

El uso de metadatos para la gestión del patrimonio digital mundial en la Sociedad de la Información

LUIS FERNANDO RAMOS SIMÓN

SILVIA COBO SERRANO

Universidad Complutense de Madrid

INTRODUCCIÓN

Los metadatos son elementos de información estructurados, embebidos en la fuente del objeto digital al que hacen referencia para otorgarle un significado, un contexto y una organización (Muñoz de Solano, 2014, p. 154). Dicho de una forma más sencilla, son datos que describen las propiedades de otros datos para facilitar su gestión y uso al vincular la información descriptiva con el objeto digital. En resumen, para un bibliotecario-documentalista, como dice Méndez Rodríguez (2002, p. 35), “los metadatos son un tipo de datos que usamos para describir el contenido, la estructura, la representación y el contexto de algún conjunto de datos específicos”.

Los metadatos comenzaron a utilizarse en 1968 en el ámbito de la informática. Más tarde, en los años setenta, se hicieron populares en Europa y Estados Unidos, donde su uso se vinculó a documentos militares. Como es bien sabido, “metadata” es una palabra de origen griego que significa “datos sobre los datos” y hoy en día se emplean en todo tipo de sectores. La información sobre metadatos define los requisitos necesarios sobre los datos en cuanto a su estructura, contenido y precisión para incorporarlos a una base

de datos. Estas descripciones constituyen un modelo en el que también se incorporan protecciones para evitar el mal uso de los datos en el *dataset*. En algunos sectores, se emplean metadatos estandarizados, como sucede en el ámbito de la información geográfica (GIS). Estos estándares pueden, por un lado, estar reconocidos por organismos internacionales de estandarización y ser un estándar dominante por su aplicación generalizada en un grupo de organizaciones o, por otro lado, solamente ser usados por algunas empresas del sector privado. Los elementos de los metadatos pueden ser separados en tres tipos: diccionario de definición de metadatos, directorio de metadatos y, por último, contenido de metadatos, también llamado metadatos de metadatos (Aalders, 2005).

Se puede decir que todos los campos científicos organizan sus servicios de información electrónica incorporando metadatos. Sin duda, los metadatos son tributarios de la larga tradición bibliotecaria de los catálogos, adaptando ese amplio saber y experiencia a las nuevas formas digitales de organización de la información. De hecho, los metadatos descriptivos más conocidos proceden del mundo bibliotecario (por ejemplo, Dublin Core, desde 1995). Los metadatos proporcionan de una forma sencilla la información más relevante sobre el contenido en donde se representan o apuntan, de tal modo que indican a las máquinas y las personas sus características y aspectos sobre lo que son, no son y lo que se puede encontrar en ellos. De ahí, la importancia de los buenos metadatos para recuperar de la información, así como para facilitar su descubrimiento y hacer visible el objeto y la colección (Dangerfield *et al.*, 2015, p. 19).

Tradicionalmente, los estudiosos de esta materia establecen varias clasificaciones de los metadatos (descriptivos, administrativos, técnicos, de conservación, funcionales, explícitos...), si bien los metadatos descriptivos y administrativos son los más comunes. Los primeros (descriptivos) se emplean para la representación e identificación de los recursos para organizar los objetos digitales (catálogos, índices...), mientras que los segundos (administrativos) se orientan hacia la gestión de los recursos en cuanto a su ubicación, uso, versiones, etcétera.

Por tanto, se pretende llamar la atención sobre la creciente importancia de los metadatos en el ámbito del patrimonio digital y su funcionalidad en la gestión de los recursos culturales.

LOS METADATOS EN EL PATRIMONIO DIGITAL

Al igual que los datos, los metadatos han alcanzado en los últimos años una notable importancia económica, como se ha mencionado. Aunque existen metadatos en todo tipo de productos y sectores, las bibliotecas y las editoriales tienen mucha experiencia en su gestión.

Algunos datos de esta realidad, referentes al sector académico, manifiestan que hay unos cincuenta millones de libros y revistas de los que podrían ser extraídas referencias bibliográficas. A pesar de lo que parece, el sector está muy atomizado y las fuentes de datos muy dispersas, por lo que hay varios intentos para tratar de crear repositorios abiertos que permitan una fácil gestión de tales referencias (Open Citation Corpus, por ejemplo).

