

# Agregación de datos bibliográficos por medio de servicios *Sparql*

## Aggregation of library data through *Sparql*

**Xavier Agenjo-Bullón**

<http://orcid.org/0000-0001-8338-8087>

Fundación Ignacio Larramendi

[xavier.agenjo@larramendi.es](mailto:xavier.agenjo@larramendi.es)

**Francisca Hernández-Carrascal**

<http://orcid.org/0000-0002-2389-0945>

Digibis

[francisca.hernandez@digibis.com](mailto:francisca.hernandez@digibis.com)

**Agenjo-Bullón, Xavier; Hernández-Carrascal, Francisca** (2019). "Agregación de datos bibliográficos por medio de servicios *Sparql*". *Anuario ThinkEPI*, v. 13, e13f02.

<https://doi.org/10.3145/thinkepi.2019.e13f02>

Publicado en *IweTel* el 5 de febrero de 2019



**Resumen:** El propósito de esta nota es describir el proceso de agregación de metadatos realizado por la *Fundación Ignacio Larramendi* por medio del servicio *Sparql* de la *Biblioteca Virtual Miguel de Cervantes*. Se describen los resultados obtenidos y se pone de manifiesto la utilidad del lenguaje *Sparql* y los servicios *Sparql* como herramienta para acceder a los datos en repositorios con grandes volúmenes de información, así como algunos de los impedimentos para su uso.

**Palabras clave:** Agregación de metadatos bibliográficos; *Sparql* 1.1; Servidores y servicios *Sparql*.

**Abstract:** This note describes the metadata aggregation process carried out by the *Ignacio Larramendi Foundation* through the *Sparql Endpoint* of the *Miguel de Cervantes Virtual Library*. The results obtained are summarized and it is showed how *Sparql* language and *Sparql Endpoint* are a useful tool to access data in large repositories. Some difficulties for its use are described.

**Keywords:** Bibliographic metadata aggregation; *Sparql* 1.1; *Sparql Endpoint*.

## 1. Introducción

La tecnología *linked open data* aplicada a las bibliotecas, archivos y museos está evolucionando hacia la productividad como describe Pace en su entrada del blog *Next* de la *OCLC* (Pace, 2018). Este autor expone su convicción de que las bibliotecas necesitan plataformas *linked open data* para embarcarse en innovaciones que no se pueden afrontar con MARC y describe algunos de los proyectos *linked open data* productivos y en producción. Para los autores de esta nota es muy importante justamente comprobar en la realidad la viabilidad productiva de algunas de las promesas de *linked open data*.

El propósito de este artículo es hacer hincapié en la utilización real de una de las recomendaciones (Hyland; Atezing; Villazón-Terrazas, 2014) que forman parte de *linked open data* y de la web semántica como es el lenguaje de búsqueda *Sparql* (*Sparql Protocol and RDF Query Language*) para acceder a la información bibliográfica. En estas mismas páginas hemos dedicado nuestra atención al enriquecimiento semántico de los registros de autoridad de la *Biblioteca Virtual de Polígrafos* por medio de vocabularios de valores como *VIAF*, *datos.bne.es*, *DBpedia*, *Wikidata*, etc. (Agenjo-Bullón; Hernán-

dez-Carrascal, 2018a). Ahora queremos describir los pros y los contras del acceso de la información bibliográfica a través de puntos de servicio *Sparql*.

Esta nota tiene su origen en un proyecto de la *Fundación Ignacio Larramendi* para agregar descripciones bibliográficas de documentos digitalizados a la *Biblioteca Virtual* de la *Escuela de Salamanca* por medio del servicio *Sparql* de la *Biblioteca Virtual Miguel de Cervantes* (en adelante *BVMC*), disponible en:

<http://data.cervantesvirtual.com/sparql>

## 2. La *Biblioteca Virtual de la Escuela de Salamanca* 3.0

La *Biblioteca Virtual de la Escuela de Salamanca* está compuesta por 20 autores<sup>1</sup> y 1.174 objetos digitales (aunque la cifra puede ir variando pues se trata de una biblioteca viva) que se corresponden con las versiones digitales de diferentes ediciones de las obras de los autores que componen esta escuela y que se han ido agregando por medio de distintas vías, como se describe en la “Introducción” y en la “Nota a esta edición” de la *Biblioteca Virtual de la Escuela de Salamanca*<sup>2</sup>.

Siguiendo la metodología de la *Fundación Ignacio Larramendi* (Agenjo-Bullón; Hernández-Carrascal; Juez-García, 2013) esos 1.174 objetos digitales se han reunido a partir de digitalizaciones propias que lleva a cabo *Digibís* (8 obras, un 0,68%) y de un 99,32% de obras digitalizadas por terceros. Por ejemplo, en relación con Francisco de Vitoria se pueden consultar 4 digitalizaciones realizadas por la propia *Fundación* y 32 proporcionadas por diferentes instituciones (tabla 1).

