

Web semántica. Informe de situación 2014

Semantic web. Situation report 2014

Juan-Antonio Pastor-Sánchez y Tomás Saorín

Pastor-Sánchez, Juan-Antonio; Saorín, Tomás (2015). "Web semántica. Informe de situación 2014". *Informes ThinkEPI 2015 sobre documentación y comunicación*, v. 1, pp. 177-188.

<http://dx.doi.org/10.3145/info.2015.12>



Juan-Antonio Pastor-Sánchez es doctor en documentación y profesor de la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia*, en el área de la construcción de servicios y sistemas de información digital. Ha desempeñado su carrera profesional como documentalista y diseñador de sistemas de información, creación de entornos de enseñanza en red y diseño web. Sus líneas de investigación se centran en la aplicación de las tecnologías de la web semántica, *linked open data*, diseño de ontologías, gestión de contenidos digitales y arquitectura de la información. Desempeña el cargo de coordinador de bibliotecas de la *Universidad de Murcia*.

<http://orcid.org/0000-0002-1677-1059>

pastor@um.es



Tomás Saorín, profesor asociado de la *Facultad de Comunicación y Documentación* de la *Universidad de Murcia*, trabaja como documentalista en la *Oficina de Transparencia y Participación Ciudadana de la Administración Pública* de la Región de Murcia. Ha trabajado en la *Biblioteca Regional de Murcia*, en el proyecto *Pregunte, las bibliotecas responden*, en el *Centro de Documentación* de la *Consejería de Trabajo y Política Social*, y en el *Servicio Regional de Empleo y Formación*. Ha actuado como asesor en el proyecto web del *Premio Mandarache de Jóvenes Lectores* de Cartagena y la comunidad sobre alfabetización informacional *AlfaRed* del *Ministerio de Cultura*, y ha sido consultor en

gestión de contenidos digitales en la *Universitat Oberta de Catalunya*. Participa en el capítulo español de *Wikimedia* para el conocimiento libre y en acciones de divulgación del movimiento *GLAM-Wiki* para bibliotecas, archivos y museos. Ha sido presidente del comité científico de las jornadas *Fesabid 2013*.

<http://orcid.org/0000-0001-9448-0866>

tsp@um.es

Universidad de Murcia, Facultad de Comunicación y Documentación
Campus Universitario de Espinardo. 30100 Murcia, España

Resumen

Este informe muestra la actividad relevante más reciente en el ámbito de la web semántica. Se analizan los aspectos más novedosos de la nueva recomendación RDF 1.1. También se abordan los aspectos más destacables del primer borrador público del documento del grupo de buenas prácticas para la publicación de datos en la web. Se analiza la evolución del marcado semántico, tanto desde el punto de vista de los estándares, como de las aplicaciones prácticas en la Web. El informe identifica a *Wikipedia* como un factor dinamizador de la web semántica, en especial a través de las iniciativas *DBpedia* y *WikiData*. Se concluye reflexionando sobre las propuestas estratégicas de *Europeana* y la *Digital Public Library of America* como claros referentes en la descripción de recursos digitales en red.

Palabras clave

Web semántica; RDF 1.1; Buenas prácticas; Datos en la Web; Mercado semántico; SKOS; *Wikipedia*; *DBpedia*; *WikiData*; *Europeana*; *Digital Public Library of America*.

Abstract

This report highlights the most important recent activity in the field of semantic web, analyzing the most innovative aspects of the new RDF 1.1 recommendation, and discussing the first public working draft of *Data on the web best practices*. The evolution of semantic markup is analyzed, both from the point of view of standards and the practical applications on the Web. *Wikipedia* is identified as a dynamic factor in the semantic web, especially through the *DBpedia* and *WikiData* initiatives. *Europeana* and the *Digital Public Library of America* are clear references in the description of networked digital resources.

Keywords

Semantic web; RDF 1.1; Data on the web; Best practices; Semantic markup; SKOS; *Wikipedia*; *DBpedia*; *WikiData*; *Europeana*; *Digital Public Library of America*.

1. Introducción

En los últimos 10 años se ha desarrollado un gran número de tecnologías bajo la denominación común de *web semántica*. La acumulación de datos, ontologías, tecnologías, proyectos, parece ir alcanzando la masa crítica para producir un ecosistema suficiente para producir el valor de nuevo cuño que se le presupone a esta evolución orgánica de la Web.

