

INTRODUCCIÓN

Los modelos de datos en la perspectiva de Europeana, la DPLA y RDA

XAVIER AGENJO

Fundación Ignacio Larramendi
xavier.agenjo@larramendi.es

FRANCISCA HERNÁNDEZ

DIGIBÍS: Producciones digitales
francisca.hernandez@digibis.com

RESUMEN: Se describen los distintos modelos de datos que han ido respaldando la creación y desarrollo de Europeana y la influencia que ésta ejerce en la normalización e interoperabilidad de los archivos, bibliotecas y museos europeos. Se analiza Europeana Data Model como una norma de facto para archivos, bibliotecas y museos en el marco de Linked Open Data, especialmente a partir de su adopción por la Digital Public Library of America. Se examinan las ventajas de integrar en un mismo proceso descriptivo los modelos de datos propios del ámbito bibliotecario, RDA y su codificación MARC 21, con las diferentes clases y propiedades de Europeana Data Model, y la conveniencia de que el software, comercial o abierto, de archivos, bibliotecas y museos responda adecuadamente para mantener un crecimiento sostenido de las colecciones digitales.

PALABRAS CLAVE: Europeana, Modelo de Datos de Europeana, Datos Abiertos Vinculados, Modelos de datos de archivos, bibliotecas y museos, Recursos Descripción y Acceso, Bibliotecas digitales.

ABSTRACT:

The paper describes the different data models that have supported the creation and development of Europeana and its influence on standardization and interoperability of archives, libraries and museums in Europe. *Europeana Data Model* is analyzed as a *de facto* standard for archives, libraries and museums in the

context of Linked Open Data, especially since its adoption by the *Digital Public Library of America*. It discusses the advantages of integrating, in the same descriptive process, models specific to library data field, RDA and MARC 21 encoding, with the different classes and properties of *Europeana Data Model*, and the convenience that software, either commercial or open source, for archives, libraries and museums respond properly in order to maintain a sustained growth of digital collections.

KEY WORDS: Europeana, Europeana Data Model, Linked Open Data, Data models for Archives, libraries and museums, Resource Description and Access, Digital libraries.

INTRODUCCIÓN

Con toda seguridad, una de las combinaciones más productivas en las últimas décadas ha sido el modelo de datos Dublin Core (DC) y el protocolo de recolección de metadatos OAI-PMH¹. Tres ejemplos serán suficientes. En junio de 2002 *OAIster* abrió sus puertas con 274.062 registros que ya en 2006 se habían transformado en 8.857.208 de registros de 680 instituciones; y en 2009 en 20.220.634 de registros de 1.082 proveedores. Un crecimiento anual de casi 2,9 millones de registros². En octubre de 2013 OCLC informa de que *OAIster* tiene más de 30 millones de registros de más de 1.500 instituciones.

En marzo de 2006, el recolector *Hispana*³ inició su andadura con 119.014 registros de objetos digitales procedentes de 25 repositorios OAI-PMH; justamente 7 años más tarde, *Hispana* ofrece 4.700.082 objetos digitales de 202 repositorios. *Hispana* es también un directorio de colecciones digitales, que pueden disponer o no de repositorios OAI. En 2006 el número de colecciones era de 55 (el 45,45 % eran repositorios OAI), mientras que en 2013 es de 602 (el 33,55 % son repositorios OAI). Estas cifras arrojan un crecimiento anual de más de 670.000 registros. En noviembre de 2008 se inauguró el portal de la Biblioteca Digital Europea denominado *Europeana*⁴ con 2 millones de objetos digitales. En octubre de 2013 ofrece ya 29.637.274 de objetos, procedentes de 2.200 instituciones de 36 países. Un crecimiento anual de más de 5,9 millones de registros.

Estas cifras muestran con claridad cómo la combinación OAI-PMH y DC ha hecho aflorar millones de objetos digitales en una década. Y desde luego la implantación del dúo OAI+DC ha producido considerables efectos tanto en los

¹ <http://www.openarchives.org/>

² Datos obtenidos a través de Wayback Machine a partir de las páginas archivadas de *OAIster*, <http://web.archive.org/web/20030217130938/http://www.oaister.org/o/oaister/reports.html> y <http://web.archive.org/web/20090302120345/http://www.oaister.org/>

³ <http://hispana.mcu.es>

⁴ <http://web.archive.org/web/20081111134916/http://dev.europeana.eu/>

proveedores de datos como en los servicios de recolección que analizaremos a continuación. Nunca estaremos suficientemente agradecidos al trabajo de Herbert Van de Sompel⁵, uno de los creadores del protocolo OAI, o al de Jeff Young de la OCLC⁶, que desarrolló OAICat, software open source para la implantación de repositorios OAI.

LA FASE RECOLECTORA

Parafraseando las fases de evolución de las sociedades humanas, podemos hablar de una primera fase de aplicación de OAI+DC basada en la recolección de registros con el esquema DC básico de 15 elementos⁷. En esta primera fase se ha realizado un gran esfuerzo por implantar el protocolo y exponer los repositorios creados de este modo. Así dicho pudiera parecer poco, pero en la realidad ha supuesto que trabajen en una dirección común archivos, bibliotecas y museos. Ésta es una de las mayores ventajas de DC: su capacidad para reunir metadatos de cualquier tipo de recurso, sea un objeto físico o electrónico, de archivos, bibliotecas o de museos. Seguramente, la construcción de Europeana no habría sido posible, por la dificultad de su gestión, si no hubiera sido gracias a esta sinergia. La misma posibilidad de reunir registros de archivos, bibliotecas y museos ha dado lugar a que los propios proveedores consideren sus datos de una forma más abierta. Formar parte de Europeana es bastante más que una cuestión de prestigio institucional; constituye el procedimiento para acrecentar la potencia informativa de los datos por el *simple* hecho de conformar un punto único de consulta.