Todas las obras se recatalogan en MARC 21/RDA para acomodarse a los criterios establecidos para la *Biblioteca Virtual de Polígrafos* que, resumiendo, consisten en describir el objeto digital y en prestar la mayor atención al control de autoridades que, como se ha dicho antes, están enriquecidas<sup>3</sup> con diferentes vocabularios de valores y bases de conocimiento (*VIAF*, *datos.bne.es*, *ULAN*, *Dbpedia*, *Wikidata*, etc).

Los procedimientos de agregación se han efectuado por distintos mecanismos, siendo el principal de ellos la herramienta de recolección automática disponible en *Digibib* que permite buscar por distintos criterios, recuperar e importar los metadatos de *Europeana* (a través de su API de búsqueda<sup>4</sup>) y de *Hispana* (a través de su servidor SRU<sup>5</sup>). *Digibib* se encarga de contrastar que no haya registros duplicados por medio de los elementos *oai:identifier*, entre otros medios.

Mediante estos dos servicios se ha tenido acceso a los registros de la gran mayoría de los proyectos de digitalización que se están llevando a cabo en Europa, aunque todavía subsisten proyectos de digitalización sin normalizar y por ello no presentes en *Hispana* o en *Europeana* y opacos en lo que se refiere al acceso a los datos, como se puede comprobar en el *Directorio de Colecciones* de la propia *Hispana*<sup>6</sup>.

En efecto, de este mecanismo de agregación masivo se habían quedado fuera importantes fuentes de información bibliográfica de obras digitalizadas como es la *BVMC*, y, lo que es también muy importante para la *Biblioteca Virtual de la Escuela de Salamanca*, los proyectos de digitalización que se están realizando en Iberoamérica —y con ello la posibilidad de recopilar las obras conservadas en bibliotecas hispanoamericanas.

Estamos pendientes de los avances que se producen en este sentido porque tendrán su repercusión en los estudios sobre la recepción de los textos de la *Escuela de Salamanca* en América. Citaremos entre ellos:

- *Biblioteca Digital* de la *Universidad Nacional Autónoma de México*;  
<http://bidi.unam.mx>
- *Biblioteca Digital Mexicana* que, al menos hace un par de años, estaba inmersa en procurar su participación en la *Digital Public Library of America*;  
<http://bdmx.mx>
- *LA Referencia* (*Red de repositorios de acceso abierto en la ciencia*) que recoge las publicaciones científicas de América Latina en acceso abierto<sup>7</sup>.  
<http://www.lareferencia.info/es>

---

**“Estamos inmersos en la transición de *linked open data* hacia la productividad y hemos de convivir con múltiples experiencias y estructuras de datos”**

---

Tabla 1. Digitalizaciones relacionadas con Francisco de Vitoria

<i>Universidad de Granada</i>	9
<i>Bayerische Staatsbibliothek</i>	8
<i>Biblioteca de la Universidad de Salamanca</i>	4
<i>Fundación Ignacio Larramendi/Digibís</i>	4
<i>Fundación Sancho el Sabio</i>	4
<i>Akademie der Wissenschaft und der Literatur</i>	3
<i>Biblioteca Valenciana Digital</i>	1
<i>Biblioteca Virtual de Aragón</i>	1
<i>Biblioteca Virtual Miguel de Cervantes</i>	1
<i>Revistas Uniandes (Universidad de los Andes)</i>	1

Estos recursos, de momento, sólo ofrecen la exportación en formatos de cita bibliográfica. Desde *LA Referencia* se ha anunciado<sup>8</sup> que durante 2019 se irán implementando las directrices de *OpenAIRE*<sup>9</sup>, lo que permitirá la integración de los repositorios nacionales en la infraestructura *OpenAIRE*. No queda claro si será posible acceder a la totalidad de los recursos por esta vía o bien habrá que acceder de forma individualizada a los recursos de cada nodo nacional.

### **3. *Data.cervantesvirtual.com*: acceso a los datos a través de *Sparql Endpoint***

Para acrecentar nuestro acervo de objetos digitales, como dice la *Wikipedia*, nuestra atención se fijó en la exposición de los datos bibliográficos de la *BVMC*. En 2015 se abrió públicamente el sitio *data.cervantesvirtual.es*<sup>10</sup> con la transformación en datos abiertos vinculados de los registros que componen la *BVMC*.

Este nuevo portal de datos tiene, aparte de otras características que trataremos más adelante, la ventaja de que por primera vez los datos de esta *Biblioteca* están accesibles para su reutilización masiva. En efecto la *BVMC* no disponía hasta *data.cervantesvirtual.com* de esta posibilidad, ya fuera en la forma de los tradicionales registros MARC 21 o a través de algún protocolo como OAI-PMH.

