

# SISTEMA DE MODELADO SEMÁNTICO PARA CATALOGACIÓN, CLASIFICACIÓN, CONSULTA Y PUBLICACIÓN EN ABIERTO DE INFORMACIÓN BIBLIOGRÁFICA

Semantic modeling system for cataloging, classification, searching, and data opening of bibliographical information

Ana-María Feroso-García, María-Isabel Manzano-García, Montserrat Mateos-Sánchez y Carlos Hernández-Tamayo



**Ana-María Feroso-García**, doctora en Ciencias de la Computación e ingeniera informática por la *Universidad de Deusto*, es profesora catedrática de Ingeniería del Software de la *Facultad de Informática* de la *Universidad Pontificia de Salamanca*. Como investigadora sus líneas de trabajo se centran en las tecnologías y la web semántica, y cuenta con publicaciones de impacto, proyectos y conferencias relacionadas con esta área. Especialmente interesada en la publicación y reutilización de datos en abierto a través de formatos semánticos, como las ontologías, incluyendo su diseño, implementación y explotación en distintos ámbitos como bibliotecas y formatos bibliográficos de catalogación. <http://orcid.org/0000-0001-7204-4414>

[aferosoga@upsa.es](mailto:aferosoga@upsa.es)



**María-Isabel Manzano-García** es licenciada en documentación por la *Universidad de Salamanca* y ha cursado los masters *Documentación Digital* y *Buscadores* de la *Universitat Pompeu Fabra*. Ha participado en diversos grupos de trabajo de organizaciones relacionadas con archivos y bibliotecas, fue presidenta de la *Asociación de Bibliotecarios de la Iglesia en España (ABIE)*, y ha publicado diversos artículos relacionados con su trabajo. Es bibliotecaria desde hace más de 24 años en la Biblioteca de la *Universidad Pontificia de Salamanca*, donde ocupa el cargo de directora del *Servicio de Biblioteca y Archivo*.

<http://orcid.org/0000-0002-8011-5734>

[mmanzanoga@upsa.es](mailto:mmanzanoga@upsa.es)



**Montserrat Mateos-Sánchez**, doctora en Informática en el área de Lenguajes y Sistemas por la *Universidad de Salamanca*, es profesora encargada de cátedra en la *Facultad de Informática* de la *Universidad Pontificia de Salamanca*. Es responsable del grupo de investigación *Maraton* sobre sistemas de información, concretamente en recuperación de información, tecnologías semánticas y datos abiertos, tecnologías móviles e *IoT*. Ha participado como investigador colaborador y principal en proyectos competitivos relacionados con sus líneas de investigación. Es autora y coautora de numerosas publicaciones científicas y ha participado como ponente en conferencias nacionales e internacionales.

<http://orcid.org/0000-0003-3934-0291>

[mmateossa@upsa.es](mailto:mmateossa@upsa.es)



**Carlos Hernández-Tamayo** es antiguo alumno del grado en Ingeniería Informática y experto universitario en *big data* por la *Universidad Pontificia de Salamanca*. Trabaja como programador junior especialista en aplicaciones de tratamiento masivo de datos en una empresa de consultoría informática sobre *business intelligence* y *analytics*. Su trabajo de fin de grado se centró en el diseño, implementación y puesta en práctica de una ontología enfocada a bibliotecas mediante la publicación en abierto de sus datos y el empleo de tecnologías semánticas (*linked open data*).

<https://orcid.org/0000-0002-9835-2914>

[chdeztamayo@gmail.com](mailto:chdeztamayo@gmail.com)

*Universidad Pontificia de Salamanca*  
Compañía, 5. 37002 Salamanca, España

## Resumen

Descripción de un nuevo sistema que plantea el diseño y formateo de información de catalogación bibliográfica para su publicación en abierto utilizando tecnologías semánticas y *linked open data*. Se proponen dos modelos ontológicos: 1) basado en la ontología bibliográfica LD4L; 2) se diseña una nueva ontología, *BiblioOntology*, para categorizar por materias los fondos de una biblioteca universitaria. Tomando como base ambas ontologías, el sistema hace posible publicar en abierto y categorizar un recurso bibliográfico partiendo de su registro MARCXML. Igualmente permitirá consultar la información por varios criterios, enriqueciéndola con los de otras fuentes de externas. Finalmente servirá como proveedor a otras fuentes, a través de un servicio *Sparql point*. La principal aportación del sistema es su utilidad para la integración, distribución, reutilización y uso abierto de los fondos bibliográficos de una biblioteca universitaria. El sistema, denominado *OpenBiblio*, está siendo utilizado en la *Universidad Pontificia de Salamanca (UPSA)*, pero igualmente resulta reutilizable en cualquier otra biblioteca universitaria. <http://dataupsa.upsa.es>

## Palabras clave

Datos enlazados abiertos; Datos enlazados; Ontologías; Web semántica; Bibliotecas; Bibliotecas universitarias; *UPSA*; *OpenBiblio*.

