

# Linked Data

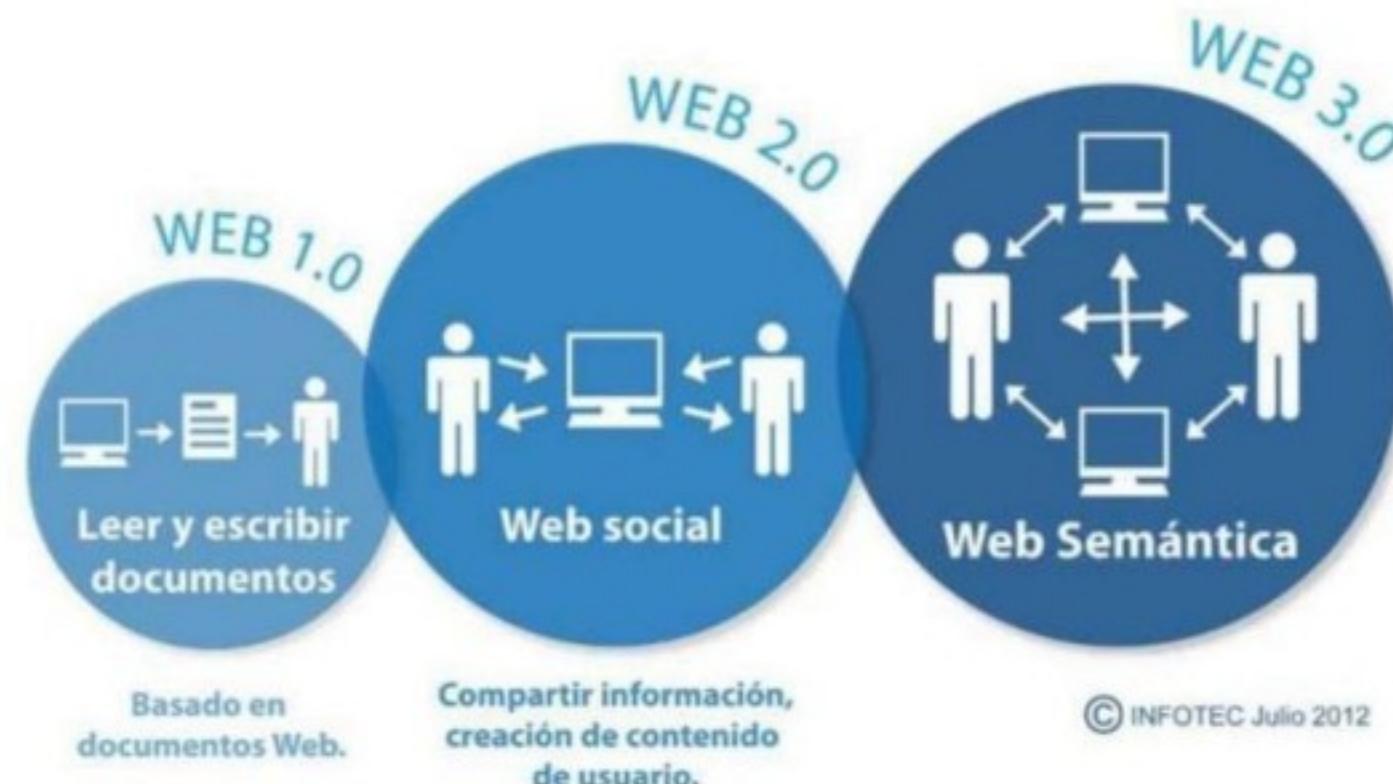
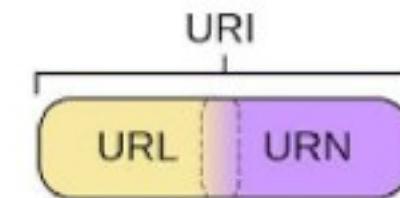
Fernando Aguilar

# Conceptos anteriores

#CSIC

- Protocolo HTTP
- Persistent Identifiers (PIPs) – URI (Uniform Resource Identifier)
- Metadatos
- XML, HTML, JSON
- Web (1.0, 2.0, Semántica)

Web 1.0, 2.0 y 3.0



- Primera comunicación entre un cliente y un servidor usando el protocolo HTTP en noviembre de 1989 (Tim Berners-Lee, CERN).
- WEB: HTTP, HTML, URL
- La Web 1.0 es Internet original. Permitía leer desde la red.
- La parte dinámica: hipervínculos que enlazaban unas páginas con otras (red, web).
- La Web se entendía como portal de información.
- Tablas a posiciones para alinear los elementos en la página.