En Wikipedia, aparecen referencias de más de 125 bases de datos que gestionan metadatos bibliográficos. Las dos principales fuentes, ambas accesibles por suscripción, son Thomson Reuters Web of Science y Scopus (Elsevier). La tercera, Google Scholar, es de libre acceso y pertenece a un operador comercial. Microsoft también intentó operar Microsoft Academic Search en 2009, pero el servicio desapareció definitivamente en 2016.

En el ámbito del acceso abierto, los principales servicios que dan acceso a referencias son las siguientes:

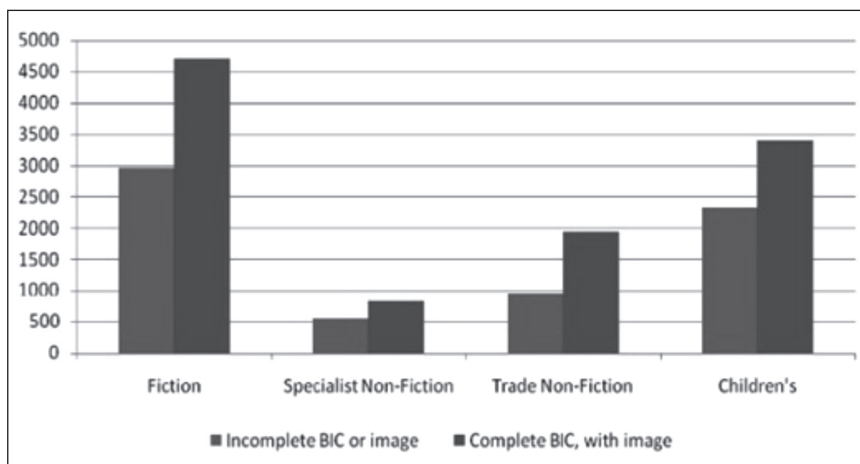
- Open Access Subset of PubMed Central, que reúne 468.805 artículos. En su conjunto, el servicio da acceso a las referencias de unos 2,1 millones de artículos.
- Preprints, en el repositorio Arxiv, que se estiman en más de 1,1 millones.
- CiteSeerX con 1,2 millones de artículos, principalmente del área de Informática.
- CitEc (Citation in Economics) contiene 545.641 documentos.

En conjunto, Shotton (2013) calcula que las referencias en acceso abierto están en torno al 4 por ciento. La fuerte presencia de las empresas comerciales y la resistencia para facilitar la puesta de datos en común dan una idea del interés económico por estos metadatos.

Para poner de relieve la importancia comercial de los metadatos, son más contundentes aún los resultados del informe elaborado para la firma Nielsen por Breedt y Walter (2012), en el que se pone de manifiesto la estrecha vinculación entre las ventas comerciales y los metadatos en el sector del libro.

La tendencia de este informe apunta que la relación entre la calidad de los metadatos y las ventas de un libro es cada vez mayor. En concreto, el informe comparó la venta de libros en el mercado inglés en 2011 según el tipo de metadatos distribuidos en el proceso de comercialización y venta: En un caso, los datos básicos del libro (autor, título, editorial, etc.); en otro, aquellos libros en los que se mejoraron los metadatos (imagen de portada, descripciones larga y corta, biografía del autor y reseña de la obra).

Figura 1. Comparación de las ventas de libros en relación a los metadatos según la temática.



Fuente: Breedt, André, y David Walter (2012, p.7).

Las principales conclusiones del análisis son:

- Las ventas con buenos metadatos y la imagen del libro son siempre muy superiores a los que tienen los datos incompletos o que carecen de imagen.
- La diferencia es más notoria cuando las ventas son en línea y no en la librería. Asimismo, cuanto mayor es el número de metadatos mejorados, se produce un mayor aumento de las ventas en línea.
- Las diferencias en las ventas son mucho mayores en los libros de ficción. En los libros especializados, las diferencias son menores. Las ventas son un 140 por ciento mayores cuando hay varios metadatos mejorados.
- Se observan diferencias claras cuando se analizan los libros con más metadatos, en donde el incremento de los mismos conlleva a un incremento en las ventas.
- El metadato más influyente es la imagen de portada. Los títulos con imagen tuvieron ventas superiores en un 268 por ciento que aquellos sin imagen.
- Se venden casi siete veces más los libros que tienen metadatos estándar, imagen y metadatos mejorados que los que no completan los metadatos estándar, no tienen imagen y no usan metadatos mejorados.