## Abstract

A new system is proposed to design and format bibliographic information using linked open data and semantic technology. Two ontological models are presented. The first one is based on the bibliographic ontology LD4L. The second one is the new ontology *BiblioOntology*, designed to categorize by subject the resources of a university library. Using either ontology, users can publish linked open data (LOD) and to categorize a resource from its MARCXML. It also allows users to query the information by various criteria and enrich it with criteria from external sources. Finally, it serves as a provider to other sources, through a *Sparql Point service*. The main contribution of the system is its usefulness for the integration, distribution, reuse, and open use of the bibliographic resources of a university library. The system is currently in use at the *Universidad Pontificia de Salamanca (UPSA)*, but it could be used by any university library. <http://dataupsa.upsa.es>

## Keywords

Linked open data (LOD); Linked data; Ontologies; Semantic Web; Libraries; Academic libraries; *UPSA*; *OpenBiblio*.

**Fermoso-García, Ana-María; Manzano-García, María-Isabel; Mateos-Sánchez, Montserrat; Hernández-Tamayo, Carlos (2018).** "Sistema de modelado semántico para catalogación, clasificación, consulta y publicación en abierto de información bibliográfica". *El profesional de la información*, v. 27, n. 2, pp. 410-418.

<https://doi.org/10.3145/epi.2018.mar.20>

## 1. Introducción

Las bibliotecas son instituciones acostumbradas a tratar con datos. Forma parte de sus tareas habituales la recopilación, tratamiento, conservación y difusión de datos de la manera más eficiente y normalizada posible. Asimismo, llevan siglos practicando el acceso abierto, la transparencia y el rigor.

Los contenidos gestionados (o "curados") en bibliotecas durante siglos no deberían quedar fuera de las listas de resultados de una búsqueda en la Web (Hallo *et al.*, 2016). Para Torre-Bastida, González-Rodríguez y Villar-Rodríguez (2015), bibliotecas y bibliotecarios deben concienciarse de que tienen mucho que aportar en el área de los datos enlazados abiertos (*linked open data*, en adelante LOD). Las legislaciones europea y nacional se han hecho eco de la importancia de que los datos y metadatos de instituciones como las bibliotecas, formen parte del universo de los datos abiertos, facilitando su búsqueda y reutilización (Sulé *et al.*, 2016). Los contenidos seleccionados y de calidad que las bibliotecas llevan siglos recopilando, describiendo y difundiendo, deben de dejar de ser islas de conocimiento y estar plenamente integradas en la Web (Tharani, 2015).

Con la tecnología de datos abiertos enlazados (LOD) (Giusti-Serra; Santarém-Segundo, 2016) los datos de las bibliotecas podrán salir de la internet profunda, a la vez que se eliminarán tareas duplicadas y se reducirán esfuerzos en muchos aspectos del trabajo de bibliotecarios, documentalistas, archiveros o profesionales de la información. El formato LOD de publicación en abierto, y en particular el formato 5 estrellas promovido por Tim Berners-Lee (2006), capacita a las máquinas para recolectar datos e información relevante que, en el caso de las bibliotecas, es rigurosa, está revisada y seleccionada y es de calidad.

En este contexto se presenta en este artículo el proyecto *OpenBiblio* de la Biblioteca de la Universidad Pontificia de Salamanca (en adelante UPSA). El proyecto tiene como principal objetivo publicar información bibliográfica que proporcione datos fidedignos de personalidades locales a partir de los registros de autoridad, así como datos bibliográficos del catálogo de la biblioteca. El modelo de diseño es aplicable a cualquier entorno de biblioteca universitaria y/o especializada, tanto desde el punto de vista del sistema de información o software que se presenta, como de sus ontologías.

<https://koha.upsa.es/cgi-bin/koha/opac-authoritiesdetail.pl?authid=242851>

Los datos serán expuestos a todos aquellos agentes externos que deseen acceder a ellos en formato de datos enlazados. Es la continuación del proyecto *DataUPSA* (Fermoso-García et al., 2016), sobre datos en abierto en formato semántico a partir de registros MARCXML de la biblioteca.

Partiendo de esta idea se ha diseñado un sistema automatizado que permitirá convertir tanto registros bibliográficos de una obra, como registros de autoridad de sus autores, desde el formato MARCXML, a un formato semántico de catalogación bibliográfica basado en LD4L.

<https://www.ld4l.org>

La elección de este formato se justifica en el siguiente apartado. El sistema también permitirá categorizar o clasificar de manera automatizada los recursos bibliográficos en un sistema de clasificación basado en la CDU (*UDC, universal decimal classification*), sistema reconocido mundialmente gestionado y mantenido por un consorcio de editores, entre los que se encuentra *Aenor*, responsable de su edición en castellano.

<http://udcdata.info>

<http://www.udcc.org>

El sistema igualmente posibilitará que las materias asignadas en la *UPSA* enlacen a listas de materias reconocidas y ya disponibles en formato LOD, como LEM (Lista de encabezamientos de materia), LCSH (*Library of Congress subject headings*) o Rameau.

<http://id.sgcb.mcu.es/Autoridades/subjectsNames.html>

<http://id.loc.gov/authorities/subjects.html>

<http://www.cs.vu.nl/STITCH/rameau>

Aprovechando la transformación a estos formatos semánticos, se implementa también un sistema de búsqueda y visualización de datos. Finalmente y para que todos los datos publicados en abierto puedan ser consumidos, buscados individual o masivamente, reutilizados y enlazados, se proporciona un *Sparql point* o servicio de consulta semántica, tanto vía web como a través de servicios *REST* (*representational state transfer*).