- En esta nueva plataforma, los usuarios tienen la posibilidad de controlar sus datos.
- Bidireccional
- Contenido generado por el “webmaster” y el usuario.
- Cualquiera puede crear contenido: comentarios, documentos, vídeos...
- Más dinamismo: bases de datos, interacción directa con el usuario.
- Web 2.0: tres conceptos... (siguiente)



Rich Internet Application (RIA) : It defines the experience brought from desktop to browser .

Service- Oriented Architecture (SOA) : defines how web 2.0 applications expose its functionality so that other application can integrate the functionality and produce a set of much richer applications. (feeds, RSS, )

Social Web : it defines how web 2.0 tend to interact much more with the end user and making the end user an integral part.

- Nombre sugerido por John Markoff del New York Times para referirse a la tercera generación de la web.
- En esta generación, todas las aplicaciones en web o smart phones se enriquecen con más características.
- Acercando el “Escritorio” a entornos online.
- Aplica el mismo principio que la Web 2.0: bidireccionalidad, interacción en dos direcciones.
- Web 3.0: más conectada, abierta, inteligente. Con tecnologías semánticas, bases de datos distribuidas, procesamiento de lenguaje natural, machine learning, machine reasoning, agentes autónomos, etc. Semantic Web, Linked Data y más...
- Prepara la información para posibles (desconocidos) reutilizaciones.
- Linked Data ≠ Open Data | Open Linked Data

# Ejemplos Web 1.0, 2.0, 3.0

#CSIC

web 2.0 :YouTube

2.0 Facebook

web 2.0: LinkedIn

web 2.0: github

1.0: IFCA

1.0 periodico online

web 1.0: Marca

1.0: páginas web con los  
comentarios desahabilitados  
2.0 : Facebook, Instagram,  
twitter

web 2.0: Wikipedia

# Ejemplos Web 1.0, 2.0, 3.0

#CSIC

Web 1.0: página personal tipo curriculum  
Web 2.0: una red social, Twitter  
Web 3.0: Amazon

