

Apertura y publicación de datos bibliográficos con formatos de datos abiertos. Aplicación a un caso práctico

Ana María Feroso García María Isabel Manzano García Alvaro Armero López ..Alvaro Hernández Hernández

(afermosoga@upsa.es)

(mmanzanoga@upsa.es)

(alvaro.armero@gmail.com)

(alvaro2hdz@gmail.com)

Facultad de Informática

Biblioteca UPSA

Facultad de Informática

Facultad de Informática

Universidad Pontificia de Salamanca (UPSA)

Abstract

Este trabajo trata de dar a conocer la información sobre autoridades y figuras ilustres que han pasado y forman parte de la historia de la UPSA y que se almacenan en el catálogo bibliográfico de la biblioteca de la Universidad. Para ello, y ahí reside la novedad, se utilizarán formatos especializados de datos abiertos enlazados (LOD) para el ámbito bibliográfico, concretamente BIBFRAME. Este formato ha sido propuesto por la Library Of Congress de los Estados Unidos, la biblioteca más importante a nivel mundial, y a pesar de su reciente creación ya está convirtiéndose en un referente y es apoyado por los agentes más importantes en el ámbito de las bibliotecas. La ventaja de utilizar estos formatos LOD para publicar la información, es que ésta puede enriquecerse con la de otros recursos externos, y a su vez nuestros datos pueden ser enlazados y enriquecer otros recursos externos. Es decir, se transformarán registros de autoridad y bibliográficos relativos a profesores y personajes ilustres de la UPSA desde el catálogo tradicional en formato MARC21 a un formato de datos abiertos. Iniciando así el primer paso de un futuro portal datos.upsa.es.

Palabras clave

Bibframe, Linked Open Data, Registros bibliográficos y de autoridad, Datos enlazados

1. INTRODUCCIÓN

La apertura de datos, eje fundamental de lo que se conoce como gobierno abierto, es una demanda actual que afecta a administraciones e instituciones. Las universidades, como instituciones también al servicio de la sociedad, al igual que las bibliotecas y en general los archivos de información cultural, no pueden ni deben quedar fuera de este movimiento aperturista y de transparencia de datos.

La apertura de datos permite poner los datos a disposición de los ciudadanos para que puedan ser reutilizados y generar valor a partir de los mismos. Para poderle sacar el mayor partido a esta reutilización es necesario su publicación en formatos lo más reutilizables posibles. Esto es lo que proporcionan las tecnologías semánticas a través de los datos enlazados o Linked Open Data. Estos a su vez facilitan la publicación, búsqueda y enriquecimiento de la información.

Partiendo de estas premisas, el objetivo de nuestro trabajo es publicar como universidad en abierto, una parte de los registros bibliográficos de autoridades de nuestra biblioteca universitaria. Más no solo publicarlos, sino para ello primero ya catalogarlos en un formato abierto, y además enlazarlos o enriquecerlos con los de otros conjuntos de datos abiertos. El sistema propuesto además pretende ser transferible como caso de éxito, a cualquier otra biblioteca universitaria a nivel nacional o internacional. Las tecnologías Linked Open Data de este modo, permiten utilizar Internet, y más concretamente la Web de datos, proyecto al que aspira a convertirse la Web actual, al servicio de la difusión de la cultura, en este caso focalizada en las bibliotecas y archivos culturales.

El carácter innovador y la relevancia de la solución aportada radica por un lado, en la utilización de un formato de catalogación ya orientado a LOD para publicar la información bibliográfica, el formato BIBFRAME, creado por la Library of Congress, LOC en adelante, y apoyado por organismos, bibliotecas y universidades de prestigio a nivel mundial. Por otro lado, por sus prestaciones, al permitir catalogar, publicar y enriquecer los datos bajo el mismo sistema de gestión bibliográfica. Por último, por su relevancia social al ser transferible y válido para ser utilizado por cualquier biblioteca, solucionando así el problema de la integración de datos bibliotecarios, tanto dentro del propio entorno, como sobre todo fuera de él, pues los datos bibliográficos se podrán enlazar y por tanto enriquecer de forma automática con otras fuentes de datos externas.

En los siguientes apartados de este trabajo trataremos primero la descripción y objetivos del proyecto llevado a cabo. A continuación el estado de la cuestión en referencia a la apertura de datos bibliográficos. Finalmente detallaremos la solución propuesta en nuestro sistema, así como las conclusiones y líneas futuras resultantes del trabajo.

2. DESCRIPCIÓN Y OBJETIVOS DEL PROYECTO

La idea de este proyecto es promover la transformación en Linked Open Data de los catálogos bibliográficos tradicionales en formato MARC. Esta idea que en principio resulta bastante ambiciosa, pretende por ello empezar a modo de caso de uso, con los registros de autoridades de la biblioteca de la Universidad de la que formamos parte todo el equipo del proyecto, la Universidad Pontificia de Salamanca. A su vez como modelo a la hora de realizar esta catalogación, se va a utilizar la iniciativa de la Library of Congress de los estados Unidos, para este ámbito, conocida como BIBFRAME (Bibliographic Framework Initiative). BIBFRAME pretende no solo establecer un modelo de catalogación e intercambio de recursos alejado de los formatos bibliográficos tradicionales como MARC, sino también ser la base de los futuros sistemas de gestión de bibliotecas y otros archivos culturales.

Una aportación fundamental de nuestro proyecto radica en demostrar en entornos reales, en concreto el entorno universitario, la utilidad en particular de este formato BIBFRAME basado en LOD, y de las tecnologías Linked Open Data en general, en el ámbito bibliográfico y por ende también en la universidad, al aplicarlo al caso de nuestra biblioteca universitaria.

