



ESTUDIOS / RESEARCH STUDIES

Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España

Andreu Sulé*, Miquel Centelles*, Jorge Franganillo*, Jesús Gascón*

*Universitat de Barcelona. Departament de Biblioteconomia i Documentació
Correos-e: sule@ub.edu, miquel.centelles@ub.edu, franganillo@ub.edu, gascon@ub.edu

Recibido: 25-02-2015; 2ª versión: 30-06-2015; Aceptado: 09-07-2015.

Cómo citar este artículo/Citation: Sulé, A.; Centelles, M.; Franganillo, J. y Gascón, J. (2016). Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. *Revista Española de Documentación Científica*, 39(1): e121. doi: <http://dx.doi.org/10.3989/redc.2016.1.1268>

Resumen: Se analiza cómo y en qué grado se aplica el modelo RDF en las principales colecciones digitales españolas de materiales patrimoniales. Se introduce este modelo y también las iniciativas de Datos Abiertos y Datos Enlazados. A continuación se examinan 51 repositorios digitales y de cada uno de ellos se determina si expresan sus registros en RDF, ofrecen un punto de consulta SPARQL consultable por agentes externos y si usan referencias como valor de las propiedades. A partir de los resultados se describen los modelos EDM de *Europeana* y *OntoWeb*. Se concluye que la aplicación de RDF es desigual, demasiado condicionada al uso de aplicaciones que convierten automáticamente los registros en tripletes RDF, que se dan pocos puntos SPARQL para la consulta externa y que el empleo de referencias se vincula al uso de aplicaciones con diferentes modelos: EDM u *OntoWeb*. Se recomienda que las colecciones han de enriquecer sus datos y determinar niveles de agregación de los datos RDF generados para difundirse, hacerse accesibles y adaptarse a la web semántica.

Palabras clave: RDF; modelos de datos; repositorios patrimoniales; consulta SPARQL; referencias; URI; *Europeana Data Model*; *OntoWeb*; datos abiertos; datos enlazados.

Implementation of the RDF data model in digital collections of Spanish libraries, archives and museums

Abstract: The article discusses how and to what extent the RDF data model is applied in major Spanish digital collections of heritage materials. This model, as well as Open Data and Linked Data initiatives, are introduced. Fifty-one digital repositories were analysed to determine whether they expressed their records in RDF, offered SPARQL query points searchable by external agents, and used references as property values. The *Europeana EDM* and *OntoWeb* models are also described. It is concluded that the use of RDF is unequal and excessively conditioned by the use of applications that automatically convert records into RDF triples. Few of the collections analysed give SPARQL points for external queries. Also, the use of references is linked to applications using different models: EDM or *OntoWeb*. Collections should enrich their data and define aggregation levels for generated RDF data in order to be disseminated, made accessible, and adapted to the semantic web.

Keywords: RDF; data models; heritage repositories; SPARQL query; references; URI; *Europeana Data Model*; *OntoWeb*; open data; linked data.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0.

1. INTRODUCCIÓN

1.1. La web semántica y el modelo de datos RDF

Las bibliotecas, archivos y museos y, más concretamente, sus depósitos digitales, atesoran generalmente ingentes cantidades de datos (organizados en forma de documentos) y metadatos (organizados en registros bibliográficos y de autoridad), que presentan altos niveles de estructuración, de tratamiento desde el punto de vista semántico y de interoperabilidad. Más allá de objetivos vinculados a la conservación y preservación de sus fondos, estos tres tipos de instituciones necesitan darlos a conocer y hacerlos más accesibles a otras instituciones similares, pero sobre todo a sus usuarios, sean particulares, instituciones o empresas. Todos estos factores resultan especialmente interesantes para la web semántica, y, más específicamente, la web de datos y su propósito de hacer los contenidos más accesibles (leíbles y comprensibles) para las máquinas.

La web semántica pone a disposición de bibliotecas, archivos y museos un modelo de datos para la representación de información en la Web, el Resource Description Framework, RDF¹, y unas tecnologías que, en conjunto, permiten dotar de significado a los datos y a las transacciones de datos en la Web. Todos estos componentes han sido estandarizados en forma de normas, especificaciones, recomendaciones, etc. por el W3C².

Una declaración RDF constituye la forma más simple de expresión de un metadato, mediante grafos de tripletas compuestas por un sujeto –la cosa, el recurso, sobre el que se declara algo–, un predicado –el aspecto que se describe– y un objeto –el valor de lo que se describe sobre el sujeto–. Los recursos que constituyen el sujeto y el predicado tienen que ser designados por identificadores uniformes de recursos, URI³, mientras que el recurso que se sitúa en el objeto de la declaración puede ser expresado mediante URI o literales⁴. Además, los recursos de los tres componentes de las declaraciones pueden vincularse a ontologías, conceptualizaciones explícitas de ámbitos concretos de la realidad, expresados por lenguajes normalizados –RDF Schema, SKOS, OWL⁵, y formados, como componentes básicos, por conceptos y relaciones entre conceptos. Al vincular los recursos situados en sujetos y objetos de declaraciones RDF a los conceptos y los predicados a las relaciones, el significado de todos ellos puede elevarse a un marco de representación más amplio que la expresión concreta de una declaración, puede ser compartido por diferentes productores y consumidores de datos, y puede ser sometido a restricciones se-

mánticas expresadas formalmente, muy útiles en el momento que son procesadas por los sistemas automáticos.

Por otro lado, la estructura de grafo permite el procesamiento de la información existente independientemente de que sea completa o detallada o del grado de estructuración de sus fuentes. Al ser una estructura simple, no está sujeta a los cambios de esquema o de formato habituales en las actualizaciones de bases de datos relacionales, y permite el fácil intercambio de datos entre sistemas o versiones diferentes, sin pérdida de datos.

1.2. Comparativa con otros modelos/sintaxis que compiten con RDF

Existen diferentes formatos, o sintaxis, que permiten la codificación de las tripletas que constituyen un grafo RDF y su publicación como datos enlazados. Son los llamados formatos de serialización de RDF, algunos de los cuales han sido normalizados por el W3C.

- Sintaxis XML para RDF, o, simplemente, RDF/XML⁶.
- Terse RDF Triple Language, abreviado Turtle, que permite expresar un grafo RDF en una forma textual compacta. TriG es una extensión de Turtle para representar conjuntos de datos RDF completos⁷.
- RDFa que permite incorporar datos estructurados y semánticamente enriquecidos de acuerdo con el modelo RDF en código HTML y XHTML, mediante un conjunto de atributos⁸.

A priori, la elección de un patrón específico de publicación de datos enlazados determina la preferencia por un formato de serialización específico. Así, en los casos en que la provisión de los datos se realiza incrustados en ficheros HTML, RDFa es la solución idónea. Para el resto de patrones de publicación, entre los que podemos situar la descarga de ficheros (*dumping*), la generación de grafos desde bases de datos relacionales, o desde almacenes de tripletas, etc., rivalizan RDF/XML, Turtle, y otros⁹. Turtle ha tenido éxito por la facilidad de escritura y lectura para las personas. No obstante, cada sistema aplica, de origen, un formato “nativo” específico. En cualquier caso, es importante tener en cuenta la existencia de mecanismos de mapeo y transformación entre formatos.

Más allá de los formatos de serialización, se han generado otros modelos de datos diferentes a RDF, que permiten representar información estructurada y semánticamente rica sobre los datos. Entre los que mayor éxito han alcanzado encontramos

los Microdatos y JSON-LD, ambos impulsados y normalizados por la W3C. A priori, ambos pueden contemplarse como competidores del modelo RDF en la consecución de la web de datos semánticamente enriquecidos.

Los Microdatos¹⁰ son grupos de pares nombre-valor —conocidos como ítems—, que se incorporan en elementos HTML5 como atributos, para describir a qué tipo de entidad hace referencia el contenido de dichos elementos, y qué propiedades se describen sobre dichas entidades. Los valores de las propiedades son, mayoritariamente, segmentos del propio contenido. Existen múltiples vocabularios para la asignación de las entidades y propiedades de las entidades; entre los más conocidos y utilizados se encuentra Schema.org, impulsado por los grandes buscadores generalistas de la web. La facilidad de aplicación de esta tecnología, que puede incorporarse en el flujo de trabajo general de la edición de contenidos en la web, y su potencialidad en la mejora de los resultados de las búsquedas en la web y del posicionamiento de los contenidos, han sido importantes factores de impulso. No obstante, en el momento actual, estas potencialidades de los Microdatos vinculadas a la búsqueda semántica y al posicionamiento web no se han desarrollado completamente. Por otro lado, los Microdatos no son legibles, en origen, por analizadores de datos enlazados y si bien se han desarrollado herramientas para la extracción de Microdatos desde HTML y su reconversión al modelo RDF¹¹, el flujo de trabajo en un contexto real puede complicarse de forma considerable.