Hay que tener en cuenta que mediante OAI+DC se pueden agregar no sólo objetos digitales correspondientes a la digitalización de los objetos culturales patrimoniales, sino también las publicaciones científicas que estudian esos objetos. Los repositorios de publicaciones científicas no son tan ajenos a las bases de datos de recursos de patrimonio cultural como a veces pudiera parecer o se quiere hacer creer; hay muchos *puntos en común* que afectan sobre todo al área de las humanidades y ciencias sociales, pero también a muchas otras áreas de las ciencias. Quizá sea necesario recordar los análisis físico-químicos aplicados a la datación de los objetos arqueológicos o a la conservación de sitios y piezas artísticas o documentales. En ocasiones la separación entre publicaciones científicas (*academic research*) y recursos de patrimonio cultural se deriva más de la diferenciación de entornos profesionales que de la lógica de la información, sobre todo siendo factible, como es, la aplicación de una normativa sencilla que permita su interrelación. La *Deutsche Digitale Bibliothek* o *Hispana* son algunas muestras de que se puede combinar sin problema ambos tipos de información.

⁵ http://en.wikipedia.org/wiki/Herbert_Van_de_Sompel

⁶ <http://pubserv.oclc.org/oaicat/jars/docs/>

⁷ <http://dublincore.org/documents/dcmi-terms/#H3>

La transformación de información en conocimiento requiere de un modelo de datos adecuado para la visualización de las interrelaciones entre los recursos, de modo que los límites del entorno profesional no sean una barrera para que los usuarios sigan el camino que les parezca más apropiado en ese mar de interrelaciones. Evidentemente, manejar decenas de millones de registros en una interfaz de búsqueda y mantener la pertinencia de los resultados es un arduo problema que se solventa a menudo con la aplicación de filtros y facetas a los resultados para que sean más manejables por los usuarios. Así, las necesidades del portal Europeana llevaron a la ampliación del modelo de datos de DC sin cualificar al *Europeana Semantic Elements* (ESE)⁸ (versión 3.2 de fecha de agosto de 2009), que no era sino una combinación de un subconjunto de *DCMI Metadata Terms*⁹ (22 elementos) junto con 13 nuevos elementos propios de Europeana. De hecho, la aplicación de ESE en el portal de Europeana se realizó internamente antes de que fuera un requisito técnico para los proveedores de datos, al igual que ha ocurrido con el Europeana Data Model. Se cumplían así dos objetivos, conservar la complejidad de los registros que los proveedores de datos podían aportar, y mejorar la presentación y acceso a los mismos en el portal de Europeana.

De la experiencia de Europeana podemos extraer algunas conclusiones. En primer lugar, ha ampliado rápidamente los elementos Dublin Core, desde 15 a 37 elementos, y ha añadido nuevos elementos propios para resaltar adecuadamente las aportaciones de los proveedores e identificar claramente aspectos como el idioma, el país de origen, los derechos y otros elementos complementarios relacionados con la administración y gestión de las agregaciones. En segundo lugar, Europeana ha tenido la virtualidad de contribuir a la normalización de los sistemas de información de archivos, bibliotecas y museos europeos de un modo nunca visto anteriormente. Efectivamente, el todo se ha convertido en mucho más que la suma de las partes. El esfuerzo normalizador se ha producido no solo en un entorno, digamos fácil, para DC como son las bibliotecas, también ha tenido un efecto similar en museos, aunque mucho menor en archivos¹⁰. En 2010 Vangelis Banos publicó un *plug-in*¹¹ para que los repositorios basados en DSpace¹² pudieran ser recolectados por Europeana, es decir según el esquema de metadatos ESE. Los repositorios OAI de las bibliotecas virtuales y digitales basadas en el ILS DIGIBIB¹³ permitían la recolección de metadatos ESE desde el mismo año 2009. Incluso, DIGIBÍB publicó una serie de instrucciones¹⁴ complementarias que solventaban algunos problemas de instalación del mencionado *plug-in* en DSpace. Son sólo dos ejemplos de

⁸ <http://pro.europeana.eu/technical-requirements>

⁹ <http://dublincore.org/documents/2012/06/14/dcmi-terms/>

¹⁰ Véase Technical Report: Archival Digital Object Ingestion into Europeana (ESE-EAD harmonization) Version 1.0, 07/08/2009 [<http://goo.gl/TRto3>]

¹¹ <http://vbanos.gr/blog/2010/02/11/dspace-plugin-for-europeana-semantic-elements/>

¹² <http://www.dspace.org/>

¹³ <http://www.digibis.com/bibliotecas-virtuales.html>

¹⁴ http://www.digibis.com/images/PDF/20101221_instalar_soporte_ese_en_dspace.pdf

cómo el software comercial y el software *open-source* se han acomodado a los requisitos de Europeana. Y esto sólo se ha producido por el interés de los proveedores de datos en formar parte de la biblioteca digital europea. Evidentemente, para que los elementos de ESE puedan ser transmitidos es necesario que estén presentes en el repositorio del proveedor de datos. Ya veremos que esto es especialmente importante en el caso del Europeana Data Model¹⁵.

Sin embargo, este esfuerzo se ha realizado también en numerosas ocasiones con un coste adicional como ha sido la duplicación de tareas, creándose ecosistemas de datos desconectados dentro de las organizaciones de los proveedores de datos. Así, se podría hablar del entorno propio de la biblioteca tradicional con sistemas de gestión bibliotecaria (en su mayoría comerciales) incapaces de gestionar adecuadamente objetos digitales y del entorno propio de los repositorios digitales no aptos para gestionar la circulación de ejemplares o con dificultades para introducir procesos como el control de autoridades. Esta impermeabilidad puede tener peores efectos en el marco de Linked Open Data. Bien sea por la lentitud en la respuesta de algunos programas comerciales o por la urgencia en dar visibilidad a los objetos digitales (digitalizados o nacidos digitalmente), el resultado ha sido una creciente y costosa complejización de la gestión de las bibliotecas. Quizá ésta sea la explicación de por qué el tamaño medio de los repositorios digitales es tan pequeño¹⁶, ya que mantener una gestión paralela para objetos físicos y para objetos digitales es difícil, del mismo modo que es difícil mantenerse actualizado tecnológicamente.