Sin embargo, la adopción de innovaciones tecnológicas es un proceso complejo lleno de interdependencias y equilibrios (**Saorín; Peset; Ferrer-Sapena**, 2013). En 2011 la web semántica se encontraba en el pico máximo de sobreexpectación según el análisis de tecnologías *Hype cycle for Web computing* que realiza anualmente *Gartner*, y establecía un plazo de 10 años para su adopción masiva. En el informe de 2014, las tecnologías de la web semántica, han entrado en lo que la consultora denomina el “abismo de la desilusión”. En esta fase se decantan los actores, proyectos y tecnologías, pero a partir de la misma se puede ir a una nueva fase de crecimiento, hacia una “Rampa de consolidación” orientada hacia unos objetivos de negocio bien centrados y convertirse en una tecnología productiva.

En este informe hemos escogido unos pocos elementos que consideramos significativos para el asentamiento de las tecnologías de la web semántica, y que pueden tener un efecto tractor sobre otros factores.

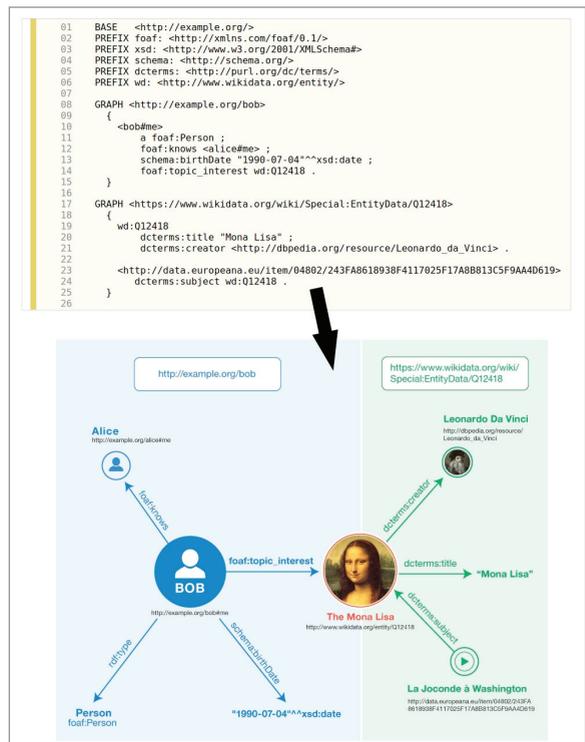
2. RDF revisitado: de triplets a quads

Durante 2014 hemos asistido a la publicación de la recomendación RDF 1.1 del W3C. Este hecho resulta de gran importancia, puesto que RDF conforma el núcleo de desarrollo de las tecnologías de la web semántica. RDF 1.1 no conlleva cambios sustanciales, sino que amplía y define determinados aspectos conceptuales y formales que cubren ciertas necesidades identificadas durante los últimos años. Las principales novedades que aporta RDF 1.1 (WC3, 2014a) podrían resumirse en los siguientes puntos:

- Se adopta el uso de IRIs (*internationalized resource identifiers*) para referenciar recursos sin limitarse al uso de caracteres ascii.
- Concreta de forma detallada la estructura de los nodos en blanco, como elementos no persistentes o con significado únicamente en el entorno de un conjunto de datos específico, lo que hace imposible su referencia desde otros conjuntos de datos externos. Para facilitar la interoperabilidad se recomienda normalizar las referencias definiendo en la medida de lo posible una IRI para cada recurso del conjunto de datos¹.
- Mejora la especificación de los literales y se definen nuevos tipos de datos que permiten especificar de un modo más detallado la semántica de los datos: cadenas literales expresadas en un idioma determinado y algunos tipos para expresar datos referidos a duración de intervalos temporales.
- Se actualiza la familia de formatos para expresar conjuntos de datos. En RDF 1.1 se distingue entre formatos basados en Turtle (N-Triples, Turtle, TriG y N-Quads), JSON-LD, RDFa y RDF/XML.

Sin duda la publicación de gran número de conjuntos de datos *linked open data* ha hecho preciso profundizar en el desarrollo de RDF. Resultan de gran interés las novedades apuntadas anteriormente, puesto que constituyen un indicador de la madurez de las tecnologías de la web semántica. Esto indica que han comenzado ciclos de revisiones enfocadas a adaptarlas a necesidades demandadas por los procesos de publicación de datos.

