

## J.9. Prospektiva de la web semántica: divergencia tecnológica y creación de mercados *linked data*

Juan-Antonio Pastor-Sánchez

20 diciembre 2011

**Pastor-Sánchez, Juan-Antonio.** "Prospektiva de la web semántica: divergencia tecnológica y creación de mercados *linked data*". *Anuario ThinkEPI*, 2012, v. 6, pp. 269-275.



**Resumen:** Se presentan algunas tendencias en el entorno de la web semántica: *Schema.org*, un proyecto desarrollado por *Google*, *Yahoo* y *Bing* para el mercado semántico mediante microdatos; *Freebase*, un banco de conocimiento para la estructuración de datos y *Kasabi*, una plataforma que opera en la modalidad "software as a service" para la publicación y comercialización de datos semánticos. Se observa cierta divergencia tecnológica en la aplicación de estándares de la web semántica con el desarrollo de nuevas propuestas como los microdatos, así como el surgimiento de un mercado *linked data* a través de plataformas de publicación que facilitan su comercialización.

**Palabras clave:** Web semántica, *Linked open data*, Microdatos, RDFa, *Schema.org*, *Kasabi*, *Freebase*, *SaaS*, Mercado de datos semánticos, Divergencia tecnológica.

**Title: Prospective for the semantic web: technological divergence and linked data markets creation**

**Abstract:** This paper shows some trends that are currently occurring in the environment of the semantic web. More specifically it presents *schema.org*, a project developed by *Google*, *Yahoo* and *Bing* for semantic markup using microdata, *Freebase*, a knowledge bank for structured information and *Kasabi*, a platform that operates into the modality of "Software as a Service" for the publication and commercial distribution of semantic data. It is possible to see some technological divergence in the application of semantic Web standards with the development of new proposals as microdata, and the emergence of a market for linked data through publishing platforms that facilitate commercialization.

**Keywords:** Semantic web, *Linked open data*, Microdata, RDFa, *Schema.org*, *Kasabi*, *Freebase*, *SaaS*, Semantic data market, Technological divergence.

El progresivo despliegue de la web semántica en los últimos años está produciendo nuevos servicios en la Web. Buena muestra de ello son las soluciones planteadas al amparo de la corriente *linked open data*, la progresiva adopción de RDF y de servicios web basados en *Sparql Endpoint*.

La elaboración y publicación de estos datos se realiza por parte de especialistas e instituciones, dentro del ámbito de proyectos de gran envergadura o sobre temas concretos. Es relevante el documento sobre vocabularios y conjuntos de datos enlazados bibliotecarios (Isaac et al., 2011)<sup>1</sup>, el catálogo "The datahub" de CKAN<sup>2</sup> o el "Opening up government" del Reino Unido<sup>3</sup>.

Existe un interés creciente por parte de organismos y empresas para publicar datos utilizando técnicas que incrementen la interoperabilidad de los mismos (Heath y Bizer, 2011), con vistas a su aplicación y explotación en futuros servicios. Otro tipo de iniciativas se centran en experimen-

tar aspectos relacionados con la disponibilidad y reutilización de cantidades masivas de datos, tal y como se hace desde el proyecto *DBpedia*. Experimentar y "jugar" con las nuevas ideas y propuestas es un paso indispensable para el desarrollo de nuevas tecnologías en la Web (Saorín, 2011).

Las iniciativas anteriores comparten una misma premisa: la publicación de datos "ad-hoc" para la web semántica, ya sea a partir de bases de datos de aplicaciones (catálogos, directorios, fondos documentales), vocabularios controlados (*Agrovoc*, *Rameau*, *Lcsh*, *STW*, *LEM*) u otro tipo de productos resultado del procesamiento de datos y fuentes de información de diversa índole, siendo *DBpedia* el núcleo central de la nube de interrelaciones en el ecosistema de *linked open data* (Bizer et al., 2009).

Otros enfoques se centran en:

– "semantización" de contenidos de sitios web creados para su publicación y consulta para el usuario final;

- servicios orientados a la publicación colaborativa de datos para la creación de bases de conocimiento estructurado;
- plataformas para la gestión de conjuntos de datos.

Son aproximaciones cuyo uso está abierto al público y están centradas en la descripción de contenidos, recursos y objetos desde una perspectiva semántica.

