

Introducción a la Web Semántica

Jose Emilio Labra Gayo

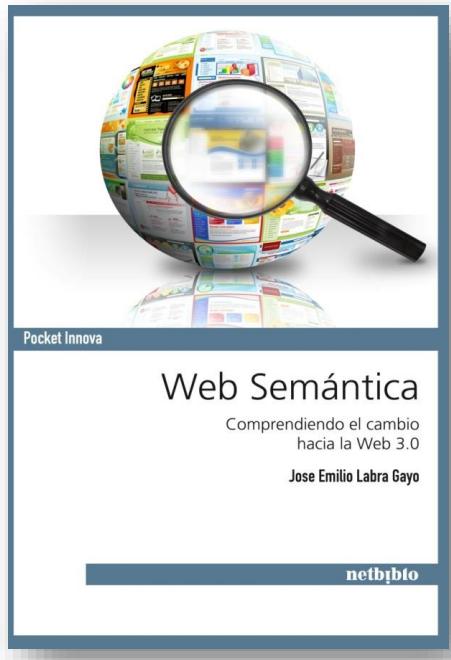
<http://labra.weso.es>

Grupo investigación WESO (WEb Semantics Oviedo)

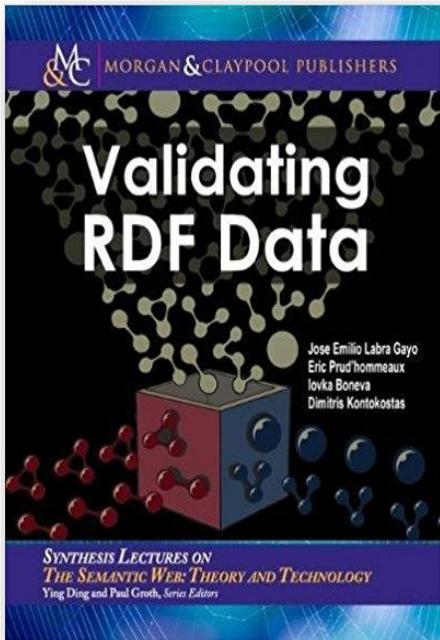
Departamento de Informática

Universidad de Oviedo

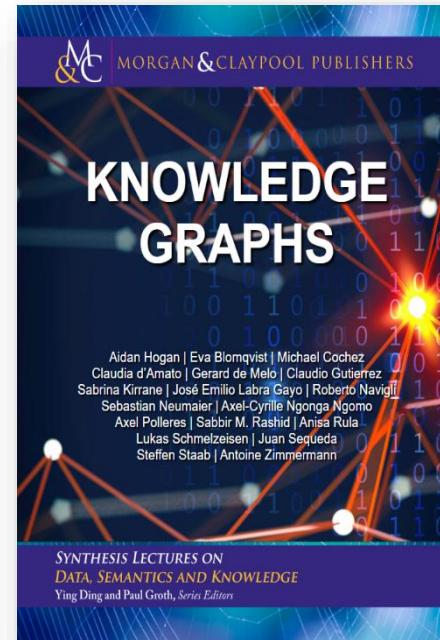
Referencias



Web Semántica
Editorial NetBiblo,
Colección Pcket Innova, 2012
<http://www.netbiblo.com>



Validating RDF Data
Ed. Springer Nature, 2018
<http://book.validatingrdf.com>
Online HTML version



Knowledge graphs
Ed. Springer Nature, 2021
<https://kgbook.org/>
Online HTML version

Web semántica

Visión de la Web como **web de datos**

No solo páginas/documentos web, sino datos

Datos enlazados

Agentes que procesan esos datos automáticamente

Campo relacionado con:

Big Data

Enormes cantidades de datos de la Web

...y muchos más datos que se van a generar!

Inteligencia Artificial

Representación del conocimiento

Inferencia de nuevo conocimiento



Tim Berners-Lee
Fuente: Wikipedia

¿Quién consume información de la Web?

Personas

Accedemos a través de un dispositivo

...y **Máquinas** (bots)

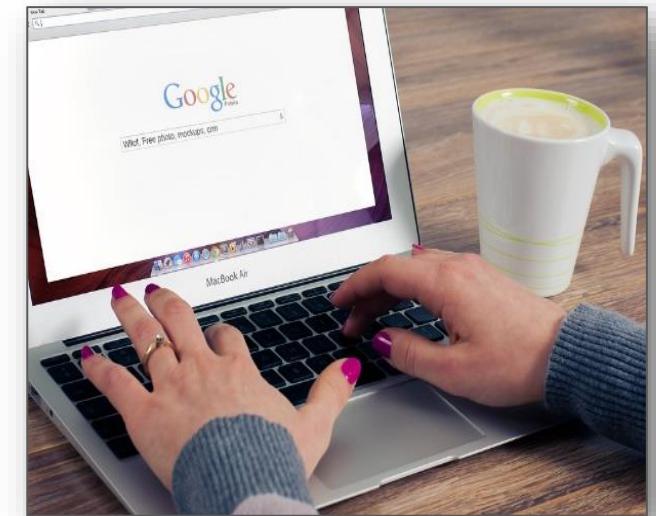
Nos muestran las páginas Web (navegadores)

...pero también analizan la información

Nos filtran contenido

Nos hacen sugerencias

...

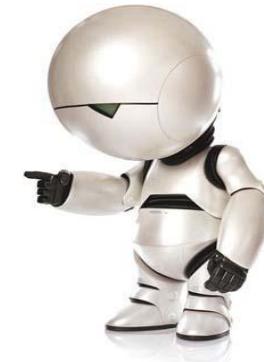


Si Google no "entiende" tu página Web es como si
no existiese

¿Personas vs Máquinas?



Creatividad, imaginación
Imprevisibles (cometemos errores)
Nos cansamos ante tareas repetitivas
Comprensión basada en contexto



Programadas para ciertas tareas
Previsibles (sin errores*)
Tareas repetitivas sin problema
Dificultad para entender el contexto

¿Información *entendible* por las máquinas?

Problema: Ambigüedad e identificación del contexto

Ejemplo: "Oviedo tiene una temperatura de 36 grados"

Oviedo Puede ser:
... Una ciudad en España
...o una ciudad en Florida
...o un jugador de fútbol

...tiene una temperatura de...

Representación para máquinas

Identificadores (URLs) en Wikidata

<http://wikidata.org/wiki/Q21414421>

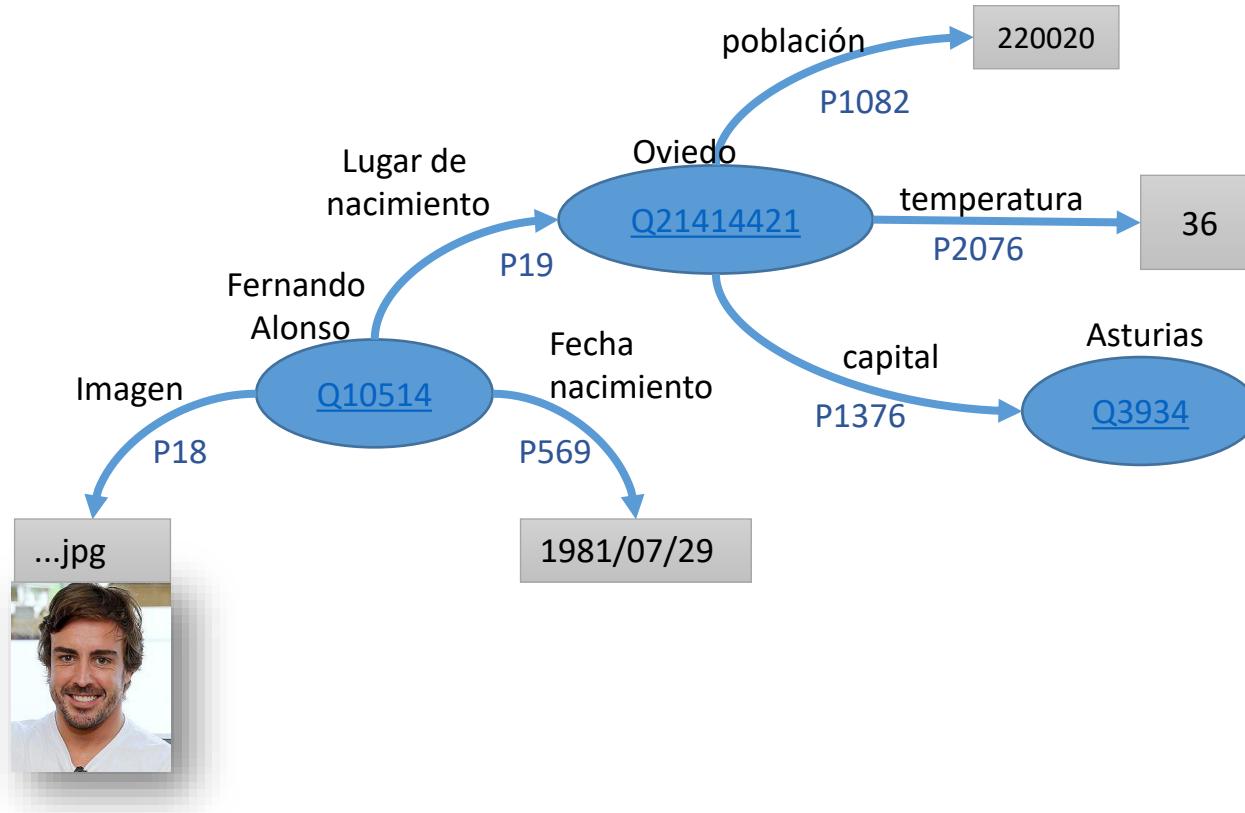
<http://www.wikidata.org/wiki/Q1813449>

<http://www.wikidata.org/wiki/Q325997>

<https://www.wikidata.org/wiki/Property:P2076>



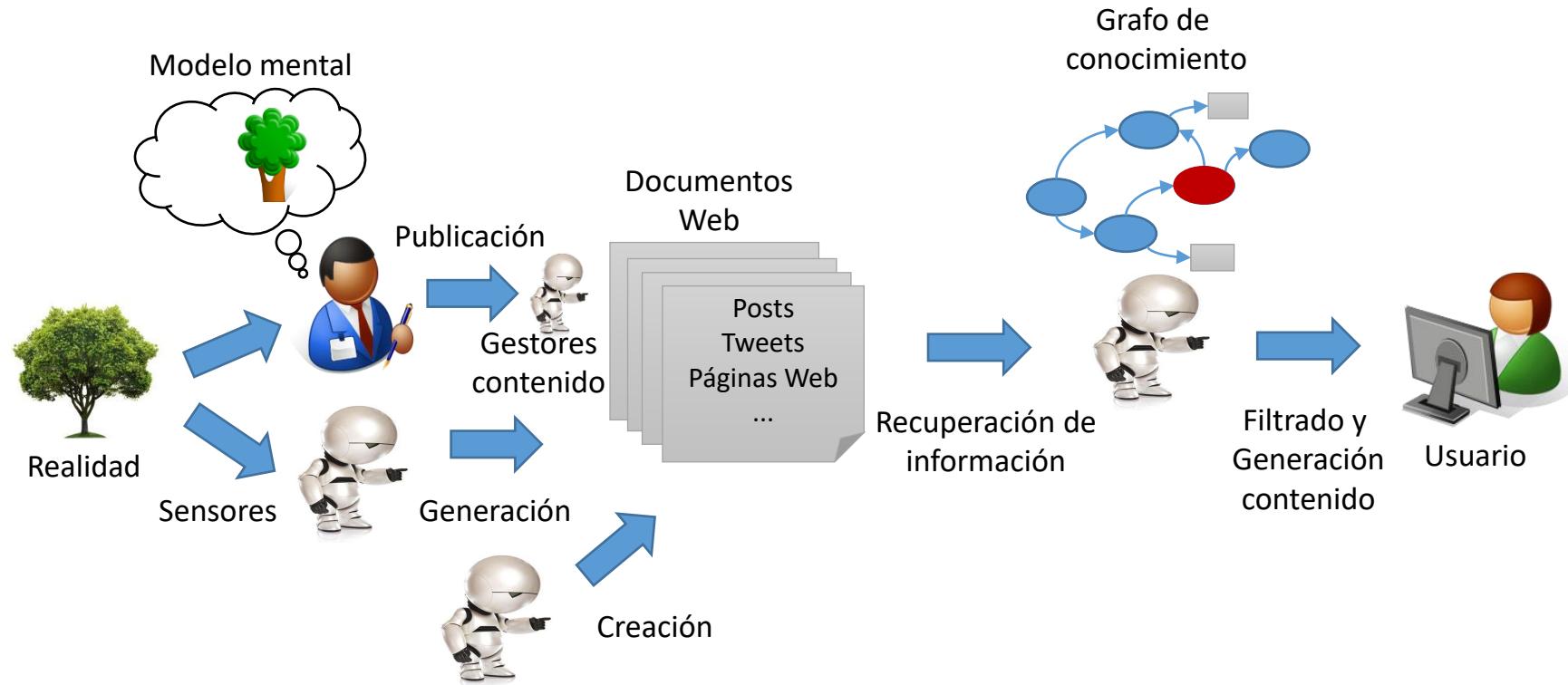
Representación de conocimiento en grafos



Estadísticas sobre Wikidata (09/05/2023)
102,972,875 items
12,671,768,950 tripletas

Wikidata es un ejemplo de grafo de conocimiento
...pero hay muchos más...

