

PUBLICAR DATOS ABIERTOS ENLAZADOS (LOD) EN ENTORNOS BIBLIOTECARIOS MEDIANTE EL GESTOR DE CONTENIDO SEMÁNTICO XIMDEX

Publish Linked Open Data in library environments with the semantic CMS Ximdex

Silvia Gomis Aliaga

Universidad de Murcia, Facultad de Comunicación y Documentación, silvia.gomis1@um.es

Resumen

Se elabora una propuesta para publicar *LOD* en un conjunto de registros bibliográficos; en concreto en la colección de grabaciones sonoras de la Biblioteca de la Universidad de Alicante. El entorno bibliotecario presenta una serie de problemáticas para adoptar *LOD*, aunque la *W3C* reconoce el potencial de las bibliotecas para el desarrollo de la Web Semántica, las iniciativas para publicar parten de grandes instituciones. Se analiza el entorno bibliotecario a través de un análisis *DAFO* que nos ofrece una visión global sobre los problemas y potencialidades de la biblioteca para adoptar *LOD*. Para la evaluación del estado actual de la colección se utiliza la métrica *Meloda*, que permitió definir la propuesta y establecer el proceso de publicación de acuerdo a los estándares de la *W3C*, dividido a su vez en dos procesos: la preparación de los datos de la colección (evaluación, modelado y generación *RDF*) y la parte técnica del proceso desarrollada en un *CMS* semántico concreto, *Ximdex* (exportación, publicación de datos en bruto y generación de enlaces). *Ximdex* no es una solución perfecta para publicar *LOD* en entornos bibliotecarios, requiere de desarrollos específicos, como *Sparql endpoint*, pero sí es una herramienta adecuada para pequeños proyectos especialmente en medianas y pequeñas organizaciones.

Palabras clave: Web Semántica, Datos Abiertos Enlazados, Bibliotecas, CMS Semánticos, Ximdex.

Abstract

A proposal has been made to publish Linked Open Data in a set of bibliographic records; particularly in the collection of sound recordings of the Alicante University. The library environment presents a number of problematic for adopting *LOD*, though *W3C* recognizes the potential of libraries for the development of the Semantic Web, initiatives to publish stem from major institutions. The library environment is analyzed through a SWOT analysis it offers an overview of the problems and potentials of the library to adopt *LOD*. For the evaluation the state of art of the collection is used *Meloda* (Metric for releasing open data), which allowed to define the proposal and establish the publishing process according to *W3C* standards, divided into two processes are used: the preparation of data Collection (evaluation, modeling and generation *RDF*) and the technical part of the process developed on a semantic *CMS*, *Ximdex* (export, publishing raw data and automatic link generation). *Ximdex* is not a perfect solution to publish *LOD* in library environments require specific developments, such as *SPARQL endpoint* but it is an effective tool for small projects focused on publishing linked open data especially in medium and small organizations.

Keywords: Open Data, Library Linked Open Data, Semantic Web, Semantic CMS, Ximdex.

1. Introducción

Desde la creación de la *World Wide Web*, alrededor de 1989, por Tim Berners-Lee y su publicación formal en 1991 el proyecto Web Semántica se presenta como una extensión de la Web actual, que pretende convertirse en una plataforma universal para el intercambio de información. En el modelo Web Semántica la información está dotada de un significado bien definido, que

permite una mejor colaboración entre humanos y máquinas. (Berners-Lee, 2000). Para la *W3C* (1) la Web Semántica proporciona un marco común que permite que los datos sean compartidos y reutilizados a través de aplicaciones, empresas y fronteras comunitarias. Es un esfuerzo colaborativo liderado por el *W3C* con la participación de un gran número de investigadores y socios industriales. Está basado en *Resource Description Framework (RDF)* e integra una va-

riedad de aplicaciones utilizando *XML* para la sintaxis y *URIs* para las denominaciones."

En 2006, Tim Berners-Lee, introduce un concepto nuevo *Linked data* dentro de la estructura de la Web Semántica y es el responsable también de establecer los requisitos o principios que caracterizan a este movimiento. Según W3C:

"los Datos Enlazados es la forma que tiene la Web semántica de vincular los distintos datos que están distribuidos en la Web, de forma que se referencian de la misma forma que lo hacen los enlaces de las páginas web."

Pastor Sánchez (2012) afirmar que "los datos enlazados usan una pequeña selección de las tecnologías de la web semántica. Estas tecnologías se encuentran implícitamente en los cuatro principios de Berners-Lee para publicar *Linked Data* que son:

1. Utilizar *URIs* para identificar con nombre únicos a los recursos.
2. Utilizar el protocolo *HTTP* para nombrar y resolver la ubicación de los datos identificados mediante esas *URIs*.
3. Representar los datos y ofrecer información sobre los recursos utilizando *RDF* y manejar *SPARQL* como lenguaje de consulta de dichos datos.
4. Incluir enlaces a otras *URIs* para permitir la localización de más datos enlazados.

RDF es un modelo que permite definir la semántica de los datos. Se compone de tres conceptos principales:

1. Los recursos, que representan cualquier cosa (lugares, personas, objetos) del mundo real y están identificados por un *URI* (*Universal Resource Identifier*).
2. Las propiedades que son atributos o relaciones para describir un recurso.
3. Las sentencias que son una combinación de los conceptos anteriores y se componen de un sujeto, un predicado y un objeto. El significado de los datos se expresa mediante un conjunto de dichas sentencias que son representadas por tripletas (sujeto, predicado, objeto) y que se escriben usando marcas *XML*. El sujeto es un recurso identificado por un *URI*, los recursos pueden representar personas, lugares o cosas en el mundo

real. El predicado describe una propiedad o atributo del recurso, y permite la definición de un nuevo concepto a través de un puntero al documento (*URI*) donde está definido dicho concepto. Es decir, una propiedad es a su vez un recurso. El objeto es el valor asignado a dicha propiedad, y puede ser un literal, u otro recurso. (Martín, Olsina, Vallecillo y Bertoa, 2002)

Para la consulta de grafos *RDF* se usa *Sparql* (*Protocol and RDF Query Language*.) Es un lenguaje estandarizado que ha sido normalizado por el *DAWG* del *W3C*. La última revisión de *Sparql* propone operadores para crear y borrar declaraciones *RDF*.

Es importante referirnos a otro concepto, que no deben entenderse como sinónimo de *Linked data* sino que lo complementa: el *Open Data* (*OD*). Así como *OD* especifica el aspecto legal de la interoperabilidad, *LD* se refiere a la interoperabilidad técnica de los datos. La complementación entre ambos da lugar al movimiento *Linked Open Data* (*LOD*). (Hernández Santiago, 2012)

En 2010 Tim Berners Lee introduce las 5 estrellas de *Linked Open Data*, cuya filosofía se basa en que además de tener datos enlazados, estos datos deben estar en abierto.

★ Disponible en la web (en cualquier formato), pero con una licencia abierta.

★★ Disponible como datos estructurados legibles por máquina (por ejemplo, un Excel).

★★★ como el segundo punto pero además en un formato no propietario (por ejemplo, CSV en lugar de Excel).

★★★★ Todo lo anterior más utilizar estándares abiertos de W3C (*RDF* y *SPARQL*) para identificar los datos, y para que la gente pueda enlazarlos.

★★★★★ Todo lo anterior, y además: Enlazar sus datos a datos externos.

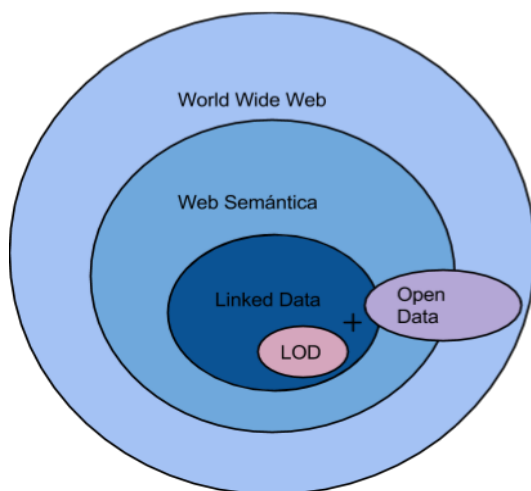


Figura 1. Relaciones de LOD en el entorno de la Web. Fuente: elaboración propia.

1.1. Situación de LOD en bibliotecas a través del análisis DAFO

El informe del grupo de trabajo *Library Linked data Incubator Group* (W3C (2011) establece una serie de limitaciones que podemos considerar como amenazas o debilidades en cuanto a los datos tradicionales de bibliotecas, como la falta de integración con los recursos de la web

“muchos catálogos de bibliotecas, y bases de datos bibliográficas y en texto completo, ya sean comerciales como gratuitas, forman parte de lo que se conoce como la “web invisible”, es decir que no son alcanzadas en los resultados de búsqueda de los buscadores” (Martín y Angelozzi, 2013).