## Schema.org

Comenzó su andadura el 2 de junio de 2011. Está desarrollado de forma conjunta por *Google*, *Yahoo* y *Bing*. A primera vista puede sorprender que estos tres motores de búsqueda, aparentemente competidores, colaboren para crear un proyecto. La expansión de las aplicaciones y servicios de la web 2.0 ha complicado el panorama de las búsquedas en la Web. Los motores de búsqueda han tenido que adaptarse a esta nueva realidad, modificando sus algoritmos de relevancia, detectando duplicaciones, evitando técnicas de *cloacking* o *stuffing*, promocionando contenidos de calidad contrastada o filtrando según la naturaleza del recurso. Todo ello sin entrar en la fiabilidad, la calidad o la usabilidad de los contenidos. La solución es obvia: incluir metadatos dentro de los propios contenidos web.

Sin embargo, anexas metadatos sobre aspectos generales de una página (descripción, autoría, palabras clave) tiene una importante limitación: no permite describir semánticamente aspectos detallados del contenido, pese a que en muchas ocasiones contenga información altamente estructurada, originalmente almacenada en bases de datos. Una página web con información sobre la cartelera de cine incluye información sobre películas, horario de sesiones, precios, ubicación de la sala, etc. Generalmente la información esta marcada para obtener un formato adecuado para su visualización y lectura por parte de personas. Pero también sería de gran utilidad que dicha página tuviera un marcado semántico, en el que todos esos datos estuvieran descritos de forma explícita y detallada, para su posterior procesamiento por parte de otras aplicaciones.

De este modo, sería muy sencillo plantear y resolver una consulta para conocer qué películas se proyectan entre las 18:00 y las 21:00 horas en un radio de 1.500 metros del lugar en el que reside el usuario. Otro ejemplo: los diferentes medios de transporte (autobuses, trenes, aviones) disponen de páginas donde se muestran horarios de salida y llegada, precio o recorrido. Este tipo de información es una muestra más de información altamente estructurada y susceptible de representarse mediante algún tipo de ontología

The screenshot shows the Schema.org website for the 'Thing' class. At the top, it says 'The most generic type of item.' Below this is a table with three columns: 'Property', 'Expected Type', and 'Description'. The table lists four properties: 'description' (Text), 'image' (URL), 'name' (Text), and 'url' (URL). Below the table, there is a section titled 'More specific types' which lists several subclasses: CreativeWork, Event, Intangible, Organization, Person, Place, and Product.

Property	Expected Type	Description
<b>Properties from Thing</b>		
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
url	URL	URL of the item.

**More specific types**

- [CreativeWork](#)
- [Event](#)
- [Intangible](#)
- [Organization](#)
- [Person](#)
- [Place](#)
- [Product](#)

para su procesamiento y posterior ejecución de inferencias: en este caso, podría ser el trayecto óptimo entre dos ciudades.

Precisamente, eso lo que ofrece *Schema.org*<sup>4</sup>, una pseudo-mega-ontología, en forma de esquema de carácter general, que intenta representar una amplia variedad de hechos: obras, eventos, intangibles, organizaciones, personas, lugares y productos. Partiendo de la superclase "thing", se definen varios cientos de clases y subclases (tipos), con multitud de propiedades, que permiten realizar descripciones detalladas y específicas sobre un determinado elemento informativo<sup>5</sup>.

El modelo de datos utilizado deriva de RDF, de manera que: los tipos se estructuran en una jerarquía múltiple (un tipo puede ser subclase de más de un tipo) dotada de herencia de propiedades, al tiempo que el dominio y el rango de éstas pueden estar constituidos por más de un tipo. El esquema general no puede modificarse por los usuarios, pero se proporciona un mecanismo para su extensión mediante la derivación de tipos y propiedades generales y, por tanto, la adaptación a necesidades concretas<sup>6</sup>.

La aplicación de *Schema.org* está íntimamente ligada al desarrollo de html5, puesto que se basa en el uso de microdatos para la inclusión de información semántica. El resultado final, es una página web que puede ser utilizada por una aplicación o agente informático para la extracción de información semánticamente relevante de forma mucho más precisa y exacta que otro tipo de técnicas como el *screen scraping* o el *web minning*.

El objetivo de *Schema.org* es ofrecer un mecanismo para el marcado semántico del código html, haciendo que una página web sea un objeto procesable para la extracción de información semán-

tica. Las implicaciones de este tipo de solución en el funcionamiento de los motores de búsqueda son considerables: puesto que estos funcionan mejor con información estructurada, las búsquedas serían más precisas, las posibilidades de filtrado aumentarían la versatilidad del proceso de recuperación de información, los resultados podrían visualizarse de forma más personalizada a los gustos de usuario, etc. Al mismo tiempo se abren nuevos horizontes en la reutilización de dichos contenidos por parte de terceros.

