero, el formato del mismo y su versión, el esquema de compresión, el esquema de codificación de píxeles, el uso previsto de la copia, y otros datos técnicos similares como los incluidos en el esquema *NISO Metadata for images in XML Schema: Technical Metadata for Digital Still Images Standard*⁶. Todas estas características conforman los metadatos técnicos de un objeto digital, cuya forma de expresión más común es como extensión dentro de un fichero METS, en la sección administrativa *amdSec*⁷.

Otro ejemplo de metadatos técnicos específicos para objetos digitales textuales, creados digitalmente u obtenidos de procesos de reconocimiento óptico de caracteres, son los establecidos por el esquema *Technical Metadata for text (textMD)*⁸ que también mantiene la *Library of Congress Network Development and MARC Standards Office* desde 2007 y que ya se encuentra en la versión 2.2. Según este esquema pueden registrarse los siguientes elementos: medida de la calidad de la salida de un proceso OCR, plataforma de hardware sobre la que un documento ha sido producido, incluyendo tipos específicos de ordenadores, escáneres, etc.; el software utilizado para la producción del texto (OCR, editor de textos); las personas que han transcrito un texto desde un medio original a otro (p.e., transcripciones orales); el juego de caracteres, incluyendo su nombre normalizado (p.e., 'ISO_8859-1:1987', 'macintosh'), el orden de bytes, el tamaño del carácter y el mecanismo de ruptura de líneas; el código de lengua según ISO 639-2 y otras normas diferentes; el tipo de fuentes; el metalenguaje del lenguaje de marcas empleado (SGML, XML, ...) y el lenguaje de marcas especificando el correspondiente esquema o DTD; los requisitos especiales de impresión; los requisitos de hardware y/o software para

⁴ Premis data Dictionary for preservation metadata. Version 2.0, March 2008.

<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>.

Véase también la reciente publicación Understanding PREMIS / Authored By: Priscilla Caplan. February 1, 2009 <http://www.loc.gov/standards/premis/understanding-premis.pdf>

En general conviene estar atento al sitio <http://www.loc.gov/standards/premis/>, e incluso suscribirse a PREMIS Implementor's Group Forum (pig@loc.gov), para seguir no sólo las actualizaciones de la normativa, sino las discusiones de los expertos, así como las implementaciones, estudios y casos reales que con mucha frecuencia aparecen en ellas.

⁵ *Exegi monumentum aere perennius (Hor. C. 3.30): preservación a largo plazo de materiales bibliográficos digitalizados* / Xavier Agenjo; Francisca Hernández. – *Boletín de la ANABAD* [en prensa].

⁶ MIX, versión 2.0 <http://www.loc.gov/standards/mix/>

⁷ En la actualidad se dispone de una excelente versión española debida al profesor Ricardo Eito Brun de *METS: introducción y tutorial* http://www.loc.gov/standards/mets/METSOverview_spa.html. Documento en el que se detalla los objetivos de esta norma y se proporciona una metodología para su utilización

⁸ TextMD, versión 2.2 <http://www.loc.gov/standards/textMD/>. El día 25 de noviembre, cuando se consulta por última vez esta página se ha publicado la version 3.0 alpha.

su visualización y cualquier nota sobre aspectos no recogidos en los elementos anteriores. Este esquema puede utilizarse tanto por sí mismo, como en forma de extensión dentro de la sección administrativa del esquema METS, al igual que en el caso mencionado para las imágenes.

Metadatos a partir de los datos

Otra de las ideas que se pretende analizar consiste en la posibilidad de utilizar el proceso de control de calidad de las imágenes escaneadas para reutilizar la información obtenida en el proceso de digitalización y sus correspondientes incidencias para la generación de ficheros MARC 21 de fondos y localizaciones, con el nivel más concreto de descripción al que se puede llegar. En efecto, muchos de los subcampos de este tipo de estructura de información hacen referencia al estado en el que se encuentra el material, que siempre será el del ejemplar, nunca el de la edición. En concreto, algunos de los subcampos del campo 852 (\$q condición física) hacen referencia directa al estado de conservación, mientras que otros se refieren a la integridad del ejemplar (863/865 \$w indicador de rupturas).