Esta situación, junto a otros factores provoca un constante descenso del número de usuarios de catálogos e interfaces web de bibliotecas (Conaway y Dickey, 2010). Los datos enlazados se plantean como una solución a la falta de visibilidad web. Otra limitación que se establece en el informe es que gran parte de los conjuntos de datos o colecciones se encuentran almacenados en sistemas de gestión bibliotecaria controladas por un número reducido de firmas comerciales, lo que dificulta que las bibliotecas adopten los datos enlazados en un entorno de producción debido a la dependencia del software comercial que utilizan para gestionar sus colecciones, el uso de software abierto para la adopción de datos enlazados en bibliotecas se con-

vierte en un factor imprescindible para la producción de este tipo de datos en bibliotecas con escasos recursos e incluso para aquellas instituciones medias que tampoco pueden permitirse desarrollos propios de software.

“Se observa una tendencia a la integración, en el núcleo de algunos CMS, de ciertas funcionalidades de la Web Semántica. Las nuevas versiones de *Open CMS*, *Drupal* o *Liferay* “absorben” o integran módulos y *plugins* relacionados con la Web Semántica. Así pues, conforme se produzca esta integración, la optimización eficiencia en la manipulación de datos *RDF* se incrementará considerablemente, lo que conducirá inevitablemente a que los CMS lleguen a utilizarse como *frameworks* para el desarrollo de interfaces entre datos de la Web Semántica y contenidos de los sitios web que gestione” (Pastor Sánchez, 2012).

Según el informe del W3C, los sistemas de gestión de contenido (al igual que los *frameworks* de aplicaciones web), han evolucionado con la extensión de la Web. De hecho, estos sistemas se construyen a menudo con un *framework* de aplicaciones Web (2), pero proporcionan funcionalidades listas para usar en la creación, edición y presentación de contenidos, tales como textos, imágenes o vídeo en la Web, así como para gestionar flujos de trabajo asociados al contenido. Por esta razón siguen las mismas buenas prácticas para el nombrado de recursos con *URIs http* (3). La amplia disponibilidad de sistemas de gestión de contenidos (CMS son sus siglas en inglés) conduce a que su uso esté muy extendido en el sector del patrimonio cultural. Algunos sistemas de gestión de contenidos, como es el caso de *Drupal*, comienzan a exponer información estructurada en bases de datos a las máquinas cliente transformándola en *HTML* por medio de *RDFa* (4).

Los *CMS semánticos* o con capacidades semánticas, facilitan la gestión y desarrollo de software, la colaboración entre investigadores y la publicación de documentos, además de operar con *Linked data*. Son sistemas eficientes en el manejo de grandes conjuntos de datos distribuidos, la visualización de datos y la construcción de herramientas de análisis (Mooney and Baenziger, 2007). Se plantean como una solución para publicar datos enlazados en bibliotecas superando la limitación que sufren actualmente las

bibliotecas que dependen de un software comercial para desarrollar aplicaciones de consumo enmarcadas dentro de la web semántica. Por otro lado, los problemas que presentan los CMS para publicar aún son evidentes: demasiada variedad tecnológica, problemas de integración con módulos y elección de los mismos. En general, resultan en demasiadas herramientas, entendidas como piezas que necesariamente han de ser ensambladas y aprendidas. Los CMS más utilizados (*Drupal* y *Wordpress*), incluyen módulos semánticos para publicar LOD (en el caso de *Drupal* se trata del módulo RDF que está instalado en su núcleo o configuración original), pero presentan los mismos problemas de dispersión tecnológica, algo que no beneficia al entorno bibliotecario que presenta cierto rechazo al cambio y falta de conocimientos técnicos.

El CMS semántico *Ximdex*, por otro lado se desarrolla específicamente para publicar LOD, se presenta como una posibilidad para enfrentarse a los problemas de adopción dentro del entorno bibliotecario, es un software listo para usar que incorpora también módulos externos, pero integrados en el CMS, no presenta problemas de dispersión tecnológica, sin imposición de una tecnología concreta o lenguaje de programación su nivel de abstracción permite no sólo la compatibilidad con sistemas actuales sino también la adopción de otros futuros.

Se han publicado más herramientas y vocabularios que conjuntos de datos enlazados. Algunos autores han resaltado una serie de problemáticas a parte de las tecnológicas en el entorno bibliotecario, lo que hace que asumir la publicación de estos datos sea complicada para gran parte de la comunidad bibliotecaria. Asimismo, se observa que las iniciativas sobre publicación de datos enlazados parten de grandes proyectos o instituciones (Europeana o las diferentes Bibliotecas Nacionales): “se percibe aún la falta de tecnologías de referencia ampliamente aceptadas y de aplicaciones de consumo de datos bibliotecarios auténticamente demostrativas del valor que se le supone a los datos enlazados” (Saorín Pérez, Peset Mancebo y Ferrer Sapena, 2013). De hecho:

“El *Library Linked data Incubator Group* considera las bibliotecas como un proveedor de contenidos con gran potencial. El origen de esta fortaleza se encuentra en la información bibliográfica de tipo referencial, que se genera mediante normativas consolidadas que

aseguran su calidad, capacidad de compartir y sostenibilidad. Además trabajan con vocabularios controlados y con ficheros normalizados de autoridades nombres de persona, entidades, etc., lo que resulta de gran utilidad para enriquecer otros datos y para favorecer el correcto desarrollo de la web semántica.” (Peset Mancebo, Ferrer Sapena y Subirats Coll, 2011).

En cuanto a las fortalezas y oportunidades dentro del entorno bibliotecario para la adopción de LOD se observa que algunas de las limitaciones o amenazas pueden convertirse en oportunidades para la publicación de datos bibliotecarios, como es el caso del marco *FRBR* (5) basado en un enfoque entidad-relación y las reglas *RDA* (6) (si bien la Biblioteca Nacional de España ha descartado su adopción por motivos presupuestarios), y las iniciativas '*Marc must die*' (8) que surgen ante la obsolescencia del *MARC* y que cuyos objetivos son:

- Hablar sobre los objetivos de los metadatos bibliográficos en el entorno actual.
- Para recoger las opiniones, enlaces, etc y demás literatura sobre por qué *MARC* no cumple con los requisitos actuales en cuanto a las necesidad que se requieren de los metadatos bibliográficos.
- Para discutir las alternativas que podrían reemplazar *MARC*.

En este punto, las bibliotecas se encuentran en un escenario crítico en cuanto a los sistemas utilizados para gestionar sus colecciones, dependientes todavía en la mayoría de los casos de la estructura de datos del formato *MARC*. La historia muestra que todas las tecnologías son pasajeras y, en concreto, la historia de las TICs enseña que los formatos específicos de datos tienen una vida especialmente corta.

Los datos enlazados separan la descripción del significado de los datos ("semántica") de su estructura específica ("sintaxis" o "formato"), lo que da como resultado que retienen su significado cuando cambian los formatos. En este sentido, son más duraderos y robustos que los formatos de metadatos que dependen de una particular estructura de datos.

Así, la obsolescencia del formato *MARC* es una oportunidad para las biblioteca (Peset Mancebo, Ferrer Sapena y Subirats-Coll, 2011), pero sigue existiendo una gran dependencia del desarrollo

de software por parte de las firmas comerciales, aunque empiezan a ofrecerse soluciones distintas en este aspecto, al igual que ocurrió con el movimiento *Open Data* y las soluciones tecnológicas para poner en marcha los repositorios con software libre.

Si resumimos todos los puntos en un análisis DAFO de la situación actual podemos concluir en cuanto a la aplicación del LOD en el entorno bibliotecario que sus principales fortalezas son: los vocabularios controlados, la normalización de la información bibliográfica y de autoridades, en cuanto a las debilidades podemos destacar: la falta de recursos económicos y humanos y la resistencia al cambio. Se perciben como oportunidades para la adopción de LOD la obsolescencia del formato *MARC*, la aparición de reglas y modelos nuevos que se adaptan a la web semántica como *RDA* y *FRBR*, el esquema de metadatos *Dublin Core* (9) y el éxito entre la comunidad bibliotecaria de *Open Data*. Las amenazas más visibles serían: la falta de herramientas tecnológicas, la dependencia de sistemas comerciales y temas legales como las licencias de uso de los datos.



Figura 2. Análisis DAFO de *Linked Open Data* en Bibliotecas. Fuente: elaboración propia

1.2. Objeto de la investigación

El objeto principal de la investigación se origina como consecuencia de las limitaciones indicadas en el informe final del W3C respecto a la falta de publicación de datos abiertos enlazados

en entornos bibliotecarios y la escasa visibilidad de sus catálogos. Así pues, los objetivos centrales que justifican abordar este trabajo son:

1. Establecer una propuesta para publicar datos abiertos enlazados en un conjunto de registros bibliográficos, concretamente de una colección de grabaciones sonoras de la Biblioteca de la Universidad de Alicante.
2. Evaluar la situación actual de la colección actual a través de la métrica Meloda, como consecuencia de la necesidad de evaluación del conjunto de datos que permita conocer los procesos necesarios para establecer la propuesta.

