

# 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

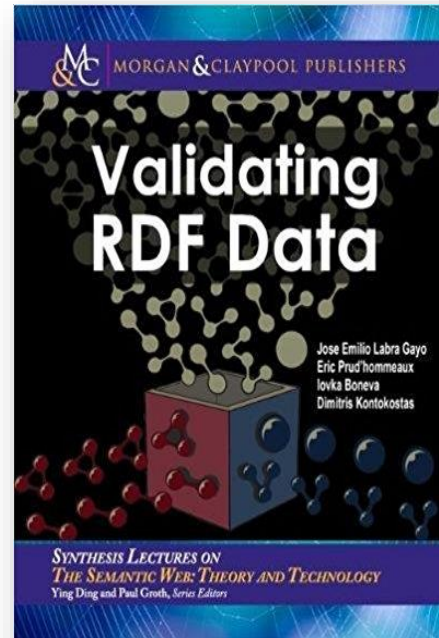
Máster Universitario Ingeniería Web

Curso 2022/23

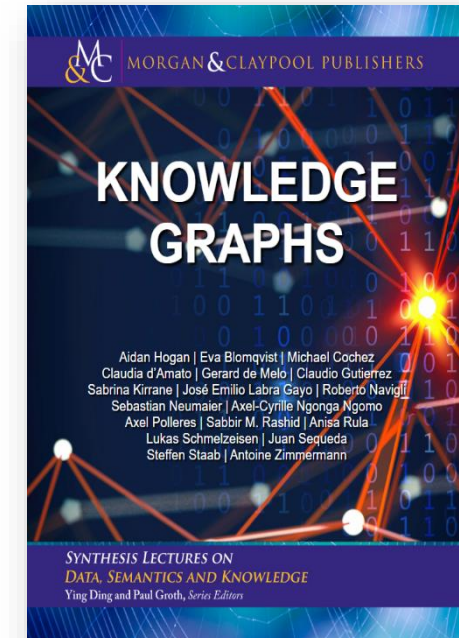
# Referencias



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



**Validating RDF Data**  
Ed. Morgan & Claypool, 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 una **web de datos**

No solo páginas/documentos web, sino datos

Datos enlazados

Campo relacionado con:

**Big Data**

Enormes cantidades de datos de la Web

...¡y 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** (programas informáticos)

Nos muestran las páginas Web (navegadores)

...pero también analizan la información (bots)

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

\*cuando están bien programadas

# ¿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

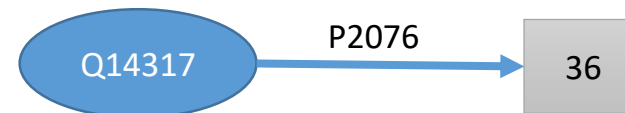
Identificadores (URIs) en Wikidata

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

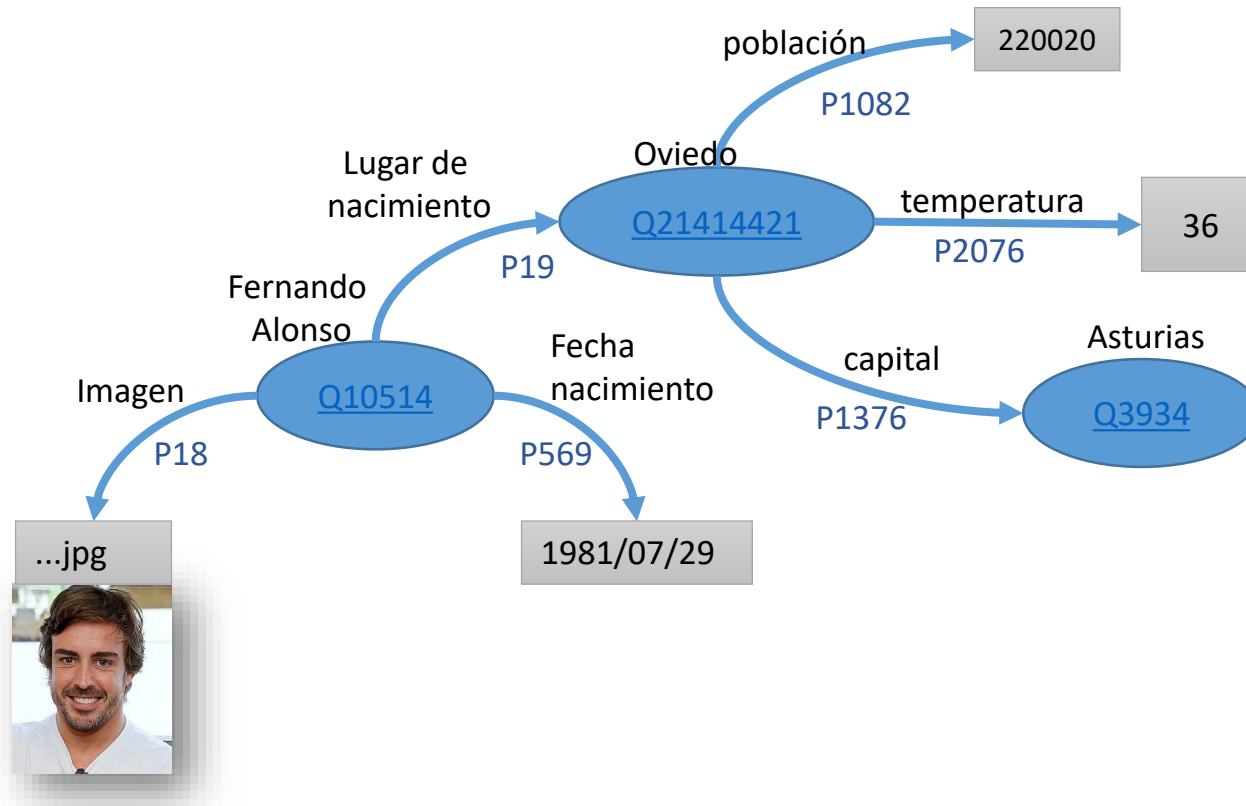
<http://www.wikidata.org/wiki/Q1813449>

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

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



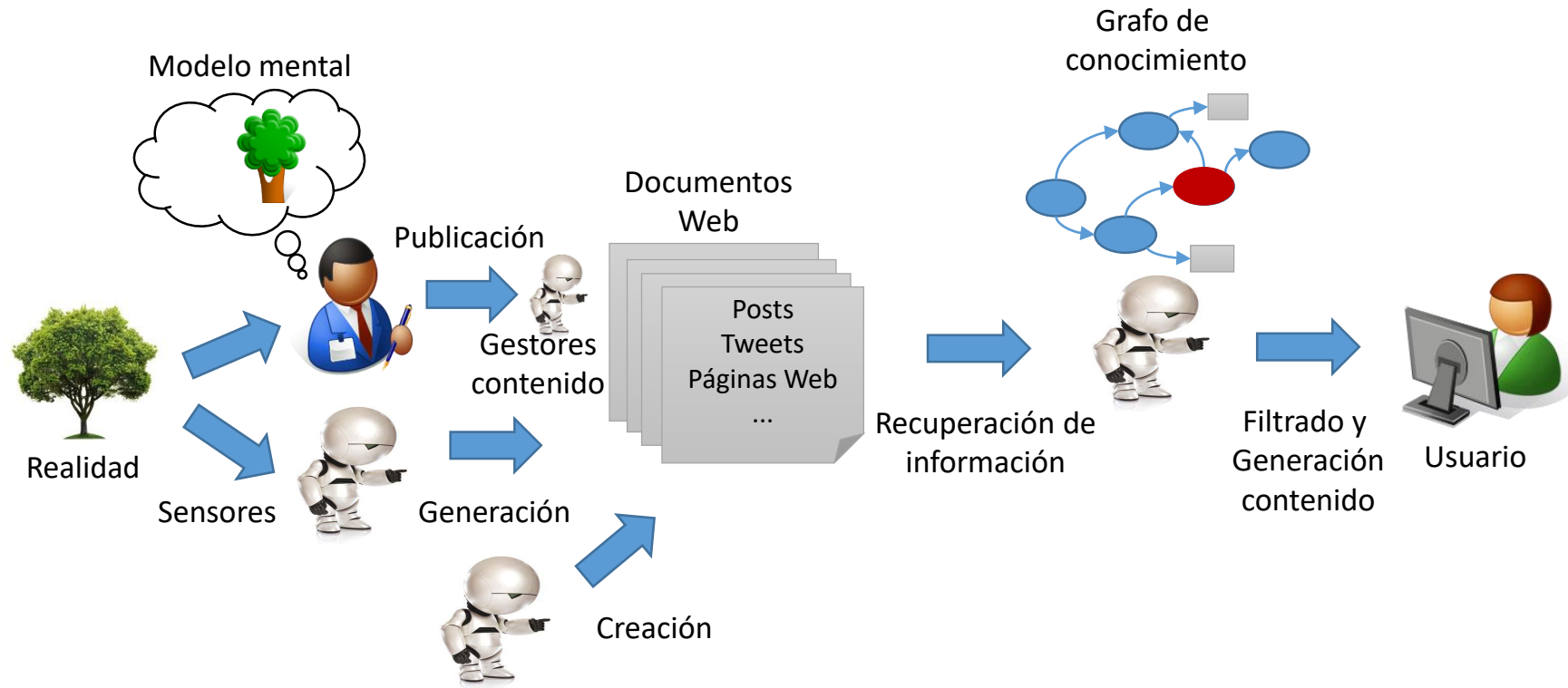
# Representación de conocimiento en grafos



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

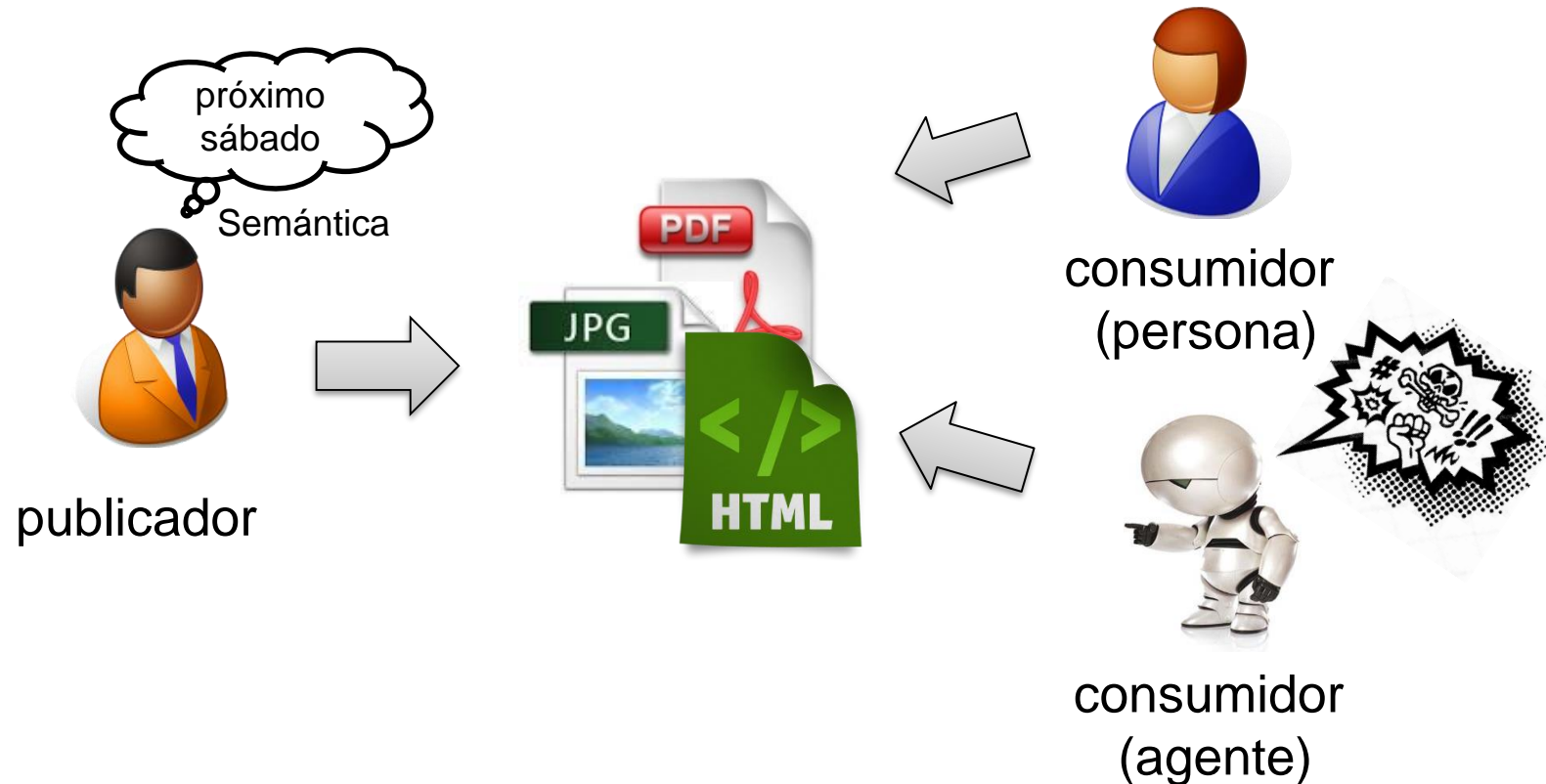
En inglés: Principio **AAA**: Anyone can say Anything about Any topic  
Fuente: Semantic Web for the Working Ontologist, D. Allemang, J. Hendler

