

Guía Breve de Linked Data

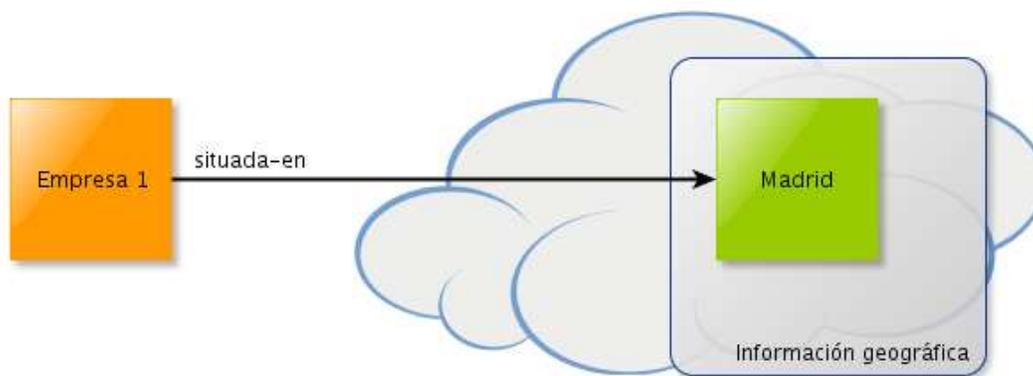
¿Qué es Linked Data?

Los Datos Enlazados es la forma que tiene la Web Semántica [DIVULGACION/GUIASBREVES/WEBSEMANTICA] de vincular los distintos datos que están distribuidos en la Web, de forma que se referencian de la misma forma que lo hacen los enlaces de las páginas web.

La Web Semántica no se trata únicamente de la publicación de datos en la Web, sino que éstos se pueden vincular a otros, de forma que las personas y las máquinas puedan explorar la web de los datos, pudiendo llegar a información relacionada que se hace referencia desde otros datos iniciales.

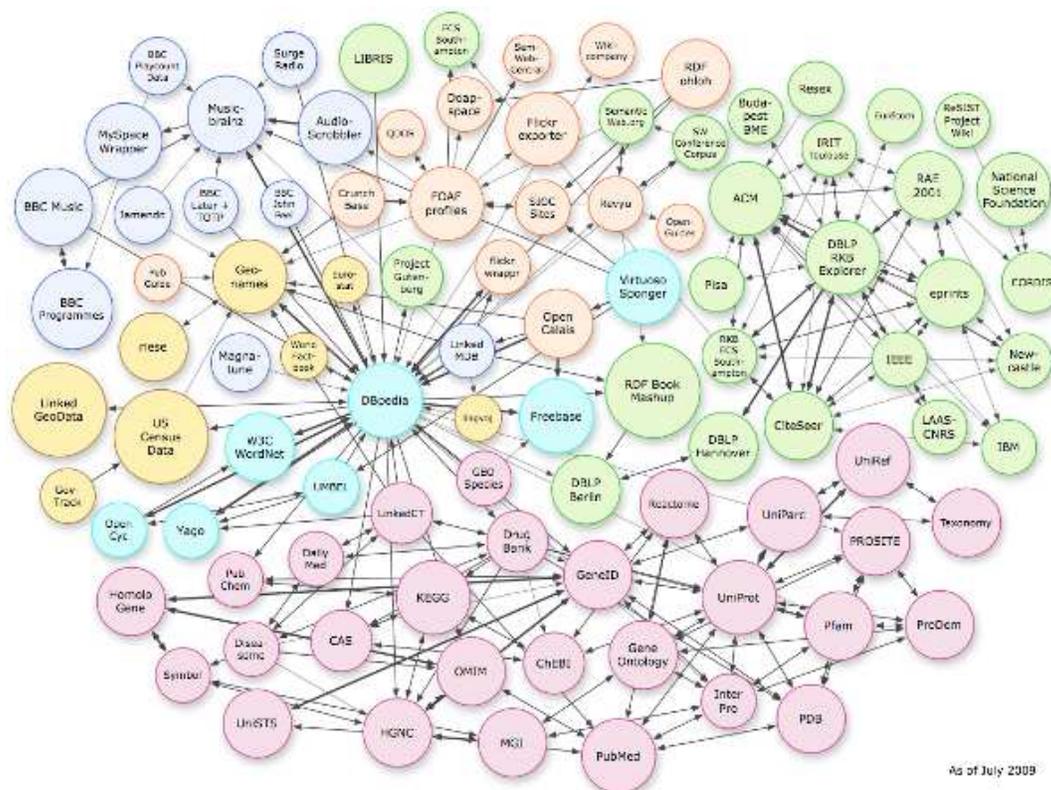
De la misma forma que la web de la web del hipertexto, la web de los datos se construye mediante documentos en la web. Sin embargo, y a diferencia de la web del hipertexto, donde los enlaces son relaciones entre puntos de los documentos escritos en HTML (LENGUAJE DE MARCADO DE HIPERTEXTO), los datos enlazan cosas arbitrarias que se describen en RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS).