Por ejemplo, la contribución de *Hispana* a *Europeana* podría tener más de 2,5 millones de objetos más si la totalidad de los 202 repositorios que recolecta fueran conformes con ESE. Obviamente, y por distintos motivos, no todos los repositorios tienen como objetivo participar en Europeana, pero parece que también el desfase tecnológico tiene su influencia. En este sentido, es fácil comprobar que los mayores proveedores de datos a *Europeana* son agregadores nacionales o transversales que han incorporado el mecanismo de recolección de una manera prácticamente automática, lo cual redundará en la visibilidad de las colecciones. El *Plan Estratégico de Europeana 2011-2015* marca como objetivo que exista un agregador nacional en cada uno de los países miembro de la Unión Europea para 2015¹⁷.

EUROPEANA DATA MODEL

Desde la publicación en 2010 de Stefan Gradmann *Knowledge = Information in Context: on the Importance of Semantic Contextualisation in Europeana*¹⁸ los esfuer-

¹⁵ <http://pro.europeana.eu/web/guest/edm-documentation>

¹⁶ Agenjo, Xavier. 2011, un gran año para las bibliotecas virtuales. Anuario ThinkEPI 2012

¹⁷ Véase p. 13. http://www.pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602

¹⁸ http://pro.europeana.eu/c/document_library/get_file?uuid=c417911-1ee0-473b-8840-bd7c6e9c93ae&groupId=10602

zos de Europeana han ido dirigidos a rehacer su modelo de datos para adecuarlo a las tecnologías de la Web Semántica y, como primera medida, de Linked Open Data.

Los objetivos claramente expuestos son los de proporcionar una búsqueda multilingüe en la que el usuario pueda obtener respuestas sobre *Who, What, Where, When*¹⁹, y unos resultados interrelacionados que le permitan seguir el camino que prefiera. Indudablemente el primer requisito de nuevo modelo era mantener la compatibilidad con *Europeana Semantic Elements*. La primera versión pública de *Europeana Data Model* fue la 5.2.2²⁰ de fecha 31 de mayo de 2011. Además de los *namespaces* de RDF, RDFS, SKOS y OAI-ORE, Europeana reutiliza *The Dublin Core namespaces for elements* (<http://purl.org/dc/elements/1.1/>), *terms* (<http://purl.org/dc/terms/>) y *types* (<http://purl.org/dc/dcmitype/>). Medio año después, el 24 de febrero de 2012 se publicó la versión 5.2.3 y en julio de 2013 la 5.2.4²¹ tras la transformación de la base de datos desde ESE a EDM que se realizó en el mismo mes.

Aunque se pueda criticar la calidad de algunos registros de Europeana, lo cierto es que se ha partido de una situación muy heterogénea. En primer lugar, se han aportado registros de archivos, bibliotecas y museos; en segundo lugar se han partido de diferentes estructuras de datos, como registros de la familia MARC no muy homogéneos (UNIMARC, IBERMARC, MARC 21 según distintas actualizaciones, etc.) y con toda seguridad ha sido necesario tratar multitud de estructuras de datos propietarias. En todos los casos Dublin Core ha actuado como modelo de datos común haciendo viable la homogeneización de los datos. Y si bien es cierto que no todos son de la misma calidad, sí lo es en cambio que los campos más relevantes (autor, *contributor*, título, materias, fechas y lugares) permiten gestionar con notable rendimiento el altísimo volumen de datos en la búsqueda e identificación. Y además estos elementos son los más importantes para cualquier tipo de usuario, puesto que otros datos como la extensión o las notas descriptivas se suplen con ventaja con la consulta de los propios objetos digitales.

Gran parte del esfuerzo de normalización de archivos, bibliotecas y museos se ha realizado en el marco de grandes proyectos europeos financiados parcialmente por la Comisión Europea como ATHENA, APENET, EUSCREEN, etc. Sin embargo, para todos ellos los *WorkProgramme* de los diferentes mecanismos de financiación (p.e., *ICT Policy Support Programme*) han tenido una influencia fundamental. Así, se ha establecido a nivel europeo, mediante la consiguiente financiación, la dirección estratégica de los objetivos y con ello se ha potenciado no sólo los tipos de documentos (p.e., impulsar un tipo de material como los archivos sonoros, filmicos, etc.), los tipos de proveedores (p.e. editores), el uso de tecnologías emergentes (*linked data* o *cloud computing*) y, por supuesto, el requisito fundamental siempre ha sido que los resultados de cualquier proyecto sean visibles en Europeana.

¹⁹ Son exactamente las etiquetas de búsqueda de Europeana

²⁰ http://pro.europeana.eu/documents/866205/13001/EDM_v5.2.2.pdf

²¹ <http://pro.europeana.eu/documents/900548/0d0ff6ec3-1905-4c4f96c8-1d817c03123c>

Por su parte los organismos que sustentan Europeana, la Europeana Foundation y la Europeana Network, han alentado la participación de los proveedores de contenido en diferentes tareas de coordinación relacionadas con los aspectos más problemáticos de la biblioteca digital europea. En este trabajo no se entrará en las cuestiones relacionadas con los derechos de propiedad intelectual, la sostenibilidad y otras que han contado con diferentes *Task Forces*²². Estos grupos de trabajo y de interés analizan, durante un periodo de tiempo definido, un problema candente y proponen soluciones. Así las *Tasks Force sobre Hierarchical Object*, sobre FRBRoo o sobre identificadores persistentes han analizado sus respectivas cuestiones. Los proveedores de contenido pueden participar en estas tareas, independientemente de su participación en proyectos europeos y/o al mismo tiempo. De este modo la normalización en archivos, bibliotecas y museos europeos ha avanzado enormemente, si se compara con la situación inicial, en un brevísimo espacio de tiempo y ha consistido en la elaboración de una normativa orientada hacia un fin, *Europeana*.

