

El Modelo de Datos de Europeana (EDM)

Martin Doerr

Institute of Computer Science, Foundation for Research and
Technology - Hellas
Creta, Grecia

Stefan Gradmann

Library and Information Science, Humboldt University
Berlín, Alemania

Steffen Henricke

Humboldt University
Berlín, Alemania

Antoine Isaac

Europeana
La Haya, Países Bajos

Carlo Meghini

Istituto di Scienza e Tecnologie dell'Informazione
Pisa, Italia

Herbert van de Sompel

Research Library, Los Alamos National Laboratory
Los Alamos, NM, USA

Meeting: 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY

10-15 Agosto 2010, Gotemburgo, Suecia

<http://www.ifla.org/en/ifla76>

Traducción al español

Ricardo Santos - Biblioteca Nacional de España

Resumen

El Modelo de Datos de Europeana (European Data Model – EDM) es un nuevo intento de estructurar y representar los datos con los que las diversas instituciones del patrimonio cultural contribuyen a Europeana. El modelo pretende ser más expresivo y flexible que el actual modelo ESE (Elementos Semánticos de Europeana – Europeana Semantic Elements), al que reemplazará. Los principios subyacentes de diseños de EDM se basan en los principios fundamentales y buenas prácticas de la Web Semántica y Linked Data, a lo que Europeana quiere contribuir. El modelo en sí se construye sobre estándares establecidos como RDF(s), OAI-ORE, SKOS y Dublin Core. Actúa como una ontología común de alto nivel que permite la interoperabilidad al mismo tiempo que mantiene los modelos de datos originales y sus perspectivas de información. Este artículo se ocupa de estos aspectos ya mencionados y de los principios de diseño que guiaron el desarrollo de EDM.

1. Introducción a Europeana

Europeana se presenta a menudo al público como un portal que da acceso a millones de objetos de patrimonio cultural de todo tipo, y aunque esta manera de presentar Europeana es cierta, no comprende algunas de las características esenciales del objetivo para las que está diseñado Europeana. Concordia et al. (2010) han tratado de dejar claro que Europeana no es un tanto un portal caracterizado únicamente por su volumen, sino que el principal objetivo de nuestra empresa es poner unos datos y unas funcionalidades valiosos disponibles a través de un API. Así se permitiría a toda clase de comunidades externas hacer de nuestras representaciones del rico tesoro cultural europeo para sus propias necesidades, y que el portal de Europeana que se ofrece a través de <http://www.europeana.eu> fuera, al final, una más de esas comunidades que hacen uso de esta riqueza de datos y funcionalidades por medio del API.

Más concretamente la idea es ofrecer una rica contextualización semántica para las representaciones de objetos, de tal manera que permita realizar, sobre los recursos, operaciones semánticas complejas que no podrían ser soportadas en la interfaz tradicional de una biblioteca digital. Con el fin de hacer posible esta funcionalidad las representaciones de objetos en Europeana necesitan estar conectadas sistemáticamente a Linked Open Data, bien en Internet o en recursos de contextualización semántica propios de Europeana, como tesauros y vocabularios estructurados migrados al estándar SKOS.

La estrategia técnica de Europeana fue, por tanto, concebida para contribuir al paradigma Linked Data y, al mismo tiempo, beneficiarse de él.

2. De ESE a EDM

El principal obstáculo a salvar antes de iniciar este camino que conduce a Linked Open Data fue el formato Elementos Semánticos de Europeana (*Europeana Semantic Elements (ESE)*), con modelo de datos subyacente simple y robusto. La creación y el uso de ESE había sido un pre-requisito indispensable para el lanzamiento del primer prototipo de Europeana en noviembre de 2008, pero desde entonces sus limitaciones se han hecho evidentes y – de un modo paradójico – han impedido pasar a un modelo semánticamente rico y funcional (probablemente ESE no debería haber sido bautizado como “semántico”)

En esencia, el principal problema con ESE es su enfoque de modelo “plano”, que no permite ni incluir enlaces a recursos externos en la web ni la extensibilidad a modelos más especializados, de mayor granularidad. Además, ESE no puede incluirse en Linked Data Namespaces de la manera que debiera para hacer a Europeana parte de esas futuras arquitecturas distribuidas de información; todo esto es solo para hacer un esbozo de los principales problemas con ESE.

En cuanto se empezó a trabajar en la capa de datos semánticos de Europeana (*Europeana Semantic Data Layer*), en el grupo 1 de tareas del proyecto EuropeanaConnect, resultó evidente que ESE tendría que ser reemplazado con urgencia, como resultado del proceso de especificación de un modelo de datos de Europeana (EDM), decidiéndose por consiguiente acelerar este proceso para obtener rápidamente resultados válidos y utilizables; esta fase de trabajo, que comenzó a finales del verano de 2009, dio como resultado la versión 5 del EDM, que es el que se explica en este artículo, y que se considera un candidato para la operacionalización.