Ahora bien, para llevar a cabo esta propuesta y poder cumplir los objetivos señalados, surge la necesidad de que los datos disponibles se encuentren en un formato abierto, para lo que debemos evaluar el estado actual de la colección en cuanto a los criterios de *Open Data* y *Linked Open Data* y establecer los procesos y tecnologías necesarias para su publicación. Así pues, son dos las acciones que se llevarán a cabo: establecer los procesos necesarios para la publicación de la colección como datos abiertos y enlazados, por un lado, y analizar y seleccionar las herramientas y tecnologías necesarias para cada parte del proceso, por otro.

1.3. Metodología

Para elaborar la metodología del trabajo se ha seleccionado la métrica Meloda (10) como primer paso para evaluar el grado de reutilización de los datos de la muestra (una colección de grabaciones sonoras).

Esta métrica está pensada para evaluar un conjunto de datos ('*datasets*') en abierto; para ponderar el nivel de reutilización de los mismos nos basamos en la conversión que ofrece la métrica para proyectos LOD basada en la calificación de las 5 estrellas de Berners-Lee para las cuatro dimensiones a evaluar

1. Dimensión legal: licencia que se asigna al conjunto de datos.
2. Estándares técnicos: estructura técnica en la que se ofrecen los datos.
3. Acceso: conexión con la información.
4. Modelo de datos: esquemas de metadatos aplicados al conjunto de datos.

La métrica se basa en las 4 condiciones que debe cumplir la información para que sea reutilizable y que coinciden con las dimensiones anteriormente citadas:

1. Inexistencia de barreras técnicas para la reutilización.
2. Posibilidad de acceso automatizado a la información.
3. Existencia de un esquema legal que permita su uso.
4. Acceso al conocimiento de la estructura de la información publicada.

Dentro de cada una de las dimensiones se establece una serie de niveles que permiten establecer en qué punto concreto y objetivo se encuentran los datos dentro de cada área y que permitió establecer los procesos para conseguir el máximo nivel de reutilización en cada caso.

El proceso de evaluación de Meloda, se basa en la ponderación según los niveles dentro de cada dimensión, estableciendo cuatro pasos : evaluación, ponderación, obtención de valor y calificación del conjunto de datos en cuanto a su reutilización.

La obtención del valor se calcula como la raíz cuarta del producto de las puntuaciones obtenidas en cada dimensión. Una vez aplicada la fórmula se puede establecer una valoración objetiva en cuanto a la reutilización del conjunto de datos:

0-25% Inadecuado para la reutilización

25-50% Reutilización básica

50-75% Reutilización avanzada con aspectos mejorables

75-100% Reutilización avanzada

El siguiente paso tras evaluar la colección fue establecer los requisitos mínimos para la propuesta de publicación a modo de guía para evaluar el CMS apropiado que permitiría publicar el conjunto de datos bajo un criterios concretos que asegurasen un nivel aceptable de adaptación de los datos a los requisitos técnicos de LOD.

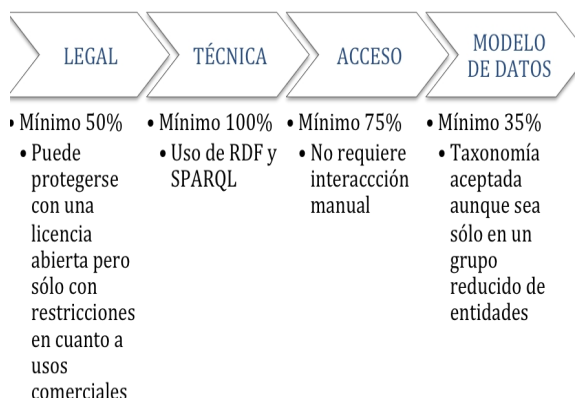


Figura 3. *Requisitos mínimos para publicar LOD.*

Fuente: elaboración propia

Para la parte de análisis descriptivo se usaron las pautas y recomendaciones de dos documentos: el informe final del W3C *Library Linked data Incubator Group* (Baker et al, 2011) que enmarcó las buenas prácticas y los objetivos que debían seguir las bibliotecas (y otros centros de información) para favorecer y publicar sus colecciones de datos; sin embargo, este informe no descendía al detalle, ya que se trataba de recomendaciones de carácter genérico.

En 2012 se publicaron las Recomendaciones *LODE v2*, elaborada por Marcia Lei-Zeng e Imma Subirats-Coll, que ayudaban en la selección de las estrategias más apropiadas para codificar de acuerdo con las necesidades de cada colección e institución que quiera publicar datos enlazados sobre registros bibliográficos.

2. Evaluación de la colección a través de las dimensiones de la métrica Meloda

La fonoteca digital de la Universidad de Alicante se encuentra integrada junto al conjunto de proyectos de conocimiento abierto de la Biblioteca de la Universidad de Alicante (11). Esta colección permite la consulta y la audición en línea de una amplia selección de fondos musicales grabados originalmente en discos de vinilo. Su objetivo es dar a conocer grabaciones de difícil acceso, así como las portadas de los discos que las contenían, como una forma de ampliar el conocimiento de una época al acercarse, de otra manera, a la música que se escuchaba en España en la segunda mitad del siglo XX.

La situación actual de la colección antes de abordar la investigación se puede resumir en los principales puntos analizados: proceso técnico de los fondos, almacenamiento y acceso (ver, Figura 4. *Catalogación, digitalización y almace-*

namiento de la colección). Esta colección se encuentra catalogada, digitalizada y accesible a través del catálogo de la biblioteca universitaria.

Fondos musicales: discos de vinilo.

- Digitalización de las colecciones de discos de vinilo cedidos por la emisora Radio Alicante Cadena Ser y de la propia biblioteca.
- Espacio temporal: desde la aparición del disco de vinilo (mediados de los años cincuenta) hasta finales de los años noventa.
- La colección está compuesta por más 3000 títulos de vinilos digitalizados.

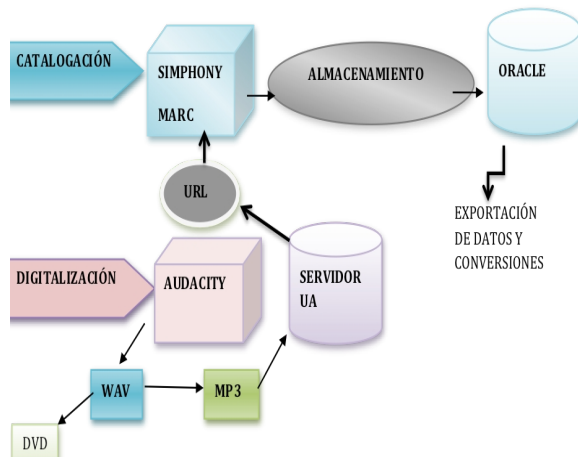


Figura 4. Catalogación, digitalización y almacenamiento de la colección. Fuente: elaboración propia

- Catalogación: se realiza a través del Gestor bibliotecario *Symphony* (12) en formato *MARC* para asociar la digitalización del audio al registro bibliográfico.
- Digitalización: grabación de formato analógico al digital, procesos donde el original digitalizado es un archivo en formato *WAV* que se almacena en un *DVD* para su preservación. También se convierte a *mp3* que es el objeto digital que se usa para el acceso y difusión vía Web. Hay que incluir la URL en el registro bibliográfico del recurso para completar la catalogación.
- Control de autoridades y materias: el control de autoridades de interpretes y grupos musicales se realiza a través del módulo de gestión propio de *Symphony*

y las materias mediante *EMBUS* (Encabezamientos de materia de la Universidad de Sevilla), dentro del mismo módulo de autoridades del programa de gestión

- Estilos musicales: para facilitar la difusión se creó un pequeño vocabulario con un código que permite agrupar las materias a través de un género mayor.
- Tras el estudio del contexto se evaluó la colección a través de la métrica Meloda.
- Dimensión Legal: se analiza la dimensión en dos niveles: para el conjunto de objetos digitales (*mp3*) y para los datos descriptivos ('datasets'),
- Conjunto de 'datasets', actualmente no existe ninguna restricción legal sobre el mismo, lo que permite poner una licencia a los datos de sólo atribución, permitiendo la reutilización comercial de los mismo.
- Conjunto de ficheros digitalizado en formato *mp3*: según la licencia actual se permite el acceso en línea para escucha en 'streaming' sin posibilidad de descarga, permitiendo enlace a las obras pero no alterar la ubicación del objeto digital dónde deberá constar mención de los derechos reservados.
- Estándares Técnicos: los datos están almacenados en una base de datos relacional creada con el SGBD Oracle. Es software propietario por lo que será necesaria una migración a otro formato más adecuado (*XML*, *RDF*, etc.).
- Acceso: la forma de acceso es a través del catálogo el acceso mediante la url única para todo el catálogo de la biblioteca (13) cada conjunto de datos tiene definidos dos tipos de enlaces: (a) Internos: autor, materias, estilos, títulos de canciones y (b) Externos: URL única para cada parte del audio y caratula de los discos. Su acceso se realiza a través de la página de aplicaciones de la Universidad.
- Modelo de datos: esta dimensión valora los modelos de datos, diferente al formato o estándar técnico en cuanto a que evalúa el nivel de descripción de cada conjunto de datos, la descripción de los datos está realizada en *MARC21*

es necesario una conversión a *Dublin Core* y el uso de ontologías.