De forma gráfica, la idea de nuestro proyecto podría verse representada en la Figura 1, “*Adaptación de recursos bibliotecarios de nuestra Biblioteca (registro de autoridades UPSA) a LOD mediante tecnologías semánticas (RDF, OWL BIBFRAME, ...), permitiendo además su enlace y enriquecimiento con otros recursos de la Web de Datos*”



Figura 1. Esquema conceptual del proyecto

2.1. Objetivos

El objetivo principal del proyecto es proponer y desarrollar un sistema automático de catalogación, publicación y enriquecimiento de datos bibliográficos, bajo el paradigma Linked Open Data, basado en la utilización del formato BIBFRAME, creado por la Library of Congress, así como demostrar su utilidad para cualquier archivo cultural o biblioteca, con su aplicación a un caso real, el registro de autoridades de la biblioteca de la Universidad Pontificia de Salamanca.

Este objetivo principal, la obtención del propio sistema de gestión bibliográfica planteado, al concretarse para el registro de autoridades de nuestra biblioteca universitaria, se materializará en los siguientes objetivos secundarios

- La transformación y catalogación en el nuevo formato BIBFRAME de los datos de las autoridades universitarias ya existentes en distintas bases de datos de nuestra biblioteca.
- El enriquecimiento o enlace de estos datos con los de otras fuentes de datos abiertos, bibliográficas, biográficas, geográficas o genéricas como DBpedia, mediante la técnicas semánticas de enlace de datos
- Visibilidad de toda esta información enriquecida sobre nuestras autoridades, de la forma más amigable posible para el usuario final.

3. ESTADO DE LA CUESTIÓN

Hecha la introducción, sobre la idea central del proyecto, es importante sentar las bases o fundamentos de los que vamos a partir.

Lo primero, definir qué es Linked Open Data, lo segundo constatar la importancia de la apertura de la cultura y el conocimiento, a través de una universidad abierta en general y de sus datos bibliográficos en particular, y luego ya centrar el tema en el formato de catalogación y enlace a utilizar para publicación en LOD de datos bibliográficos, BIBFRAME, por comparación también con otros formatos y modelos bibliográficos.

- **Linked Open Data o datos abiertos enlazados**

Linked Open Data (LOD) responde como su nombre indica a la unión de dos conceptos, datos abiertos y datos enlazados.

La apertura de datos es un elemento fundamental de lo que también se conoce como gobierno abierto. Las administraciones e instituciones ponen gratuitamente a disposición de los ciudadanos e instituciones sus datos, para que estos los reutilicen y generen nuevos servicios de valor añadido.

Linked data, o datos enlazados en castellano, hace referencia a la utilización de las tecnologías semánticas a la hora de publicar la información en la Web, concretamente utilizando el formato RDF, estándar para publicación de datos en la Web. Este estándar permite que los datos puedan interoperar.

Cuando ambos conceptos se unen resulta el término Linked Open Data, que significa que los datos abiertos se publican en formato de datos enlazados o formato semántico. Esto es lo que Tim Berners Lee denomina formato cinco estrellas (<http://5stardata.info/es/>). Esto en la práctica significa que los datos abiertos se pueden enlazar o enriquecerse con los de otras fuentes de datos abiertas. Esta utopía de la "nube de información global" que persigue la Web de Datos o Web Semántica, puede ser una realidad a través de los datos vinculados en bibliotecas, creándose una plataforma que permita acceder a la información de estas fuentes culturales de forma gratuita y sin costes de infraestructura.

- **Universidad y bibliotecas abiertas**

Tanto universidades como bibliotecas, como el resto de instituciones culturales, están inmersas en un profundo proceso de cambio, provocado en gran medida por las tecnologías de la información. Estas instituciones deben adherirse al movimiento aperturista y colaborativo de su información, y Linked Open Data, es el mejor instrumento para conseguirlo. LOD por sus importantes ventajas en la gestión de datos, principalmente ofrece la posibilidad de compartirlos, de hacerlos reutilizables y por tanto de favorecer el enriquecimiento potencial de la información al hacer interconectables los conjuntos de datos, fin último de la Web semántica.

“Nos encontramos en los inicios de un probable camino hacia el gobierno abierto, empezando por la transparencia, que se practica idealmente mediante la apertura de datos. Esta nueva forma de relación entre las instituciones, y la ciudadanía está siendo poco a poco demandada por ésta, pero sobre todo ha sido identificado como necesario, y beneficioso, en algunas instancias políticas supranacionales, o por algunas administraciones nacionales o locales pioneras.

Las universidades, no debieran quedar rezagadas en este camino, complementario a la arquitectura del conocimiento que está en su naturaleza, y que también está adquiriendo nuevas formas.”

Con estas palabras finaliza en sus conclusiones el documento de la CRUE-TIC titulado “Hacia una universidad abierta” () en sus recomendaciones para el Sistema Universitario de Educación (SUE), documento que hemos querido tomar también como referencia en nuestro proyecto.

El objeto de este documento es que dentro de las universidades también se promueva y facilite lo que se denomina gobierno abierto, que en este caso también incluye conocimiento abierto, para alcanzar como objetivo último la construcción de una universidad abierta.

Esto en la práctica se concreta fundamentalmente en promover que las universidades se unan a las políticas de transparencia abriendo sus datos institucionales de forma coherente.

Ya existen universidades que han empezado esta andadura, más aún podemos ser pioneros en esta tarea, y esto también es una motivación.