Máquinas en la Web



Información en la Web manipulada constantemente por máquinas
Web Semántica \Rightarrow facilitar esa manipulación

Características de la web

No centralizada

Difícil garantizar integridad de la información

Información Dinámica

La información existente cambia

Mucha información

Un sistema no puede pretender acaparar toda la información

Es abierta

Open World Assumption

Principio CCC

Cualquiera puede decir Cualquier cosa sobre Cualquier tema

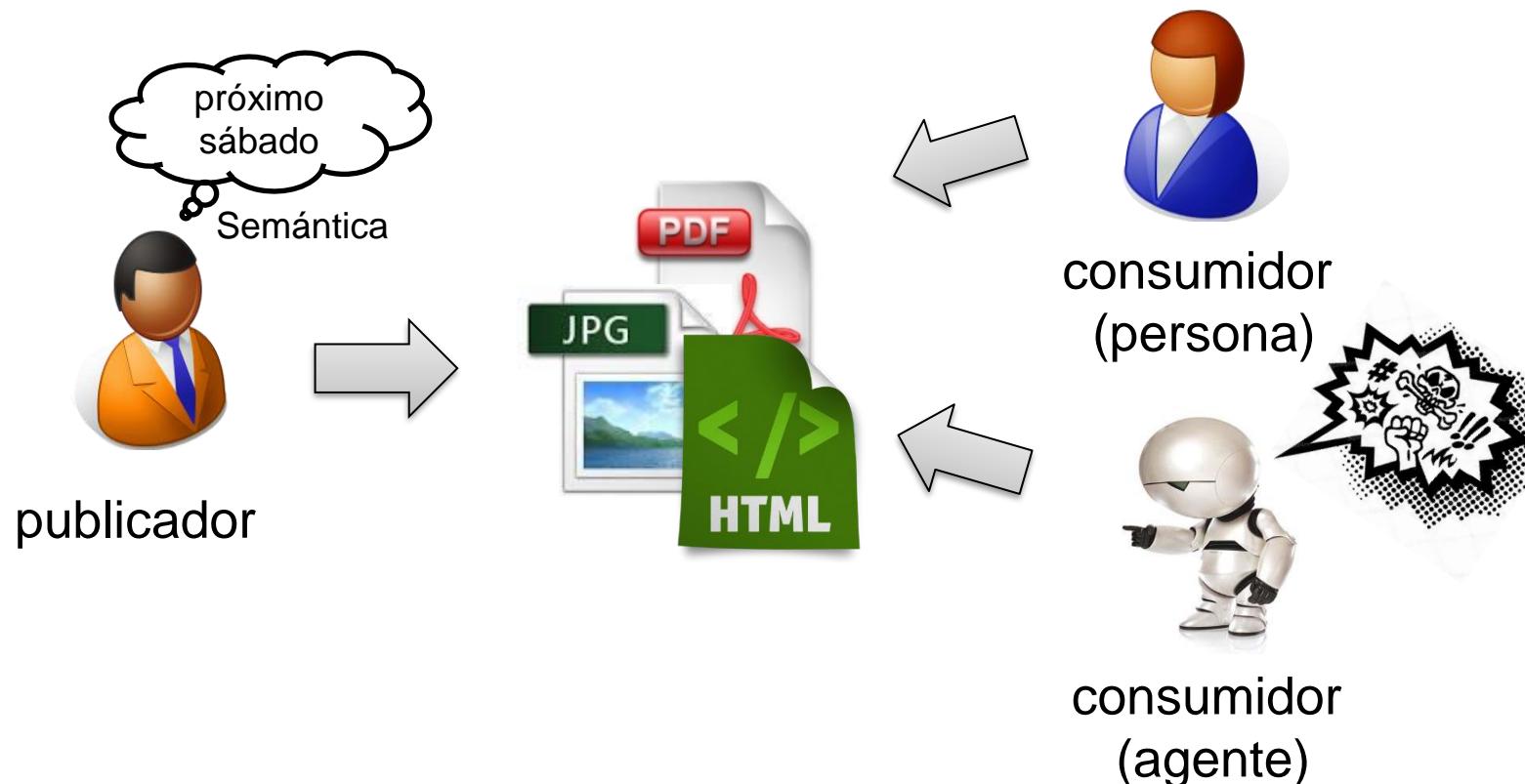
En inglés: Principio AAA: Anyone can say Anything about Any topic
Fuente: Semantic Web for the Working Ontologist, D. Allemang, J. Hendler

Problema de la pérdida semántica

Pérdida de semántica en el proceso de publicación

La persona que va a publicar tiene más información

Esa información se pierde en el proceso



El problema de HTML

HTML tiene como objetivo publicar hipertexto

Etiquetas HTML legibles por los navegadores

Información dentro de marcas = lenguaje natural

Las máquinas no entienden el lenguaje natural

```
<p>Evento:  
<ul>  
  <li>Nombre: Concierto</li>  
  <li>Fecha: Próximo sábado</li>  
</ul>  
</p>
```

```
<p>իրադարձություն:  
<ul>  
  <li>տիպ: համերգ</li>  
  <li>ամսաթիվ: հաջորդ շաբաթ</li>  
</ul>  
</p>
```

El problema de XML

XML da un paso más hacia la solución

Se pueden definir vocabularios específicos

Pueden crearse aplicaciones que los procesan

Sin embargo, los documentos XML no se integran fácilmente si son de otros vocabularios

```
<event>
  <name>Concierto</name>
  <date>Próximo sábado</date>
</event>
```

```
<event>
  <name>համերգ</name>
  <date>հաջորդ շաբաթ</date>
</event>
```

```
<իրադարձություն>
  <տիպ>համերգ </տիպ>
  <ամսաթիվ>հաջորդ շաբաթ</ամսաթիվ >
</իրադարձություն>
```

¿Y JSON?

Más o menos...lo mismo que XML

JSON tiene un modelo jerárquico similar a XML

Los nombres de los campos son cadenas de texto

```
{ "event":  
  { "name": "Concierto" ,  
    "date": "Próximo sábado"  
  }  
}
```

```
{ "event":  
  { "name": "համերգ" ,  
    "date": "հաջորդ շաբաթ"  
  }  
}
```

```
{ "իրադարձություն":  
  { "տիպ": "համերգ" ,  
    "ամսաթիվ": "հաջորդ շաբաթ"  
  }  
}
```

RDF

Formato basado en grafo

Unidad básica = tripleta (sujeto, predicado, objeto)

Base para representación de conocimiento

Los predicados son URIs globales

```
prefix schema: <http://schema.org/>
prefix xsd:     <http://www.w3.org/2001/XMLSchema#>
prefix :         <http://example.org/>

:myEvent a schema:Event ;
    schema:startDate "2014-05-23T20:00"^^xsd:dateTime .
```

Estado actual de la web semántica

La web semántica fue propuesta en 2001

¿Dónde está?



3 ejemplos
Grafos de conocimiento
Wikidata
Datos abiertos enlazados

Grafos de conocimiento

- Noción actual: Popularizada por Google, 2012*

The screenshot shows a Google search results page for the query "tim berners lee". At the top, there's a search bar with the query, a microphone icon, a camera icon, and a magnifying glass icon. Below the search bar are navigation links for "Todo", "Imágenes", "Noticias", "Vídeos", "Libros", and "Más", followed by "Herramientas". A status message indicates "Aproximadamente 4.040.000 resultados (0,35 segundos)".

The main result is a knowledge graph card for "Tim Berners-Lee", described as a "Científico de la computación". The card features a profile picture of Tim Berners-Lee and five smaller images below it. To the right of the images is a section titled "Información" containing his biography, birthplace, education, awards, children, parents, and spouse.

Below the knowledge graph card, there's a link to "Tim Berners-Lee - Wikipedia, la enciclopedia libre" with a snippet of text about him. Further down, there are sections for "Desarrollo de su carrera", "Proyectos actuales", and "Publicaciones". At the bottom of the card, there's a section for "Otras preguntas de los usuarios".

Información

Timothy "Tim" John Berners-Lee, es un científico de la computación británico, conocido por ser el padre de la World Wide Web. Estableció la primera comunicación entre un cliente y un servidor usando el protocolo HTTP en diciembre de 1990. [Wikipedia](#)

Nacimiento: 8 de junio de 1955 (edad 67 años), Londres, Reino Unido

Premios: Premio de Tecnología del Milenio, Premio Japón, MÁS

Educación: The Queen's College (1973–1976), Emanuel School (1969–1973), Sheen Mount Primary School

Hijos: Ben Berners-Lee, Alice Berners-Lee

Padres: Conway Berners-Lee, Mary Lee Woods

Cónyuge: Rosemary Leith (m. 2014), Nancy Carlson (m. 1990–2011)

<https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>

Grafos de conocimiento



Grafos de conocimiento

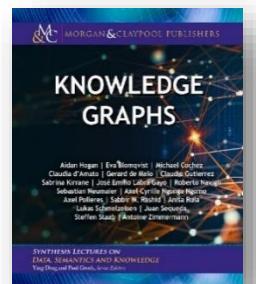
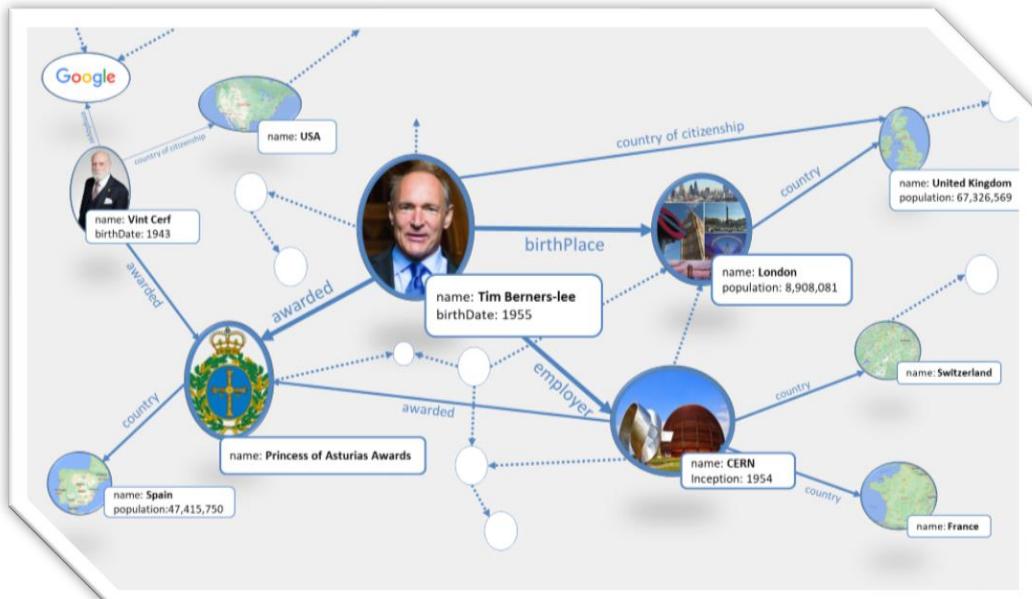
Grafo de conocimiento = un **grafo de datos**

que pretende acumular y transmitir

Conocimiento del mundo real

*cuyos nodos representan **entidades** de interés y*

*cuyas aristas representan **relaciones** entre entidades*



<https://kgbook.org/>

Aplicaciones de grafos de conocimiento

Mejorar resultados de búsqueda

Responder preguntas

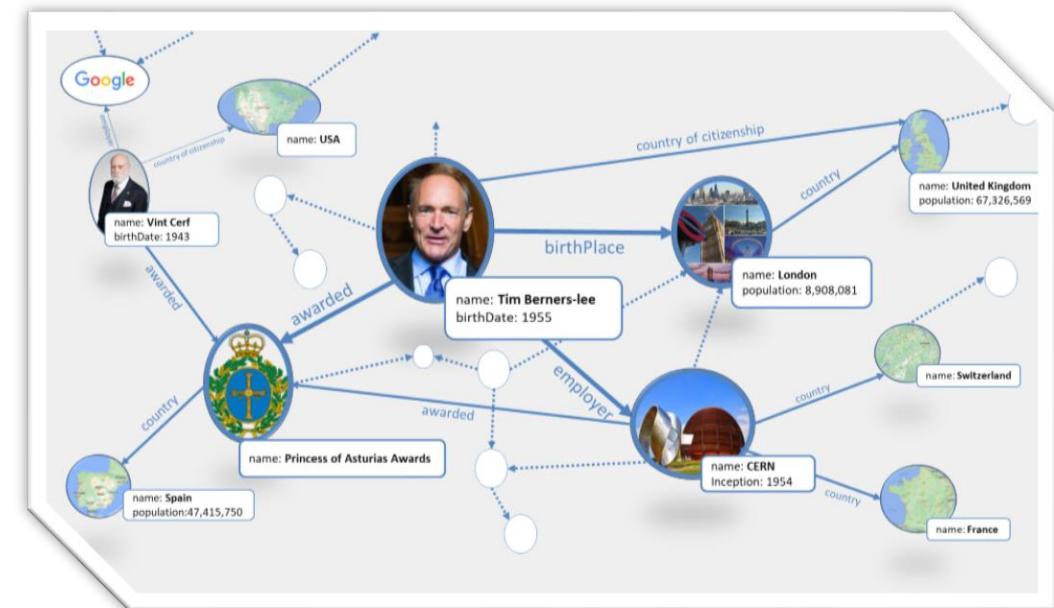
Gobierno de datos

Gestionar datos heterogéneos

Sistemas de recomendación

Chatbots y NLP

...



Google knowledge panel

Google search results for "oviedo".

Search bar: oviedo

Navigation: All | Images | Maps | News | Videos | More | Settings | Tools

About 78,400,000 results (1.05 seconds)

<https://www.oviedo.es> ▾ Translate this page
Ayuntamiento de Oviedo - oviedo.es
Web institucional del Ayuntamiento de Oviedo: Toda la información política, administrativa, cultural y festiva.

<https://en.wikipedia.org/wiki/Oviedo> ▾
Oviedo - Wikipedia
Oviedo is the capital city of the Principality of Asturias in northern Spain and the administrative and commercial centre of the region. It is also the name of the ...
Country: Spain Postal code: from 33001 to 33013
Autonomous community: Asturias Province: Asturias
History · Architecture · Culture · Climate

Images for oviedo
asturias, spain, downtown, cathedral, ciudad

A grid of 12 images showing various landmarks and scenes from Oviedo, including the cathedral, streets, and modern architecture.

Report images

Oviedo
City in Spain

Oviedo is a town in northwest Spain between the Cantabrian Mountains and the Bay of Biscay. The capital of Asturias, it's known for its medieval old town, the site of the Gothic Oviedo Cathedral with its 9th-century Holy Chamber. The Archaeological Museum of Asturias, in a nearby convent, displays regional artifacts. The Fine Arts Museum of Asturias includes 2 palaces and has an expansive Spanish art collection. — Google

Elevation: 232 m

Weather: 11°C, Wind N at 11 km/h, 93% Humidity [weather.com](#)

Population: 220,020 (2018) [Instituto Nacional de Estadística](#)

Local time: Monday 17:42

Province: [Asturias](#)

Official language(s): Spanish

Plan a trip

Things to do

3-star hotel averaging €42, 5-star averaging €92

Upcoming Events

University: [University of Oviedo](#)

Google rich snippets



recetas de cachopo



<https://www.pequerecetas.com> › receta ▾ [Translate this page](#)

[Recetas CACHOPO ASTURIANO. Cómo hacer cachopos de ...](#)



Cómo hacer un **cachopo** asturiano de ternera, uno de los platos típicos de Asturias. **Receta** tradicional paso a paso para hacer **cachopos** de forma fácil.

★★★★★ Rating: 4.4 · 7 votes · 25 min

[Cómo hacer cachopo de...](#) · [Consejos para hacer...](#)

<https://recetasdecocina.elmundo.es> › ... ▾ [Translate this page](#)

[Cachopo de ternera. Receta asturiana paso a paso | Receta ...](#)



Feb 19, 2020 — **Receta** asturiana paso a paso. El **cachopo** asturiano, una maravillosa comida tradicional de la cocina asturiana que cada vez cobra más ...

★★★★★ Rating: 3 · 37 reviews · 30 min

<https://www.directoalpaladar.com> › r... ▾ [Translate this page](#)

[Cachopo asturiano: la receta de escalope de ternera con ...](#)



May 1, 2021 — **Cachopo** de ternera. **Receta** tradicional asturiana, con fotos paso a paso del proceso de elaboración. Consejos para que el empanado quede ...

★★★★★ Rating: 4.3 · 14 reviews · 25 min

<https://www.recetasdeescandalo.com> › ... ▾ [Translate this page](#)

[Cachopo asturiano de ternera con pimientos del piquillo y ...](#)



Aprende a preparar esta **receta** típica asturiana siguiendo todos los trucos y consejos para que consigas una ternera jugosa, un rebozado crujiente y el queso ...

★★★★★ Rating: 3.3 · 21 votes · 1 hr

Cómo conoce
Google el rating,
tiempo, los votos,
etc...