La *BVMC* está presente en *Europeana* desde el inicio de esta biblioteca digital. En la actualidad ofrece 70.971 registros, volumen que no ha cambiado desde 2014. Esto suponía que la agregación de datos de la *BVMC* a través de la API de *Europeana* ofrecía una foto fija de hace casi 5 años. *Hispana* tampoco podía servir de vía de acceso porque la *BVMC* no dispone de repositorio OAI-PMH y sólo aparece en el *Directorio de Colecciones Digitales*.

<http://hispana.mcu.es/es/comunidades/registro.do?id=61>

La ausencia de repositorio OAI-PMH, o de acceso a los datos por medio de APIs, SRU u otros, impedía que esta gran biblioteca formara parte de las fuentes de agregación de la *Biblioteca Virtual de Polígrafos*.

La publicación de *data.cervantesvirtual.com* interesó a la *Fundación Ignacio Larramendi*, como lo hubiera hecho cualquier otra actuación similar, por un doble motivo:

- se podía acceder libremente a los datos de esta biblioteca;
- permitía explorar el acceso a datos abiertos vinculados a través de un servicio *Sparql* con un caso práctico, integrado en las tareas habituales de la *Fundación Ignacio Larramendi*, es decir, en producción.

Desde la *Fundación Ignacio Larramendi* y *Digibís* teníamos especial interés en llevar a cabo este análisis puesto que cada vez son más las instituciones que disponen de esta utilidad como se ha ido reflejando en las diferentes encuestas sobre la implantación de *linked open data* llevadas a cabo por *OCLC Research* (**Smith-Yoshimura**, 2018).

Hay que decir que el uso de los datos abiertos vinculados a través de un servicio *Sparql* no requiere ningún tipo de relación formal puesto que en sí se trata de una tecnología para la consulta de datos abiertos —lo que, por cierto, es lo mismo que cuando se aplica a los repositorios OAI (la O es de *open*, claro).

La experiencia ha resultado muy positiva ya que ha permitido obtener 250 registros nuevos, lo que ha supuesto aumentar el número de ejemplares digitales en un 21,30%. Después de esta prueba de descarga de registros para los 20 autores de la *Biblioteca Virtual de la Escuela de Salamanca*, la *Fundación Ignacio Larramendi* y *Digibís* están utilizando la misma estrategia<sup>11</sup> para los 1.000 autores que componen la totalidad del proyecto *Biblioteca Virtual de Polígrafos* y, en concreto, para la *Biblioteca Virtual de Novatores* que se presentará en los primeros meses de 2019.

Sin embargo, podemos extraer algunas conclusiones que en su mayoría no tienen que ver, en principio, con el lenguaje de búsqueda *Sparql*. Efectivamente, para componer una búsqueda en *Sparql* se requieren fundamentalmente dos cosas:

---

**“El uso de los datos abiertos vinculados a través de un servicio *Sparql* no requiere ningún tipo de relación formal puesto que en sí se trata de una tecnología para la consulta de datos abiertos”**

---

---

**“Para componer una búsqueda en *Sparql* se requiere tener una cierta familiaridad con la construcción de búsquedas según este lenguaje y conocer al detalle el modelo de datos que utiliza un determinado proveedor”**

---

- tener una cierta familiaridad con la construcción de búsquedas según este lenguaje<sup>12</sup>;
- conocer al detalle el modelo de datos que utiliza un determinado proveedor para saber cómo se debe interrogar.

No es el objetivo de esta nota mostrar la forma en que deben construirse las preguntas *Sparql* para lo que hay multitud de tutoriales en la Web (la propia *BVMC* ofrece uno<sup>13</sup>). Pero sí es importante señalar que para las descripciones bibliográficas la *BVMC* ha seguido fundamentalmente la ontología de RDA (*RDA Registry*)<sup>14</sup> y su modelo de referencia es FRBR<sup>15</sup>.

De las clases RDA hemos reutilizado 'obras' y 'manifestaciones', porque contenían todas las propiedades necesarias para la *Fundación Ignacio Larramendi* y eliminaban la redundancia de las propiedades de las 'expresiones', básicamente consistentes en el título e idioma de una obra, ya contenidos en la 'manifestación'.

<https://bit.ly/2TfnK3v>

No hemos conseguido acceder a ningún ítem.