## Freebase

Su objetivo es la creación de una gran base de conocimiento y no, como en el caso de *Schema.org*, el marcado semántico de páginas html. Cualquiera puede registrarse como usuario e introducir datos siguiendo un esquema propio o reutilizar alguno ya existente en *Freebase*. Los esquemas incluyen tipos y propiedades que permiten describir objetos de información, que en la terminología de *Freebase* se denominan "topics". Los datos pueden introducirse manualmente o utilizar algunas herramientas para su carga masiva (**Bollacker, 2008**).

Tras habituarse al entorno, terminología y funciones de *Freebase*, se tiene la sensación de estar trabajando en una *wiki* de carácter semántico, en la que los "topics" desempeñan el papel de los artículos de *Wikipedia* y en la que cualquier miembro de la comunidad puede realizar aportaciones y correcciones sobre los datos. Actualmente contiene más de 20 millones de "topics". La licencia de uso de dichos contenidos es *Creative Commons Attribution (CC-BY)*.

Uno de los aspectos más interesantes de *Freebase* es el enorme potencial para la creación de aplicaciones, en cualquier plataforma, que permita la explotación de sus datos. Existe una api formada por un conjunto de servicios web REST muy sencillos, que permiten la consulta y la escritura (previa identificación como usuario autorizado) de datos. *Freebase* utiliza Json (como formato para el intercambio de datos) y MQL, como lenguaje de consulta<sup>8</sup>. Estamos ante un servicio que no sólo aporta una plataforma para la gestión y consulta de los datos, sino que además ofrece un amplio soporte al desarrollo de aplicaciones.

Se ofrece un soporte muy básico para RDF, permitiendo acceder a la descripción de cada "topic" mediante RDF/XML. También puede descargarse un volcado completo de *Freebase* en formato TSV.

Un indicio del impacto que está teniendo este tipo de servicios es la adquisición de *Metaweb* (la compañía que desarrolló *Freebase*) por parte de *Google* en 2010.

## Kasabi

*Freebase* y *Schema.org* tienen aproximaciones propias, que difieren en muchos aspectos de las soluciones desarrolladas para la web semántica en el seno del W3C. Ninguna de ellas utiliza de un modo claro el modelo de datos propuesto por RDF (esencial en la web semántica) ni hace uso de otras tecnologías que podrían ser aplicadas, como RDFa (en el caso de *Schema.org*) o Sparql (en el caso de *Freebase*).

Ambas ofrecen una aplicación inmediata y sencilla de algunos planteamientos de la web semántica, y ofrecen mecanismos para extender los esquemas de descripción generales. Sin embargo, su énfasis en simplificar los procesos de gestión y uso de los datos les hace optar por no utilizar ciertos aspectos normativos del W3C.

Otros tipos de usuarios tal vez prefieran atenerse a las diferentes recomendaciones del W3C y puedan generar sus propios datos en un formato que cumpla dichas especificaciones. Estos usuarios podrían generar salidas en RDF/XML o *Turtle* a partir





**Datasets**

Guided search

Click a term to initiate a search

**Category**

- Government (48)
- Geography (42)
- Uncategorised (32)
- Travel (27)
- Media (21)
- Publishing (18)
- Linking (16)
- Health (13)
- Music (12)
- Science (12)
- Education (11)
- Business (9)
- Commerce (8)
- Food (3)

**Results**

**Adelaide Metro**  
Dataset has 6 Standard APIs  
Published by Ian Davis in Geography Travel

**aim25omp**  
Dataset has 6 Standard APIs  
Published by Rory McNicholl in Uncategorised

**Airlines**  
Dataset has 6 Standard APIs  
Published by Kasabi Administrator in Travel

**Airports**  
Dataset has 6 Standard APIs  
Published by Kasabi Administrator in Travel

**aligaballa**  
Dataset has 6 Standard APIs  
Published by ali abd el-g hany gab... in Health Uncategorised

**BBC**  
Dataset has 2 Custom APIs and 6 Standard APIs  
Published by Leigh Dods in Music Media

**BBC Music**  
Dataset has 2 Custom APIs and 6 Standard APIs

de bases de datos propias, y precisarían de una plataforma *software as a service (SaaS)*<sup>9</sup> para gestionar dichos datos, visualizarlos o definir *Sparql Endpoints* (Heath, 2008). Esto es lo que ofrece *Kasabi*<sup>10</sup>: una plataforma para la gestión y diseminación de datos RDF.