Cuando se trata, además, de un tipo de publicación particularmente dañada por el paso del tiempo, el uso reiterado y sobre todo la calidad del papel (como ya se ha señalado antes), sólo la codificación detallada de este formato permitirá hacer referencia a todas y cada una de las páginas de todos y cada uno de los niveles de numeración y cronología (hasta 6) en las que puede describirse una publicación periódica. Por lo tanto, la transferencia de datos desde el sistema de producción de los objetos digitales, durante el escaneo de los mismos, puede asegurar una tarea que completada en el proceso de verificación y control de calidad, pueda recoger las peculiaridades de cada uno de los objetos que forma la colección, es decir, el ejemplar. En efecto, el control de calidad de las imágenes obtenidas en el proceso de digitalización arroja frecuentes incidencias que pueden realimentar la descripción de los ejemplares, pero también la necesaria organización de las imágenes en directorios lógicos (por volúmenes, por números, etc.) llevan a la detección de datos relativos a la descripción de la edición y del ejemplar.

Así, pueden detectarse errores de numeración y datación típicas de la prensa histórica; por ejemplo, diferentes ediciones encuadradas juntas y con un mismo patrón (853/855), o con diferente patrón de numeración y cronología, y, sobre todo, lagunas en la publicación de un determinado título, motivadas por faltas en la colección que posee una biblioteca (863 \$W) o por interrupciones en la publicación (853 \$w). Igualmente, mediante una combinación de los procesos de control de calidad y de estructuración de los ficheros, pueden obtenerse datos tales como la presencia de índices, suplementos, números monográficos, etc. Todo ello puede y debe llevar, mediante la codificación de los resultados de los procesos mencionados, a una mayor riqueza en la descripción bibliográfica. Esto es absolutamente imprescindible en proyectos cooperativos en los que el resultado es la suma de los fondos conservados en las instituciones participantes y en los que la única comprobación física comprehensiva puede realizarse sobre la copia digital.

También muy rico para determinados datos técnicos resulta el campo 007 para los recursos electrónicos y digitales. Hay que tener en cuenta y no olvidar nunca que el resultado de la digitalización de una partitura... es un objeto digital y que el resultado de la digitalización de un mapa es otro objeto digital. Es decir, todos los procesos de digitalización, sea cual sea el material que se digitalice es un objeto digital, simple o complejo. Así, el recurso electrónico 007/00=c (recursos electrónicos) incluye valores fundamentales, como la resolución de la imagen expresada en bits, el formato del fichero, la fuente de la que procede el original, el nivel de compresión, que, en general, se pueden realizar automáticamente mediante las oportunas tablas de variables predefinidas y las características técnicas de la digitalización.

En la sorprendente, por lo abrupta, adopción por la Biblioteca Nacional de España del formato MARC21, que no conversión puesto que la base de datos institucional no se corresponde de forma precisa con la aplicación de este formato, se ha generado un mapeo⁹ considerablemente precisa y

⁹ Todavía deja algunos cabos sueltos que será necesario concretar en el futuro. Sin embargo, www.bne.es/esp/servicios/equivalencias.htm

unificada de estos términos utilizando los definidos en las *Reglas de Catalogación*¹⁰. Es decir, que una buena parte de los outputs producidos durante el proceso de digitalización que se generan desde el inicio de la toma de las imágenes hasta su alocaión en su correspondiente fichero, pueden ir o mejor dicho deberían de ir, tanto a los campos mencionados de los registros bibliográficos o de fondos, como a las correspondientes etiquetas o bien del esquema METS o bien del esquema PREMIS.

No estará de más mencionar que tanto METS como PREMIS prescriben en algunos casos formas diferentes de codificar los mismos hechos, procesos y objetos (por ejemplo, los metadatos técnicos de un objeto digital pueden detallarse siguiendo PREMIS, o siguiendo MIX tanto como una extensión de PREMIS o de METS) y que el resultado final es una combinación de estructuras METS/PREMIS que deben ser fijadas en un perfil específico. La nueva versión de PREMIS 2.0 ha eliminado alguna de las redundancias existentes entre ambas estructuras (METS y PREMIS) y así se ha comentado en un importante artículo reciente¹¹. En este sentido, es especialmente importante el elemento *source* de la sección *admSec* de METS para establecer las relaciones correspondientes con la descripción del original. Igualmente, PREMIS incluye un apartado para el establecimiento de relaciones complejas entre un objeto digital y la fuente, y entre diferentes versiones digitales. Por ejemplo, el perfil *ECHO Dep Generic METS Profile for Preservation and Digital Repository Interoperability*¹² elige esta segunda posibilidad por la capacidad de PREMIS para este tipo de relaciones.