Schema.org

Los publicadores añaden conocimiento estructurado
Motivación: más visitas (SEO)
Google puede generar mejores resultados

```
<div itemscope itemtype="https://schema.org/Recipe">
  <span itemprop="name">Cachopo asturiano</span> de
    <span itemprop="author">Juanón</span>,
    Tiempo: <meta itemprop="prepTime" content="PT25M">25 minutos
    <div itemprop="nutrition"
      itemscope itemtype="https://schema.org/NutritionInformation">
      Nutrición:
        <span itemprop="calories">240 calorías</span>,
        <span itemprop="fatContent">9 gramos grasa</span>
    </div>
    . . .
</div>
```

<https://twitter.com/linkedktk/status/1654490911601053708>



Adrian Gschwend
@linkedktk

...

50% of all websites in common crawl dataset contain triples in some form. Semantic Web won, people just didn't notice.

webdatacommons.org/structureddata/

Traducir Tweet

3.2. Extraction Results from the October 2022 Common Crawl Corpus	
The October 2022 Common Crawl Corpus is available on Amazon S3 in the bucket commoncrawl under the key prefix /crawl-data/CC-MAIN-2022-48/.	
Extraction Statistics	
Crawl Date	October 2022
Total Data	82.71 Terabyte (compressed)
Parsed HTML URLs	3,048,746,652
URLs with Triples	1,518,609,988
Domains in Crawl	33,820,102
Domains with Triples	14,235,035
Typed Entities	19,072,628,514
Triples	86,462,816,435
Size of Extracted Data	1.6 Terabyte (compressed)

4:19 p. m. · 5 may. 2023 · 2.963 Reproducciones

<http://webdatacommons.org/structureddata/>

Tipos de grafos de conocimiento

Grafos de conocimiento abiertos

Multi-dominio: Wikidata, Dbpedia, Freebase, YAGO, ...

Dominio específico

Académico: Open citations, SciGraph, Microsoft Academic Knowledge Graph, ...

Ciencias naturales: UniProt, PubChem, PDB, ...

Gobierno: EU Knowledge graph, ...

...

Grafos de conocimiento empresariales

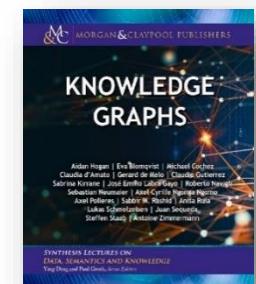
Búsqueda en la web: Google, Bing...

Comercio electrónico: AirBnb, Amazon, eBay, Uber,...

Redes sociales: Linkedin, Facebook,...

Finanzas: Banca d'Italia, Bloomberg, Wells Fargo, Capital One,... <https://kgbook.org/>

...



Wikidata

Wikidata fue creada en 2012 como grafo de conocimiento colaborativo

Objetivo inicial:

Dar soporte a cajas de información multilingüe en Wikipedia

Se está utilizando por numerosos sistemas

Grafo de conocimiento de uso común



Wikidata

Página de Wikipedia en inglés de Tim Berners-Lee

WIKIPEDIA The Free Encyclopedia

Article Talk Read View source View history Search Wikipedia

Tim Berners-Lee

From Wikipedia, the free encyclopedia

Sir Timothy John Berners-Lee, OM, KBE, FRS, FREng, FRSA, DFBCS, RDI (born 8 June 1955),^[1] also known as **TimBL**, is an English computer scientist best known as the inventor of the World Wide Web. He is a Professorial Fellow of Computer Science at the University of Oxford^[2] and a professor at the Massachusetts Institute of Technology (MIT).^{[3][4]} Berners-Lee proposed an information management system on 12 March 1989,^{[5][6]} then implemented the first successful communication between a Hypertext Transfer Protocol (HTTP) client and server via the Internet in mid-November.^{[7][8][9][10][11]}

Berners-Lee is the director of the World Wide Web Consortium (W3C), which oversees the continued development of the Web. He co-founded (with his then wife-to-be Rosemary Leith) the World Wide Web Foundation. He is a senior researcher and holder of the 3Com founder's chair at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL).^[12] He is a director of the Web Science Research Initiative member of the at social ment. He as Foreign if the 100 as the age Next received the ing the Web

Tim Berners-Lee
OM KBE FRS FREng FRSA FBCS



Born Timothy John Berners-Lee 8 June 1955 (age 67) London, England

Other names TimBL TBL

Education The Queen's College, Oxford (BA)

Known for Invention of the World Wide Web

Spouse(s) Nancy Cadeen

Página en Bihari de Tim Berners-Lee

<http://www.wikidata.org/entity/Q80>

WIKIDATA

Tim Berners-Lee (Q80)

English computer scientist, inventor of the World Wide Web (born 1955)

TimBL | Sir Tim Berners-Lee | Timothy John Berners-Lee | TBL | T. Berners-Lee | T Berners-Lee | Tim Berners Lee | T.J. Berners-Lee | Sir Timothy John Berners-Lee

» In more languages

Statements

instance of human edit

+ 1 reference edit

+ add value

image edit



Sir Tim Berners-Lee (cropped).jpg
570 x 713; 178 KB

point in time 2014

media legend Tim Berners-Lee in 2014.
(English)

Tim Berners-Lee i 2014.
(Norwegian Bokmål)

Tim Berners-Lee i 2014.
(Norwegian Nynorsk)

* 0 references edit

+ add reference

+ add value

sex or gender male edit

+ 2 references edit

+ add value

Wikidata: algunas características

Collaborativo: cualquiera pueda editar

Licencia abierta

Actualmente (01/2023): 101m items, 1,8b ediciones

Co-editado por humanos y bots: 23k usuarios activos, 343 bots

Servicio consulta abierto: Public SPARQL endpoint

Dumps disponibles: 109Gb comprimidos

Software que da soporte a Wikidata = Wikibase



Datos abiertos y datos enlazados

Tendencias *Open*

Producir datos es tecnológicamente fácil

Movimientos:

Open Software

Open Content

Open Data

Open Science

Open Government

Viejos modelos afectados

Música, Cine, finanzas,...

¿Educación?

¿Gobierno?

...



¿Porqué?

Razones para los gobiernos

Transparencia

Liderazgo

Gobierno como catalizador

Fomentar participación

Nuevas iniciativas y Apps

Razones para los ciudadanos

Nos pertenecen

Creados con dinero público

Queremos mejores servicios

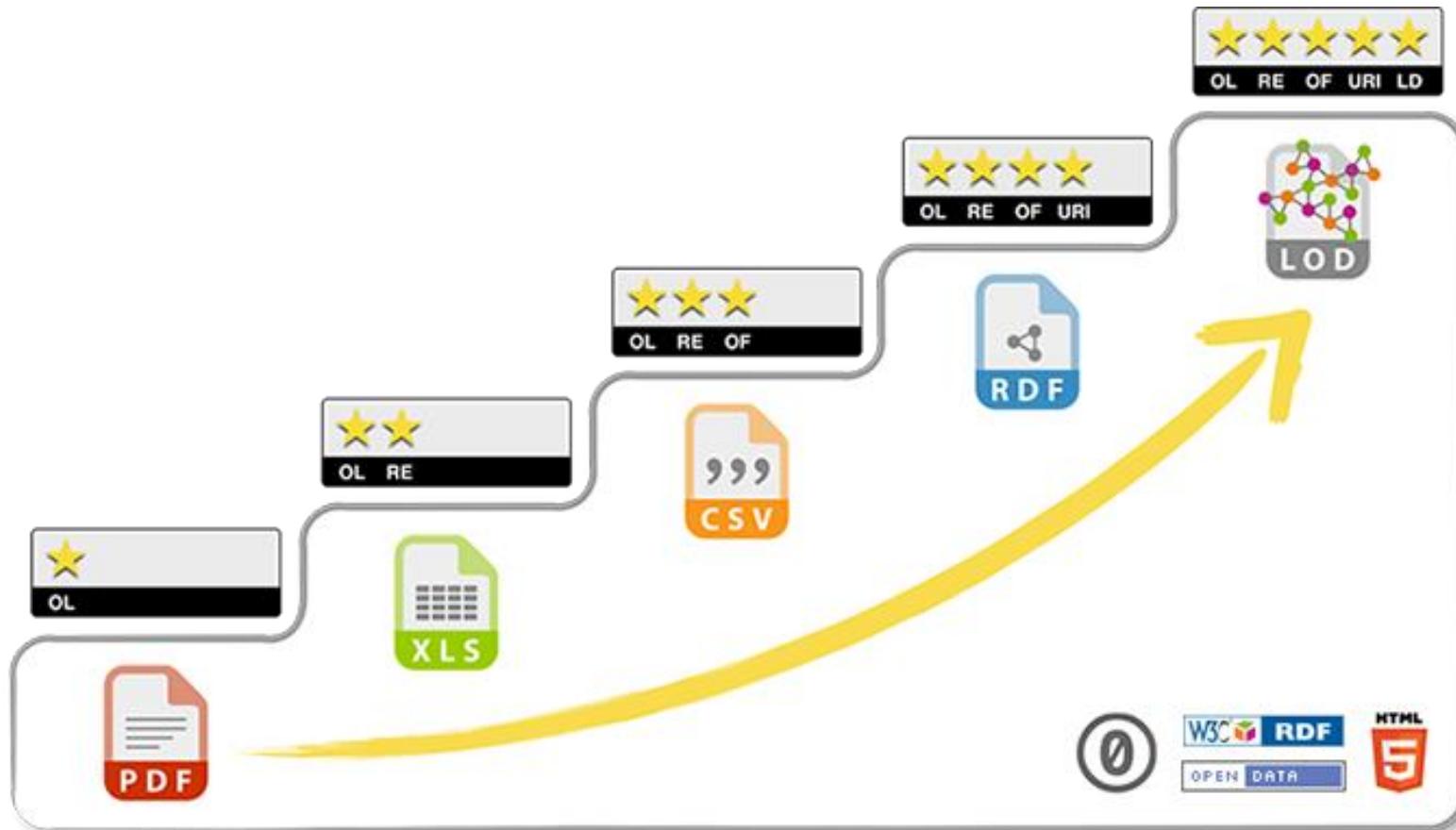


Modelo de Estrellas*

<http://5stardata.info/>

- ★ **Publicar** los datos
(en cualquier formato)
- ★★ Utilizar **formato estructurado**
(Excel en lugar de imágenes escaneadas)
- ★★★ Usar formatos **no propietarios**
(CSV en lugar de Excel)
- ★★★★ Usar **URLs para identificar** datos
(otros sistemas puedan enlazar nuestros datos)
- ★★★★★ **Enlazar con otros** datos externos
(proporcionar contexto)

Modelo de estrellas



Formatos no estructurados



Formatos “caja negra”: Imágenes, videos, música, etc.

Formatos binarios: PDF, PS, etc.

Requieren técnicas de tratamiento de la señal, reconocimiento de patrones, etc.





Formatos estructurados

Los datos tienen una estructura

Ejemplo: Hojas de cálculo

Problema con formatos propietarios

Requieren herramientas que no son públicas

No resuelven el problema de la pérdida semántica

¿Qué significa cada celda?





Formatos no propietarios

Utilizar formatos abiertos estructurados

Ejemplos: CSV, HTML, JSON, XML...

Permiten procesamiento automático

Pero no resuelven el problema de la pérdida semántica

URIs para identificar datos



Utilizar URIs para identificar datos

Negociación de contenido

Devolver diferentes representaciones

Ejemplo:

HTML para personas con navegadores

RDF para sistemas automáticos



Ejemplo: RDF



<http://www.sepe.es/datos/desempleo/Asturias/Allande/2013/10>



¿Varias representaciones de lo mismo?

La arquitectura de la web separa recurso de representación

Ejemplo: Bolsa de patatas fritas



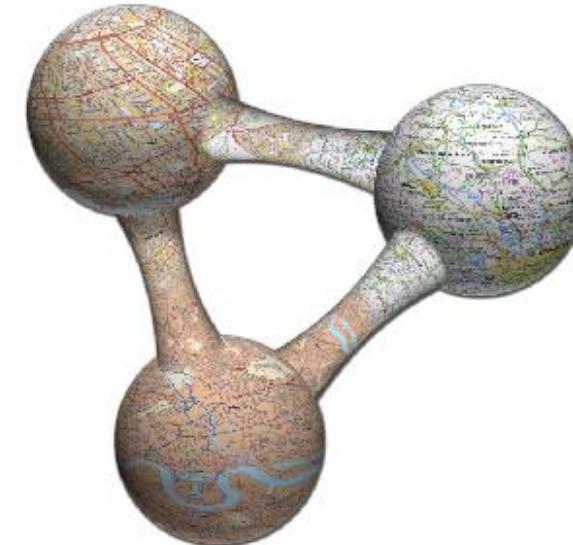
Enlazar con otros datos ★★★★☆

Las representaciones que se devuelven incluyen enlaces a otros datos

Permite:

Reutilizar y descubrir datos

Aplicaciones "*no previstas*"



Ejemplo: RDF bien enlazado



<<http://www.sepe.es/datos/desempleo/Asturias/Allende/2013/10>>

RDF?