*Kasabi* está basado en la plataforma *Talis*<sup>11</sup> y, además de ofrecer soporte para el almacenamiento de conjuntos de datos (*datasets*), dispone de un completo repertorio de apis (su punto fuerte) para la gestión y acceso a los datos. Dichas apis son muy flexibles, siendo posible personalizar ciertos parámetros definiendo vistas, filtros, formatos y estilos de visualización. Las apis se basan en la definición servicios web REST o de *Sparql Endpoints*. El acceso a las apis, y por tanto a los datos para su explotación y reutilización, se realiza a través de una clave asociada al conjunto de datos en cuestión<sup>12</sup>.

Tras registrarse como usuario es posible crear nuevos conjuntos de datos (hasta cinco en esta versión beta), configurar ciertos aspectos (estado de publicación, categorías temáticas, tipo de licencia de los datos, etc.) y mantener los datos. Éstos pueden añadirse directamente a un *dataset*, indicando una uri o mediante extracción RDFa de un url concreto. El modelo utilizado es RDF. *Kasabi* sigue las recomendaciones del W3C en cuanto al uso de tecnologías de la web semántica.

## Divergencia tecnológica

Uno de los aspectos más llamativos de *Schema.org* y *Freebase* es el planteamiento de una divergencia en relación con ciertos planteamien-

tos del W3C. El modelo RDF prácticamente no se contempla y, si se hace –visualización RDF de los “topics” en *Freebase*–, es mediante una aplicación muy general.

En el caso de *Schema.org* el modelo de descripción es jerárquico. El uso de microdatos frente a RDFa supone un motivo de confusión. Este punto resulta un tanto polémico puesto que se plantea una posible barrera entre html y xhtml, de forma que los microdatos se utilicen en html5 y RDFa en xhtml<sup>13</sup>.

Esto puede resultar problemático debido a las diferencias de modelo entre ambas opciones: mientras que los microdatos tienen una estructura jerárquica, RDFa utiliza el modelo general de RDF basado en grafos (Sporny, 2011). Por este motivo es posible realizar el mapeado de una estructura de microdatos a RDFa, pero no a la inversa. RDFa es mucho más flexible ya que permite la combinación de múltiples vocabularios y esquemas RDF, mientras que *Schema.org* se centra en una estructura *ad-hoc* de tipos y propiedades.

Así pues, el uso de los microdatos y su capacidad de interoperabilidad en el futuro es incierta. La sencillez de su sintaxis, frente a la mayor complejidad de RDFa, aducida por los responsables de *Schema.org*, no es motivo suficiente para optar por dicha solución. Sobre todo si se tiene en cuenta que el uso de una u otra tecnología estará mediado por un sistema de gestión de contenidos (CMS), que realizará el mapeado entre las estructuras internas de los contenidos y los elementos de los esquemas de metadatos correspondientes, y la consiguiente generación automática de código html/xhtml.

Sería recomendable que en un futuro se alcanzaran tres objetivos:

- separar las especificaciones de microdatos y RDFa de un lenguaje de marcado concreto; de esta forma ambas opciones podrían usarse tanto en html5 como en xhtml;
- los motores de búsqueda deberían considerar ambas opciones;
- ir más allá del esquema propuesto en *Schema.org*; además de no obligar a adoptar un esquema concreto, podrían incorporarse descripciones de recursos con vocabularios RDF con una semántica bien definida.

Otro punto de separación lo constituyen los mecanismos de extensión de los esquemas propuestos tanto por *Freebase* como por *Schema.org*. La capacidad de ultra-personalización que ofrecen ambas iniciativas constituye un arma de doble filo: por un lado ofrece libertad total para la derivación de elementos del esquema, pero por otro puede suceder que se represente cierto tipo de información con elementos definidos por el usuario, en lugar de con los tipos y propiedades propios de *Schema.org* para ello.