A continuación se analizan los conceptos mencionados, después se describen los modelos ontológicos y la arquitectura del sistema propuesto, así como sus funciones asociadas. Las conclusiones finales analizan las aportaciones del sistema.

## 2. Conceptos relacionados

### 2.1. *Linked open data* (LOD)

Son un formato o conjunto de normas para publicar (es decir, enlazar) datos en abierto de la manera más eficiente posible utilizando tecnologías semánticas.

<http://linkeddata.org>

Permiten conectar unos datos de la Web con otros con los que no existía una relación directa previamente. La relación se establece gracias al valor semántico de estos datos, valor a su vez especificado a través de etiquetas o URIs asociadas. Los datos enlazados permiten enriquecer y añadir valor a

los mismos, al interconectarlos con otros con los que se relacionan semánticamente, complementando y ampliando su significado e información asociada.

Gracias a la utilización de normas y estándares como RDF o OWL, los datos son interoperables y se pueden enlazar y enriquecer unos con otros en la Web.

<https://www.w3.org/RDF>

<https://www.w3.org/OWL>

La Web es considerada como un conjunto de conceptos entrelazados, no como páginas aisladas comunicadas a través de enlaces web explícitos. Es lo que se conoce como formato de publicación de datos 5 estrellas de **Berners-Lee** (2006).

En nuestro caso concreto y como segunda fase, se ha utilizado el formato *Linked data for libraries* (LD4L). Se trata de un formato LOD específico de catalogación bibliográfica, evolución de *Bibframe*. En una fase anterior del proyecto se utilizó *Bibframe*, y en nuestro sistema se pueden utilizar ambas ontologías.

<https://www.ld4l.org>

Los datos enlazados permiten enriquecer y añadir valor a los mismos, al interconectarlos con otros con los que se relacionan semánticamente

Hay otras experiencias de LOD para bibliotecas. *OCLC* propone la suya basándose en el formato *Schema.org*, un proyecto piloto que resulta muy específico para los servicios propios de *OCLC* como el catálogo *WorldCAT*, y que por ello es menos reutilizable en otros contextos o catálogos.

Según **Godby** (2013) *Bibframe* se adapta mejor al entorno bibliográfico, con una descripción más detallada, mientras que *Schema.org*, con menor profundidad descriptiva, facilita la recuperación y uso en la web semántica. En ambos formatos existe una cierta incapacidad para expresar relaciones y en ambos casos se está avanzando hacia una mayor capacidad descriptiva. En el caso de *OCLC* y *Schema*, usando una versión extendida o apoyándose en otras ontologías como DDC (*Dewey Decimal Classification*), *VIAF* para autores, o *FAST* para materias (basadas en una versión más ágil de LCSH).

<http://bib.schema.org>

*Bibframe*, con la aparición de nuevas versiones ha mejorado la descripción de obras (*works*) y relaciones. En el apartado de *Bibframe* de la web de la *Library of Congress* hay actualizaciones continuas de vocabularios, aplicaciones y contenidos, con la aparición de nuevos formatos derivados de él como LD4L.

Estamos en un momento de muchas iniciativas, todas en evolución, también LD4L. En nuestro caso nos hemos decantado por esta última iniciativa porque da soporte también a datos sobre investigación e investigadores asociados a las obras. Este componente de descripción de la investigación es algo que deseamos contemplar también en nuestro sistema.

## 2.2. LD4L

*Linked data for libraries* (LD4L) es un proyecto de las bibliotecas de las universidades *Cornell*, *Harvard*, y *Stanford*, y la *Library of Congress*.  
<https://www.ld4l.org>

Su objetivo es facilitar la búsqueda de información académica en particular y bibliográfica en general dentro de estas universidades. Trata de establecer un modelo semántico con el objetivo final de que acabe convirtiéndose en un formato bibliográfico estándar de facto, más concretamente en un formato LOD compuesto por descripciones bibliográficas y biográficas de personas o entidades, interoperable con otras fuentes.

LD4L, ontología de base del proyecto, se basa en formatos bibliográficos de referencia como *Bibframe*, añadiendo información sobre identificación de organizaciones académicas y sus miembros

Según **Morillo-Calero** (2016), LD4L sigue en desarrollo con trabajos en colaboración de las citadas universidades, además de *Iowa*, financiados por la fundación *Mellon*. Como ya se ha indicado, LD4L se basa en modelos semánticos ya existentes tales como *Bibframe* o *VIVO*.  
<https://www.loc.gov/Bibframe>  
<http://vivoweb.org>

El primero, promovido por la *Library of Congress*, se considera un estándar de catalogación bibliográfica en formato semántico, frente al tradicional MARCXML. Ante la cierta incapacidad de MARCXML para denotar relaciones, *Bibframe* apareció como su evolución natural. La transformación de MARC a MARCXML y después a RDF y en concreto a *Bibframe*, ha sido un gran avance y abre grandes expectativas a las bibliotecas y a los nuevos programas de gestión bibliotecaria.