JSON-LD¹² es una alternativa a los formatos de serialización de RDF que está ganando terreno de forma rápida en la publicación de datos enlazados. Esto se debe a su doble naturaleza. Por un lado, está basado en el modelo de datos JSON, lo que le confiere la capacidad de modelar los datos en forma de texto comprensible por humanos (pares atributo-valor), para transmitir datos entre servidores y aplicaciones web, de forma independiente a cualquier lenguaje de programación específico. Por otro lado, su modelo de datos es, en gran medida, compatible con RDF; las limitaciones del original JSON han sido superadas proporcionando apoyo a la referencia a recursos mediante el uso de URI, y proporcionando un mecanismo de vinculación de términos de vocabularios a sus URI correspondientes¹³. La diferencia fundamental se encuentra en el hecho que JSON-LD admite tripletes con sujetos, predicados y objetos expresados por URI, nodos vacíos y literales en cualquiera de las tres posiciones¹⁴. JSON-LD está sirviendo como pasarela para la incorporación de datos enlazados en

contextos críticos para su desarrollo y éxito; como son entornos JavaScript, servicios Web y bases de datos NoSQL, sin que se ponga en peligro la continuidad de RDF, ya que, siguiendo la tendencia de otros formatos y modelos de datos, se han desarrollado especificaciones para la interpretación de JSON-LD como RDF, y para la serialización de RDF como JSON-LD.

Podemos concluir que la selección de modelos de datos diferentes, aunque muy próximos, a RDF, como son Microdatos y JSON-LD, para el desarrollo de aplicaciones de datos enlazados está profundamente vinculada a preferencias de contextos y comunidades de uso. Lo que pareciera, a priori, una amenaza puede verse como una oportunidad de expansión, apoyada por la posibilidad de transformación entre todos los modelos de datos y formatos de serialización.

1.3. RDF, Open Data y Linked Data

Para la publicación —y, por lo tanto, también el consumo— de datos RDF resultan fundamentales dos iniciativas complementarias: Datos Abiertos (Open Data) y Datos Enlazados (Linked Data)¹⁵. La primera iniciativa no prescribe, *per se*, la adopción del modelo de datos RDF. Se trata de una apuesta por la interoperabilidad de los datos desde una perspectiva legal. De acuerdo con la definición de Open Bibliographic Data¹⁶, los Datos Abiertos son, en esencia, usables, reusables y redistribuibles libremente y se manifiestan mediante la aplicación de licencias y formatos abiertos. Esta iniciativa ha tenido un impacto decisivo en el ámbito de la administración pública, cuyo marco jurídico fue definido por primera vez a escala europea con la *Directiva 2003/98/CE*¹⁷, y en España, con su transposición al derecho propio en la *Ley 37/2007*¹⁸. Es importante tener en cuenta que en estas primeras disposiciones, los datos y metadatos de bibliotecas, archivos y museos quedaban explícitamente excluidos del amparo de la regulación. No obstante, la revisión de aquella primera directiva mediante la *Directiva 2013/37/UE*¹⁹ ha enmendado esta situación al considerar los beneficios que pueden aportar los datos generados en esas instituciones, no sólo desde un punto de vista cultural y educativo, sino también en oportunidades de negocio. Esta directiva debe transponerse al derecho propio de España, como muy tarde el 18 de julio de 2015, y, en el momento de elaborar este artículo, el *Proyecto de Ley por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público* ha sido publicado en el *BOCG*²⁰ y se ha encomendado su aprobación por el organismo legislativo.

La modificación del marco normativo de los Datos Abiertos en España va a tener implicaciones relevantes en relación con el desarrollo de esta iniciativa para las Administraciones y organismos del sector público, en general, y, especialmente, para las bibliotecas, archivos y museos.

Por un lado, los nuevos textos normativos abren explícitamente el marco de actuación de los Datos Abiertos tanto a los documentos —o datos— como los metadatos que estos tienen vinculados. En lo que se refiere a los formatos, establecen que, siempre que sea posible, tanto los datos como los metadatos deben ser proporcionados en formatos abiertos y legibles por máquina, y deben cumplir estándares y normas abiertas. Además, los metadatos deben ofrecerse con los niveles más elevados de precisión y desagregación.

Por otro lado, es importante tener en cuenta que una motivación fundamental para la incorporación de estas instituciones al marco normativo de los Datos Abiertos es la constatación de que sus fondos, y, especialmente, los que resultan de proyectos de digitalización, son clave para el desarrollo de productos y servicios en diferentes sectores empresariales (se citan, concretamente, el aprendizaje y el turismo). De hecho, el proyecto de ley española insta a las Administraciones Públicas a velar por la reutilización de los datos para fines tanto comerciales como no comerciales (artículo 4.2), e incorpora medidas de impulso y flexibilización del régimen administrativo de la reutilización de los datos y metadatos (diversificación de las modalidades de reutilización, limitación de las condiciones incorporadas en las licencias, etc.)

Especialmente relevante, en relación con el impulso de la reutilización de datos y metadatos, es el requerimiento a las instituciones a adoptar una actitud proactiva en lo que se refiere a los mecanismos de búsqueda y recuperación de los datos por parte de los usuarios finales. En concreto, el artículo 4.5 del proyecto de ley española prevé medidas que comprometen a todas las Administraciones y organismos del sector público: la creación de sistemas de gestión documental enlazados con sistemas ofrecidos por otras administraciones que faciliten la recuperación de los datos por parte de los ciudadanos, y la provisión de herramientas de búsqueda de datos abiertos para su reutilización. Adicionalmente, ordena a la Administración General del Estado mantener “un catálogo de información pública reutilizable, correspondiente al menos al ámbito de la Administración General del Estado y a sus organismos públicos vinculados o dependientes.” (*Proyecto de Ley por la que se modifica la Ley 37/2007*, p. 5).

La segunda iniciativa a la que hemos hecho referencia, Datos Enlazados, aborda la perspectiva técnica de la interoperabilidad y es una apuesta clara por el modelo de datos RDF y por otras dos tecnologías de la web semántica: los URI y el lenguaje de búsqueda en datos RDF denominado SPARQL²¹. Ya son clásicos los cuatro principios rectores de este método de exponer, compartir e interconectar datos en la Web enunciados originalmente por Berners-Lee (2006):

1. Deben utilizarse URI para identificar los recursos en cada uno de los componentes de las declaraciones RDF: en el sujeto, en la propiedad y, con prioridad sobre la alternativa de los literales, en el objeto.
2. Deben utilizarse URI creados de acuerdo con el protocolo HTTP, de forma que puedan ser consultados y desreferenciados en la web por las personas y, sobre todo, por los sistemas automáticos.
3. Debe proporcionarse información útil sobre los recursos identificados con los URI desreferenciables, para cuando alguien los consulte. Con esta finalidad, es necesario utilizar estándares de la web semántica, como RDF y SPARQL.
4. Deben establecerse enlaces con otros recursos (utilizando sus URI) en el momento de publicar datos en la web, de forma que se puedan descubrir más datos.

La publicación de datos enlazados significa la creación de conjuntos de datos (*data sets*) en el contexto de una o más entidades, donde se agregan y mantienen las declaraciones RDF, formando un grafo global, sin los límites de los registros (bibliográficos y de autoridades) y catálogos locales donde se generaron inicialmente. De esta forma, están listos para:

- Enlazarse, en origen, con datos RDF de otros depósitos digitales e incluso de otros conjuntos de datos ajenos a las colecciones documentales.
- Ser enlazados por conjuntos de datos externos que quieran ser enriquecidos.
- Ser consumidos directamente por aplicaciones que los necesiten para la generación de nuevos recursos y servicios, mediante descargas masivas (*dump*) o mediante consultas selectivas de datos con el uso de tecnologías de consulta de datos RDF, ajenas al paradigma de la web semántica (RESTful) o específicamente diseñadas para ésta (puntos de acceso SPARQL).

Los beneficios de la aplicación del modelo de Datos Abiertos Enlazados²² a los datos de institucio-

nes del patrimonio cultural han sido señalados por el W3C Library Linked Data Incubator Group²³. El ahorro de costes en el enriquecimiento de los datos que se ponen a disposición de los usuarios, el incremento de la difusión y accesibilidad de las colecciones y, más concretamente, su (re)utilización en proyectos culturales y educativos y también en iniciativas empresariales. En conjunto, es una buena oportunidad para impulsar la visibilidad de estas colecciones y su posicionamiento en el ámbito de los recursos de información. También es verdad que la adaptación al modelo de datos RDF y la iniciativa de publicación en forma de Datos Abiertos Enlazados imponen unas condiciones técnicas, legales y políticas que será necesario abordar, en especial en el momento en que los datos y metadatos de bibliotecas, archivos y museos deban incorporarse al conjunto de los datos abiertos de la administración del estado.

El modelo se ha venido aplicando y se ha impuesto como paradigma en la comunidad que trabaja en bibliotecas digitales. Como ejemplos, a parte de los de centros españoles que se comentan más adelante, se puede citar el *Zentrale Verzeichnis Digitalisierter Drucke*²⁴, depósito centralizado de 108 colecciones digitales alemanas, mayoritariamente de fondos patrimoniales y antiguos, con todos los metadatos recogidos e integrados en una base de datos única de metadatos para acelerar las búsquedas. Entre estas colecciones destaca la *Digitale Bibliothek* del Münchener Digitalisierungszentrum (Bayerischen Staatsbibliothek)²⁵, con cerca de 907.500 títulos digitalizados. Ambas trabajan con el protocolo OAI y recopilan metadatos Dublin Core y METS. También son destacables las colecciones digitales de todo tipo de documentos desarrolladas por British Library Labs²⁶, muchas de ellas de valor patrimonial. En este caso, la biblioteca ofrece libremente el acceso a los metadatos en diferentes formatos, según cada colección, entre ellos RDF/XML básico, OWL o SKOS con acceso SPARQL (por ejemplo, en la *British National Bibliography*²⁷).