Esbozo de metodología

A la vista de lo que antecede quizá no sea necesario insistir en la obligación de que el objeto digital esté perfectamente descrito desde el punto de vista bibliográfico para favorecer posteriormente su recuperación automatizada, sino también en que uno de los vínculos de unión entre un original y su copia digital se realiza a través de la descripción bibliográfica, debiendo permanecer esta a lo largo del tiempo de una forma lo más correcta posible para garantizar que la copia digital no surge por sí misma, sino como copia de un determinado original, ya sea como imágenes solo, como texto o como imágenes y texto. Hay que hacer especial hincapié en la importancia de contar con puntos de acceso normalizados, lo que implica la necesidad de generar registros de autoridades en MARC 21 o siguiendo el esquema MADS¹³. Es decir, si se considera desde un punto de vista amplio el proceso de creación de un objeto digital como fruto de la captura de las imágenes y la posterior creación de un fichero, que no sólo pueda almacenarse y transmitirse a través de la web, sino que, además, contenga el germen ya muy desarrollado de lo que constituirá el correspondiente registro de preservación, formado por todos los metadatos administrativos referidos a ellos, se puede perfectamente superar la dialéctica que se establece entre METS y PREMIS e incluir así mismo, los aspectos descriptivos o catalográficos que ya existieron en el inicio de ese proceso biblioteconómico.

Es decir, que si en un primer momento la descripción bibliográfica y la catalogación tenía por objeto tanto la identificación del material bibliográfico, es posible que debamos definir un nuevo paradigma en el cual la creación de un nuevo registro bibliográfico, en este caso para un objeto digital, contenga así mismo -y además de la identificación y localización del objeto-, las características necesarias para su preservación a largo plazo, codificables posteriormente según PREMIS. Lógicamente, no se trata tanto de atribuir unas nuevas tareas a este proceso bibliográfico, sino en procurar que en la medida de lo posible éstos se retroalimenten unos de otros, de forma tal que exista

¹⁰ *Reglas de catalogación*. – 1ª ed., 6ª reimp. – Madrid : Dirección General del Libro, Archivos y Bibliotecas, 2007. – 630 p.

Aunque en esta 6ª impresión el texto no presenta variaciones, resultan especialmente interesantes los prólogos de Rogelio Blanco (p. VI), Director General del Libro, Archivos y Bibliotecas y de Rosa Regás (p.IX), Directora de la Biblioteca Nacional, por lo que declaran de perspectivas de futuro.

¹¹ Guenther, Rebecca S.: Battle of the Buzzwords. Flexibility vs. Interoperability When Implementing PREMIS in METS. *D-Lib Magazine*. July/August 2008 Volume 14 Number 7/8 .ISSN 1082-9873. <http://www.dlib.org/dlib/july08/guenther/07guenther.html>

¹² <http://www.loc.gov/standards/mets/profiles/00000015.html>

¹³ Metadata Authority Description Schema. <http://www.loc.gov/standards/mads/>

cierto tipo de unicidad en este proceso funcional. Catalogación, localización y preservación pueden ser distintos lados de un polígono que probablemente deberá constituir el objeto, en este caso virtual, con el que habrá de entenderse tanto el bibliotecario como el usuario (y los sistemas automatizados) en un futuro próximo.

Un nuevo panorama

En esta línea, las FRBR¹⁴ y las RDA¹⁵ dedican una considerable atención, dentro de una perspectiva de programación orientada a objetos (OOP, Object Oriented Programming) desde un punto de vista informático, a las características del ítem, del cual se hace no sólo una identificación estática sino una traslación de sus atributos a través del tiempo, entre los cuales no serán los menos importantes aquellos que hagan referencia a su perdurabilidad. Aunque este enfoque está dirigido como ya se ha señalado desde un primer momento a los objetos digitales creados a partir de un ejemplar producido de manera tradicional, técnicas similares pueden utilizarse, incluyendo los metadatos PREMIS, para objetos nacidos ya digitalmente, como se señala en el trabajo de Angela Dappert y Markus Enders, *Using METS, PREMIS and MODS for Archiving eJournals*¹⁶.