Por otro lado, resulta necesario definir el concepto de patrimonio cultural en el ámbito digital. En este sentido, entendemos el nuevo patrimonio documental digitalizado del que son titulares las bibliotecas, los archivos, los museos y todas las tecas, con mención especial por su importancia a las filmotecas y fototecas o colecciones fotográficas. Este conjunto documental es un patrimonio creciente, cuya presencia en Internet ya es visible, aunque sólo se haya digitalizado y puesto en línea una pequeña parte del total de documentos que conservan estas instituciones.

La carta sobre la preservación del patrimonio digital (Unesco, 2003), en su artículo primero, considera patrimonio digital a los “recursos de carácter cultural, educativo, científico o administrativo

e información técnica, jurídica, médica y de otras clases, que se generan directamente en formato digital o se convierten a éste a partir de material analógico ya existente”.

El objetivo señalado por la carta en la conservación de este patrimonio es el de hacerlo accesible al público, así como elaborar estrategias encaminadas a su preservación, si bien distingue entre los elementos en dominio público —cuyo acceso no debería estar sujeto a requisitos poco razonables— y el patrimonio privado para el que propugna un equilibrio entre los derechos de sus titulares y el interés del público por tener acceso a dichos elementos. En el aspecto institucional, la carta otorga un papel determinante en esa función a las bibliotecas, archivos y museos (filmotecas y tecas en general) para llevar a cabo esas tareas en colaboración con el sector privado (creadores, editores, productores, distribuidores, fabricantes de equipos y software...).

Los metadatos —que, como se ha mostrado, son claves en el sector cultural comercial— también son esenciales para construir el patrimonio cultural digital. Así se pone de manifiesto en algunos informes, en los que se destaca el papel de los metadatos como uno de los activos más valiosos para la reutilización dentro del sector cultural, junto con los archivos, las obras en dominio público inéditas y las fotografías (Davies, 2009b).

Para lograr un patrimonio digital de alcance mundial, es preciso que los metadatos sean libres, como lo han entendido European (más de 53 millones de objetos digitales) y la Digital Public Library of America (DPLA) (13,3 millones de objetos en mayo de 2016), probablemente las dos iniciativas más importantes a nivel mundial en este ámbito.

La discusión sobre el libre intercambio de metadatos está justificada por el alto coste que conllevan la recopilación y organización de los metadatos en las grandes instituciones culturales. Los pocos datos que se conocen ponen de relieve que los ingresos por este concepto se cuentan entre los más importantes obtenidos por las instituciones culturales en países como Reino Unido o Alemania (Davies, 2009a).

FUNCIONALIDAD DE LOS METADATOS EN LA GESTIÓN DE LOS RECURSOS CULTURALES

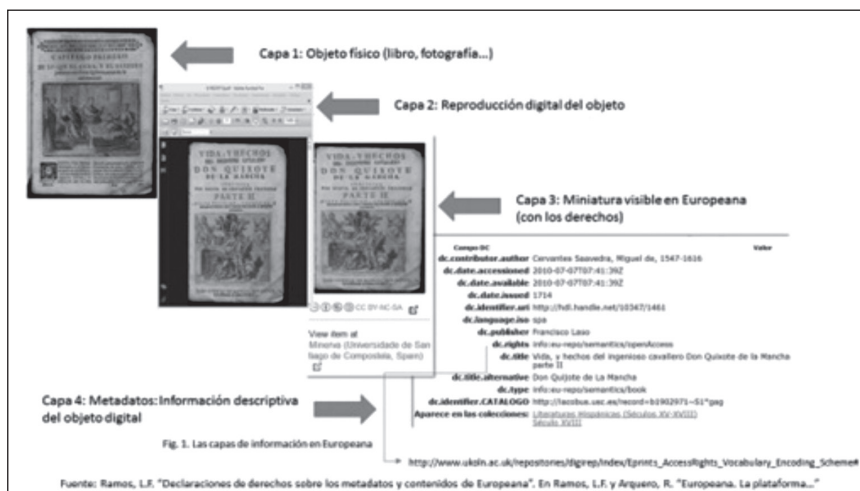
Es importante distinguir entre el acceso abierto a los metadatos que gestionan un recurso digital del acceso y el contenido mismo. Ambos aspectos se indican en los metadatos, por eso a veces se crea confusión. Por ejemplo, cuando de una base de datos que contiene metadatos se dice que es de libre acceso o abierta, parece indicarse que tanto el acceso a la base de datos como al contenido son libres, cuando en absoluto es así.