Como decíamos al principio, estamos inmersos en la transición de *linked open data* hacia la productividad y hemos de convivir con múltiples estructuras de datos, como es el formato MARC 21, y con distintas experiencias<sup>16</sup> de *frbrización* de los registros como las de la *BVMC* (o la *Biblioteca Nacional de España* en *datos.bne.es*), así como con otros modelos de datos no basados en FRBR como es *Europeana Data Model*<sup>17</sup>, que reutiliza los espacios de nombres OAI, entre otros:

- *Object Reuse and Exchange (ORE)*;  
<http://www.openarchives.org/ore/terms>
- *Simple Knowledge Organization System (SKOS)*;  
<http://www.w3.org/2004/02/skos/core>
- *Dublin Core*.  
<http://purl.org/dc/elements/1.1>  
<http://purl.org/dc/terms>

La *Biblioteca Virtual de Polígrafos*, como otras instancias de *Digibib*, utiliza el formato MARC 21 para la importación masiva de datos y para su transformación dinámica a RDF conforme a la ontología *Europeana Data Model*.

Quiere esto decir que la *BVMC* ha transformado sus registros desde una estructura MARC 21 a la ontología de RDA y que la *Fundación Ignacio Larramendi* ha tenido que invertir esa transformación para obtener tras un cierto proceso de manipulación de datos registros MARC 21. Al final, después del proceso de tratamiento de las descripciones de *data.cervantesvirtual.com* se tiene un fichero de registros MARC 21 en *ISO 2709* que la mayoría de los sistemas tratan sin dificultad, y en el caso que nos ocupa, *Digibib* lo hace llevando a cabo además unas validaciones precisas.

Por los motivos señalados hasta la aplicación de *linked open data* en la *BVMC* no se podía disponer de esos registros en ningún formato. La apertura de esos datos se ha producido por medio de *RDA Registry* en el caso de la *BVMC*, mientras otras bibliotecas han adoptado otros modelos. En todos los casos es necesario conocer al detalle el modelo utilizado, si bien suele ser familiar porque de uno u otro modo todos tienen algún parentesco cercano con MARC 21 y se enmarcan en los vocabularios del mundo bibliotecario. Esto es lógico, obviamente, por dos cosas:

- la estructura de origen impone y va a imponer unas condiciones específicas a las transformaciones que se realicen;
- el modelo de datos de destino va a ejercer también sus imposiciones.

Hemos iniciado el camino de superación del formato MARC 21, pero aún no hemos llegado a un destino que garantice una normalización bibliográfica como la que ha cumplido y está cumpliendo el formato MARC 21, fin con el que se está desarrollando *Bibframe*. Con toda seguridad *Bibframe* no será la única estructura de datos bibliográficos, aunque sí lo sea para la catalogación, y convivirá con *Schema.org* y *Europeana Data Model*, así como con otras estructuras de datos de otros entornos como *Cerif* o *VIVO*.

La reutilización de *linked open data* va a requerir de algunas transformaciones, como en el caso de la

---

**“Hemos iniciado el camino de superación del formato MARC 21, pero aún no hemos llegado a un destino que garantice una normalización bibliográfica. Con toda seguridad *Bibframe* no será la única estructura de datos bibliográficos”**

---

integración de datos en un particular sistema a través de cualquier API de acceso a los datos. De hecho, solo en contadas ocasiones ha sido posible integrar registros bibliográficos MARC 21 de forma masiva sin hacer algún tratamiento previo. Prueba de ello es la misma existencia del magnífico programa *MarcEdit* que facilita la vida a muchos bibliotecarios

a la hora de tratar registros MARC 21 para su integración en un sistema. Desde luego las aplicaciones basadas en MARC 21, pero cerradas a su acceso son ya una vía muerta desde hace algunos años.

Volviendo a nuestra prueba, el resultado de la búsqueda en el servicio *Sparql* de *data.cervantesvirtual.com* fue la obtención de 396 obras y manifestaciones. Quizá no esté tan presente que la *BVMC* está constituida, al igual que la *Biblioteca Virtual de Polígrafos* —salvando las distancias— no sólo por los objetos digitales propios sino por la contribución de otras instituciones como universidades españolas y sus proyectos de digitalización.

Sin embargo, fue necesario eliminar un 37% de los registros obtenidos por distintos motivos, principalmente porque las reproducciones en las instituciones de origen eran inaccesibles, bien porque los enlaces estuvieran rotos o bien porque implicaban requisitos no soportados por las versiones actuales de los navegadores. Otro de los motivos de eliminación fue la falta de integridad de las reproducciones en las que se accedía únicamente a la digitalización de las portadas y los preliminares.

A pesar de estos pequeños inconvenientes, el resultado del proceso de agregación no puede ser más alentador. La tabla 2 ofrece un resumen de la distribución de la procedencia de los objetos digitales en el que hemos agrupado a las instituciones para que proporcionen una visión general.