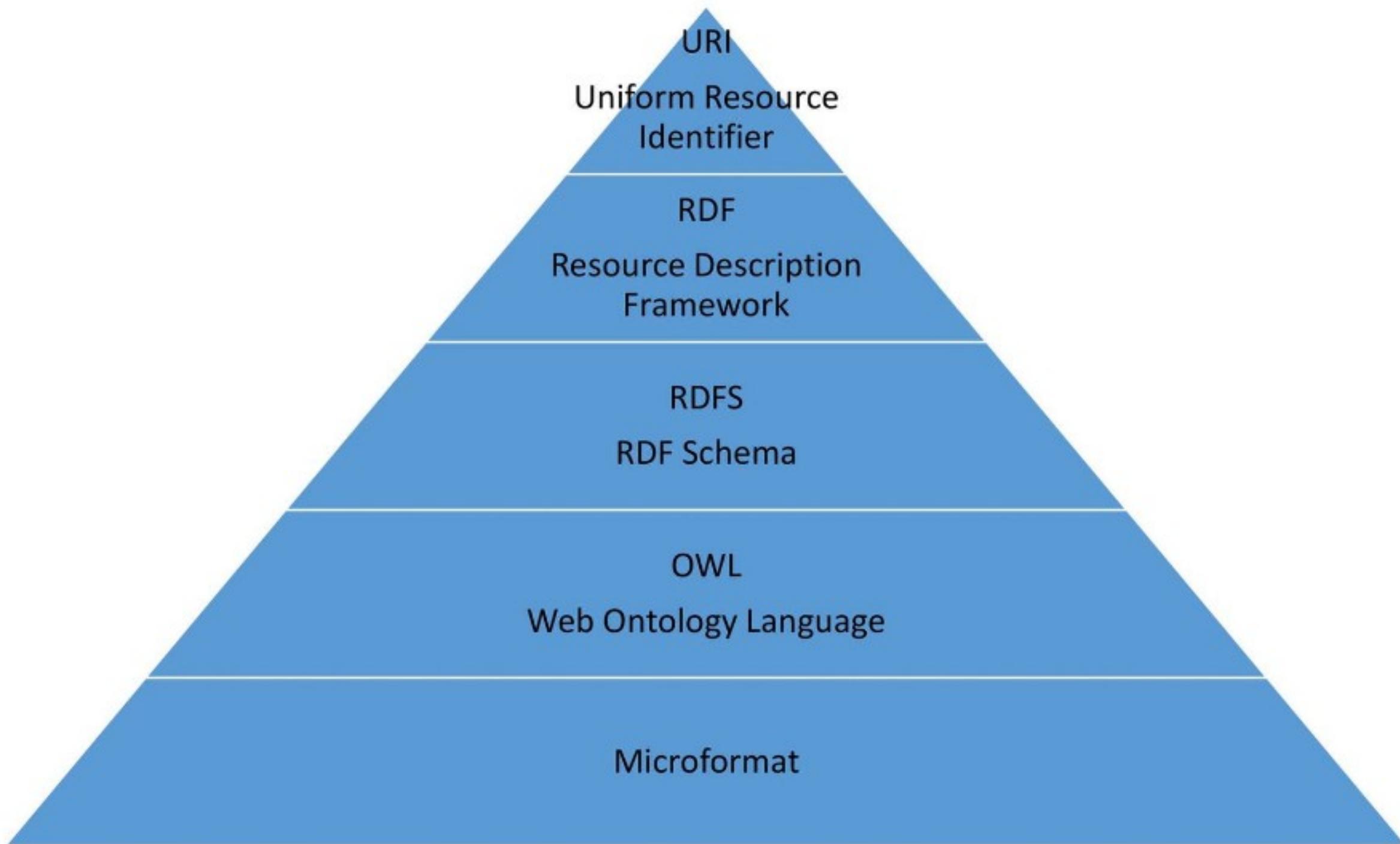
3.0 Retrievr

web 2.0: Twitter

:)

# Semantic web as a component of web 3.0

#CSIC



- Uniform resource identifier (URI) es la forma para identificar todos los puntos de contenido, independientemente si son una página, un texto, un vídeo, mp3, gif o un programa. Recurso digital.
- Sirve para identificar el recurso. Sin URI, no podemos hablar de web semántica.
- La forma más común de una URI es la dirección de una página web, establecida de una cierta forma: URL o Uniform Resource Locator.
- Una URI describe:
  - El mecanismo para acceder al recurso
  - El servidor donde está alojado
- Ejemplo: [http://www.w3.org/Icons/WWW/w3c\\_main.gif](http://www.w3.org/Icons/WWW/w3c_main.gif)

- Resource Description Framework (RDF) de la World Wide Web Consortium (W3C) es una serie de especificaciones originalmente diseñadas para modelos de metadatos.
- Este framework es utilizado para representar datos en la “nube” LOD (linking open data), LD, como ficheros XML entre otros formatos.
- Es también una forma de definir recursos utilizando un framework particular.
- LOD : Linked Open Data. Los datos en la “nube” LOD cloud son abiertos y libremente accesibles. Libres en el sentido de que los datos pueden ser creados, modificados por los propios usuarios.
- Protegé...

- RDF Schema (Resource Description Framework Schema, variously abbreviated as RDFS, RDF(S), RDF-S, or RDF/S)
- Es un conjunto de clases con ciertas propiedades utilizando el “extensible knowledge representation language” RDF
- Proporciona elementos básicos para la descripción de ontologías, llamadas también vocabularios RDF.
- Define las estructuras de los RDF, relaciones, elementos, etc.
- Semantic Extension RDF .

- Las ontologías definen los términos utilizados para describir y representar un área de conocimiento. Son utilizadas por personas, bases de datos y aplicaciones que necesitan compartir cierta información bajo un dominio (por ejemplo, un área de conocimiento, como la medicina).
- El lenguaje OWL (Web Ontology Language) está diseñado para ser utilizado por aplicaciones que necesitan procesar contenidos de información en lugar de solamente presentar información a humanos.
- Se definen un esquema de relaciones para casos específicos (por ejemplo, en ciencias o para aplicaciones específicas de empresas, como comercio electrónico).