Más concretamente, durante los últimos años ha proliferado la publicación de conjuntos de datos estructurados, en donde se mezclaban elementos de diferente na-



Sentencias RDF en formato TriG y grafos resultantes. Puede observarse la agrupación de sentencias RDF en diferentes grafos.

<http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225>

turalidad en un mismo fichero. Esto ha hecho evidente la necesidad de utilizar una macroestructura que permita agrupar diferentes tipos de datos al tiempo que permita distinguirlos fácilmente. RDF ha evolucionado en este sentido, puesto que la nueva especificación permite agrupar las tripletas en torno a grafos. Estos grafos también se referencian mediante IRIs y así las sentencias RDF pueden contener cuatro elementos, en vez de los tres elementos de la especificación original. De hecho, Sparql ya contempló en su momento el uso de grafos como un mecanismo que permitía almacenar de forma separada los datos en un mismo *triplestore* para su posterior recuperación.

“La publicación de gran número de conjuntos de datos *linked open data* ha hecho preciso profundizar en el desarrollo de RDF”

Este nuevo enfoque precisa de formatos adecuados. Tal es el caso de N-Quads (W3C, 2014b) y TriG (W3C, 2014c) que amplían N-Triples y Turtle respectivamente. La utilidad de estos formatos

resulta evidente por ser mucho más compactos y adecuados para un escenario de consumo masivo de datos semánticos. No hay que olvidar que RDF/XML resulta considerablemente más verboso y además precisa de una carga computacional para su procesamiento mucho mayor que el requerido por otros formatos como TriG. La recomendación de JSON-LD (W3C, 2014d) supone dotar de un formato estándar para el consumo de datos por parte de aplicaciones web cuyo procesamiento se realiza mediante el motor Javascript en la parte del navegador.

3. Buenas prácticas, vocabularios y ontologías para la publicación de datos abiertos enlazados

Parece que poco a poco se van definiendo de un modo más preciso los principios originarios de los *linked open data* enunciados en 2006 por Berners-Lee. Resulta muy relevante el borrador del trabajo del W3C sobre buenas prácticas para la publicación de datos en la web (W3C, 2015a)². Este documento recomienda 27 prácticas agrupadas en ocho secciones que podrían resumirse del siguiente modo:

1. Metadatos: se debe proporcionar información sobre el contenido, estructura, actualización y licencias de uso de los datos. Para ello han de usarse estándares y formatos adecuados, facilitando su descubrimiento, acceso y reutilización y por supuesto todo lo anterior indicando la procedencia y asegurando la calidad de los datos.
2. Identificación de los datos: esto conlleva el uso de URIs perdurables en el tiempo para referenciar adecuadamente conjuntos de datos (incluso una URI para cada versión). En cualquier caso es deseable que dichas referencias puedan resolverse para acceder a los datos a través del protocolo HTTP (de-referenciación).
3. Formatos de datos: facilitar los datos en múltiples formatos estándares abiertos para facilitar su reutilización, sin las limitaciones de los formatos propietarios y reduciendo los costes que implican los procesos de transformación de datos.
4. Vocabularios de datos: los esquemas o modelos de datos (y sus diferentes versiones) utilizados para representar los datos deben estar correctamente documentados y compartidos entre la comunidad y con un correcto nivel de formalización semántica, procurando su reutilización.

5. Datos sensibles: esencial no solamente para salvaguardar el derecho a la intimidad de las personas, sino también asegurando la seguridad de la organización. Todo ello informando de las condiciones de acceso a dichos datos con la adecuada documentación de los conjuntos de datos.
6. Acceso a los datos: debe contemplarse tanto la descarga completa de los conjuntos de datos como el acceso mediante servicios RESTful con sus correspondientes APIs. Además el acceso a los datos debería ser en tiempo real y actualizados.
7. Preservación de los datos: para ello deben utilizarse URIs persistentes, garantizando la futura consistencia de la interconexión de conjuntos de datos (*linked data*) y manteniendo siempre los datos en los formatos originales en que se publicaron aunque se adapten posteriormente nuevos formatos.
8. Retroalimentación: los usuarios deben poder indicar las posibles mejoras y aplicaciones de los conjuntos de datos.

Los casos que se han tenido en cuenta para la elaboración de estas prácticas resultan de gran interés para el desarrollo de nuevos proyectos de publicación de datos. En dicho análisis también identifican algunos vocabularios de gran relevancia para la publicación de metadatos de conjuntos de datos. Las ontologías PROV y ORG, disponibles respectivamente en:

<http://www.w3.org/TR/prov-o>

<http://www.w3.org/TR/vocab-org>

son esenciales para realizar una integración de información de diferentes fuentes y para describir la organización de la que proceden.