También hemos agregado objetos digitales del proyecto *La Escuela de Salamanca: una colección digital de fuentes y un diccionario de su lenguaje jurídico-político* que llevan a cabo la *Akademie der Wissenschaften und der Literatur*, la *Goethe-Universität*, *Institut für Philosophie* y el *Max-Planck-Institut für europäische Rechtsgeschichte*. <https://www.salamanca.school/es/project.html>

Esta agregación se presentó en detalle en la comunicación realizada por los autores de esta nota, “Visibilidad y accesibilidad digital de la Escuela de Salamanca en un entorno de linked open data” (Agenjo-Bullón; Hernández-Carrascal, 2018b), a la conferencia internacional *The School of Salamanca: A Case of Global Knowledge Production?* (Fundación Ignacio Larramendi, 2018) que se celebró en Buenos Aires los días 24 a 26 de octubre de 2018.

#### 4. Búsquedas federadas

Esta estrategia de incrementar las obras de un conjunto de autores seleccionados se va a seguir a partir del servidor *Sparql* de la *Biblioteca Nacional* y de las APIs de búsqueda de la *Biblioteca Digital Hispánica* y de la *Biblioteca Digital del Patrimonio Iberoamericano* (BNE, 2018). <http://datos.bne.es/sparql>

Desde luego el punto más a favor es que se puede obtener el enlace a los objetos digitales mediante sentencias SPARQL puesto que *datos.bne.es* ya dispone de esta información.

Lógicamente, de forma análoga a la búsqueda en el servidor *Sparql* de *datos.cervantesvirtual.com*, se pueden lanzar preguntas similares en los servicios *Sparql* de *datos.bne.es* (*Biblioteca Nacional de España*), de *Europeana* u otros. <http://sparql.europeana.eu>

Bastaría con repasar los catálogos *datos.gob.es*, *Datahub*, o *Google Dataset Search*, entre otros, para obtener información de otros servicios de acceso a los datos.

---

**“Estos procedimientos *linked open data* no dejan de recordar a la labor bibliográfica clásica”**

---

Tabla 2. Procedencia de los objetos digitales

Universidades españolas, especialmente <i>Granada</i> , <i>Salamanca</i> y <i>La Rioja</i>	40,62%
<i>Bayersische Staatsbibliothek</i>	19,49%
<i>Bibliotecas de Comunidades Autónomas</i>	11,42%
<i>Fundación Ignacio Larramendi</i>	7,98%
<i>Biblioteca Virtual Miguel de Cervantes</i>	5,44%
<i>Biblioteca Digital Hispánica</i>	4,44%
<i>Biblioteca Virtual de Patrimonio Bibliográfico</i>	3,90%
Distintas instituciones europeas ( <i>Gallica</i> , <i>Goethe Universität</i> , etc.)	3,5%
Otras instituciones españolas	1,84%
Instituciones portuguesas	1,10%
Instituciones americanas	0,27%

<https://datos.gob.es/es/catalogo>  
<https://datahub.io/>  
<https://toolbox.google.com/datasetsearch>

Estos procedimientos *linked open data* no dejan de recordar a la labor bibliográfica clásica, puesto que el bibliógrafo selecciona las fuentes de información, las bibliografías, los índices, etc., que sirven a su propósito y a partir de ellos anotar las referencias bibliográficas de su interés.

Y dado que *Sparql* 1.1 permite construir búsquedas federadas el próximo paso a dar es lógicamente explorar los resultados que se pueden obtener consultando desde un cliente *Sparql* diferentes servidores en una misma búsqueda (Coombs, 2016). Esta posibilidad requiere aún de la complejidad adicional de seleccionar, descargar e instalar el cliente<sup>18</sup>, lo cual no es ni inmediato, ni fácil de llevar a cabo si no se cuenta con un cierto apoyo informático. Este es uno de los motivos por los que se está extendiendo el uso de *Linked Data Fragments* que facilitará la búsqueda federada en distintos conjuntos de datos abiertos vinculados a través de un servicio web como se puede probar en *Comunica* (Taelman et al., 2018). <http://linkeddatafragments.org>

De forma paralela, *Wikidata* ha creado un utilísimo servicio de consulta *Sparql*<sup>19</sup> que permite interrogar no sólo los datos de la propia *Wikidata* sino también realizar búsquedas federadas sobre un conjunto delimitado de servidores *Sparql*<sup>20</sup>, entre ellos los de la *Biblioteca Nacional de España*, la *BVMC*, la *British Library* o *Europeana*.

The screenshot shows the Wikidata Query Service interface. At the top, there are navigation links for 'Examples', 'Help', and 'More tools'. The main area is divided into a 'Query Helper' sidebar on the left and a query editor on the right. The query editor contains a federated SPARQL query that searches for authors named 'Vitoria, Francisco de' across different data sources. Below the query editor, there is a table of results with columns for 'autor', 'nombreautor', 'obra', and 'titulo'. The table shows four results, each with a unique identifier, the author's name, the work title, and a brief description of the work.