LD4L tiene el mismo núcleo que la ontología *Bibframe*, pero enriquece la ontología con otras como *VIVO* o *FOAF*.  
<http://xmlns.com/foaf/spec>

*VIVO* complementa al anterior, permitiendo entre ambos añadir información para la identificación de organizaciones académicas, sus miembros y estamentos, lo que mejora la descripción y comprensión de documentos académicos y de investigación. Gracias a esta combinación resulta LD4L, una ontología más completa que las anteriores que puede enriquecerse de datos geográficos, biográficos e institucionales, de producción científica.

La mayor amplitud descriptiva de LD4L al añadir más detalle a las descripciones del formato *Bibframe*, y el aval de las instituciones que lo respaldan, nos ha hecho decantar por su uso en el proyecto. Se ha diseñado un software para convertir el formato tradicional MARCXML al nuevo modelo semántico basado principalmente en LD4L.

A la ontología estándar se le han añadido algunas clases y propiedades para adaptarla a nuestras necesidades. También se ha tenido en cuenta otro modelo ontológico, el *DataCite*.

## 2.3. DataCite

Es un esquema de metadatos, entre los que está el DOI (*digital object identifier*), para la publicación de datos de investigación. Fue creado en 2009 con el objetivo de facilitar el acceso a datos de investigación en internet.  
<https://www.datacite.org>

Es un modelo relativamente sencillo que incluye conceptos como *Orcid* para identificación de investigadores, así como informaciones sobre becas y fondos para la financiación de la investigación.  
<https://orcid.org>

En nuestro caso *DataCite*, cuyo esquema semántico se muestra en la figura 1, se utiliza para complementar LD4L, añadiendo aspectos como el citado *Orcid* de los investigadores.

## 2.4. SKOS

Es una iniciativa del W3C que utiliza un formato semántico basado en RDF y OWL para representar el contenido de esquemas conceptuales como listas de encabezamiento, tesauros o taxonomías.  
<https://www.w3.org/TR/SKOS-reference>

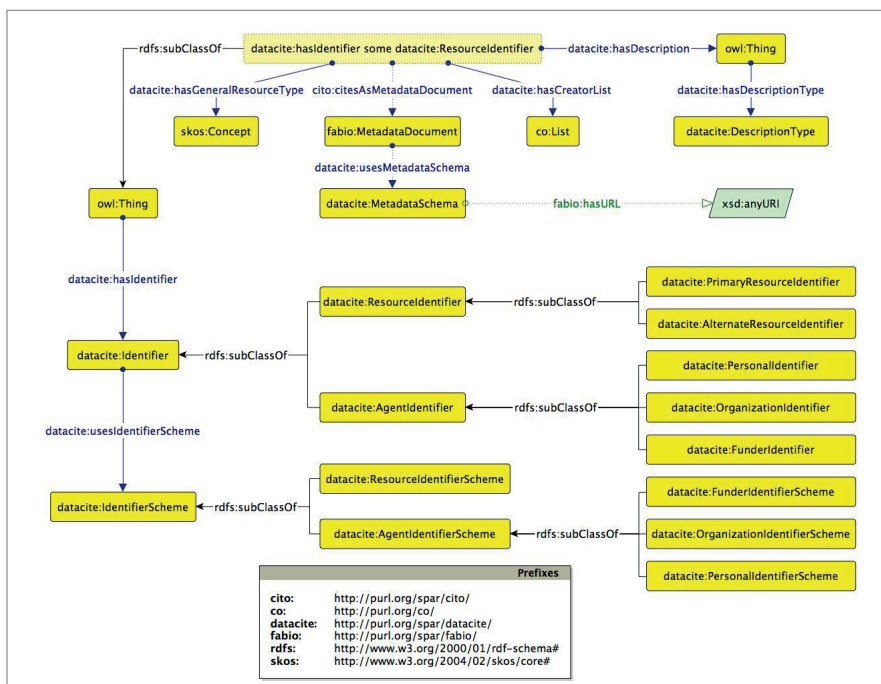


Figura 1. Modelo ontológico *DataCite*  
<https://www.ld4l.org/ld4l-2014/ontology-overview>



Tanto la lista de materias LEM como el sistema de clasificación *UDC* anteriormente mencionados, tienen su implementación en formato SKOS. Esto será aprovechado desde nuestro sistema para enlazar categorías y materias con el concepto SKOS asociado en estas listas si existe, y con ello también averiguar y navegar al mismo tiempo por términos relacionados por proximidad conceptual a través de etiquetas del tipo *SKOS:broader*, *SKOS:closer* o *SKOS:prefLabel*.

### 3. Arquitectura y componentes de *OpenBiblio*

El software *OpenBiblio* es uno de los resultados principales de nuestro trabajo. En este sistema se distinguen dos perfiles de usuarios:

- personal bibliotecario;
- resto de posibles usuarios.

Los dos perfiles distinguen dos partes en el sistema, cuyas funciones son diferentes.

La parte del usuario final está pensada para facilitar las búsquedas y enriquecer los resultados con nuevas fuentes de información. Se permitirá hacer búsquedas sobre los fondos de la biblioteca por autoridad, obra o materia concreta. También se ha desarrollado un *Sparql point* para las consultas semánticas.

