



## **Agregación y difusión de registros de autoridad en Linked Open Data<sup>1</sup>**

Authors<sup>2</sup>:

**Xavier Agenjo**

Director de proyectos de la Fundación Ignacio Larramendi y director de la Biblioteca Virtual de Polígrafos

[xavier.agenjo@larramendi.es]

**Francisca Hernández**

Bibliotecaria. Consultora de DIGIBÍS, Producciones Digitales

[francisca.hernandez@digibis.com]

**Andrés Viedma**

Arquitecto de software, DIGIBÍS

Producciones Digitales

[andres.viedma@digibis.com]

**Meeting:**

**Cataloguing Section**

---

### **Abstract:**

*Se describe, mediante el análisis de la Biblioteca Virtual de Polígrafos, la agregación y difusión de los datos de registros de autoridad a través de Linked Open Data. El objetivo es reunir información, datos, textos digitales y sitios web sobre polígrafos, españoles, hispanoamericanos, brasileños y portugueses de todos los tiempos. Como tales estos agregados reúnen información sobre el pensamiento, la filosofía, la política, la ciencia, etc. de España, Hispanoamérica, Portugal y Brasil, escrita en cualquier lengua (latín, árabe, hebreo, español, portugués...) y en cualquier momento (desde Séneca en el primer siglo antes de Cristo hasta la actualidad).*

*La columna vertebral del sistema son los autores. Para cada autor se crea un registro de autoridad conforme a MARC21/RDA que se enriquece con datos biográficos, se categorizan atributos específicos para enriquecer las relaciones y favorecer la navegabilidad del sitio (profesión, ocupación, género, membresía, fechas de nacimiento y muerte, lugares de nacimiento y muerte, lenguas y alfabetos). En ese sentido, cada registro de autoridad agrega información procedente de múltiples fuentes y*

---

<sup>1</sup> Todas las URL y cifras de este texto han sido consultadas y actualizadas a fecha 30 de abril de 2011.

<sup>2</sup> Agradecemos a César Juanes, del departamento de I+D de DIGIBÍS, la revisión del texto de esta comunicación.

vocabularios. También para cada autor se describe, siguiendo el MARC/RDA (12 update), las versiones digitales de sus obras. De forma análoga, cada autor se relaciona con otros autores, así como traductores, editores o comentaristas, como procedimiento para seguir la transmisión textual de sus obras, utilizando registros de autor título estructurados según el modelo Obra- Expresión- Manifestación - Item conforme a la próxima propuesta de MARBI.

La creación de esas autoridades ha tenido lugar mediante el cotejo de diferentes fuentes de información de la máxima fiabilidad (ficheros de autoridades, enciclopedias, diccionarios biográficos, etc.). La relación de un autor concreto con sus correspondientes datos se ha realizado manualmente hasta 2010. Desde 2011 la Fundación Ignacio Larramendi está desarrollando diferentes herramientas para, a partir de fuentes Linked Open Data, obtener todo tipo de información. La comunicación detalla el procedimiento para obtener URIs de los recursos LOD de forma tal que se generen búsquedas automáticas contra servidores SRU/OpenSearch o SPARQL Endpoint, o realizando este enriquecimiento semántico a través de ficheros disponibles en LOD. La exportación de los datos se realiza en MARC 21, así como Europeana Data Model 5.2.1, SKOS y el VIAF, para enriquecerlos semánticamente a partir de vocabularios tales como LCSH, VIAF, GeoLinkedData (del Instituto Geográfico Nacional) y GeoNames.

---

## **1. Introducción: La Biblioteca Virtual de Polígrafos**

Las siguientes reflexiones sobre agregación y difusión de registros de autoridades en un entorno Linked Open Data no parten de la elucubración teórica, sino que tienen su base en una práctica bibliotecaria actualmente en producción. Estos trabajos catalográficos se realizan en la Biblioteca Virtual de Polígrafos<sup>3</sup> que forma parte de las Bibliotecas Virtuales de la Fundación Ignacio Larramendi<sup>4</sup>. El objetivo fundamental de la Biblioteca Virtual de Polígrafos es dar una especial relevancia y significación a la obra de los pensadores españoles, portugueses, brasileños e hispanoamericanos y esta es una de las características principales del proyecto bibliográfico ponerlas en relación con otras obras de características similares, para hacerlas ampliamente accesibles internacionalmente.

Se quiere resaltar desde el principio, aunque parezca obvia, la importancia de disponer de un diseño bibliográfico previo, aspecto que debe prevalecer sobre las cuestiones tecnológicas, máxime en un tema tan novedoso como Linked Open Data. Precisamente, son estos objetivos bibliográficos los que marcan los

---

<sup>3</sup> <http://goo.gl/3kn00>

<sup>4</sup> Para un estado de la cuestión de este proyecto puede consultarse la comunicación Agenjo Bullón, Xavier y Hernández Carrascal, Francisca. *La Biblioteca Virtual Larramendi: fuente de información bibliográfica para el pensamiento iberoamericano en la Web 3.0*. En: *Jornadas Virtuales Iberoamericanas de Bibliotecología*. [<http://goo.gl/VbwtN>] Estas páginas fueron redactadas en otoño de 2010, habiéndose producido desde entonces algunos cambios importantes. Muchos de estos cambios pueden verse en Agenjo, Xavier y Hernández, Francisca. *La Biblioteca Virtual Ignacio Larramendi desde la perspectiva LOD y EDM* que se presentó en el I Seminario Internacional de la Biblioteca de Galicia [<http://goo.gl/uFrXx>].

requisitos funcionales tanto de la práctica catalográfica como del desarrollo específico de la Biblioteca Virtual de Polígrafos. El objetivo del proyecto es reunir y relacionar información bibliográfica, o información a secas, sobre el pensamiento iberoamericano desde una perspectiva histórica, estableciendo a su vez relaciones con otras 'civilizaciones', por usar el término de Toynbee<sup>5</sup>.

### Las fuentes de información para registros de autoridad

Es evidente que una gran parte de las fuentes de información que resultan útiles para los objetivos del proyecto bibliográfico no están disponibles en soporte digital, o, estándolo, carecen de una estructura apropiada. Y en otros casos, en los que el recurso está disponible en la Web, el diseño de su interfaz le resta capacidad y rendimiento informativo. Por ejemplo, el *Diccionario Biográfico Español de la Real Academia de la Historia* proporciona una información sobre varios miles de personajes capitales de la historia y de la cultura hispana, entendiéndola en el sentido más amplio, pero de una forma tecnológicamente paupérrima. Otra fuente de información fundamental es *Hombres y documentos de la filosofía española* de Gonzalo Díaz y Díaz, publicado en 7 volúmenes entre 1980 y 2003, que peca del mero acarreo erudito, con escaso criterio heurístico, pero que facilita el acceso a una ingente cantidad de información<sup>6</sup>. Lo mismo podría decirse del magnífico *Diccionario de filosofía* de Ferrater Mora<sup>7</sup>, mucho más riguroso desde el punto de vista intelectual, aunque más escaso en lo que se refiere a la información propiamente dicha.

Es una lástima que magníficos repertorios bibliográficos, publicados en su día en papel y todavía protegidos por los derechos de autor, sean inaccesibles en la Web y en especial a las tecnologías como Linked Open Data. Se da la paradoja de que los grandes proyectos de digitalización retrospectiva se detienen, según los países y la legislación correspondiente, en torno a los 50, 70 e incluso 80 años posteriores a la muerte del autor, con lo que es posible acceder fácilmente en la Web a contenidos completamente anticuados, mientras otros más valiosos y actualizados no están disponibles. Cuando se habla de los problemas de Open Access suele olvidarse uno elemental, que no se trata de que la información

---

<sup>5</sup> Como se recordará Toynbee en su *A Study of History* permitía determinar hasta 21 civilizaciones diferentes y aunque esta información puede parecer superada e incluso la aplicación del concepto civilización por parte de Huntington en su célebre libro (y a los efectos que nos interesan en el no tan célebre que siguió le después) lo cierto es que parecen existir compartimentos estancos entre las culturas que sin duda alguna el intercambio de información, especialmente planteado como se lleva a cabo en Tim Berners-Lee en su seminal *Design Issues: Linked Data* [<http://www.w3.org/DesignIssues/LinkedData>], deberían solventar.

<sup>6</sup> Puede consultarse en la *Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico* [<http://goo.gl/NHiVW>]. La transcripción de los textos no parece ser integral.

<sup>7</sup> Ferrater, Mora J. *Diccionario de Filosofía*. Madrid: Alianza, 1979. Se trata de la edición definitiva de Ferrater Mora, posteriormente ha aparecido una edición ampliada por la Cátedra Ferrater Mora bajo la dirección del profesor Josep María Terricabras que –afortunadamente– identifica con un asterisco lo que añade

digital esté accesible o no de forma abierta, sino sencillamente de que no existe tal información digital.