En el documento mencionado se dan las pautas sobre por qué datos empezar a publicar, así como recomendaciones de cómo hacerlo y qué estándares seguir. En este sentido queremos alcanzar y comenzar nuestra andadura de apertura de datos, utilizando desde el comienzo los formatos 5 estrellas (http://www.ted.com/talks/tim_berniers_lee_the_year_open_data_went_worldwide), los formatos semánticos, los más completos y reutilizables. Es decir no se tratará únicamente de empezar a abrir ciertos conjuntos de datos, sino además hacerlo en los formatos de más alto nivel, los formatos semánticos como RDF o OWL, para presentar no solo datos, sino ir un paso más allá y representar conocimiento. Este conocimiento además si es posible enlazarlo y enriquecerlo, aunque en este aspecto será fundamental que otras universidades también se suban al carro de publicar sus datos en abiertos, y tratando de usar los mismos formatos.

Otro de los objetivos de la apertura de datos es además servir de fuentes de datos con la que trabajar para crear aplicaciones o servicios de valor añadido para la sociedad en general, y para la comunidad universitaria en particular en este caso.

Partiendo de las recomendaciones del mismo documento referenciado sobre qué datos comenzar a abrir, en el apartado sobre la apertura de conjuntos de datos sobre infraestructuras y servicios universitarios, se mencionan entre otros el catálogo bibliográfico de la biblioteca de la universidad, y en estos datos es precisamente en los que nos queremos centrar, por el servicio que con ello podemos prestar a la universidad y a la difusión de sus recursos, en particular, pero también a las bibliotecas y a la cultura en general.

De forma más concreta y a la vista de las necesidades planteadas por la dirección de nuestra biblioteca, parte esencial del proyecto es dar a conocer la información sobre nuestras autoridades y figuras ilustres que han pasado y formado parte de nuestra historia, y cuyos datos se almacenan en nuestra biblioteca en los registros de autoridad de su catálogo.

Compartir datos permite la creación de una red global de conocimiento que desde el punto de vista de la biblioteca, propicia la construcción de una plataforma de red de ricos metadatos plenamente accesibles y utilizables por cualquiera. En el ámbito interno, los datos vinculados mejorarán los procesos y servicios, por ejemplo la catalogación

- **Propuestas de migración de las bibliotecas a LOD**

BIBFRAME, como se puede confirmar en el documento “Avances de BIBFRAME en 2014” (<http://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2015.64>), está consiguiendo grandes avances en los últimos dos años. y está de enorme actualidad en el mundo bibliotecario. Pero antes de centrarnos en este formato, haremos un breve repaso de algunos referentes en formatos bibliográficos para contextualizar el tema.

El formato MARC21 es un formato estándar muy estructurado que se viene utilizando en las bibliotecas de todo el mundo y también es el utilizado en la Biblioteca de la UPSA. Dada su estructura es capaz de describir con gran minuciosidad cualquier tipo de documento. Sin embargo este formato que ha descrito ya tantos millones de documentos y por tanto tan rico en datos, no puede ser utilizado en la web semántica y enlazar con otros datos abiertos. De ahí que desde hace varios años se está buscando migrar este formato a otros que permitan aprovechar la ingente cantidad de información que contienen los catálogos de bibliotecas en el mundo del Linked Open Data (LOD). Para ello instituciones pioneras como la Library of Congress y consorcios como OCLC (Consortio de bibliotecas que mantiene el Worldcat, el

mayor catálogo bibliográfico del mundo), que marcan las tendencias en todo el mundo en cuanto a bibliotecas, llevan trabajando en estos formatos desde hace tiempo.

Otros formatos que intentan acercarse a los datos vinculados y de hecho han influido en el diseño de BIBFRAME son por ejemplo:

- FRBR: (<https://www.loc.gov/catdir/cpsp/Que-es-FRBR.pdf>) es un esquema conceptual para la descripción bibliográfica. Su modelo se adapta con facilidad a las estructuras de datos vinculados, porque su estructura modular encaja con el modelo Linked data.
- MARC XML (<http://www.loc.gov/standards/marcxml/>) pretende ser la adaptación de MARC a las nuevas normas tecnológicas, pero no es interoperable para poder hablar de un modelo de intercambio común. Esto es una limitación para enlazar las bibliotecas con el resto de instituciones pertenecientes al ámbito de la información.
- El estándar ISBD (International Book Description) de la IFLA (<http://www.ifla.org/files/assets/hq/publications/series/44-es.pdf>), también se encuentra en pleno proceso de renovación para mejorar la portabilidad de datos en la Web semántica y propiciar la interoperabilidad de este estándar con otras normas, pero como se menciona, está aún por consolidar y definir.
- RDA (<https://www.oclc.org/es-americalatina/rda/about.html>), es un nuevo código de catalogación que ya utiliza IRIs y ofrece características que favorecen su integración con FRBR.
- Schema.org (<https://schema.org/>) es un vocabulario genérico desarrollado por las grandes compañías como Google, Microsoft o Yahoo, para potenciar la semántica de los contenidos y mejorar los contenidos de la Web. Al ser genérico también se puede utilizar para bibliotecas, aunque al mismo tiempo ya se está trabajando en una extensión del mismo, Bib Extended, más específico para contenidos bibliotecarios.
- ALIADA (<http://www.scanbit.net/es/aliada-open-source-tool-share-data-europes-libraries-museums/>) es una herramienta de publicación automática de datos e bibliotecas y museos bajo el paradigma Linked Data, que convierte los datos bibliográficos a RDF, pero sin seguir ningún estándar predefinido para el entorno bibliotecario, como ocurre con BIBFRAME.

BIBFRAME tiene influencias de RDA, FRBR, e incluso se están promoviendo cambios en Schema.org para su integración con BIBFRAME. De esta forma BIBFRAME es el formato que está liderando como formato de catalogación e intercambio de registros bibliográficos en el nuevo entorno LOD para bibliotecas.

- Qué es BIBFRAME?