autor	nombreautor	obra	titulo
<http://data.cervantesvirtual.com/person/19248>	Vitoria, Francisco de, 1483-1546	<http://data.cervantesvirtual.com/work/105004>	Utrum Papa possit dispensare in statutis Concilii generalis, et abrogare illa. [Francisci de Vitoria, O.P., De potestate Papae et Concilii Relectio]
<http://data.cervantesvirtual.com/person/19248>	Vitoria, Francisco de, 1483-1546	<http://data.cervantesvirtual.com/work/105008>	Tractatus de magia. [F. de Vitoria, Relectio de magia]
<http://datos.bne.es/resource/00891721>	Vitoria, Francisco de	<http://datos.bne.es/resource/002921211>	The principles of political and international law in the work of Francisco de Vitoria
<http://data.cervantesvirtual.com/person/38549>	Vitoria, Francisco de (O.P.), 1483-1546	<http://data.cervantesvirtual.com/work/103569>	Summa Sacramentorum Ecclesiae ex doctrina F. Francisci a Vitoria...

Figura 1. Búsqueda federada a través de *Wikidata Query Service*

## 5. Entonces ¿cuál es el problema?

Dejando de lado que construir esas ecuaciones de búsqueda en la propia interfaz de un servicio *Sparql*, o en una interfaz web como *Wikidata* o *Comunica*, no se caracterizan por su amigabilidad, aunque con un poco de entrenamiento esta dificultad se solventa, el problema reside sobre todo en la transición de unos registros catalogados en diferentes épocas y codificados en formato MARC a un modelo de datos con unos planteamientos muy distintos.

Hay que dejar una cosa muy clara, los autores de esta nota no le ponen ni un pero a la *BVMC*, cuya utilidad pública es extraordinaria, y nos parece un gran avance haber dado el paso de publicar en *linked open data* sus registros (en el caso de las manifestaciones se eleva a 211.887 registros, recuérdese que en *Europeana* sólo están accesibles 70.971, es decir el 33%). Todo ello es admirable, pero igual que un usuario normal interroga la

**“Sparql es una magnífica forma de acceder a la información bibliográfica semántica”**

*BVMC*, o cualquier otro sistema, y se encuentra con muchas funcionalidades muy buenas y algunas no tan buenas, el usuario accede a esa información *LOD* utilizando justamente las herramientas que la propia *BVMC* pone a su disposición, pero desde luego hay que dedicarle su esfuerzo.

Es decir, si se parte de una base de datos descrita en MARC 21 y se obtiene como resultado tras diferentes tratamientos un fichero en MARC 21 ¿Cuál es la dificultad? Pues justamente los tratamientos que es necesario realizar para llegar desde descripciones RDF según *RDA Registry* hasta registros MARC 21.

Ya hemos mencionado que estamos en una etapa de transición, es muy probable que en unos años estas transformaciones sean innecesarias y podamos utilizar directamente descripciones RDF según un vocabulario común como promete *Bibframe*<sup>21</sup>.

El hecho de que no haya ninguna descripción de estas dificultades, que nosotros sepamos, pone de manifiesto a nuestro entender que las aplicaciones *linked open data* que disponen de *Sparql* no tienen demasiados usuarios, incluso seríamos capaces de decir que casi ninguno.

El *Ministerio de Cultura* implementó en diversas fases la *Lista de Encabezamientos de Materia* en SKOS con la vinculación de los encabezamientos en castellano, catalán, gallego y vasco, así como en inglés, alemán y francés, y dispone de su correspondiente servicio *Sparql*.

<http://lid.sgcb.mcu.es>

Lo mismo se puede decir del proyecto de los *Tesoros del Patrimonio Cultural de España* de la *Subdirección General de los Museos Españoles*.

<http://tesoros.mecd.es/tesoros>

¿Cuál es el uso real de estas fantásticas aplicaciones? Pues pensamos que muy poco. Pocas son las bibliotecas que disponen de sistemas de gestión que les permitan vincular o reutilizar los encabezamientos de materia en SKOS y todavía mucho menos en los museos. De hecho, *Domus* no puede tratar la versión semántica de los *Tesoros del Patrimonio Cultural de España* (*Digimús* sí). Hace un par de años *Digibís* y *Serdoc*, que es la empresa que implementó tanto la *Lista de Encabezamientos de Materia* en SKOS como los *Tesoros del Patrimonio Cultural de España*, realizaron un proyecto conjunto para la *Biblioteca Dixital de Galicia* para gestionar y publicar en *linked open data* la *Lista de Encabezamientos de Materia en Galego*, de gran utilidad para las bibliotecas gallegas puesto que contienen los descriptores de esa lista vinculados con los encabezamientos en castellano, en catalán y en vasco y, lo que es todavía más importante, con otras listas en inglés, francés y alemán.