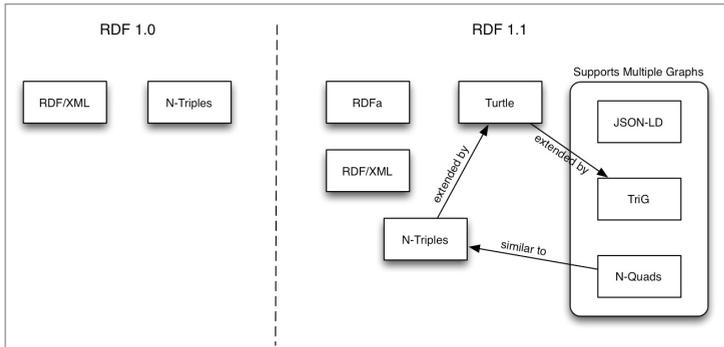
Podría decirse que la descripción de los metadatos sobre conjuntos de datos es uno de los puntos centrales de las buenas prácticas para su publicación. Teniendo en cuenta precisamente que estos metadatos suelen conformar catálogos de datos, la primera recomendación del vocabulario DCAT (*W3C, 2014e*) aporta una herramienta para una descripción normalizada tanto de los conjuntos de datos como de los propios catálogos.

También están presentes los *big data*, puesto que la recomendación del vocabulario *Data cube* (*W3C, 2014f*) se enfoca a la publicación de grandes cantidades de datos multi-dimensionales que

“La descripción de los metadatos sobre conjuntos de datos es uno de los puntos centrales de las buenas prácticas para su publicación”

usualmente se utilizan para representar datos estadísticos con diferentes niveles de agregación. Al parecer la recomendación R2RML para definir mapeados entre bases de datos relacionales y conjuntos de datos RDF no ha tenido la aceptación que se esperaba de ella. Tal vez se precise un enfoque más sencillo, como el consumo a través de la web de datos tabulares, generalmente originados en ficheros de hojas de cálculo o de texto plano separados por comas (CSV). Con respecto a esto se trabaja en dos frentes: por un lado la representación de los datos tabulares mediante RDF y por otro su representación mediante JSON-LD para su consumo directamente por el navegador³.

No hay que olvidar que cualquier tipo de datos, independientemente de su nivel de agregación o estructuración, precisa de sistemas para la clasificación o la descripción temática de su



Comparativa de los formatos de serialización de RDF 1.0 y RDF 1.1
<http://www.w3.org/TR/rdf11-new/serialization-formats.png>

contenido. En la web semántica, la tecnología desarrollada para ello es SKOS. Sin embargo, la publicación de la norma *ISO* sobre tesauros (*ISO*, 2011 y 2013) puso sobre la mesa ciertas carencias de SKOS en escenarios más exigentes. Por este motivo se hizo evidente la necesidad de definir una serie

de equivalencias entre SKOS y la norma *ISO*, que han conformado un documento con la especificación ISO-THES (Isaac; De-Smedt, 2015). También es reseñable el proyecto de código abierto SKOS-Lucene para dotar de soporte SKOS a *Apache Lucene* y *Solr*:
<https://github.com/behaskos/skos-lucene>

4. El mercado semántico

Parece que la confusión suscitada por la divergencia entre XHTML y HTML5 poco a poco va desapareciendo. Para entender esto, debemos considerar que HTML es en realidad un lenguaje abstracto para describir documentos y aplicaciones. Para concretar dicho lenguaje en una sintaxis concreta tenemos dos opciones en función de las necesidades del tipo de aplicación: la sintaxis de HTML (HTML5) y la sintaxis XHTML definida mediante XML (XHTML5).

Actualmente, los formatos de marcado semántico más extendidos son los microformatos, microdatos y RDFa. Sin duda el W3C apuesta por RDFa, tal y como lo demuestra la intensa actividad normativa al respecto para su aplicación tanto en HTML como XHTML (W3C, 2015b, 2015c y 2015d). Además, se sigue trabajando en el desarrollo de RDFa Lite (W3C, 2015e), como una versión más simple y sencilla de aplicar en el marcado semántico de documentos. A finales de 2013, el proyecto *Web data commons* apunta que el número de documentos HTML que contienen RDFa es mayor que aquellos con microdatos. Sin embargo, el mayor volumen de información estructurada (número de tripletas) se corresponde al marcado realizado con microdatos:

<http://webdatacommons.org/structureddata/2013-11/stats/stats.html>

Esta tendencia de uso preferente de microdatos sobre RDFa podría deberse a una estrategia de posicionamiento de *Google* en este tipo de tecnologías. De hecho este buscador muestra de un modo especial en los resultados de búsqueda aquellas páginas que utilizan la combinación microdatos-*schema.org*. Esto resulta especialmente relevante en algunos servicios, como *Google shopping*, puesto que los editores de contenidos de productos, en su afán por tener una visibilidad máxima, prefieren utilizar evidentemente el marcado demandado por *Google*.