3. Principios de diseño de EDM

3.1 Estándares de W3C y Web Semántica

En su primera etapa, la WWW consistía esencialmente en un gran aplicación hipertextual con multitud de “documentos” interconectados (= páginas web), a cada una de las cuales se le asignaba un Identificador Único de Recurso (*Unified Resource Identifier (URI)*). Este escenario cambia con la web semántica, que, entre otras cosas, incluye ahora representaciones de entidades del mundo real (llamados recursos no informativos; véase <http://www.w3.org/TR/cooluris>), a los que se les asigna igualmente un URI. En un reciente intento de clarificar los fundamentos de la Web Semántica, se han desarrollado los conceptos de “Linked Data” (<http://linkeddata.org>) o “Web de Datos”, que introducen la idea de un documento web (un “recurso de información”) que representa al recurso no informativo, y la idea de una redirección sistemática entre ambos.

Un cambio crucial que traerá la adopción de EDM a Europeana será la compatibilidad con el paradigma de la Web Semántica. En los últimos años, el Consorcio World Wide Web (W3C) ha introducido un número de estándares para posibilitar representar y compartir en la web datos estructurados y entendibles por máquinas. El primero de estos estándares es RDF (*Resource Description Framework*), que permite representar información estructurada acerca de cualquier recurso mediante simples declaraciones triples (sujeto, predicado, objeto). La visión que hay detrás de RDF es que los recursos se pueden describir por medio de conexiones entre ellos semánticamente expresivos. Por ejemplo, la tripleta (ec:ulises, ex:author, ex:james_joyce)¹ describe el libro Ulises mediante la conexión entre su identificador y otro que representa a James Joyce, usando un link llamado autor que expresa la relación entre el libro y el autor. EDM sigue fielmente este enfoque de descripción mediante tripletas.

Los tipos de links son cruciales para la capacidad de RDF de transmitir conocimiento semánticamente expresivo. Estas propiedades, junto con otras tipologías que pueden atribuirse a los recursos sujeto y objeto, son definidas en las ontologías. La palabra “ontología” se entiende aquí como sinónimo de conceptualización (Gruber, 1993), en oposición al significado que esta palabra tiene en filosofía: “la ciencia de lo que es, de las clases y estructuras de objetos, propiedades, acontecimientos, procesos y relaciones en todas las áreas de la realidad” (Smith, 2003). Las ontologías, que se definen por medio de RDF Schema (RDFS) y del estándar Web Ontology Language (OWL), contienen tanto definiciones informales, en forma de documentación leíble por humanos, y como definiciones formales, en forma de reglas y restricciones que permiten detectar inconsistencias o deducir nuevos hechos a partir de otros datos. Por ejemplo, una ontología puede definir clases para libros, pinturas y personas, o una propiedad ‘autor’, y declarar formalmente que todos los recursos conectados a libros mediante la propiedad ‘autor’ serán del tipo ‘persona’. Puede igualmente definir un objeto de otra clase como una superclase de autor y pintura. Mediante un motor de inferencia que trabaje con los datos de una colección de libros y pinturas, y buscando por todos los

¹ ex: es un nombre inventado de un espacio de nombre.

objetos creados por una persona, se podrían recuperar todos ellos sin conocer a priori su tipología, una característica crucial para la integración de información.

El enfoque de Web Semántica permite la combinación de varias ontologías en las mismas descripciones. Una puede implementar diferentes visualizaciones sobre las mismas afirmaciones, o construir afirmaciones que combinen diferentes vocabularios específicos para satisfacer los requisitos de una aplicación más general. Así, un catálogo de libros en puede reutilizar una ontología para la descripción de libros para representar los datos básicos de los registros de libros, y una ontología de personas para describir los autores de los libros, a la manera de un fichero de autoridad.

EDM reutilizará alguna de ontologías de referencia ya existentes, como el estándar Simple Knowledge Organization System (SKOS), apoyado por el W3C. SKOS define un modelo para representar los elementos de un sistema de organización del conocimiento como puede ser un tesoro, un esquema de clasificación o similar. SKOS consta de una clase principal para describir conceptos. Para adaptar a un enfoque de modelado basado en el concepto estándares como el ISO2788, SKOS dispone de propiedades tanto para las designaciones de dichos conceptos (por ejemplo, `skos:prefLabel` para la designación preferente de un concepto, `skos:alt Label` para las alternativas), como para las relaciones semánticas entre los conceptos (`skos:narrower` [término específico], `skos:broader` [término general], `skos:related` [término relacionado]) y para la documentación general de conceptos (`skos:scopeNote` [nota de alcance], `skos:definition` [definición], etc.). Y lo que es más importante, a través de la propiedad `skos:exactMatch`, SKOS está preparado para trazar correspondencias entre esquemas conceptuales diferentes, por ejemplo, enlazar conceptos de diferentes tesauros que están semánticamente relacionados. Técnicamente, esto posibilita a las aplicaciones navegar a través de una capa semántica de conceptos de diferentes fuentes, aprovechando tales redes conceptuales para acceder a objetos que están descritos originalmente utilizando conceptos diferentes conceptos, pero relacionados.