[http://biblioteca.galiciana.gal/gldatos\\_abiertos/datos\\_abiertos.cmd](http://biblioteca.galiciana.gal/gldatos_abiertos/datos_abiertos.cmd)

Por eso, pensamos que esta nota puede tener su utilidad, porque por un lado informa de la posibilidad de enriquecer un catálogo bibliográfico a partir de los objetos digitales de una biblioteca que disponga de *linked open data* y *Sparql* y, por otro lado, llamar la atención a los responsables de esos servicios para mitigar esas disfuncionalidades que creemos advertir, entre las que señalaremos los frecuentes *time-out*, los tiempos de respuesta propios de un servicio muy poco utilizado y el añadido de un modelo de datos complejo que previsiblemente no será el sucesor del formato MARC 21.

Es evidente que la implementación del protocolo OAI-PMH es caso de éxito como prueban los 8 millones de registros de *Hispana* o la realidad de *Europeana*, *OAIster*, la *Digital Public Library of America* y otros. Para que se pueda decir lo mismo de la tecnología *linked open data* es necesario llevarla a la práctica, ponerla en producción como ha hecho la *BVMC* con su servicio *Sparql* en [data.cervantesvirtual.com](http://data.cervantesvirtual.com) (o [datos.bne.es](http://datos.bne.es) o [datos-abertos.galiciana.gal](http://datos-abertos.galiciana.gal)) y como ha hecho la *Fundación Ignacio Larramendi* como consumidor de datos.

Consideramos que esta nota, era su intención, pone de manifiesto que *Sparql* es una magnífica forma de acceder a la información bibliográfica semántica. Estamos tan convencidos de su utilidad que se va a aplicar a todos los autores que conforman la *Biblioteca Virtual de Polígrafos*. Y para ello se utilizarán todos aquellos servicios *Sparql* que cada vez más las bibliotecas están poniendo a disposición de cualquier usuario. Ni más ni menos que al igual que ocurre con el enriquecimiento semántico, dando así confirmación a la entrada de Pace en el blog “*Next*” que citábamos al principio, *linked open data* en producción.

---

**“*Sparql* permite construir búsquedas federadas”**

---

---

**“El problema reside en la adaptación de los datos existentes a un nuevo modelo”**

---

## 6. Notas

1. Autores de la *Escuela de Salamanca*.  
[http://www.larramendi.es/esc\\_sal/i18n/cms/elemento.do?id=ms/esc\\_sal/paginas/Poligrafos.html](http://www.larramendi.es/esc_sal/i18n/cms/elemento.do?id=ms/esc_sal/paginas/Poligrafos.html)
  2. *Escuela de Salamanca 3.0*. Presentación  
[http://www.larramendi.es/esc\\_sal/es/micrositios/inicio.do](http://www.larramendi.es/esc_sal/es/micrositios/inicio.do)
  3. *Biblioteca Virtual de la Escuela de Salamanca 3.0*. Nota a esta edición digital.  
<https://bit.ly/2slWKUg>
  4. *Europeana Search API*. <https://pro.europeana.eu/resources/apis/search>
  5. *Digibib SRU/SRW*  
<http://hispana.mcu.es/i18n/sru/sru.cmd>
  6. Colecciones. *Hispana*.  
<http://hispana.mcu.es/es/comunidades/directorio.do>
  7. Participan en los respectivos nodos nacionales Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, México y Perú
  8. Nuevas directrices de *OpenAIRE* para repositorios de literatura. *LA Referencia*.  
<http://www.lareferencia.info/es/component/k2/item/222-presentamos-las-nuevas-directrices-de-openaire-para-repositorios-de-literatura>
  9. *OpenAIRE* Guidelines. <https://guidelines.openaire.eu/en/latest/>
  10. Datos enlazados. <http://data.cervantesvirtual.com/about?lang=es>
  11. *Biblioteca Virtual de la Escuela de Salamanca 3.0*. Nota a esta edición digital  
[http://www.larramendi.es/esc\\_sal/es/cms/elemento.do?id=ms%2Fesc\\_sal%2Fpaginas%2FNota\\_a\\_esta\\_edicion\\_digital.html](http://www.larramendi.es/esc_sal/es/cms/elemento.do?id=ms%2Fesc_sal%2Fpaginas%2FNota_a_esta_edicion_digital.html)
  12. *Sparql* 1.1 Overview.  
<https://www.w3.org/TR/sparql11-overview/>
  13. Tutorial de inicio a *Sparql*.  
<http://data.cervantesvirtual.com/blog/2018/10/25/tutorial-de-inicio-a-sparql>
  14. *RDA Registry*. El Registro RDA contiene datos enlazados y representaciones de la web semántica de los elementos y designadores de relaciones aprobados por el *RDA Steering Committee (RSC)*.  
<https://www.rdaregistry.info>
  15. Datos enlazados.  
<http://data.cervantesvirtual.com/about?lang=es>
  16. Aunque no contiene todas las ontologías y vocabularios utilizados en la práctica en archivos, bibliotecas y museos GLAM *flavoured Semantic Web List*, elaborada por Nicola Carboni, es una relación bastante completa y actualizable.  
<https://github.com/ncarboni/GLAM-Semantic-Web>
  17. Definition of the *Europeana Data Model* v. 5.2.8. <https://bit.ly/2L9kFkF>
  18. Algunos clientes *Sparql*:  
<https://help.poolparty.biz/pp6/developer-guide/basic-advanced-server-apis/poolparty-s-sparql-endpoint/available-sparql-clients>
  19. *Wikidata Query Service*.  
<https://query.wikidata.org/>
- Véase también por la calidad de las entradas:  
[https://www.wikidata.org/wiki/Wikidata:SPARQL\\_query\\_service/es](https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/es) *Wikidata:Servicio de consultas SPARQL*
20. *Wikidata Query Service*. User Manual. *Sparql Federation endpoints*.  
[https://www.mediawiki.org/wiki/Wikidata\\_Query\\_Service/User\\_Manual/SPARQL\\_Federation\\_endpoints](https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual/SPARQL_Federation_endpoints)
  21. *Bibframe* model, vocabulary, guidelines, examples, notes, analyses.  
<https://www.loc.gov/bibframe/docs/index.html>