# 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

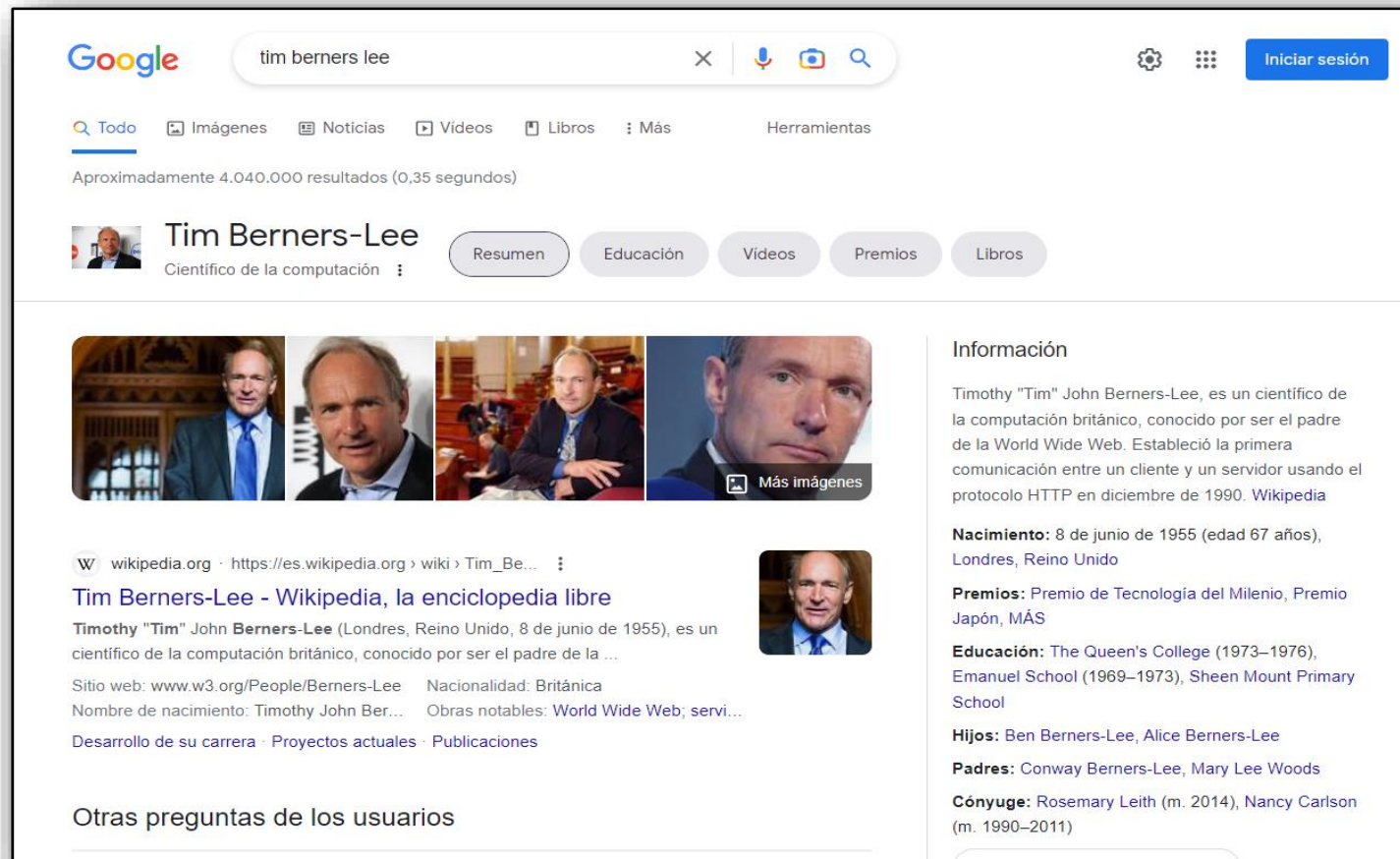
Grafos de conocimiento

Wikidata

Datos abiertos enlazados

# Grafos de conocimiento

- Noción actual: Popularizada por Google, 2012\*



The screenshot shows a Google search for "tim berners lee". The search bar is at the top with the text "tim berners lee" and a search button. Below the search bar, there are navigation tabs for "Todo", "Imágenes", "Noticias", "Videos", "Libros", and "Herramientas". The search results show "Aproximadamente 4.040.000 resultados (0,35 segundos)".

The main result is for "Tim Berners-Lee", identified as "Científico de la computación". There are buttons for "Resumen", "Educación", "Videos", "Premios", and "Libros". Below this, there are four small images of Tim Berners-Lee, with a "Más imágenes" button.

Below the images, there is a snippet from Wikipedia: "Tim Berners-Lee - Wikipedia, la enciclopedia libre". The snippet reads: "Timothy 'Tim' John Berners-Lee (Londres, Reino Unido, 8 de junio de 1955), es un científico de la computación británico, conocido por ser el padre de la ...". Below the snippet, there are links for "Sitio web", "Nacionalidad", "Nombre de nacimiento", "Obras notables", "Desarrollo de su carrera", "Proyectos actuales", and "Publicaciones".

On the right side, there is an "Información" section with the following details:

- 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)

At the bottom, there is a link for "Otras preguntas de los usuarios".



# Grafos de conocimiento



# Grafos de conocimiento

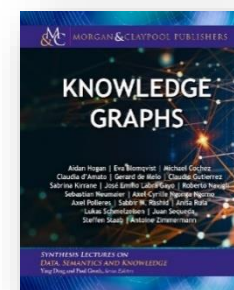
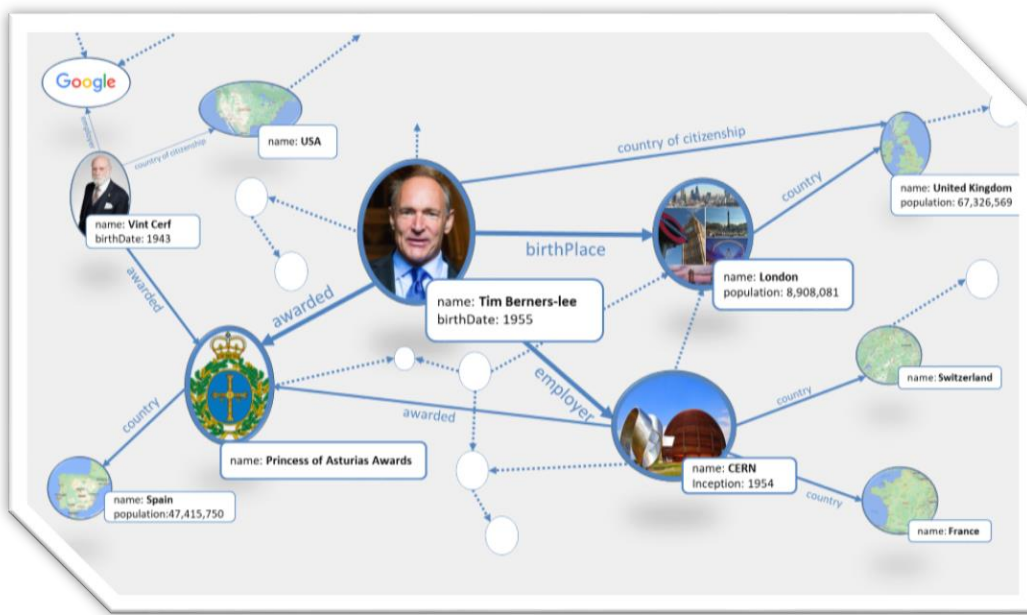
Grafo de conocimiento = un **grafo de datos**

*que pretende acumular y transmitir*

**konocimiento 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

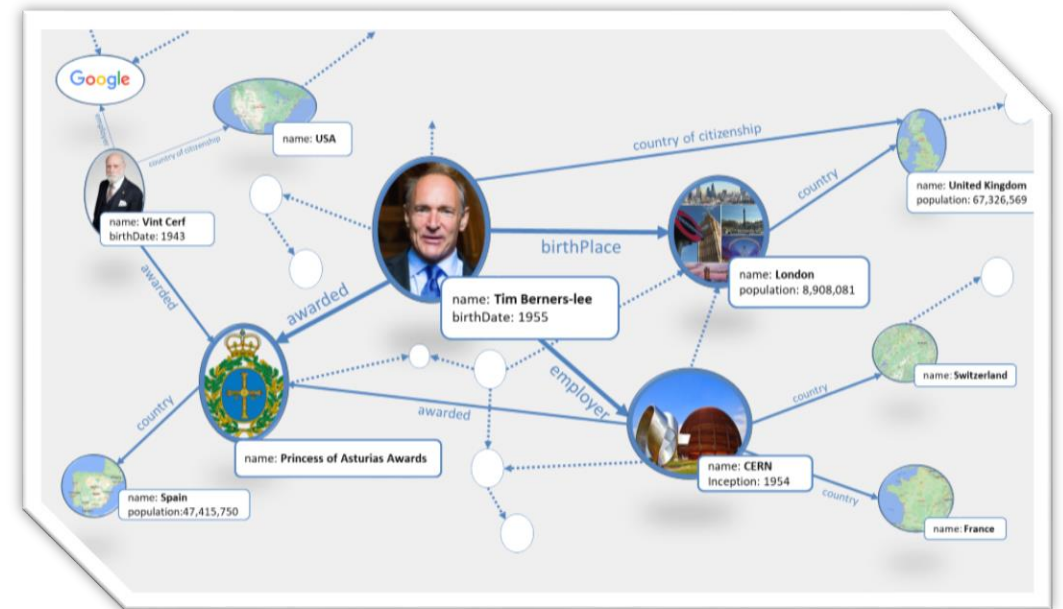
Gobierno de datos

Gestionar datos heterogéneos

Sistemas de recomendación

Chatbots y NLP

...



# Google knowledge panel



oviedo



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



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/](http://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](https://www.amazon.com/s?ref=asc_df_crawl-data-CC-MAIN-2022-48/) 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	19,072,828,514
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

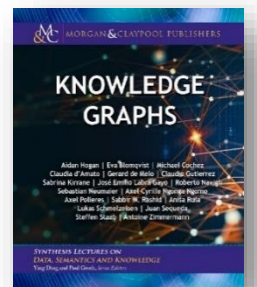
Búsqueda en la web: Google, Bing...

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





# Wikidata

Página de Wikipedia en inglés de Tim Berners-Lee

The screenshot shows the English Wikipedia article for Tim Berners-Lee. The article text is partially visible, mentioning his birth on 8 June 1955 and his role as the inventor of the World Wide Web. A light blue box highlights a portion of the article text.

The screenshot shows the Hindi Wikipedia article for Tim Berners-Lee. The title is 'टिम बर्नर्स-ली'. The article text is partially visible, mentioning his birth on 8 June 1955 and his role as the inventor of the World Wide Web. A light blue box highlights a portion of the article text.

Página en Bihari de Tim Berners-Lee

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

The screenshot shows the Wikidata page for Tim Berners-Lee (Q80). The page displays the Wikidata logo, the name 'Tim Berners-Lee (Q80)', and various statements. A light blue box highlights the 'instance of' statement, which is 'human'. Another light blue box highlights the 'image' statement, which is 'Sir Tim Berners-Lee (cropped).jpg'. The page also shows a list of statements, including 'point in time' and 'media legend', and a list of references.