Esta circunstancia, que no debe perderse de vista en un congreso de bibliotecarios, entre los cuales la bibliografía debe ser, más que en ningún otro caso, una disciplina fundamental, nos hace olvidar que, con frecuencia, las mejores ediciones críticas que cotejan manuscritos y variantes y disponen de estudios preliminares, etc., no están accesibles en la red. Desde luego, existen admirables proyectos, cada vez más frecuentes, en los cuales la *ecdótica digital* se aplica significativamente, pero son una minoría. Por su extraordinaria calidad citaremos el proyecto *Mark Twain Papers & Project*<sup>8</sup> por lo que tiene de ejemplo de una edición digital de un autor relevante.

Por fortuna, dentro de la Biblioteca Virtual de Polígrafos se ha intentado al menos recurrir a un especialista, denominado polígrafista, con el objeto de presentar una visión panorámica del estado de la cuestión de los estudios bibliográficos sobre cada autor (polígrafo) en particular. Además, en general, se ha establecido un mínimo de fuentes de información que deben ser consultadas en todos los casos: *Diccionario Biográfico de la Real Academia de la Historia*<sup>9</sup>, *Fichero de Autoridades de la Biblioteca Nacional*<sup>10</sup>, *Virtual International Authority File*<sup>11</sup> (que incluye las autoridades de la Biblioteca Nacional de España y de la Biblioteca Nacional de Portugal), *Library of Congress Authorities*<sup>12</sup>, y los impresos ya mencionados *Hombres y documentos de la Filosofía Española* y *Diccionario de Filosofía*. Lógicamente se utilizan además otras fuentes de información como *Stanford Encyclopedia of Philosophy*<sup>13</sup>, *The Catholic Encyclopedia*<sup>14</sup>, *Enciclopedia católica*<sup>15</sup>, *Jewish virtual Library*<sup>16</sup>, *Islamic philosophy Online*<sup>17</sup>, *Proyecto Filosofía en Español*<sup>18</sup>, *Biblioteca Miralles*<sup>19</sup>, *Biblioteca Saavedra y Fajardo*<sup>20</sup>, *Biblioteca Virtual Miguel de Cervantes*<sup>21</sup>, *Proyecto Sarmiento*<sup>22</sup>, entre otras.

---

<sup>8</sup> <http://bancroft.berkeley.edu/MTP/>

<sup>9</sup> <http://www.rah.es/diccBiografico.htm>

<sup>10</sup> <http://catalogo.bne.es/uhtbin/authoritybrowse.cgi>

<sup>11</sup> <http://viaf.org/>

<sup>12</sup> <http://authorities.loc.gov/>

<sup>13</sup> <http://plato.stanford.edu/>

<sup>14</sup> <http://www.newadvent.org/cathen/>

<sup>15</sup> <http://ec.aciprensa.com/>

<sup>16</sup> <http://www.jewishvirtuallibrary.org/>

<sup>17</sup> <http://www.muslimphilosophy.com/>

<sup>18</sup> <http://www.filosofia.org>

<sup>19</sup> <http://www.bibliotecamiralles.org/escritores.html>

## De los registros de autoridad a los registros bibliográficos

Entre las fuentes generales se ha incluido preferentemente el VIAF, a partir del cual se han localizado los autores en una serie de fuentes de información autorizadas como la Library of Congress, la British Library o la Deutsche Bibliothek. Lo que constituye una gran ventaja de VIAF es la posibilidad de navegar desde la agrupación de registros de autoridades a cada uno de los encabezamientos constituyentes, procedentes de las distintas fuentes que contribuyen a VIAF. Esto permite priorizar la elección de encabezamientos según su procedencia, la Biblioteca Nacional de España para autores españoles, la Biblioteca Nacional de Portugal para autores portugueses, la de Brasil<sup>23</sup> para los brasileños y así sucesivamente. Y, en último término, como nexo común, los procedentes de la Library of Congress. Lo esencial es que, a partir de una búsqueda puede establecerse una nueva y muy importante funcionalidad de carácter universal consistente en formar un agregado digital de todos los registros bibliográficos vinculados con los registros de autoridades. Con ello se localizan los nombres de un autor y sus variantes, lo que permite llevar a cabo un enriquecimiento muy completo gracias a los registros bibliográficos asociados, realizados en la mayoría de los casos por Agencias Bibliográficas Nacionales.

VIAF permite conectarse en estos momentos a casi una veintena de ficheros de autoridad a través de los cuales se puede acceder a un gran número de catálogos bibliográficos. Es muy importante señalar que sólo en algunos casos se vinculan los campos 100 de autoridades con los campos 6XX de las descripciones bibliográficas, y que si esta funcionalidad estuviera siempre presente sería posible acceder a otras obras sobre un autor<sup>24</sup>, utilizando VIAF a modo de concentrador o *hub* de recursos bibliográficos no sólo *de* un autor, sino también sobre las *obras* de ese mismo autor.

## Fuentes de información para descripciones bibliográficas

De igual forma que se han definido una serie de fuentes para los autores, se ha establecido el mismo mecanismo para las recopilaciones bibliográficas y así se han seleccionado el *Catálogo Colectivo de Patrimonio Bibliográfico*<sup>25</sup>, la *Heritage of the Printed Book Database* (HPB) del Consortium of European Research Libraries

---

<sup>20</sup> <http://saavedrafajardo.um.es/Biblioteca/IndicesW.nsf/Inicio?OpenForm&m=2>

<sup>21</sup> [www.cervantesvirtual.com](http://www.cervantesvirtual.com)

<sup>22</sup> <http://www.proyectosarmiento.com.ar/>

<sup>23</sup> No incluida aún en VIAF

<sup>24</sup> Resulta particularmente interesante el acceso a través de VIAF a los archivos y bibliotecas de Canadá [<http://www.collectionscanada.gc.ca/>], puesto que en este caso y aunque dentro de los propósitos y objetivos de la Biblioteca Virtual de Polígrafos no ha sido posible encontrar, al menos por el momento, alguno de los autores que conforman el proyecto, sí sería posible hacerlo si la metodología se extendiera.

<sup>25</sup> <http://www.mcu.es/bibliotecas/MC/CCPB/index.html>

(CERL)<sup>26</sup>, el *Novum Regestrum*<sup>27</sup> y el *WorldCat*. Mención aparte merece la *Biblioteca Virtual Menéndez Pelayo*<sup>28</sup> ya que la Biblioteca Virtual de Polígrafos continúa la labor bibliográfica y recopilatoria de la obra de los principales pensadores españoles y su influencia, que constituyó el tema de *La ciencia española*<sup>29</sup> con la que Menéndez Pelayo respondía a la famosa pregunta de Monsieur Masson<sup>30</sup> *Que doit-on à l'Espagne? Et depuis deux siècles, depuis quatre, depuis dix, qu'a-t-elle fait pour l'Europe?*.

Además, hay que mencionar dos proyectos de gran calado que se están realizando de forma paralela en el tiempo, aunque con distintas dimensiones y que afectan, no sólo por su contenido, sino también desde el punto de vista de los requisitos funcionales, al desarrollo de la Biblioteca Virtual de Polígrafos. Estos proyectos son Hispana<sup>31</sup> y Europeana<sup>32</sup>. Sobre Hispana no hay muchas publicaciones que documenten su creación y evolución, lo cual es de extrañar dadas las características y dimensiones del proyecto, si bien se cuenta con numerosas presentaciones<sup>33</sup>. Cuando se redactan estas páginas [30 abril 2011] Hispana recoge 3.181.786 objetos digitales procedentes de 146 repositorios españoles. Sobre Europeana hay, evidentemente, mucha más documentación, pero para el propósito de esta comunicación importa señalar el nuevo modelo de datos *Definition of the Europeana Data Model Elements*, version 5.2.1<sup>34</sup>, actualizado el 7 de marzo de 2011, y las *Functional specification for the Europeana Danube release*, publicados el 31 de agosto de 2010<sup>35</sup>.