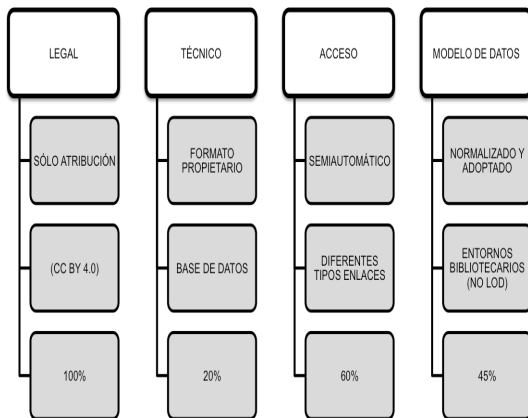


Figura 5. Evaluación de la colección a través de las dimensiones de Meloda. Fuente: elaboración propia

Aplicando la fórmula obtenemos que el conjunto de datos se encuentra en una fase de reutilización básica. El siguiente paso fue comparar los niveles de Meloda a los mínimos necesarios para publicar la colección como datos abiertos y enlazados, lo que permitirá establecer los requisitos necesarios para la adecuación de la colección a los estándares W3C.

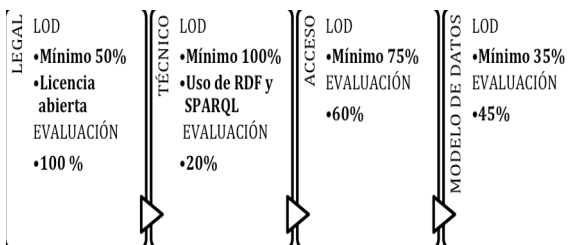


Figura 6. Requisitos mínimos para publicar LOD comparados con la evaluación de la colección. Fuente: elaboración propia.

3. Gestión de contenidos semánticos en Ximdex

Ximdex CMS es un sistema de gestión de contenidos semántico que permite la manipulación y generación de contenido, datos y aplicaciones web en diferentes tecnologías: PHP, J2EE, .NET, XML/ XSLT, JS, RDF, JSON, RoR, etc. Ha sido desarrollado por *Open Ximdex Evolution S.L* (14), empresa sevillana especializada en el desarrollo de sistemas de gestión de contenidos. Fue fundada en el año 2010 con el objetivo de crear un sistema de gestión de contenidos semántico y de software libre. *Ximdex*, está principalmente escrito en *PHP* y *JAVA* y usa ex-

tensivamente *XML* y *XSLT* para representar y tratar el contenido; desde 2014 se ha convertido en una plataforma que sirve para fusionar tanto datos como contenidos integrando buscador, y análisis y visualización de datos.

Se puede definir como “gestor distribuido para la gestión de datos, contenidos y servicios convirtiendo los mismos a diferentes formatos finales en múltiples tecnologías e incluso usando simultáneamente estas” (Prieto, 2014). Es un software neutral, libre y abierto, tanto en el uso de la herramienta al no limitar el número de proyectos, usuarios, idiomas, canales, tecnologías y lenguajes de programación (java, python, PHP, XML, XHTML, HTML5, .net, Ruby, django, etc.), como de su estructuración de la información, sin limitación alguna de definición o modificación del código fuente.

Permite elegir los componentes tecnológicos, establecer que contenidos son estáticos y cuales dinámicos, uso o no de base de datos, el tipo de conector a emplear, que servidor de aplicaciones se usa, usando estándares (XML, XSLT). *Ximdex* se sustenta en los siguientes 4 pilares:

1. **Abstracción:** para separar los contenidos tanto de la presentación como de la lógica de negocio, mejorar la reutilización de los mismos, representar servicios que pueden incluirse y personalizarse por el redactor sin programar, etc.
2. **Significado:** vinculado a los elementos de información que componen un contenido o servicio (párrafos tipo entrada o documentos tipo noticia, elementos multimedia, etc) lo que permite mejorar la recuperación en el buscador, automatizar la sugerencia de enlaces e imágenes relacionadas, etc.
3. **Acceso abierto a la información** permitiendo publicar datos en bruto para su posible reutilización y marcado semántico (RDF, microformatos, búsqueda semántica).
4. **Estructura libre:** sin imposición de una tecnología concreta o lenguaje de programación.

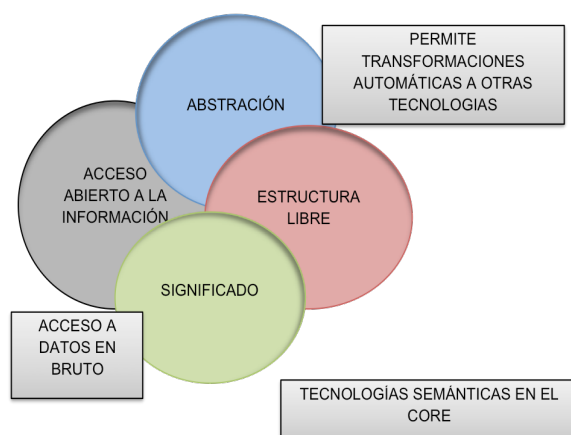


Figura 6. *Pilares de Ximdex*. Fuente: elaboración propia basada en la descripción de Ximdex

En la descripción de los módulos semánticos de *Ximdex* encontramos una serie de módulos a parte de los propios de cualquier gestor de contenidos imprescindibles para publicar datos abiertos y enlazados mediante el gestor:

- *Ximporta* proporciona un mecanismo de importación masiva de documentos desde formatos típicos como MS-Office, OpenOffice, PDF, etc. además de incluir el documento en el gestor, proporciona una translación automática a un esquema XML previamente definido, lo que garantiza el pleno uso del contenido en Ximdex de forma similar a si hubiera sido creado desde cero o su reutilización por otros módulos del sistema. Así, unido al módulo *Xlyre* para la gestión de datos abiertos enlazados, permite agregarlos, transformarlos, anotarlos y enriquecerlos semánticamente (mediante técnicas de entrelazado LOD contra servidores de ontologías, Freebase, Apache Stanbol, DBpedia, etc.), filtrarlos, reutilizarlos en otros 'datasets' (o colecciones de datos) y/o publicarlos en distintos formatos (RDFa, XML, html, PDF, etc.).
- *Ximtax* es un módulo que permite la creación y edición de metainformación asociada tanto a recursos internos gestionados por Ximdex (proyectos, contenidos, formularios, fichas, noticias, etc.) como a recursos externos (imágenes, páginas Web, etc.). A través de este entorno gráfico, el usuario puede añadir y editar libremente metainformación a los contenidos gestionados. Toda esta in-

formación semántica es recogida internamente en base a la ontología *Schema.org*, permitiendo la interoperabilidad con elementos anotados por otros sistemas estándar, mediante la traducción entre los distintos esquemas de representación. Permite la explotación de la metainformación del contenido proporcionando, por ejemplo, rutas alternativas de navegación para un portal, enriqueciendo los contenidos existentes, o mejorando el proceso de localización de los contenidos.

- *Ximfind* es un módulo que proporciona funcionalidades de indexación y búsqueda de los contenidos. La información indexada por este módulo puede ser de tipo no estructurado (HTML, PDF, MS Office Word, etc.), estructurado (XML) o de tipo semántico (RDF). Combina diversas tecnologías de recuperación de la información e indexación (nativa XML, a partir de crawlers, de procesado de metainformación, de microformatos, etc.) para proporcionar un acceso centralizado. Además, sus profundas raíces en la Web Semántica permiten vincular los resultados de las búsquedas a ontologías y taxonomías. Incluye un lematizador optimizado para castellano así como un interfaz de administración. Es un módulo independiente de Ximdex y puede ser utilizado por separado. Con Ximdex permite ampliar las capacidades de búsqueda dentro del gestor.
- El módulo Ximedit, integrado en la base, es un editor visual para XML que muestra en tiempo real como quedará el documento una vez publicado en el portal y, lo que es más importante, garantiza que el documento esté correctamente estructurado en base al esquema definido. En unión al módulo Xowl aplica tecnologías semánticas para enriquecer, de forma automática, nuestros contenidos, proporcionando imágenes y textos en consonancia con lo ya escrito. A partir de las relaciones y entidades semánticas permite la publicación dinámica de la información semantizada ('dynamic semantic publishing'), sintetizando de forma automática portales multicanal de información estructurada.