*OpenBiblio* mejora LD4L añadiendo información sobre el curriculum del investigador (*Orcid*)

El perfil de bibliotecario podrá convertir un recurso en MARCXML a un formato semántico basado en LD4L. Igualmente se puede clasificar e indizar el recurso en un modelo propio, usando nuestra clasificación y listas de encabezamientos de materias, y a su vez enlazarlo y apoyarlo en LEM, *LCSH* y *UDC*. Esta clasificación se logra nuevamente extrayendo información del registro MARCXML asociado, pero ahora sobre la categoría o signatura *UPSA* de la obra y de sus materias asociadas. Igualmente se tiene acceso al mismo *Sparql point* en el subsistema anterior accesible tanto online como a través de un servicio web de tipo *REST*. Este servicio facilita enlazar otras fuentes con nuestro conjunto de datos abiertos de la biblioteca.

Para suministrar estas funciones se ha confeccionado el software *OpenBiblio*, que convierte los datos del formato MARCXML, tanto al modelo semántico basado en LD4L como al modelo propio para catalogación *BiblioOntology*.

Esto implica elaborar la ontología propia LD4L\_UPSA como ampliación de la ontología LD4L para adaptarla a nuestras necesidades, añadiendo una nueva clase que permite representar datos de investigación como el identificador *Orcid*.

La otra aportación de interés es el diseño de la nueva ontología *BiblioOntology*, que hace posible clasificar los recursos bibliográficos por su tema. Inicialmente fue aplicada al sistema de categorización temática propio de la universidad presentada como caso de uso, pero con posibilidad de ampliarse a cualquier área temática del saber y/o sistema de catalogación de cualquier universidad, pues se puede actualizar dinámicamente con nuevas categorías.

Atendiendo a las dos ontologías dominantes, LD4L\_UPSA y *BiblioOntology*, el sistema también puede considerarse dividido desde el punto de vista lógico en dos subsistemas, según la ontología fundamental con la que se trabaje en cada parte. El módulo *Sparql point* sin embargo es común para ambos subsistemas.

Por otro lado, desde el punto de vista del software se distingue la interfaz a utilizar por el personal de la biblioteca y la del usuario final, ambas con funciones diferentes.

Vamos a ver en detalle cada una de estas aportaciones principales, partiendo de la arquitectura del sistema, representada en la figura 2. En ella podemos ver representados los tres subsistemas mencionados:

- publicación en LOD;
- categorización;
- *Sparql point*.

### 4. Subsistema para publicación de datos bibliográficos como LOD

Es la parte del sistema que trabaja con el formato LD4L, tanto para convertir registros a este formato como para facilitar las búsquedas y enlace de datos. La base de este subsistema será la propia ontología que se utiliza como modelo semántico. Primero se analiza esta ontología y luego los módulos o funciones que hacen uso de ella, el de conversión de MARCXML al formato LOD, LD4L, y el de consulta de

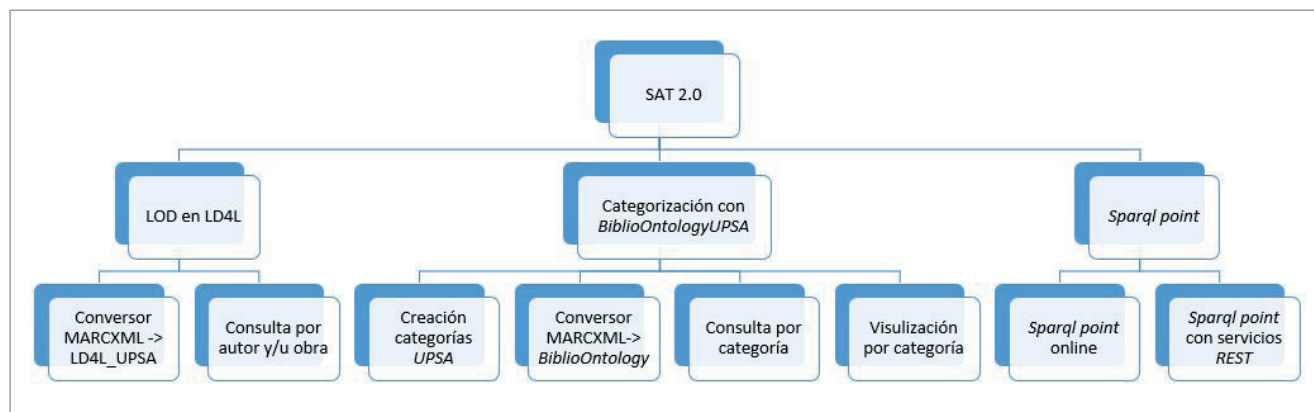


Figura 2. Arquitectura *OpenBiblio*



De manera similar se establece una correspondencia entre los datos de un libro, en realidad cualquier obra bibliográfica o instancia de obra “*Work*”, en formato MARCXML, y nuestra ontología, mediante las propiedades de la clase Libro. Se han añadido las *ObjectProperties* necesarias que consiguen relacionar a Libro como una instancia de obra, con su ISBN, título, autor o lugar de edición: *hasIsbn*, *Writeln*, *hasAuthorName*, *hasTitle*.