No es casualidad que los desarrollos que se llevan a cabo para la Biblioteca Virtual de Polígrafos y para la Fundación Ignacio Larramendi estén tan

---

<sup>26</sup> Heritage of the Printed Book Database (HPB) del CERL  
<http://www.cerl.org/web/en/resources/hpb/main>

<sup>27</sup> <http://goo.gl/fqtHA>. Véase, además, *Novum Regestrum: el Catálogo Colectivo del Patrimonio Bibliográfico Iberoamericano* de Xavier Agenjo Bullón, Francisca Hernández Carrascal, Boletín de la ANABAD, ISSN 0210-4164, Tomo 44, Nº 4, 1994, págs. 127-142  
<http://dialnet.unirioja.es/servlet/articulo?codigo=50938>

<sup>28</sup> <http://www.larramendi.es/i18n/bvmpelayo/inicio.cmd>

<sup>29</sup> Menéndez Pelayo, Marcelino. *La ciencia española*. En: *Biblioteca Virtual Menéndez Pelayo*.  
<http://goo.gl/iB5m>

<sup>30</sup> Nicolas Masson de Morvilliers (1740-1789)

<sup>31</sup> <http://hispana.mcu.es>

<sup>32</sup> <http://europeana.eu>

<sup>33</sup> *Hispana y las iniciativas del Ministerio de Cultura* María Antonio Carrato Mena *Jornada de Difusión de EuropeanaLocal*, 17 nov. 2010 [<http://hdl.handle.net/10421/4765>]. *La aplicación del Modelo de Datos de Europeana a la Biblioteca Virtual de Patrimonio Bibliográfico : bvpb.mcu.es*, María Luisa Martínez-Conde *Jornada de Difusión de EuropeanaLocal*. 17 nov 2010 URI: <http://hdl.handle.net/10421/4783>. La más reciente es Carrato, María Antonia. *Hispana. I Seminario Internacional de la Biblioteca de Galicia* [<http://goo.gl/6xIqq>]

<sup>34</sup> <http://goo.gl/ojIL>

<sup>35</sup> <http://goo.gl/P6jme>

fuertemente relacionados con las especificaciones funcionales de Europea. La razón es esta: de forma completamente mayoritaria España participa en Europea mediante Hispana (con 1.367.808 objetos digitales, a gran distancia de la Biblioteca Virtual Cervantes<sup>36</sup> que aporta sólo 19.062, aunque este segundo proyecto esté mucho más publicitado). Se da la circunstancia de que tanto Hispana como la mayoría de los proveedores de datos recolectados por Hispana<sup>37</sup> se realiza con el ILS DIGIBIB, actualmente en la versión 6.0, pero que ya estará en la versión 7.0 cuando se presente esta comunicación en Puerto Rico.

Tampoco se debe pasar por alto el hecho de que la empresa DIGIBÍS<sup>38</sup>, que lleva a cabo los desarrollos del programa de gestión bibliotecaria DIGIBIB y el archivístico DIGIARCH, así como de un repositorio OAI denominado OAsIs, es una empresa propiedad de la Fundación Ignacio Larramendi, que utiliza sistemáticamente a la Biblioteca Virtual de Polígrafos como *testbed* para sus desarrollos. Por ello (y si así lo desean los clientes) muchas de las funcionalidades aquí descritas afectarán de forma considerable a una treintena de importantes bibliotecas españolas y más de un centenar, si se cuentan las que recogen colectivamente la Biblioteca Virtual de Patrimonio Bibliográfico<sup>39</sup> y la Biblioteca Virtual de Prensa Histórica.

Cuando esta comunicación se lea en Puerto Rico habrá entrado ya en producción la fase Danubio del proyecto Europea, previsto para el 1 de mayo de 2011. Naturalmente, no podemos estar seguros de que esto sea definitivamente así, pero de cualquier modo, el hecho de que se produzca algún retraso es completamente independiente de lo que se ha venido planificando en la Biblioteca Virtual de Polígrafos, puesto que lo verdaderamente decisivo ha sido la opción tomada por Europea para desarrollar nuevas funcionalidades sobre una estructura de información semántica. Por estas características, tanto Europea como Hispana son dos fuentes primordiales de información para la Biblioteca Virtual de Polígrafos, así como dos ejemplos metodológicos que, como es el caso de Hispana, además se retroalimentan.

---

<sup>36</sup> <http://www.cervantesvirtual.com/>

<sup>37</sup> Biblioteca Virtual de Prensa Histórica, Galiciana: Biblioteca Digital de Galicia, Biblioteca Virtual de Andalucía, Gredos (Universidad de Salamanca, Spain), Biblioteca Digital de Madrid, Biblioteca Digital de Castilla-La Mancha, Centro de Documentación de Fundación MAPFRE, Biblioteca Digital de Castilla y León, Biblioteca Virtual del Patrimonio Bibliográfico, Biblioteca Virtual de Derecho Aragonés, Biblioteca Digital Real Academia de la Historia, Catálogo Colectivo de la Red de Bibliotecas de los Archivos Estatales, Biblioteca Virtual del Principado de Asturias, Archivo de la Imagen de Castilla La Mancha, Biblioteca Valenciana Digital, Biblioteca Virtual de Aragón, Fundación Sancho el Sabio, Biblioteca Regional de Murcia, Biblioteca Digital de Aranjuez, Universidad de La Laguna, Biblioteca Virtual de la Diputación de Zaragoza, Biblioteca Virtual de La Rioja, Fundación Ignacio Larramendi, Biblioteca Virtual de la Real Academia Nacional de Farmacia.

<sup>38</sup> <http://www.digibis.com>

<sup>39</sup> Puede verse especialmente en la Biblioteca Virtual de Patrimonio Bibliográfico el subdominio Iberoamérica en las colecciones de la BVPB [<http://goo.gl/8Jyfa>]. También puede resultar de utilidad la búsqueda a texto completo en la Biblioteca Virtual de Prensa Histórica [<http://prensahistorica.mcu.es>].

De momento, Europeana ha liberado una API basada en OpenSearch<sup>40</sup> que permite la integración en un sitio Web la búsqueda y visualización de registros de Europeana. Esta API, que funciona muy eficazmente y se ha incorporado a la interfaz web de la Biblioteca Virtual de Polígrafos, lanza una consulta a partir de un término de búsqueda inicial en la base de datos de Europeana. Es decir, que la misma consulta realizada sobre una base de datos de varios miles de registros, se ejecuta también sobre la base de datos de Europeana que cuenta con cerca de veinte millones objetos digitales (17.901.019).

## **2. Agregados digitales**

Según la metodología establecida para la Biblioteca Virtual de Polígrafos, el núcleo sobre el que gira todo el sistema de información lo constituyen los denominados 'Agregados digitales', que incluso de una forma tipográfica se aproximan mucho a las tradicionales entradas enciclopédicas e incluso pretenden alcanzar el nivel de complejidad y completitud de algunas entradas de la Wikipedia (o DBpedia, como se verá más adelante).

El término agregación o agregados tiene diferentes usos según el entorno en el que se aplique y puede generar cierta indeterminación en el lector. Así, en los procesos de recolección de metadatos, los recolectores agregan metadatos y pueden denominarse agregadores, como así ocurre en Europeana. Sin embargo, el concepto de agregado digital que se utiliza en la Biblioteca Virtual de Polígrafos está mucho más acorde con la definición que realiza *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE)<sup>41</sup>, por la cual un agregado constituye un recurso en sí mismo, formado por un conjunto de otros recursos relacionados entre sí. Un agregado digital sería, pues, el conjunto de datos y recursos informativos reunidos en torno a un determinado autor, cuyo núcleo de información o nodo principal está formado por un registro de autoridad de persona.

El proceso de control de autoridades que realiza la Fundación Ignacio Larramendi es mucho más amplio que el que tradicionalmente se lleva a cabo en las grandes bibliotecas o agencias bibliográficas. No insistiremos en la función de identificación de las entidades *personas* sino las tareas de contextualización, al modo definido por FRAD<sup>42</sup>. Con ello no sólo se registran datos no habituales hasta ahora en los ficheros de autoridad, sino también se identifican y enlazan los registros de autoridad con otras fuentes, recursos y datos disponibles en la red. En este sentido, el registro de fuentes de información de las cuáles se obtienen datos (670) es no sólo una justificación para la elección de los nombres y sus variantes, sino también un registro de

---

<sup>40</sup> <http://www.version1.europeana.eu/web/api>

<sup>41</sup> <http://www.openarchives.org/ore/>

<sup>42</sup> [http://www.ifla.org/files/cataloguing/frad/frad\\_2009-es.pdf](http://www.ifla.org/files/cataloguing/frad/frad_2009-es.pdf)



fuentes de información de recursos bibliográficos seleccionados acerca de un determinado autor.

De un modo similar a como lo realiza VIAF cada registro de autoridad queda definido, siguiendo el formato MARC 21, por su encabezamiento principal (1XX), las variantes de nombre (4XX), las relaciones con otros nombres (5XX) y las equivalencias en otras lenguas de los encabezamientos principales (7XX). Este último aspecto es de gran importancia para la Biblioteca Virtual de Polígrafos puesto que todos los autores están seleccionados no sólo por la importancia dentro de la cultura iberoamericana, sino por la aportación de la cultura iberoamericana a la cultura universal, motivo por el cual es absolutamente necesario disponer de las equivalencias en otras lenguas para seguir el hilo de la transmisión de los textos a nivel mundial, así como de la transmisión y repercusión de las ideas. Pero la misma situación se da en España dónde coexisten junto al castellano o español otros idiomas oficiales como el catalán, el valenciano, el gallego y el vasco.