3.1. Evaluación de *Ximdex* mediante los resultados para publicar LOD

Se realizó una evaluación a partir del análisis anterior para establecer los mínimos necesarios en las diferentes dimensiones evaluadas y las necesidades detectadas en la evaluación de la colección

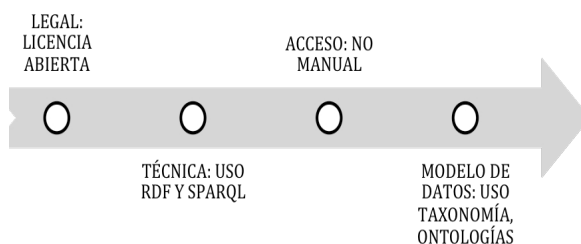


Figura 7. Requisitos mínimos para publicar datos abiertos enlazados Fuente: elaboración propia

- Marco Legal: puede protegerse con una licencia abierta con restricciones sólo en cuanto a usos comerciales, en cuanto a la colección debemos separar dos tipos de licencias la del registro bibliográfico y el conjunto de audios descritos. *Ximdex* permite establecer el tipo de licencia para cada conjunto de 'datasets' dentro de la sección Open Data de *Xylire*. El tipo de licencia debe quedar representado dentro del modelo de datos, como elemento de los metadatos descriptivos.
- Estándares Técnicos: uso de RDF y Sparql. Actualmente, los datos están almacenados en una base de datos objeto-relacional Oracle. Es software propietario, por lo que en esta dimensión estamos en el primer nivel y será necesario una migración a otro formato adecuado a LOD (XML, RDF, etc.).
- En *Ximdex* el módulo *Ximporta* permite la transformación automática de conjuntos de datos (distribuciones en un mismo formato) a un esquema XML previamente definido, junto con el módulo de datos abiertos enlazados *Xlyre* permite agregarlos, transformarlos, anotarlos y enriquecerlos semánticamente.
- En cuanto a Sparql, el buscador *Xfind* tiene carácter comercial y no tiene implementado *Sparql Endpoint* de forma pública para evitar problemas de escalabilidad, pero está previsto activarlo para la siguiente versión de *Ximdex* será publicada en el tercer trimestre de 2015.

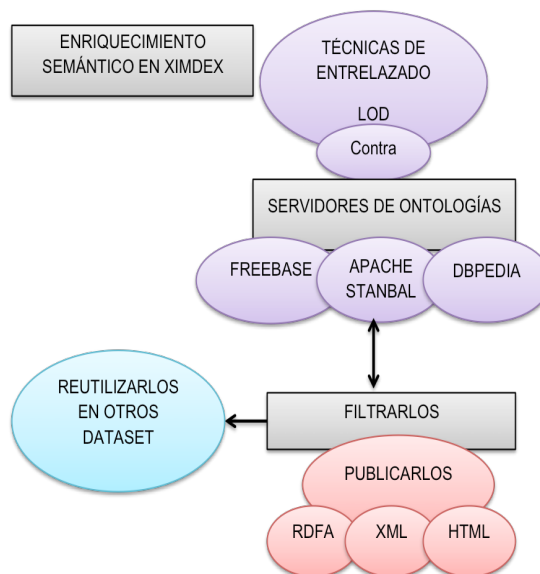


Figura 8. Proceso de enriquecimiento semántico y enlazado externo en *Ximdex* y formatos de publicación Fuente: elaboración propia.

- Acceso: en esta dimensión nos encontramos entre dos niveles. Por un lado tenemos un punto de acceso para el conjunto de datos (entendido como toda la colección), pero no para cada conjunto individual (registro bibliográfico). Por otro lado, dentro de un conjunto de datos sí encontramos URL para partes del documento pero no puntos de acceso para cada tripleta de interés del registro bibliográfico, los CMS se construyen con *Frameworks Web* siguen las mismas buenas prácticas para el nombrado de recursos con URIs *http*.

4. Resultados

Los resultados principales obtenidos del análisis y evaluación basados en los objetivos de la investigación (establecer una propuesta para publicar datos abiertos enlazados para un conjunto de registros bibliográficos de la colección estudiada), fue conocer los procesos necesarios para publicar LOD, se analizaron las características de una manera descriptiva y sistemática a través de Meloda, que nos permitió conocer la situación actual en cuanto al conjunto de datos y establecer los requisitos mínimos necesarios para publicar siguiendo los estándares de la W3C.

Era necesario seleccionar las herramientas y tecnologías necesarias para cada parte del proceso. Los resultados del estudio previo permitie-

ron establecer los problemas para publicar *LOD* en entornos bibliotecarios, así como las pautas y normas para llevarlo a cabo. Los resultados proporcionaron una perspectiva general de la situación que permitía seleccionar las herramientas necesarias adecuadas a los problemas concretos que presenta el entorno, los principales problemas del entorno son:

1. Falta de recursos económicos y humanos.
2. Resistencia al cambio.
3. Falta de conocimientos técnicos.
4. Falta de herramientas tecnológicas.
5. Dependencia de sistemas comerciales.
6. Licencias de los datos.
7. Los principales proyectos parten de grandes instituciones.

El *CMS* semántico *Ximdex* se postula como una posible solución para publicar en entornos bibliotecarios. El *W3C*, consideraba en su informe a los *CMS* como herramientas válidas para publicar conjuntos de datos, ya que se adaptaban a los estándares generales de la Web Semántica. Por otro lado, los problemas que presentaban los *CMS* para publicar aún eran evidentes debido a su amplia variedad tecnológica y a la complicada integración con módulos y la elección de los mismos. Esos principales problemas detectados se pueden resumir en:

- Exportación de los datos almacenados en el entorno de bases de datos relacionales y software propietario.
- Falta de abstracción de los *CMS*, dependencias tecnológicas, lenguajes de programación.
- Dificultad de aprendizaje.
- Dificultad de elección en cuanto a módulos y de evaluación en cuanto a integración de los módulos.
- Versiones beta.

En general existe demasiada dispersión tecnológica, muchas herramientas entendidas como piezas que - necesariamente - han de ser ensambladas y aprendidas.

Se analizó el *CMS semántico Ximdex* con la suficiente abstracción y con módulos integrados. Después del análisis se comparó con los resultados de la evaluación y la propuesta de mínimos a aplicar a la colección. Los resultados en esta fase son la propuesta técnica y el prototipo

para publicar la colección como datos abiertos y enlazados.

Los resultados del análisis y evaluación de ambas dimensiones nos proporcionan por un lado los procesos para adaptar y publicar la colección en los dos niveles y, por el por otro lado, la propuesta para publicar *LOD*.

4.1. Proceso para publicar datos abiertos enlazados

Se pueden definir dos dimensiones dentro del proceso para publicar *LOD*: proceso para preparar los datos y la parte técnica de exportación y publicación en *Ximdex*.

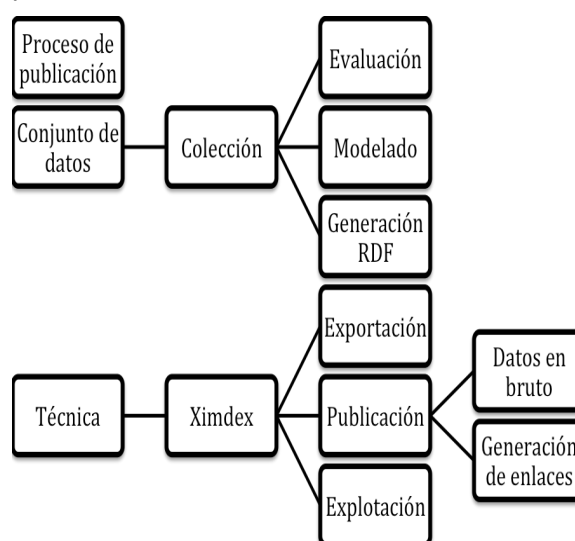


Figura 9. Fases del proceso para adaptar el conjunto de datos y publicar *LOD*. Fuente: elaboración propia

4.1.1. Evaluación

La evaluación de la colección a través de Meloda permitió establecer los problemas y necesidades para publicar. Los principales problemas son actualmente la dependencia de software y la adaptación a los estándares de la *W3C*.

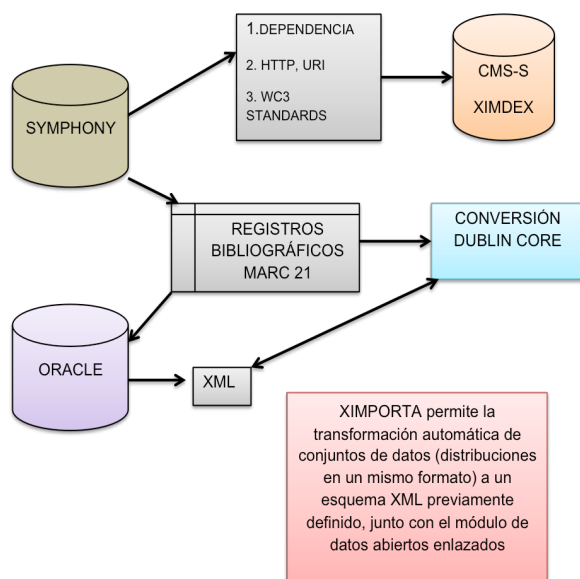


Figura 10. Problemas y necesidades de la colección para publicar LOD en Ximdex. Fuente: elaboración propia

4.1.2. Modelado

Los resultados principales de la evaluación sobre el modelado permitieron establecer las necesidades de la colección en cuanto al modelado de datos, la conversión del formato Marc21 a *Dublin Core*, mediante las Recomendaciones LODE y la conversión definida por la *Library of Congress*, definir el esquema DC XML y crear la distribución en *Ximdex* transformándola a través de *Ximporta*.