BIBFRAME es un formato de catalogación basado en LOD especialmente diseñado para registros bibliográficos y archivísticos. BIBFRAME establece un modelo conceptual, con cierta relación con FRBR, pero simplificándolo. Pretende diferenciar los conceptos de las manifestaciones físicas (*Work* frente a *Instance*) Establece una relación bidireccional entre la descripción de contenidos de bibliotecas y los metadatos expuestos en la Web de datos. Está específicamente diseñado para datos vinculados, lo que supone, compartirlos, enriquecerlos y mejorar la recuperación de información.

BIBFRAME propone un modelo compatible con el resto de participantes en el mundo de la información, no solo las bibliotecas, también por ejemplo los museos y archivos, por lo que posibilita la interoperabilidad entre las instituciones del patrimonio cultural.

Finalmente, la experiencia y aplicación práctica de BIBFRAME en algunas universidades de prestigio, y en general la participación universitaria en la definición de BIBFRAME, también ha sido un factor positivo, al que nosotros también queremos seguir aportando con nuestro proyecto.

BIBFRAME es un proyecto reciente, propuesto por LOC en 2011, pero que ya está convirtiéndose en un referente y es apoyado por los agentes más importantes en el ámbito de las bibliotecas. Son muy significativas las bibliotecas a nivel mundial que están dando los primeros pasos con el nuevo formato de catalogación, como la Deutsche Bibliothek, que ya ha incorporado la representación de datos en BIBFRAME a su catálogo en línea, la National Library of Medicine que está elaborando el perfil de aplicación propio que incluya el uso de diferentes vocabularios como extensiones de BIBFRAME. Las bibliotecas universitarias están representadas entre los implementadores con proyectos como *Linked Data for Libraries (LD4L)*, en el que participan la Cornell University Library, el Harvard Library Innovation Lab y las Stanford University Libraries para crear un modelo para el desarrollo de un almacén de información semántica sobre recursos académicos. En España, por el momento, solo el CSUC (Consorti

de Serveis Universitaris de Catalunya), de la Universidad de Catalunya, la Fundación Ignacio Larramendy y DIGIBIS, están intentado abordar la iniciativa BIBFRAME. Por lo que sería interesante que nosotros nos uniéramos a este grupo tan privilegiado de instituciones pioneras. Esto avala a su vez, la relevancia del proyecto.

El modelo de datos de BIBFRAME se basa en las siguientes clases, también representadas en la Figura 2:

- *Creative Work*: trabajo creativo de un autor. Es el punto central de la descripción BIBFRAME.
- *Instance*: materialización de un trabajo creativo de cualquier índole, definido por sus propiedades y relaciones específicas del recurso.
- *Authority*: autoridades de la obra, que en BIBFRAME se consideran personas, lugares, temas del recurso,...
- *Annotation*: aporta información relevante y adicional sobre la obra o instancia de un recurso (revisiones, resúmenes, tablas de contenido, ...)