El mercado semántico, tal como indica el informe de *Gartner* mencionado en la introducción, se encuentra ahora mismo en una fase emergente y su aplicación está ya produciéndose en la

actualidad. Los dos mayores riesgos a los que se podría enfrentar el mercado semántico son la dispersión de formatos y la heterogeneidad de vocabularios o modelos descriptivos. El primero de ellos se está solventando debido a que la mayor parte de contenidos presentes en la Web se publican mediante sistemas de gestión de contenidos que pueden utilizar varios formatos de marcado semántico de manera automática (Pastor, 2012; Pastor; Orduña; Saorín, 2013).

Schema.org intenta solventar la heterogeneidad de vocabularios, puesto que es un proyecto en el que colaboran *Yahoo*, *Google*, *Bing* y *Yandex*, de forma que se sientan unas bases comunes utilizables por todos los buscadores.

<http://schema.org>

Se trata una decisión muy inteligente: en la Web, los estándares abiertos y públicos facilitan la creación de nuevos servicios, mientras que las soluciones propietarias conducen al arrinconamiento y a la inversión de recursos en proyectos que posteriormente son abandonados.

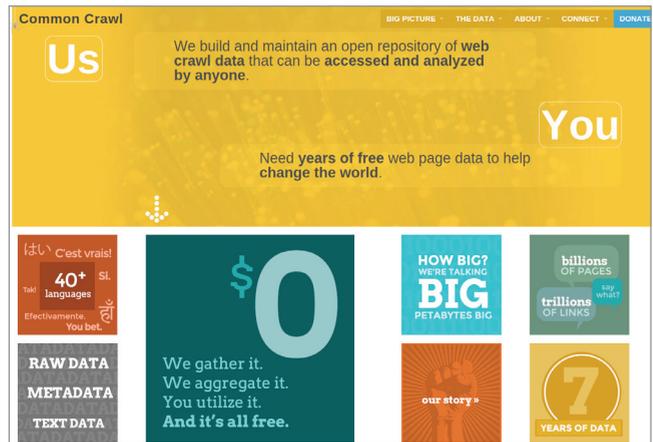
Pero no debemos detenernos en el mercado semántico como un factor de evolución de los motores de búsqueda. Se trata de una tecnología con muchas aplicaciones. El proyecto *Common crawl* ha recopilado cientos de terabytes de datos estructurados a partir de la análisis de miles de millones de páginas web para su explotación de forma totalmente abierta:

<http://commoncrawl.org>

De hecho, los archivos de 2015 contienen 1.820 millones de páginas y su volumen es de 139 TB. La explotación de estos conjuntos de datos puede proporcionar nuevas herramientas o usos de la Web, que hasta en este momento no somos capaces de vislumbrar (Meusel; Petrovski; Bizer, 2014).

5. El papel de *Wikipedia* como nodo central de la web semántica: de *DBpedia* a *WikiData*

Wikipedia no sólo se ha convertido en una referencia utilizada por millones de usuarios sobre conceptos y hechos, con un óptimo posicionamiento en buscadores, sino que además alimenta el nodo central de la web de datos enlazados, por cantidad de datos y enlaces:



Sitio web de *Common crawl*
<http://commoncrawl.org>

“Los dos mayores riesgos a los que se podría enfrentar el mercado semántico son la dispersión de formatos y la heterogeneidad de vocabularios o modelos descriptivos”

DBpedia. La *Wikipedia* semántica es un esquema fluido y compartido para la organización del conocimiento humano, en el cual en la actualidad conviven dos enfoques complementarios: *DBpedia* y *WikiData*:

<http://es.dbpedia.org>

<http://www.wikidata.org>

El punto de partida fue la ontología *DBpedia* sobre la cual, usando técnicas de análisis automatizado, se mapearon los artículos de *Wikipedia*, primero a partir de la edición inglesa, luego extendido a otros idiomas. En español se realizó esta semantización desde las *I Jornadas esDBpedia*, en noviembre de 2011. Posteriormente se realizaron otros trabajos para mejorar la calidad de los datos para convertir *esDBpedia* en el núcleo de los datos semánticos del idioma español. En 2014 el *Ministerio de Economía y Competitividad* financió la mejora de la infraestructura y servicios al usuario del nodo mantenido desde la *Red Temática Española de Linked Data*.