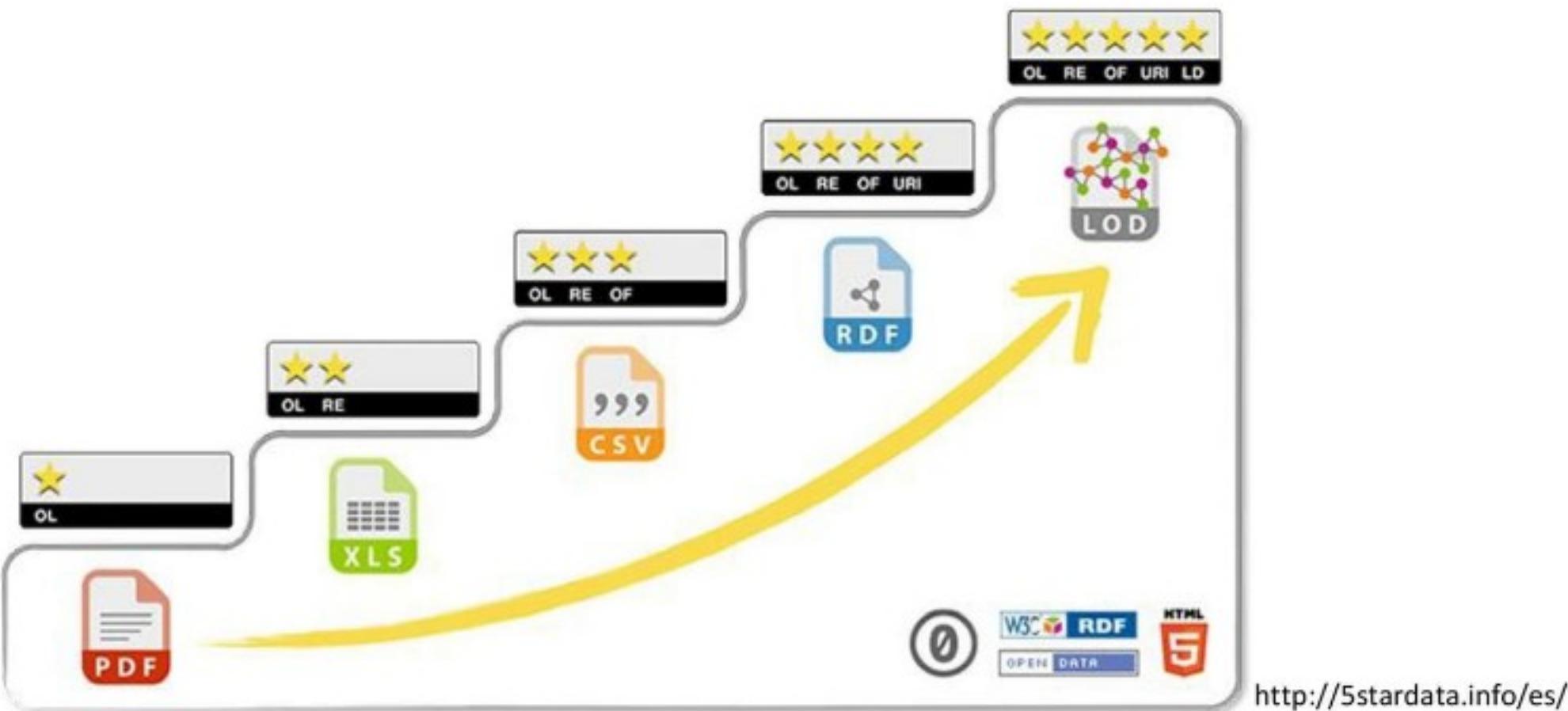
- Los microformatos (Microformats) son pequeños patrones de HTML para representar entidades publicadas como personas, eventos, posts en blogs, revisiones y etiquetas en páginas web.
- Microformats son la forma más simple y rápida de proporcionar una API a la información de tu sitio web.
- Microformats intentan resolver problemas simples adaptándose a los patrones de uso (e.g. XHTML, blogging).
- Ejemplo: embeder información relevante en tu web, para bots, extraer información semántica, etc.

```
<ul class="vcard">
    <li class="fn n">
        <strong class="given-name">Guillermo</strong> <strong class="additional-name">García</strong><br />
        <span class="org">El Ejemplo S. A.</span>
    </li>
    <li class="adr">
        <span class="street-address">Calle falsa 1</span><br />
        <span class="locality">una ciudad</span>,
        <abbr class="region" title="Una región">UR</abbr>,
        <span class="postal-code">94301</span>,
        <abbr class="country-name" title="Un país">UP</abbr>
    </li>
    <li class="tel"><strong class="type" title="Teléfono del trabajo">Work</strong>; <span class="value">604-555-1234</span>
    </li>
    <li class="url"><strong class="type" title="Sitio web oficial del trabajo">Work</strong>; <a href="http://ejemplo.com/">http://ejemplos.com/</a></li>
</ul>
```