BIBFRAME

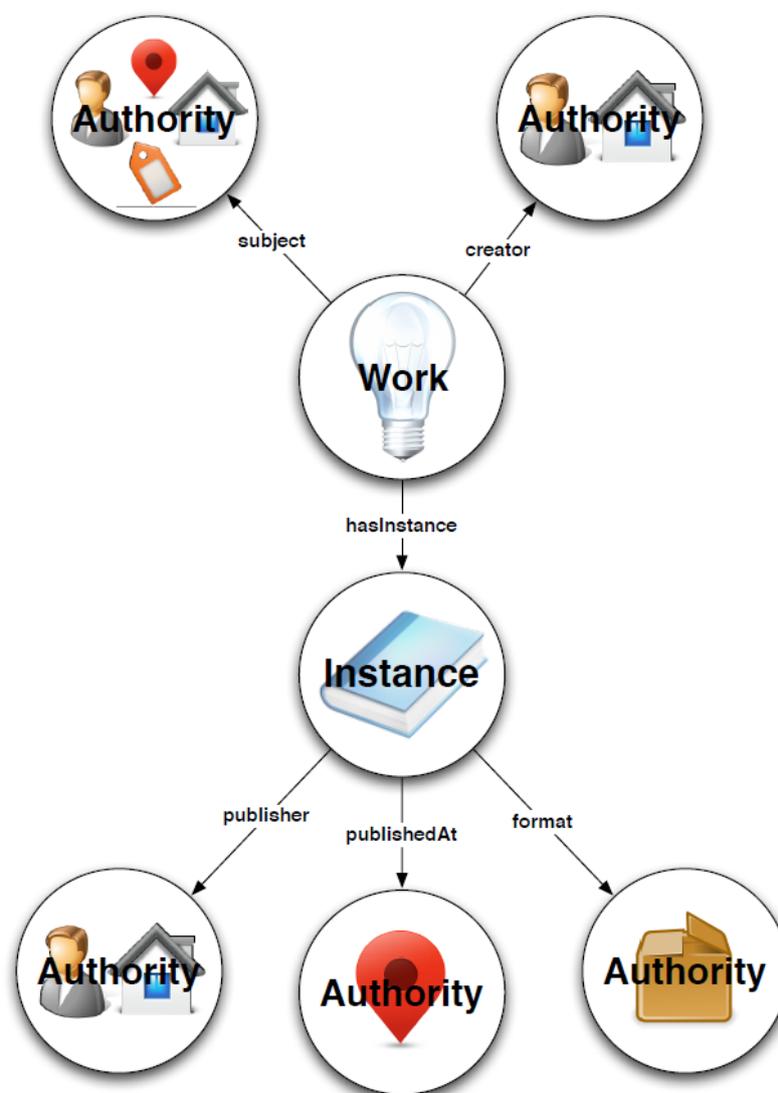


Figura 2. Representación gráfica de las clases del modelo BIBFRAME

4. Sistema de apertura y consulta de datos bibliográficos en formato semántico

El sistema desarrollado se articula en dos módulos principales. Por un lado el módulo que permite convertir cualquier archivo de entrada en formato MARCXML al formato bibliográfico de datos abiertos BIBFRAME. Por otro lado el segundo módulo será el que aprovechando los datos ya convertidos a BIBFRAME, permite enlazarlos con los de otras fuentes de datos abiertas, generalistas como DBPedia, y bibliográficas, como algunas bibliotecas nacionales tales como la española, francesa o alemana, y otros enlaces bibliográficos relacionados, aunque ya no en formatos abiertos, como VIAF o ISNI. Además el enlace con las dos primeras tipos de fuentes, DBpedia y las Bibliotecas Nacionales, se realizarán mediante tecnologías y consultas en formato semántico, demostrando así el poder de estas tecnologías. Este segundo módulo o servicio estará disponible como servicio de consulta público para cualquier usuario que desee consultar datos de las autoridades de la Biblioteca UPSA, en este caso.

La arquitectura del sistema desarrollado podría representarse en la siguiente representación gráfica. La Figura 3 representa estas dos partes: en amarillo, el conversor MARC21 a BIBFRAME nos servirá para generar un catálogo de documentos BIBFRAME, este se alimenta de archivos de obras y autores MARC21. Por otra parte, el Portal de Datos Abiertos Enlazados en azul, este último, consumirá tanto datos de nuestro repositorio BIBFRAME, como de las diferentes burbujas que encontramos en el gráfico con prefijo enlace, de las cuales, si coinciden como la Biblioteca Nacional Española o Wikipédia (DBpedia) con portales de datos abierto consumirá a través de consultas SPARQL sus datos.

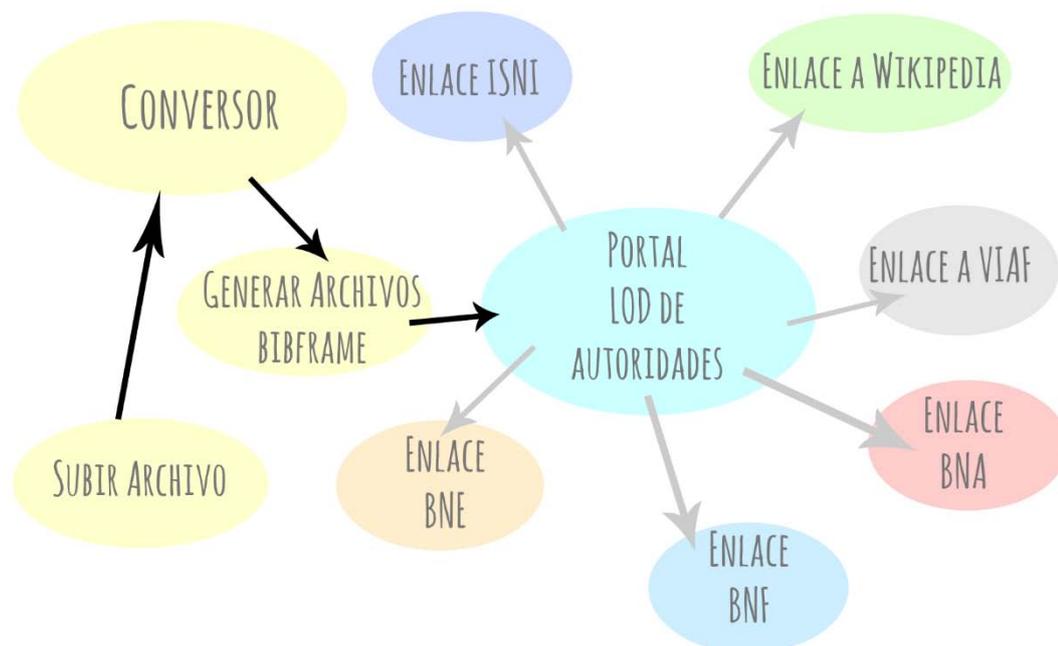


Figura 3. Arquitectura del Sistema

A continuación se describirán con algo más de detalle los dos módulos mencionados del sistema descrito.

4.1. Conversión de MARCXML a BIBFRAME

El formato MARC21 (<http://www.bne.es/es/Micrositios/Guias/Marc21/resources/Docs/Marc21.pdf>) define una serie de campos codificados que permiten describir un registro bibliográfico. La Tabla 1 podría servir de resumen de estos campos.

Por otro lado en el apartado de “Estado de la Cuestión” de este trabajo, también se han definido de forma esquemática los fundamentos de BIBFRAME. La idea de nuestro servicio es establecer la correspondencia entre campos de MARC21 y conceptos o etiquetas de BIBFRAME tales como Analizados ambos formatos y sus correspondencias se ha implementado un conversor que permite convertir un fichero MARCXML dado como entrada, a un fichero de salida en formato BIBFRAME, utilizando la interfaz que se muestra en la Figura 4.