# Wikidata: algunas características

Colaborativo: 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



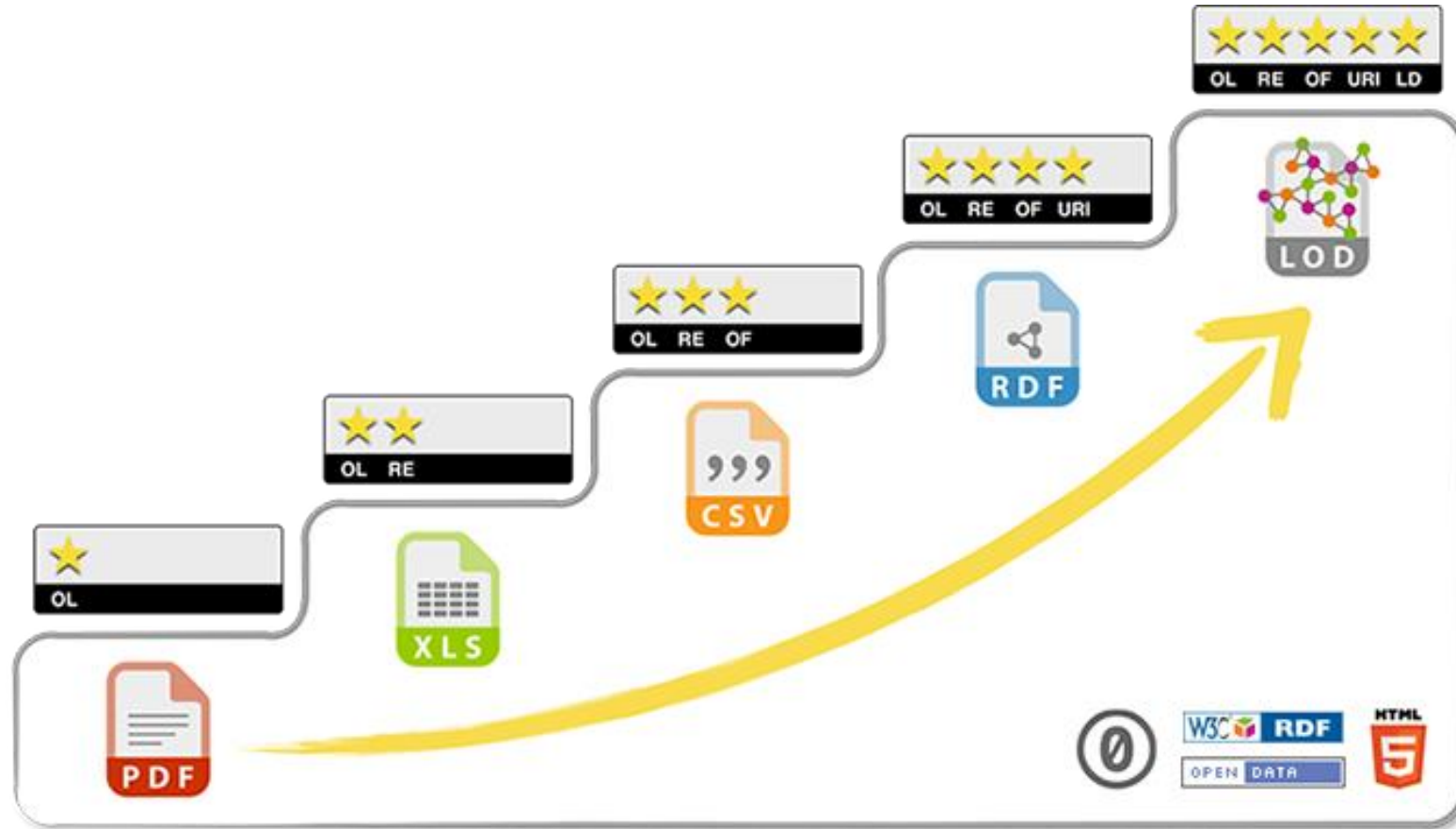
# Modelo de Estrellas\*

<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 **URIs para identificar** datos  
(otros sistemas puedan enlazar nuestros datos)
- ★★★★★ **Enlazar con otros** datos externos  
(proporcionar contexto)

\* Enunciado por Tim Berners-Lee en Gov 2.0 Expo 2010

# Modelo de estrellas





# Formatos no estructurados

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

Formatos binarios: PDF, PS, etc.

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



```
FF 08 FF E0 00 10 44 46 49 46 00 01 01 01 00 60 00 60 00 00 FF DB 00 43 00 02 02 02 02 01 02 02 02 02 03 02 02 03 03 06 04 03 03 03  
05 04 06 08 07 09 08 08 07 08 08 09 0A 0D 0B 09 0A 0C 0A 08 09 0B 0F 0B 0C 0D 0E 0F 0E 09 0B 10 11 10 0E 11 0D 0E 0E 0E FF DB 00 43  
03 03 03 03 07 04 04 07 0E 09 09 09 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E  
0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E 0E FF C0 00 11 08 01 E0 01 E0 03 01 22 00 02 11 01 03 11 01 FF C4 00 1F 00 00 01 05 01 01 01 01  
00 00 00 00 00 00 00 01 02 03 04 03 04 07 08 09 0A 0B FF C4 00 86 13 00 09 02 03 05 02 04 03 06 06 04 04 00 00 01 07 02 02 03 00 04 11  
31 41 0E 23 51 61 07 22 71 14 32 81 91 A1 0B 23 42 81 C1 16 52 D1 F0 14 39 62 73 82 09 0A 16 17 18 19 1A 25 26 27 28 29 2A 34 36 37  
43 44 45 46 47 48 49 4A 53 54 56 56 57 58 59 5A 63 64 65 66 67 68 69 6A 73 74 75 76 77 78 79 7A 83 84 85 86 87 88 89 8A 92 93 94 95 96  
9A A2 A3 A4 A5 A6 A7 A8 A9 AA B2 B3 B4 B5 B6 B7 B8 B9 BA C2 C3 C4 C5 C6 C7 C8 C9 CA D2 D3 D4 D5 D6 D7 D8 D9 DA E1 E2 E3 E4 E5 E6 E7 E8  
F2 F3 F4 F5 F6 F7 F8 F9 FA FF C4 00 1F 01 00 03 01 01 01 01 01 01 01 01 01 01 00 00 00 00 00 01 02 03 04 05 06 07 08 09 0A 0B FF C4 00  
02 01 02 04 04 03 04 07 05 04 04 00 01 02 77 00 01 02 09 11 04 05 21 31 06 32 41 51 07 61 71 13 22 32 81 08 14 42 91 A1 B1 C1 09 23 33  
62 72 D1 0A 16 24 34 E1 25 F1 17 18 19 1A 26 27 28 29 2A 35 36 37 38 39 3A 43 44 45 46 47 48 49 4A 53 54 55 56 57 58 59 5A 63 64 65 66  
6A 73 74 75 76 77 78 79 7A 82 83 84 85 86 87 88 89 8A 92 93 94 95 96 97 98 99 9A A2 A3 A4 A5 A6 A7 A8 A9 AA B2 B3 B4 B5 B6 B7 B8 B9 BA  
C5 C6 C7 C8 C9 CA D2 D3 D4 D5 D6 D7 D8 D9 DA E2 E3 E4 E5 E6 E7 E8 E9 EA F2 F3 F4 F5 F6 F7 F8 F9 FA FF DA 00 0C 03 01 00 02 11 03 11 00  
F4 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
58 5E 5D A3 15 83 20 61 C0 34 00 EA 29 33 C6 69 54 86 E9 40 05 14 A4 4D 19 34 83 93 C5 00 14 B3 B6 9F 61 40 82 68 01 B4 E2 0E D2 50 01  
01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01  
01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01 45 14 50 01  
6A EA 0E FD 8C 3D 3B D3 63 3B AE 00 63 86 03 BA 80 E8 19 01 51 B9 87 6A E6 BC 41 E2 BF 07 78 5A C5 EF 36 BB E8 ED 1D 14 90 A6 B6 4D 3B  
2E 48 3B 33 E4 E3 A4 D6 E2 CA 00 EC 7D 2B E6 08 4F DA BF C0 93 F9 B6 4D 3A 5B 94 89 15 F6 87 CF 6A F6 5D 33 E2 87 81 B5 88 63 6B 6D  
1C 02 C2 A9 45 A0 3D 01 5F 81 E0 01 D2 SA 23 6C 92 38 AA 76 F3 D9 E8 AD 78 3B C6 91 4F AD AD BA B6 52 75 81 48 61 F8 A1 80 A1 1F 9D DC  
91 D0 D2 18 A4 23 EF 73 40 21 3E FD 48 0C 11 65 C3 6F FC 2A 7C D8 3E 46 7C 86 A1 9B B6 4E C0 29 3C 71 4D A4 07 9F 4A 5E DA 5C 02 BA 28  
51 48 5A 60 14 51 48 3B 80 51 45 14 5C 02 BA 28 A9 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00 A2 BA 28 00  
93 4B C6 05 D0 93 29 51 10 C5 BE 53 81 2E 69 1B 6A 03 E5 9F 9A 96 B5 01 C1 0E FC 15 C5 3D 80 AC 01 3D 09 89 2B 4D 7E 2D FF 00 7A B1  
48 D1 AC A5 9F 84 BF 89 36 8C 80 88 81 60 37 08 6C 8D A3 76 6B 27 59 D7 94 BF 0F E9 E8 73 AA E8 47 4C A0 47 0E C0 57 CA 9F 11 BF 6A 8D  
48 2D 0F 12 DE 0C 85 76 3C 6E 8E 14 F1 A7 C5 EF 19 78 E6 FA 4F DE 0B D7 4B 67 27 01 1F 86 0A 2D 7D B1 F1 3B F6 B1 D2 74 8B 28 74 D7 0F
```



# 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

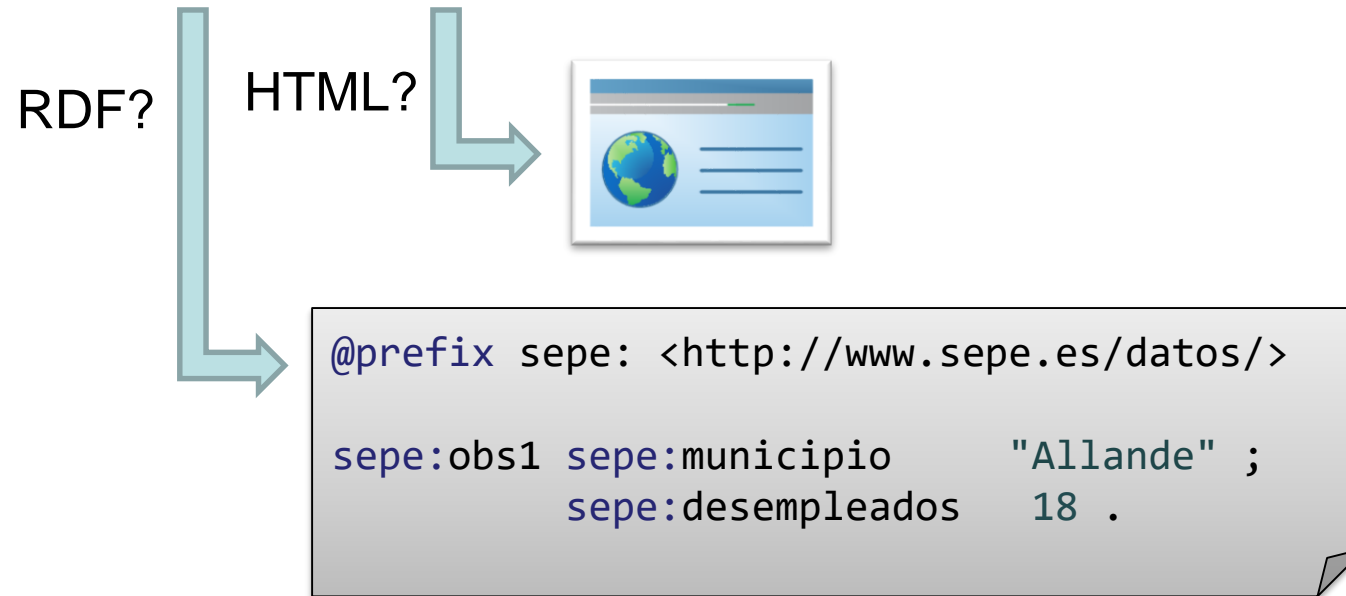
RDF para sistemas automáticos



# 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



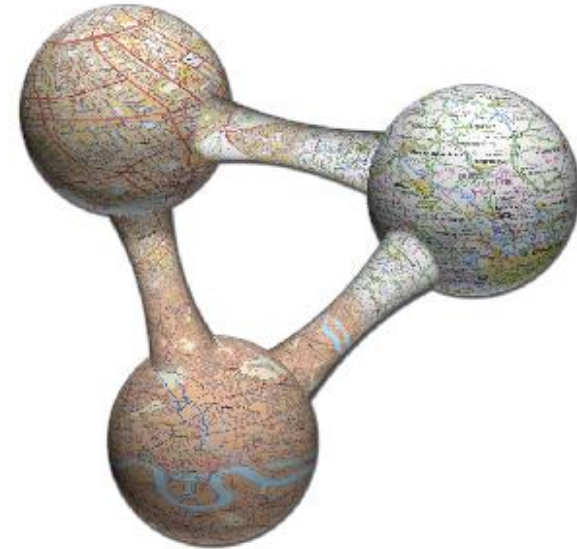
# Enlazar con otros datos ★ ★ ★ ★ ★

Las representaciones que se devuelven incluyen enlaces a otros datos

Permite:

Reutilizar y descubrir datos

Aplicaciones "*no previstas*"



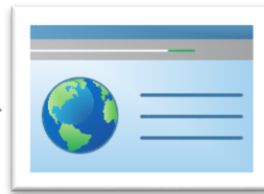
# Ejemplo: RDF bien enlazado



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

RDF?