1.4. Herramientas y recursos de interés para explotar las posibilidades de la web semántica

Existen diferentes iniciativas orientadas a facilitar la exploración de las oportunidades y beneficios de publicar datos de repositorios como datos enlazados.

El ciclo de vida de los datos enlazados comprende un conjunto de fases cíclicas, que incluyen la creación de datos, el enlace (reconciliación) con datos RDF externos, el modelado mediante vocabularios/ontologías, y la publicación para facilitar su consumo. Diferentes metodologías han sido

propuestas para concretar y sistematizar dichas fases. En paralelo a la definición del ciclo de vida, se ha desarrollado un sector de herramientas informáticas para la gestión y automatización de fases aisladas, o de varias de ellas de forma conjunta. Una fuente interesante para la identificación y selección de estos tipos de herramientas es Semantic Web Development Tools²⁸ un directorio anotado que es mantenido por la comunidad de usuarios del wiki sobre web semántica del W3C.

Open Refine²⁹ es una herramienta que permite el tránsito por diversas fases del ciclo de vida de los datos enlazados, y que, si bien no ha sido desarrollada para la gestión específica de datos de bibliotecas, archivos y museos, ha tenido un importante impacto en estos ámbitos. A priori, Open Refine facilita la limpieza de grandes cantidades de datos en formatos estructurados, y su enriquecimiento mediante consultas a bases de datos externas. Mediante la incorporación de extensiones como RDF Refine, es posible enlazar los datos originales con datos RDF de conjuntos externos, y transformarlos a RDF mediante el mapeo con vocabularios/ontologías generadas "ad hoc", o importadas. A partir de aquí, es posible exportar los datos transformados a herramientas de explotación e integración de datos RDF.

En el ámbito específico de las bibliotecas y los museos, cabe destacar ALIADA³⁰, que estará disponible como prototipo final en octubre de 2015. Es una aplicación que facilita el tránsito por todas las fases del ciclo de vida de los datos enlazados, desde la transformación de datos de origen en formatos MARC y LIDO, a datos RDF a través de ontologías específicas de esos dos ámbitos (FRBRoo para datos de biblioteca y CIDOC-CRM para datos de museos), hasta su publicación en diversos tipos de plataformas, como, por ejemplo, CKAN. En el momento de escribir este artículo, es posible la descarga e instalación de la versión ALIADA v. 1.0.

Además de estas herramientas, es importante tener en cuenta casos de uso representativos de diferentes recetas de publicación de datos enlazados a partir de colecciones de bibliotecas, archivos y museos. Ya el W3C Library Linked Data Incubator Group realizó una recopilación de estos casos de uso que se incluye en su informe final³¹. Más recientemente, Mitchell (2013) explora tres iniciativas de publicación de datos enlazados orientadas a las comunidades del patrimonio cultural: *Europeana*³², *The Digital Public Library of America*³³, y *BIBFRAME*³⁴, una iniciativa liderada por la Library of Congress. El valor de estos casos es que el análisis de sus similitudes y diferencias permiten inferir buenas prácticas para la aplicación en otros proyectos del sector.

2. OBJETIVOS

Dadas la relevancia de las tecnologías de la web semántica, su relación con los Datos Enlazados y su importancia para las colecciones digitales de entidades patrimoniales, cabe preguntarse cuál es su grado de utilización actual en bibliotecas, archivos y museos. Este trabajo pretende analizar la aplicación del modelo de datos para metadatos RDF en las principales colecciones digitales españolas con una orientación patrimonial y, con ello, ofrecer una visión panorámica del grado de implementación de esta tecnología en las colecciones digitales españolas y, por consiguiente, de su mayor o menor adaptación a la web semántica.

3. METODOLOGÍA

Para evaluar el nivel de aplicación de las tecnologías de la web semántica se ha examinado un conjunto de repositorios con el objetivo de determinar en qué medida expresan sus registros en RDF, ofrecen un punto de consulta SPARQL consultable por agentes externos, y usan referencias como valor de las propiedades, ya sea mediante un URI local o uno externo. En primer lugar, se ha identificado qué programa se emplea en cada repositorio para gestionar la colección digital, ya que el software puede condicionar el modelo de datos. A continuación, se ha averiguado si los registros se expresan en RDF y, en caso afirmativo, si el repositorio ofrece alguna forma de consulta. Por último, se han analizado los resultados de las consultas para describir los modelos mediante los cuales se articulan los datos.

Para la selección de colecciones digitales se han seguido los mismos criterios de selección que en Sulé y otros, 2011; es decir, se han analizado las colecciones recogidas en el directorio de colecciones digitales de *Hispana*³⁵ de carácter exclusivamente patrimonial y que participan en la recolección de datos OAI. Aplicando estos criterios de selección, el 4 de abril de 2014 identificamos un total de 51 colecciones, que se enumeran en un apéndice al final del estudio.

La obtención de los datos de las colecciones digitales se ha realizado aplicando diferentes métodos de análisis según la información que se quería recopilar. La identificación del programa de gestión del repositorio se ha llevado a cabo, principalmente, por medio del análisis directo del sitio web. Cuando esto no ha sido suficiente, se ha enviado un correo electrónico a su responsable solicitando dicha información.

La confirmación, o no, de que los datos contenidos en las colecciones se expresan también según el modelo de datos para metadatos RDF y la

identificación de su forma de consulta (OPAC, API o SPARQL) se ha realizado también por medio de la combinación de la observación directa de los repositorios y del envío de correos electrónicos a los responsables de las colecciones. También se han consultado catálogos de datos abiertos de las administraciones públicas –como son *datos.gov.es*³⁶ y los catálogos específicos de los gobiernos autonómicos y municipales– por si contuvieran los datos RDF de las colecciones.

Por último, los detalles sobre el modelo de datos, las clases y las propiedades utilizados por las colecciones que expresan sus datos en RDF se han obtenido produciendo, cuando así ha sido posible, el fichero RDF/XML correspondiente, generando su DTD con XMLSpy XML Editor³⁷ y analizando los datos resultantes.

A continuación, se presentan los resultados del estudio y su valoración, comentando los programas utilizados en la gestión, si expresan los datos en RDF y cómo pueden consultarse. Se muestran los modelos de datos utilizados, concretamente EDM y OntoWeb, de los cuales se describen las clases y propiedades, para pasar a una evaluación global de la aplicación de estos modelos.

4. RESULTADOS

4.1. Programa de gestión de las colecciones

Los programas empleados para gestionar las 51 colecciones digitales seleccionadas son, ordenados según el número de colecciones que los usan:

- DIGIBIB: 17
- DSpace: 14
- CONTENTdm: 4
- Pandora: 4
- Fedora: 2
- Invenio: 1
- Nubarchiva³⁸: 1
- Microsoft Access: 1
- Desarrollo propio: 5
- Desconocido³⁹: 2

Como se ve, hay un claro predominio de DIGIBIB⁴⁰ (33%) y de DSpace⁴¹ (27%). El primero es un programa propietario de la empresa DIGIBIS orientado a la gestión de colecciones, físicas o digitales, que cuenta con una importante presencia en centros españoles⁴². Basado en MARC21, permite la conversión automática y transparente para el usuario de re-

gistros MARC en Europeana Data Model⁴³ (EDM) para su recolección por *Hispana* y *Europeana*. Por el contrario, DIGIBIB no dispone de ningún punto SPARQL.

Por su parte, DSpace es un programa de código abierto orientado a la gestión de colecciones digitales con una gran implementación internacional en repositorios institucionales. Basado en Dublin Core, DSpace es compatible también con OAI-PMH y desde la versión 5.0 permite convertir y almacenar los datos en RDF, así como publicarlos por medio de un punto de consulta SPARQL o serializados en RDF/XML, Turtle o N-Triples⁴⁴. En relación con Europeana Data Model, vale la pena mencionar el excelente trabajo que está llevando a cabo la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte que ha desarrollado el plugin *DSpace2Europeana* que permite a las instalaciones de DSpace interoperar con *Europeana* "a través de añadir la capacidad de exportación de registros de DSpace en los formatos ESE (Europeana Semantic Elements) y EDM (Europeana Data Model)"⁴⁵ (Ibai Sistema, 2012).

4.2. Registros expresados en RDF

En relación con la expresión de los datos en RDF, los resultados obtenidos son:

- Expresan sus registros en RDF: 26
- No expresan sus registros en RDF: 16
- Sin respuesta: 9

De las 26 colecciones que expresan sus registros en RDF, la distribución según el programa de gestión es la siguiente:

- DIGIBIB: 17
- DSpace: 7
- Fedora: 2

Es decir, todas las colecciones gestionadas con DIGIBIB permiten expresar sus registros bibliográficos en RDF, mientras que de las que son gestionadas con DSpace únicamente hemos tenido constancia de siete que lo hacen (de las siete restantes, tres han confirmado que no, y de otras cuatro no hemos obtenido respuesta). Las dos colecciones gestionadas con Fedora⁴⁶ también permiten expresar los registros bibliográficos en RDF.