## 7. Referencias

**Agenjo-Bullón, Xavier; Hernández-Carrascal, Francisca; Juez-García, Patricia** (2013). "La *Escuela de Salamanca* desde el punto de vista de la web semántica y la información en la Red". En: *Crisis de la modernidad y filosofías ibéricas. X Jornadas Internacionales de Hispanismo Filosófico* Universidade de Santiago de Compostela, 13-15 abril.

Fundación Ignacio Larramendi; Departamento de Filosofía de la Universidad de Santiago de Compostela y Asociación de Hispanismo Filosófico, pp. 199-215.  
<http://eprints.rclis.org/19039/>

**Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca** (2018a). "Registros de autoridades, enriquecimiento semántico y Wikidata". *Anuario ThinkEPI*, v. 12, pp. 361-372.  
<https://doi.org/10.3145/thinkepi.2018.61>

**Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca** (2018b). "Visibilidad y accesibilidad digital de la Escuela de Salamanca en un entorno de linked open data". En: *La Escuela de Salamanca, ¿un ejemplo de producción global de conocimiento?* Max Planck Institute for European Legal History. Buenos Aires, 24-26 de octubre de 2018.  
<http://www.larramendi.es/fundacion/wp-content/uploads/2018/pdf/visibilidad-accesibilidad-bv-es-3.pdf>

*BNE* (2018). "La búsqueda en la Biblioteca Digital del Patrimonio Iberoamericano y en la Biblioteca Digital Hispánica al alcance de un clic". *Biblioteca Nacional de España. Noticias*, 30 octubre.  
<http://www.bne.es/es/AreaPrensa/noticias2018/1030-API-busqueda-BDH-BDPI.html>

**Coombs, Karen** (2016). "Federated queries with Sparql". *OCLC developer network. News*, April 22.  
<https://www.oclc.org/developer/news/2016/federated-queries-with-sparql.en.html>

*Fundación Ignacio Larramendi* (2018). "Conferencia Internacional 'The School of Salamanca: A case of global knowledge production?'". *Fundación Ignacio Larramendi, News*, 31 octubre.  
<http://www.larramendi.es/fundacion/conferencia-internacional-the-school-of-salamanca-a-case-of-global-knowledge-production-2>

**Hyland, Bernadette; Ateazing, Ghislain; Villazón-Terrazas, Boris** (2014). *Best practices for publishing linked data*. W3C Working Group Note 09 January 2014.  
<https://www.w3.org/TR/ld-bp>

**Pace, Andrew K.** (2018). "Linked data in libraries: From disillusionment to productivity". *Next*, November 8.  
<http://www.oclc.org/blog/main/linked-data-in-libraries-from-disillusionment-to-productivity>

**Smith-Yoshimura, Karen** (2018). "Analysis of 2018 International linked data survey for Implementers". *Code{4}lib Journal*, n. 42.  
<https://journal.code4lib.org/articles/13867>

**Taelman, Ruben; Van-Herwegen, Joachim; Sande, Miel-Vander; Verborgh, Ruben** (2018). "Comunica: a modular *Sparql* Query Engine for the Web". En: Vrandečić, Denny (ed.). *Proceedings of the 17th International Semantic Web Conference*, pp. 239-255.  
<https://comunica.github.io/Article-ISWC2018-Resource>

**Xavier Agénjo-Bullón**  
Fundación Ignacio Larramendi  
[xavier.agenjo@larramendi.es](mailto:xavier.agenjo@larramendi.es)

**Francisca Hernández-Carrascal**  
*Digibis*  
[francisca.hernandez@digibis.com](mailto:francisca.hernandez@digibis.com)