Los esquemas propuestos por *Schema.org* y *Freebase* son realmente extensos, su comprensión es compleja, y el ámbito de aplicación de algunos elementos descriptivos puede resultar confuso. Si un *webmaster* no comprende un elemento –o no localiza alguno que le resulte apropiado– definirá uno nuevo. Esto puede producir una dispersión semántica en la descripción de recursos que puede provocar resultados no deseados en los procesos de búsqueda. El uso de esquemas más pequeños y sencillos proporciona una visión modular de la web semántica que resulta mucho más adecuada en la descripción de recursos.

En el caso de *Freebase* encontramos otro aspecto que se separa de las tecnologías de la web semántica: MQL. Aunque en algunos puntos MQL es similar a Sparql, su sintaxis es totalmente distinta ya que se basa en Json, y su potencialidad es inferior a la de Sparql. En realidad MQL (*metaweb query language*) no es un lenguaje propiamente dicho se trata de una api que permite definir ciertos patrones como condiciones que deben cumplir los datos a recuperar. El grado de interoperabilidad de los datos de *Freebase* sería mayor si cada conjunto de datos contara con un *Sparql endpoint*.

## Mercado *linked data*, SaaS y apis: evoluciones necesarias

La web semántica precisa de ciertos elementos que aumenten su radio de acción y contribuyan a incrementar su despliegue. *Kasabi* sugiere y lleva a la práctica algunas ideas y conceptos que pueden resultar de gran interés a los profesionales y usuarios que operen en este entorno.

Hablar de un mercado de datos a partir de la web semántica y de *linked open data* puede resultar un contrasentido, pero resulta necesario hacerlo y referirse a *linked data* (sin el "open") para crear nuevos modelos de negocio y oportunidades.

Dicho mercado podría matizarse, así como los servicios y productos objeto de la compraventa: podrían comprarse los datos en su totalidad o parte de ellos, con mayor o menor nivel de granularidad (con la consiguiente diferencia de tarifas), ofrecer un producto básico gratuito y otro producto comer-

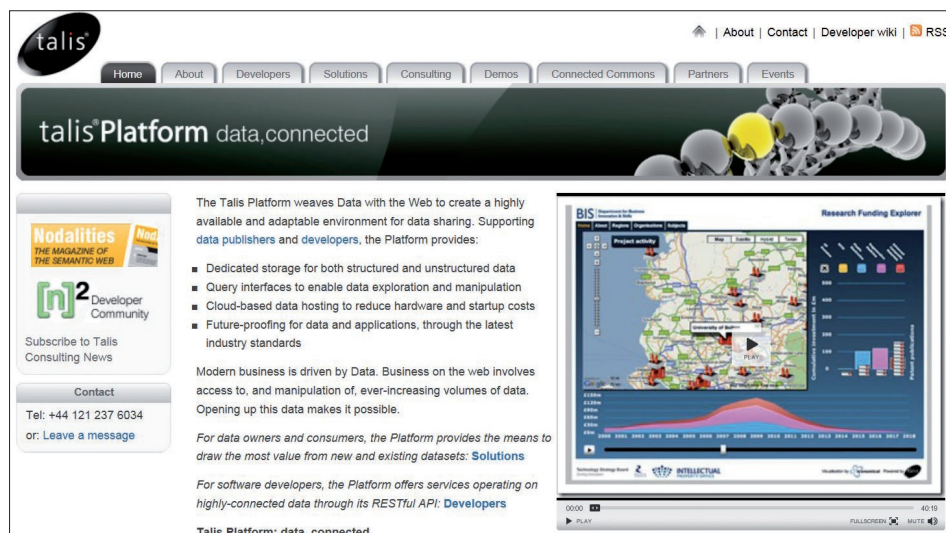
cial mucho más avanzado, con datos añadidos producto de un procesamiento de los datos originales. Comparando este nuevo mercado con el de los medios de comunicación en internet, es fácil llegar a la conclusión de que el camino a seguir sería muy similar. Evidentemente hay aspectos que habría que afinar en este modelo: derechos de explotación de datos en función de su origen, licencias de uso, inclusión de publicidad, etc.

La aportación de conjuntos de datos de organismos e instituciones públicas desempeñaría un papel fundamental, constituyendo núcleos de datos enlazados abiertos a su uso y explotación públicos. La sociedad invierte en la gestión digital y cotidiana de dichos datos, y en un futuro inmediato deberá ser obligación de las administraciones públicas la difusión de los mismos de forma desagregada, no en forma de documentos, ya que de este modo se consigue mayor versatilidad para su reutilización.