Entra otras ontologías que son de especial interés para EDM se incluye Dublin Core (DC) y Friend-of-a-Friend (FOAF). Dublin Core proporciona un vocabulario compacto para describir las características esenciales de objetos culturales (creadores, relaciones con otros recursos, indización por materias, etc.) de un modo muy acorde con la Web Semántica que se ajusta a un amplio rango de necesidades. DC se utilizó como la base de ESE: manteniéndolo como parte del nuevo modelo hace posible, por tanto, la compatibilidad con todos los datos ya existentes de Europeana. Permite también a las instituciones suministradoras de datos que no deseen proporcionar descripciones más minuciosas de un vehículo simple para los datos que envían. Y, finalmente, permite que los datos en EDM sean más susceptibles de compartir y reutilizar, porque DC se usa en muchas aplicaciones, que podrían adaptarse fácilmente para utilizar los datos en EDM.

FOAF es una ontología utilizada para describir personas en RDF, cuyo origen está en los requisitos de descripción de perfiles web de personas. Podría ser válida, por lo tanto, aunque con algunas adaptaciones o extensiones, para describir las muchas personas que tienen un papel crucial en el contexto de los objetos de Europeana.

3.2 El entorno “Intercambio y reutilización de objetos”

Las representaciones de objetos típicas de Europeana serán entidades compuestas de varias partes, como por ejemplo metadatos de atributos, una vista en miniatura y una página estática en html (solo por poner un ejemplo bastante sencillo). Por esta y otras razones se escogieron las especificaciones de OAI Object & Exchange (OAI-ORE) (<http://www.openarchives.org/ore/1.0/to>) como el marco estructural para la modela de la ontología EDM. OAI-ORE define un enfoque para la identificación y descripción de conjuntos de recursos Web. Con el fin de hacer referencia de modo unívoco a un conjunto de recursos Web, se introduce un nuevo Recurso que representa a una colección de otros Recursos. Este nuevo Recurso, llamado Agregación (*Aggregation*), tiene un URI como cualquier otro recurso en la web. Dado que una Agregación es una construcción conceptual, se modela como un recurso no informativo que no tiene una Representación propia, sino que se describe por otro Recurso, llamado Mapa del Recurso (*Resource Map*); éste tiene también en URI y una representación, entendible por el ordenador, que contiene los detalles sobre la Agregación. En esencia, un Mapa del Recurso transmite lo que la Agregación describe (la relación *ore:describes* de la figura 1), y lista los Recursos Agregados que son parte de la Agregación (la relación *ore:aggregates* de la figura 2, una subpropiedad de *dcterms:hasPart*). Además, un Mapa del Recurso puede expresar relaciones y propiedades que afectan a todos los Recursos Agregados, así como metadatos relativos al propio mapa. Por ejemplo, la figura 2 muestra que se incluyen la autoría y fecha de modificación del Mapa del Recurso (las relaciones *dcterms:creator* y *dcterms:modified*, respectivamente). Un Mapa del Recurso puede también expresar las relaciones que la Agregación, los Recursos Agregados y el propio mapa tienen cualquier otro recurso.

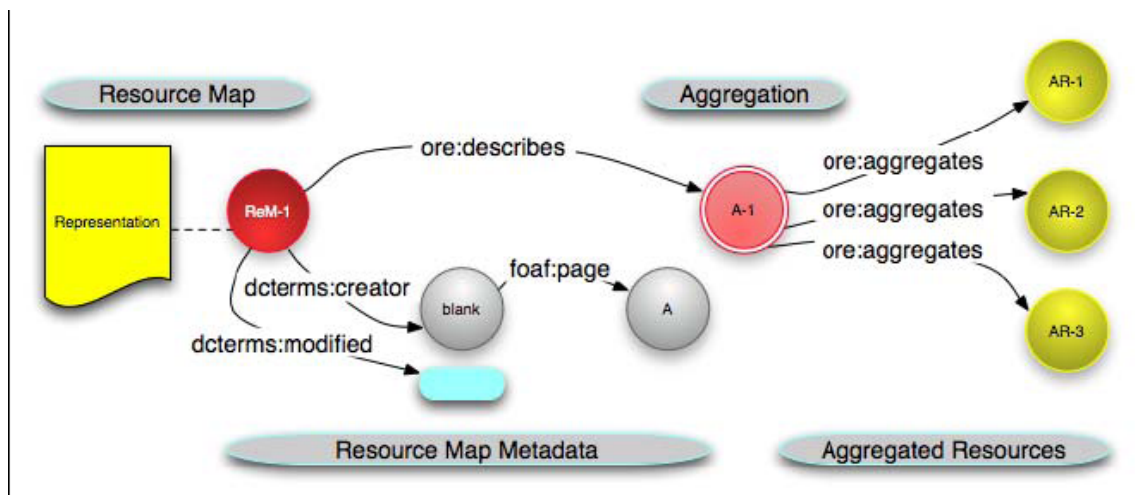


Figura 1: Los componentes básicos del modelo de datos OAI-ORE

Además, el modelo de datos permite expresar que un Recurso Agregado es en sí mismo una Agregación (agregaciones anidadas). Para ello, se expresa una relación *ore:isDescribedBy* (la inversa de *ore:describes*), que es un subpropiedad de *rdfs:seeAlso*, entre el susodicho Recurso Agregado y un Mapa del Recurso que declara que este Recurso Agregado en sí mismo una agregación. Por motivos de navegación, el modelo de datos también permite que un Mapa del Recurso expresa que un Recurso