### Encabezamientos multilingües

Se quiere llamar la atención sobre la necesidad de que el formato MARC 21 cuente con algún elemento que permita la codificación del idioma de los encabezamientos. Por desgracia, el *Discussion Paper 2001-DP05 Multilingual Authority Records in the MARC 21 Authority Format*<sup>43</sup>, parece que no ha concluido en ninguna propuesta, al menos de momento. Los campos 7XX permiten registrar encabezamientos equivalentes en otros idiomas, pero no tienen una forma para indicar el idioma del encabezamiento, lo cual es especialmente necesario en ficheros de autoridad multilingües. Otros mecanismos presentan también limitaciones; la posición 008/08 sólo permite codificar los idiomas inglés y/o francés como lengua del catálogo y el 040 \$b que permite sólo establecer el idioma de catalogación que en muchas ocasiones difiere del idioma del encabezamiento.

Esta situación está razonablemente superada en MADS y desde luego en la propuesta de MADS/RDF<sup>44</sup>, pero no debe olvidarse que la mayor parte de los registros de autoridad se encuentran tan solo en formato MARC y que, por tanto, será difícil trasladar de forma inmediata este tipo de datos al entorno Linked Open Data. Así, VIAF ha heredado este problema y mientras es posible reconocer la institución, y país, de procedencia de un registro, no puede saberse, salvo por deducción, en qué idioma está expresado un encabezamiento. Esto es especialmente importante en los casos de encabezamientos elaborados en una misma área lingüística que pueden ser muy similares o indistinguibles (p.e., Geografía en español y Geografia [sin tilde] en italiano o catalán).

---

<sup>43</sup> <http://www.loc.gov/marc/marbi/2001/2001-dp05.html>

<sup>44</sup> MADS/RDF Primer. Final Public Review Document. [<http://www.loc.gov/standards/mads/rdf/>]

## Variantes de nombres de persona

La búsqueda de información en otros ficheros de autoridad tiene no sólo la intencionalidad de establecer relaciones de equivalencia lingüística entre los encabezamientos, sino también de consignar la mayor cantidad posible de variantes de los nombres y de relaciones con otros nombres. Este procedimiento, adecuado al comportamiento del ILS DIGIBIB, permite que al realizar importaciones de registros bibliográficos estos queden perfectamente enlazados con su correspondiente registro de autoridad y controla además la generación de entradas de autoridad no válidas a partir de los campos 4XX. El proceso en sí es más lento en el inicio, en el momento de la creación del registro de autoridad, pero es mucho más productivo a medio plazo y ofrece mucha mayor calidad. Además, debe recordarse que el proyecto bibliográfico originario se consigue mucho mejor si se analizan las fuentes partiendo desde el núcleo central de los registros de autoridad.

Este procedimiento es especialmente importante si tenemos en cuenta que se trata de un conjunto de autores que abarcan desde Séneca (Córdoba 4 a.C.- Roma 65 d.C.) hasta Martín de Riquer (Barcelona, 1914-) y que muchos de ellos han utilizado diferentes nombres según el idioma en el que hayan escrito, por no mencionar que estos nombres han sido traducidos a diferentes idiomas junto con sus obras (p.e., Averroes es el nombre latino de Ibn Rus). Disponer, por tanto, de una serie de variantes de nombres, adecuadas además a los recursos que se desea importar, permite que cada inclusión de un recurso necesite un número menor de modificaciones en los registros de autoridad al estar ya consignadas un gran número de variantes.

### **3. Registros de autoridad**

La actualización nº 12 del formato MARC 21, publicada en octubre de 2010, consolida una serie de campos, siguiendo la filosofía de *Resource Description and Access*, que resultan esenciales para la agregación de información sobre personas. Nos referimos, como es obvio, a los campos 3XX. Desde un punto de vista histórico<sup>45</sup>, los elementos que en un primer momento tuvieron como objetivo básico diferenciar entre homónimos, poco a poco fueron adquiriendo un valor constitutivo en sí mismo. Así, los diferentes subcampos del campo 100 que sólo se recomendaba utilizar para diferenciar autores con el mismo nombre, pasaron a adquirir un criterio de obligatoriedad, muy en particular el \$d, pues contextualizaba inmediatamente desde un punto de vista cronológico a un autor, lo que ayudaba en muchísimas ocasiones a relacionarle con la obra y deshacer posibles errores. Hasta la adopción de RDA por MARC 21<sup>46</sup> no había elementos específicos para atributos de persona tan expresivos como el lugar de nacimiento. Justamente estos atributos, referidos no tanto al nombre de la

---

<sup>45</sup> Agenjo Bullón, Xavier y Palá Gasós, Pilar. *El fichero de autoridades del Catálogo Colectivo del Patrimonio Bibliográfico*. En: *Boletín de la ANABAD*, ISSN 0210-4164, Tomo 37, Nº 4, 1987, págs. 593-606. [ <http://dialnet.unirioja.es/servlet/articulo?codigo=801041> ]

<sup>46</sup> <http://www.loc.gov/marc/RDAinMARC29.html>

persona sino a la persona, contribuyen a darle a los registros MARC un aspecto esencialmente enciclopédico, casi dbpédicos, si se nos permite el neologismo.

Para los fines de la Biblioteca Virtual de Polígrafos la publicación de RDA supuso un fortalecimiento de sus líneas de trabajo. En primer lugar, en la recuperación de información sobre los autores, siguiendo las tareas de usuario definidas en las FRAD como contextualización y búsqueda. Y en segundo lugar, en la forma de estructurar los recursos de información siguiendo el modelo Work-Expression-Manifestation-Item<sup>47</sup>. Por este motivo se aplicaron inmediatamente las sucesivas actualizaciones del formato MARC 21 hasta el punto de que el programa DIGIBIB ha acompasado la liberación de sus actualizaciones a la publicación de las actualizaciones de MARC 21.

Para cada autor, en la medida de lo posible, se incluye, además de la forma canónica del nombre, sus variantes y equivalencias en otras lenguas, la biografía del autor (678), las fechas de nacimiento y muerte (046), los lugares de nacimiento y defunción (370) y de cualquier otra actividad (370), su ocupación (372 y 374), el idioma utilizado (377) o el género (375). Son atributos de la entidad *persona* que tienen como finalidad última sustentar la búsqueda y navegación semánticas entre los datos. Estos atributos permiten que el sitio Web de Biblioteca Virtual de Polígrafos ofrezca la posibilidad de buscar autores por las fechas en las que desarrollaron su actividad, por los lugares asociados a su biografía, por sus profesiones y ocupaciones, y no sólo por sus obras, lo que es relativamente frecuente en los OPAC Web.

#### Registros de autoridad en otros esquemas de metadatos

De igual interés para la *BIBLIOTECA VIRTUAL DE POLÍGRAFOS* es la información de entidades, personas y familias procedentes de los archivos. La necesidad de que los datos de archivos, bibliotecas y museos converjan está ya muy asumida y con algunos ejemplos de notable calidad. Es ya frecuente que se plantee unificar bajo una única interfaz de consulta este tipo de datos, lo que contribuye a potenciar la capacidad informativa de los mismos. Muy recientemente en DIGIBÍS se ha llevado a cabo un trabajo de estas características con el proyecto de la Junta de Castilla-La Mancha dedicado al Archivo de la Imagen (fotografías)<sup>48</sup> y al Archivo de la palabra (archivos sonoros)<sup>49</sup>. Para todo ello se ha realizado un mapeo sistemático<sup>50</sup> de MARC 21 a EAD y a EAC-CPF, así como a ISAD (G) y a ISAAR (CPF).

---

<sup>47</sup> MARC Discussion Paper No. 2011-DP03: Identifying Work, Expression, and Manifestation records in the MARC 21 Bibliographic, Authority, and Holdings Formats  
[\[http://www.loc.gov/marc/marbi/2011/2011-dp03.html\]](http://www.loc.gov/marc/marbi/2011/2011-dp03.html)

<sup>48</sup> [http://clip.jccm.es/archivo\\_de\\_la\\_imagen/es/micrositios/inicio.cmd](http://clip.jccm.es/archivo_de_la_imagen/es/micrositios/inicio.cmd)

<sup>49</sup> [http://clip.jccm.es/archivo\\_de\\_la\\_palabra/es/micrositios/inicio.cmd](http://clip.jccm.es/archivo_de_la_palabra/es/micrositios/inicio.cmd)

<sup>50</sup> [http://www.sedic.es/p\\_boletin\\_novedades\\_mensual.asp#Digibis](http://www.sedic.es/p_boletin_novedades_mensual.asp#Digibis)

Igualmente, y como era de esperar, en la Biblioteca Virtual de Polígrafos ya se ha manifestado la necesidad de vincular materiales bibliográficos y archivísticos relacionados, por ejemplo, con Agustín de Montiano y Luyando (1697-1794), historiador, crítico y dramaturgo español que ha pasado a la memoria colectiva por ser el primer director de la Real Academia de la Historia, fundada en 1735, puesto que se conserva su archivo y forma parte de la Bibliotecas Virtuales Larramendi<sup>51</sup>.