En definitiva, las tablas 1 y 2 representan la base del módulo de conversión de la información del catálogo de recursos bibliográficos y de autoridades, desde MARCXML a formato *open linked data*. Este módulo de conversión, como se ha comentado, se puede utilizar directamente para convertir cualquier recurso bibliográfico definido en MARCXML procedente de otra biblioteca, a formato semántico basado en LD4L.

### 4.3. Consulta de recursos por obras y autoridades

Los recursos, una vez convertidos al propio formato LD4L gracias a otro de los módulos de *OpenBiblio*, se pueden consultar por autor u obra. Se pueden consultar desde dos opciones, según el formato previo de conversión: el estándar *Bibframe* (ya disponible desde la primera versión del proyecto), o el formato LD4L.

<http://dataupsa.upsa.es/DataUpsaWeb>

<http://dataupsa.upsa.es/DataUpsaWebLD4L>

Otra aportación fundamental es que la información consultada se muestra enriquecida, no sólo con los datos deducidos del registro MARCXML original, sino también ampliados con datos externos enlazados con técnicas semánticas.

Concretamente de cada autoridad, si existe, se mostrará la información en *DBPedia*, la *Wikipedia* en formato semántico. Igualmente a través del identificador *Orcid* de cada autor se enlazará con sus datos de investigación.

<http://es.dbpedia.org>

<https://es.wikipedia.org>

## 5. Subsistema de catalogación de recursos bibliográficos por categorías y materias

La segunda parte o subsistema principal de nuestro sistema es la que permite clasificar temáticamente los recursos. Esto posibilitará hacer consultas más específicas a nuestros datos, por ejemplo para conocer y representar las obras de la biblioteca sobre una categoría o tema. Esta información se obtendrá nuevamente de los campos correspondientes de los registros MARCXML. Para lograr esta indización se diseña una ontología nueva, *BiblioOntology*.

En la biblioteca de la UPSA, la categorización de obras se hace usando una clasificación propia basada en la CDU, que se utiliza también como signatura topográfica.

Por otro lado, cada obra también puede tener asignadas varias materias, algunas incluidas en las LEM, *LCSH*, o *Rameau*. Este dato también se asocia a dicha obra en la nueva ontología.

Una vez los datos del catálogo se hayan convertido a dicha ontología, el sistema permitirá enriquecerlos con fuentes externas en modelos semánticos como SKOS.

Tabla 1. Modelo conversión clase “Autoridad” de MARCXML a LD4L\_UPSA

MARCXML	LD4L_UPSA
100a	Instance
024a-viaf	Viaf
024a-orcid	Orcid
024a-isni	Isni
100d	Date
370a	Place

Tabla 2. Modelo conversión clase “Libro” de MARCXML a LD4L\_UPSA

MARCXML	LD4L
245a	Instance, Work, CoverTitle
020a	Isbn10
100a	Author
260a	Place

La ontología diseñada se utilizará inicialmente para categorizar las obras del catálogo UPSA y según modelo propio de categorización. Sin embargo está preparada para actualizarse y ampliarse con nuevas categorías referidas a cualquier área de conocimiento y biblioteca. Dinámicamente se ha partido ya de unas categorías creadas propias de la UPSA y sus áreas de conocimiento, pero el personal bibliotecario puede seguir ampliando la ontología en cualquier momento con nuevas categorías, convirtiéndose así en un modelo escalable y reutilizable por cualquier otra institución.

Una vez diseñada la ontología, el software también permite categorizar cualquier recurso bibliográfico del repositorio de la biblioteca, creando para ello una instancia del mismo en la ontología *BiblioOntology*. Partiendo de la información del registro MARCXML asociado al recurso, se crea una instancia para ese recurso dentro de la clase *Libro* de la ontología. Dicha instancia contendrá el ISBN, categoría y materias asociadas a la obra bibliográfica. Una vez categorizado un recurso u obra bibliográfica también puede enlazar, ente otros, con la instancia asociada en el modelo LD4L\_UPSA y ampliar su información con otros datos como el título de la obra o autores de la misma u otros. Igualmente la obra se podrá localizar cuando se consulte por su categoría y a su vez enlazar si es el caso con información sobre sus materias asociadas.

A continuación se describe la ontología *BiblioOntology*, el conversor a la misma desde MARCXML, el módulo de actualización de la ontología con nuevas categorías, así como el sistema final de consultas por categoría de recursos y de publicación de los recursos encontrados enriquecidos, si es el caso con los datos aportados por el sistema LD4L y las listas externas de materias. Por último se incluye un sistema de visualización para las consultas mencionadas, que permitirá comprobar gráficamente el número de recursos por categoría UPSA. Todas estas funciones pueden reutilizarse en cualquier otra biblioteca.