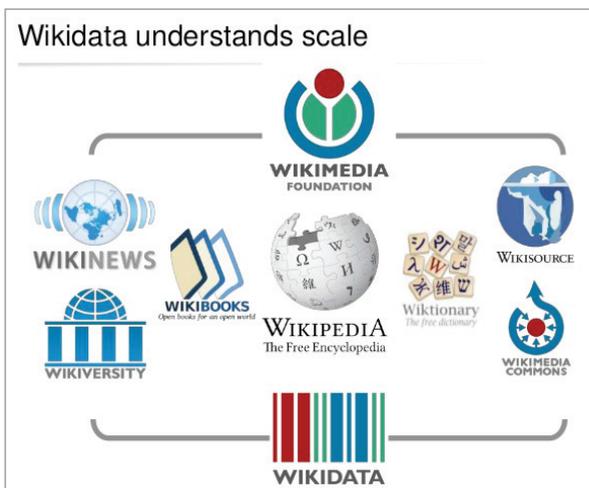
DBpedia añade una capa semántica externa a los contenidos de *Wikipedia*; las *infoboxes* y otros elementos de los artículos que responden a cierta estructura formalizada. De este modo la actividad editorial colaborativa plasmada en tablas, fichas y categorías, es forzada a convertirse en significados formalizados. Esta materia prima semántica multidisciplinar y en constante actualización, ofrece una potencialidad enorme para transformar el funcionamiento de la búsqueda y explotación de información en la Web, dado que el actual estado del arte de las tecnologías de la web semántica ha conseguido un equilibrio entre la capacidad de descripción y la capacidad de cálculo que lo hace eficiente (Rico, 2013). El esfuerzo en formalizar significados puede empezar a producir valor a gran escala y en diferentes dominios. Más de 100 millones de datos semánticos están entrando en ebullición.

En el ámbito del control de autoridades, la vinculación con *DBpedia* desde la versión semántica enriquecida de vocabularios y listas de valores permitirá colocar los tradicionales silos bibliotecarios como piezas en la corriente principal. Un ejemplo con amplias repercusiones en el ámbito del patrimonio cultural es el *Getty vocabulary program ontology* que integra desde

2014 al espacio *linked open data* sus dos principales vocabularios, el *Art & architecture thesaurus (AAT)* y el *Getty thesaurus of geographic names (TGN)*, y tiene previsto incorporar a lo largo de 2015 también la *Union list of artist names (ULAN)* y *Cultural objects name authority (CONA)*. Usa las ontologías SKOS, SKOS-XL, ISO 25964; DC, DCT, BIBO, FOAF y PROV.

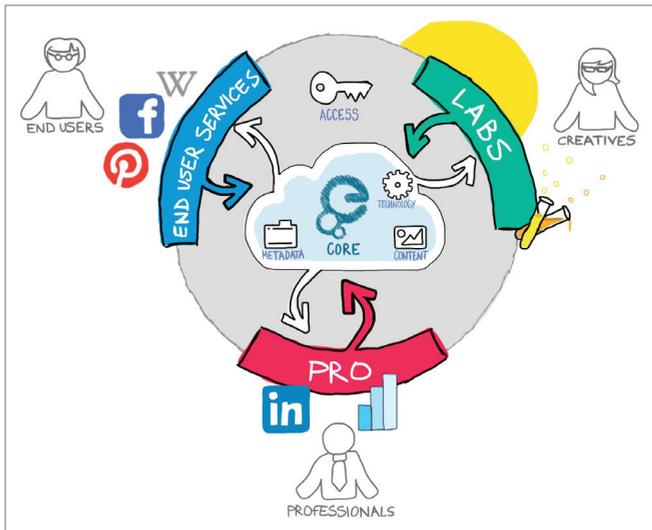
<http://vocab.getty.edu>

El impulso semántico ha llegado también dentro de la comunidad *Wikimedia*, que desde 2012 viene desarrollando el proyecto *WikiData* para proporcionar una base



Alcance del proyecto *WikiData*.