Quizá a muchos lectores les sorprenderá saber que hasta el nacimiento de Europeana, y la asunción por parte de la *Agenda Digital para Europa* (2010)²³ de la biblioteca digital europea, la interoperabilidad de los archivos, bibliotecas y museos europeos ha sido muy fragmentaria y circunscrita a los proyectos desarrollados en el marco de los programas ICT, e-ContentPlus u otros. Con demasiada frecuencia el fruto del trabajo de estos proyectos se desvanecía tras el final del mismo, y los esfuerzos de normalización de las grandes organizaciones no gubernamentales del sector como IFLA, ICA e ICOM, tenían unos efectos mínimos. Es significativo que en el entorno de las publicaciones académicas y científicas no se haya producido este hecho y a pesar de los esfuerzos de proyectos como DRIVER²⁴, COAR²⁵ u OpenAire²⁶, no ha llegado a consolidarse un portal similar a Europeana que proporcione continuidad a las tentativas de normalización.

Es esperable y deseable que la normalización e interoperabilidad continúe, la puesta en funcionamiento de las *Task Forces* de Europeana así lo hace prever, puesto que es mucha la cantidad de información que ha aflorado y mucha más la que tiene que aflorar. El proyecto ENUMERATE²⁷, una *Thematic Network* financiada por la Comisión Europea dentro del programa ICT-PSP para los años 2011-2013 tiene como cometido proporcionar datos fidedignos sobre el estado de la digitalización en Europa, sobre sus costes, su acceso y preservación. La primera encuesta²⁸ arroja como cifra general que cerca del 20% de las colecciones que se necesitan digitalizar se encuentra ya digitalizada, por lo que se puede ver que Europeana no ha hecho más que empezar.

²² <http://pro.europeana.eu/network/task-forces/overview>

²³ <http://ec.europa.eu/digital-agenda/>

²⁴ <http://www.driver-repository.eu/>

²⁵ <http://www.coar-repositories.org/>

²⁶ <http://www.openaire.eu/>

²⁷ <http://enumerate.eu/>

²⁸ En el momento de redactar este texto se ha lanzado la segunda encuesta general de ENUMERATE que sin duda modificará en algo esta cifra.

Sin embargo, se puede apreciar que aún persisten escollos para lograr una continuidad en la aportación normalizada de objetos digitales. En la medida que las colecciones digitales y sus metadatos no estén incorporadas a la rutina habitual de los archivos, bibliotecas y museos, incluso formando parte de los mismos programas comerciales de gestión, las aportaciones de los proveedores crecerán durante la duración de un proyecto y se estancarán a la finalización del mismo. A pesar de las grandes promesas del software libre, lo cierto es que éste requiere de profesionales especializados para su implantación y mantenimiento y que en muchas ocasiones supone la instalación de un conjunto de piezas complejo para distintas operaciones no integradas, lo que añade aún más complejidad. Por su parte, el software comercial, al menos en la experiencia española, y es fácil ver que en muchos otros países, no responde con agilidad a las necesidades de las bibliotecas, archivos y museos digitales. Prueba de ello es que el *Business Plan* de 2013²⁹ marca como objetivo que 14 países miembros de la UE aumenten su contribución.

LA DIGITAL PUBLIC LIBRARY OF AMERICA Y SU PERFIL DE APLICACIÓN DE METADATOS

En abril de 2013 se lanzó oficialmente la *Digital Public Library of America*³⁰ que tras siete meses ofrece más de 4,5 millones de registros de bibliotecas, archivos y museos. Ya en la *Europeana Tech Conference*, celebrada en Viena en 2011, Maura Marx hizo una exposición sobre la DPLA³¹ y quedó claro para todos los asistentes que su modelo iba a ser muy similar al de *Europeana*³². La experiencia (y éxito) de *Europeana* ha tenido una gran influencia en la DPLA. Ambos proyectos partían de situaciones muy parecidas en cuanto a los objetivos, pero muy diferentes en cuanto al estado de la normalización y la interoperabilidad; el multilingüismo europeo planteaba además una complejidad añadida. Ambos proyectos convergían en situar a Linked Open Data como una tecnología factible para gestionar un enorme volumen de datos y extraer su potencia informativa.

El 25 de octubre de 2011 el *W3C Library Linked Data Incubator Group*³³ publicó su informe final. Al día siguiente se publicaron el *Europeana Data Model Primer*³⁴, el *Europeana Data Model Mapping Guidelines* versión 1.0.1, actualmente en la versión 2.0³⁵, y *The Europeana Data Model for Cultural Heritage*, que ha sido sustituido por *General EDM factsheet and Presentation*³⁶. El importante documento

²⁹ <http://pro.europeana.eu/documents/858566/9d4632d3-3f6d-4162-ba29-27a9a739946d>

³⁰ <http://dp.la>

³¹ <http://www.europeanaconnect.eu/europeanatech/index.php?section=programme&page=programme>

³² Véase el texto de la DPLA en este mismo boletín: La colaboración multiplica el valor: el modelo de asociación internacional de la DPLA y Europeana / Amy Rudersdorf, Emily Gore.

³³ <http://www.w3.org/2005/Incubator/ldd/>

³⁴ <http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>

³⁵ <http://pro.europeana.eu/documents/900548/60777b88-35ed-4bae-8248-19c3696b81fb>

³⁶ <http://pro.europeana.eu/documents/900548/2f66c73a-f5c9-49cc-a8e1-4b3ed763bc63>

*D3.4 Final Technical & Logical Architecture and future work recommendations*³⁷ de Stefan Gradmann se publicó también ese mes.

Es de suma importancia que la DPLA haya reutilizado en gran parte el modelo de datos de Europeana pues contribuirá a facilitar la colaboración entre ambos proyectos y entre los participantes en los mismos en otros ámbitos. Además, con toda seguridad centrará el amplio panorama de los esquemas y ontologías en uso, o en desarrollo, en el mundo de los *Library Linked Data*, que no ha hecho más que crecer en los últimos años. No deja de recordar a la explosión de formatos MARC de los años 80 y 90 que se alivió un tanto con la armonización de los formatos CANMARC, UKMARC y USMARC en 1997 y que dio lugar al formato MARC21. Esperemos que ocurra lo mismo que entonces.