HTML?



```
prefix sepe: <http://www.sepe.es/datos/>
prefix dbo: <http://dbpedia.org/ontology/>
```

```
sepe:obs1 sepe:municipio  
sepe:desempleados
```

```
dbo:allande ;  
23 .
```

```
dbo:allande dbo:areaTotal 342.24 ;  
rdf:type <http://.../municipalitiesInAsturias> ;  
dbo:country <http://.../Spain> ;  
dbo:populationTotal 2106 ;  
...
```

Principios Linked Open Data

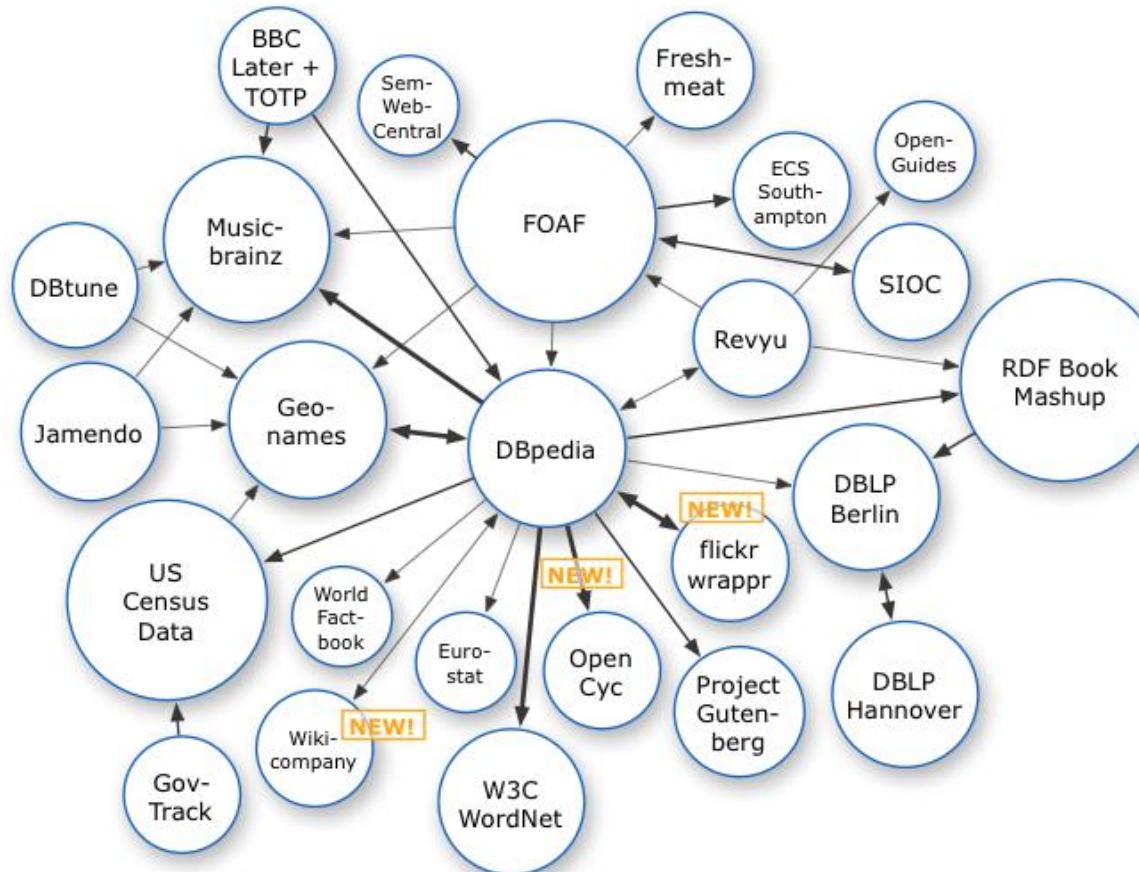


Buenas prácticas para publicar datos en la web

1. Utilizar URIs para denotar cosas
2. Permitir que las URIs sean dereferenciables
3. Proporcionar información útil
Para personas y máquinas (HTML, RDF)
4. Incluir enlaces a otras cosas relacionadas

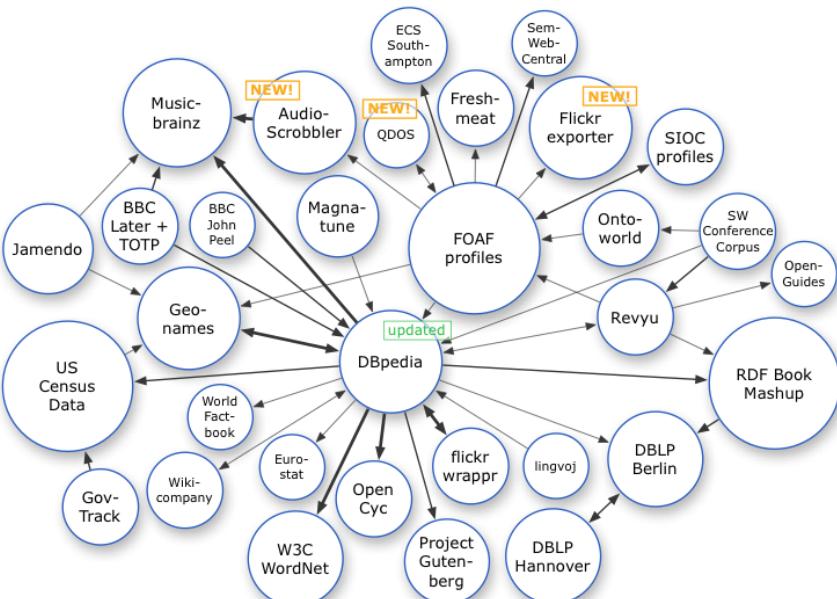
Fuente: <https://www.w3.org/DesignIssues/LinkedData.html>

LOD (2007)

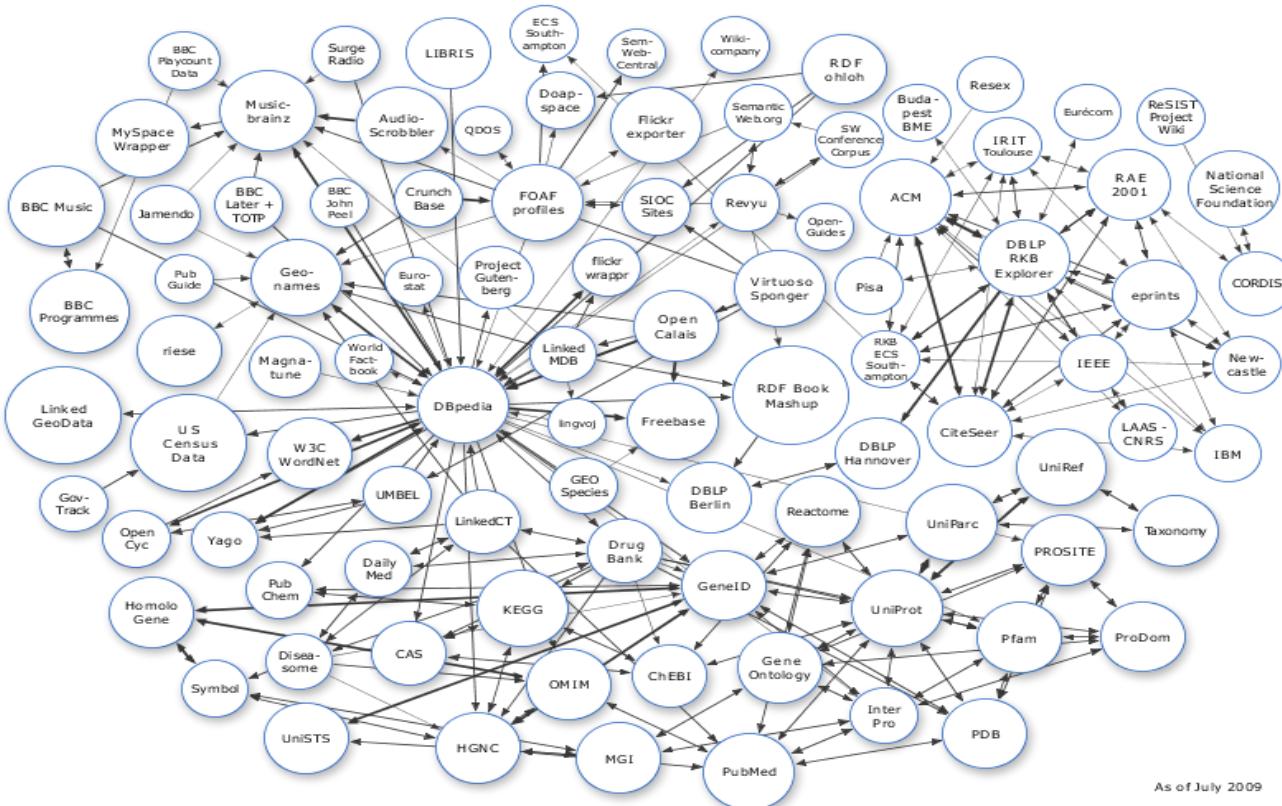


Evolución: <https://lod-cloud.net/>

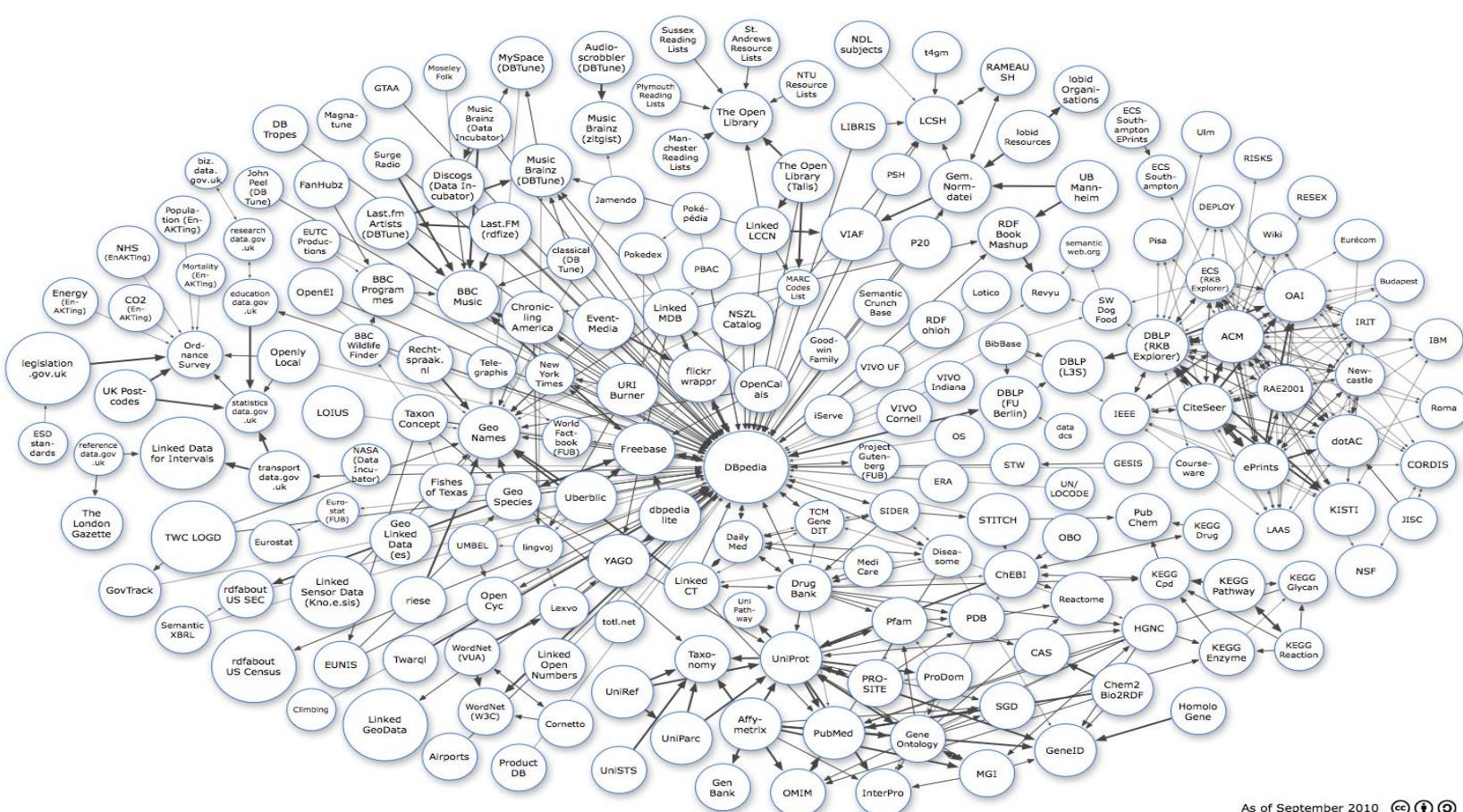
LOD (2008)



LOD (2009)

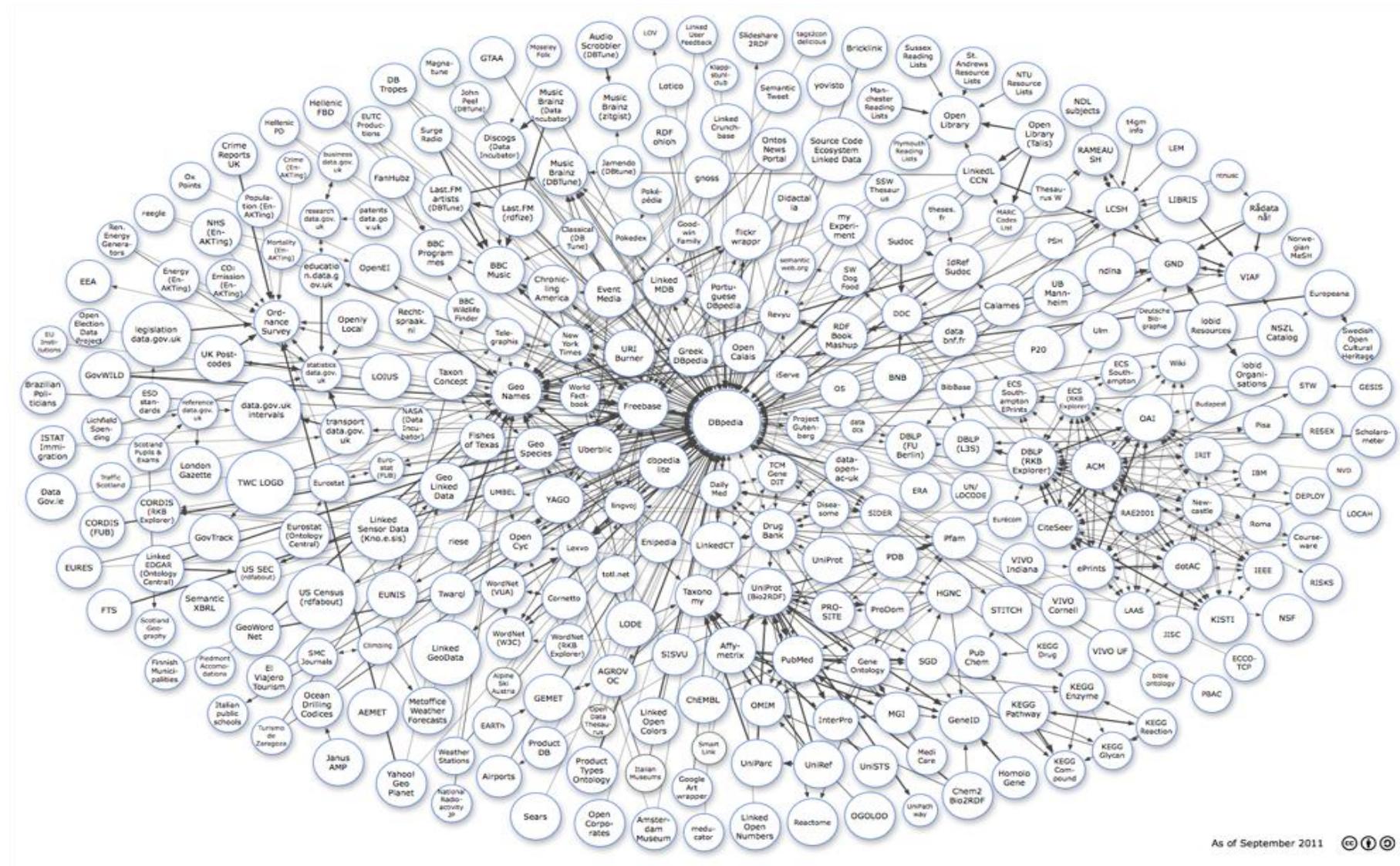


LOD (2010)