#### **4. Vinculación de datos**

La Biblioteca Virtual de Polígrafos participa en Europeana, a través de Hispana, como proveedor de datos y, como se ha dicho, utiliza su base de datos como fuente de información. Por ello, se ha planteado como objetivo estratégico la adopción del modelo propuesto por el Europeana Data Model (EDM)<sup>52</sup> y las tecnologías Linked Open Data. Tanto EDM como Linked Open Data están abriendo el campo de actuación de una forma velocísima y proporcionando ingentes cantidades de información como en el caso de VIAF<sup>53</sup> o la DBpedia<sup>54</sup>. Desde la propuesta de Tim Berners-Lee de Linked Data en 2006 hasta el momento actual se han publicado datos en estructuras semánticas y reutilizables a un ritmo como seguramente no se había producido en la Web anteriormente, probablemente con la salvedad de los registros bibliográficos en MARC 21.

Para todo lo relacionado con la aportación bibliotecaria a LOD conviene recordar que dentro del W3C se creó en 2010 el Library Linked Data Incubator Group<sup>55</sup> (LLD). Este grupo tiene como cometido de preparar diferentes informes sobre el estado de la cuestión de las ontologías para archivos, bibliotecas y museos, vocabularios y *datasets* de interés para estos ámbitos y el uso que la comunidad bibliotecaria está realizando con ellos. El seguimiento de la actividad de este grupo, así como de Linking Open Data<sup>56</sup>, informará convenientemente a cualquier interesado.

Para la Biblioteca Virtual de Polígrafos Linked Open Data (LOD), y las tecnologías anejas, representa la oportunidad de realizar plenamente el

---

<sup>51</sup> De hecho, fue un antecesor de Ignacio de Larramendi y Montiano (1921-2001) propulsor, entre otras muchas iniciativas de las Bibliotecas Virtuales Larramendi o de DIGIBÍS. Véase además, Hernando de Larramendi, Ignacio. *Mecenazgo cultural de Ignacio Hernando de Larramendi y Montiano : crónica y testimonios*. [<http://www.larramendi.es/i18n/consulta/registro.cmd?id=954>]

<sup>52</sup> Doerr, Martin, Gradmann, Stefan, Henniecke, Steffen, ISAAC, Antoine, Meghini, Carlo, and SOMPLE, Herbert van de. *The Europeana Data Model (EDM)*. En: *76<sup>th</sup> IFLA General Conference and Assembly*, 10-15 August, 2010, Gothenburg, Sweden. <http://goo.gl/7eLs>

<sup>53</sup> <http://ckan.net/package/viaf>

<sup>54</sup> <http://dbpedia.org/About>

<sup>55</sup> <http://www.w3.org/2005/Incubator/llid/>

<sup>56</sup> <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

planteamiento de los 'agregados digitales' como nodo de vínculos de recursos de información. De hecho, la Biblioteca Virtual de Polígrafos es un caso de uso analizado por el LLD<sup>57</sup>. Se trata ya no sólo de capturar información, sino de enlazarla a nodos apropiados de la nube LOD, conforme a unas reglas de juego en las que es esperable que cada cual cumpla su papel. Cada recurso disponible en un conjunto de datos LOD está identificado por una URI con especiales características<sup>58</sup>. Esta URI arrastra a su vez toda una red de relaciones, atributos, con otros recursos que definen, según el modelo que se esté utilizando (ontología o esquema RDF), las entidades y propiedades de un ámbito funcional determinado. Pero con ser esto importante, lo que verdaderamente añade valor es que permite identificar que una entidad en un entorno es la misma o muy similar a otra entidad definida en otro entorno. Así, Maimónides es un *filósofo* para la DBpedia y el *nombre de una persona* en VIAF. El procedimiento consiste en enlazar datos por medio de esas URIs y en disponer de una serie de herramientas<sup>59</sup> que permitan localizar esos datos, reutilizarlos, deducir unos datos de otros a partir de las relaciones y reglas de inferencia, actualizar los datos originales y servirlos nuevamente modificados y ampliados. Y así sucesivamente, de modo iterativo.

### VIAF y DBpedia

En 2010 la Biblioteca Virtual de Polígrafos inició el proceso de vincular los registros de autoridades con los recursos similares existentes en LOD. Obviamente, las fuentes de información señaladas en la metodología son aquí también las fuentes para la obtención de enlaces. Así, se ha consignado en el campo 024 de los registros MARC 21 la URI de las entidades de VIAF (Maimónides `rdf:about="viaf/100185495"`), que agrupan los encabezamientos establecidos para este autor procedentes de 7 ficheros de autoridad distintos, con sus 255 variantes de nombre, pero también enlazado al contenido de la DBpedia (<http://dbpedia.org/resource/Maimonides>). Si el bibliotecario tiene experiencia y conocimientos en la materia, es decir un catalogador senior, deducirá del entramado de relaciones de VIAF hasta 12 encabezamientos distintos, de otros tantos países y lenguas, correspondientes a la entidad Maimónides<sup>60</sup>.

Evidentemente, el proceso de adquisición de estos datos de forma manual es muy preciso, pero lento, y pronto se consideró la posibilidad de automatizarlo, valorando las ventajas y desventajas de hacerlo así. Indudablemente, el número de autores que conforman el proyecto bibliográfico de la BVP, en torno a 1.000, permitiría que se hiciera manualmente, pero la productividad en tiempo y en

---

<sup>57</sup> [http://www.w3.org/2005/Incubator/ld/wiki/Use\\_Case\\_Polymath\\_Virtual\\_Library](http://www.w3.org/2005/Incubator/ld/wiki/Use_Case_Polymath_Virtual_Library)

<sup>58</sup> <http://www.w3.org/TR/2007/WD-cooluris-20071217/>

<sup>59</sup> <http://www.w3.org/wiki/SemanticWebTools>

<sup>60</sup> Ibn Maymun, Musà o, el mucho más familiar para los judíos ortodoxos, Rambam

datos sería mucho mayor realizado de forma automatizada. Además, una vez diseñadas y desarrolladas las herramientas necesarias podrían ser incorporadas como funcionalidades específicas del ILS DIGIBIB. Se espera que en el momento de la lectura de esta comunicación, 13-18 de agosto de 2011, estas herramientas conformarán ya un módulo adicional en DIGIBIB.

### SKOS: otra vía a Linked Open Data

Las mismas vinculaciones se han llevado a cabo utilizando SKOS. El resultado es indicar que un determinado concepto es igual o muy similar a otros conceptos en otros vocabularios. A través de los encabezamientos de materia de los registros bibliográficos, y la vinculación de conceptos con otros vocabularios también disponibles en SKOS, es posible obtener un número también muy amplio de obras sobre un autor. Además, se puede automatizar el proceso de búsqueda de recursos de información, recorriendo la trama de relaciones entre conceptos de diferentes vocabularios, pudiéndose recuperar al mismo tiempo los recursos de información vinculados a esos conceptos.

En estos momentos, la Biblioteca Virtual de Polígrafos está trabajando, en pruebas, con la *Lista de Encabezamientos de Materia para Bibliotecas Públicas* que creó en su día la Dirección General del Libro y Bibliotecas y que posteriormente se ha enriquecido, cruzándola con el fichero de autoridades de la Biblioteca Nacional<sup>61</sup> y su equivalencia con los encabezamientos de materia de la Library of Congress<sup>62</sup>. Dado que este proceso es análogo realizado para vincular los conceptos SKOS de RAMEAU<sup>63</sup> con los LCSH y con los Schlagwortnormdatei (SWD)<sup>64</sup> será factible establecer esas relaciones conceptuales que, sin duda, no poseeran una granularidad demasiado grande, pero que, tratadas y cruzadas con un número de autores determinado sí podrán ofrecer una vinculación clara de conjuntos de información considerablemente amplios.