Agregado de una Agregación dada es también parte de otra Agregación. Esto se consigue mediante la relación *ore:isAggregatedBy* (la inversa de *ore:aggregates*) establecida entre el Recurso Agregado y la otra Agregación. Además, el uso de identificadores no basados en ningún protocolo, que se pueden expresar mediante URIs (como por ejemplo un URN), es bastante común para referirse a bienes de patrimonio cultural. Para apoyar esta práctica, existe la relación *ore:similarTo*, que se establece entre una Agregación y un recurso, de alguna manera equivalente, identificado por una URI no basada en ningún protocolo. La especificidad de *ore:similarTo* se halla entre *rdfs:seeAlso* y *owl:sameAs*. Debe hacerse constar que la comunidad de Linked Data aún debate la necesidad por otras expresiones de equivalencia diferentes de *owl:sameAs* (http://events.linkedata.org/ldow2010/papers/ldow2010_paper09.pdf)

El URI que representa a un Recurso Agregado como elemento dentro de una Agregación particular no es diferente del URI que le identifica como un Recurso independiente de dicha Agregación. Sin embargo, en muchos casos prácticos – incluyendo Europeana – es necesaria una distinción entre referirse a un recurso en sí, y referirse al mismo recurso como Recurso Agregado dentro de una Agregación. La cita en contexto, y el seguimiento de la procedencia de los datos son algunos ejemplos. Para llevar a cabo esta diferenciación, OAI-ORE introducen el concepto de Proxy. Un Proxy es un Recurso que representa a un Recurso Agregado en el contexto de una Agregación específica. El URI de un Proxy proporciona un mecanismo para expresar un Recurso en su contexto. La Figura 2 muestra las relaciones *ore:ProxyFor* y *ore:ProxyIn* entre un Proxy con un Recurso Agregado y con una Agregación, respectivamente. Ilustra también cómo citar el Recurso Agregado es diferente de citar su Proxy: el primero cita el Recurso “como es”, y el segundo lo cita como existe en el contexto de una Agregación específica. Con el fin de que todo esto funcione en la web sin problemas, y para proporcionar la información de contexto a clientes OAI-ORE, la resolución de http URIs asignados a Proxys debe conducir al Recurso Agregado, y la respuesta debe incluir un HTTP Link Header (<https://datatracker.ietf.org/doc/draftnottingham-http-link-header/>) que conduzca a la Agregación.

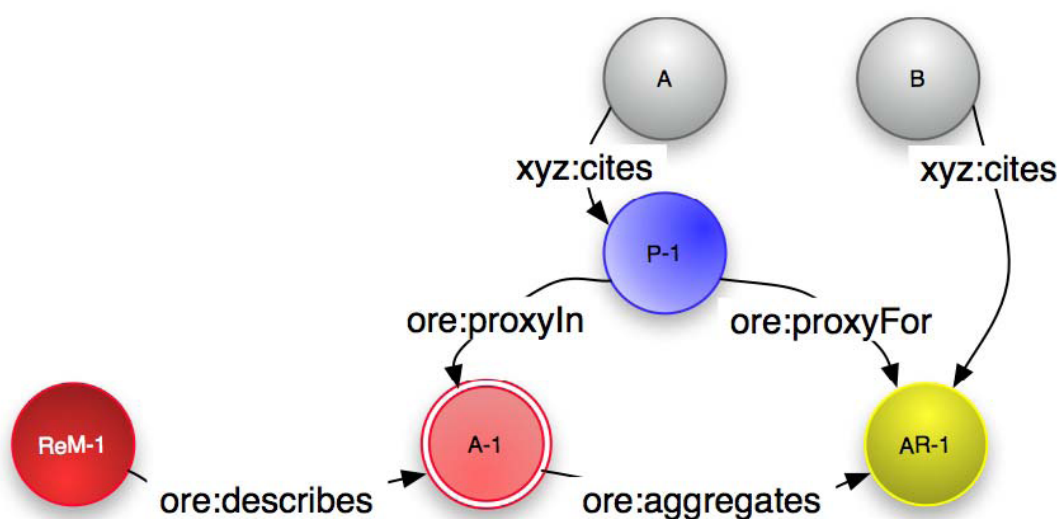


Figura 2: Proxys en OAI-ORE: referenciar a un Recurso Agregado en su contexto

3.3 EDM y Linked Open Data

El objetivo de Linked Data es hacer posible el compartir datos estructurados en la Web. Para este fin, Linked Data tiene en la arquitectura de la web su columna vertebral., y en RDF su lenguaje de representación. Hace recomendaciones sobre cómo hacer accesible en la Web los datos RDF, con la visión general de la web como datos cooperativos que se conoce habitualmente como Web Semántica.