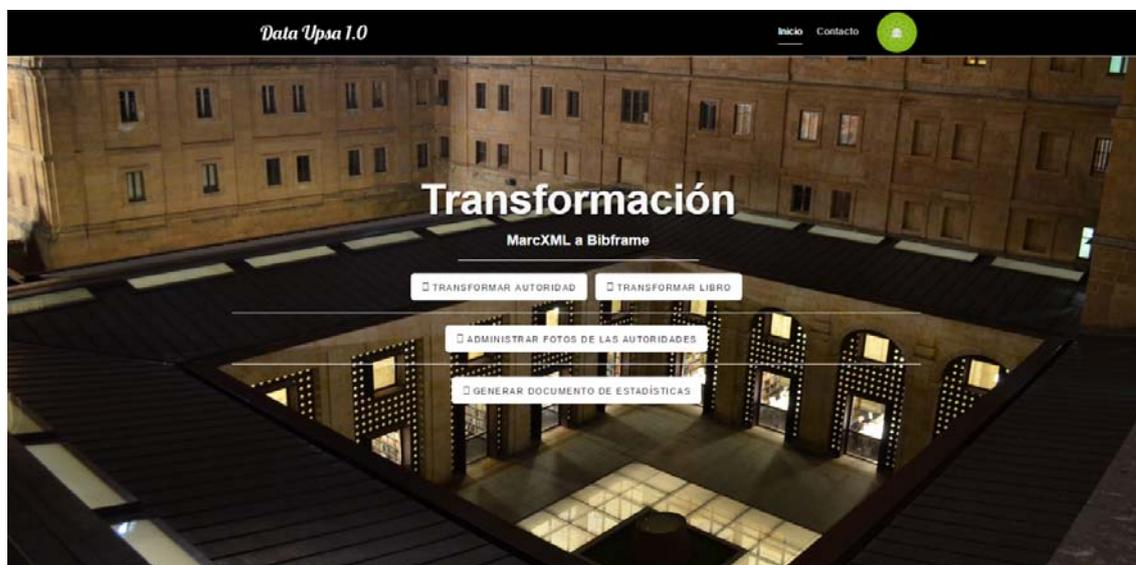


Figura 4. Interfaz conversor MARCXML-BIBFRAME

Los códigos finales que utilizamos del formato MARCXML están explicados en la siguiente Tabla 1.

Código en MARC21	TAG	Significado
"001"		Número de control
"003"		Identificador del número de control
"005"		Fecha y hora de la última transacción
"020"	A	Número Internacional Normalizado para Libros (ISBN)
"024"	2	Otros identificadores normalizados (VIAF/ISNI)
"040"	A	Centro catalogador de origen
"040"	B	Lengua de catalogación
"040"	D	Centro modificador
"040"	E	Normas de descripción
"100"	A y D	Punto de acceso principal-Nombre de persona
"100"	9	Código de catalogación del autor
"245"	A	Mención de título
"245"	B	Resto del título
"245"	C	Mención de responsabilidad, etc.
"260"	A	Lugar de publicación, distribución, etc.
"260"	B	Nombre del editor, distribuidor etc.
"260"	C	Fecha de publicación, distribución etc.
"300"	A	Extensión
"370"	A	Lugar asociado
"490"	A, P y V	Mención de serie

"504"	A	Nota de bibliografía, etc.
"505"	A	Nota de contenido con formato
"520"	A	Sumario, etc.
"600"	A	Nombre de persona
"600"	C	Títulos y otros términos asociados al nombre
"600"	9	Código de persona
"610"	9	Código de organización
"610"	A	Nombre de entidad corporativa
"650"	9	Código de materia o de nombre geográfico
"650"	A	Término de materia o nombre geográfico como elemento inicial
"650"	V	Subdivisión de forma
"700"	9, A, D, C y E	Punto de acceso adicional - Nombre de persona
"999"	C y D	Código de catalogación propia de la UPSA en Koha

Tabla 1. Descripción campos MARC21

Por otro lado, la siguiente Figura 5, muestra a través de un ejemplo concreto, la relación entre campos MARC21 y BIBFRAME al hacer la conversión


```

<?xml version="1.0" encoding="UTF-8"?>
<collection
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/
    schema/MARC21slim.xsd"
  xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>02306nam a220034971 4500</leader>
    <controlfield tag="001">285671</controlfield>
    <controlfield tag="003">esslup</controlfield>
    <controlfield tag="005">20160315164553.0</controlfield>
    <controlfield tag="008">140616s2014 sp ||||| |||| 00| 0 spa d</controlfield>
    <datafield tag="020" ind1=" " ind2=" " >
      <subfield code="a">978-84-7299-470-6</subfield>
    </datafield>
    <datafield tag="040" ind1=" " ind2=" " >
      <subfield code="a">esslup</subfield>
      <subfield code="b">spa</subfield>
      <subfield code="c">esslup</subfield>
      <subfield code="e">rda</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="3">
      <subfield code="a">La luz de la fe :</subfield>
      <subfield code="b">comentarios a la encíclica Lumen Fidei del Papa Francisco /</subfield>
      <subfield code="c">Profesores de la Facultad de Teología de la Universidad Pontificia de
        Salamanca</subfield>
    </datafield>
    <datafield tag="260" ind1=" " ind2=" " >
      <subfield code="a">[Salamanca] : </subfield>
      <subfield code="b">Universidad Pontificia de Salamanca, </subfield>
      <subfield code="c">2014</subfield>
    </datafield>
    <datafield tag="300" ind1=" " ind2=" " >
      <subfield code="a">223 páginas</subfield>
    </datafield>
    <datafield tag="336" ind1=" " ind2=" " >
      <subfield code="2">rdacontent</subfield>
      <subfield code="a">Texto (visual)</subfield>
      <subfield code="b">txt</subfield>
    </datafield>
    <datafield tag="337" ind1=" " ind2=" " >
      <subfield code="2">rda-media</subfield>
      <subfield code="a">sin mediación</subfield>
      <subfield code="b">n</subfield>
    </datafield>
  </record>
</collection>