Por ejemplo, suponiendo que un directorio de empresas publica información especializada relativa a las organizaciones, como su tamaño o área profesional, es posible que desee indicar también información sobre la localización. Ya que en la web existen sitios con grandes bases de datos geográficas, con información pormenorizada sobre las localizaciones, el directorio de empresas puede hacer referencia a los datos geográficos que están dispuestos por esa fuente externa. De esta forma, los datos iniciales de la organización se enriquecen con información que ofrecen los expertos en el ámbito geográfico.



¿Para qué sirve?

Linked Data permite construir la Web de los datos, una gran base de datos interconectados y distribuidos en la Web. Los datos se vinculan y se exploran de una forma similar a la utilizada para vincular los documentos HTML (LENGUAJE DE MARCADO DE HIPERTEXTO).



[[HTTP://ESW.W3.ORG/SWEIOG/TASKFORCES/COMMUNITYPROJECTS/LINKINGOPENDATA](http://ESW.W3.ORG/SWEIOG/TASKFORCES/COMMUNITYPROJECTS/LINKINGOPENDATA)]

Este grafo representa a distintos conjuntos de datos de diversos tipos, organizados mediante colores por dominios. Estos conjuntos de datos están conectados entre sí de forma que componen la "Nube de Linked Data" o "Nube de Datos Enlazados".

¿Cómo funciona?

Los Datos Enlazados, como parte de la Web Semántica, se basa en la aplicación de ciertos principios básicos y necesarios, que fomentarán el crecimiento de la Web, tanto a nivel de los documentos HTML (LENGUAJE DE MARCADO DE HIPERTEXTO) (vista clásica de la Web), como a nivel de los datos expresados en RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS) (vista de la Web Semántica).

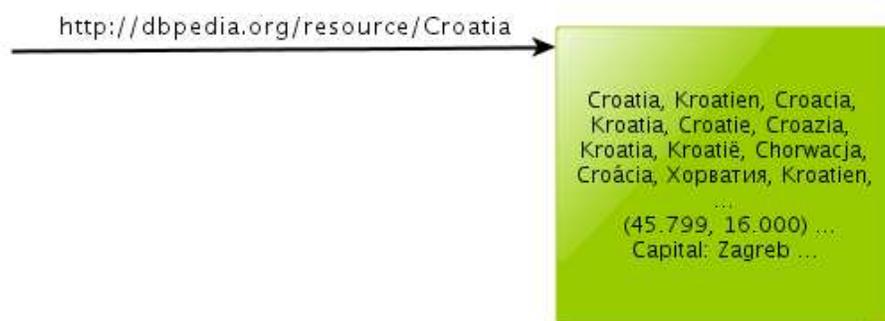
1. Usar URI (IDENTIFICADOR DE RECURSOS UNIFORME) s para identificar las cosas
2. Usar URI (IDENTIFICADOR DE RECURSOS UNIFORME) s HTTP
3. Ofrecer información sobre los recursos usando RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS)
4. Incluir enlaces a otros URI () s

Para conseguir tener los datos interconectados, como si la Web fuese una gran base de datos, se deben respetar los cuatro pasos anteriores para conseguir tener los datos interconectados. Gracias a esta interconexión, se permite reutilizar la información de cualquier manera esperada o inesperada, lo que ofrece un valor añadido a la Web.

Usar URI () s para identificar las cosas

Al nombrar los conceptos o cosas mediante URI () s, se ofrece una abstracción del lenguaje natural y así se consigue evitar ambigüedades y así ofrecer una forma estandar y unívoca para referirnos a cualquier recurso.

Un ejemplo de esto puede observarse en la información geográfica. Las posición de los lugares puede representarse mediante coordenadas, información que puede ser fácilmente interpretable por las personas o de forma automática. El problema surge cuando debemos referirnos a un lugar por su topónimo, ya que éste puede variar en función del idioma (Croacia, Croatia, Hrvatska, क्रोशिया, etc), de su representación (Republika Hrvatska, Rep. de Croacia, etc.), u otros factores (Croacia, antigua República Yugoslava). Si usásemos el nombre para referirnos a los lugares, las múltiples acepciones que podría adoptar, dificultaría el tratamiento automatizado de la información.