HTML?



```
@prefix sepe: <http://www.sepe.es/datos/>  
sepe:obs1 sepe:municipio <http://dbpedia.org/resource/Allende>;  
          sepe:desempleados 23 .
```

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

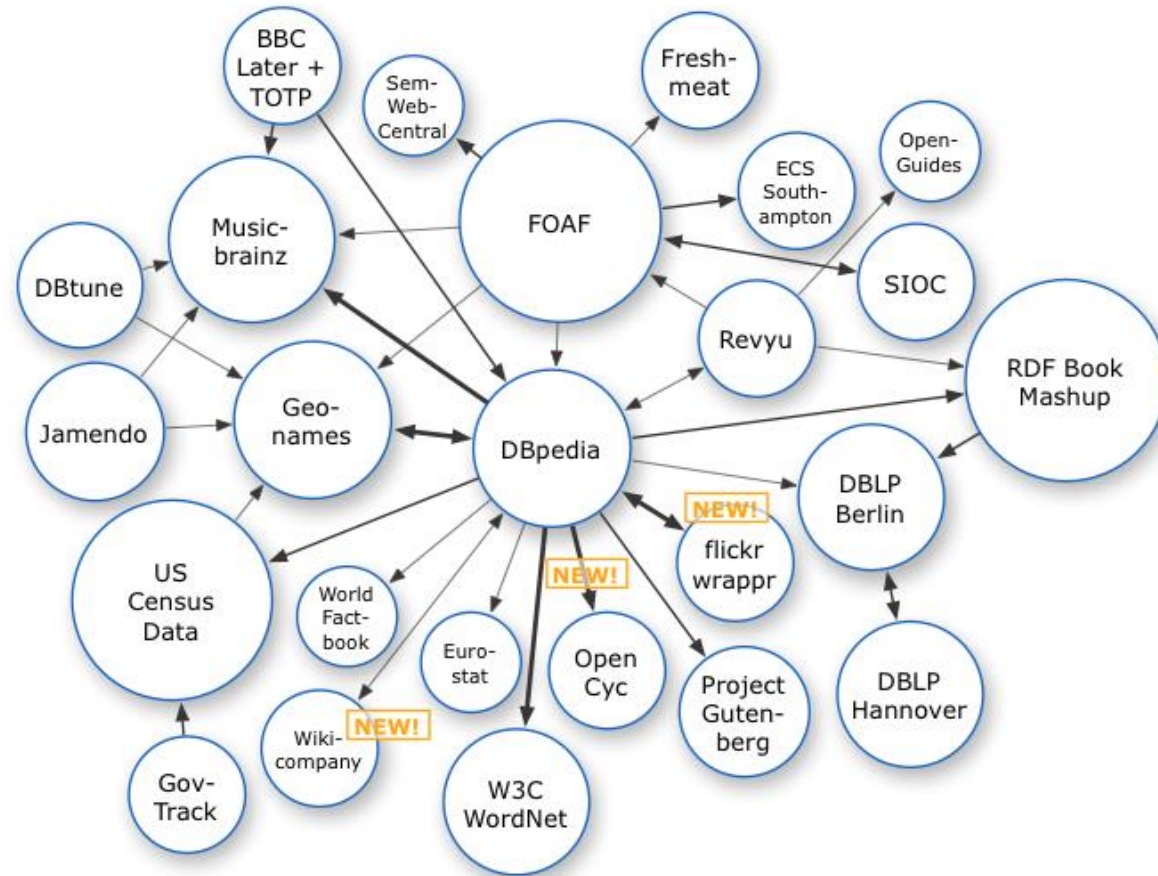
# Principios Linked Open Data

## 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>

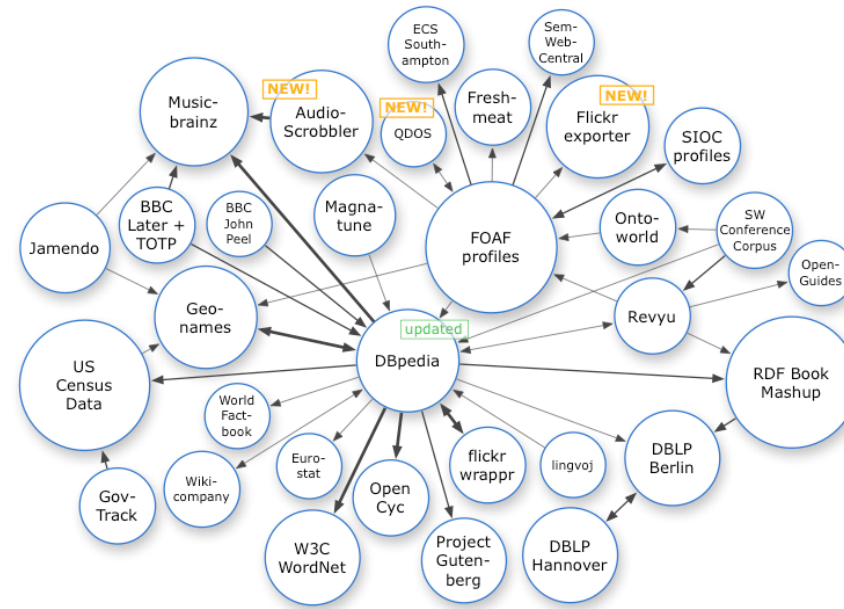
# LOD (2007)