Podemos observar la importancia de los metadatos en la construcción del patrimonio digital al analizar cómo cumplen su papel en Europeana y en la DPLA. Estas plataformas en esencia sólo mantienen una base de datos compuesta por la información catalográfica y una imagen digital en miniatura de los objetos de la colección, en forma de capas. A partir de este esquema, la búsqueda del usuario conduce al enlace con el objeto digital.

La complejidad de estos procesos técnicos de gestión de los metadatos es notable. Desde la creación de Europeana y la implantación del modelo de datos EDM (European Data Model), el enriquecimiento de los metadatos ha continuado siendo una de las principales vías para mejorar la calidad de los datos, el contexto de los objetos digitales y la información multilingüe. Estas actividades en la actualidad se llevan a cabo a través de los distintos grupos de trabajo para mejorar los distintos procesos de enriquecimiento de los datos y hacer más transparentes los intercambios de información (Europeana, 2013 y 2015; Dangerfield *et al.*, 2015).

El símil de las capas, utilizado por Keller (2011), nos permite mostrar de una forma sencilla los distintos niveles de abstracción que van desde el objeto material (por ejemplo, un cuadro de pintura), su imagen digital y después la representación miniaturizada, acompañada de los datos descriptivos del objeto (por ejemplo, título, autor, museo, titular de los derechos, etc.). Europeana organiza este esquema en cuatro capas, tal como se muestra en la figura 2.

Figura 2. Capas de los objetos digitales en las instituciones culturales.



- 1^a Capa de objetos físicos: es el objeto, como un libro, una pintura, escultura u otro elemento del patrimonio cultural o científico.
- 2^a Capa de reproducciones digitales: corresponde al objeto anterior digitalizado por cualquier procedimiento, ya sea escaneado, fotografía o cualquier otra forma de archivo digital. Si el objeto tiene origen digital, por ejemplo, una imagen o un libro digital, la primera y la segunda capa coinciden.
- 3^a Capa de miniaturas (*previews*, vistas previas o miniaturas): es una visualización en baja resolución que permite una rápida identificación de las características generales del objeto, ya sea una pequeña imagen digital, una secuencia de imagen o video, o una muestra de cualquier cualidad de la imagen digitalizada. Como hemos visto para el caso de los libros comerciales, es un elemento atractivo para captar el interés de los usuarios hacia el contenido y actúa a modo de metadato complementario. Es un archivo de imagen, generalmente en formato jpg o la portada de un archivo pdf de pequeño tamaño, menos de 200 píxeles. Los proveedores

de los objetos digitales deben proporcionar un *link* a una imagen de mayor tamaño y calidad que la miniatura misma. Cuando no se puede usar una miniatura adecuada al objeto representado, entonces se debe sustituir por una imagen genérica que representa uno de los formatos comunes: texto, imagen, vídeo, sonido y 3D.

- 4^a Capa de metadatos: contiene la información sobre los objetos mediante una descripción catalográfica del objeto físico y digital de las capas anteriores.

Esta capa de información de metadatos es de libre acceso y uso para todos los usuarios, lo que significa que el proveedor de los metadatos renuncia a los derechos sobre los mismos mediante la iniciativa de licencia CC0. Esta licencia, impulsada por la organización Creative Commons, tanto en estas plataformas de recursos culturales como en otros entornos de datos abiertos, ha favorecido a nivel internacional la apertura, la interoperabilidad y los estándares comunes, lo que está cambiando la forma de ver el sector cultural y el uso del contenido digital en línea. De hecho, la DPLA —abierta en abril de 2013— está basada en el European Data Model (Modelo de Datos de Europeana) y opera con el mismo esquema de derechos en cuanto a los metadatos, “lo que supone que los dos masivos conjuntos de datos de ambos lados del Atlántico sean interoperables” (Cousins y Daley, 2013, p. 35). En efecto, a los más de cincuenta millones de objetos digitales, hay que sumar los más de siete millones que ya ofrece el portal estadounidense. De hecho, es posible la búsqueda conjunta sobre ambas plataformas. La búsqueda está disponible en: <http://www.digibis.com/dpla-europeana>.