Siguiendo el razonamiento, se plantea además la posibilidad de elaborar encabezamientos multilingües, utilizando las posibilidades con las que cuenta el formato MARC21 para establecer puntos de acceso equivalentes en otras lenguas. De esta manera, el objeto digital contará así mismo con puntos de acceso en otras lenguas y por lo tanto el acceso multilingüe se aplicará tanto al registro bibliográfico como al objeto digital. Volvemos a recordar en este punto que muchos de los proyectos de digitalización son proyectos cooperativos, del que es muestra Europea¹⁷. Así mismo, podrá contener si se considera oportuno los esquemas de metadatos correspondientes a las estructuras de intercambio basadas en Dublin Core gracias a la facilidad con la que después de una programación eficaz es posible obtener dinámicamente los esquemas de metadatos según Dublin Core ISO 15.836:2003, para alimentar, tanto repositorios OAI-PMH, como servidores SRU¹⁸.

Empaquetamiento de metadatos

Este amplio conjunto de informaciones implica así mismo contar con un contenedor de metadatos que agrupe todos los tipos de metadatos que hemos ido mencionando y por eso se insiste en la necesidad del uso del esquema METS, que permite estructurar y utilizar etiquetas para objetos digitales multipágina, y que agrupará tanto la descripción bibliográfica en formato MARC XML (incluyendo los encabezamientos multilingües) como los metadatos administrativos, la relación de ficheros, el mapa estructural de los mismos y los metadatos de preservación según el esquema PREMIS¹⁹.

¹⁴ www.ifla.org/VII/s13/FRBR/index.htm

Además de la lista de discusión, resulta interesante syndicar The FRBR Blog <http://www.frbr.org/feed>

¹⁵ <http://www.rdaonline.org/> dónde se puede consultar el extensísimo borrador final [RDA Constituency Review Draft] y además se anuncia para principios de febrero de 2009 la posibilidad de descargar una pequeña aplicación de demostración de las funcionalidades de las RDA, que ya fue presentada en el IFLA Satellite Meeting on RDA celebrado el 8 de agosto de 2008 en Quebec, y que presumiblemente incorporará las últimas modificaciones que ha experimentado la norma.

¹⁶ Dappert, Angela & Enders, Markus: *Using METS, PREMIS and MODS for Archiving eJournals*. D-Lib Magazine. September/October 2008 Volume 14 Number 9/10 ISSN 1082-9873. <http://www.dlib.org/dlib/september08/dappert09dappert.html>

¹⁷ <http://www.europeana.eu/portal/> [consultado el 15 de febrero de 2009]

¹⁸ Para tener una idea clara de cuáles son las perspectivas de futuro de esta normativa, conviene consultar el artículo Search Web Services - The OASIS SWS Technical Committee Work : The Abstract Protocol Definition, OpenSearch Binding, and SRU/CQL 2.0 / . D-Lib Magazine. January February 2009. <http://www.dlib.org/dlib/january09/denenberg01denenberg.html>

¹⁹ Es interesante seguir la evolución e implantación de la norma ISO 25577:2008 MarcXchange, que pretende dar una estructura común en XML a los registros bibliográficos, autoridades, fondos, clasificación e información a la comunidad en ISO 2709. También se precisa su utilización en METS, SRU y OAI.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43005

Entre paréntesis puede mencionarse la nueva funcionalidad con la que cuenta la ISO y que permite utilizar RSS para estar atentos a la evolución de la normativa. De hecho, existen agregadores para cada norma o para los distintos

Hay que tomar en consideración también la importancia que tiene la posibilidad de empaquetar conjuntamente dentro del esquema METS, tanto lo que se refiere a los derechos de autor que pudiera tener el objeto digital (METSRights), como poder encapsular el esquema PREMIS, cuyo objetivo es fijar los eventos y las características técnicas de los mismos, mediante los cuales se produjo el proceso digitalización y la creación del objeto digital mismo, con vistas a su futura ingesta en un repositorio ISO 14.721:2003.

En este sentido, resulta sumamente importante la aplicación de JHOVE (*JSTOR/Harvard Object Validation Environment*)²⁰ puesto que no sólo, y como su nombre indica, valida las complejas estructuras que pueden encapsularse, como se ha venido diciendo, en METS sino que son capaces de generar a su vez una estructura MIX que es justamente la que se recomienda para llevar a cabo la ingesta de los objetos digitales en repositorios OAIS. La nueva versión de JHOVE denominada *JHOVE 2: A Next-Generation Architecture for Format-Aware Digital Preservation Processing*²¹, se encuentra actualmente en fase de discusión, y según se anuncia mejorará considerablemente estas funcionalidades.