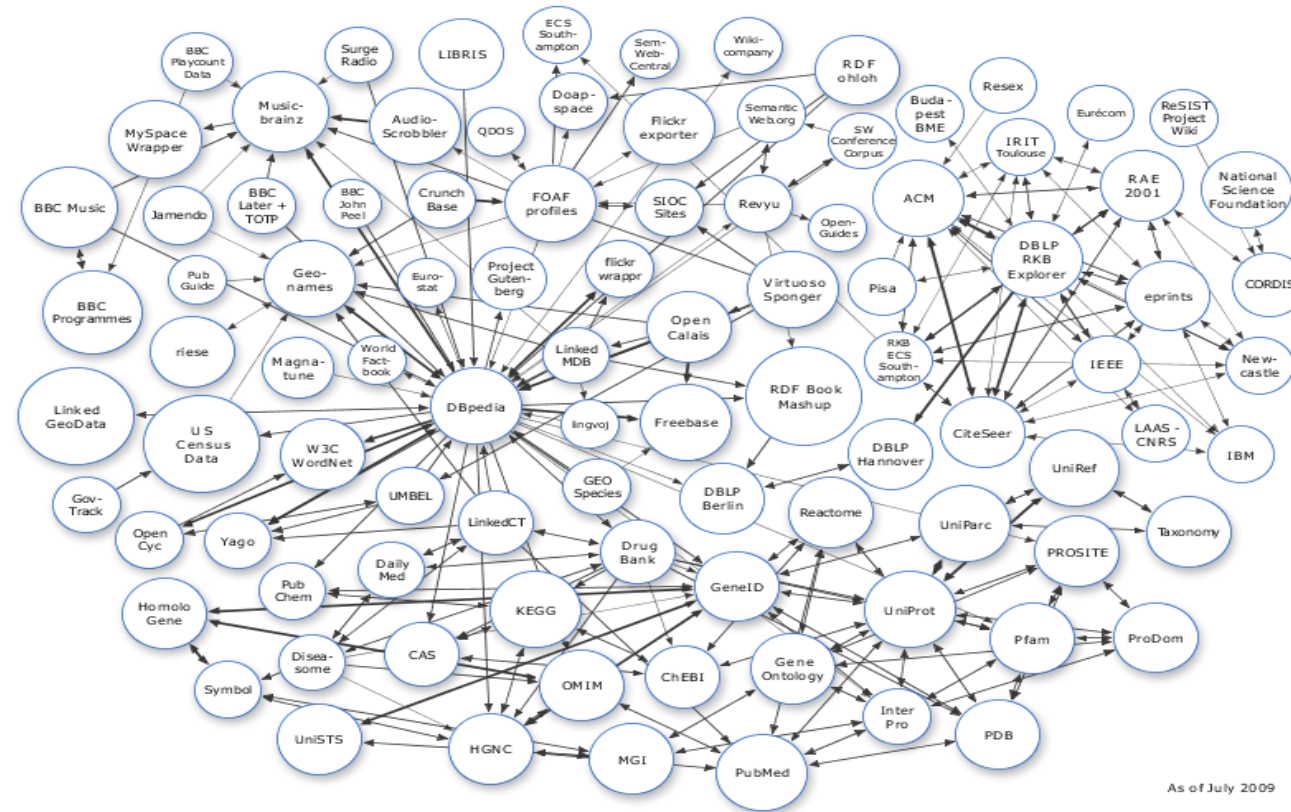
Evolución: <https://lod-cloud.net/>



# LOD (2008)



# LOD (2009)

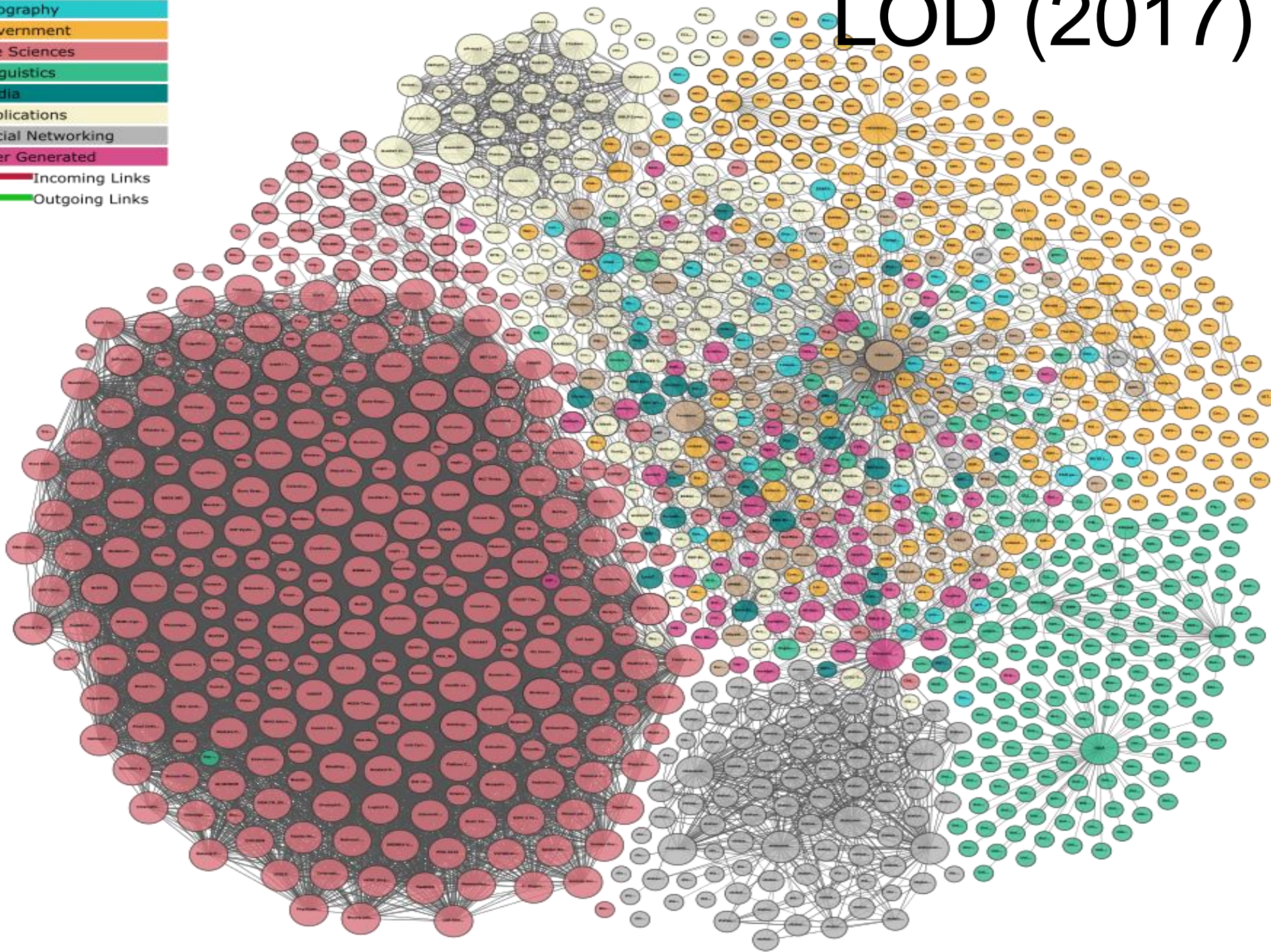




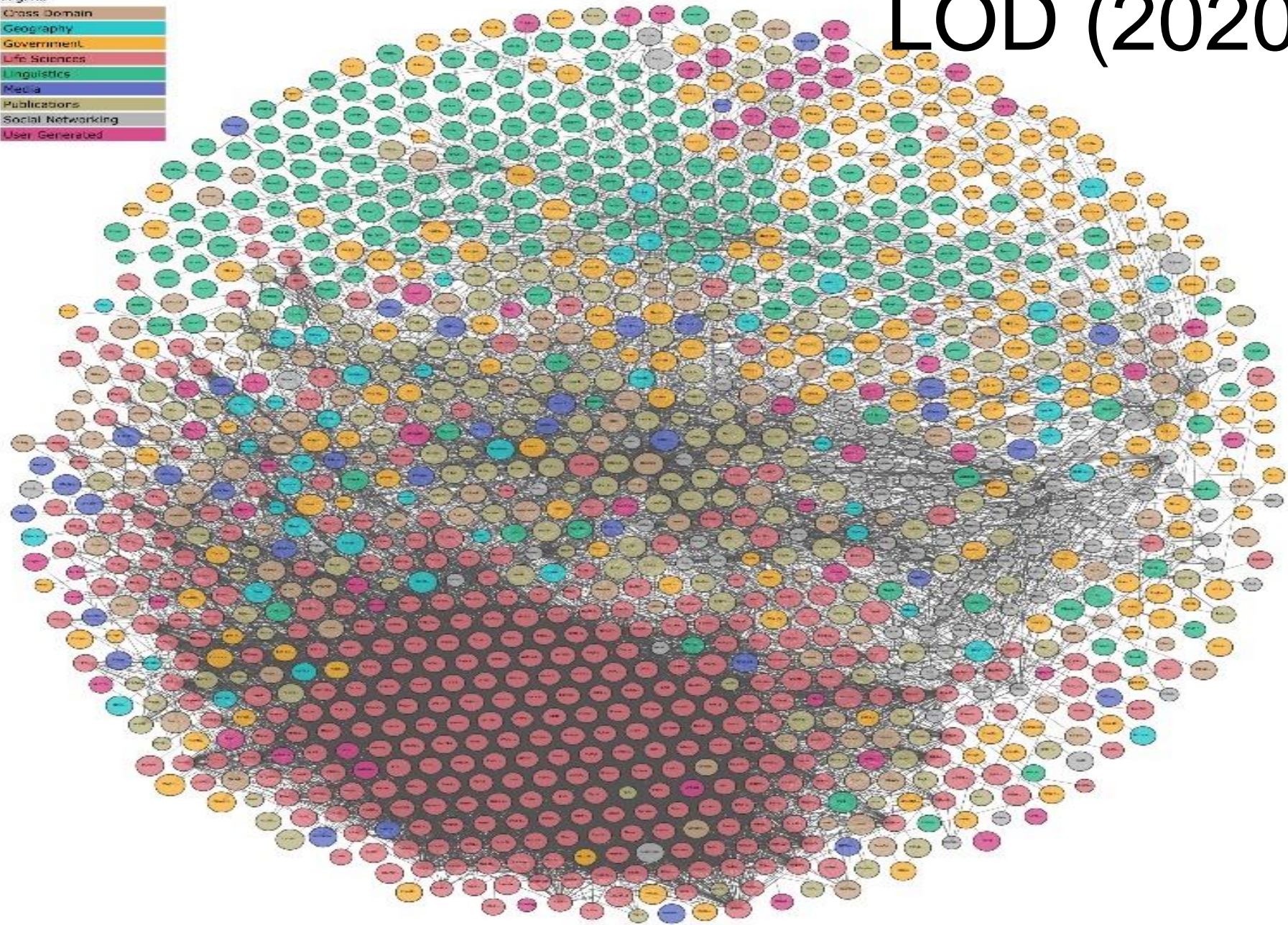




# LOD (2017)



# LOD (2020)



# Linked open data

## Ejemplos de iniciativas

[data.gov.uk](http://data.gov.uk)

[data.worldbank.org](http://data.worldbank.org)

[data.gov](http://data.gov)

[datos.gob.es](http://datos.gob.es)

[datos.gijon.es](http://datos.gijon.es)

...

[datos.bcn.cl](http://datos.bcn.cl)

[data.webfoundation.org](http://data.webfoundation.org)