En relación a los catálogos de datos abiertos de las administraciones públicas (*datos.gov.es* y catálogos específicos de los gobiernos autonómicos y municipales), hemos constatado que, en general, estos depósitos no referencian las colecciones digitales analizadas y que cuando lo hacen no dan cuenta de sus datos RDF.

4.3. Punto de consulta

Respecto a la forma de consulta de los datos expresados en RDF, los resultados obtenidos son:

- SPARQL endpoint: 1
- API (GET method): 21
- OPAC: 3
- Sin respuesta: 1

En este caso, la única colección donde hemos podido constatar que dispone de un punto de consulta SPARQL es el *Fondo histórico digitalizado de la UNED*⁴⁷ gestionado con Fedora, si bien por motivos de seguridad su consulta requiere de nombre de usuario y contraseña, y no hemos podido analizar sus datos.

En cuanto a las 21 colecciones que facilitan la consulta de todos sus datos expresados en RDF a través de una API, el resultado es un fichero XML obtenido mediante una petición "verb=ListRecords" según los parámetros del protocolo OAI-PMH⁴⁸. De todas ellas, catorce corresponden a colecciones gestionadas con DIGIBIB (es decir, no todas) y siete con DSpace.

Por último, hay que mencionar que tres de las colecciones que expresan sus datos en RDF únicamente permiten consultarlos desde el mismo OPAC y de forma individual, es decir, registro por registro. Se trata de tres colecciones gestionadas con DIGIBIB que no disponen de las funcionalidades de publicación en Linked Open Data y en EDM.

Así pues, el análisis de los modelos de datos, las clases y las propiedades de los registros expresados en RDF de las colecciones digitales españolas se ha tenido que limitar a las 21 que lo hacen a través de una API con el método GET. En el resto de casos, como se ha comentado, no hemos obtenido respuesta de la institución, no hemos podido acceder a los datos (SPARQL con claves de acceso) o sólo se pueden consultar los registros individualmente uno por uno (lo que imposibilita garantizar unos resultados exhaustivos para toda la colección).

4.4. Modelo de datos, clases y propiedades

En relación con el modelo en que son expresados los datos, el análisis de los ficheros obtenidos con las consultas GET, realizadas en los meses de mayo y junio de 2015, muestra que la mayoría de las colecciones, 17 de 21, siguen el Europeana Data Model (EDM): 14 de ellas gestionadas con DIGIBIB y tres con DSpace. Las cuatro restantes, gestionadas con DSpace, siguen el extinto modelo *OntoWeb*⁴⁹. En los dos casos, se utiliza XML para la serialización de los datos.

Puesto que ambos modelos establecen especificaciones de clases y de propiedades que pueden ser empleadas de forma diferente, pasamos a analizarlos de forma separada.

4.4.1. Colecciones digitales que siguen el modelo EDM

Como se ha mencionado, son 17 las colecciones que expresan sus datos siguiendo European Data Model. EDM establece tres clases básicas (*core classes*): edm:ProvidedCHO, edm:WebResource y edm:Aggregation. Para nuestro estudio nos interesa la clase edm:ProvidedCHO, puesto que contiene los metadatos que representan a los objetos originales en sí mismos. Sus propiedades obligatorias son:

- dc:title o dc:description
- dc:language para objetos textuales
- dc:subject o dc:type o dc:coverage o dcterms:spatial
- edm:type

A continuación se muestra en la tabla I en cuántas colecciones aparecen las propiedades de la clase edm:ProvidedCHO que se han identificado en las 17 bases de datos analizadas, así como la naturaleza de sus valores. Para una mejor valoración de los resultados, en la columna Datatype se indica el tipo de dato que autoriza EDM para cada elemento⁵⁰.

La primera evidencia es que hay seis propiedades que aparecen en todas las colecciones (columna "No" igual a 0): dc:creator, dc:description, dc:language, dc:title, dc:type y edm:type. De estos seis elementos, cinco corresponden a las propiedades obligatorias mencionadas anteriormente, mientras que dc:creator no. Esta coincidencia viene en gran medida determinada porque cumplir con las especificaciones de EDM es un requisito para que las colecciones puedan integrarse en *Europeana*.

Si nos fijamos ahora en la naturaleza de los valores de las propiedades, y más concretamente en la de aquellas que EDM especifica que puede ser una referencia, vemos que de las 26 posibles únicamente once propiedades se dan en uno o más repositorios con un URI local, destacando por su frecuencia dc:creator, dc:subject, dcterms:spatial, dc:contributor, dcterms:isPartOf y dc:publisher. En algunos casos (dc:creator, dcterms:spatial y dc:Publisher) la mayoría de los repositorios expresan los valores tanto con un URI local o con un literal, dependiendo ello de si el concepto tiene asociado en el catálogo un registro de autoridad o no.

Mención especial merece el caso de la propiedad dc:language en la *Biblioteca Virtual del Patrimonio Bibliográfico* y *IURIS Digital*, puesto que en algunos registros su valor se da en forma de URI externo (en concreto el URL del código ISO 639 de la web de la Library of Congress⁵¹) a pesar de que EDM recomienda el uso de códigos literales de dos o tres letras.

Además de las clases básicas, EDM incluye también cuatro clases contextuales (*contextual classes*): edm:Agent, edm:Place, edm:TimeSpan y skos:Concept. Estas clases contextuales permiten dar información sobre estas entidades de una manera diferenciada del objeto descrito (edm:ProvidedCHO). Los resultados del uso de estas clases contextuales en las 17 colecciones analizadas se pueden ver en la tabla II.

Esta clase aparece en 15 de los 17 repositorios que siguen el modelo EDM. En estas 15 colecciones, la propiedad skos:prefLabel aparece en todas ellas (columna "No" igual a 0), hecho que concuerda con la recomendación de EDM. Tras ella, las más utilizadas son owl:sameAs (nueve colecciones), skos:altLabel (ocho colecciones) y skos:note (seis colecciones). El resto aparecen en tres o menos colecciones.

Por lo que hace a las propiedades que EDM establece que su valor ha de ser una referencia, vemos que su uso no está muy extendido en las colecciones analizadas: owl:sameAs se da únicamente en nueve de ellas y edm:isRelatedTo en tres. En el caso de edm:isRelatedTo los valores son en todos siempre URI locales (referencias a otros agentes de la misma colección). Por el contrario, en la propiedad owl:sameAs hemos encontrado URI externos que referencian al mismo agente en otras bases de datos. En la tabla III se muestran los catálogos referenciados ordenados de mayor a menor frecuencia.

Como se puede observar en la tabla IV, en los repositorios analizados los enlaces externos más frecuentes de la clase contextual Agent son con *VIAF* y con *DBpedia*.

En este caso, la clase contextual Place aparece en catorce de las 17 colecciones analizadas. El análisis también refleja un uso mucho menor de las propiedades establecidas por EDM. De hecho, excepto skos:prefLabel, que se encuentra en las catorce colecciones, el uso de las otras propiedades se reduce a cuatro o menos colecciones.

En las tres colecciones que hemos encontrado la propiedad owl:sameAs, en dos de ellas se referencia a la *Lista de Encabezamientos de Materia para las Bibliotecas Públicas en SKOS* y en una al *Library of Congress Name Authority File* (tabla V).

Tabla I. Propiedades de la clase edm:ProvidedCHO

	Datatype	No	URI local	URI externo	URI local o literal	URI externo o literal	Literal
dc:creator	Literal o ref	0	3	0	12	0	2
dc:description	Literal o ref	0	0	0	0	0	17
dc:language	Literal	0	0	0	0	2	15
dc:title	Literal	0	0	0	0	0	17
dc:type	Literal o ref	0	0	0	0	0	17
edm:type	Literal	0	0	0	0	0	17
dc:date	Literal o ref	1	0	0	0	0	16
dc:publisher	Literal o ref	1	0	0	10	0	6
dc:subject	Literal o ref	2	11	0	3	0	1
dc:contributor	Literal o ref	3	13	0	0	0	1
dcterms:spatial	Literal o ref	3	2	0	12	0	0
dcterms:tableOfContents	Literal	3	0	0	0	0	14
dcterms:temporal	Literal o ref	3	0	0	2	0	12
dc:rights	Literal o ref	4	0	0	0	0	13
dcterms:alternative	Literal	4	0	0	0	0	13
dcterms:isPartOf	Literal o ref	6	8	0	2	0	1
dcterms:isReferencedBy	Literal o ref	8	0	0	0	0	9
dc:relation	Literal o ref	12	0	0	1	0	4
dcterms:replaces	Literal o ref	13	1	0	0	0	3
dcterms:extent	Literal o ref	14	0	0	0	0	3
dcterms:isReplacedBy	Literal o ref	14	1	0	0	0	2
dc:coverage	Literal o ref	15	0	0	0	0	2
dcterms:created	Literal o ref	15	0	0	0	0	2
dcterms:hasFormat	Literal o ref	15	0	0	0	0	2
dcterms:hasPart	Literal o ref	15	2	0	0	0	0
dc:format	Literal o ref	16	0	0	0	0	1
dc:identifier	Literal	16	0	0	0	0	1
dc:source	Literal o ref	16	0	0	0	0	1
dcterms:issued	Literal o ref	16	0	0	0	0	1
dcterms:medium	Literal o ref	16	0	0	0	0	1
edm:currentLocation	Ref	16	0	0	0	0	1
owl:sameAs	Ref	16	0	0	0	0	1

