

Linked Open Data y DIGIARCH. La aplicación de Europeana Data Model

Fecha de envío: 30 de octubre de 2017

Autores:

Xavier Agenjo Bullón orcid.org/0000-0001-8338-8087	Fundación Ignacio Larramendi xavier.agenjo@larramendi.es
Francisca Hernández Carrascal orcid.org/0000-0002-2389-0945	DIGIBÍS francisca.hernandez@digibis.com
Susana Hernández Rubio orcid.org/0000-0001-8940-1713	DIGIBÍS susana.hernandez@digibis.com
Montserrat Martínez Guerra orcid.org/0000-0002-9943-8103	DIGIBÍS montse.martinez@digibis.com
Maribel Campillejo Suárez orcid.org/0000-0002-9283-0188	DIGIBÍS maribel.campillejo@digibis.com
Fernando Román Ortega orcid.org/0000-0002-8619-1674	DIGIBÍS fernando.roman@digibis.com

Resumen:

DIGIBÍS ha apostado por la tecnología *Linked Open Data*ⁱ desde el año 2000. De hecho, el año 2004, dentro del *Programa para el Fomento de la Investigación Técnica* (PROFIT) del Ministerio de Industria, Energía y Competitividad, la Fundación Ignacio Larramendi y DIGIBÍS presentaron un proyecto llamado *Biblioteca Virtual de Polígrafos y la Web semántica*. Siguiendo esta estela, el programa DIGIBÍS ha incorporado la tecnología *Linked Open Data* a su proyecto empresarial estableciendo vínculos con los vocabularios de valores más importantes, siguiendo las recomendaciones del W3C-LLD (*World Wide Web Consortium Library Linked Data Incubator Group*)ⁱⁱ y, sobre todo, publicando los registros conforme a la ontología del *Europeana Data Model* (EDM)ⁱⁱⁱ. En 2015, la Asociación Española de Normalización y Certificación, AENOR, ha reconocido nuevamente la labor de innovación tecnológica de DIGIBÍS^{iv}; en esta ocasión la certificación en I+D+i ha sido para el proyecto "*Archivos Digitales y Datos Abiertos*".

Esta estrategia ha dado muy buenos frutos puesto que una gran cantidad de bibliotecas virtuales y digitales españolas, implementadas con el programa DIGIBÍS, aportan contenidos, primero a *Hispana* y después a *Europeana* con esas características semánticas. De hecho, a fecha de 31 de octubre de 2017, en *Europeana*^v el agregador nacional *Hispana* ofrece 2.732.959 de objetos digitales, de los que una parte muy importante está basada (66,53 %) en el software DIGIBÍS.

Esta misma estrategia se ha seguido en DIGIARCH^{vi}, el Sistema de Gestión Digital Archivística. Para ello, DIGIARCH incorpora la posibilidad de exportar los registros archivísticos conforme al formato EAD3 (*Encoded Archival Description*)^{vii}, de forma pionera en España, y publica la información en este esquema de metadatos, pero también en RDF conforme a la ontología EDM. Internamente, se ha mapeado la información para que EAD3 actúe de forma análoga a MARC 21 en el mundo

bibliotecario, como formato común de intercambio (al igual que LIDO, *Lightweight Information Describing Objects*^{viii} para los museos).

Así, los registros de las descripciones archivísticas se publican tanto en la Web como en el repositorio OAI-PMH y, por supuesto, en Linked Open Data, tanto en EAD3 como en EDM. De hecho, las distintas instalaciones que corren sobre DIGIARCH como Galiciana, Archivo Dixital de Galicia^{ix}, Ciconia (que reunirá próximamente los fondos de los archivos históricos provinciales de Badajoz y de Cáceres), o el Archivo de la Fundación Norman Foster^x, pueden recolectarse en EAD3 y EDM a través de las correspondientes instancias del programa DIGIARCH. En los dos casos, DIGIBIB y DIGIARCH, publican la información, tanto de forma individualizada como a través del repositorio OAI-PMH (*Open Archive Initiative-Protocol for Metadata Harvesting*) de forma transparente. De hecho, también el programa DIGIMÚS cuenta con las mismas funcionalidades.

No sólo Europea, sino la Web en general y la Web Semántica en particular están facilitando, y presionando, a los archivos, bibliotecas y museos a abrir, publicar y conectar su información. Estos procesos han reavivado el papel de los archivos como receptores y creadores de una inmensa cantidad de datos, documentos y procedimientos, lo cual puede verse en la activa participación de los archiveros en el análisis, desarrollo e implantación de las políticas de gestión de los documentos electrónicos. La combinación de las leyes de transparencia^{xi} y de reutilización de la información del sector público^{xii} ha dado lugar a la publicación de grandes volúmenes de datos en distintos sectores muy útiles para los archivos, y viceversa. Ejemplo de ello son el Catálogo Nacional de Datos Abiertos, datos.gob.es^{xiii}, (2010), que ofrece en la actualidad 16.040 conjuntos de datos distintos, o los catálogos de las Comunidades Autónomas españolas y de los ayuntamientos^{xiv}, que publican diferentes conjuntos de datos abiertos^{xv}. Los archivos no están bien representados aún, aunque las propias descripciones de archivos, bibliotecas y museos constituyen una de las fuentes más importantes para la difusión de la información del sector público (preámbulo de la Ley 18/2015).