### Tiempo y espacio

Otros dos elementos juegan un papel esencial en la agregación y vinculación de los datos referidos a un autor como son los datos cronológicos y los geográficos. En efecto, un autor vive y muere en unas fechas determinadas y trabaja y se relaciona socialmente con sus contemporáneos durante determinados periodos, escribe libros que se publican, traducen y editan en fechas muy concretas<sup>65</sup>. La

---

<sup>61</sup> *Autoridades de la Biblioteca Nacional* [Recurso electrónico]/ Biblioteca Nacional ; software, Chadwyck Healey -- Número 1 (nov. 1996)- -- Madrid : Biblioteca Nacional : Chadwyck-healey España, 1996- CD-ROM.

<sup>62</sup> <http://id.loc.gov>

<sup>63</sup> <http://www.cs.vu.nl/STITCH/rameau/>

<sup>64</sup> <http://www.d-nb.de/standardisierung/normdateien/swd.htm>

<sup>65</sup> Existe ahora un nuevo formato que va a intentar homologar las muy diferentes maneras en que las diferentes civilizaciones han fijado las cronologías e incluso las cifras con las que las han representado,

posibilidad de interrelacionar y vincular a los autores mediante esta nueva categoría cronológica hace que ese conjunto de obras y autores vinculados por nombres y materias, queden ya considerablemente precisados también cronológicamente, si bien es seguro que la dispersión o la incoherencia será visible. La codificación del tiempo, presente en numerosos elementos de los registros de autoridad o bibliográficos, con frecuencia necesitan de calificativos o atributos para poder expresar la probabilidad, inexactitud, estimación, etc., tan frecuente en las descripciones de personas y de recursos, por no mencionar los eventos. Este tipo de datos es fundamental para la generación de líneas del tiempo o para la realización de operaciones y deducciones con fechas, uso que puede verse en WorldCat Identities<sup>66</sup>.

De igual manera se está trabajando en la agregación y vinculación de datos geográficos. De hecho, existe en Linked Open Data una importante iniciativa para este tipo de datos, GeoNames<sup>67</sup>, que además es el único caso conocido hasta la fecha que ha tenido su réplica en España, mediante el proceso de transformación de los repertorios geográficos del Instituto Geográfico Nacional a RDF y en LOD<sup>68</sup>. Nuevamente se produce el proceso de agregación y así nuestros autores han *nacido* y han *muerto* en determinado *lugar*, se han relacionado con sus contemporáneos en determinados *sitios*, tal vez escuelas, universidades, órdenes religiosas, ejércitos, sociedades científicas o culturales, grupos políticos, situados o relacionados con determinado *lugar*. Por lo tanto, y a través de ellos, se podrán establecer nuevamente las oportunas agregaciones y vinculaciones.

En el caso de GeoNames y GeoLinkedData el proceso se ha realizado de una forma automática, transformando los datos seleccionados en una estructura de información importable en DIGIBIB. Estos datos se han mapeado a MARC, con el objetivo añadido de cumplimentar automáticamente los campos 752 y 151. De esta forma se han obtenido no sólo los nombres disponibles en estos vocabularios sino también las coordenadas geográficas. Es indudable que debe existir posteriormente un trabajo manual para solucionar los problemas que puedan presentarse a la hora de diferenciar jurisdicciones de ubicaciones geográficas y de ubicación geográfica de jurisdicciones antiguas. Del mismo modo que, como se verá más adelante, una cosa es gestionar *personas* y otra *nombres de personas*, es evidente que gestionar jurisdicciones, nombres

---

cuando no mediante el abecedario, el alfabeto, el alifato, etc. Extended Time/Date Format ETDf  
[<http://www.loc.gov/standards/datetime/>]

<sup>66</sup> <http://www.worldcat.org/identities/>

<sup>67</sup> <http://www.geonames.org/ontology/documentation.html>

<sup>68</sup> Esta iniciativa se ha puesto en marcha con la publicación de diversas fuentes de información procedentes del Instituto Geográfico Nacional, haciéndolas disponibles como bases de conocimiento RDF conforme a los principios de Linked Data.

<http://geo.linkeddata.es/web/guest;jsessionid=6E6A9C1E1DAF40F81005E8F4DA3A27E1>

geográficos y ubicaciones geográficas obliga a disponer de estructuras de datos diferenciadas.

## 5. Uso de datos en LOD

En el sitio web de VIAF pueden realizarse manualmente búsquedas concretas y capturar los registros obtenidos en su versión MARC XML o RDF. A través del VIAF API<sup>69</sup> pueden realizarse búsquedas a través de SRU y de OpenSearch, obteniéndose los resultados en MARC XML, RDF y otra gran variedad de formatos y esquemas. Igualmente VIAF<sup>70</sup> está disponible como Linked Open Data. Siguiendo las recomendaciones para publicar datos<sup>71</sup> en LOD frecuentemente los *datasets* disponen de un servidor SPARQL Endpoint<sup>72</sup>, que permite realizar búsquedas en SPARQL<sup>73</sup> o bien pueden utilizarse clientes SPARQL<sup>74</sup> que permiten interrogar ficheros RDF. Esto quiere decir que VIAF o DBpedia<sup>75</sup> puede ser interrogados vía SPARQL y que por tanto las preguntas que pueden realizarse permiten hacer un barrido lógico contra esos datos y recuperar más información relacionada que de las maneras anteriormente expuestas. Por medio de SPARQL es posible definir una búsqueda mucho más compleja; pongamos un ejemplo posible en la DBpedia, pero no en VIAF, ‘filósofos influidos por Maimónides’. Por todo ello, la Biblioteca Virtual de Polígrafos está dando los pasos, repetimos, conforme a un plan bibliográfico determinado para aprovechar la información que se está vertiendo en la forma de *datasets* de LOD<sup>76</sup>. Y esto lo está haciendo tanto desde el punto de vista de consumidor de datos como desde el de productor de datos.

Así, el primer paso ha sido la aplicación del Europeana Data Model como respuesta a las transformaciones que Europeana tiene planificadas para su fase Danubio. Es posible que Europeana pueda realizar tratamientos más o menos

---

<sup>69</sup> <http://www.oclc.org/developer/documentation/virtual-international-authority-file-viaf/using-api>

<sup>70</sup> <http://ckan.net/package/viaf>

<sup>71</sup> *How to Publish Linked Data on the Web*. <http://www4.wiwiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

*Best Practice Recipes for Publishing RDF Vocabularies*. W3C Working Group Note 28 August 2008. <http://www.w3.org/TR/swbp-vocab-pub/>

<sup>72</sup> <http://www.w3.org/wiki/SparqlEndpoints>

<sup>73</sup> SPARQL Query Language for RDF. [<http://www.w3.org/TR/rdf-sparql-query/>]. En la actualidad el W3C está trabajando en SPARQL 1.1 Federated Query. *This specification defines the syntax and semantics of a SPARQL 1.1 Federated Query extension for executing queries distributed over different SPARQL endpoints*. [<http://www.w3.org/2009/sparql/docs/fed/service>]

<sup>74</sup> <http://www.w3.org/wiki/SparqlImplementations>.

<sup>75</sup> <http://ckan.net/package/dbpedia>

<sup>76</sup> <http://ckan.net/package>. Véase también la recopilación realizada por el LLD <http://ckan.net/group/lld> o DOIs as Linked Data <http://inkdroid.org/journal/2011/04/25/doi-as-linked-data/> o Linked Periodicals Data <http://periodicals.dataincubator.org/.html>



automatizados para transformar la información que recibe de los proveedores, al nuevo modelo de datos EDM, pero hay que considerar que estos son ya varios miles y que trabajan con más de 25 lenguas, lo que hace suponer que los tratamientos automáticos sólo podrán afectar a un porcentaje indeterminado de los datos, con unos resultados seguramente imposibles de verificar y con un gran riesgo de realizar equivalencias erróneas. Independientemente de que un procedimiento de este tipo pueda ser práctico y pueda resolver la vinculación entre una parte de los datos, no puede extenderse a la totalidad, ni puede mejorarse sin la intervención de los propios proveedores de datos. Situándonos en un caso típico, los nombres geográficos son muy a menudo nombres de jurisdicción que no se corresponden con las coordenadas geográficas actuales y así, igualar la Castilla del siglo XV a la Castilla actual es solamente una aproximación. En consecuencia, la Biblioteca Virtual de Polígrafos ha iniciado un proceso de publicación de sus datos vinculados internamente en LOD.