Tabla II. edm:Agent

	Datatype	No	URI local	URI externo	URI local o externo	Literal
skos:prefLabel	Literal	0	0	0	0	15
owl:sameAs	Ref	6	0	8	1	0
skos:altLabel	Literal	7	0	0	0	8
skos:note	Literal	9	0	0	0	6
edm:isRelatedTo	Ref	12	3	0	0	0
rdaGr2:gender	Literal	12	0	0	0	3
edm:begin	Literal	13	0	0	0	2
edm:end	Literal	13	0	0	0	2
rdaGr2:dateOfEstablishment	Literal	14	0	0	0	1
rdaGr2:dateOfTermination	Literal	14	0	0	0	1

Tabla III. edm:Agent / owl:sameAs

VIAF¹	10
DBpedia²	7
ISNI³	2
datos.bne.es⁴	2
LCNA⁵	1
FAST Linked Data⁶	1
Katalog der Deutschen Nationalbibliothek⁷	1

¹ VIAF: *The Virtual International Authority File*. <https://viaf.org/>. [Consulta: 21/06/2015].

² DBpedia. <http://wiki.dbpedia.org/>. [Consulta: 21/06/2015].

³ ISNI. <http://www.isni.org/>. [Consulta: 21/06/2015].

⁴ DATOS.BNE.ES. <http://datos.bne.es/>. [Consulta: 21/06/2015].

⁵ Library of Congress Name Authority File. <http://id.loc.gov/authorities/names.html>. [Consulta: 21/06/2015].

⁶ FAST Linked Data: *FAST Authority File*. <http://experimental.worldcat.org/fast/>. [Consulta: 21/06/2015].

⁷ Katalog der Deutschen Nationalbibliothek. <https://portal.dnb.de/opac.htm> [Consulta: 21/06/2015].

Tabla IV. edm:Place

	Datatype	No	URI local	URI local o literal	URI local o externo	URI externo	Literal
skos:prefLabel	Literal	0	0	0	0	0	14
skos:altLabel	Literal	10	0	0	0	0	4
owl:sameAs	Ref	11	0	0	1	2	0
skos:note	Literal	12	0	0	0	0	2
dcterms:hasPart	Ref	12	2	0	0	0	0
dcterms:isPartOf	Ref	12	1	1	0	0	0
wgs84_pos:lat	Floating point	13	0	0	0	0	1
wgs84_pos:long	Floating point	13	0	0	0	0	1

Tabla V. edm:Place / owl:sameAs

Lista de Encabezamientos de materia para las Bibliotecas Públicas¹	2
LCNA	1

¹ *Lista de Encabezamientos de Materia para las Bibliotecas Públicas en SKOS*. <http://id.sgcb.mcu.es/lem/>. [Consulta: 21/06/2015].

La clase edm:TimeSpan es, con mucha diferencia, el elemento de EDM con menor presencia en las colecciones analizadas. Únicamente se encuentra en dos de ellas. Además, en estos dos casos la única propiedad que se incluye es skos:prefLabel.

La última clase analizada es skos:Concept que se encuentra en 15 de las 17 colecciones analizadas. Como se puede ver en la tabla VII, sólo hay una

propiedad (skos:prefLabel) que aparece en todas ellas, mientras que skos:broader lo hace en catorce.

skos:closeMatch, la propiedad que permite dar el URI de un concepto similar de otra base de datos, se encuentra en siete repositorios: en cinco de ellos se referencia a la *Lista de Encabezamientos de materia para las Bibliotecas Públicas*, y en dos a la *Library of Congress Subject Headings* y a la *DBpedia*.

Tabla VI. edm:TimeSpan

	Datatype	No	Literal
skos:prefLabel	Literal	15	2

Tabla VII. skos:Concept

	Datatype	No	URI local	URI local o externo	URI externo	Literal
skos:prefLabel	Literal	0	0	0	0	15
skos:broader	URI	1	14	0	0	0
skos:closeMatch	URI	8	0	1	6	0
skos:altLabel	Literal	10	0	0	0	5
skos:note	Literal	12	0	0	0	3
skos:related	URI	12	3	0	0	0
skos:narrower	URI	14	1	0	0	0
skos:notation	URI	14	0	0	0	1

Tabla VIII. skos:Concept / skos:closeMatch

Lista de Encabezamientos de materia para las Bibliotecas Públicas	5
LCSH¹	2
DBpedia	2

¹ *Library of Congress Subject Headings*. <http://id.loc.gov/authorities/subjects.html>. [Consulta: 21/06/2015].

4.4.2. Colecciones digitales que siguen el modelo OntoWeb

Como se ha comentado en el apartado 4.4, hay cuatro colecciones (todas ellas gestionadas con DSpace) que siguen el modelo de datos OntoWeb. OntoWeb Ontology fue diseñada en el marco del proyecto "OntoWeb: ontology-based information exchange for knowledge management and electronic commerce", financiado por la Unión Europea a través del programa IST (*Tecnologías de la Sociedad de la Información*), entre junio de 2001 y mayo de 2004. La ontología constituía la columna vertebral del portal del mismo nombre, cuyo propósito era integrar los ámbitos de la investigación y la industria para mejorar el intercambio y los servicios de información mediante el uso de ontologías. Al mismo tiempo, OntoWeb también pretendía fortalecer la influencia europea en los procesos de normalización en diferentes áreas de la web semántica.

En su forma original, algunos de los componentes de OntoWeb son utilizados por DSpace en el diseño de una tabla de equivalencias para facilitar la inclusión de datos formalizados en RDF en la exposición de metadatos estructurados a través de OAI-PMH. En este mecanismo se utiliza la clase `ow:Publication` para enmarcar el registro de metadatos descriptivos de los objetos, que se expresan mediante propiedades de Dublin Core.

En ninguna de las cuatro colecciones analizadas en este apartado hemos encontrado clases de contexto (`edm:Agent`, `edm:Place`, `edm:TimeSpan` o `skos:Concept`), por lo que en la tabla IX se muestran únicamente las propiedades de la clase

`ow:Publication` ordenadas, de mayor a menor, por el número de repositorios donde aparecen.

Tabla IX. `ow:Publication`

	No	Literal
dc:date	0	4
dc:description	0	4
dc:rights	0	4
dc:type	0	4
dc:title	0	4
dc:identifier	0	4
dc:creator	1	3
dc:language	1	3
dc:subject	1	3
dc:publisher	2	2
dc:source	2	2
dc:contributor	3	1
dc:relation	3	1
dc:format	3	1

Como se puede apreciar hay seis propiedades que aparecen en las cuatro colecciones: `dc:date`, `dc:description`, `dc:rights`, `dc:type`, `dc:title` y `dc:identifier`. El resto lo hace con menor frecuencia.

Vale la pena destacar que, en este caso, el valor de todas las propiedades en todas las colecciones es literal, es decir, en ningún caso hemos encontrado un URI (local o externo).

5. CONCLUSIONES

La principal conclusión que se extrae de estos resultados es que la aplicación del modelo de datos RDF en las colecciones digitales analizadas es muy desigual y, en general, poco desarrollada. Únicamente 26 de las 42 colecciones analizadas expresan sus datos en RDF, y de estas 26 la mayoría (17) están gestionadas con DIGIBIB, lo que nos da una idea de hasta qué punto la publicación de datos en RDF puede estar condicionada al uso de aplicaciones que incorporen, ya de serie, funcionalidades para convertir de manera automática sus registros en tripletes RDF.

Un segundo resultado que demuestra el poco desarrollo de la aplicación de tecnologías de la web semántica es la casi inexistencia de puntos de consulta SPARQL con los que permitir que agentes externos puedan consultar y recuperar de manera automática los datos RDF de las colecciones. Únicamente nos consta que una colección, el *Fondo histórico digitalizado de la UNED*, ofrece dicho punto de consulta SPARQL, si bien por motivos de seguridad su consulta requiere de nombre de usuario y contraseña. Del resto de colecciones, en tres casos sólo se pueden obtener los datos RDF consultando directamente el OPAC (registro por registro), mientras que en 21 se puede obtener un fichero RDF/XML con todos los registros expresados en RDF a través de una consulta GET.

Un tercer criterio de evaluación es el uso de referencias como valor de las propiedades, ya sea en forma de URI local o externo. En este caso hay una primera y clara diferencia según el modelo de datos empleado, ya sea éste EDM u OntoWeb. EDM establece para ciertas propiedades que su valor ha de ser o puede ser una referencia, mientras que OntoWeb no establece ningún criterio en este sentido. Esta diferencia de modelos de datos determi-

na que en las cuatro colecciones OntoWeb ninguna propiedad contenga en ningún caso una referencia como valor (siempre es un literal), mientras que en las 17 de EDM encontramos en diversas propiedades valores en forma de URI.