El nuevo entorno en el que trabajan los archivos está caracterizado por el desarrollo de la Administración Electrónica; por los propios documentos y expedientes electrónicos; por la aplicación de las políticas europeas y españolas sobre digitalización y preservación digital; y por la obligatoriedad de producir, intercambiar y reutilizar datos, especialmente los procedentes del sector público. La fuerza de estos requisitos no se encuentra solo en su conveniencia técnica o tecnológica, sino, esencialmente, en que deben ser la respuesta obligada a las prescripciones legales, normativas y reguladoras emanadas de las instituciones políticas y administrativas de España y de la Unión Europea, como las estrategias Europa 2020^{xvi} o el Mercado Único Digital^{xvii} y las diferentes Directivas europeas sobre Administración Electrónica y Gobierno Abierto, que junto a la legislación española (Ley 19/2013, de transparencia, acceso a la información pública y buen gobierno, la Ley 18/2015 sobre reutilización de la información del sector público, la Ley 39/2015 del Procedimiento Administrativo Común de las Administraciones Públicas y la Ley 40/2015 de Régimen Jurídico del Sector Público y sus desarrollos complementarios), definen que las tecnologías de la información y las comunicaciones deben aplicarse de forma general, para mantener la interoperabilidad de servicios, datos y aplicaciones.

La incorporación de la tecnología Linked Open Data a DIGIARCH permite publicar los datos de un sistema de archivos de forma abierta y en RDF conforme a la ontología EDM, por lo que los sistemas construidos sobre DIGIARCH cumplen con las 5 estrellas Linked Open Data^{xviii}, con la Norma Técnica de Interoperabilidad para la

Reutilización de Recursos de Información^{xix}, y con el Goldbook del Portal Europeo de Datos^{xx}. Así, la combinación de EAD3 y EDM hace posible que un sistema de gestión de archivos combine el tratamiento de la documentación histórica tradicional y los documentos digitales, y estar en disposición de contribuir con sus descripciones normalizadas a Hispana, a Europeana, o al Portal de Archivos Europeo; pero también al Catálogo de datos de datos.gob.es, o al Portal de Datos Europeos, y estar presentes en la Web Semántica, produciendo y consumiendo datos abiertos vinculados.

De forma totalmente transparente el sistema archivístico de Galicia o el Archivo de la Fundación Norman Foster (los archivos de Extremadura lo harán en breve), ya están incorporados a la tecnología *Linked Open Data*, y por ello a la *Web semántica*, dándole al maravilloso patrimonio histórico que conservan los archivos el realce que merecen. Todo ello, cumpliendo además, las recomendaciones de la Comisión Europea recogidas en el informe *Digitisation, online accessibility and digital preservation* (junio 2016), puesto que esa información recolectada, primero por los agregadores de las comunidades autónomas, serán a su vez recolectados por *Hispana* y desde ahí a *Europeana* proporcionando la visibilidad que la Comisión y los ciudadanos del mundo requieren.

ⁱ W3C Guía Breve de Linked Data. <https://www.w3c.es/Divulgacion/GuiasBreves/LinkedData>

ⁱⁱ Documentación Técnica del W3C-LLD

http://www.larramendi.es/es/cms/elemento.do?id=estaticos%2Fpaginas%2Fdocumentacion_tecnica_la_m.html

ⁱⁱⁱ Definition of the Europeana Data Model v5.2.8. 06/10/2017

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

^{iv} <http://www.digibis.com/sobre-nosotros/i-d-i/certificados-aenor-proyectos-i-d-i-digibis.html>

^v <https://www.europeana.eu/portal/es/explore/sources.html>

^{vi} <http://www.digibis.com/software/digiarch>

^{vii} <http://www.loc.gov/ead/EAD3taglib/index.html>

^{viii} <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>

^{ix} <http://archivo.galiciana.gal/arpadweb/gl/inicio/presentacion.cmd>

^x <http://archive.normanfosterfoundation.org/en/inicio/presentacion.do>

^{xi} Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.

BOE número 295, de 10/12/2013. <https://www.boe.es/buscar/act.php?id=BOE-A-2013-12887>

^{xii} Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. BOE número 164, de 10 de julio de 2015.

https://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-7731

^{xiii} <http://datos.gob.es>

^{xiv} UNE 178301:2015. Ciudades Inteligentes. Datos Abiertos (Open Data). (Versión corregida en fecha 2015-07-29).

<http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0054318#.Vx5mvPI9670>

^{xv} <http://datos.gob.es/es/iniciativas>

^{xvi} https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester/framework/europe-2020-strategy_es

^{xvii} https://ec.europa.eu/growth/single-market_es

^{xviii} <http://5stardata.info/es/>

^{xix} https://www.boe.es/diario_boe/txt.php?id=BOE-A-2013-2380

^{xx} <https://www.europeandataportal.eu/es/providing-data/goldbook>