```

```

<?xml version="1.0" encoding="UTF-8"?><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-
  rdf-syntax-ns#" xmlns:bf="http://bibframe.org/vocab/" xmlns:rdfs="http://www.w3.org/
  2000/01/rdf-schema#" xmlns:bf2="http://bibframe.org/vocab2/" xmlns:madsrdf="http://
  www.loc.gov/mads/rdf/v1#" xmlns:relators="http://id.loc.gov/vocabulary/relators/">
  <bf:Work rdf:about="C:\DataUpsa\resources\authorities\1189256\285671\285671work.rdf">
  <rdf:type rdf:resource="http://bibframe.org/vocab/Text"/>
  <bf:authorizedAccessPoint>Profesores de la Facultad de Teología de la Universidad
    Pontificia de Salamanca La luz de la fe : comentarios a la encíclica Lumen Fidei del
    Papa Francisco /</bf:authorizedAccessPoint>
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\358063\358063person.rdf"/>
  >
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\82564\82564person.rdf"/>
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\349736\349736person.rdf"/>
  >
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\245684\245684person.rdf"/>
  >
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\1189256\1189256person.rdf"/>
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\242767\242767person.rdf"/>
  >
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\82412\82412person.rdf"/>
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\737945\737945person.rdf"/>
  >
  <bf:contributor rdf:resource="C:\DataUpsa\resources\authorities\242851\242851person.rdf"/>
  >
  <bf:authorizedAccessPoint xml:lang="x-bf-
    hash">profesoresdelafacultaddeteologiadelauiversidadpontificiadesalamancalaluzdelafe
    comentariosalaenciclicalumenfideideipapafranciscopawork</bf:authorizedAccessPoint>
  <bf:workTitle rdf:resource="C:\DataUpsa\resources\authorities
    \1189256\285671\285671title.rdf"/>
  <bf:title xml:lang="x-bf-sort">La luz de la fe :</bf:title>
  <bf:subject rdf:resource="C:\DataUpsa\resources\authorities
    \1189256\285671\285671topic.rdf"/>
  <bf:subject rdf:resource="C:\DataUpsa\resources\authorities
    \1189256\285671\285671organization.rdf"/>
  <bf:series rdf:resource="C:\DataUpsa\resources\authorities
    \1189256\285671\285671series.rdf"/>
  <bf:language rdf:resource="http://id.loc.gov/vocabulary/languages/spa"/>
  <bf:contentCategory rdf:resource="C:\DataUpsa\resources\authorities
    \1189256\285671\285671category1.rdf"/>
  </bf:Work>
  <bf:Work rdf:about="C:\DataUpsa\resources\authorities\1189256\285671\285671series.rdf">
  <bf:title>ser::Ensayos de teología ; 7</bf:title>
  <bf:authorizedAccessPoint>Ensayos de teología ; 7</bf:authorizedAccessPoint>
  </bf:Work>

```

Figura 5. Ejemplo Correspondencia MARC21 BIBFRAME

Tras la transformación de los archivos se procede a su almacenamiento en un repositorio de datos, con una estructura y nomenclatura de archivos prefijada para que luego estos datos ya en formato BIBFRME puedan reutilizarse.

4.2. Enriquecimiento de datos bibliográficos

El segundo servicio ofrecido por nuestro sistema, como se ha comentado, es un portal de abiertos cuyos datos resultan de la conversión del servicio descrito en el apartado anterior.

La intención de este portal es mostrar la información contenida en los archivos de nuestro repositorio y además enlazar esta información con otras fuentes externas. Así se consigue enriquecerla mediante consultas a otras fuentes de datos abiertos mediante tecnologías o consultas semánticas.

Ya que el punto principal de nuestro repositorio son las autoridades, el portal se centra en ellas. De forma más concreta, en la página principal, Figura 6, aparece un listado de las autoridades almacenadas en el repositorio con su fotografía correspondiente siempre que exista, en caso de no tener fotografía se indica mediante una imagen que no hay una imagen para ese autor en el sistema. En la página principal, Figura 7, podemos observar los diferentes autores y sus fotografías asociadas que hemos obtenido de las funcionalidades del conversor.



Figura 6. Inicio del Portal de datos abiertos

Una vez dentro de la página de un autor determinado se mostrarán varios enlaces:

- Información sobre el autor: en este apartado mostraremos información personal de la autoridad, como su fecha de nacimiento, curriculum o lugares importantes a lo largo de su vida (lugar de nacimiento, de residencia...). En este punto en el caso de que el autor tenga información en la página de datos abiertos DBpedia se mostrara esa información, también se mostrara, en caso de tenerlos, enlaces a la Biblioteca Nacional Española, la Biblioteca Nacional Alemana y la Biblioteca Nacional Francesa. Figura 8
- Obras: a través de este enlace podemos acceder a un listado con las obras del autor. En ese listado podemos ver el título de la obra, así como su resumen en caso de tenerlo. Cada obra tendrá asociado un botón para acceder a más información de la misma, donde veremos su lugar de publicación, extensión, editorial...
- Influenciado por: este enlace mostrara una página en la que aparecerán aquellos autores que han influenciado al autor seleccionado, sobre estos autores podemos ver su nombre, un abstract y un enlace a su página de Wikipedia.
- Otros enlaces externos del autor: este apartado está compuesto por tres enlaces, un enlace a la Biblioteca Nacional Española, un enlace a VIAF y otro a ISNI. Estos enlaces solo aparecerán si

existen en nuestros archivos o en los enlaces a las fuentes externas los datos correspondientes a estos campos.

Para la realización del portal hemos utilizado, como se ha comentado, tanto datos propios como datos pertenecientes a otras fuentes externas.

Para la extracción de los datos de nuestros archivos BIBFRAME hemos utilizado las ya mencionadas consultas SPARQL, mediante estas consultas hemos extraído los datos para completar las páginas de nuestros autores y obras.

Para enlazar con puntos de datos abiertos se utilizan las consultas SPARQL, las consultas se realizan sobre unos servicios denominados SPARQL Points que permiten acceder al conocimiento almacenado en los archivos RDF mediante las consultas SPARQL.