<http://microformats.org/wiki/h-card>

# 5 Estrellas Datos Abiertos

Tim Berners-Lee, el inventor de la Web e iniciador de los Datos Enlazados (Linked Data), sugirió un esquema de desarrollo de 5 estrellas para Datos Abiertos. A continuación te mostramos ejemplos para cada escalón o nivel de estrellas y te explicamos los costos y beneficios involucrados en cada caso.



# 5 Estrellas Datos Abiertos

#CSIC

- ★ Publica tus datos en la Web (con cualquier formato) y bajo una licencia abierta<sup>1</sup>
- ★★ Publícalos como datos estructurados (ej: Excel en vez de una imagen de una tabla escaneada)
- ★★★ Usa formatos no propietarios (ej: CSV en vez de Excel)
- ★★★★ Usa URIs para denotar cosas, así la gente puede apuntar a estas
- ★★★★★ Enlaza tus datos a otros datos para proveer contexto

# Actualmente: recursos de datos distribuídos en diversos formatos

- Datos « encerrados » en distintos formatos: APIs, Datos en HTML o desestructurados
- 2710 Databases de neurociencia(Neuroscience Information Framework)



### Metadatos:

- DC: Autor, Título, etc.
- EML: Administrativos, Contextuales, Estructurales.
  - Atributos
  - Proyecto

### Metadatos:

- DC: Autor (URI, type:Person), Título.
- EML: Administrativos, Contextuales, Estructurales.
  - Atributos (type:Press, Unit)
  - Proyecto (type:project fundedBy EC ...)

### Datos

date	Temp	Press	Cond
2010-01-27 11:44:00	2.4712041875	2.6513564375	0.03858075
2010-01-27 14:14:00	3.0359613913043	3.5679646521739	0.039838869565217
2010-01-27 15:00:00	3.0426879833333	3.5414305833333	0.039633033333333
2010-01-27 16:00:00	3.0361835517241	3.545420637931	0.039581586206897

### Datos (Machine – Human)

date	Temp	Press	Cond
2010-01-27 11:44:00	2.4712041875	2.6513564375	0.03858075
2010-01-27 14:14:00	3.0359613913043	3.5679646521739	0.039838869565217
2010-01-27 15:00:00	3.0426879833333	3.5414305833333	0.039633033333333
2010-01-27 16:00:00	3.0361835517241	3.545420637931	0.039581586206897

Búsquedas: datos del mismo proyecto, datasets con mismos atributos, proyectos similares...