La publicación del *DPLA Metadata Application Profile v3*³⁸ ha puesto de manifiesto hasta qué punto el modelo de datos de *Europeana* ha sido útil para la DPLA. De hecho, en los grupos de trabajo de Europeana Network ya ha surgido alguna voz promoviendo que EDM se convierta en un estándar internacional, idea que indudablemente se habrá visto muy fortalecida con la publicación del *DPLA Metadata Application Profile*. Su primera frase no puede ser más significativa “*This version is designed to build on the experience of the Europeana Data Model (EDM). In addition, it incorporates feedback from the DPLA community and digital hub pilot participants to create a balanced framework that allows us to accommodate existing and emerging data models for library, archive, and museum resources*”. Es una muestra de la capacidad descriptiva del *Europeana Data Model* para representar los *Cultural Heritage Objects*, para utilizar la propia definición del modelo, y sus objetos digitales. Es muy probable que se produzcan modificaciones en este perfil de aplicación de metadatos, pero también es previsible que todo ello se haga de forma coordinada con *Europeana*. De hecho, esto se menciona explícitamente en el *Business Plan* de 2013³⁹ en el que se cita la necesidad de trabajar coordinadamente con otras instituciones: *Promote knowledge-sharing on R&D activities in the various Europeana-related projects. Support appropriate R&D project proposals and continue to actively engage with R&D activities from relevant projects and other initiatives (e.g. Digital Public Library of America (DPLA), the World Wide Web Consortium (W3C), Dublin Core) and the private sector*.

Las *Task Forces*⁴⁰ de Europeana, que proponen recomendaciones para el conjunto de la red, han dado lugar ya a algunas propuestas de modificación. El 28 de marzo de 2013 la *Hierarchical Objects Task Force* publicó *Recommendations for the representation of hierarchical objects in Europeana*⁴¹ que se han incluido en la

³⁷ <http://pro.europeana.eu/documents/10602/370691/D3.4+final.pdf>

³⁸ Digital Public Library of America. Metadata Application Profile, Version 3. Release date: 2013-02-08. Document URI: <http://dp.la/about/map>

³⁹ <http://pro.europeana.eu/documents/858566/9d4632d3-3f6d-4162-ba29-27a9a739946d>

⁴⁰ <http://pro.europeana.eu/network/task-forces/overview#EuropeanaTech>

⁴¹ <http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Taskforce+on+hierarchical+objects>

última versión de EDM (5.2.4). Lo más significativo es que se trata de recomendaciones tanto para los proveedores de contenido como para la propia Europeaana. También han publicado sus conclusiones las *task forces* sobre *Persistent Identifiers* (PIDs) y *EDM - FRBRoo Application Profile*.

Uno de los aspectos más interesantes de las recomendaciones de la *Hierarchical Objects Task Force* es su relación con los datos de los archivos, hasta ahora muy insuficientemente representados en Europeaana (de forma conjunta 43%, Bibliotecas 40%, Museos 10%, Audiovisuales 6%, y Archivos 1%)⁴². Por el contrario, parece que los archivos sí van a estar fuertemente representados en la DPLA como puede verse en la información aportada por los National Archives and Records Administration (NARA)⁴³. En este sentido resulta muy importante el hecho de que los NARA hayan anunciado su intención de utilizar de forma sistemática tanto *Encoded Archival Description* como *Encoded Archival Context*. Además está anunciada una nueva versión de EAD tras más de 10 años desde la última versión (2002)⁴⁴.

Los autores de este trabajo han realizado un mapeo entre MARC 21 y EAC, todavía inédito, para su aplicación en los programas DIGIBIB y DIGIARCH que ha desarrollado DIGIBÍS. Ambos programas disponen de repositorios OAI recolectables por Europeaana, tanto en ESE como en EDM. Es muy revelador el caso del archivo familiar Montiano⁴⁵ que comparte el fichero de autoridades en MARC 21/RDA de la *Biblioteca Virtual de Polígrafos*⁴⁶ y que transmite datos en EDM. El Archivo Montiano ha sido uno de los casos de estudio de la *Task Force de Hierarchical Objects*.

La situación de los museos está más avanzada. El proyecto ATHENA (*Access to Cultural Heritage Networks Across Europe*)⁴⁷ aportó a Europeaana 2.543.302 registros y desencadenó un esfuerzo de normalización de los museos europeos que contribuiría finalmente al surgimiento del esquema *LIDO - Lightweight Information Describing Objects Version 1.0 (LIDO)* del ICOM-CIDOC Working Group Data Harvesting and Interchange⁴⁸. El proyecto *MIMO Musical Instruments Museum Online*⁴⁹ ha aportado a su vez 46.483 registros en EDM, generados a partir de LIDO. Dado que la Smithsonian Institution forma parte de la DPLA será muy interesante ver cuáles serán los mapeos que se utilicen para adaptarse al *DPLA Metadata Application Profile*. Indudablemente, el trabajo realizado en ATHENA (que continuará en ATHENA Plus⁵⁰), en MIMO y en otros proyectos será de

⁴² D3.7: Content Strategy Update. MS11: Content Acquisition & Development Plan Merged and renamed to: Collections and Data Analysis, Strategy and Plan. Project Acronym: Europeaana v2 [Project Acronym: Europeaana v2].

⁴³ <http://www.archives.gov/>

⁴⁴ <http://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/ead-revision>

⁴⁵ <http://pro.europeana.eu/documents/468623/09602a17-0f79-4145-9012-46adf7125585>

⁴⁶ <http://pro.europeana.eu/polymath-edm>

⁴⁷ <http://www.athenaeurope.org/>

⁴⁸ <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/>

⁴⁹ <http://www.mimo-international.com/>

⁵⁰ <http://www.athenaplus.eu/>

gran utilidad y pronto se verá el uso concreto de EDM para representar los objetos museísticos.