Hemos enlazado con dos puntos de datos abiertos que son DBPedia y la Biblioteca Nacional Española gracias a la información de nuestros archivos. Para poder realizar las consultas a puntos de datos abiertos externos necesitamos información sobre el objeto de la consulta, en el caso de los autores su nombre o su código de identificación en algún registro internacional (VIAF, ISNI) y en el caso de los libros su título o su ISBN.

En cuanto a la información sobre las autoridades hemos utilizado nuestros datos en la información del autor y en los enlaces a VIAF e ISNI.

En la información de las obras hemos utilizado las consultas a nuestros archivos para obtener toda la información relativa a estas.

A modo de ejemplo, si entramos al apartado de Olegario González de Cardedal, nos encontraremos con una tabla con subapartados: Información sobre el autor; Obras; Influenciado por; y enlaces externos en caso de que los tuviese dicho autor (Figura 8).



Figura 7. Página principal de autor

En la página de información sobre el autor, tendremos su curriculum y sus lugares asociados de nacimiento y vida, por otra parte, nos encontraremos con los datos que obtenemos de DBPedia del autor, y además obtendremos de la Biblioteca Nacional Española los enlaces a la Francesa y a la Alemana.

Data Upsa 1.0 Inicio Contacto

Olegario González de Cardedal

Curriculum:
Catedrático 2004

Lugares Importantes(nacimiento, vida):
Cardedal, Ávila

Más datos en dbpedia

Abstract:
Olegario González de Cardedal (Cardedal, La Lastra del Cano, (Ávila), 2 de octubre de 1934) es un sacerdote y teólogo católico español.@es

Lugar de nacimiento:
La Lastra del Cano
Provincia de Ávila
Cardedal

Fecha de nacimiento:
1934

Nacionalidad:
Española

Materias:
Pensador
Real Academia de Ciencias Morales y Políticas
Escritor
Teología
Intelectual
Teólogo

Enlaces externos:
Wikipedia Biblioteca Nacional Española Biblioteca Nacional Francesa Biblioteca Nacional Alemana

Figura 8. Página Información de autor

En el listado de obras veremos todas las obras que tenemos en nuestro repositorio de datos BIBFRAME de ese mismo autor, con sus respectivos enlaces para obtener más información de la obra.

5. CONCLUSIONES

Nuestro sistema de conversión acceso y enriquecimiento de datos bibliográficos mediante tecnologías semánticas da respuesta a los objetivos planteados en el trabajo.

- La transformación de los datos de formato MARC a BIBFRAME manteniendo toda la información posible y publicándolos en formatos abiertos.
Se han transformado todos los datos entre los formatos MARCXML y BIBFRAME. Además se ha completado la información de los BIBFRAME con información que la Library of Congress no añade en sus BIBFRAME, como pueden ser los campos VIAF, ISNI, curriculum, etc.
- Visibilidad y consulta de datos:
Se permite consultar a las autoridades de nuestra universidad y visualizar no solo los datos que de ellos ya disponemos, sino además enriquecidos con información complementaria acerca de los mismos. La mayoría de nuestros datos de autores tienen relevancia en la Biblioteca Nacional Española y en repositorios como ISNI y VIAF, además sobre algunos de ellos también se encuentra información en DBPedia.

En definitiva el proyecto permite demostrar la utilidad de los nuevos formatos de datos bibliográficos para datos abiertos, en particular BIBFRAME, y sobre todo mostrarlo a través de un caso y aplicación real que demuestra esta utilidad y define sus posibilidades como herramienta de futuro, especialmente cuando se amplíe la web de datos. Se trata además de un caso o sistema real que ya se encuentra actualmente en explotación en la Biblioteca de la Universidad Pontificia de Salamanca.

5.1. Líneas futuras

Las líneas futuras y ampliación de nuestro trabajo se consideran en dos ámbitos.

- Ampliar y dar difusión al proyecto para su reutilización y explotación en otras bibliotecas
- Buscar nuevas fuentes de datos abiertas con las que enlazar para enriquecer los datos de nuestra biblioteca y/o cualquier otra que se una a nuestro proyecto.

6. BIBLIOGRAFÍA

Agenjo, X., & Hernández-Carrascal, F. (2015, Avances de BIBFRAME en 2014. *Anuario ThinkEPI*, 9 doi:<http://dx.doi.org/10.3145/thinkepi.2015.64>

Biblioteca Nacional (España). Formato Marc21 para registros bibliográficos. Consultado el 15 de mayo de 2016 en <http://www.bne.es/es/Micrositios/Guias/Marc21/resources/Docs/Marc21.pdf>

Hausenblas, M.5 * datos abiertos. Consultado el 15 de mayo de 2016 en <http://5stardata.info/es/>

IFLA. (2011). ISBD. Descripción bibliográfica internacional normalizada. Consultado el 15 de mayo de 2016 en <http://www.ifla.org/files/assets/hq/publications/series/44-es.pdf>

OCLC. Acerca de RDA. Consultado el 15 de mayo de 2016 en <https://www.oclc.org/es-americalatina/rda/about.html>

Scanbit. Aliada, una herramienta Open Source para publicar datos enlazados en abierto en bibliotecas y museos. Consultado el 15 de mayo de 2016 en <http://www.scanbit.net/es/aliada-open-source-tool-share-data-europes-libraries-museums/>

The Library of Congress. Marcxml Schema Official Site. Consultado el 15 de mayo de 2016 en <http://www.loc.gov/standards/marcxml/>

The Library of Congress. Marcxml Schema. Official Site. Consultado el 15 de mayo de 2016 en <http://www.loc.gov/standards/marcxml/>

Tillet, B. (2004). ¿Qué es FRBR?. Un modelo conceptual del universo bibliográfico. Consultado el 15 de mayo de 2016 en <https://www.loc.gov/catdir/cpsa/Que-es-FRBR.pdf>

-