### 5.1. Ontología *BiblioOntology*

En la ontología hay cuatro clases principales: *Libro*, *Categoría*, *Identificador* y *Materia*. A su vez están relacionadas en-



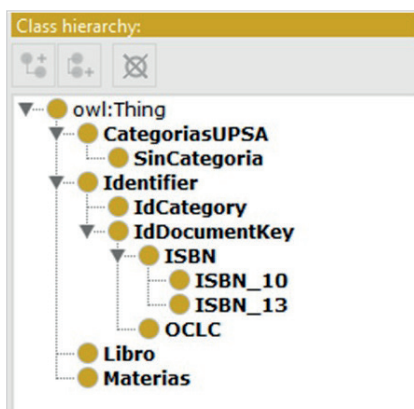


Figura 5. Jerarquía de clases que se empleará para la ontología UPSA

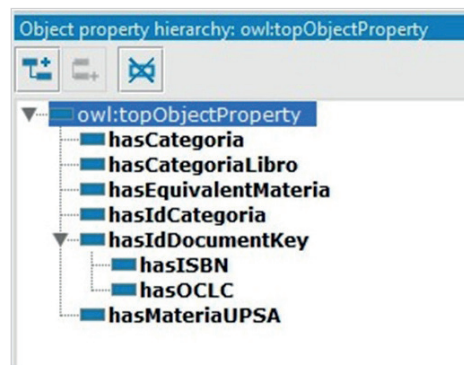


Figura 6. Propiedades ontología BiblioOntology

tre sí, fundamentalmente a través de la clase *Libro*, que es la que tendrá asociada tanto identificadores como materias y categorías que determinan de qué trata su contenido. La jerarquía de clases se muestra en la figura 5.

Para asociar estas características a un libro se definirá una serie de *ObjectProperties*, como se muestra en la figura 6. El contenido final de las instancias de libro con sus datos, a su vez se obtiene del registro asociado al libro en MARCXML, gracias al conversor de MARCXML a *BiblioOntology* implementado para esta tarea.

OpenBiblio permite la publicación en abierto gracias a la ontología LD4L, y la clasificación de las obras gracias a la ontología BiblioOntology

Las *ObjectProperties* de la figura 6 cuando se categoriza un libro permitirán respectivamente:

- *hasCategoria*: asociar a una materia la categoría a la que pertenece;
- *hasCategoriaLibro*: asociar una categoría a un libro;
- *hasEquivalentMateria*: relacionar materia UPSA con otras externas;
- *hasMateriaLibro*: asociar materia a libro;
- *hasIDCategoria*: asociar un ID o signatura a su categoría;
- *hasIDDocument*: de la que a su vez deriva el *hasISBN* y *hasOCLC*, para asociar los correspondientes identificadores a su libro.

### 5.2. Incorporación de categorías al modelo semántico de categorización UPSA

El software implementado hace posible crear el sistema de categorización de la propia biblioteca de la UPSA con el identificador o signatura asociada a cada categoría, a las que se asociarán luego los libros de la misma.

Para agilizar esta tarea y no tener que crear una a una cada categoría en la ontología, se puede partir de un CSV que contendrá el nombre de cada categoría y su signatura, que normalmente simbolizará también su clasificación jerárquica. El sistema procesa automáticamente el CSV y crea las categorías como instancias en la ontología.

En el sistema actual inicialmente ya se ha partido de un CSV con el tesoro o clasificación del área temática de *Biblia*, pero como se menciona, la propia ontología se puede incrementar con otros CSV con nuevas temáticas o categorías. Esto convierte esta propuesta ontológica en reutilizable en otras bibliotecas, permitiendo adaptar la ontología propuesta al propio sistema de categorización de la institución en la que se utilice.

### 5.3. Clasificación de una obra en el modelo semántico de categorización UPSA. Conversor MARCXML a BiblioOntology

El subsistema consigue clasificar cualquier obra o libro partiendo del registro MARCXML que lo describe, y transformarlo en una instancia de la clase libro de nuestra ontología. Más concretamente las características que nos interesan de cada libro y el dato o campo del que se deduce en MARCXML para crear la instancia de *BiblioOntology*, son las que se indican en la tabla 3.

Openbiblio convierte a la UPSA en fuente de datos abiertos accesibles (Sparql point)

Desde estos campos MARCXML se crea la correspondiente instancia de libro en el sistema, con sus propiedades o entidades asociadas: título, identificador del libro con su categoría UPSA, y adicionales si existiesen, así como la materia o materias asociadas a la obra, de existir éstas.

Al partir el conversor de un registro MARCXML habitual en bibliotecas, se podrá utilizar en cualquier entorno bibliotecario.

Tabla 3. Modelo conversión de MARCXML a BiblioOntology

MARCXML	BiblioOntology
245a	Título libro
020a	ISBN
035a	OCLC
953a	Categoría (signatura)
6XX/a/x/z	Materias



#### 5.4. Sistema de consultas y visualización por tema de recursos bibliográficos UPSA

Finalmente, *OpenBiblio* permite la consulta por categorías UPSA, de los recursos bibliográficos ya convertidos a nuestro formato LOD.

Por cada recurso encontrado se muestran no sólo los datos en la propia ontología, sino que además estos datos se enriquecen y amplían, enlazando con el subsistema o formato LD4L\_UPSA del mismo recurso. Mediante consultas semánticas se puede ampliar información sobre la categoría y materias de datos del propio recurso en formato SKOS a partir de la lista LEM de encabezamiento de materias del *Ministerio de Cultura*, o de aquellas otras que existan una vez convertido el recurso bibliográfico al formato propio LOD.

#### 6. Sparql point

El último subsistema es el que pone de manifiesto que se trata de un sistema LOD a todos los efectos, e implica que los datos publicados en abierto puedan ser consultados mediante técnicas semánticas, lenguaje *Sparql*, para poder enlazar dinámicamente con otros conjuntos de datos semánticos si es el caso. Este servicio se ofrece tanto online, para que el usuario realice directamente su consulta *Sparql* al sistema, como a través de un servicio web o servicio REST, para que también se puedan hacer consultas desde otros sistemas software.