RESOURCE DESCRIPTION AND ACCESS (RDA)⁵¹

Hemos visto cómo los avances en la creación de grandes bibliotecas digitales, o puntos únicos de consulta a los diferentes tipos de objetos del patrimonio cultural, han supuesto una extensión de la normalización a la que los proveedores de contenidos han respondido por su interés en mantener esos sistemas de consulta. Pero hemos de considerar también los cambios en las normas que se están produciendo en sectores específicos, como el bibliotecario, y los efectos que tendrá en la normativa de los grandes proyectos como Europea o la DPLA. El acontecimiento de mayor importancia ha sido la adopción por parte de un amplio conjunto de bibliotecas nacionales de los *Resource Description and Access (RDA)*⁵².

Los datos creados conforme a las reglas RDA y su codificación en MARC21 van a constituir una de las fuentes de información fundamentales para datos sobre personas e instituciones. Uno de los cambios sustanciales de EDM sobre ESE es la introducción de las clases contextuales (*Agent, Place, TimeSpan, Concept, ConceptScheme, Event, Physical Thing*), del mismo modo que es un gran cambio la introducción de los campos 046, 368 y 37X⁵³, de fuerte contenido semántico, en los registros MARC. Como ejemplo, los registros de autoridad de nombres de personas deben incluir información propia de la persona, lugares de nacimiento y muerte, fechas de nacimiento y muerte, ocupación y profesión, género, etc. Todos los elementos codificables según *RDA Group 2 Elements (rdaGr2)*⁵⁴ están recogidos, de una u otra manera, en las propiedades de la clase *Agent*. EDM ha reutilizado las propiedades de *rdaGr2*, pero subsiste alguna incertidumbre con respecto a la perdurabilidad del *namespace* RDA pues como se dice en *Open Metadata Registry*: “This is the provisional registration of the *RDA Group 2 Element Vocabulary*, managed by the *DCMI/RDA Task Group*”⁵⁵, lo cual resulta muy importante para los desarrollos de software que se realicen, ya que introduce una duda sobre el hecho del mantenimiento del vocabulario. Por el contrario, la existencia de Europea como organización y de sus *Task Forces* suponen una garantía de mantenimiento y continuidad, como puede verse en la última propuesta *Europeana Collection Profile* de octubre de 2013⁵⁶. No es difícil entender por qué la Library of Congress ha optado por desarrollar la *Bibliographic Framework Transition Initiative (BIBFRAME)*⁵⁷ en lugar de reutilizar ontologías de dudosa perdurabilidad.

⁵¹ <http://www.niso.org/news/events/2013/dcmi/rda/>

⁵² <http://www.loc.gov/aba/rda/>

⁵³ <http://www.loc.gov/marc/RDAinMARC.html>

⁵⁴ <http://metadataregistry.org/schema/show/id/15.html>

⁵⁵ El subrayado es nuestro

⁵⁶ <http://pro.europeana.eu/documents/900548/90890f64-777b-40d6-af78-593864a41072>

⁵⁷ <http://bibframe.org>

Por otro lado, RDA introduce un amplio conjunto de valores para la codificación de las funciones que una persona (o entidad) puede realizar en relación con una obra, que se trasladan a Dublin Core como las propiedades genéricas *dc:creator* y *dc:contributor*. Se perderá, por tanto, precisión en las conversiones de MARC a Dublin Core y, en consecuencia, se ve la necesidad de refinar estas propiedades para poder acomodar la lista de funciones establecida en el §e de los X00⁵⁸, aunque sea de una forma simplificada. Otra pregunta que surge es hasta qué punto será necesario crear nuevas clases en el modelo de datos Dublin Core o si es más conveniente reutilizar las clases contextuales de EDM (y DPLA), como parece lógico.

Todo ello nos lleva nuevamente a que EDM, o la conjunción de EDM y *DPLA Metadata Application Profile* se conviertan en una norma oficial. Europea viene ejerciendo esta función desde la primera versión de *Europeana Data Model*, pero parece necesario un mayor compromiso institucional, al menos para quienes tienen que tomar las decisiones de implantar el software necesario para cumplir el modelo de datos. Desde el punto de vista de los autores de este trabajo resulta trascendental que el modelo de datos permita recoger toda la nueva información RDA/MARC 21⁵⁹. Éste es, sin duda, el objetivo de la última *Task Force* de Europea que se ha constituido *EDM mappings refinements and extensions*⁶⁰.

Es obvio que para los proveedores de datos disponer del software que les permita realizar las operaciones habituales en archivos, bibliotecas y museos y al mismo tiempo transmitir sus datos conforme a la normativa de Europea o DPLA de una forma integrada es lo más ventajoso. Y también lo es para estos grandes proyectos, pues los datos fluirán regularmente y no a trompicones. Dicho brevemente, la situación ideal es la de un sistema de gestión que convierta automáticamente, y de forma transparente al usuario, los datos catalogados al modelo de datos propio de Europea y de la DPLA (o ambos) y que al mismo tiempo comparta las reglas del universo bibliotecario, archivístico o museístico. De otro modo, no será sostenible económicamente el proceso. Este es el planteamiento de desarrollo de los productos de software de DIGIBÍS y que se manifiesta en que el 62 % de los objetos digitales españoles en Europea proceden de

En busca de esa normativa internacional para el intercambio de información sobre el patrimonio cultural hemos de mencionar al *CIDOC Conceptual Reference Model (CIDOC-CRM)*⁶¹, que es norma ISO desde 2006⁶² y que actual-

⁵⁸ <http://www.loc.gov/marc/relators/relaterm.html>

⁵⁹ Xavier Agenjo, Francisca Hernández and Andrés Viedma. Data Aggregation and Dissemination of Authority Records through Linked Open Data in a European Context. *Cataloging & Classification Quarterly*, Volume 50, no. 8, 2012

⁶⁰ <http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+force+on+EDM+mappings+refinements+and+extensions>

⁶¹ La última versión es la 5.1 de noviembre de 2012. <http://www.cidoc-crm.org/>

⁶² ISO 21127:2006. Information and documentation – A reference ontology for the interchange of cultural heritage information.