La visión de la Web Semántica ha dado forma, en gran medida, al diseño de EDM. En concreto, ha llevado a la adopción de RDF como el meta-modelo de EDM, y a la decisión de convertir a cualquier objeto de interés en el espacio de Europeana (ya sea un objeto de patrimonio cultural, o una entidad de contextualización como una persona, un lugar, un concepto, etcétera) en un recurso, identificado por un URI HTTP. Esta decisión permite la normalización de los valores en las descripciones de Europeana al formato URI HTTP, un estándar de facto soportado por la arquitectura de la web.

Linked Data añade a esta visión una dimensión fundamental, porque a través de Linked Data Europeana puede utilizar los URIs HTTP es su dominio al mismo tiempo que permite el acceso a descripciones estructuradas de sus correspondientes objetos. Estos link actúan por lo tanto como conectores de la información de Europeana con la información de otras instituciones, permitiendo a Europeana reunir conocimiento adicional sobre personas, lugares, conceptos y otros. No hace falta decir que se espera que este conocimiento juegue un importante papel en la mejora de la usabilidad de Europeana en aspectos tan vitales como el funcionamiento de las funcionalidades de sugerencia y hallazgos de nuevos contenidos por el usuario.

3.4 Abstracción de las relaciones semánticas relevantes

EDM no ciñe la representación de los metadatos aportados a un esquema común, sino que actúa como una ontología común de alto nivel según la cual los metadatos, compatibles entre sí, se expresan según sus modelos originales. Busca amplitud más que seleccionar los campos más comunes. Esto permite integrar las distintas perspectivas y necesidades de información de las diferentes comunidades que aportan datos a Europeana. La riqueza original de estándares específicos como LIDO, CIDOC CRM, MARC o EAD se puede, por tanto, preservar, mientras que el acceso por relaciones implícitas de alto nivel garantizará una recuperación y una precisión nunca antes visto sobre datos aportados en diferentes formatos.

El reto particular era hallar un conjunto de relaciones semánticas que fueran muy relevantes para la estructura y el acceso a la información sobre objetos socio-culturales, y que fueran lo suficientemente abstractas como para abarcar miles de relaciones potenciales más especializadas, pero a la vez expresivas como para cubrir la brecha

existente entre las búsquedas por palabra clave y el acceso a campos específicos de cada disciplina. Los esquemas básicos como DC, VRA y la ontología CIDOC CRM son muy genéricos pero ninguno cubre el otro aspecto satisfactoriamente.

Excluyendo la gestión de la identidad, finalmente distinguimos cinco relaciones semánticas fundamentales más especializadas, por ejemplo, en relación a Dublin Core:

- Clasificación en categorías, que se pueden expresar con SKOS
- La separación de algo en partes, y la incorporación de un recurso de información dentro de otro.
- Semejanza, es decir, la relación entre cosas o recursos de información que comparten características comunes, ya sea por casualidad, por influencia, o por derivación unas de otras, tal y como se describen en FRBR (Doerr y LeBoeuf, 2007).
- Aquellas entidades o ideas a los que una cosa o recurso de información representa, introduce, se refiere o de las que trata. (*Aboutness*).
- Historia de un objeto, es decir, las cosas, personas, lugares, acontecimientos con los que ha tenido contacto, ha coincidido, o ha coexistido. De un modo más analítico, todas las relaciones se pueden explicar o ampliar en función de su presencia con acontecimientos relacionados.

Todos los valores relacionados con estas relaciones se pueden representar, en Linked Open Data, por medio de URIs; por ejemplo, VIAF (www.viaf.org) para personas, repertorios geográficos para lugares. Las personas pueden representarse también en EDM mediante la ontología FOAF y por sus relaciones con acontecimientos.

4. EDM y los esquemas de representación específicos de cada comunidad

RDF, el “lenguaje de la Web”, permite declarar subordinación entre propiedades; en otras palabras, que una relación utilizada para asociar algún objeto con un valor dado implica (es “subpropiedad de”) para dicho valor ciertas relaciones más generales, del mismo modo que lo hacen, en un tesauro, los “términos generales”. Consecuentemente, se puede buscar un objeto que tenga esta asociación, bajo una relación más general, en principio no declarada en los datos. Esta característica de RDF no ha sido aprovechada comúnmente por los formatos de metadatos diseñados por comunidades de usuarios, pero en realidad es el método más importante para integrar esquemas de representación específicas de cada comunidad con otros esquemas más genérico o “básicos”. Por ejemplo, “dcterms:created” implica y precisa “dc:date”. El modelo EDM supone las generalizaciones de propiedades de metadatos más radicales que se ha propuesto hasta ahora.