Transformar los registros bibliográficos MARC 21 con los que trabaja su ILS a Europeana Semantic Elements 3.3.1 y de ahí a Europeana Data Model presenta una dificultad relativa, pero si únicamente se realiza este proceso se ofrecerán datos a Europeana sin enlazar con fuentes externas o vocabularios comunes sobre los que Europeana pueda realizar una vinculación exacta de los datos. La metodología de incorporar las URIs de VIAF en los registros de autoridad se ha revelado muy útil puesto que en la generación de EDM puede ofrecerse una referencia a un vocabulario externo para la clase *Agent*. El segundo proceso de vinculación se ha realizado a través de los LCSH disponibles en *id.loc.gov*. Así, la Biblioteca Virtual de Polígrafos permite publicar sus encabezamientos de materia en SKOS, pero vinculados a LCSH. Por esta vía, la propiedad *dc:subject* de los recursos de la *Biblioteca Virtual de Polígrafos* se relacionará con *skos:Concept*, pero vinculados a LCSH. Esto permite indirectamente la vinculación con RAMEAU<sup>77</sup> y SWD<sup>78</sup>, como mínimo. Dicho de otro modo, los encabezamientos de materia accesibles en una estructura semántica y asociados a otros encabezamientos de materia iguales en inglés, francés y alemán. Es evidente la ventaja de la vinculación de diferentes *datasets* con uno o varios vocabularios determinados.

Además de esto, lo que no deja de ser paradójico, es que este mismo procedimiento puede valer para la interconexión de datos a nivel nacional. Uno de los efectos de que LCSH haya sido considerada como fuente principal de información para la elaboración de encabezamientos de materia en diferentes bibliotecas españolas es que a través de esa relación es posible realizar la vinculación entre las listas de encabezamientos de materia españolas que, no debe olvidarse, pueden estar presentes en varios idiomas. Este es un efecto colateral de LOD bastante beneficioso y, aunque sea dando un rodeo, permitirá interconectar y vincular datos entre los catálogos bibliográficos y de

---

<sup>77</sup> <http://ckan.net/package/stitch-rameau>

<sup>78</sup> <http://ckan.net/package/dnb-gemeinsame-normdatei>

autoridades españolas, que han permanecido bastante aislados e impermeables, mucho más de lo deseable y de lo comprensible.

También aquí tenemos que señalar otro efecto beneficioso de Europea en la mejora de los sistemas de información de un país. El Ministerio de Cultura español que mantiene Hispana ha iniciado también su proceso de adaptación a EDM y, como se ha dicho, una de las primeras actividades ha sido la transformación de la Lista de Encabezamientos de Materia para Bibliotecas Públicas (LEM) a SKOS, vinculando las entradas posibles con LCSH. Lógicamente, la Biblioteca Virtual de Polígrafos completará su proceso de vinculación de datos enlazando sus encabezamientos de materia con las LEM. Es de esperar que a través de este dataset pueda realizarse también la vinculación con encabezamientos de materia en gallego, catalán y vasco. A fecha 30 de abril de 2011 la Biblioteca Virtual de Patrimonio Bibliográfico, implementada sobre el ILS DIGIBIB 7.0 ya permite la exportación de registros conforme a EDM 5.2.1.

Estos procesos confluirán tanto en la aplicación completa de EDM como en la publicación de los *datasets* de la Biblioteca Virtual de Polígrafos en LOD. En este sentido, la publicación de datos en LOD deberá ser pausada correspondiéndose con el avance del proceso bibliográfico de modo que se publique en LOD aquellos datos que estén establecidos con un grado razonable de seguridad y suficientemente vinculados<sup>79</sup>. Del mismo modo que el proyecto bibliográfico de la Biblioteca Virtual de Polígrafos conlleva una selección crítica de fuentes, el proyecto biblioteconómico de su publicación en LOD implica una selección similar de *datasets*. No puede haber un entorno mejor que este Congreso de IFLA para recordar que los árboles de esquemas de metadatos, ontologías y alineamiento de ontologías, no pueden hacernos perder la visión del bosque bibliotecario, de los objetivos bibliográficos del sistema de información que se va a publicar en LOD.

## **6. Procesos automáticos**

Como consumidora de información la Biblioteca Virtual de Polígrafos está diseñando diferentes aplicaciones para la captura, integración y difusión de datos. Ya hemos mencionado la reutilización de datos de GeoNames o de LEM. Además, en el momento en el que se redacta esta comunicación se están diseñando otras aplicaciones para la captura de *datasets* de LOD por medio de procedimientos semiautomatizados. Estos procesos pueden implicar la captura de tripletes relacionadas con uno o varios autores o la obtención únicamente de la URI del recurso. Estos datos se almacenan en un repositorio intermedio sobre el cual se pueden realizar operaciones de búsqueda, selección y modificación, de modo que una vez contrastado puedan incorporarse a la base de datos de autoridades o bibliográfica, siendo posible elegir la subida de los atributos correspondientes a una entidad o únicamente su URI y actualizar los registros base.

---

<sup>79</sup> Weibel, Stuart. *Principles of Linked Data Recast*. Weibel Lines [<http://goo.gl/sYSDe>]

El mismo procedimiento se realizará con las descripciones de recursos bibliográficos y de información. Este proceso se iniciará próximamente ya que presenta el problema añadido de la correcta identificación de obras. Si la identificación de nombres de personas en los ficheros de autoridades ha heredado una serie de problemas como la falta de datos suficientes para la identificación, en el caso de las obras esta falta de datos es aún mayor. Así la falta de títulos uniformes y entradas de autor-título uniforme y la ausencia de relación entre los títulos de las obras y sus traducciones hace especialmente complicada la realización de procesos automatizados o semiautomatizados.

Por otra parte, no todas las fuentes de datos de autoridad o bibliográficas se encuentran disponibles en LOD, panorama que está cambiando día a día, por lo que la Biblioteca Virtual de Polígrafos hace uso de servidores SRU sobre los cuáles puede lanzar determinados perfiles de búsqueda. Es posible configurar estos perfiles de búsqueda de modo que puedan variarse las preguntas en función del o los elementos buscados. La ventaja del diseño que se ha realizado es que los resultados obtenidos se pueden almacenar en un repositorio OAI que realiza funciones de recolección periódica a partir de los resultados de los servidores SRU. No se nos escapan las grandes posibilidades que este sistema puede proporcionar para crear un servicio de recolección a nivel nacional o, incluso, a nivel internacional. El tremendo progreso de Europeana hasta la fecha ha estado basado en la recolección de metadatos por el protocolo OAI y que un sistema similar, que bien pudiera denominarse *Americanae*, puede iniciarse por este procedimiento. La principal ventaja de este procedimiento sería la recolección selectiva de metadatos y esperamos que obtenga muy buenos resultados tanto en recursos españoles como internacionales.

## **7. Enciclopedia de autoridades**

Digamos que con este modelo se ha pasado de los registros de autoridad de nombres de persona a los registros de personas, lo cual no es una cuestión nominal, sino de largo alcance. De hecho, algunos de los problemas de catalogación que habrá que resolver próximamente será el de conjugar los atributos propios de las *personas* con los atributos propios de los *nombres de personas*. Un vistazo a algunas ontologías y *datasets* como DBpedia o VIAF nos muestran las diferentes definiciones de clases y atributos que ambas realizan para establecer los nombres de las personas.

Si el centro del diagrama de la nube Linked Open Data está la DBpedia, es decir la Wikipedia, parecería normal que la presentación de nuestras fichas enciclopédicas de autor recordasen o tuvieran una estructura similar a la de la Wikipedia, pero utilizando el formato MARC como base. Despreciar la inmensa cantidad (miles de millones) de registros MARC existentes en el mundo sería absurdo. Por un lado, se cuenta con la extraordinaria granularidad y versatilidad de MARC 21 que puede convertirse fácilmente en estructuras XML. Es posible, desde luego, que cualquier sistema pueda obtener los datos necesarios de cualquier *dataset* disponible en LOD, pero no debe olvidarse que se trata no sólo de enlazar y vincular datos, sino también de mantenerlos. En

este sentido las estructuras de DBpedia, VIAF, MADS, MARC 21 y EAC-CPF pueden acomodarse de modo que pueda efectuarse un *mashup* de información, pero esto puede ser insuficiente, puesto que la misma vinculación de datos puede dar lugar a la obtención de datos contradictorios, erróneos, inexactos o redundantes. Es decir, es necesaria una nueva estructura de datos para los registros de personas y nombres de personas destinada a la gestión y mantenimiento de datos.

La estructura de las entradas de autoridad podría ser una combinación de esquemas con base en el formato MARC, que compartiera además el análisis de MADS/RDF, EAC-CPF<sup>80</sup> e incluyera algunos de los atributos de DBpedia<sup>81</sup>. Es muy posible que la influencia que está ejerciendo la DBpedia haga necesario ampliar el número de atributos que definen a una persona. En un somero vistazo a la ontología de DBpedia observamos algunas propiedades de la clase persona, de verdadero interés para la Biblioteca Virtual de Polígrafos como 'influenced', 'influencedBy' o 'philosophicalSchool'.