El uso de dichas aportaciones por parte de organizaciones privadas para su procesamiento con vistas a incorporar valor añadido, generaría nuevos focos de negocio y por tanto empleo y crecimiento económico, además de difundir las actividades de las administraciones públicas en este sentido<sup>14</sup>.

En este modelo de negocio resulta fundamental la existencia de plataformas SaaS de uso gratuito con ciertas limitaciones y que cubrirían las necesidades de gestión y publicación de datos de la mayor parte de entidades públicas o de cualquier otra procedencia, siempre y cuando tuvieran licencias de uso tipo *Creative Commons*. Para otro tipo de requisitos asociados a la explotación comercial de dichos datos mediante acceso restringido a estas plataformas, se podría disponer de servicios de pago con un amplio catálogo de apis: limitación del acceso a

<http://kasabi.com>



<http://www.talis.com/platform>

*datasets* mediante claves de autenticación, aplicación de estilos visuales, sistemas de búsqueda avanzados, incorporación de fuentes de datos externas, definición de ontologías, ejecución de inferencias etc.

En cuanto a las apis para la reutilización de datos, también puede observarse que se están incorporando los principios de *crowdsourcing* para su desarrollo. Este modelo no es nuevo, puesto que, en gran medida, la implantación y evolución de los CMS se debe al desarrollo de su catálogo de *plugins* y módulos (mediante este modelo (Leimeister et al., 2009).

En el caso de las plataformas para la publicación de datos semánticos, se debería seguir un modelo similar, que enriqueciera las aplicaciones y con capacidad de reutilización de los datos. De este modo, serían los propios editores y usuarios de datos (y no únicamente los programadores) los que marcarían las tendencias y necesidades en el ámbito de la web semántica.

## Conclusiones: apertura, semantización e integración

El conocimiento estratégico del *momentum* actual de la web semántica puede ayudar a decidir el modo en el que una comunidad, proyecto u organización puede participar en este entorno aportando datos o dotando a sus proyectos de ciertas características que pueden ayudar a un mejor posicionamiento y difusión de sus contenidos y actividades.

Los proyectos analizados, aunque mejorables (en especial *Schema.org* y *Freebase*), constituyen aportaciones de gran interés. *Schema.org* muestra la relevancia que tendrán los procesos de "semantización" de la Web convencional, y por tanto la creación de vínculos, y la consi-

guiente integración, con conjuntos de datos RDF. *Freebase* muestra un posible camino evolutivo del modelo *wiki*, en donde comunidades de usuarios participan en la gestión descriptiva de objetos de información de manera estructurada, y tal vez podría ser un referente para *Wikipedia*. *Kasabi* ofrece una visión esquemática de cómo serían las plataformas de publicación de datos semánticos, con las que pueden desarrollarse nuevos modelos de negocio.

Hay una tendencia general a facilitar la publicación de datos semánticos, y un escenario en el que surgen tecnologías divergentes con enfoques alternativos sobre la Web semántica. Las organizaciones deben encontrar su posición en esta nueva web semántica más abierta, mientras que los desarrolladores deben considerar la aplicación o evolución de tecnologías ya existentes para no volver a pasar por un camino ya recorrido.

Resulta imprescindible abrir nuevos focos de debate y desarrollo sobre tecnologías y aplicaciones de la Web semántica y *linked data*, porque la web semántica es de todos.

## Referencias bibliográficas

**D'Aquin, Mathieu et al.** "Toward a new generation of semantic web applications". *Intelligent systems, IEEE*, mayo de 2008, v. 23, n. 3, pp. 20-28.  
<http://dx.doi.org/10.1109/MIS.2008.54>

**Bizer, Christian et al.** "DBpedia: A crystallization point for the web of data". *Journal of web semantics: science, services and agents on the world wide web*, 2009, v. 7, n. 3 pp. 154-165.  
<http://dx.doi.org/10.1016/j.websem.2009.07.002>

**Bollacker, Kurt et al.** "Freebase: a collaboratively created graph database for structuring human knowledge". En: *Procs. of the 2008 ACM Sigmod intl conf on management of data*.  
<http://147.46.216.176/w/images/9/98/SC17.pdf>  
<http://dx.doi.org/10.1145/1376616.1376746>

**Franklin, Michael; Halevy, Alon; Maier, David.** "From databases to dataspace: a new abstraction for information management". *ACM Sigmod Record*, dic. 2005, v. 34, n. 4, pp. 27-33.  
<http://www.cs.washington.edu/homes/alon/files/dataspaceDec05.pdf>  
<http://dx.doi.org/10.1145/1107499.1107502>



**Heath, Tom.** "How will we interact with the Web of data?" *Internet computing, IEEE*, sept.-oct., 2008, v. 12, n. 5, pp. 88-91.