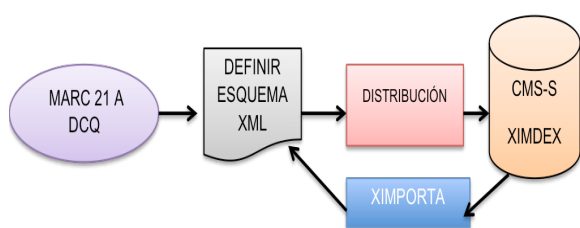


Figura 11. Proceso de modelado para adaptar el conjunto de datos a LOD. Fuente: elaboración propia

Para el modelado se usaron las *Recomendaciones LODE v2*, elaboradas por Marcia Lei-Zeng e Imma Subirats-Coll, que nos ayudaron en la selección de las estrategias más apropiadas para codificar de acuerdo con las necesidades de la colección. Usa diagramas para facilitar la toma de decisiones y escoger las propiedades más usadas en la descripción de los recursos.

El documento está basado en una estructura que tras unas recomendaciones de carácter general y un repaso a los esquemas de metadatos más importantes, despliega una serie de árboles de decisiones a partir de los cuáles es posible establecer propuestas para las propiedades individuales.

Así, se proporcionan árboles para información sobre el título, para la entidad responsable, diferenciando entre creador, contribuidor y editor; para las características físicas; para la información sobre los ejemplares y su localización, etc.

En la siguiente figura se muestra un ejemplo de la conversión de *Marc21 a DCQ* de la propiedad título y título alternativo para la colección.

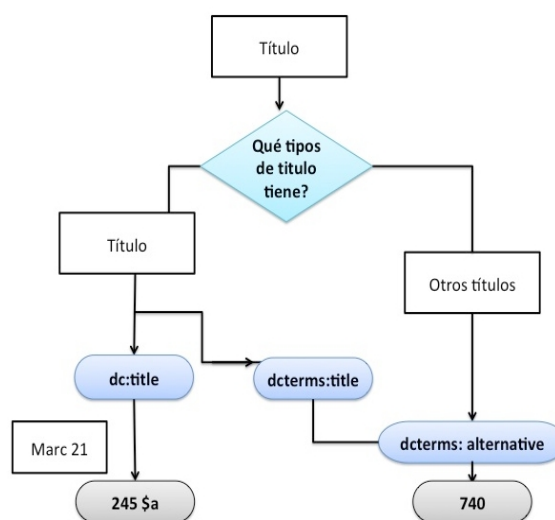


Figura 12. Ejemplo del proceso de modelado para la conversión de la colección de marc21 a DCQ. Fuente: elaboración propia mediante árboles de decisión de LODE v.2

La siguiente imagen muestra un ejemplo de la codificación en *Marc21 XML* formato en el que se encontraba la colección y había que convertir a *DCQ* para exportar los conjuntos de datos al repositorio semántico de *Ximdex*. En la figura 14 se puede observar un ejemplo de una parte del documento XSL usado para la conversión a *DCQ* para la etiqueta título.

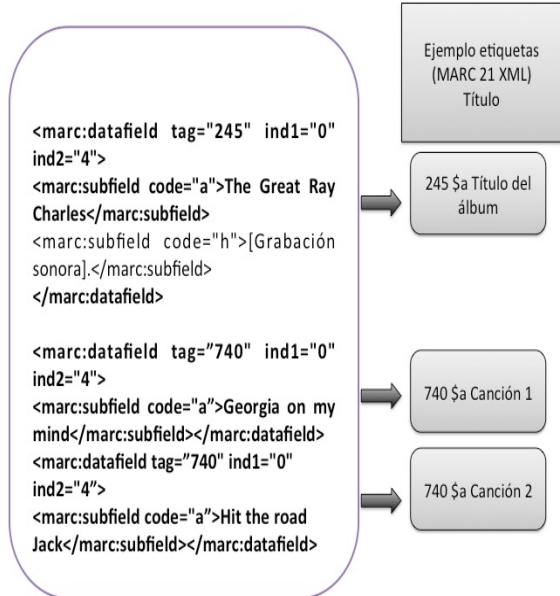


Figura 13. Ejemplo de etiquetas título Marc21.xml. Fuente: elaboración propia

```

<xsl:for-each select="marc:datafield[@tag=245]">
<dc:title>
<xsl:call-template name="subfieldSelect">
<xsl:with-param name="codes">abfghk</xsl:with-param>
</xsl:call-template>
</dc:title>
</xsl:for-each>
    
```

Conversión de Marc21 a DCQ ejemplo xsl (elemento: title)

Figura 14. Ejemplo de conversión mediante XSLde Marc21 a DCQ. Fuente: <http://www.loc.gov/standards/marcxml/xslt/MARC21slim2OAIDC.xsl>

4.1.3. Generación RDF

Los principales resultados en cuanto a la generación de RDF es que Ximdex permite, una vez importados los conjuntos de datos, tratar de forma manual la inserción de tripletas para todo el conjunto de datos este proceso se realiza mediante un sistema de actualización en caliente de tripletas que permite añadir y eliminar las mismas al repositorio semántico sin interrupción del servicio de datos.

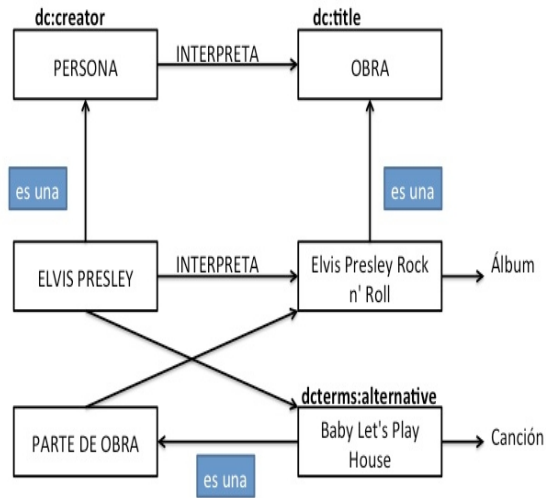


Figura 15. Ejemplo selección de tripletas para el elemento título de una obra musical. Fuente: elaboración propia

RDFa (7) permite añadir declaraciones RDF a documentos web en forma de propiedades de etiquetas de marcado (X)HTML(8) De este modo pueden extraerse datos RDF de dichos documentos las principales ventajas de utilizar RDFa son:

- Se hace innecesario separar secciones XML y HTML de los mismos contenidos
- Reutilización de datos
- Se pueden añadir campos adicionales
- Independencia de los estándares de publicación web

En cuanto a Sparql Xfind el buscador, tiene carácter comercial, no tiene implementado Sparql Endpoint (15) de forma pública para evitar problemas de escalabilidad, pero está previsto activarlo para la siguiente versión de Ximdex prevista para el tercer trimestre de 2015.

4.1.4. Exportación

El primer paso del proceso técnico es seleccionar como exportar el conjunto de datos desde la actual base de datos Oracle a la plataforma de Ximdex, existen dos posibles formas de exportación de los datos

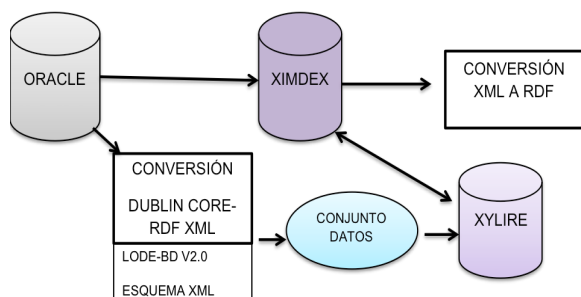


Figura 16. Proceso para la exportación del conjunto de datos a Ximdex. Fuente: elaboración propia

La ejecución puede ser manual o automática mediante cron. Para ello es requisito indispensable que exista el fichero y la ruta: [\[PROJECT\]/last-import/last.ini](#) El contenido de dicho fichero debe ser una línea donde se inicializa la variable 'last_imported' al valor de la última sesión importada, por ejemplo: last_imported = 60

El proceso se puede dividir en dos partes fundamentales:

- Implementación de esquemas y plantillas en *Ximdex*.
- Recolección de documentos *XML* y su importación en *Ximdex*.