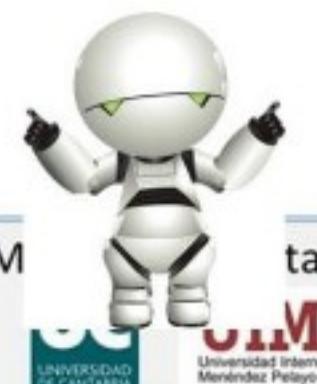
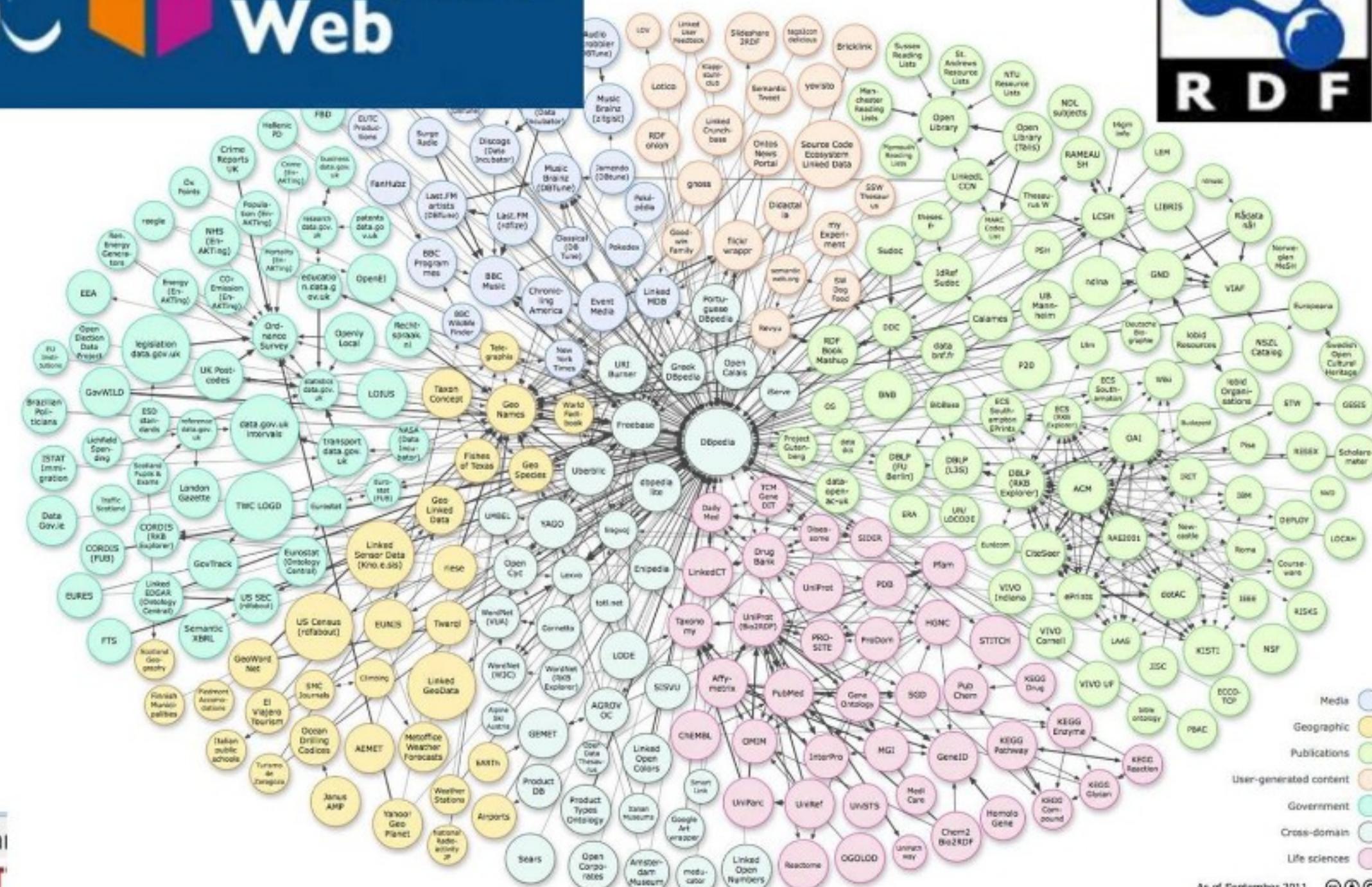
# Un problema Global