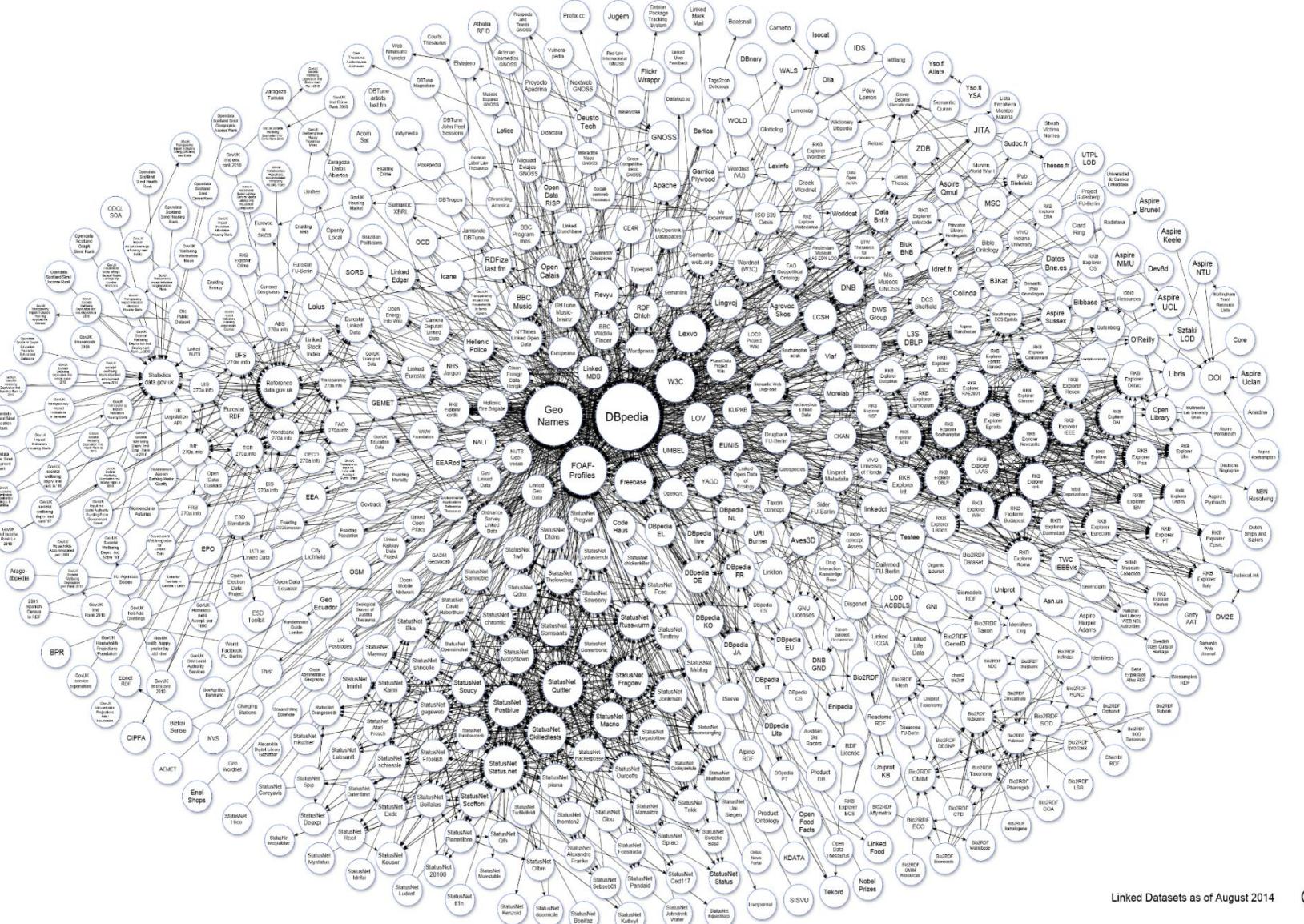


As of September 2010

LOD (2011)



LOD (2014)



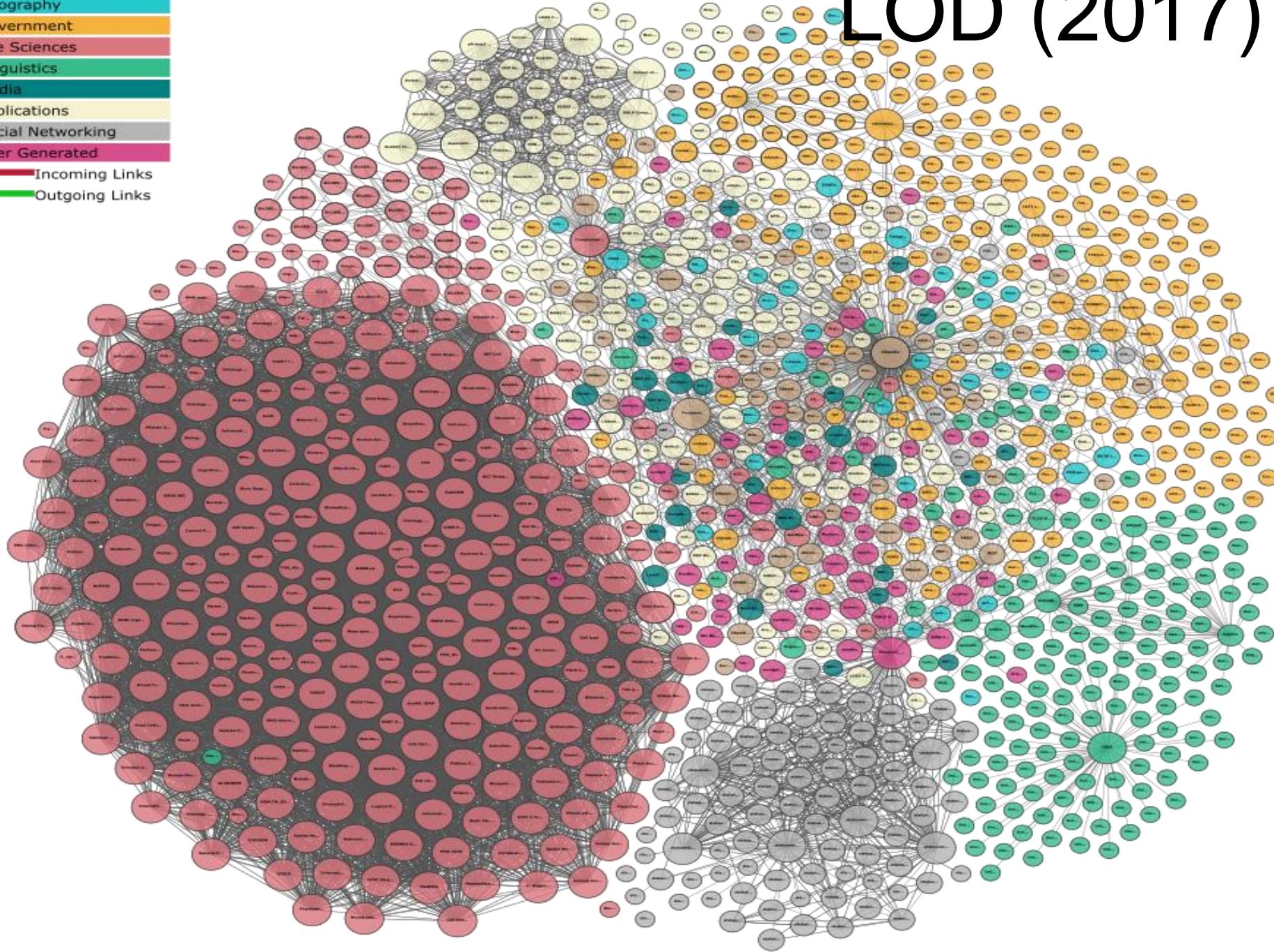
Linked Datasets as of August 2014



LOD (2017)

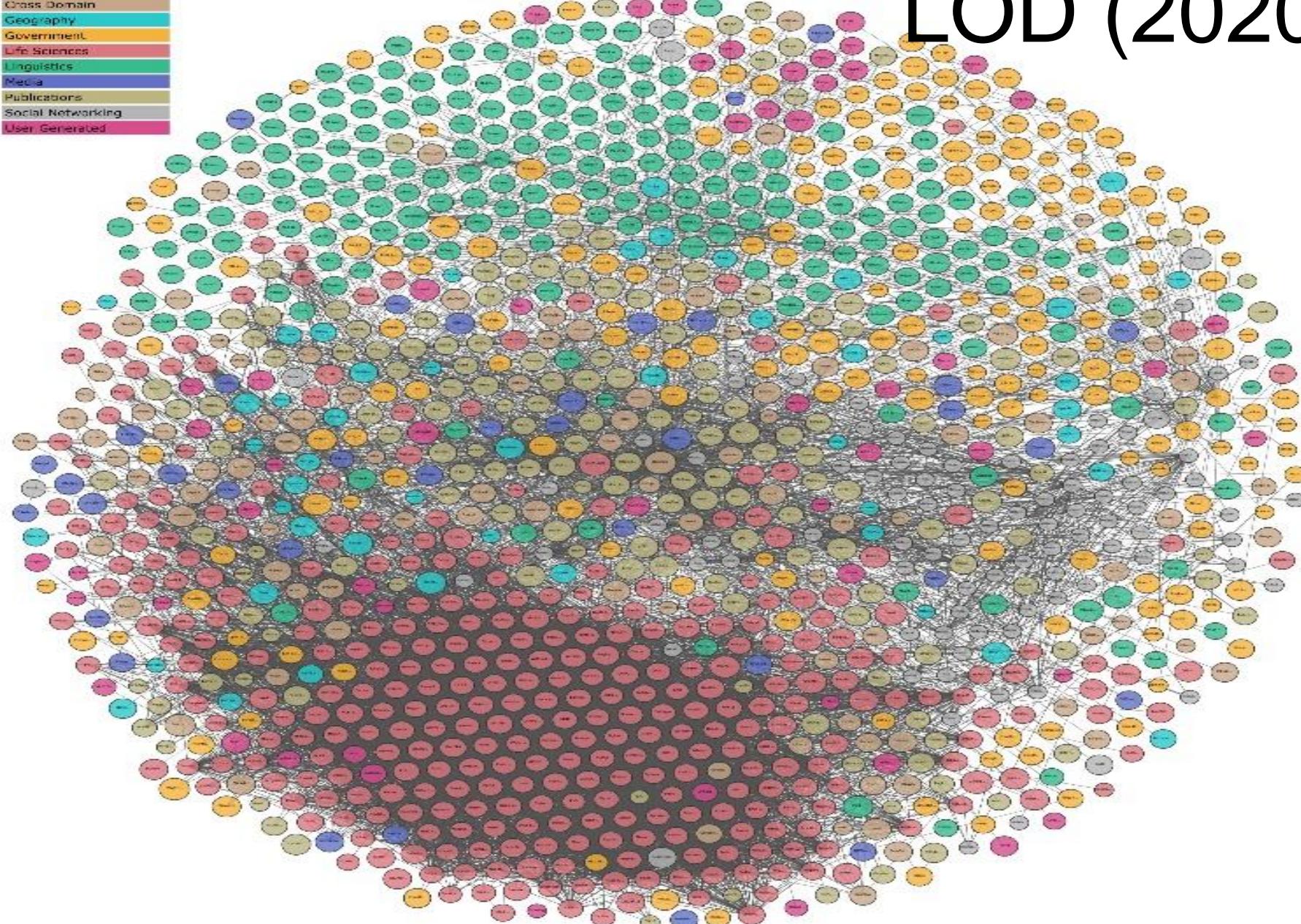
Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated
- Incoming Links
- Outgoing Links



LOD (2020)

Legend	
Cross Domain	
Geography	
Government	
Life Sciences	
Linguistics	
Media	
Publications	
Social Networking	
User Generated	



Linked open data

Ejemplos de iniciativas

data.gov.uk

data.worldbank.org

data.gov

datos.gob.es

datos.gijon.es

...