<http://tomheath.com/papers/heath-2008-how-will-we-interact-with-the-web-of-data.html>

<http://dx.doi.org/10.1109/MIC.2008.101>

**Heath, Tom; Bizer, Christian.** "Linked data: evolving the Web into a global data space". *Synthesis lectures on the semantic web: theory and technology*, 2011, v.1, n. 1, pp. 1-136.

<http://linkeddatatoolkit.com/book>

**Isaac, Antoine et al.** "Library Linked Data Incubator Group: datasets, value vocabularies, and metadata element sets". *W3C Incubator Group Report*, 25 de octubre de 2011.

<http://www.w3.org/2005/Incubator/ld/XGR-ld-vocab-dataset-20111025>

**Leimeister, Jan Marco et al.** "Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition". *Journal of management information systems*, 2009, v. 26, n. 1, pp. 197-224.

[http://www.uni-kassel.de/fb7/libw/leimeister/pub/JMLS\\_124.pdf](http://www.uni-kassel.de/fb7/libw/leimeister/pub/JMLS_124.pdf)

<http://dx.doi.org/10.2753/MIS0742-1222260108>

**Manku, Gurmeet-Singh; Jain, Arvind; Sarman, Anish-Das.** "Detecting near-duplicates for web crawling". En: *Proceedings of the 16<sup>th</sup> intl conf on World Wide Web ACM New York*, 2007.

[http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/research.google.com/es/pubs/archive/33026.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/es/pubs/archive/33026.pdf)

<http://dx.doi.org/10.1145/1242572.1242592>

**Pastor-Sánchez, Juan-Antonio.** *Tecnologías de la web semántica*. Colección El profesional de la información, n. 1, Barcelona: Editorial UOC, 2011. ISBN: 978 84 9788 474 7

**Saorín, Tomás.** "Cómo Linked open data impactará en las bibliotecas a través de la innovación abierta". *Anuario ThinkEPI*, 2012, v. 6.

<http://www.thinkepi.net/como-linked-open-data-impactara-en-las-bibliotecas-a-traves-de-la-innovacion-abierta>

**Sporny, Manu.** "An uber-comparison of RDFa, microdata and microformats". *The beautiful, tormented machine*, 25 junio, 2011.

<http://manu.sporny.org/2011/uber-comparison-rdfa-md-uf>

## Notas

1. Puede consultarse una traducción de dicho documento en:

<http://skos.um.es/Incubator/ld/XGR-ld-vocabdataset>

2. <http://thedatahub.org>

3. <http://data.gov.uk>

4. <http://schema.org>

5. El esquema completo puede consultarse en: <http://schema.org/docs/full.html>

6. <http://schema.org/docs/documents.html>

7. <http://freebase.com>

8. El manual de MQL también constituye una buena referencia de la arquitectura de *Freebase/Metaweb*: [http://wiki.freebase.com/wiki/MQL\\_Manual](http://wiki.freebase.com/wiki/MQL_Manual)

9. Más información sobre *software as a service* en: [http://en.wikipedia.org/wiki/Software\\_as\\_a\\_service](http://en.wikipedia.org/wiki/Software_as_a_service)

10. <http://kasabi.com>

11. <http://www.talis.com/platform>

12. <http://kasabi.com/doc/api>

13. <http://www.w3.org/html/wg/tracker/issues/76>

14. La clase política y la sociedad en general, ya está haciéndose eco de esta necesidad:

<http://www.elmundo.es/elmundo/2011/12/19/navegante/1324300705.html>

\* \* \*

## Supermercados de linked data

Tomás Saorín



No me convence la traducción de "mercados linked data". Entendemos "mercado" como un concepto económico con unas leyes, agentes y dinámicas de interacción.

El término original es "marketplace", que es más tangible: la plaza del mercado, el mercado de abastos.

Quizá empleando "supermercados de linked data" nos acercamos mejor a la idea: lugares fácilmente accesibles al consumidor, con una política de precios clara y barata, con mucho trasiego y una oferta surtida.

El concepto es clave: los datos que pueden ser adquiridas de forma cómoda y transparente para producir servicios.

[tsp@um.es](mailto:tsp@um.es)