Europeana preve que cualquier comunidad pueda declarar su perfil de aplicación propio (Heery y Patel, 2000) con el fin de enriquecer la precisión de EDM para su subconjunto particular de datos. Todas las relaciones particularmente relevantes expresadas en dichos perfiles deben construirse, de modo explícito, sobre una o más de las propiedades de EDM, posibilitando así poder ser utilizadas por usuarios que incluso no conozcan la semántica específica de la comunidad. Y, a la inversa, Europeana ha procurado tener en cuenta requisitos de metadatos de las comunidades, es decir, formatos de metadatos, cuya base son grandes conjuntos de datos reales, con los que se puede contribuir a Europeana, y verificar su compatibilidad con EDM. En último extremo, EDM es el resultado de la generalización de todas estas contribuciones.

Las comunidades relevantes para Europeana se pueden subdividir en bibliotecas, bibliotecas digitales, archivos, museos y el sector audiovisual. La comunidad de museos es la que tiene actualmente la mayor diversidad interna en cuanto a sub-disciplinas y formatos de metadatos, en consonancia con la diversidad de objetos físicos que gestionan. Los museos exponen poca cantidad de metadatos en Dublin Core. En los Estados Unidos, el más popular es VRA y CDWA, ya que incluyen los datos de localización física de los objetos. Estos tres últimos formatos están todos contemplados en EDM. Sin embargo, los metadatos de museos plenamente satisfactorios son muy complejos, por lo que se ha alcanzado un amplio acuerdo con CIDOC CRM (ISO21127:2006), que tiene una rica semántica común de elementos de metadatos en un modo compatible con RDF pero que no prescribe ningún campo. No ha sido hasta fecha muy reciente que los principales actores del sector museístico han acordado un formato explícito para recolectar datos de museos, llamado LIDO. Es compatible con CRM y ya está soportado para la producción de grandes conjuntos de datos en el proyecto europeo ATHENA. LIDO nace como una reformulación de CDWA-Lite centrado en acontecimientos. Los datos LIDO se pueden convertir a EDM, pero también a CRM, que es un lenguaje mucho más rico. Otros cientos de formatos de otros museos se pueden convertir también a CRM, que es un formato integrado en EDM. Por todo ello, los representantes de la comunidad museística han votado por CRM como el perfil de

aplicación para museos bajo EDM. Con metadatos en CRM se pueden producir también metadatos en Dublin Core (Kakali et al. 2007)

Para la comunidad archivística, las descripciones a nivel de colección como EAD juegan un importante papel, lo cual encaja perfectamente dentro de EDM, en particular el uso que *ore:aggregation* permite para describir el concepto de “fondo” de archivo. El Consejo Internacional de Archivos acaba de empezar la discusión acerca un modelo conceptual similar a FRBR o a CRM. Entretanto, con CRM se pueden describir con más detalle que con EDM los hechos históricos asociados con los contenidos archivísticos (Stasinopolou et al. 2007). Además, las descripciones a nivel de colección en Dublin Core son bastante apropiadas y cada vez son más comunes para las descripciones archivísticas. En cuando al sector bibliotecario, el formato dominante es aún MARC, pero se pueden aportar metadatos en el formato MODS, más simple, y podría ser utilizado como perfil de aplicación bajo EDM, pero Dublin Core tiene ya un parecido muy cercano a los registros MODS. Desde finales de los 90, con FRBR, el sector bibliotecario ha enfocado su interés hacia metadatos más ricos. Aunque hay ya varias implementaciones de FRBR, la semántica precisa de “Obras” y “Expresiones” es aún muy controvertida y la aceptación práctica en las bibliotecas va aún muy atrasada. Sin embargo, no existe disputa acerca de la necesidad de agregación de contenidos junto con jerarquías derivativas, y puede ser implementado evitando la controversia de los niveles de agregación. FRBRoo, la interpretación ontológica de FRBR (Doerr y LeBoeuf, 2007), ha identificado relaciones básicas, asumida en la propiedades de “similaridad” de EDM. Por ello, EDM puede representar las nociones básicas no conflictivas de FRBR.

5. Enfoque centrado en el objeto y en evento.

De modo natural, el diseño de los esquemas de metadatos se ha concentrado mayormente en asignar atributos directamente al objeto de la colección a la manera de “finding aids”. Tras este enfoque está una tradición de siglos de sabia y exitosa biblioteconomía, basada en el paradigma que el usuario conoce el tema sobre el que trata el objeto, o una propiedad material del mismo, o cualquier asociación del objeto con otro con el que esté relacionado, como su autor o editor. Se asume igualmente que una vez que el usuario accede al objeto, sus necesidades de información se satisfacen con el objeto en sí. Este paradigma se puede, y ha sido transferido completamente a colecciones digitales y, aún más allá, si tenemos en cuenta que además de recuperar los metadatos, también se puede buscar por el contenido.

Con el tiempo nuestros sistemas de información se han vuelto más poderosos, y la gente se da cuenta que los metadatos tienen, por sí solos, un enorme valor como información documental. Al mismo tiempo, con el extraordinario tamaño de las colecciones modernas, la creciente sinonimia limita significativamente el poder de los meros atributos para identificar cosas. Finalmente, se han empezado a describir grandes colecciones de objetos “no verbales”, como imágenes o los objetos de un museo, cuya comprensión depende completamente de los metadatos.