datos.bcn.cl

data.webfoundation.org



Beneficios de Linked Open Data

Datos accesibles

Evitar pérdidas semánticas al publicar

Facilitar automatización de tareas

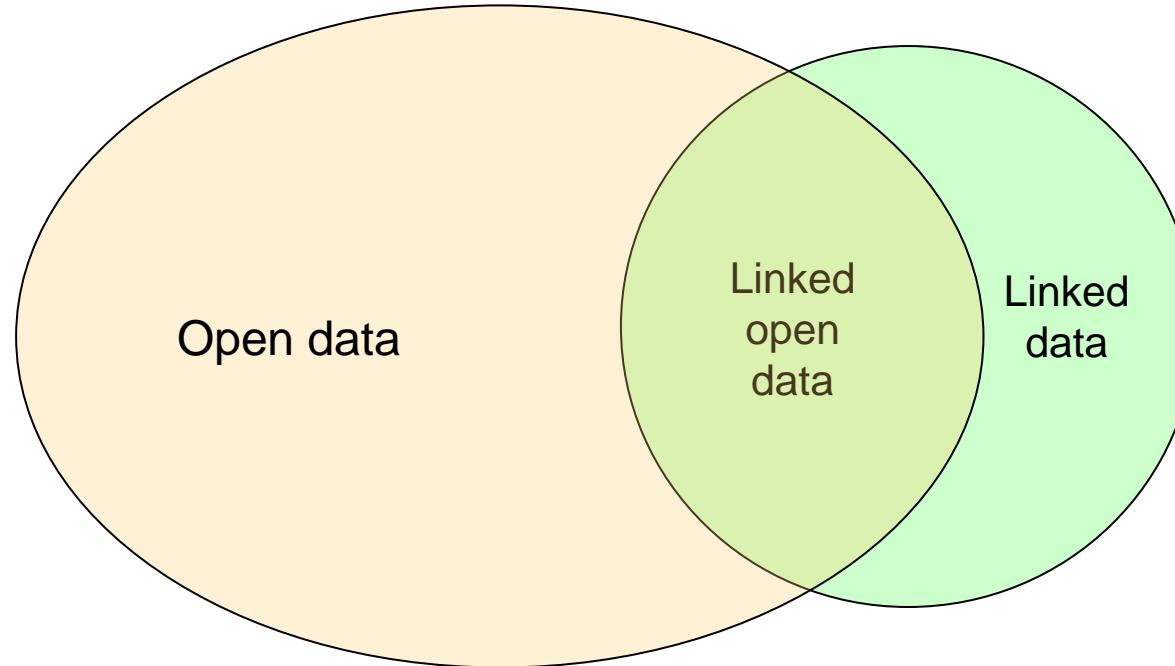
Datos enlazados

Reutilización de datos

Integración de aplicaciones

La mejor manera de explotar tus
datos se le ocurrirá a otro

Linked data vs Open data vs LOD



No todos los datos enlazados son abiertos

No todos los datos abiertos cumplen con los principios LOD

Principios FAIR

Propuestos para publicación de datos científicos

4 principios

(F)indable

(A)ccessible

(I)nteroperable

(R)eusable

Más información: <https://www.go-fair.org/fair-principles/>

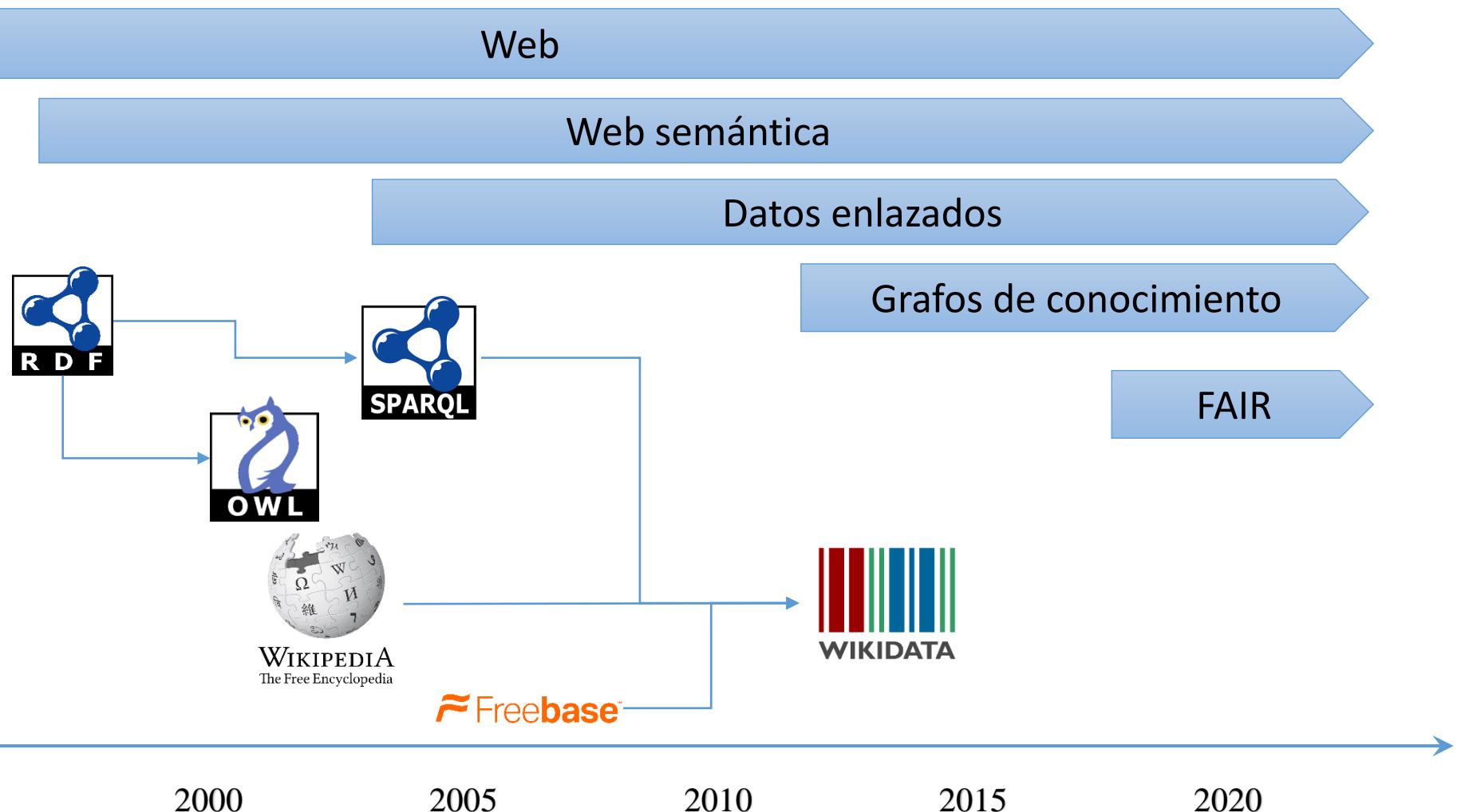
Similar a Linked Open Data

No es necesario que los datos sean abiertos

Ejemplo. Datos de pacientes

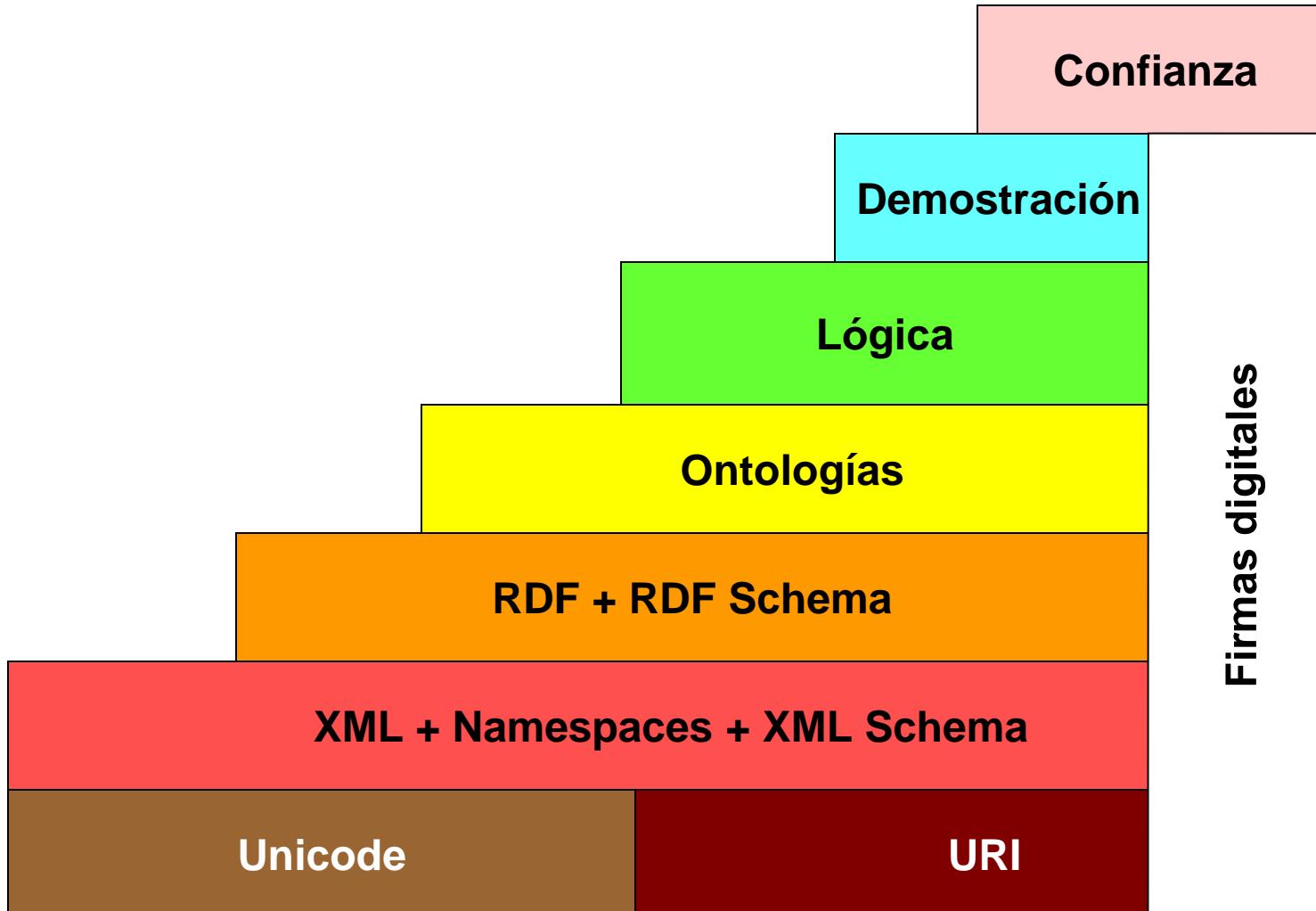
En FAIR no se hace tanto énfasis en tecnologías RDF

Evolución



Tecnologías Web Semántica

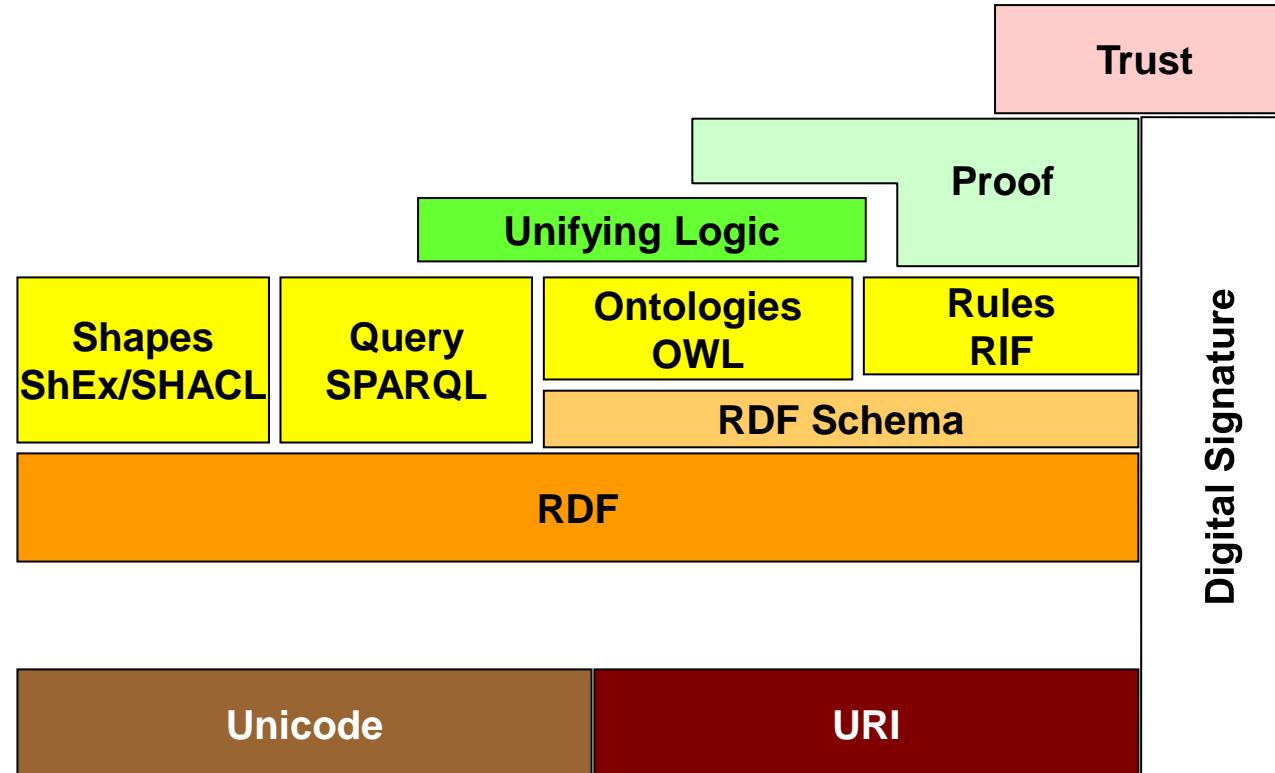
Tarta de la Web (versión inicial)



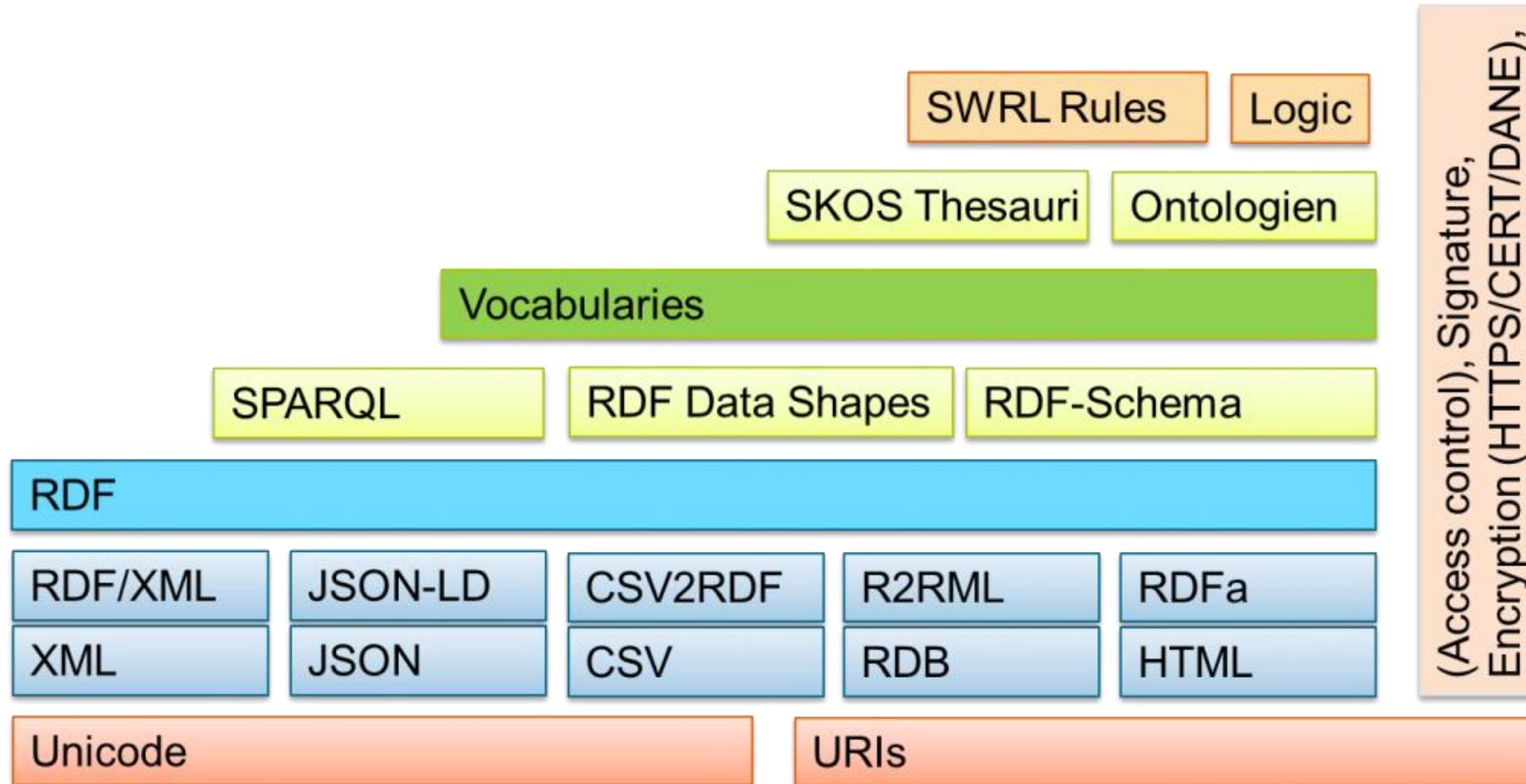
Versión propuesta por Tim Berners Lee, año 2000
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slides10-0.html>

Tecnologías Web Semántica

Cambios en la tarta...