Por lo que concierne al enriquecimiento de los registros con enlaces externos a otras bases de datos, la valoración tampoco es muy positiva. De las 17 colecciones EDM, únicamente nueve dan un URI externo como valor de la propiedad owl:sameAs de la clase edm:Agent, tres en la clase edm:Place y ninguna en la clase edm:TimeSpan. En la clase skos:Concept sólo siete colecciones dan un URI externo como valor de la propiedad skos:closeMatch.

Por todo ello, valoramos que las colecciones digitales españolas deberían hacer un doble esfuerzo enriqueciendo sus datos RDF con enlaces externos y creando puntos de consulta SPARQL para, de esta forma, reforzar sus actuales objetivos de difusión y accesibilidad de las colecciones y, a su vez, desarrollar nuevas metas en el marco de los Datos Abiertos Enlazados.

También sería conveniente determinar los niveles de agregación (local, nacional, europeo) de los datos RDF generados por cada colección. A priori, sería posible considerar únicamente los niveles de la propia colección o del conjunto de las colecciones de una institución, pero, desde nuestro punto de vista, también sería conveniente considerar niveles superiores, como los agregadores nacionales (*Hispana*) o supranacionales (*Europeana*). De esta forma, aumentaría la proyección y difusión de la información contenida en las colecciones. Todo ello, no lo olvidemos, en un momento en que, como ya se ha comentado, los datos y metadatos de bibliotecas, archivos y museos se incorporarán al ámbito de los datos abiertos de la administración del estado.

6. NOTAS

- [1] La versión actualmente vigente es RDF 1.1, cuya especificación está formada por un conjunto de recomendaciones y otros documentos no normativos disponibles en: *RDF current status*. http://w3.org/standards/techs/rdf#w3c_all [Consulta: 21/06/2015].
- [2] Todos los documentos normativos de estas tecnologías y otros documentos informativos para el seguimiento de su evolución pueden consultarse en: *W3C data activity: building the Web of Data*. <http://w3.org/2013/data> [Consulta: 21/06/2015].
- [3] En la versión RDF 1.1 se incorpora otro modelo de identificador, *Internationalized Resource Identifier* o IRI, definido conjuntamente por el World Wide Web Consortium e Internet Engineering Task Force. Resulta de una generalización de URI que permite un margen más amplio de caracteres *Unicode*. Se prevé que progresivamente sustituya a los URI.
- [4] En rigor, el sujeto y el objeto también admiten nodos en blanco (*blank nodes*).
- [5] Las especificaciones de estos lenguajes y otros documentos informativos pueden ser consultados en: *Vocabularies*. <http://w3.org/standards/semanticweb/ontology> [Consulta: 21/06/2015].
- [6] Esta sintaxis de serialización ha sido definida por la recomendación *RDF 1.1 XML Syntax*.

- <http://w3.org/TR/2014/REC-rdf-syntax-grammar-20140225> [Consulta: 21/06/2015].
- [7] Las recomendaciones de W3C que definen estos dos formatos de serialización son, respectivamente, *RDF 1.1 Turtle: Terse RDF Triple Language*. <http://w3.org/TR/turtle> [Consulta: 21/06/2015] y *RDF 1.1 TriG: RDF Dataset Language*. <http://w3.org/TR/2014/REC-trig-20140225> [Consulta: 21/06/2015].
- [8] Las recomendaciones de W3C que definen este lenguaje de serialización son, para XHTML, *XHTML+RDFa 1.1: Support for RDFa via XHTML Modularization*. 3rd ed. <http://www.w3.org/TR/xhtml-rdfa> [Consulta: 21/06/2015], y, para HTML4 y HTML5, *HTML+RDFa 1.1: Support for RDFa in HTML4 and HTML5*. 2nd ed. <http://w3.org/TR/html-rdfa> [Consulta: 21/06/2015].
- [9] Por ejemplo, el formato JSON de RDF (RDF/JSON), cuya sintaxis ha sido definida por la recomendación *RDF 1.1 JSON Alternate Serialization (RDF/JSON): W3C Working Group Note 07 November 2013*. <http://w3.org/TR/rdf-json> [Consulta: 21/06/2015]. No debe confundirse con JSON-LD, que se considera de uso preferente a RDF/JSON.
- [10] La especificación de W3C que define el mecanismo de Microdatos en HTML es HTML Microdata: W3C Working Group Note 29 October 2013. <http://w3.org/TR/microdata> [Consulta: 21/06/2015].
- [11] Véase, por ejemplo, *Microdata to RDF: Transformation from HTML+Microdata to RDF*. 2nd ed. <http://w3.org/TR/microdata-rdf> [Consulta: 21/06/2015]. Incluso se han desarrollado iniciativas para dar apoyo al despliegue y uso del vocabulario Schema.org, con una orientación específica a los Datos Enlazados.
- [12] La especificación de W3C que lo define es *JSON-LD 1.0: A JSON-based Serialization for Linked Data*. <http://w3.org/TR/json-ld> [Consulta: 21/06/2015].
- [13] En concreto, la propiedad de JSON-LD que facilita esta operativa es @context. Más información y ejemplos pueden localizarse en Mitchell (2013, p. 24).
- [14] Este tipo de tripleta es denominada "generalized RDF triple".
- [15] Todos los documentos normativos de estas tecnologías y otros documentos informativos para el seguimiento de su evolución pueden consultarse en: *Linked Data*. <http://w3.org/standards/semanticweb/data> [Consulta: 21/06/2015].
- [16] *Open Bibliographic Data*. <http://opendefinition.org/bibliographic> [Consulta: 21/06/2015].
- [17] Unión Europea. Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público. *DOUE-L*, nº 345, de 31.12.2003, p. 90-96.
- [18] España. Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. *BOE*, nº 276, 17 noviembre 2007, p. 47160-47165.
- [19] Unión Europea. Directiva 2013/37/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013 por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. *DOUE-L*, nº 175, de 27 de junio de 2013, p. 1-8.
- [20] España. Congreso de los Diputados. "Proyecto de Ley por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *BOCG. Congreso de los Diputados. Serie A: Proyectos de Ley*, nº 150-1, 30 de abril de 2015, p. 1-11.
- [21] La versión actualmente vigente de SPARQL es 1.1. La especificación de este lenguaje y otros documentos informativos, pueden ser consultados en: *SPARQL corrent status*. http://w3.org/standards/techs/sparql#w3c_all [Consulta: 21/06/2015].
- [22] Es el resultado de la confluencia de Datos Abiertos (interoperabilidad legal) y Datos Enlazados (interoperabilidad técnica).
- [23] *W3C Library Linked Data Incubator Group*. W3C. <http://w3.org/2005/Incubator/ld> [Consulta: 21/06/2015].
- [24] *Kollektionen/Sammlungen*. Zentrales Verzeichnis Digitalisierter Drucke. <https://zvdd.de/dms/browsen> [Consulta: 21/06/2015].
- [25] *Digitale Bibliothek*. Münchener Digitalisierungszentrum. http://www.digitale-sammlungen.de/index.html?c=digitale_sammlungen&l=de [Consulta: 21/06/2015].
- [26] *Digital collections*. British Library Labs. <http://labs.bl.uk/Digital+Collections> [Consulta: 21/06/2015].
- [27] *Free data services*. British Library. <http://www.bl.uk/bibliographic/datafree.html> [Consulta: 21/06/2015].
- [28] *Semantic Web Development Tools*. <http://w3.org/2001/sw/wiki/Tools> [Consulta: 21/06/2015].
- [29] *Open Refine: a free, open source, powerful tool for working with messy data*. Open Refine. <http://openrefine.org/> [Consulta: 21/06/2015].
- [30] *ALIADA: Automatic publication under Linked Data Paradigm of library DATA*. ALIADA. <http://www.aliada-project.eu> [Consulta: 21/06/2015].