De esta forma, si utilizamos un identificador único como <http://dbpedia.org/resource/Croatia> [[HTTP://DBPEDIA.ORG/RESOURCE/CROATIA](http://dbpedia.org/resource/Croatia)], cualquier aplicación se podría referir al mismo lugar, independientemente de la ambigüedad del lenguaje natural.

Usar URI₀s HTTP

Ya que existen muchos esquemas de URI₀s, se pretende el uso de URI₀s sobre HTTP (p.e., <http://dbpedia.org/resource/Croatia> [[HTTP://DBPEDIA.ORG/RESOURCE/CROATIA](http://dbpedia.org/resource/Croatia)]) para asegurar que cualquier recurso pueda ser buscado y accedido en la Web. Debe tenerse en cuenta que los URI₀s no son sólo direcciones, son identificadores de los recursos.

Ofrecer información sobre los recursos usando RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS)

Una vez que se busca y se accede a un recurso identificado mediante una URI₀ HTTP, se debe obtener información útil sobre dicho recurso, representada mediante descripciones estándares en RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS). Se pretende que para cualquier conjunto de datos o vocabulario, se ofrezca información relativa a la información que representa.

De esta forma, si una aplicación desea obtener información sobre un concepto identificado mediante una URI₀, cuando hace una llamada HTTP para obtener el recurso, debería obtener información fácilmente procesable en formato RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS). De la misma forma, si se proveen puntos de consulta avanzada, como SPARQL, el resultado ante una consulta podrá ser interpretado de forma automática.



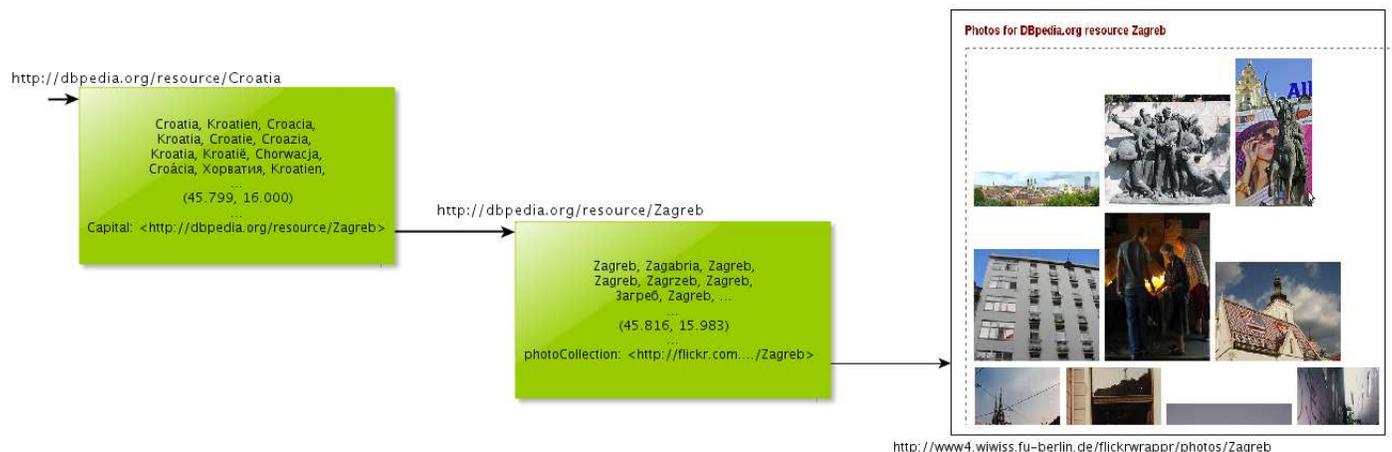
El recurso que identifica a "Croacia" incluye información sobre esta nación. Si cualquier aplicación quisiera hacer uso de la información que provee la DBpedia [[HTTP://DBPEDIA.ORG](http://dbpedia.org)] sobre este recurso, simplemente debería obtener el recurso identificado como "<http://dbpedia.org/resource/Croatia> [[HTTP://DBPEDIA.ORG/RESOURCE/CROATIA](http://dbpedia.org/resource/Croatia)]" y obtendría todos los datos relativos a Croacia (población, extensión del país, topónimos en distintos idiomas, etc.).

Incluir enlaces a otros URI₀s

La cuarta regla, enlazar datos en cualquier lugar, es necesaria para conectar los datos que tenemos en sitios web de forma que no se queden aislados y así se pueda compartir información con otras fuentes externas y que otros sitios puedan enlazar los datos propios de la misma forma que se hace con los enlaces en HTML (LENGUAJE DE MARCADO DE HIPERTEXTO).

A través de la utilización de enlaces a recursos provenientes de sitios más especializados en determinados dominios, se ofrece un valor añadido a la información que se provee.