#### 7. Conclusiones

Hay que destacar las principales aportaciones de *OpenBiblio* como sistema de formateo o publicación de información en formato semántico, que además hace posible enriquecer esta información y/o actuar como fuente de datos para otras fuentes de datos externas. El sistema es asimilable y reutilizable en cualquier otra biblioteca.

Para ello se han diseñado dos ontologías que posibilitan modelar semánticamente obras y categorías en bibliotecas. La ontología adaptada LD4L\_UPSA, que tiene como base a LD4L, ontología promovida por instituciones de referencia del sector de las bibliotecas universitarias (*Stanford*, *Harvard* o *Cornell*). La otra ontología utilizada es el diseño propio *BiblioOntology* y ayuda a categorizar semánticamente un recurso bibliográfico.

Se ha implementado la aplicación online *OpenBiblio*, apoyada en estas ontologías. Para ello primero se proporciona un sistema de conversión del formato original de los recursos bibliográficos, MARCXML, tanto a LD4L\_UPSA, como a *BiblioOntology*.

En paralelo y haciendo uso de ambas ontologías, se provee para el usuario final un sistema de consulta a los recursos bibliográficos por libro, autor o tema (categoría), con diferentes formatos visuales al mostrar los resultados de las consultas. Finalmente y como todo sistema de datos enlazados en abierto, se provee un *Sparql point* para enlazar desde fuentes externas a nuestras fuentes propias.

El modelo ontológico del proyecto *OpenBiblio* es generalizable y reutilizable en cualquier biblioteca. Desde el formato

MARCXML se puede pasar a formato LOD mediante nuestra ontología basada en LD4L, LD4L\_UPSA. Por otro lado, *BiblioOntology* puede ser utilizada por otras instituciones para categorizar por materias sus recursos y enlazar así con índices de materias reconocidas. En este sentido resultan formatos LOD y modelos ontológicos que se pueden incluso estandarizar para su reutilización en bibliotecas, especialmente las universitarias.

#### 8. Referencias

Berners-Lee, Tim (2006). *Linked data. Design issues*. <http://www.w3.org/DesignIssues/LinkedData.html>

Feroso-García, Ana; Manzano-García, María-Isabel; Armero-López, Álvaro; Hernández-Hernández, Álvaro (2016) "Apertura y publicación de datos bibliográficos con linked open data. Un caso práctico". En: *XV Workshop Rebiun de proyectos digitales "Datos y bibliotecas"*. Castellón, 29-30 septiembre. <http://hdl.handle.net/10234/163387>

Giusti-Serra, Liliana; Santarém-Segundo, José-Eduardo (2017). "O catálogo da biblioteca e o linked data". *Em questão*, v. 23, n. 2, pp. 167-185 <http://dx.doi.org/10.19132/1808-5245232.167-185>

Godby, Carol-Jean (2013). *The relationship between Bibframe and OCLC's linked-data model of bibliographic description: A working paper*. Ohio: OCLC Research. ISBN: 978 1 55653 460 7 <https://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf>

Hallo, María; Luján-Mora, Sergio; Maté, Alejandro; Trujillo, Juan (2016). "Current state of linked data in digital libraries". *Journal of information science*, v. 42, n. 2, p. 117. <https://goo.gl/cNPQSU> <https://doi.org/10.1177/0165551515594729>

Morillo-Calero, María-Jesús (2016). *Informe de asistencia a la conferencia anual de la ALA 2016 y a las reuniones del MARC Advisory Committee* (Departamento de Proceso Técnico). Madrid: BNE. [http://www.bne.es/export/sites/BNWEB1/webdocs/Inicio/Perfiles/Bibliotecarios/ALA/ala\\_annual\\_orlando\\_2016.pdf](http://www.bne.es/export/sites/BNWEB1/webdocs/Inicio/Perfiles/Bibliotecarios/ALA/ala_annual_orlando_2016.pdf)

Sulé, Andreu; Centelles, Miquel; Franganillo, Jorge; Gascón, Jesús (2016). "Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España". *Revista española de documentación científica*, v. 39, n. 1, e121 <https://doi.org/10.3989/redc.2016.1.1268>

Tharani, Karim (2015). "Linked data in libraries: A case study of harvesting and sharing bibliographic metadata with Bibframe". *Information technology & libraries*, v. 34, n. 1, pp. 5-19. <https://doi.org/10.6017/ital.v34i1.5664>

Torre-Bastida, Ana-Isabel; González-Rodríguez, Marta; Villar-Rodríguez, Esther (2015). "Datos abiertos enlazados (LOD) y su implantación en bibliotecas: iniciativas y tecnologías". *El profesional de la información*, v. 24, n. 2, p. 113-120. <https://doi.org/10.3145/epi.2015.mar.04>

8ª Conferencia internacional sobre revistas científicas

**CRECS 2018**

2-3-4 de mayo

<http://crecs.info>



**Universidad de La Costa. Barranquilla (Colombia)**



**SPRINGER NATURE**