Todo lo cual ha originado un interés creciente en guardar un registro más expresivo y más coherente de la proveniencia y de la historia de los objetos. Desde mediados de los 90, varias comunidades comenzaron a entender (IndeCs, CRM, ABC, OPM) que la complejidad de describir relaciones de tipo histórico puede ser normalizada y radicalmente simplificada si, en vez de colocar a los objetos en el centro de la descripción, situamos en su lugar los acontecimientos históricos, que son “mediadores” en cierto sentido de todas las relaciones dinámicas entre persona, cosas, tiempo y espacio. Además de proporcionar más detalles, permite detectar con gran precisión los objetos que están relacionados a través de una historia común, esto es, comparten una participación en dichos acontecimientos.

Describir los acontecimientos explícitamente puede, en el peor de los casos, duplicar el tamaño de algunos metadatos con la inclusión de, a menudo, múltiples atributos pertenecientes a un evento compartido (como “fecha de creación, lugar de creación, creador”), pero su poder expresivo se incrementa considerablemente. Por ejemplo, sin una descripción de acontecimientos no se podría hacer una búsqueda precisa de objetos antiguos de Egipto importados a Crete en la Edad del Bronce. La documentación de acontecimientos es un requisito de la comunidad de patrimonio cultural. Las descripciones, del mismo ejemplar, centradas en el objeto o centradas en el acontecimiento, se pueden transformar de una en otra con una pérdida controlada de precisión en el caso de las centradas en el objeto. El modelo Europeana representa estas transformaciones internamente. El equilibrio entre complejidad y poder expresivo de ambos aspectos se puede por tanto ajustar en función de las necesidades del usuario y la calidad de la fuentes de datos.

Es este sentido, la relación “ha coincidido” es una poderosa innovación de EDM. Sirve de enlace de modo muy elegante entre los modelos “centrados en el objeto” y “centrados en el acontecimiento”, y amplía el atributo general Dc:date a las otras categorías fundamentales de recursos no informativos (acontecimientos, agentes, lugares, objetos materiales e inmateriales), consiguiendo con ello una cobertura semántica que no es posible con ningún otro estándar hasta el momento. Esta cobertura garantiza la extensibilidad, sin pérdida en la capacidad de acceso, mediante las relaciones semánticas fundamentales implícitas en la mayoría de campos de datos especializados.

6. Validación y potencial de EDM

Con la número 5, EDM ha alcanzado una versión estable. El grupo de expertos consideró que sería necesario en este momento comenzar a evaluar y validar el modelo con el fin de incluirlo en el lanzamiento de la versión “Danubio” de Europeana; la validación debía basarse en ejemplos reales desde una variedad de diferentes perspectivas.

Por ellos, Europeana organizó cuatro “reuniones de comunidades” con representantes de archivos, archivos audiovisuales, bibliotecas y museos. Cada comunidad aportó una muestra de datos de sus colecciones. El objetivo era averiguar si los diferentes estándares de cada comunidad podían ser satisfactoriamente mapeados en EDM. Los resultados fueron muy esperanzadores. Desde luego, cada comunidad anotó diversos detalles a mejorar, pero fundamentalmente se referían a la presentación y recuperación de los datos. Como ontología común de alto nivel, EDM se mostró muy flexible y estable y capaz de acomodarse a las clases y propiedades de cada comunidad. Esta serie de encuentros concluirán con una reunión de expertos en Pisa en junio donde se incorporarán a EDM las propuestas de las comunidades.

6.1 Validación

La comunidad de archivos proporcionó archivos de ejemplos de *finding aids* para material archivístico, codificados en EAD. La característica distintiva de estas descripciones archivísticas es la profunda estructura jerárquica y el potente enfoque en descripciones pormenorizadas y con mucha información contextual. Las propiedades de EDM para la descomposición en partes y la incorporación a un todo demostraron su potencia para tratar descripciones de colecciones que contienen varios niveles de sub-descripciones donde cada nivel intermedio contiene información contextual.

Los museos proporcionaron principalmente ejemplos codificados en *museumdat* y *LIDO*. El enfoque, muy basado en eventos, de museumdat/LIDO encaja muy bien en EDM. Sus clases y propiedades basadas en eventos ofrecieron posibilidades de modelados lo suficientemente flexibles para integrar las ricas descripciones basadas en eventos de LIDO por medio de la creación de eventos, con sub-clases y sub-propiedades como especialización de las clases y propiedades de EDM. La comunidad de museos, de todos modos, sugirió reemplazar, cuando sea posible, para el perfil de aplicación de museos, las clases y propiedades de EDM por las de CRM, y utilizar las entidades de CRM que no tienen un equivalente en EDM.

Los archivos audiovisuales constituyen una comunidad muy heterogénea que aportan objetos muy diversos y que aplican estándares de codificación diferentes. En el dominio audiovisual, la entidad que se describe y su naturaleza son, a menudo, difíciles de identificar. No hay un enfoque claro como ocurre con museos (LIDO) o archivos (EAD) en la comunidad audiovisual y el material que trata contienen muchos más objetos nacidos digitales. Sin embargo, EDM fue capaz de integrar la diversidad y riqueza de los ejemplos aportados.