# Beneficios de Linked Open Data

## Datos accesibles

- Evitar pérdidas semánticas al publicar

- Facilitar automatización de tareas

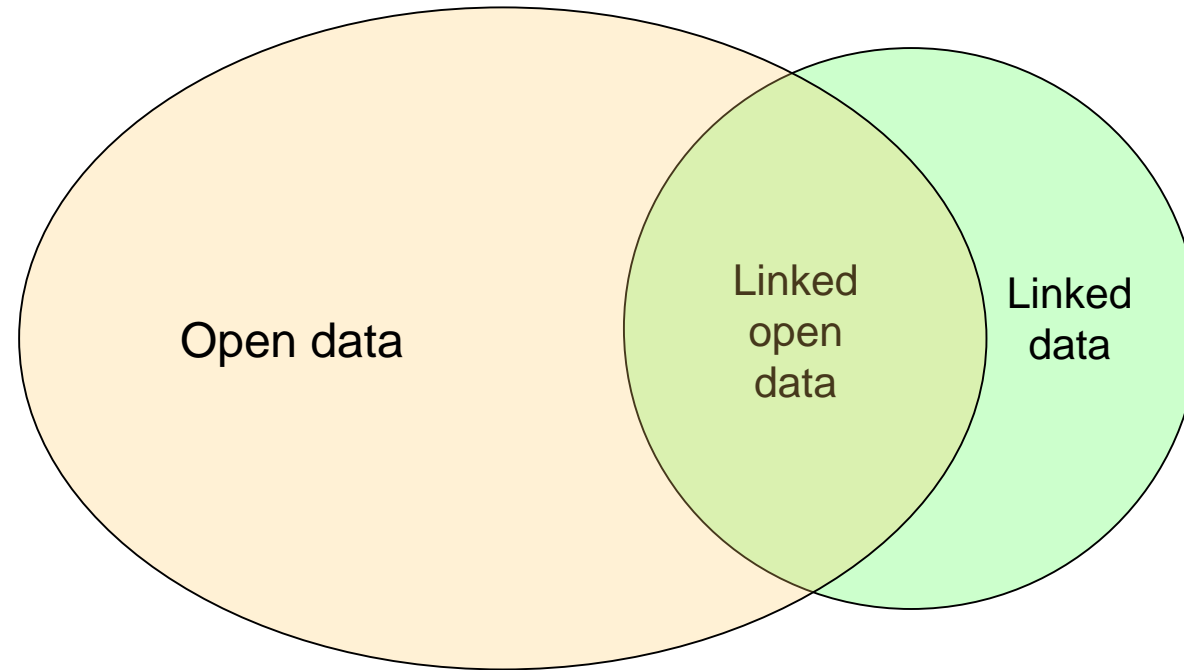
## Datos enlazados

- 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/>

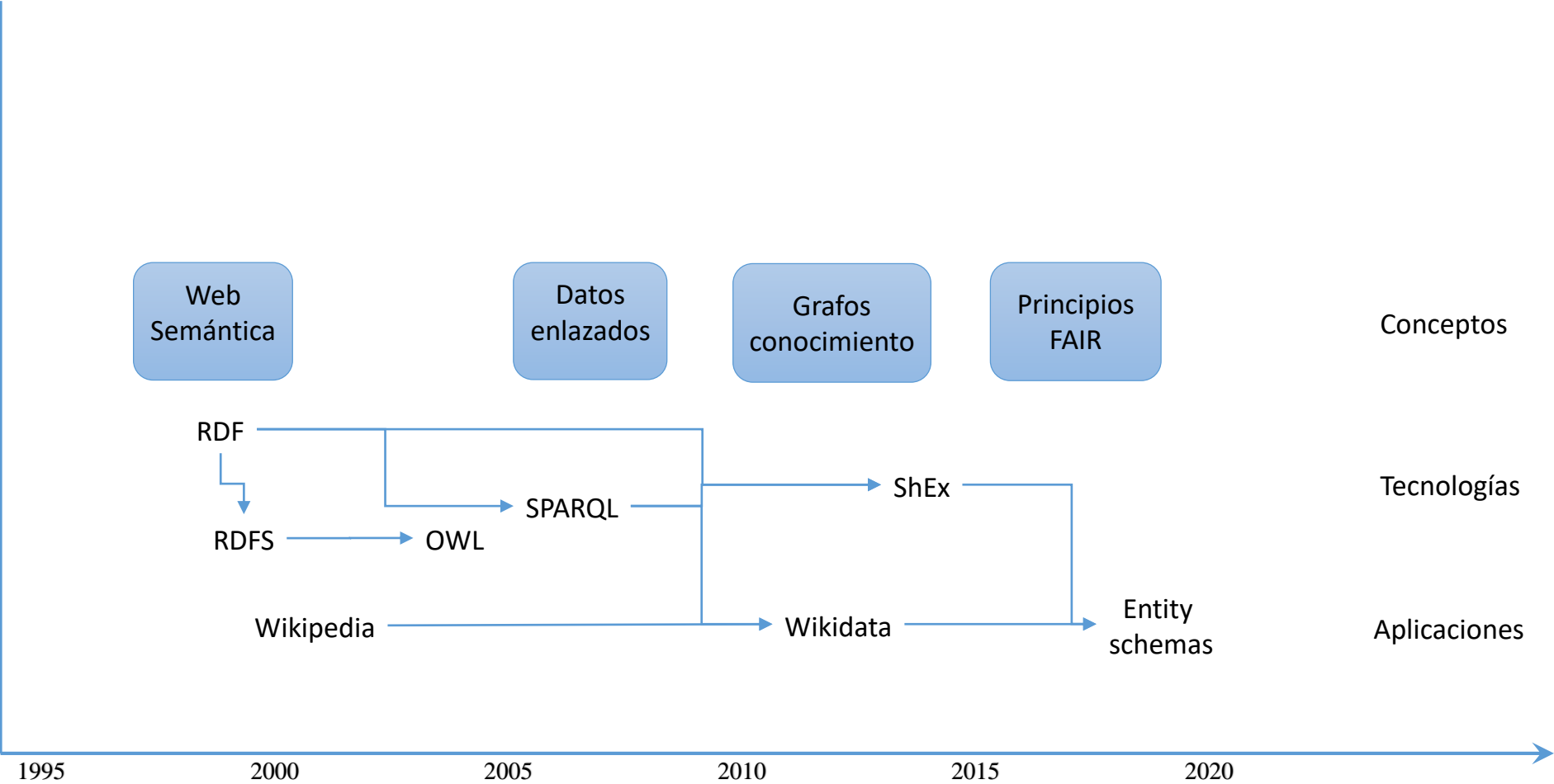
Similar a Linked Open Data

No es necesario que los datos sean abiertos

Ejemplo. Datos de pacientes

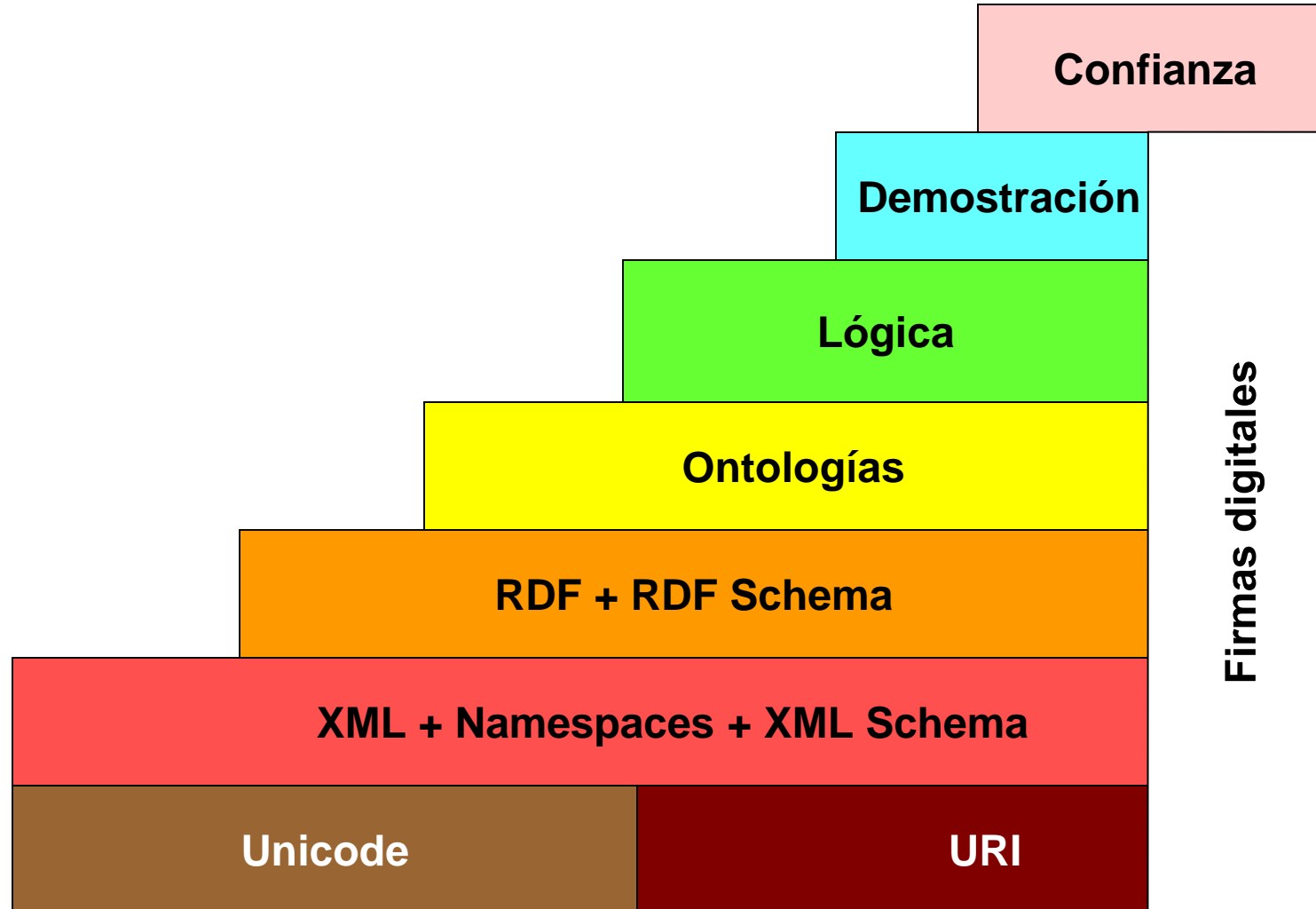
En FAIR no se hace tanto énfasis en tecnologías RDF

# Línea temporal conceptos Web Semántica



# Tecnologías Web Semántica

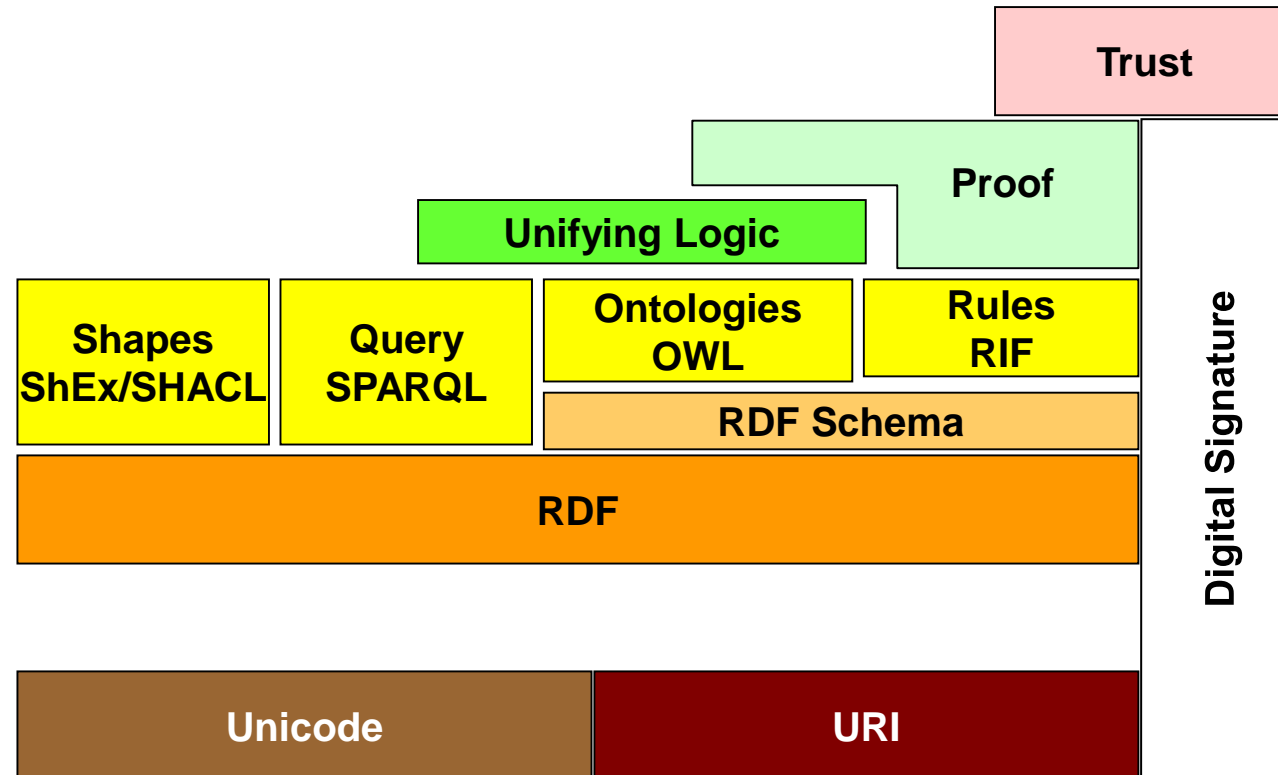
## Tarta de la Web (versión inicial)



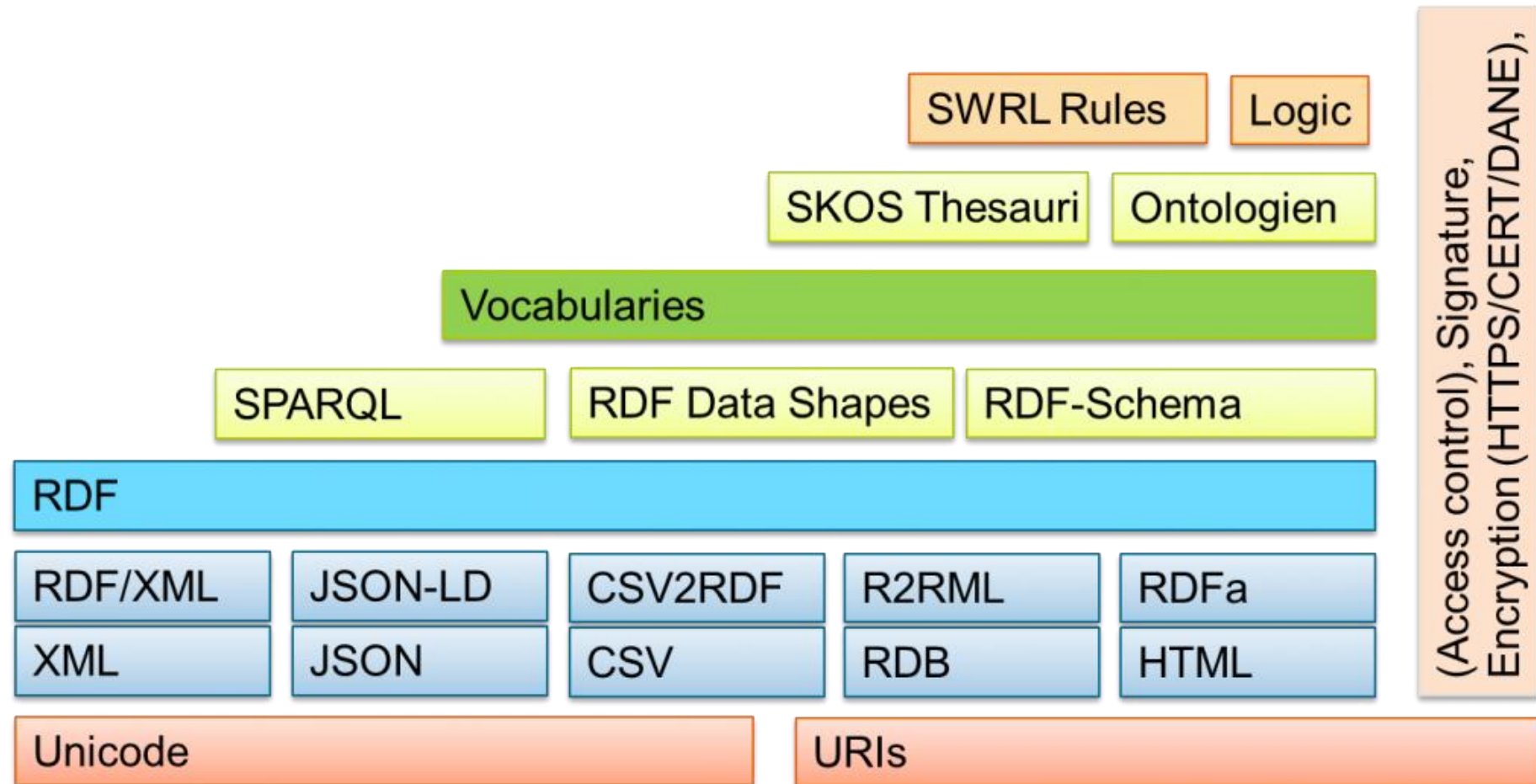
Versión propuesta por Tim Berners Lee, año 2000  
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

# Tecnologías Web Semántica

## Cambios en la tarta...



# Nuevas versiones de la tarta



# Algunas tecnologías

**RDF**  
Descripción datos

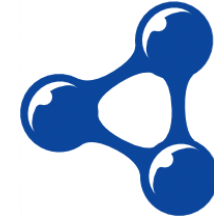
**SPARQL**  
Consultas

**SHEX - SHACL**  
Validación

**OWL - RDFS**  
Inferencias



# RDF



Resource Description Framework (1998)

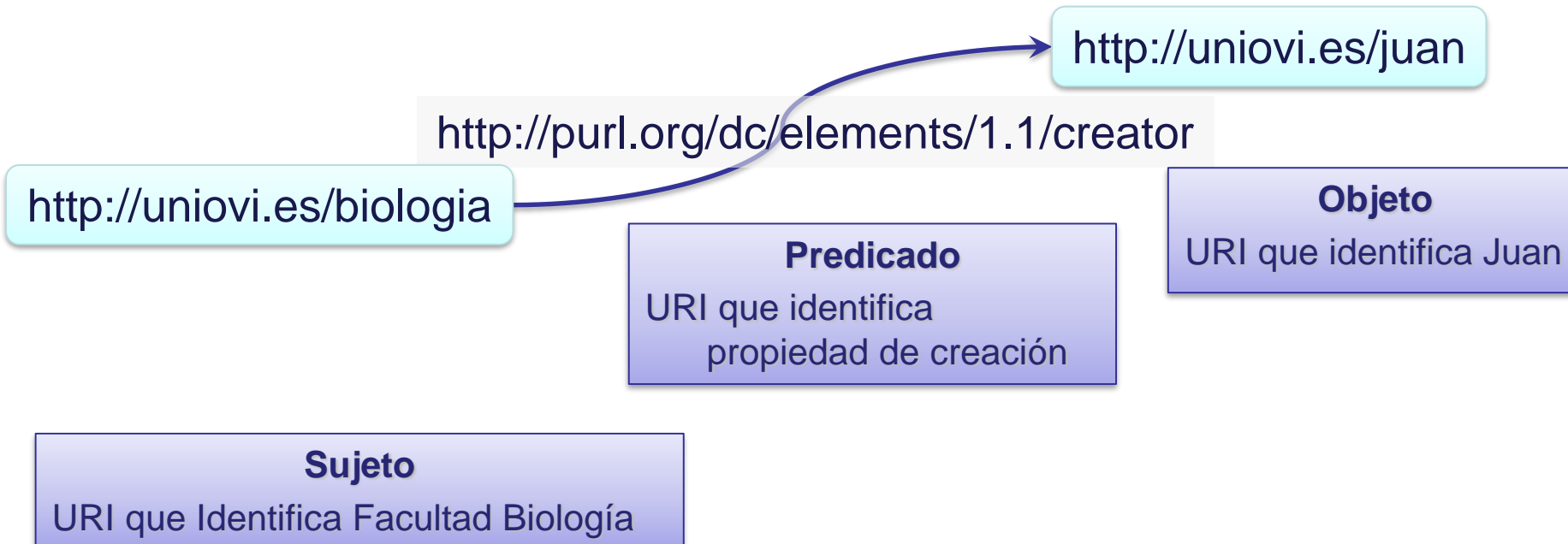
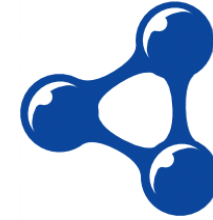
Descripción de recursos