Esquemas y plantillas

- *Ximdex* define sus esquemas de documentos *XML* basado en el formato *RelaxNG* definido por *Oasis* (como se observa en la figura 17 un paso imprescindible para la posterior conversión a *RDF* era definir el esquema *XML* en el que están descritos el conjunto de datos, la definición se realiza a través de *RelaxNG* que también permitirá validar cada nuevo *XML* que se incorpore al repositorio)
- Documentos *XSL* para transformar los documentos *XML* en páginas *HTML* para su correcta visualización en navegadores. Estas plantillas nos permiten recorrer todos los datos contenidos en los *XML* además de operar con ellos con el objetivo de añadirlos a la estructura de elementos *HTML* que se desee generar.

```

<?xml version="1.0" encoding="UTF-8"?>
<grammar xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://relaxng.org/ns/structure/1.0">
  <dc:title>The library element</dc:title>
  <dc:creator>Eric van der Vlist</dc:creator>
  <dc:subject>library, book, RELAX NG</dc:subject>
  <dc:description>This RELAX NG schema has been written as an example to show how
    Dublin Core elements may be used.</dc:description>
  <dc:date>2003-01-30</dc:date>
  <dc:language>en</dc:language>
  <dc:rights>Copyright Eric van der Vlist, Dyomedeia.
    During development, I give permission for non-commercial copying for
    educational and review purposes. After publication, all text will be
    released under the Free Software Foundation GFDL.</dc:rights>
  ...
</grammar>
  
```

Figura 17. Definición del esquema DCQ en RelaxNG Fuente: disponible en:

<http://books.xmlschemata.org/relaxng/relax-CHP-13-SECT-2.html>

4.1.5. Publicación

Los pasos previos a la publicación a parte de la instalación son:

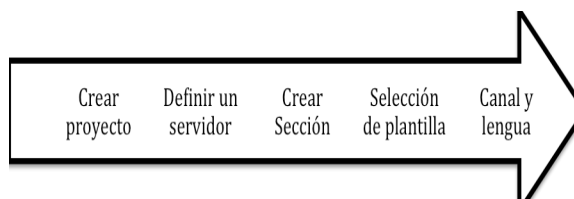


Figura 18. Pasos previos a realizar en Ximdex anteriores a la publicación Fuente: elaboración propia

Los resultados en cuanto a publicación se pueden dividir en dos:

- Publicación de datos en bruto a través del módulo Xylire.
- Enriquecimiento semántico (generación de enlaces) a través de XOWL.

Las distribuciones dentro del módulo Xylire nos permiten crear fichas tipo, con campos para cada uno de los registros bibliográficos, la idea de XOWL es un "agregador de 'datasets' semánticos" por tanto semantiza los datos con su propio modelo ontológico. Es sencillo soportar vocabularios u otras ontologías aunque hay que realizar un 'mapping' (por ejemplo, SKOS o Schema (10) se han cargado en otros proyectos desarrollados en Ximdex).

XOWL permite enlazar datos con servidores ontológicos como Bdpedia de una manera automática y publicarlos en diversos formatos (Ver: Figura 8. Proceso de enriquecimiento semántico y

enlazado externo en Ximdex y formatos de publicación)

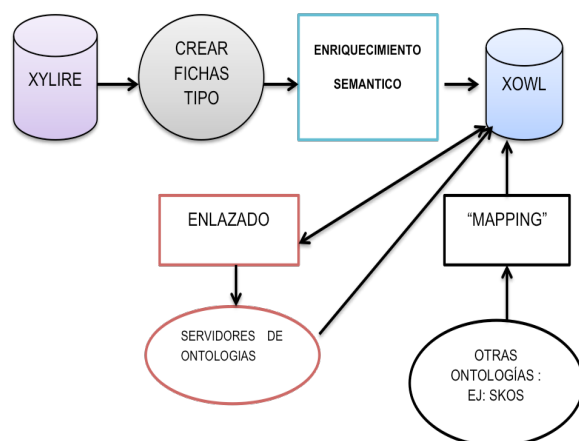


Figura 19. Proceso para la publicación del conjunto de datos en Ximdex. Fuente: elaboración propia

5. Conclusiones

Linked data se plantea como una solución a la falta de visibilidad web para las colecciones bibliotecarias. En un principio la Web Semántica y *LOD* parecen transparentes a los usuarios, pero las posibilidades en cuanto a aplicaciones de consumo y explotación de los datos no lo son.

La adopción por parte del entorno se hace crucial, el éxito de esos proyectos, la experimentación y visibilidad de aplicaciones, puede hacer que los datos enlazados en bibliotecas se adopten, de la misma manera que se asumió la necesidad de que el conocimiento fuera abierto y se adoptara de manera masiva mediante el uso de repositorios Open Data. Se percibe también una brecha en cuanto a la publicación, ya que la mayor parte de las iniciativas parten de grandes instituciones con más recursos para enfrentarse al cambio de paradigma, por ello era necesario encontrar soluciones tecnológicas que facilitarían la publicación en instituciones medianas y pequeñas más reacias al cambio, que no sólo facilitará su publicación, sino la evaluación de sus componentes.

Se seleccionó una colección dentro del catálogo de la Biblioteca de la Universidad de Alicante. Es más fácil comenzar a experimentar y realizar aplicaciones en subconjuntos de los catálogos, colecciones atractivas que puedan reflejar la potencialidad de los datos enlazados en bibliotecas y promover la adopción por parte de instituciones similares.

Los CMS se plantean como una solución para publicar datos en bibliotecas y superar una de las limitaciones que sufren actualmente las bibliotecas dependientes de software propietario que no permiten desarrollo de terceros y que dificulta compartir experiencias y proyectos piloto. Los problemas que presentaban los CMS para publicar aún son evidentes demasiada variedad tecnológica, problemas de integración con módulos. La dispersión tecnológica: demasiadas herramientas, entendidas como piezas que necesariamente han de ser analizadas, ensambladas y aprendidas, y que no facilitan la publicación en el entorno que nos ocupa con una fuerte resistencia al cambio y falta de conocimientos tecnológicos, a decidir publicar las colecciones bibliográficas como *LOD*.

Ximdex es un CMS nativo semántico las características principales son el marcado semántico automático, la generación de páginas que incluyen ya (HTML y RDF) en la misma URL mediante *RDFa*, publicación de open data (colecciones de datos en bruto) y la posibilidad de implementar ontologías a través del lenguaje OWL incipiente, aunque no permite actualmente importar ontologías, su arquitectura puede incluirlo fácilmente aunque a través de un mapeo.

Una de las características que hace interesante al *CMS Ximdex* y lo distingue de otros CMS con mayor adopción es la independencia tecnológica de lenguajes de programación y la posibilidad de conectar los datos externos almacenados de la plataforma a través de cualquier bases de datos o repositorio. Tiene un alto nivel de abstracción que permite no sólo la compatibilidad con sistemas actuales sino también la posibilidad de adoptar otros futuros.

Ximdex no es una solución perfecta para publicar *LOD* en entornos bibliotecarios requiere de desarrollos específicos para que resulte más fácil publicar registros bibliográficos, falta desarrollo en cuanto a *Sparql endpoint* que actualmente no es público (previsto para la próxima versión del *CMS*) y la necesidad de conectar automáticamente con otros servidores ontológicos propiamente bibliotecarios que permitan aprovechar y reutilizar los vocabularios ya creados en nuestro campo. Una de las fortalezas de *Ximdex* es la integración de módulos, algunos externos al core pero testeados por los propios desarrolladores, no tiene dispersión tecnológica, los desarrollos por terceros son probados antes de incluirlos dentro del *CMS*.

La conversión a *RDF* y *RDFa* son prácticamente automáticas aunque hay que definir el esquema *XML* y seleccionar las tripletas manualmente, para desechar las que no resultan interesantes, el enriquecimiento y enlazado a otros datos LOD es automático y la generación de páginas se puede realizar en cualquier formato.

La evaluación con Meloda ha permitido conocer en que situación se encontraba la colección y proponer los requisitos mínimos necesarios para la publicación. Se han establecido los procesos necesarios a través de la herramienta *Ximdex* para poder realizar una propuesta de publicación para la colección estudiada mediante el *S-CMS*

En la fase de explotación, la investigación no ha desarrollado una interfaz gráfica para explotar los datos, para diseñar un modelo para la explotación se estudiaron dos proyectos ya en producción; el catálogo de datos abiertos de la *BNE* y *Linked jazz*, este apartado necesita un análisis más profundo que el realizado, es necesario adaptar las características de visualización a la colección, no se ha desarrollado lo suficiente en el estudio y podría abordarse en proyectos futuros.