Nuevas versiones de la tarta



Algunas tecnologías

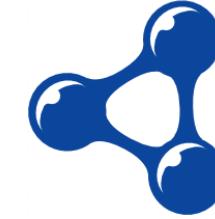
RDF
Descripción datos

SPARQL
Consultas

SHEX - SHACL
Validación

OWL - RDFS
Inferencias

RDF



Resource Description Framework (1998)

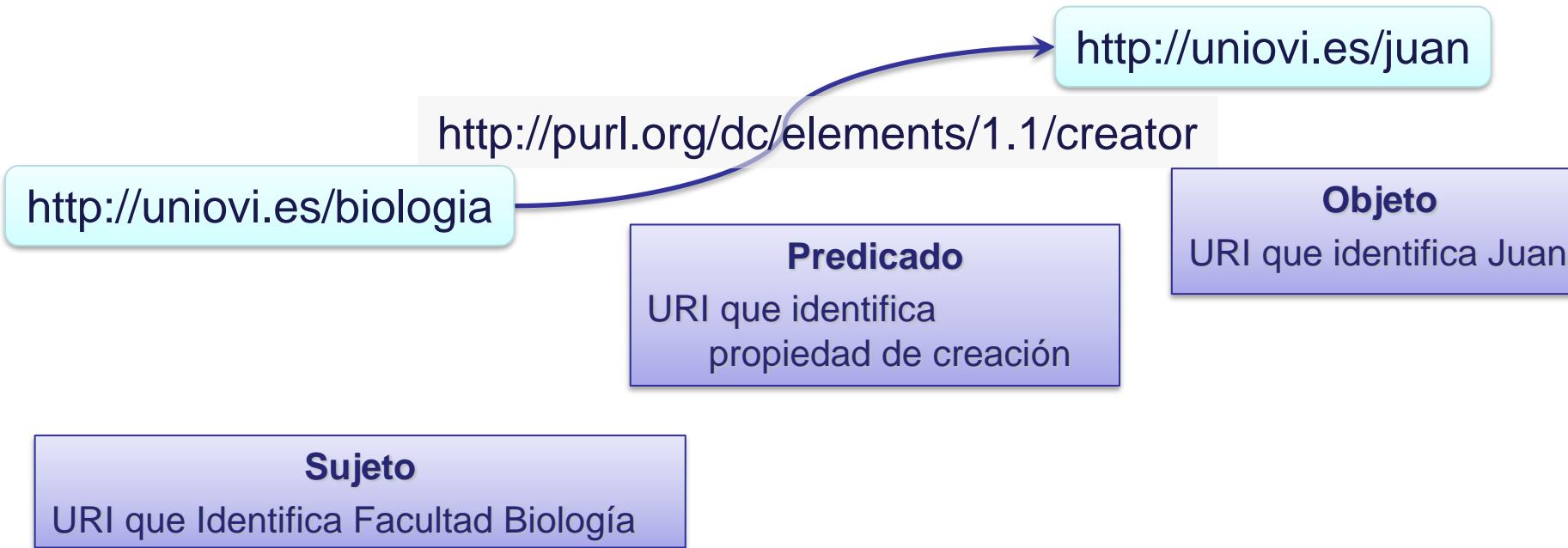
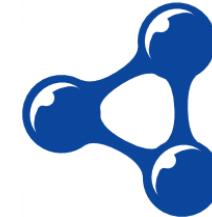
Descripción de recursos

Recurso = se identifica con URI

Se basa en tripletas

Sujeto → Predicado → Objeto

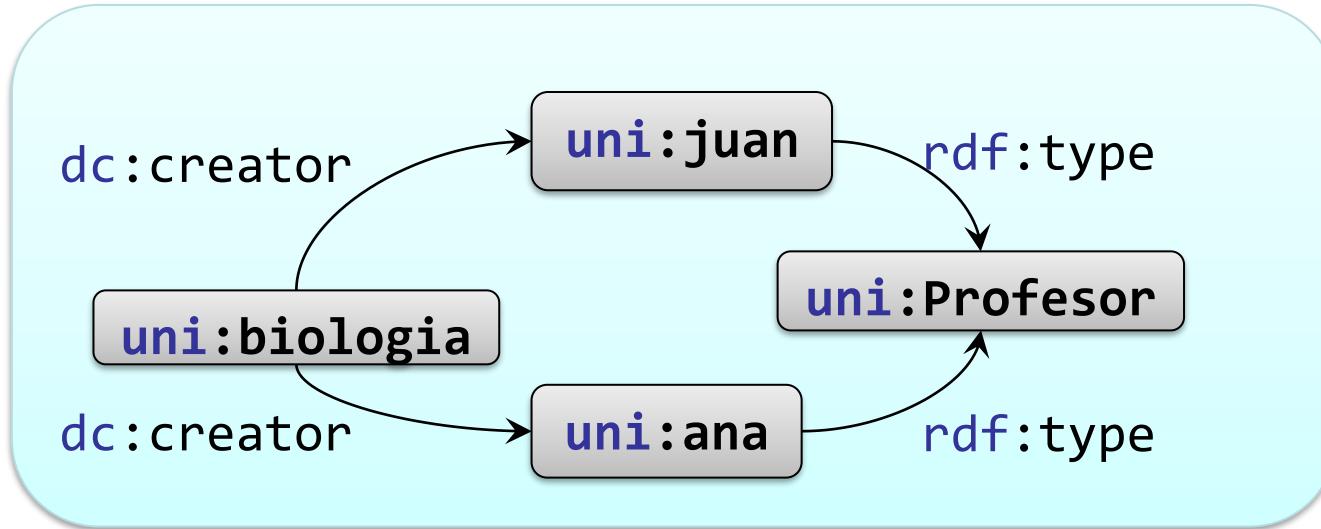
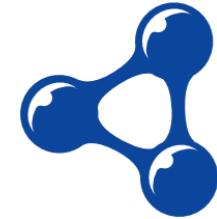
Tripletas RDF



RDF en notación Turtle

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.  
@prefix uni: <http://uniovi.es/> .  
  
uni:biologia dc:creator uni:juan .
```

Grafo RDF



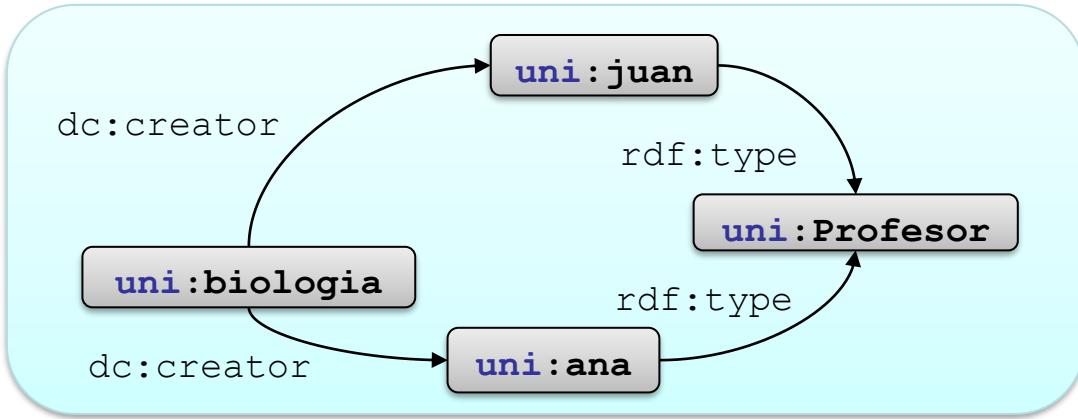
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix uni: <http://uniovi.es/> .  
@prefix dc: <http://purl.org/dc/elements/1.1/> .
```

uni:biologia	dc:creator	uni:juan .
uni:biologia	dc:creator	uni:ana .
uni:juan	rdf:type	uni:Profesor .
uni:ana	rdf:type	uni:Profesor .

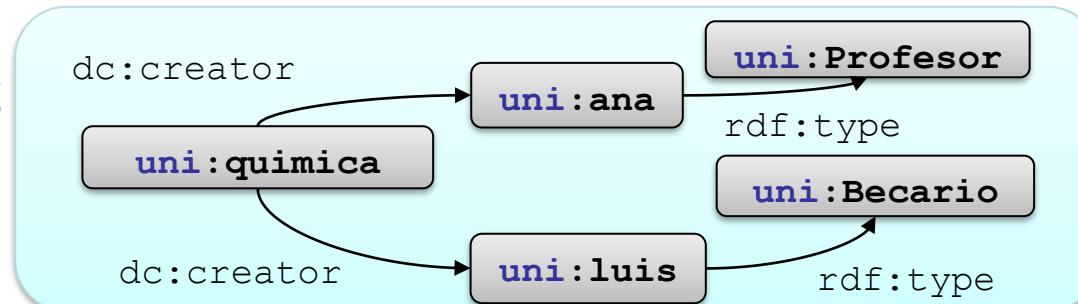
RDF es composicional



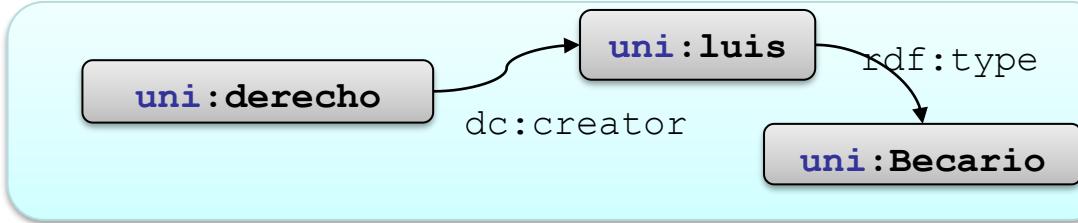
Grafo 1



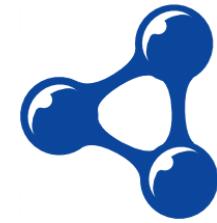
Grafo 2



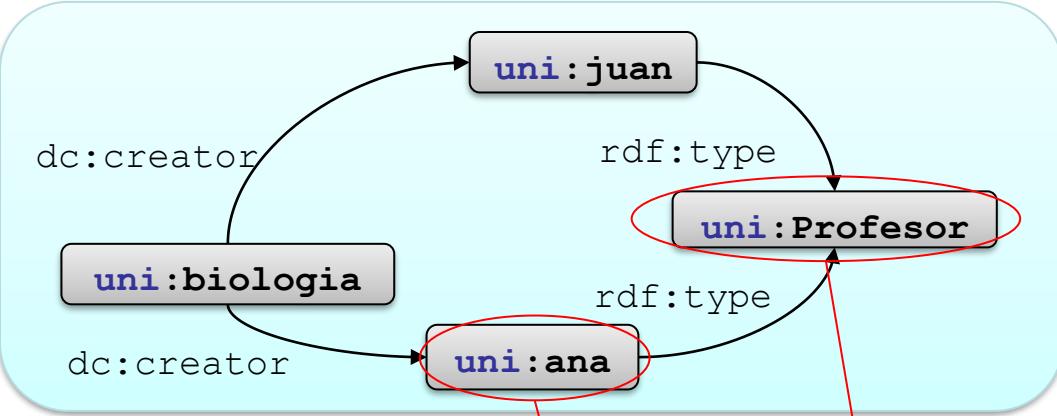
Grafo 3



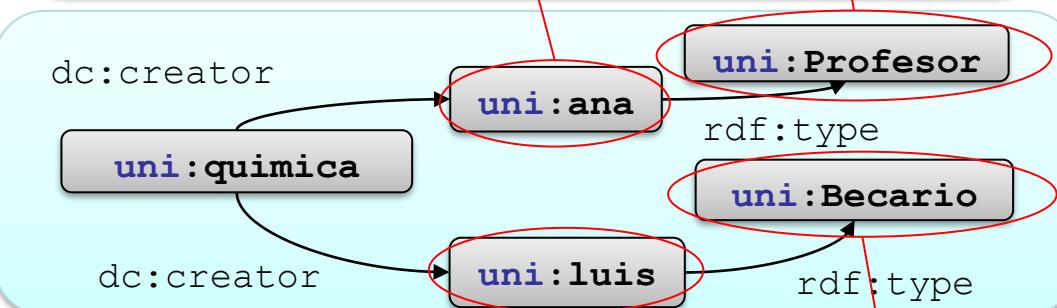
RDF es composicional



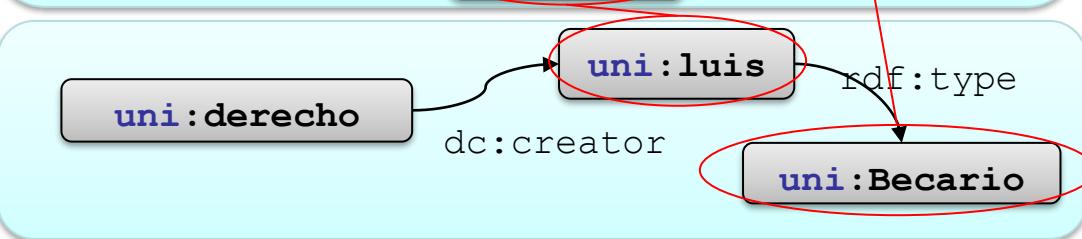
Grafo 1



Grafo 2



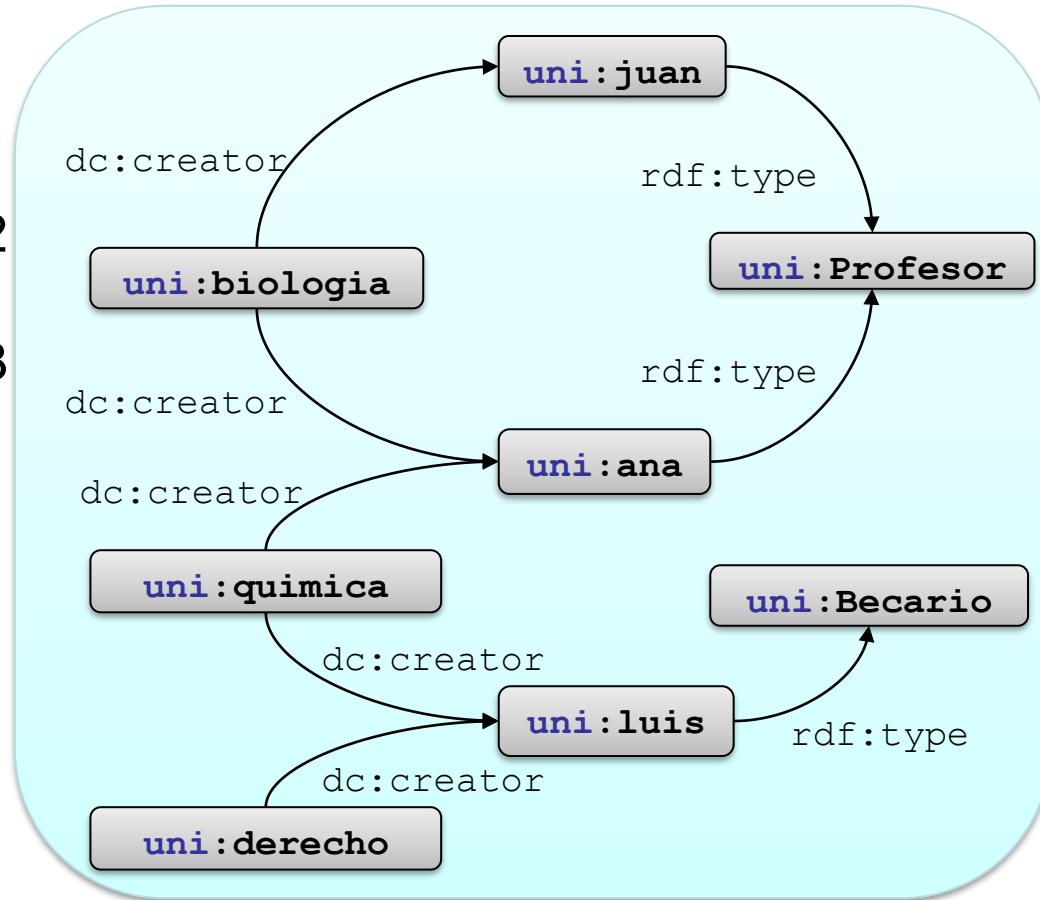
Grafo 3



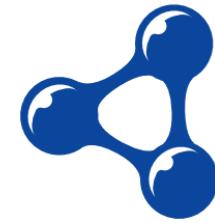
RDF es composicional



Grafo 1
+
Grafo 2
+
Grafo 3



Formatos RDF



Numerosos formatos y sintaxis:

N3

RDF/XML

N-Triples

Turtle

json-ld

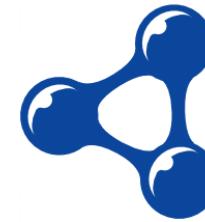
RDFa

etc.