mente está en fase de revisión. A pesar de la temprana existencia de esta norma, su utilización en Europea se ha concretado en la reutilización de algunas clases (o en la definición de equivalencias entre algunas clases y propiedades de EDM a CIDOC-CRM y no del modelo completo. De hecho, tanto EDM como el *Metadata Application Profile* de la DPLA parecen estar llamados a cumplir las funciones de CRM. Una de las primeras decisiones de *Europeana Data Model*⁶³ fue la posibilidad de adoptar tanto una visión centrada en los *objetos* como en los *eventos*. Mientras que la visión centrada en objetos es la utilizada en bibliotecas, los eventos son más propios del ámbito museístico, sobre todo después de la publicación de LIDO. El propio CRM permite ‘atajos’ para abreviar el modelado de la realidad por medio de eventos. Indudablemente, desde el punto de vista de un catalogador es más fácil decir que ‘X es el ilustrador de la obra Y’, que decir, ‘un determinado evento de creación (E65 Creation), llevado a cabo (P14 carried out by) por un (E39 Actor) ha creado (P94 has created) una (E38 Image)’. Tanto Europea como la DPLA deben estar preparadas para admitir ambos tipos de representaciones.

La implementación sistemática de la RDA se inició en los Estados Unidos el 31 de marzo de 2013, y paulatinamente se han incorporado las bibliotecas nacionales de Canadá, Reino Unido, Alemania y Australia, y pronto lo hará Francia. Alemania no sólo ha tenido que modificar sus antiguas reglas de catalogación RAK (*Regeln für die Alphabetische Katalogisierung*), sino también su modelo de datos para el intercambio de información bibliográfica MAB (*Maschinelles Austauschformat für Bibliotheken*) por, como es lógico, MARC21. Además, no hace muchos meses que se puso en producción quizá el mayor proyecto de objetos digitales en Europa como es la *Deutsche Digitale Bibliothek* (DDB)⁶⁴.

Por otra parte, el modelo del *Bibliographic Framework Transition Initiative* (BIBFRAME) está en una fase preliminar, aunque ya se pueden ver los primeros borradores de definiciones. Así, en el caso de la clase *Agent*⁶⁵ aún no se han creado las propiedades que recojan la información contenida en los campos 37X de MARC/RDA. La experiencia de la *Biblioteca Virtual de Polígrafos*⁶⁶ muestra hasta qué punto son útiles desde un punto de vista semántico los campos 37X.

Parece que la multiplicidad de modelizaciones de la clase *Agent* (o *Person*) dentro de los modelos de datos EDM, RDA, MADS, BIBFRAME, por no mencionar FOAF, está reclamando un sistema simple al modo de SKOS⁶⁷, un mínimo denominador común que no pierda los valores semánticos establecidos por RDA y que tan útiles van a ser en la contextualización, en la búsqueda y en la navegación de los grandes bases de datos de patrimonio cultural y científico. Para la *Biblioteca Virtual de Polígrafos* los registros de autoridad de nombres

⁶³ Europeana Data Model Primer. <http://goo.gl/AY8QJ>

⁶⁴ <https://www.deutsche-digitale-bibliothek.de/>

⁶⁵ <http://bibframe.org/vocab/Agent.html>

⁶⁶ <http://goo.gl/x3jOz4>

⁶⁷ <http://www.w3.org/2004/02/skos/>

de persona, transformados en registros de persona por la aplicación de MARC/RDA se han convertido en un recurso informativo con carácter propio.

La multiplicidad de modelos de datos también afecta a la reutilización de los mismos. Los modelos de datos de *Europeana*, ESE y sobre todo EDM, han supuesto una convergencia de los datos de archivos, bibliotecas y museos, pero también una unificación de las estructuras de datos en ámbitos concretos, por ejemplo en las bibliotecas. La panoplia de formatos MARC en toda Europa dificultó enormemente el intercambio de registros, muy parecidos, pero no exactamente iguales. Esta misma situación puede darse en la reutilización de datos Linked Open Data.

REUTILIZACIÓN DE LOS DATOS

Se pueden descargar⁶⁸ 20 millones de objetos digitales de *Europeana* en EDM a través del proyecto *Europeana Linked Open Data (LOD)*⁶⁹. España participó en el proyecto inicial con los registros procedentes de la *Biblioteca Virtual de Patrimonio Bibliográfico* y los de la *Biblioteca Virtual de Prensa Histórica*. También se pueden explorar los *datasets* de *Europeana* a través de un *SPARQL endpoint*⁷⁰, conforme con SPARQL 1.1⁷¹, aunque su operatividad no parece estable.

Las ventajas de SPARQL son innumerables para realizar búsquedas complejas, para obtener datos y para construir conversiones entre modelos de datos⁷², entre otras cosas. Sin embargo, no puede decirse que sea usable⁷³ para la mayoría de los usuarios, incluidos archiveros, bibliotecarios y museólogos. Además de dominar el lenguaje, requiere conocer en profundidad el o los modelos de datos a consultar, y por lo tanto se ve la necesidad de desarrollar aplicaciones que permitan explotar convenientemente estos datos. De otro modo, el enorme desarrollo de Linked Open Data para que los datos sean tratados por máquinas redundará en una enorme brecha digital entre los humanos y las máquinas, o entre los humanos que manejan SPARQL y los que no pueden hacerlo.

Tal vez uno de los mayores aciertos de *Europeana* haya sido desarrollar una API (*Application Programming Interface*)⁷⁴ que permite ejecutar una consulta a su base de datos. Un ejemplo podría ser útil. Si se pregunta por Luis Vives en una biblioteca de digamos 100.000 registros puede que se obtengan 200 respuestas que se ajusten a nuestra petición; ahora bien si simultáneamente se ejecuta esa consulta en la base de datos de *Europeana* (30 millones de objetos digitales)

⁶⁸ <http://data.europeana.eu/download/2.0/>

⁶⁹ <http://pro.europeana.eu/linked-open-data>

⁷⁰ <http://europeana.ontotext.com/sparql>

⁷¹ <http://www.w3.org/TR/sparql11-query/>

⁷² Véase las SPARQL CONSTRUCT queries definidas en Dublin Core to PROV Mapping. W3C Working Draft 12 March 2013. <http://www.w3.org/TR/prov-dc/>

⁷³ <http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/index.html>

⁷⁴ <http://pro.europeana.eu/api>

es sumamente probable que el número de respuestas sea mayor en uno o dos grados de magnitud. En la *Biblioteca Virtual de Polígrafos* de la Fundación Ignacio Larramendi se ha instalado la API de Europeana con el resultado descrito. Además es posible reutilizar los datos porque Europeana API permite obtener los registros completos en JSON (*JavaScript Object Notation*)⁷⁵. Del mismo modo, la DPLA ha publicado su API⁷⁶ con el mismo fin. La ventaja de construir aplicaciones usables sobre estas APIs⁷⁷ puede verse en *Europeana/DPLA Query*⁷⁸ desarrollada por DIGIBÍS para consultar en una misma pregunta ambas bases de datos.