<http://www.slideshare.net/andrewsu/20140116-gmod-short>



El plan estratégico de *Europeana* hasta 2020: de portal a plataforma.

<http://strategy2020.europeana.eu/it>

datos de *DBpedia* se actualicen también desde *WikiData*. Al estar esta última descargable en RDF, tecnológicamente será viable con unas complicaciones razonables, aunque no triviales.

<http://wiki.dbpedia.org/meetings/Amsterdam2014>

“La web semántica tiene efectos colaterales en posicionamiento, tráfico, y en consecuencia, en algunas de las formas convencionales de retorno de la inversión, léase publicidad online, reputación y visibilidad”

un uso indirecto, que altera el régimen normal de visitas e impactos directos en el sitio web. La web semántica tiene efectos colaterales en posicionamiento, tráfico, y en consecuencia, en algunas de las formas convencionales de retorno de la inversión, léase publicidad online, reputación y visibilidad.

6. Conclusiones: la web semántica como plataforma

Hemos dejado para el final el comentario de dos grandes proyectos culturales que apuestan por el enriquecimiento de metadatos a través de datos enlazados, *Europeana* y *Digital Public Library of America*. Ambos proyectos están llamados a suponer una transformación en la forma de modelar los registros de información sobre contenidos digitales, usando modelos de datos altamente expresivos y alineados con los estándares de la web semántica. *Europeana data model v.5.2.6* (diciembre 2014) y *Metadata application profile v.4* (marzo 2015), disponibles en:

la definición colaborativa de propiedades. Desde el punto de vista epistemológico, tiene un indudable valor que *WikiData* almacene no sólo las afirmaciones (*statements*), sino también sus fuentes, reflejando la diversidad del conocimiento y reforzando la regla de hierro de *Wikipedia* de la verificabilidad.

Por otro lado, *DBpedia* es un nodo activo de la infraestructura de los *linked open data*, y *WikiData* no. En la agenda del 1st *DBpedia community meeting* de 2014 en Amsterdam estuvo la coordinación entre ambos proyectos, de forma que los

Queremos, no obstante, resaltar una paradoja en un escenario de una *Wikipedia semántica*: ya podemos apreciar cómo *Google* es capaz de aprovechar sus contenidos y ofrecer respuestas directas en los resultados de búsquedas, con lo que parte del uso actual deriva hacia

<http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation>
<http://dp.la/info/developers/map>

están llamados a ser el patrón oro de la descripción de recursos digitales en red. Es de señalar además que en sus planes estratégicos⁴ se conciben ya como plataformas más que como meros portales agregadores, entendiendo a la perfección la transformación que supone ser un proveedor de datos en la web semántica, donde la riqueza viene de la posibilidad de la existencia de servicios digitales innovadores y de valor a partir de fuentes de datos combinables y accesibles en línea. El lema de *Europeana* “Los portales son para visitar, las plataformas son para construir” es esencial para entender el presente inmediato de la extensión de la web semántica, siendo necesarios proyectos que pongan en valor la potencialidad que reside en los datos semánticos.

Notas

1. Este proceso se denomina “*Bnode skolemization*” y consiste en convertir nodos en blanco en IRIs, creando de este modo referencias persistentes más estables para la elaboración de consultas y el consumo de datos RDF.

<http://www.w3.org/wiki/BnodeSkolemization>

2. No hay que olvidar que una cosa son las buenas prácticas de los procesos de edición y otra su implementación mediante algún tipo de software o plataforma *linked data*. Para entender este punto es imprescindible consultar los documentos:

<http://www.w3.org/TR/ldp>

<http://www.w3.org/TR/ldp-bp>

3. Resulta de gran interés consultar los siguientes documentos sobre el modelo de representación de datos tabulares en la Web y la aplicación de RDF y JSON-LD para ello:

<http://www.w3.org/TR/csv2rdf>

<http://www.w3.org/TR/csv2json>

<http://www.w3.org/TR/tabular-metadata>

<http://www.w3.org/TR/tabular-data-model>

4. En este sentido es especialmente relevante el *Plan estratégico 2015-2017* de la *Digital Public Library of America*:

http://dp.la/info/wp-content/uploads/2015/01/DPLA-StrategicPlan_2015-2017-Jan7.pdf

y el *Plan estratégico de Europa 2015-2020*:

http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Strategy%202020.pdf

8. Referencias

Isaac, Antoine; De Smedt, Johan (2015). ISO 25964 SKOS extension.