Este aspecto puede apreciarse mucho mejor en el subproyecto de la Biblioteca Virtual de Polígrafos denominado Escuela de Salamanca. Bajo esta denominación colectiva se agrupa un conjunto de profesores universitarios españoles y portugueses (en su mayoría teólogos, juristas y economistas) que participaron de manera directa en el renacimiento del pensamiento que se produjo en el siglo XVI y principios del XVII, a raíz del descubrimiento y la actuación en el Nuevo Mundo, con raíz en la labor intelectual y pedagógica de Francisco de Vitoria, por citar un caso (y un nombre eminente), en la Universidad de Salamanca. El objetivo último no es otro que seguir la influencia de estos precursores de algunas ramas del derecho, de la política y especialmente de la economía<sup>82</sup>. La Edad Moderna supuso un cambio importante en el concepto del hombre en sociedad y es justamente la Escuela de Salamanca la que aborda estos problemas desde nuevos planteamientos, así, Francisco de Vitoria, Domingo de Soto, Martín de Azpilcueta, Tomás de Mercado o Francisco Suárez, intentaron reconciliar la doctrina tomista con el nuevo orden social y económico. Así, nos encontramos un grupo de autores que

---

<sup>80</sup> Aunque se ha obviado a lo largo de esta comunicación, no podemos dejar de mencionar la reciente publicación de CIDOC-ICOM Linked Open Data Recommendation for Museums. [<http://www.cidoc-crm.org/URIs and Linked Open Data.html>]

<sup>81</sup> <http://mappings.dbpedia.org/server/ontology/classes/Person>

<sup>82</sup> Agenjo Bullón, Xavier; Hernández Carrascal, Francisca y Juez García, Patricia: La Escuela de Salamanca considerada desde el punto de vista de la Web semántica y la información en la red. Ponencia presentada el día 14 de abril de 2011 en las *X Jornadas de la Asociación de Hispanismo Filosófico: Crisis de la modernidad y filosofías ibéricas*, celebradas en Santiago de Compostela los días 13-15 de abril de 2011. [<http://goo.gl/uFrXx>]

dieron origen a la ciencia económica<sup>83</sup>, ya que, preocupados por la legitimidad moral de la conquista, desarrollaron teorías muy novedosas<sup>84</sup>.

Si se recapitula veremos que los autores de la Biblioteca Virtual de Polígrafos, escritores de varios miles de obras, que habíamos podido localizar en multitud de bibliotecas, utilizando al VIAF como procedimiento de vincular las obras con sus autores y que más tarde habíamos ampliado, estableciendo una navegación virtual entre las materias de esos libros, utilizando vocabularios o listas de encabezamientos como las LEM o los LCSH, las RAMEU o los SWD, están ahora interrelacionados y agregados espacial y cronológicamente, formando una red fácilmente visible. Se ha establecido una vinculación enorme a través de los datos y, sobre todo, de los metadatos referidos, no tanto al *nombre de la persona* como a la *persona* misma y, qué duda cabe, que al igual que se tratan *personas* pueden tratarse *corporaciones*. Deberemos sumar ahora las especificidades de la *obra* porque en una gran mayoría de los casos, la *obra* a través de sus distintas *expresiones* se habrá *manifestado* gracias a *impresores* y *editores*, *libreros* incluso (y *bibliotecas* donde se conservarán los *item* o *ejemplares*) y todas esas entidades tendrán así mismo su representación tanto en el espacio como en el tiempo. Así, y siguiendo con la Escuela de Salamanca, veremos hasta qué punto el Convento de San Esteban (40° 57' 38"N 5° 39' 47" O) jugó un papel esencial para docentes y discentes, e incluso como editorial e imprenta, labor, que por cierto mantiene hasta el presente. Aquel término procedente de la época de las escuelas de documentación, las capillas invisibles, se tornan claramente visibles.

### Redefinición de funcionalidades de los ILS

Desde luego, las posibilidades que ofrece y ofrecerá la vinculación de datos va a llevar a una completa redefinición de las funcionalidades de los ILS. En primer lugar, como se ha mencionado, para permitir la incorporación de elementos de información o atributos no existentes en el formato MARC para su gestión y exportación como LOD. En segundo lugar, permitir la selección de conjuntos de datos completos o seleccionados procedentes de fuentes externas; lo que significará disponer de la capacidad de actualizar total o parcialmente un registro. Por ejemplo, es posible, como ocurre en la Biblioteca Virtual de Polígrafos que únicamente se quiera seleccionar la URI de un determinado recurso. En tercer lugar, será necesario modificar las interfaces de consulta permitiendo la recuperación y mash-up en línea de otras fuentes a través de APIs basadas en SRU/OpenSearch o incluso realizando combinaciones con SPARQL. En último lugar de esta lista, no exhaustiva, deberá ser posible gestionar otro tipo de informaciones no muy utilizadas como las propias descripciones de los datasets de LOD.

---

<sup>83</sup> Schumpeter Schumpeter, Joseph A. *History of Economic Analysis*. New York: Oxford University Press, 1954.

<sup>84</sup> Grice-Hutchinson, Marjorie. *The School of Salamanca: Readings in Spanish Monetary Theory, 1544-1605*. Oxford: Clarendon Press, 1952.

## 8. Disponibilidad de datasets

Debe tenerse en cuenta que además de la disponibilidad de los *datasets* es muy importante, como siempre lo ha sido, la selección crítica de estos *datasets*, que a fin de cuentas no son otra cosa que recursos de información bibliográfica con una nueva estructura; el análisis crítico de la calidad de los datos desde el punto de vista intelectual y teórico permanece. Aparte de lo que se requiera para registrar datasets según CKAN<sup>85</sup> o VOID<sup>86</sup>, es evidente que es necesario disponer de información acerca, no sólo de la ontología o estructura de datos, sino también de la frecuencia de actualización, del número de entradas y relaciones, de la periodicidad de actualización, así como de las fuentes utilizadas para realizar cualquier fusión de datos. Podemos decir, que seguramente la experiencia más inmediata de la vinculación de datos con fuentes externas, como por otra parte es ya muy sabido por los bibliotecarios y bibliógrafos, está en la variedad de aserciones sobre una misma *cosa* que, como hemos dicho, a veces son inexactas, otras erróneas, redundantes y en otras se carece de garantía de calidad acerca de la fuente.

La web semántica y LOD proponen una mayor exactitud en las navegaciones, pero se puede apreciar que, aún habiendo en la actualidad pocas fuentes disponibles en LOD, es fácil encontrar *cosas* definidas de diferentes maneras, con muy distintos atributos y que la vinculación *per se* no contribuirá a mejorar la exactitud de la navegación. Debe recordarse en este punto lo que se ha dicho sobre la necesidad de transformar en datasets de LOD una enorme cantidad de fuentes bibliográficas de primera calidad que permanecen aún sin digitalizar y sin estructurar por no ser aún de dominio público. Consideramos que esta es una responsabilidad que bibliotecas y bibliotecarios no pueden desdeñar y que proyectos masivos de digitalización que convergen en Europeana e Hispana van realizando a buen ritmo. *Das Ding im sich, das ding am sich y das ding bei sich*, como se planteaba el idealismo alemán del siglo XIX.

## 9. Conclusiones

No se nos oculta que todo el trabajo ha gravitado en torno a dos proyectos que cuando se cierra la revisión última de este trabajo, el 30 de abril de 2011, no pueden aún conocerse. Por un lado, la puesta en marcha de la fase Danubio de Europeana y el resultado de la aplicación experimental de las RDA en un amplio conjunto de bibliotecas norteamericanas. Si a esto le añadimos que las conclusiones del grupo de trabajo W3C Library Linked Data Incubator Group no están aún disponibles, es obvio que nuestros resultados no son más que un conjunto de acciones realizadas en paralelo al desarrollo e implantación de este nuevo paradigma. Quizá ahí radique el mayor interés de nuestra aportación, puesto que el proyecto de la Biblioteca Virtual de Polígrafos ha llevado a cabo acciones similares a los de la implantación de RDA y participa activamente en

---

<sup>85</sup> Comprehensive Knowledge Archive Network. [<http://ckan.net/package/new>]

<sup>86</sup> *Describing Linked Datasets with the VOID Vocabulary* [<http://www.w3.org/TR/void/>]

Europeana (y es un caso de estudio del LLD). Por lo tanto nuestra propuesta de trabajo real, accesible y disponible a través de la Web para quien quiera consultarla, puede considerarse como una aportación más para conseguir que la información (y no sólo la información bibliográfica) alcance la máxima accesibilidad y, sobre todo, cumpla los objetivos propuestos por Tim Berners-Lee cuando enunció Linked Open Data. Este es el sentido último de esta comunicación.