Recurso = se identifica con URI

Se basa en tripletas

Sujeto → Predicado → Objeto

# Tripletas RDF

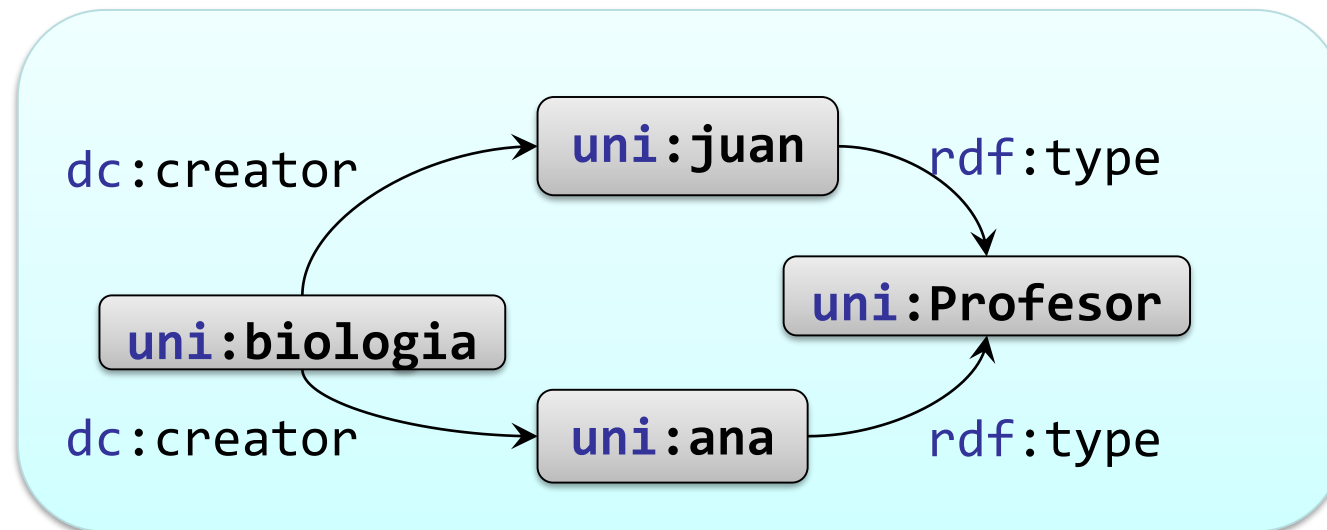
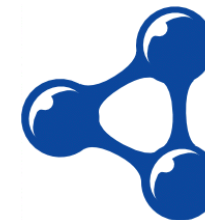


## RDF en notación Turtle

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

uni:biologia dc:creator uni:juan .
```

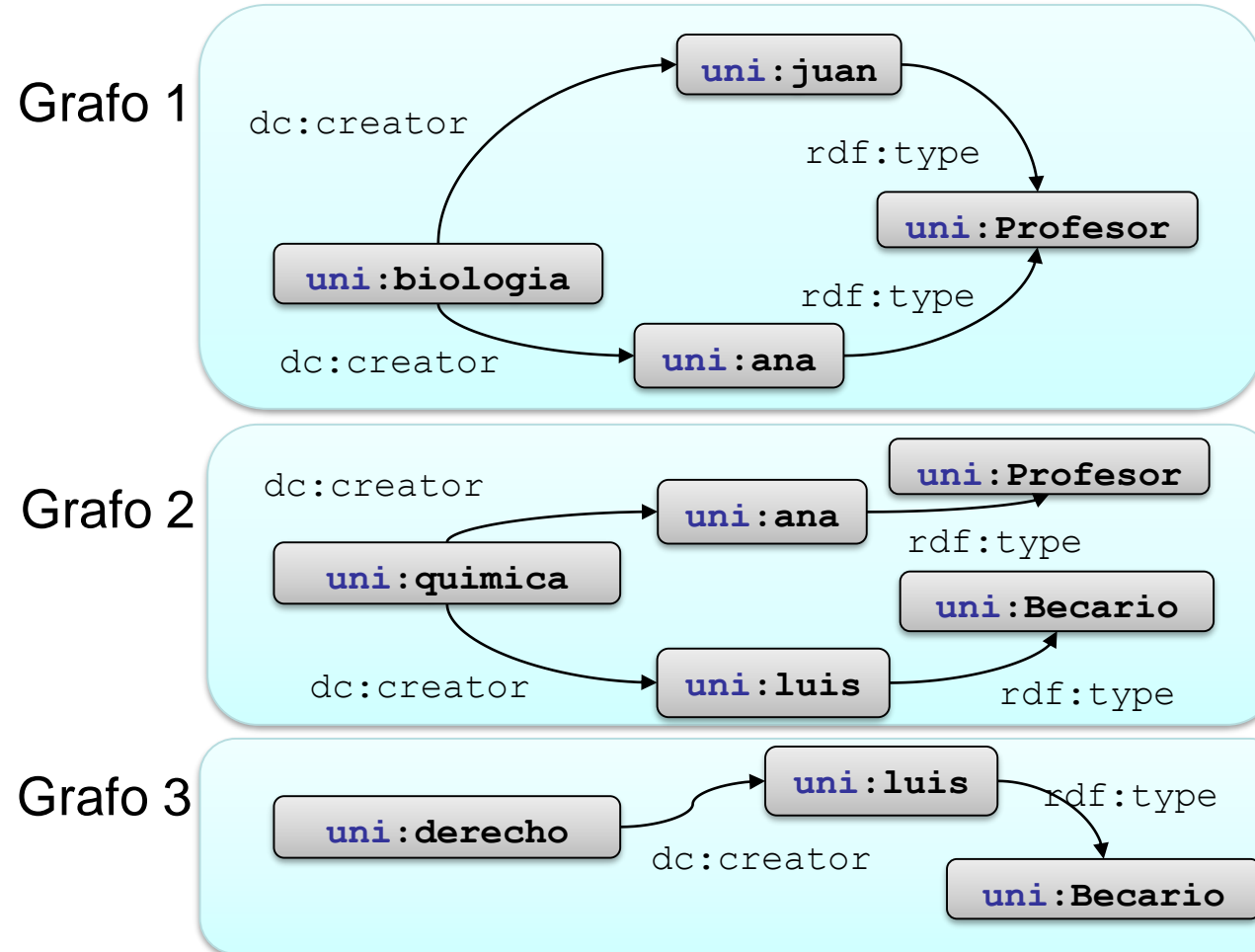
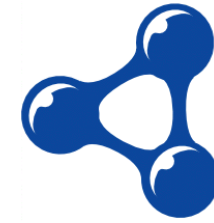
# Grafo RDF



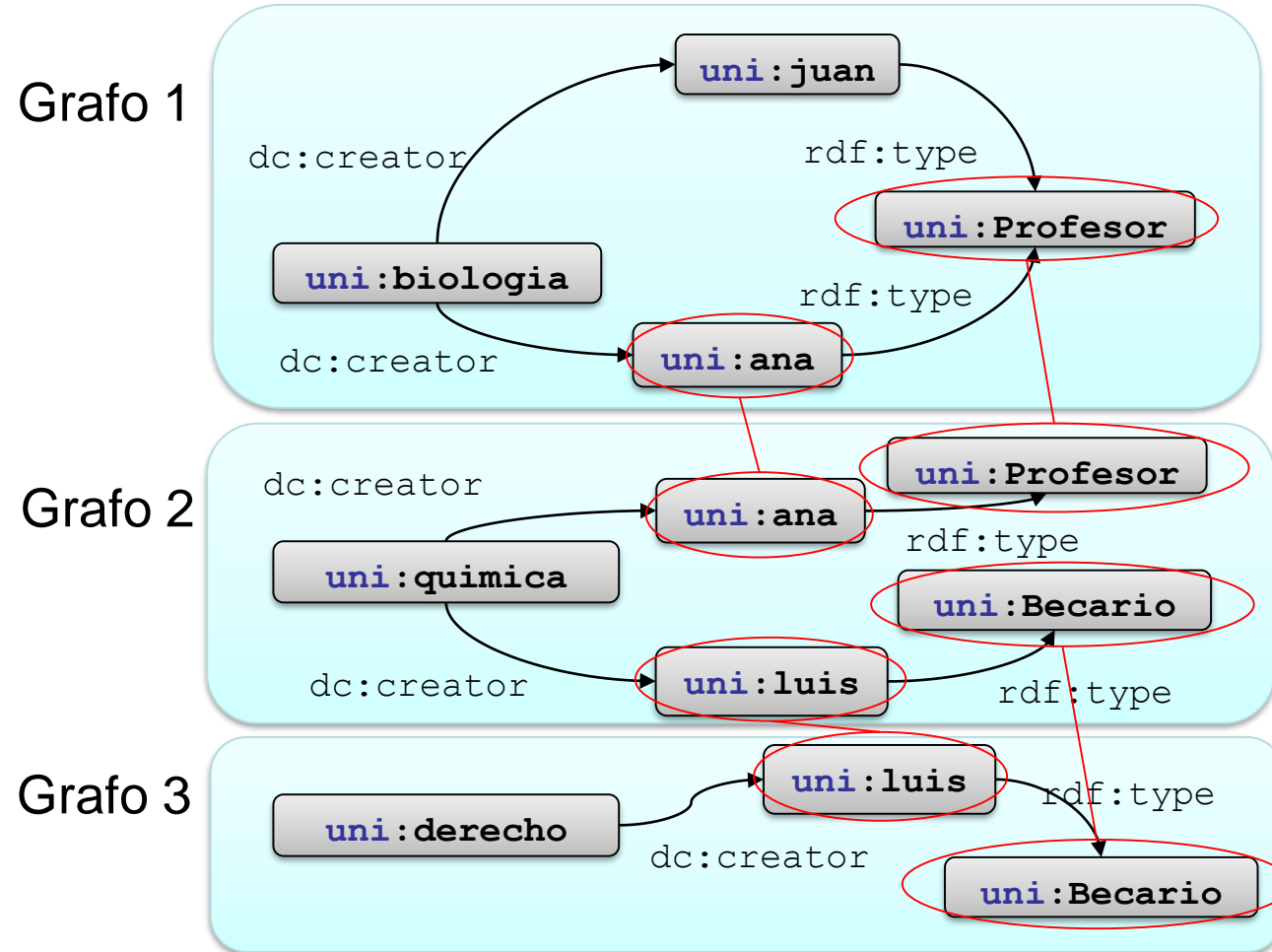
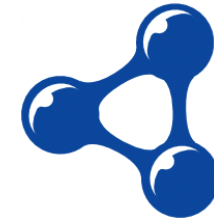
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix uni: <http://uniovi.es/> .  
@prefix dc: <http://purl.org/dc/elements/1.1/> .
```

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

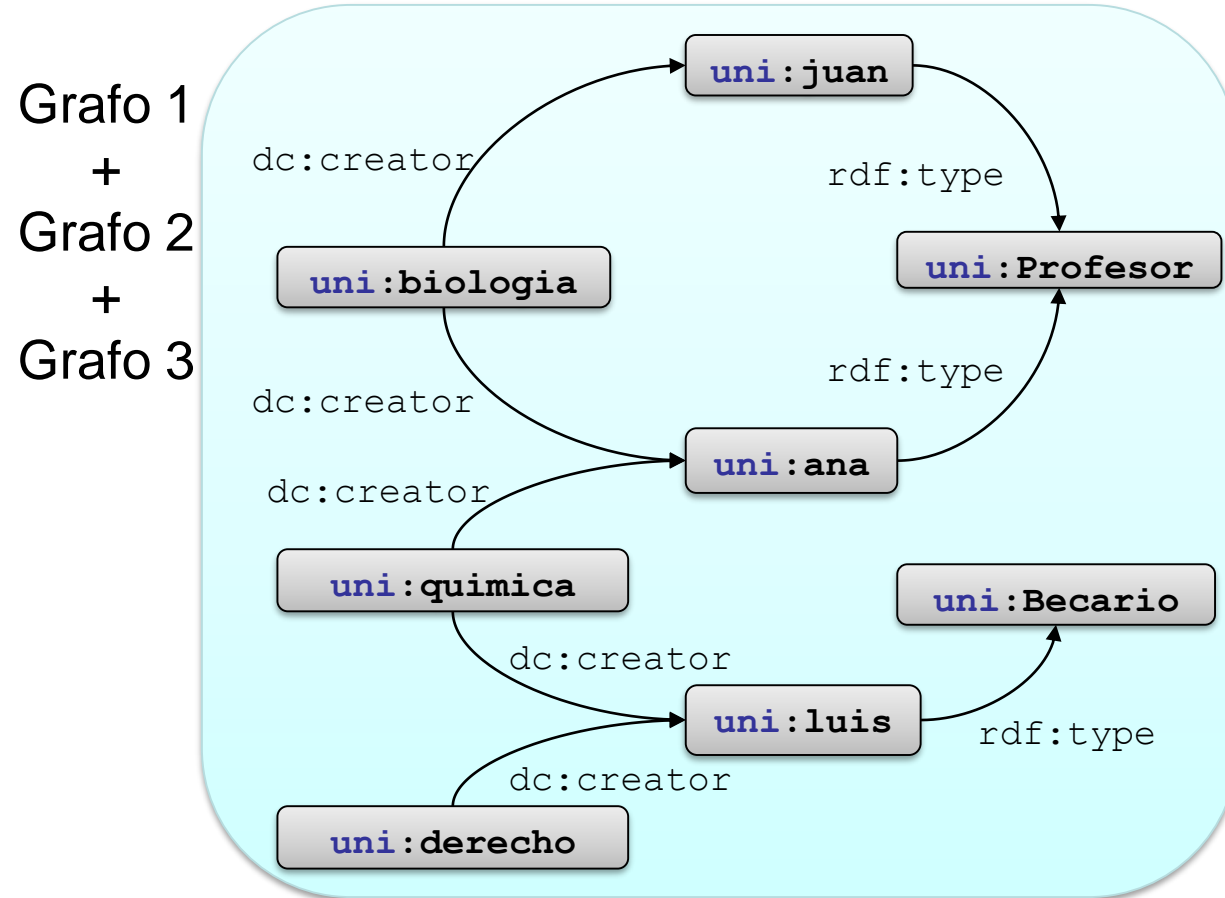
# RDF es composicional



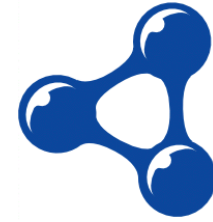
# RDF es composicional



# RDF es composicional



# Formatos RDF



Numerosos formatos y sintaxis:

N3

RDF/XML

N-Triples

Turtle

json-ld

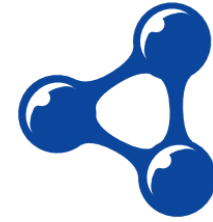
RDFa

etc.