#CSIC

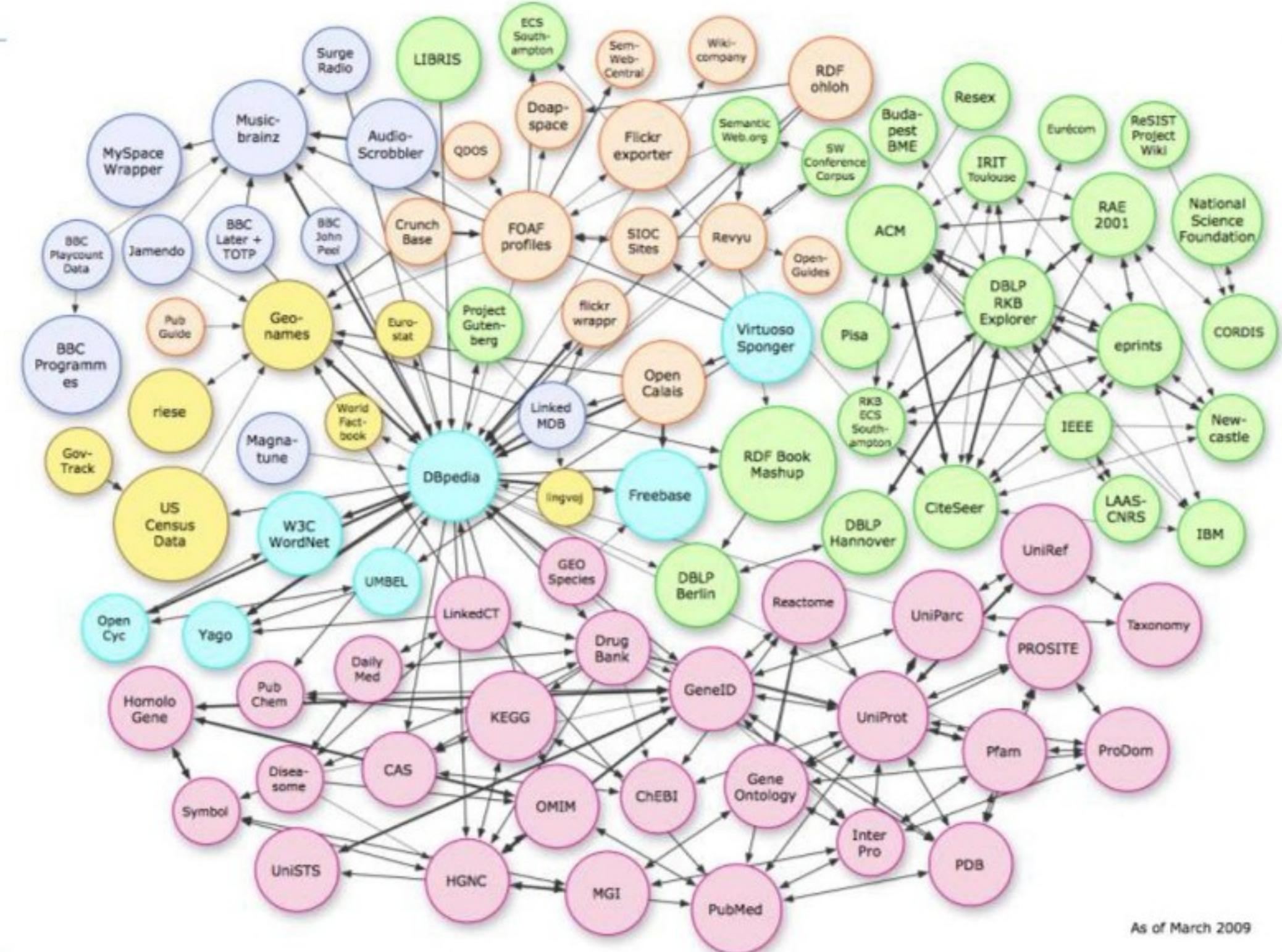
- World Wide Web: un espacio global de documentos
- Interconexión mediante enlaces (links)
- Datos « Escondidos » en HTML: Sencillo de entender por humanos. No tanto por máquinas.
- Diversidad de APIs Web.
- Complicado de acceder y enlazar datos
- Necesidad de semántica para transformar el espacio global de documentos en un espacio global de datos interconectados.