Notas

- (1) The World Wide Web Consortium (W3C): es una comunidad internacional que trabaja para desarrollar estándares web, en: www.w3.org/2001/sw/
- (2) Un framework para aplicaciones web es un marco de trabajo diseñado para apoyar el desarrollo de sitios web dinámicos, aplicaciones y servicios web.
- (3) URIs [http](http://) : Un identificador de recursos uniforme es una cadena de caracteres que identifica los recursos de una red de forma unívoca, en el caso de [http](http://), es el protocolo de acceso al recurso. *What do HTTP URIs Identify?* En: <http://www.w3.org/DesignIssues/HTTP-URI.html>
- (4) *RDFa*: es un conjunto de extensiones de *XHTML* propuestas por W3C para introducir semántica en los documentos. *RDFa* aprovecha atributos de los elementos meta y link de *XHTML* y los generaliza de forma que puedan ser utilizados en otros elementos, en: <http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/>
- (5) *FRBR*: Requisitos funcionales para registros bibliográficos, es un modelo conceptual de entidad-relación desarrollado por la IFLA en: <http://www.ifla.org/files/assets/cataloguing/frbr/frbr-es.pdf>
- (6) *RDA*: Es el nuevo código de catalogación que ha sustituido a *AACR2*, se basa en los modelos conceptuales desarrollados en los últimos años por

IFLA, particularmente *FRBR* y *FRAD*, en: <http://www.loc.gov/aba/rda/>

- (8) "Marc must die" iniciativa para promover la sustitución de *Marc* por otros estándares más adecuados a las necesidades actuales de catalogación, en: http://marc-must-die.info/index.php/Main_Page
- (9) Dublin Core (DC) modelo de metadatos elaborado y auspiciado por la DCMI (Dublin Core Metadata Initiative) en: <http://dublincore.org>
- (10) Meloda: métrica para evaluar la reutilización de datos abiertos, Metric For Relasing Open Data (MELODA). Version 3.10 <http://www.meloda.org/full-description-of-meloda/>
- (11) Fonoteca de la Biblioteca de la Universidad de Alicante, en: <http://biblioteca.ua.es/es/fonoteca/fonoteca-digital/fonoteca.html>
- (12) Sistema de Gestión Bibliotecaria Symphony: Gestor bibliotecario de software propietario para la gestión de bibliotecas físicas: <http://www.sirsidynix.com/>
- (13) Acceso único a todo el fondo del catálogo de la Biblioteca de la Universidad de Alicante: <http://gaudi.ua.es/uhtbin/cgiirsi/?ps=iKhECuyo-KY/x/250160063/38/1/X/BLASTOFF>
- (14) Open Ximdex Evolution S.L. desarrolla aplicaciones en la nube para la gestión de contenidos y datos atendiendo a los últimos estándares del Consorcio W3C, en <http://www.Ximdex.com/es/>
- (15) SPARQL Query Language for RDF en <http://www.w3.org/TR/rdf-sparql-query/>
- (16) Schema.org: son un conjunto de etiquetas y atributos (micro-datos) añadidos al código HTML que le indican a los buscadores el significado semántico de nuestros contenidos <http://schema.org/>

Referencias

- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., & Zeng, M. (2011). *Library Linked data incubator group final report*. Disponible en: <http://www.w3.org/2005/Incubator/ldd/XGR-Lld-20111025/> [20-10-2014]
- Berners-Lee, T. (2000) *Tejiendo la red*. Madrid: Editorial Siglo XXI.
- Bizer, C., Heath, T. & Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.
- Bratsas, C., Bamidis, P., Dimou, A., Antoniou, I., & Ioannidis, L. (2012). Semantic CMS and wikis as platforms for Linked Learning. In *2nd Int. Workshop on Learning and Education with the Web of Data-24th Int. World Wide Web Conference*. Disponible en <http://ceur-ws.org/Vol-840/03-paper-26.pdf> [09-11-2015]
- Connaway, L.S., & Dickey, T.J. (2010). The digital information seeker. Report of findings from selected OCLC, RIN and JISC user behaviour projects.

- Disponible en <http://www.webarchive.org.uk/way-back/archive/20140615023510/http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf> [09-11-2015]
- Eíto-Brun, R. (2014). *Gestión de contenidos : procesos y tecnologías para gestionar activos de información* (Vol. 17). Editorial UOC.
- Ferrer-Sapena, A., Peset Mancebo, F. y Aleixandre-Benavent, R. (2011). Acceso a los datos públicos y su reutilización: Open data y open government. *El Profesional de la Información*, 20(3), 260-269. Disponible en <http://eprints.rclis.org/20997/> [09-11-2015]
- Navarro Galindo, J.L. (2013). *FLERSA: un sistema semántico de gestión de contenido web (S-CMS)*. Tesis Doctoral. Universidad de Granada. Disponible en: <http://hera.ugr.es/tesisugr/2122481x.pdf> [09-11-2015]
- Hannemann, J. and Kett, J. Linked data for libraries. In: *76th IFLA general conf and assembly*, 10-15 August 2010, Gothenburg, Sweden. Disponible en: <http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf> [20-10-2014]
- Hernández Santiago, C. (2012). *La aplicación de Linked data en la agrupación "Datos Bibliográficos": análisis de los casos de uso presentados en el Library Linked data Incubator Group Final Report*. Disponible en: http://gredos.usal.es/jspui/bitstream/10366/121123/1/HERN%C3%81NDEZ%20SANTIAGO.%20CARMEN%20_TFM.pdf [10-11-2015]
- Martín, S.G. & Angelozzi, S.M. (2013). Datos abiertos enlazados y libros abiertos: Impacto en las bibliotecas y en el desarrollo de la sociedad de la información. En *42 JAIIO Jornadas Argentinas de Informática SSI 2013 - 11º Simposio sobre la Sociedad de la Información*. Córdoba, Argentina, 19 de Septiembre. Disponible en: <http://eprints.rclis.org/20153/> [10-11-2015]
- Martín, M.A., Olsina Santos, L.A., Vallecillo, A. y Bertoa, M.F. (2002). Hacia un enfoque semántico para la catalogación de métricas. En *VIII Congreso Argentino de Ciencias de la Computación*. Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/23031/Documento_completo.pdf?sequence=1 [10-11-2015]
- Mooney, Sean D. and Baenziger, P.H. (2007). Extensible open source content management systems and frameworks: a solution for many needs of a bioinformatics group. *Briefings in bioinformatics*, n. 9, pp. 68-74.
- Ortiz-de-Urbina-Criado, M., Abella, A. y De-Pablos-Heredero, C. (2014). Meloda, métrica para evaluar la reutilización de datos abiertos. *El Profesional de la Información*, 23(6), 582-588. Disponible en: <http://www.elprofesionaldelainformacion.com/contenidos/2014/nov/04.pdf> [1-11-2015]
- Pastor-Sánchez, J.A. (2012). Los CMS como pieza fundamental en el despliegue de la Web Semántica. *Anuario Think EPI*, v. 6, p. 184-189. Disponible en: <http://www.thinkepi.net/los-cms-como-pieza-fundamental-en-el-despliegue-de-la-web-semantica> [1-11-2015]
- Pastor-Sánchez, J.A. (2013). *Seminario sobre tecnologías de la web semántica para la publicación de vocabularios SKOS. XVIII Encuentros Internacionales Sobre Sistemas De Información y Documentación. Ibersid 2013*. Disponible en: <http://digitum.um.es/xmlui/bitstream/10201/36136/1/IBERSID-2013-taller.pdf> [24-11-2014]
- Pastor-Sánchez, J.A. (2015, 16 de enero). Tendencias en el ámbito de las bibliotecas digitales universitarias. Correo electrónico enviado a: IWETEL@listservrediris.es Disponible en: <https://deintresparaelbibliotecario.wordpress.com/2015/01/20/tendencias-en-el-ambito-de-las-bibliotecas-digitales-universitarias-por-juan-antonio-pastor-sanchez/> [10-11-2015]
- Peset Mancebo, M.F., Ferrer Sapena, A. y Subirats-Coll, I. (2011). Open data y Linked open data: su impacto en el área de bibliotecas y documentación. *El Profesional de la Información*, 20(2), 165-173. Disponible en: <http://www.elprofesionaldelainformacion.com/contenidos/2011/marzo/06.pdf> [10-11-2015]
- Prieto, J. Semantic Content & Data Management with Ximdex CMS. (2014, 15 octubre) OPENEXPO CMS, 15 de octubre de 2014.
- Saorín Pérez, T. (2012). Cómo Linked Open Data impactará en las bibliotecas a través de la innovación abierta. *Anuario ThinkEPI*, 6, p. 288-292. Disponible en: <http://eprints.rclis.org/16913/1/2012%20Anuario%20ThinkEpi%20288-292-J12-linked-open-data-bibliotecas-Saorin.pdf> [10-11-2015]
- Saorín Pérez, T., Peset Mancebo, F. & Ferrer-Sapena, A. (2013). Factores para la adopción de Linked Data e implantación de la Web Semántica en bibliotecas, archivos y museos. *Information Research*, 18 (1) paper 570. Disponible en: <http://InformationR.net/ir/18-1/paper570.html> [10-11-2015]
- Subirats-Coll, I., & Zeng, M.L. (2013). LODE-BD recommendations 2.0: Report on how to select appropriate encoding strategies for producing linked open data (LOD) enabled bibliographical data. Disponible en: <http://travesia.mcu.es/portaInb/jspui/handle/10421/7362> [10-11-2015]
- Ximdex. (2014) *Guía para la publicación de datos abiertos y enlazados: versión 4*. Disponible en: http://www.Ximdex.com/lib/bib/OXE_OpenData_Guide_ES.pdf [06-06-2015]
- W3C *Guía Breve de Web Semántica*. Disponible en: <http://www.w3c.es/Divulgacion/GuiasBreves/Web-Semantica> [06-06-2015]