Algo a tener en cuenta es que los enlaces de los recursos mediante URI ()s, pueden hacerse localmente y a través de toda la red. Por ejemplo, el recurso de la DBpedia [HTTP://DBPEDIA.ORG] que representa a Croacia, puede tener una propiedad que representa la capital del país, Zagreb, que también está representada mediante RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS) e identificada por un URI () unívoco similar al de Croacia. En este caso, Zagreb se representa como <http://dbpedia.org/resource/Zagreb> [HTTP://DBPEDIA.ORG/RESOURCE/ZAGREB]. De esta forma, ya aparecen dos recursos enlazados, aunque se encuentran en el mismo servidor.



Algunos de los recursos relacionados con los lugares son las fotografías, por lo que el recurso “Croacia” y “Zagreb” incluyen una propiedad que enlaza el lugar con las fotografías relacionadas, que se encuentran almacenadas en un almacén como Flickr [HTTP://WWW.FLICKR.COM], que a su vez ofrece su información en RDF (). De esta forma, se enlazan los datos de lugares con las fotografías.

Gracias a estos mecanismos, cualquier recurso es susceptible de ser enriquecido con cualquier tipo de información especializada, incluso la que no se espera que sea combinable. De la forma inversa, al publicar información en RDF () y utilizando URI ()s, cualquiera podría hacer referencia a esos datos.

Ejemplos

El grupo de comunicación **BBC** [HTTP://WWW.BBC.CO.UK/EN] utiliza tecnologías semánticas basadas en los principios de Linked Data para sitios como **BBC Music** [HTTP://WWW.BBC.CO.UK/MUSIC/EN] (música), **BBC Programmes** [HTTP://WWW.BBC.CO.UK/PROGRAMMES/EN] (programas), o **BBC Wildlife Finder** [HTTP://WWW.BBC.CO.UK/NATURE/ANIMALS/] (fauna salvaje) confían en conjuntos de datos externos, que están publicados en la web. Estos conjuntos de datos proceden de sitios como Musicbrainz [HTTP://MUSICBRAINZ.ORG/EN] (base de datos sobre música) o la Wikipedia [HTTP://EN.WIKIPEDIA.ORG/EN] (o su abstracción semántica, DBpedia).

Los sitios de la **BBC** muestran una vista agregada de esta información obtenida de varias fuentes y puesta en el contexto de la **BBC**. Además, la **BBC** también crea contenido semántico, identificado mediante URI ()s, de cualquier elemento que se considera interesante. Además, al crear representaciones en RDF (INFRAESTRUCTURA DE DESCRIPCIÓN DE RECURSOS) de estos identificadores, se permite a cualquier desarrollador que aproveche estos datos de la **BBC** para construir sus propias aplicaciones.

BBC Music ofrece un identificador para cualquier artista que pueda aparecer en los programas de la **BBC**, eventos de la **BBC**, etc. y así ser reutilizada su información. **BBC Music** se sustenta gracias a la base de datos de música Musicbrainz [HTTP://MUSICBRAINZ.ORG/EN] y la Wikipedia [HTTP://EN.WIKIPEDIA.ORG/EN], aprovechando la información existente en estas fuentes de datos, mejorándola y enriqueciéndola con enlaces dentro del sitio de la **BBC**. **BBC Music** considera a la Web como un sistema gestor de contenidos con la información distribuida.

BBC Music combina sus propios datos con los obtenidos desde Musicbrainz y la Wikipedia

Más información sobre Linked Data

Guía Breve de Web Semántica [./WEBSEMANTICA]

Guía básica en castellano sobre Web Semántica en general.

Linked Data - Design Issues [HTTP://WWW.W3C.ORG/DESIGNISSUES/LINKEDDATA.HTMLEN]

Aspectos básicos de diseño que definen Linked Data y la forma de aplicar estos principios (en inglés).

Proyecto LinkingOpenData [HTTP://ESW.W3C.ORG/SWEOIG/TASKFORCES/COMMUNITYPROJECTS/LINKINGOPENDATAEN]

Proyecto, dentro de la Comunidad de Educación y Difusión de la Web Semántica del W3C (en inglés).

Tutorial Linked Data [HTTP://WWW.W3C.ORG/2010/TALKS/0525-LINKED-DATA/EN]

Presentación de un tutorial del W3C sobre Linked Data (en inglés).

Actividad de Web Semántica del W3C [HTTP://WWW.W3C.ORG/2001/SW/EN]

Toda la información sobre Web Semántica en el W3C (en inglés).

Wiki con recursos sobre la Web Semántica del W3C [HTTP://ESW.W3C.ORGEN]

Wiki con herramientas, ayuda, tutoriales, etc. acerca de la Web Semántica (en inglés).