Otro aspecto crítico, por su coste en tiempo y dinero, como es la generación de ficheros o esquemas de metadatos tras el reconocimiento óptico de caracteres (OCR) y la necesidad de normalizar este proceso mediante la combinación METS/ALTO²². En este caso, el mapa estructural de los ficheros METS contendrá por cada página del original la relación de ficheros de imagen (con identificación de su versión de uso) y su correspondiente fichero texto obtenido del proceso de OCR. Para estos ficheros texto el esquema ALTO (Analyzed Layout and Text Object) permite fijar de manera normalizada las coordenadas de cada uno de las unidades textuales (párrafos, palabras o caracteres) representados dentro del fichero, según su referencia a la posición de los píxeles en los ficheros de imagen. Así, cada carácter tiene una posición horizontal y una posición vertical, un ancho y una longitud que se refieren a la posición de los píxeles en la imagen. Como ya se ha señalado antes, textMD permite proporcionar información sobre las características técnicas de ficheros texto, como son los ficheros ALTO, y en consecuencia puede usarse dentro del elemento PREMIS <additionalTechnicalCharacteristics>

La necesidad de gestionar todos los procesos relacionados con la creación de objetos digitales y su preservación favorecerá en un futuro próximo, el intercambio de información según METS y el uso de programas de análisis y consulta de una información procesada y representada de forma análoga, buscando siempre la integración de procesos y la normalización de los mismos, tanto en sus datos, como en sus metadatos.

comités, tanto técnicos como científicos. Así es perfectamente posible seguir la evolución de la normativa del TC 46 Information and Documentation y del SC4 Technical Interoperability.

Lo mismo puede decirse de la norma ISO 20775:2009 *Information and documentation — Schema for holdings information* que especifica un esquema que cubre los fondos y localizaciones de todos los tipos de recursos, físicos y electrónicos, como texto impreso, imágenes, grabaciones sonoras, videos, electrónicos, etc., editados como publicaciones monográficas o seriadas. Está diseñada para ser utilizada como esquema para preguntas y respuestas en sistemas de búsqueda y recuperación tales como SRU, Z39.50 or Open Search.

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39735

²⁰ <http://hul.harvard.edu/jhove/>

²¹ <http://hul.harvard.edu/jhove/JHOVE2-proposal.doc>

²² En *The Digital Blog* se hacen varias referencias a importantes proyectos que se han ajustado a la combinación METS/ALTO (<http://www.godigitalblog.com/tag/mets-alto/>), así como al proyecto europeo METAe (<http://metae.uibk.ac.at>) que fue el origen de este esquema.

El esquema ALTO (Analyzed Layout and Text Object) puede consultarse en su ubicación original <http://www.ccs-gmbh.com/alto> y en http://www.loc.gov/ndnp/alto_1-1-041.xsd

METS/ALTO se está utilizando en numerosos proyectos españoles como la Biblioteca Virtual del Prensa Histórica (<http://prensahistorica.mcu.es>), la Biblioteca Digital de Castilla-La Mancha (<http://bidicam.jccm.es>), y Galiciana, la Biblioteca Digital de Galicia (<http://galiciana.bngalica.org>), por citar solo algunos de los más destacados.

Entre los proyectos internacionales puede mencionarse, el NDNP (National Digital Newspaper Project) de la Library of Congress (<http://www.loc.gov/ndnp/>) el Databank of Digital Daily newspapers de la Koninklijke Bibliotheek (<http://www.kb.nl/hrd/digi/ddd/index-en.html>) o el Australian Newspapers Digitisation Program (http://www.nla.gov.au/ndp/project_details/).

Se recomienda, pues, integrar en el aplicativo todos los procesos citados. El gran nivel de detalle de la catalogación de impresos y ejemplares puede conseguirse a partir del escaneo al nivel de página, y la gran riqueza de generación de diferentes esquemas de metadatos, puede llevarse a cabo por medio de un sistema de información integrado que considere a los objetos multimedia una parte más del aplicativo. Así, desde el punto de vista de la integración de los procesos, podrán eliminarse las duplicaciones de esfuerzos para realizar generaciones paralelas o consecutivas de metadatos.