...pero...

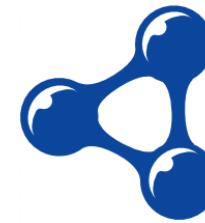
¡Lo más importante es el modelo de grafo!

SPARQL



- Simple Protocol and RDF Query Language
 - Lenguaje de consultas para la web semántica
 - Encaje de grafos
 - Extrae información de modelos RDF
 - Un protocolo
 - Define un mecanismo para invocar un servicio
 - También define un vocabulario para resultados

SPARQL



Ejemplo:

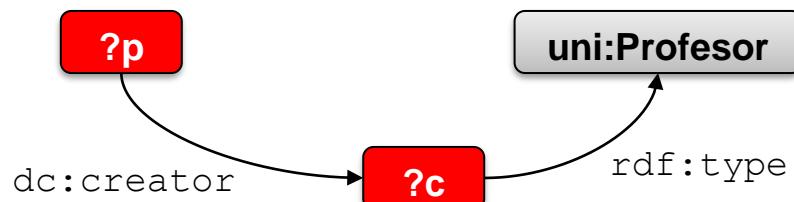
Buscar páginas cuyo autor sea un profesor

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix uni: <http://uniovi.es/>
prefix dc: <http://purl.org/dc/elements/1.1/>

SELECT ?p ?c WHERE {
    ?p dc:creator ?c .
    ?c rdf:type uni:Profesor.
}
```

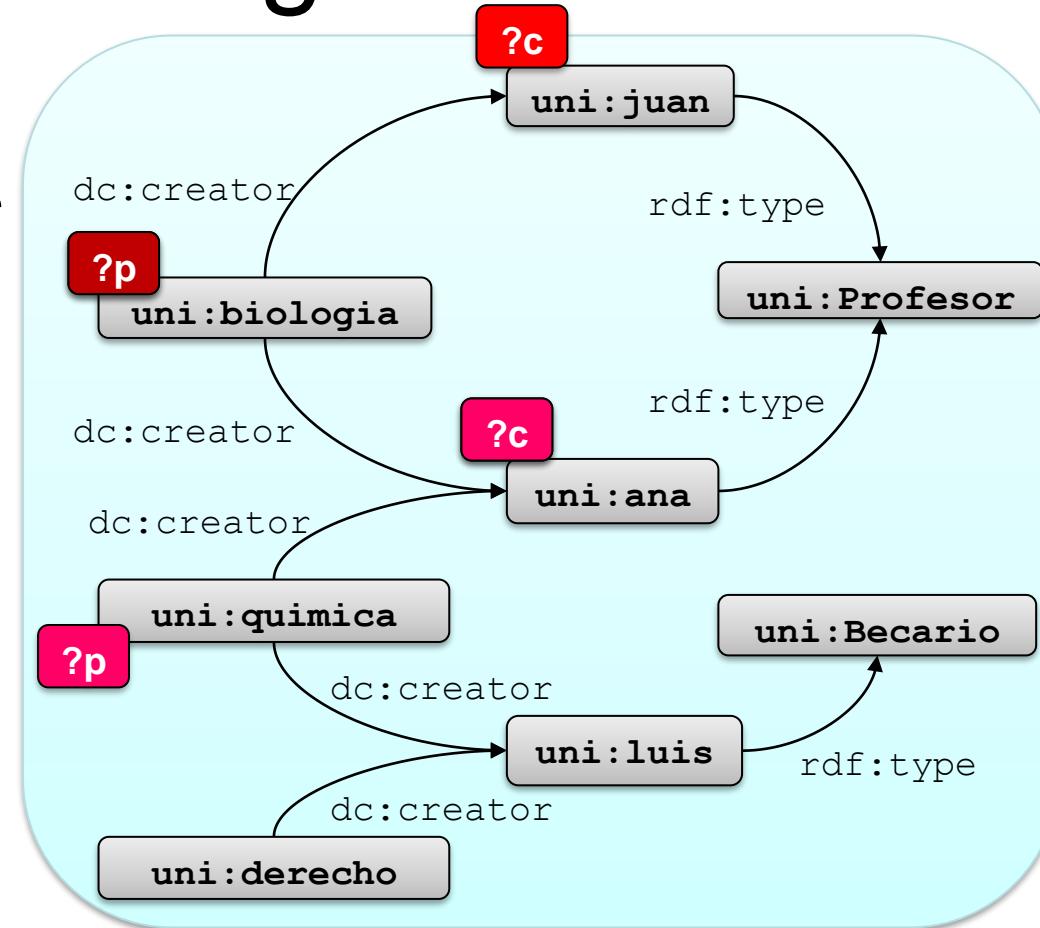
Encaje de grafos

```
SELECT ?p ?c WHERE {  
  ?p dc:creator ?c .  
  ?c rdf:type uni:Profesor  
}
```

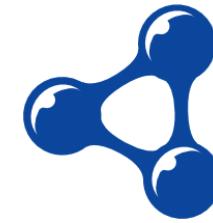


Resultados

?p	?c
uni:biologia	uni:juan
uni:biologia	uni:ana
uni:quimica	uni:ana



RDF Schema



Añade un vocabulario de esquema a RDF

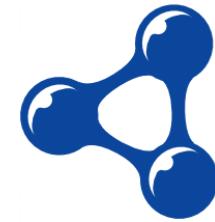
Class, Property, Resource,...

type, subClassOf, subPropertyOf,...

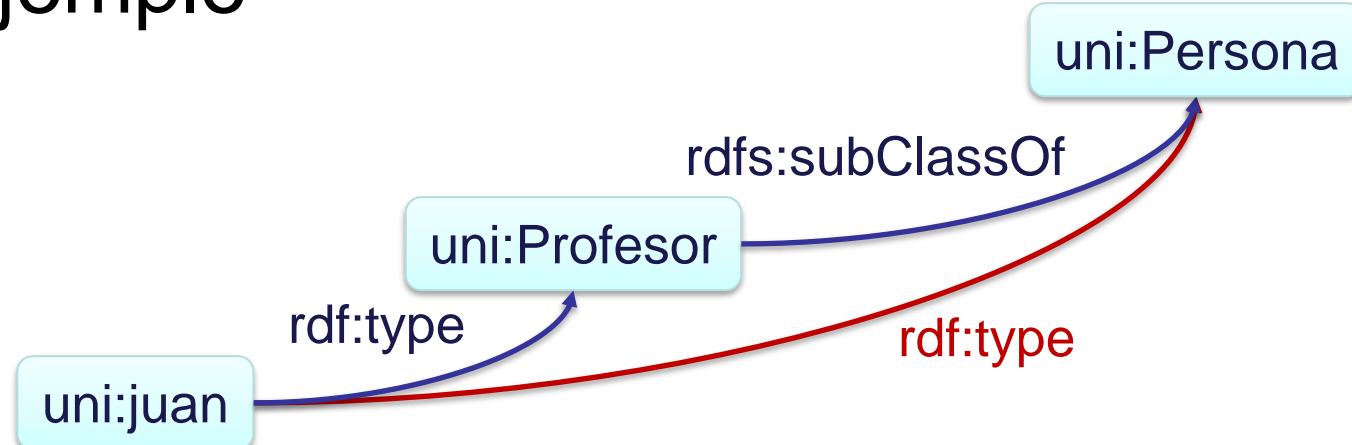
range, domain,...

RDF Schema permite **inferencias**

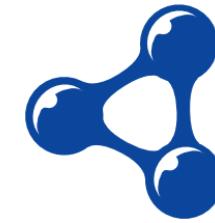
RDF Schema



- Ejemplo



SPARQL + Inferencia



Combinar SPARQL e inferencia

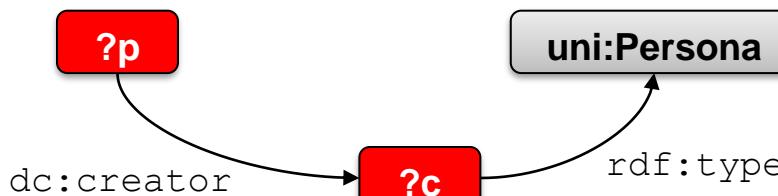
Ejemplo:

Páginas cuyo autor sea una persona

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix uni: <http://uniovi.es/> .  
@prefix dc: <http://purl.org/dc/elements/1.1/> .  
  
SELECT ?p ?c WHERE {  
    ?p dc:creator ?c .  
    ?c rdf:type uni:Persona.  
}
```

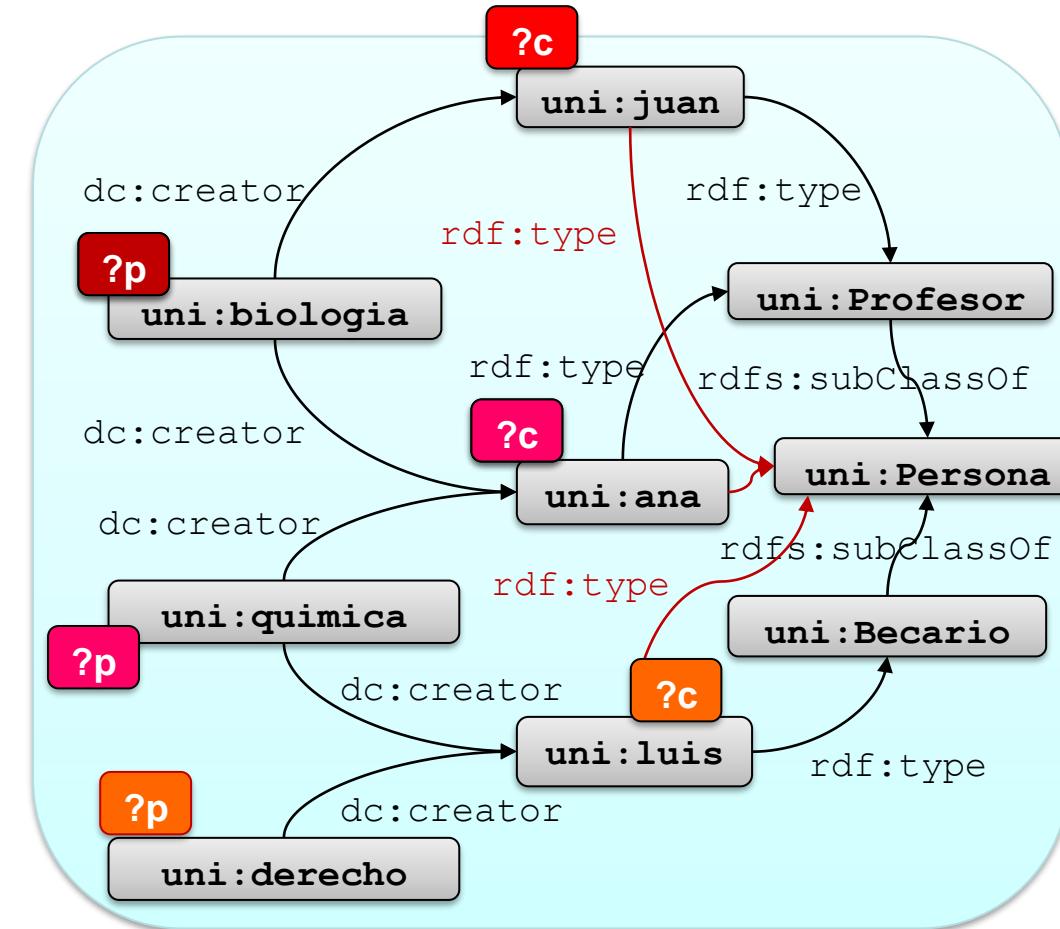
SPARQL con inferencia

```
SELECT ?p ?c WHERE {  
  ?p dc:creator ?c .  
  ?c rdf:type uni:Persona .  
}
```



Resultados

?p	?c
uni:biologia	uni:juan
uni:biologia	uni:ana
uni:quimica	uni:ana
uni:derecho	uni:luis





Ontologías

RDF Schema permite hacer inferencias sencillas
Poca expresividad

OWL (Web Ontology Language)
Añade más expresividad
Formalizar dominios concretos: ontologías

Expresividad vs Complejidad



Mitos de la Web Semántica

- Navegador inteligente
- Una nueva Web
- El cerebro global
- La gran verdad: Una única ontología
- Una etiqueta para cada cosa
- Nadie querrá compartir datos
- Demasiada apertura
- Moda pasajera
- No hay *Killer application*



El navegador inteligente

Mito:

El objetivo es conseguir sistemas que naveguen por internet de forma inteligente

Realidad:

Objetivo = desarrollar tecnologías que faciliten el procesamiento automático de la información de la Web y su integración

No es Inteligencia Artificial pero sí se utilizan técnicas de esa disciplina

Una nueva Web

Mito:

La Web Semántica (\approx Web 3.0) es una nueva versión de la web que obligará a cambiar todo lo que ya hay

Realidad:

Se propone transición gradual.

Las tecnologías ofrecerán valor añadido.

El cerebro global

Mito:

El proyecto de la Web semántica generará un cerebro global

Realidad:

La web semántica facilitará un mejor uso de los datos de la web.

Sí es un camino hacia la inteligencia colectiva

La gran verdad

Mito:

Se propone la creación de **una única ontología** con todo el conocimiento de la humanidad

Realidad:

Múltiples ontologías para diferentes dominios

Facilitar la integración

Mejorar la descripción de dominios

Una etiqueta para cada cosa

Mito:

El objetivo es asignar una etiqueta similar a RFID para cada cosa

Realidad

No es factible que cada cosa conlleve sus propios metadatos

Descripciones de recursos externas a ellos

Nadie querrá compartir datos

Mito:

Los proveedores de información no tendrán motivación para adoptar tecnologías nuevas

Realidad:

Lo harán cuando encuentren un retorno de inversión adecuado
Posicionamiento semántico

<http://schema.org>

Principales buscadores indexan datos estructurados
Google, Yandex, Yahoo, Bing

Demasiada apertura

Mito:

Si abrimos datos de bases de datos, los perdemos

Realidad:

Hay tecnologías para limitar acceso

Declarar de dónde provienen los datos

Establecer propiedad legal de los datos

Moda pasajera

Mito:

Mito1: La Web semántica es algo nuevo

Mito 2: La Web semántica es algo viejo

Realidad:

Planteada ya en 1994, visión a largo plazo

Exceso de entusiasmo vs escepticismo

Casos de éxito: RSS, microformatos, XBRL,...

"A little semantics goes a long way"

No hay *killer application*

Mito:

No se ha desarrollado una *killer application*

Realidad:

¿Es necesaria?

¿*Linked Open Data*?

Retos

Proyecto Web semántica:

Primera fase = producción



Segunda fase = consumo



Calidad es cada vez más importante



Fin de la Presentación