...pero...

¡Lo más importante es el modelo de grafo!

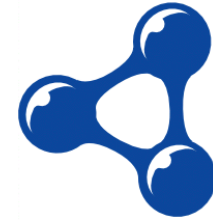
# SPARQL



- 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



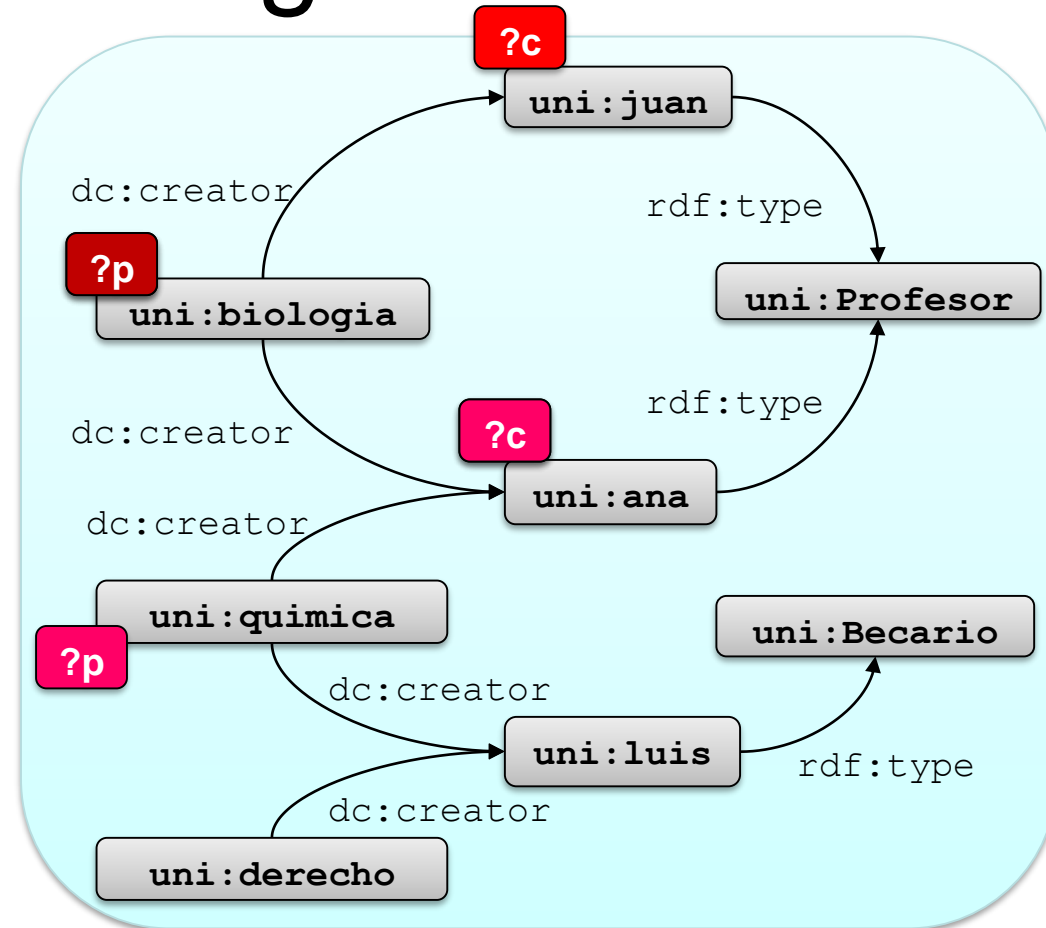
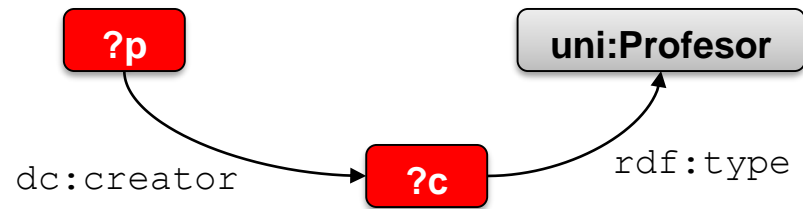
- Ejemplo:
- *Buscar páginas cuyo autor sea un profesor*

```
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

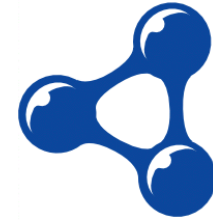
```
SELECT ?p ?c WHERE {  
  ?p dc:creator ?c .  
  ?c rdf:type uni:Profesor  
}
```



## Resultados

<code>?p</code>	<code>?c</code>
<code>uni:biologia</code>	<code>uni:juan</code>
<code>uni:biologia</code>	<code>uni:ana</code>
<code>uni:quimica</code>	<code>uni:ana</code>

# RDF Schema



Añade un vocabulario de esquema a RDF

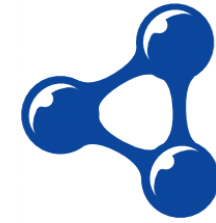
Class, Property, Resource,...

type, subclassOf, subPropertyOf,...

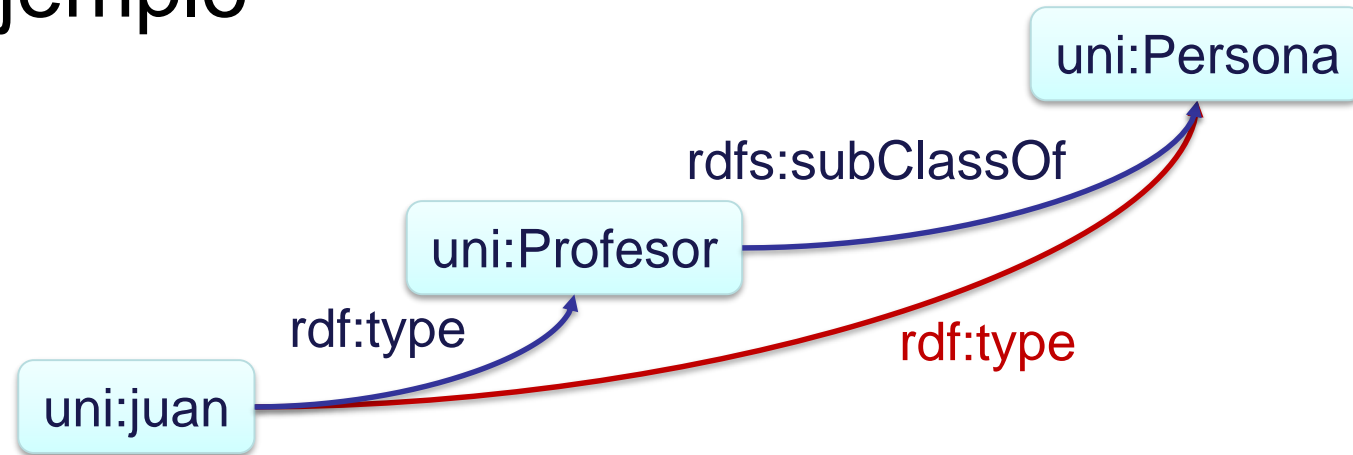
range, domain,...

RDF Schema permite **inferencias**

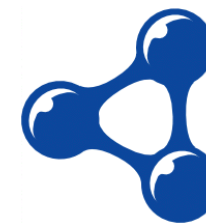
# RDF Schema



- Ejemplo



# SPARQL + Inferencia



Combinar SPARQL e inferencia

Ejemplo:

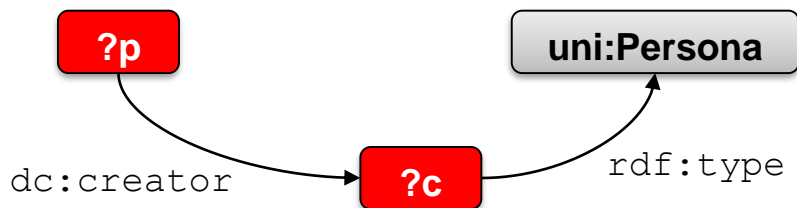
*Páginas cuyo autor sea una persona*

```
@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.
}
```

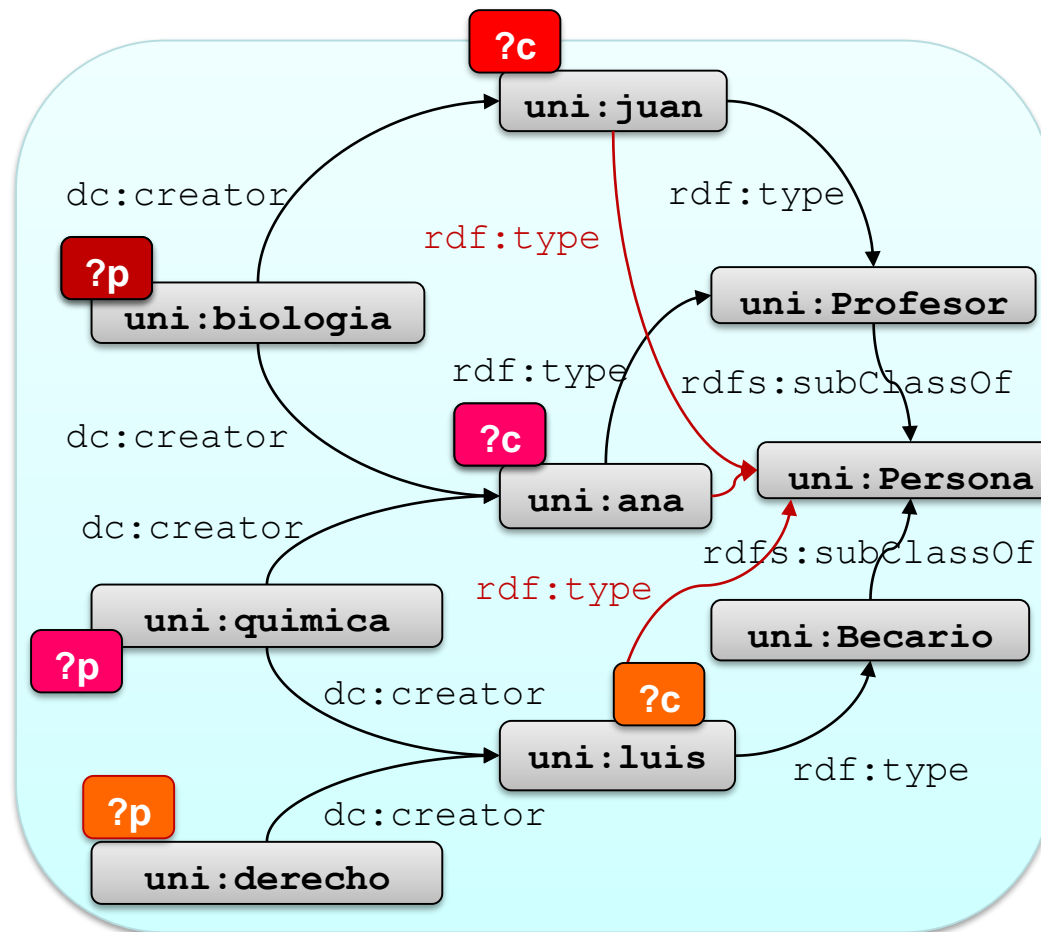
# SPARQL con inferencia

```
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*