La comunidad bibliotecaria aportó un número de ejemplos, en ocasiones muy complejos, que fueron satisfactoriamente mapeados a EDM. Resultó evidente, no obstante, que el modelado de datos bibliotecarios en EDM se beneficiaría enormemente de una extensión del modelo que incluyera la categorización en obra, expresión, manifestación y ejemplar del grupo 1 de FRBR. Los expertos bibliotecarios convinieron que la introducción de RDA, una vez esté operativo, intensificaría substancialmente la necesidad de incluir categorías FRBR, quizá como parte del perfil de aplicación de la comunidad.

La conclusión de todos los encuentros de las comunidades fue que EDM, una vez validado e implementado, será una buena herramienta integradora para modelar los diferentes objetos culturales de muy distinta procedencia.

6.2 Potencial

El nuevo “Modelo de datos de Europeana” (EDM) reemplazará los “Elementos Semánticos de Europeana” (ESE) que actualmente subyace en el espacio de datos de Europeana. EDM ofrece una mayor expresividad y flexibilidad y permite representación más rica y verdaderamente semántica de los millones de objetos de todo tipo de las comunidades de patrimonio cultural en Europeana.

En comparación con anteriores modelos de datos EDM consigue un alto nivel de abstracción. Es la generalización de propiedades de metadatos más radical hasta el momento en el área del patrimonio cultural y no constriñe la representación de los metadatos introducidos a un esquema común. EDM integra ontologías bien establecidas como SKOS, Dublin Core y FOAF con el fin de permitir descripciones ricas e interoperables de objetos de Europeana. Como ontología de alto nivel permite integrar las distintas perspectivas de información y las necesidades de las diferentes comunidades que aportan datos a Europeana y conserva la riqueza original de los estándares de dichas comunidades, como LIDO, CIDOC CRM, MARC o EAD.

EDM utiliza RDF(s) como su meta-modelo y URIs para identificar información estructurada acerca de los objetos de patrimonio cultural. Las especificaciones de OAI Object Reuse & Exchange (OAI-ORE) proporcionan el marco estructural del modelado de la ontología EDM. Esta arquitectura abierta de EDM hace a Europeana compatible con el paradigma de la web semántica y le permite ser parte de la emergente comunidad de Linked Open Data. De hecho, EDM supone, para las instituciones relacionadas con el patrimonio cultural, una vía para la migración sus arquitecturas de información, actualmente cerradas, a otros entornos abiertos e interrelacionados – para beneficio tanto de dichas instituciones como para la comunidad entera de Internet.

El intento de validación de los encuentros de comunidades demostró que EDM tiene el potencial de funcionar con éxito como una ontología de alto nivel para muchos tipos diferentes de modelos de datos más especializados, proveniente de dominios diferentes. La posibilidad de las comunidades de acordar un perfil de aplicación propio en EDM para incrementar la precisión de sus colecciones particulares de datos, debería permitir la integración de cualquier modelo de datos sin pérdida de datos.

Por último, EDM ofrece nuevos escenarios de uso para los diversos datos integrados en Europeana. Por ejemplo, la arquitectura basada en EDM permite la contextualización de representaciones de objetos creando nuevos escenarios heurísticos para expertos – más específicamente dentro de las “humanidades digitales”-, pero esto sería objeto de otro artículo...

7. Bibliografía

- Cesare Concordia, Stefan Gradmann, Sjoerd Siebinga (2010): Not just another portal, not just another digital library: A portrait of Europeana as an application programinterface. En: International Federation of Library Associations and Institutions 36(1), pp. 61–69. (<http://dx.doi.org/10.1177/0340035209360764>)
- Martin Doerr, Patrick LeBoeuf (2007): Modelling Intellectual Processes: the FRBR - CRM Harmonization. En: C. Thanos, F. Borri, and L. Candela (eds.): Digital Libraries: R&D, LNCS 4877, pp. 114–123, 2007. (First DELOS Conference on Digital Libraries, February 2007 Tirrenia, Pisa, Italy)
- Thomas R. Gruber (1993): Toward Principles for the Design of Ontologies Used for Knowledge Sharing. En: Nicola Guarino and Roberto Poli (eds.): Formal Ontology in Conceptual Analysis and Knowledge Representation.
- Constantia Kakali, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies. En: Proceedings of the International Conference on Dublin Core and metadata Applications (DC-2007), pp.128-139. 27-31 Agosto, 2007. Singapur.
- Rachel Heery, Manjula Patel (2000): Application profiles: mixing and matching metadata schemas. Ariadne 25. (<http://www.ariadne.ac.uk/issue25/appprofiles/intro.html>)
- Barry Smith (2003): Ontology. En: L. Floridi (ed.): Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, pp. 155–166.
- Thomais Stasinopoulou, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Ontology-based Metadata Integration in the Cultural Heritage Domain. En: Proceedings of the 10thInternational Conference on Asian Digital Libraries, pp.165-175. Hanoi, Vietnam, 10-13 Diciembre, 2007.