<http://pub.tenforce.com/schemas/iso25964/skos-ths>

ISO 25964-2:2011. *Thesauri and interoperability with other vocabularies*. Part 1: *Thesauri for information retrieval*. Geneva: International Organization for Standardization.

ISO 25964-2:2013. *Thesauri and interoperability with other vocabularies*. Part 2: *Interoperability with other vocabularies*. Geneva: International Organization for Standardization.

Meusel, Robert; Petrovski, Petar; Bizer, Christian (2014). The webdata commons microdata, RDFa and microformat dataset series. En: *The semantic web–ISWC 2014*. Springer International Publishing, pp. 277–292.

Pastor-Sánchez, Juan-Antonio (2012). “Los CMS como pieza fundamental en el despliegue de la web semántica”. *Anuario ThinkEPI*, 2012, n. 6, pp. 184-189.

<http://recyt.fecyt.es/index.php/ThinkEPI/article/view/30416>

Pastor-Sánchez, Juan-Antonio; Orduña-Malea, Enrique; Saorín, Tomás (2013). "Marcado semántico automático en gestores de contenidos: integración y cuantificación". *El profesional de la información*, v. 22, n. 5, pp. 381-391.

<http://dx.doi.org/10.3145/epi.2013.sep.02>

Pérez, Sarah (2012). *Wikipedia's next big thing: Wikidata, a machine-readable, user-editable database funded by Google, Paul Allen and others.*

<http://techcrunch.com/2012/03/30/wikipedias-next-big-thing-wikidata-a-machine-readable-user-editable-database-funded-by-google-paul-allen-and-others>

Rico, Mariano (2013). "Evolución de la información y su captación en la web del futuro". *Encuentros multidisciplinares*, v. 15, n. 45, pp. 52-59.

Saorín, Tomás; Peset, Fernanda; Ferrer-Sapena, Antonia (2013). "Factores para la adopción de *linked data* e implantación de la web semántica en bibliotecas, archivos y museos". *Information Research*, 2013, v. 18, n. 1, paper 570.

<http://informationr.net/ir/18-1/paper570.html>

Vrandečić, Denny; Krötzsch, Markus (2014). "Wikidata: A free collaborative knowledge base". *Communications of the ACM*, v. 57, n. 10, pp. 78-85.

<http://dx.doi.org/10.1145/2629489>

W3C (2014a). RDF 1.1 Primer. *W3C Working group note*, 24 June.

<http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624>

W3C (2014b). RDF 1.1 N-Quads: A line-based syntax for RDF datasets. *W3C Recommendation*, 25 February.

<http://www.w3.org/TR/2014/REC-n-quads-20140225>

W3C (2014c). RDF 1.1 TriG. *RDF dataset language.*

<http://www.w3.org/TR/2014/REC-trig-20140225>

W3C (2014d). JSON-LD 1.0: A JSON-based serialization for linked data. *W3C Recommendation*, 16 January.

<http://www.w3.org/TR/2014/REC-json-ld-20140116>

W3C (2014e). Data catalog vocabulary (DCAT). *W3C Recommendation*, 16 January.

<http://www.w3.org/TR/2014/REC-vocab-dcat-20140116>

W3C (2014f). The RDF data cube vocabulary. *W3C Recommendation*, 16 January.

<http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116>

W3C (2015a). Data on the web best practices. *W3C First public working draft*, 24 February.

<http://www.w3.org/TR/2015/WD-dwbp-20150224>

W3C (2015b). RDFa 1.1 Primer - 3rd ed.: Rich structured data markup for web documents. *W3C working group note*, 17 March.

<http://www.w3.org/TR/2015/NOTE-rdfa-primer-20150317>

W3C (2015c). HTML+RDFa 1.1 - 2nd ed.: Support for RDFa in HTML4 and HTML5. *W3C Recommendation*, 17 March.

<http://www.w3.org/TR/2015/REC-html-rdfa-20150317>

W3C (2015d). XHTML+RDFa 1.1 - 3rd ed.: Support for RDFa via XHTML modularization. *W3C Recommendation*, 17 March.

<http://www.w3.org/TR/2015/REC-xhtml-rdfa-20150317>

W3C (2015e). RDFa Lite 1.1 - 2nd ed. *W3C Recommendation* 17 March.

<http://www.w3.org/TR/2015/REC-rdfa-lite-20150317>