Entre los objetivos básicos de Europeana está, por un lado, que los metadatos financiados con fondos públicos estén ampliamente disponibles para todos y, por otro, que las instituciones proveedoras de contenido no se introduzcan nuevos derechos después de la digitalización de objetos que están en dominio público, en línea con las recomendaciones del Comité de sabios (Comisión Europea, 2011).

CONCLUSIONES

Tras el estudio de la temática, las principales conclusiones que se pueden extraer son las siguientes:

- Los metadatos identifican, orientan y condicionan los contenidos.
- Los metadatos son decisivos para asegurar la colaboración a largo plazo entre las instituciones al facilitar el acceso, la recuperación y organización de los contenidos.
- En el ámbito del patrimonio digital, los metadatos tienen una notable importancia económica en los sectores comerciales, de ahí su atomización y la dificultad de ponerlos en acceso abierto. Por ejemplo, en el sector del libro está demostrada la eficacia de los metadatos mejorados para aumentar las ventas.
- El papel clave de los metadatos en el sector cultural-comercial demuestra su importancia para construir el patrimonio cultural digital, de ahí la importancia de favorecer la reutilización libre de estos metadatos en la economía digital, a pesar del alto coste que supone su recuperación y organización en las grandes instituciones culturales.
- Las grandes plataformas de acceso en línea al patrimonio cultural, como Europeana y DPLA, disponen una capa de metadatos que contiene una descripción catalográfica del objeto físico y digital de las capas anteriores. Estos datos son de libre acceso para todos a través de la licencia libre CC0. Esta modalidad, impulsada por Creative Commons, favorece la apertura de datos y unos estándares comunes.
- Como conclusión general, los metadatos deben ser libres y contener orientaciones sencillas sobre los derechos de los contenidos.

REFERENCIAS BIBLIOGRÁFICAS

- Aalders, H. (2005). An Introduction to Metadata for Geographic Information. En H. Moellering, H. Aalders y A. Crane. *World Spatial Metadata Standards*. Elsevier.
- Breedt, André, y D. Walter (2012). *White Paper: The Link Between Metadata and Sales*. Disponible el 24 de julio de 2016 en: http://www.nielsenbookdata.co.uk/uploads/3971_Nielsen_Metadata_white_paper_A4%281%29.pdf.
- Comisión Europea (2011). *The New Renaissance: Report of the 'Comité des Sages': Reflection Group on Bringing Europe's Cultural Heritage Online*. Disponible el 24 de julio de 2016 en: www.ace-film.eu/wpcontent/uploads/2011/02/CdS_Final_report.pdf.
- Cousins, J. y B. Daley (2013). Europeana: el patrimonio cultural de Europa disponible en línea para todo el mundo, *Boletín de Anabad*, 63, 3.
- Dangerfield, M. C. et al. (2015). *Report and Recommendations from the Task Force on Metadata Quality*. Disponible el 24 de julio de 2016 en: http://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf.
- Davies, Rob (2009a). *PSI in the Cultural Sector*. European PSI Platform. Topic Report No: 4
- . (2009b). *Economic and Social Impact of the Public Domain: EU Cultural Institutions and the PSI Directive*. Informe para la Comisión Europea. Rightscom/Comisión Europea.
- Europeana (2015). *Report on Enrichment and Evaluation*. Disponible el 24 de julio de 2016 en: http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf.

- Europeana (2013). *Europeana Data Model Primer*. Disponible el 24 de julio de 2016 en: http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.
- Keller, P. (2011). *The European Licensing Framework, Next Steps*. Disponible el 24 de julio de 2016 en: <http://www.bnl.public.lu/fr/actualites/communiques/2011/11/europeana/keller--2011-11-28.pdf>.
- Méndez Rodríguez, E. (2002). *Metadatos y recuperación de información. Estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón: Trea.
- Muñoz de Solano, B. (2014). Metadatos en Europeana. En L.F. Ramos Simón y R. Arquero Avilés. *Europeana. La plataforma del patrimonio cultural europeo*, p. 153-176.
- Shotton, D. (2013). Open citations. *Nature*, 502, 295:7.
- UNESCO (2003). *Carta sobre la preservación del patrimonio digital*. Disponible el 24 de julio de 2016 en http://portal.unesco.org/es/ev.php-URL_ID=17721&URL_DO=DO_TO PIC&URL_SECTION=201.html.