- [31] Library Linked Data Incubator Group (2011). *Use Cases*. W3C. <http://w3.org/2005/Incubator/1ld/wiki/UseCases> [Consulta: 21/06/2015]. La información recoge muestras hasta el 2011.
- [32] *Europeana*. <http://europeana.eu> [Consulta: 21/06/2015].
- [33] *The Digital Public Library of America*. Library of Congress. <http://dp.la/> [Consulta: 21/06/2015].
- [34] *BIBFRAME: Bibliographic Framework Initiative*. Library of Congress. <http://loc.gov/bibframe/> [Consulta: 21/06/2015].
- [35] *Hispana: directorio de colecciones digitales*. Ministerio de Educación, Cultura y Deporte. <http://hispana.mcu.es/es/comunidades/directorio.cmd> [Consulta: 21/06/2015].
- [36] *datos.gob.es*. Ministerio de Hacienda y Administraciones Públicas, Ministerio de Industria, Energía y Turismo. <http://datos.gob.es/> [Consulta: 21/06/2015].
- [37] *XMLSpy XML Editor*. Altova. <http://altova.com/xmlspy.html> [Consulta: 21/06/2015].
- [38] Una extensión de Alfresco.
- [39] No se pudo averiguar por la consulta directa del repositorio ni la organización respondió al correo que enviamos solicitando esta información.
- [40] *DIGIBIB para bibliotecas*. DIGIBÍS. <http://www.digibis.com/es/software/digibib.html> [Consulta: 21/06/2015].
- [41] *DSpace*. DSpace. <http://dspace.org> [Consulta: 21/06/2015].
- [42] *Bibliotecas virtuales producidas por DIGIBÍS*. DIGIBÍS. Última actualización: 8 de Enero de 2015. <http://digibis.com/bibliotecas-virtuales/producidas-por-digibis.html> [Consulta: 21/06/2015].
- [43] *Definition of the Europeana Data Model v5.2.5*. 22/05/2014. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.2.pdf [Consulta: 21/06/2015].
- [44] *Linked (Open) Data*. DuraSpace Wiki. <https://wiki.duraspace.org/display/DSDO-C5x/Linked+%28Open%29+Data> [Consulta: 21/06/2015].
- [45] *Dspace2Europeana*. Subdirección General de Coordinación Bibliotecaria. <http://www.meecd.gob.es/cultura-mecd/areas-cultura/bibliotecas/recursos-profesionales/8.html> [Consulta: 21/06/2015].
- [46] *Fedora* (<https://getfedora.org>) es un programa de código abierto que permite construir repositorios digitales.
- [47] *Fedora Resource Index Query Service*. <http://62.204.194.43:8080/fedora/risearch> [Consulta: 21/06/2015].
- [48] *Open Archives Initiative Protocol for Metadata Harvesting*. <http://openarchives.org/pmh> [Consulta: 21/06/2015].
- [49] Kim, H. H. (2005). ONTOWEB: implementing an ontology-based web retrieval system. *Journal of the American Society for Information Science and Technology*, 56.11, 1167-1176.
- [50] *Europeana Data Model: mapping guidelines v2.2*. 2014, p. 14-21. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.2.pdf [Consulta: 21/06/2015].
- [51] *ISO 639-1: Codes for the Representation of Names of Languages - Part 1: Two-letter codes for languages*. Library of Congress. <http://id.loc.gov/vocabulary/iso639-1.html> y *ISO 639-2: Codes for the Representation of Names of Languages - Part 2: Alpha-3 Code for the Names of Languages*. Library of Congress. <http://id.loc.gov/vocabulary/iso639-2.html> [Consulta: 21/06/2015].

7. REFERENCIAS

- Berners-Lee, T. (2006). *Linked Data*. W3C. <http://w3.org/DesignIssues/LinkedData.html> [Consulta: 21/06/2015].
- Heath, T.; Bizer, C. (2011). *Linked Data: evolving the Web into a global data space* (1st ed.) San Rafael, Ca.; Morgan & Claypool. 136 p. <http://linkeddata-book.com/editions/1.0/> [Consulta: 21/06/2015].
- Ibai Sistemas (2012). *Informe de adaptación de DSpace a Europeana. Fase Danubio: Europeana Data Model (EDM). Parte I: panorama general*. Madrid: Ministerio de Educación, Cultura y Deporte. Subdirección General de Coordinación Bibliotecaria. 91 p. http://travesia.mcu.es/portaln/jspui/bitstream/10421/6301/5/Informe_adaptacion_DSpace_EDM_I.pdf [Consulta: 21/06/2015].
- Ibai Sistemas (2012). *Informe de adaptación de DSpace a Europeana. Fase Danubio: Europeana Data Model (EDM). Parte II: propuesta de solución técnica*. Madrid: Ministerio de Educación, Cultura y Deporte. Subdirección General de Coordinación Bibliotecaria. 32 p. http://travesia.mcu.es/portaln/jspui/bitstream/10421/6301/6/Informe_adaptacion_DSpace_EDM_II.pdf [Consulta: 21/06/2015].

- Kim, H. H. (2005). ONTOWEB: implementing an ontology-based web retrieval system. *Journal of the American Society for Information Science and Technology*, 56 (11), 1167-1176. <http://dx.doi.org/10.1002/asi.20220>
- Library Linked Data Incubator Group (2011). *Library Linked Data Incubator Group Final Report*. W3C. <http://w3.org/2005/Incubator/llid/XGR-llid-20111025> [Consulta: 21/06/2015]
- Mitchell, E. T. (2013). Chapter 2: Building Blocks of Linked Open Data in Libraries. *Library Technology Reports*, 49 (5), 11–25.
- Nilsson, M.; Powell, A.; Johnston, P.; Naeve, A. (2008). *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. DCMI. <http://dublincore.org/documents/dc-rdf> [Consulta: 21/06/2015]
- RDF 1.1 Concepts and Abstract Syntax (2014). W3C. <http://w3.org/TR/rdf11-concepts> [Consulta: 21/06/2015]
- RDF 1.1 Primer (2014). W3C. <http://w3.org/TR/rdf11-primer> [Consulta: 21/06/2015].
- Suárez-Figueroa, M.C.; Gómez-Pérez, A.; López-Cima, A. (2004). *Portal ontology*. Version 1.0. 36 p. (Deliverable D1.6.2). <http://knowledgeweb.semanticweb.org/semanticportal/deliverables/D1.6.2.pdf> [Consulta: 21/06/2015].
- Sulé Duesa, A.; Estivill Rius, A.; Gascón García, J. (2011). Interfaces de consulta en las colecciones digitales patrimoniales españolas. *Anales de documentación*, vol, 14 (2). <http://revistas.um.es/analesdoc/article/view/113931> [Consulta: 21/06/2015].
- Wood, D.; Zaidman, M.; Ruth, L.; Hausenblas, M. (2014). *Linked Data: Structured Data on the Web*. Shelter Island, NY; Manning. 276 p.
- Yu, L. (2011). *A developer's guide to the semantic web*. Heidelberg; Springer. 608 p. <http://dx.doi.org/10.1007/978-3-642-15970-1>

8. ANEXO

Tabla X. Información sobre las colecciones digitales (datos de febrero de 2015)

Colección digital	URL	Entidad responsable
<i>Arias Montano: repositorio Institucional de la Universidad de Huelva</i>	http://rabida.uhu.es/dspace	Universidad de Huelva
<i>ARCA: arxiu de revistes catalanes antigues</i>	http://www.bnc.cat/digital/arca/	Biblioteca de Catalunya
<i>Biblioteca digital de Castilla y León</i>	http://bibliotecadigital.jcyl.es	Junta de Castilla y León
<i>Biblioteca digital de Castilla-La Mancha</i>	http://bidicam.castillalamancha.es	Biblioteca de Castilla-La Mancha
<i>Biblioteca digital de fondo antiguo</i>	https://biblioteca.unizar.es/colecciones/fondo-antiguo	Universidad de Zaragoza
<i>Biblioteca digital de la ciudad de Murcia</i>	http://murcia.es/bibliotecadigitaldemurcia	Archivo Municipal de Murcia
<i>Biblioteca digital de la Comunidad de Madrid</i>	http://bibliotecavirtualmadrid.org	Bibliotecas de la Comunidad de Madrid
<i>Biblioteca digital de la Real Academia de la Historia</i>	http://bibliotecadigital.rah.es/dgbrah	Real Academia de la Historia
<i>Biblioteca digital Floridablanca</i>	http://bibliotecafloridablanca.um.es/bibliotecafloridablanca	Universidad de Murcia. Biblioteca General
<i>Biblioteca digital hispánica</i>	http://www.bne.es/es/Catalogos/BibliotecaDigitalHispanica/Inicio/index.html	Biblioteca Nacional de España
<i>Biblioteca digital Memoria de Madrid</i>	http://memoriademadrid.es	Ayuntamiento de Madrid
<i>Biblioteca histórica de Santa Cruz</i>	http://www.bhsc.uva.es:8080/BHSC/fondos.jsp	Universidad de Valladolid
<i>Biblioteca navarra digital</i>	https://administracionelectronica.navarra.es/binadi/busqueda.aspx	Servicio de Bibliotecas del Gobierno de Navarra
<i>Biblioteca Saavedra Fajardo de pensamiento político hispánico</i>	http://saavedrafajardo.org	Universidad de Murcia
<i>Biblioteca valenciana digital</i>	http://bv2.gva.es	Biblioteca Valenciana
<i>Biblioteca virtual de Andalucía</i>	http://www.bibliotecavirtualdeandalucia.es	Junta de Andalucía. Consejería de Educación, Cultura y Deporte
<i>Biblioteca virtual de Aragón</i>	http://bibliotecavirtual.aragon.es	Gobierno de Aragón. Dirección General de Cultura