En esa misma línea conviene mencionar que también OCLC ha desarrollado una API⁷⁹ para buscar, identificar y seleccionar los encabezamientos de materia formateados en SKOS y vinculados a LCSH y GeoNames en su servicio FAST (*Faceted Application of Subject Terminology*)⁸⁰. Parece que una integración de las APIs de FAST y de la DPLA obtendría magníficos resultados como ya se puede apreciar en la integración entre FAST y WorlCat. Igualmente los registros del *Virtual International Authority File*, de enorme utilidad, están accesibles a través de la correspondiente API⁸¹.

La vinculación de datos que propone LOD es de enorme interés para la constante retroalimentación de un sistema, puesto que junto con los datos podemos utilizar también los URIs asociados por medio de las propiedades *owl:sameAs*⁸² o *skos:closeMatch*. Mediante la vinculación a vocabularios de valores o recursos comunes puede identificarse una misma cosa por medio de diferentes descripciones de la misma, y por supuesto con sus nombres en múltiples idiomas. Europeana está realizando este proceso de dos formas. Una, cruzando los datos con vocabularios como Semium Time⁸³, GeoNames⁸⁴, o el tesoro multilingüe GEMET⁸⁵. Y otra, a través de la recomendación de determinados vocabularios de valores para cada una de las clases como puede verse en la clase *edm:place*⁸⁶. Es decir, la reutilización de datos vinculados supone tanto la posibilidad de enriquecer los datos, mejorar la navegación y realizar búsquedas multilingües dentro de un sistema, como poder obtener las redes de vínculos establecidos por otros.

Ahora, la pregunta es ¿estos procesos debe realizarlos sólo Europeana (o la DPLA) o el proveedor de contenidos debe contribuir a estos procesos? Desde

⁷⁵ <http://es.wikipedia.org/wiki/JSON>

⁷⁶ <http://dp.la/info/developers/codex/>

⁷⁷ <http://pro.europeana.eu/pro-blog/-/blogs/1558732>

⁷⁸ <http://www.digibis.com/dpla-europeana/>

⁷⁹ <http://oclc.org/developer/services/fast-linked-data-api>

⁸⁰ <http://www.oclc.org/research/activities/fast.html>

⁸¹ <http://www.oclc.org/developer/services/viaf>

⁸² Véase <http://sameas.org/>

⁸³ <http://semium.org/time/>

⁸⁴ <http://sws.geonames.org/>

⁸⁵ <http://www.eionet.europa.eu/gemet>

⁸⁶ <http://europeanalabs.eu/wiki/EDMObjectTemplatesProviders>

nuestro punto de vista debe ser el proveedor de datos el que ajuste los vínculos deseados y seleccione los recursos a los que vincularse para cumplir con sus objetivos informativos, independientemente de que Europeana o la DPLA realicen los enriquecimientos y vinculaciones que consideren necesarios para el funcionamiento informativo del sistema. Aún más, la verdadera visibilidad de los datos se consigue extremando el cuidado de los propios datos en su enriquecimiento semántico y en su vinculación con fuentes externas. Y para ello debe disponer de las herramientas, el conocimiento de los modelos de datos y de los recursos LOD disponibles. El medio más útil para hacerlo y participar plenamente en los sistemas de información del mundo LOD es la transformación de los ficheros de autoridad y la reutilización de los datos de autoridades, de todo tipo, disponibles en LOD. Es difícil pensar que sin la colaboración de los proveedores de contenido la aplicación de EDM y LOD en Europeana pueda realizarse únicamente por los medios de ésta.

CONCLUSIONES

El camino recorrido hasta llegar a Europeana ha requerido de un proyecto común con una organización y un soporte político y financiero por parte de la Comisión Europea, pero también de los esfuerzos individualizados de los proveedores de datos en torno a un conjunto de normas. La inauguración de la DPLA y el desarrollo de un modelo conjunto de datos entre Europeana y la DPLA va a conducir a que su uso se extienda mundialmente. Corea del Sur y Nueva Zelanda ya han firmado acuerdos de colaboración con Europeana y es previsible que toda América, como ya lo está haciendo Europeana, preste una especial atención a la DPLA. Luego, el modelo de datos que se construya es de un enorme interés mundial.

Estamos seguros que aún transcurrirá algún tiempo para que BIBFRAME sea una realidad eficaz para englobar todo lo que supone tanto la catalogación MARC21/RDA como el decantado DC. Con toda seguridad, próximamente tendremos ocasión de contemplar modelos, programas, iniciativas y propuestas que busquen ese objetivo común, pues la capacidad de vincular no sólo los datos bibliotecarios sino también los archivísticos y los museológicos es sólo una pequeña parte de todo lo que permite y promete el modelo LOD. Una de las recomendaciones del WC3 LLD es justamente la de vincular y reutilizar los *datasets* de otras áreas que pueden interactuar con la información LOD-LAM (*Linked Open Data in Libraries, Archives & Museums*), y viceversa abrir sus datos a otras áreas. Esto es especialmente importante en ámbitos como el de la justicia, la salud o la administración electrónica, especialmente esta última, cuya importancia política a nadie se le escapa⁸⁷.

⁸⁷ <http://ec.europa.eu/digital-agenda/en/ict-enabled-public-sector-innovation-through-open-government>