Colección digital	URL	Entidad responsable
<i>Biblioteca virtual de derecho aragonés</i>	http://derechoaragones.es	Gobierno de Aragón
<i>Biblioteca virtual de la provincia de Málaga</i>	http://bibliotecavirtual.malaga.es	Diputación de Málaga
<i>Biblioteca virtual de la Real Academia Nacional de Farmacia</i>	http://bibliotecavirtual.ranf.com	Real Academia Nacional de Farmacia
<i>Biblioteca virtual de La Rioja</i>	http://bibliotecavirtual.larioja.org/bvrioja	Gobierno de La Rioja. Biblioteca de La Rioja
<i>Biblioteca virtual de prensa histórica</i>	http://prensahistorica.mcu.es	Ministerio de Educación, Cultura y Deporte
<i>Biblioteca virtual del patrimonio bibliográfico</i>	http://bvpb.mcu.es	Ministerio de Educación, Cultura y Deporte
<i>Biblioteca virtual del Principado de Asturias</i>	http://www.bibliotecavirtual.asturias.es	Principado de Asturias. Consejería de Cultura y Turismo
<i>Bilboko liburutegi digitala</i>	http://bilbao.net/bld	Ayuntamiento de Bilbao. Red de Bibliotecas Municipales de Bilbao
<i>Calaix</i>	http://calaix.gencat.cat	Generalitat de Catalunya. Departament de Cultura i Mitjans de Comunicació
<i>Cartoteca digital</i>	http://cartotecadigital.icc.cat	Institut Cartogràfic de Catalunya
<i>DARA: documentos y archivos de Aragón</i>	http://www.sipca.es/dara	Gobierno de Aragón. Dpto. de Educación Cultura y Deporte
<i>DIGIBUG: repositorio institucional de la Universidad de Granada</i>	http://digibug.ugr.es	Universidad de Granada
<i>Fondo fotográfico Universidad de Navarra</i>	http://coleccionfff.unav.es/bvunav	Universidad de Navarra
<i>Fondo histórico digitalizado de la UNED</i>	http://e-spacio.uned.es/fez/list.php?community_pid=bibliuned:Fondohistoricodigitalizado	Universidad Nacional de Educación a Distancia
<i>Fondos digitalizados de la Universidad de Sevilla</i>	http://fondosdigitales.us.es	Universidad de Sevilla
<i>Fons Sol-Torres</i>	http://soltorres.udl.cat	Universitat de Lleida
<i>Fototeca del IPCE</i>	http://ipce.mcu.es/documentacion/fototeca.html	Instituto del Patrimonio Cultural de España
<i>Galiciana</i>	http://galiciana.bibliotecadegalicia.xunta.es	Consellería de Cultura, Educación e Ordenación Universitaria da Xunta de Galicia
<i>Hemeroteca Digital</i>	http://hemerotecadigital.bne.es	Biblioteca Nacional de España
<i>IURIS Digital</i>	http://bvrajyl.rajyl.es/i18n/estaticos/contenido.cmd?pagina=estaticos/presentacion	Real Academia de Jurisprudencia y Legislación
<i>Memoria digital de Canarias</i>	http://mdc.ulpgc.es	Universidad de Las Palmas de Gran Canaria
<i>Memòria digital de Catalunya</i>	http://mdc1.cbuc.cat	Consorci de Serveis Universitaris de Catalunya
<i>Memoria digital vasca</i>	http://memoriadigitalvasca.es	Fundación Sancho el Sabio
<i>Patrimonio bibliográfico lacunense</i>	http://hermes.bbtk.ull.es/pandora	Universidad de La Laguna
<i>Proyecto CARMESÍ: archivos históricos de la Región de Murcia</i>	http://www.regmurcia.com/carmesi	Fundación Integra
<i>Recursos de Investigación de la Alhambra</i>	http://www.alhambra-patronato.es/ria	Patronato de la Alhambra y el Generalife. Archivo-Biblioteca
<i>Red digital de colecciones de museos de España: CER.es</i>	http://ceres.mcu.es	Ministerio de Educación, Cultura y Deporte
<i>REGIRA</i>	http://www.ddgi.cat/web/servei.seam?serveiId=1042&nivellId=202&cid=708&categoriaId=2	Diputació de Girona. Servei de Biblioteques
<i>Repositorio de objetos digitales y aprendizaje</i>	http://roda.gobex.es	Junta de Extremadura. Consejería de Cultura
<i>Repositorio documental de la Universidad de Salamanca</i>	http://gredos.usal.es/jspui	Universidad de Salamanca

Colección digital	URL	Entidad responsable
<i>Repositorio documental de la Universidad de Valladolid</i>	http://uvadoc.uva.es	Universidad de Valladolid
<i>Repositorio institucional Arjona y Cubas</i>	http://repositorio.racordoba.es/jspui	Real Academia de Ciencias, Bellas Letras y Nobles Artes de Córdoba
<i>Repositorio institucional de la Universidad de Málaga</i>	http://riuma.uma.es	Universidad de Málaga
<i>Repositorio institucional de la Universidad de Oviedo</i>	http://digibuo.uniovi.es/dspace	Universidad de Oviedo

Colección digital	Programa	RDF	Modelo	SPARQL	API	OPAC
<i>Arias Montano: repositorio Institucional de la Universidad de Huelva</i>	DSpace	Sin respuesta	--	--	--	--
<i>ARCA: arxiu de revistes catalanes antigues</i>	ContentDM	No	--	--	--	--
<i>Biblioteca digital de Castilla y León</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca digital de Castilla-La Mancha</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca digital de fondo antiguo</i>	Invenio	No	--	--	--	--
<i>Biblioteca digital de la ciudad de Murcia</i>	DSpace	No	--	--	--	--
<i>Biblioteca digital de la Comunidad de Madrid</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca digital de la Real Academia de la Historia</i>	DIGIBIB	Sí	EDM	No	No	Sí
<i>Biblioteca digital Floridablanca</i>	DSpace	Sin respuesta	--	--	--	--
<i>Biblioteca digital hispánica</i>	Pandora	Sin respuesta	--	--	--	--
<i>Biblioteca digital Memoria de Madrid</i>	Desarrollo propio	No	--	--	--	--
<i>Biblioteca histórica de Santa Cruz</i>	Desarrollo propio	Sin respuesta	--	--	--	--
<i>Biblioteca navarra digital</i>	Desarrollo propio	No	--	--	--	--
<i>Biblioteca Saavedra Fajardo de pensamiento político hispánico</i>	Desconocido	Sin respuesta	--	--	--	--
<i>Biblioteca valenciana digital</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual de Andalucía</i>	DIGIBIB	Sí	EDM	No	No	Sí
<i>Biblioteca virtual de Aragón</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual de derecho aragonés</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual de la provincia de Málaga</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual de la Real Academia Nacional de Farmacia</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual de La Rioja</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual de prensa histórica</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual del patrimonio bibliográfico</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Biblioteca virtual del Principado de Asturias</i>	DIGIBIB	Sí	EDM	No	Sí	Sí

Colección digital	Programa	RDF	Modelo	SPARQL	API	OPAC
<i>Bilboko liburutegi digitala</i>	DSpace	No	--	--	--	--
<i>Calaix</i>	DSpace	Sí	OntoWeb	No	Sí	No
<i>Cartoteca digital</i>	ContentDM	No	--	--	--	--
<i>DARA: documentos y archivos de Aragón</i>	Nubarchiva	No	--	--	--	--
<i>DIGIBUG: repositorio institucional de la Universidad de Granada</i>	DSpace	Sí	OntoWeb	No	Sí	No
<i>Fondo fotográfico Universidad de Navarra</i>	DIGIBIB	Sí	EDM	No	No	Sí
<i>Fondo histórico digitalizado de la UNED</i>	Fedora	Sí	--	Sí	No	No
<i>Fondos digitalizados de la Universidad de Sevilla</i>	Desarrollo propio	No	--	--	--	--
<i>Fons Sol-Torres</i>	DSpace	Sí	OntoWeb	No	Sí	No
<i>Fototeca del IPCE</i>	MSAccess	No	--	--	--	--
<i>Galiciana</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Hemeroteca Digital</i>	Pandora	No	--	--	--	--
<i>IURIS Digital</i>	DIGIBIB	Sí	EDM	No	Sí	Sí
<i>Memoria digital de Canarias</i>	ContentDM	No	--	--	--	--
<i>Memòria digital de Catalunya</i>	ContentDM	No	--	--	--	--
<i>Memoria digital vasca</i>	DSpace	Sí	EDM	No	Sí	No
<i>Patrimonio bibliográfico lacunense</i>	Desconocido	Sin respuesta	--	--	--	--
<i>Proyecto CARMESÍ: archivos históricos de la Región de Murcia</i>	Pandora	Sin respuesta	--	--	--	--
<i>Recursos de Investigación de la Alhambra</i>	DSpace	Sí	OntoWeb	No	Sí	No
<i>Red digital de colecciones de museos de España: CER.es</i>	Desarrollo propio	No	--	--	--	--
<i>REGIRA</i>	Pandora	No	--	--	--	--
<i>Repositorio de objetos digitales y aprendizaje</i>	Fedora	Sí	--	No	Sin respuesta	No
<i>Repositorio documental de la Universidad de Salamanca</i>	DSpace	Sí	EDM	No	Sí	No
<i>Repositorio documental de la Universidad de Valladolid</i>	DSpace	Sí	EDM	No	Sí	No
<i>Repositorio institucional Arjona y Cubas</i>	DSpace	Sin respuesta	--	--	--	--
<i>Repositorio institucional de la Universidad de Málaga</i>	DSpace	No	--	--	--	--
<i>Repositorio institucional de la Universidad de Oviedo</i>	DSpace	Sin respuesta	--	--	--	--