## Semantic web – Linked Data

#CSIC



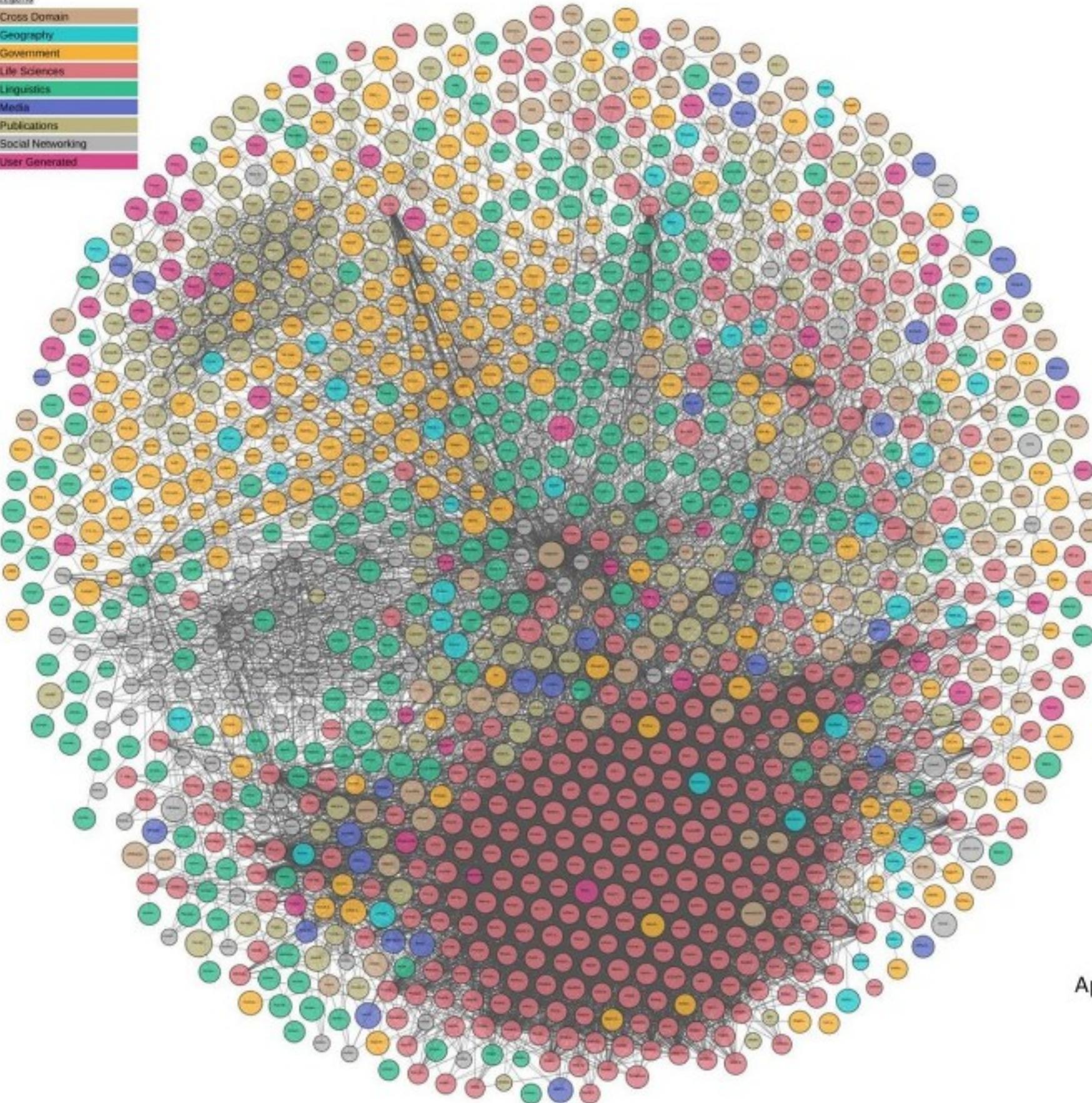
#CSIC



As of March 2009

# #CSIC

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



April 2019



# 4 Reglas básicas Linked data



1. Usar **URIs** para nombrar recursos.
2. Usar **HTTP** URIs para que se puedan buscar esos nombres.
3. Cuando alguien busca una URI, proporcionar información útil usando los estándares (**RDF**, SPARQL, Metadatos).
4. Incluir enlaces a otras URIs, para poder descubrir más recursos.

HTTP + URIs + RDF = Web of Data

# Name things with URIs

#CSIC

Use URI instead of URN (Uniform Resource Name) and DOIs

Separate the URI representing the real object or concept from its description

## Example

Real Person

[http://www.esciencedatafactory.com/people/yann\\_le\\_fran](http://www.esciencedatafactory.com/people/yann_le_fran)

Description RDF (for machines)

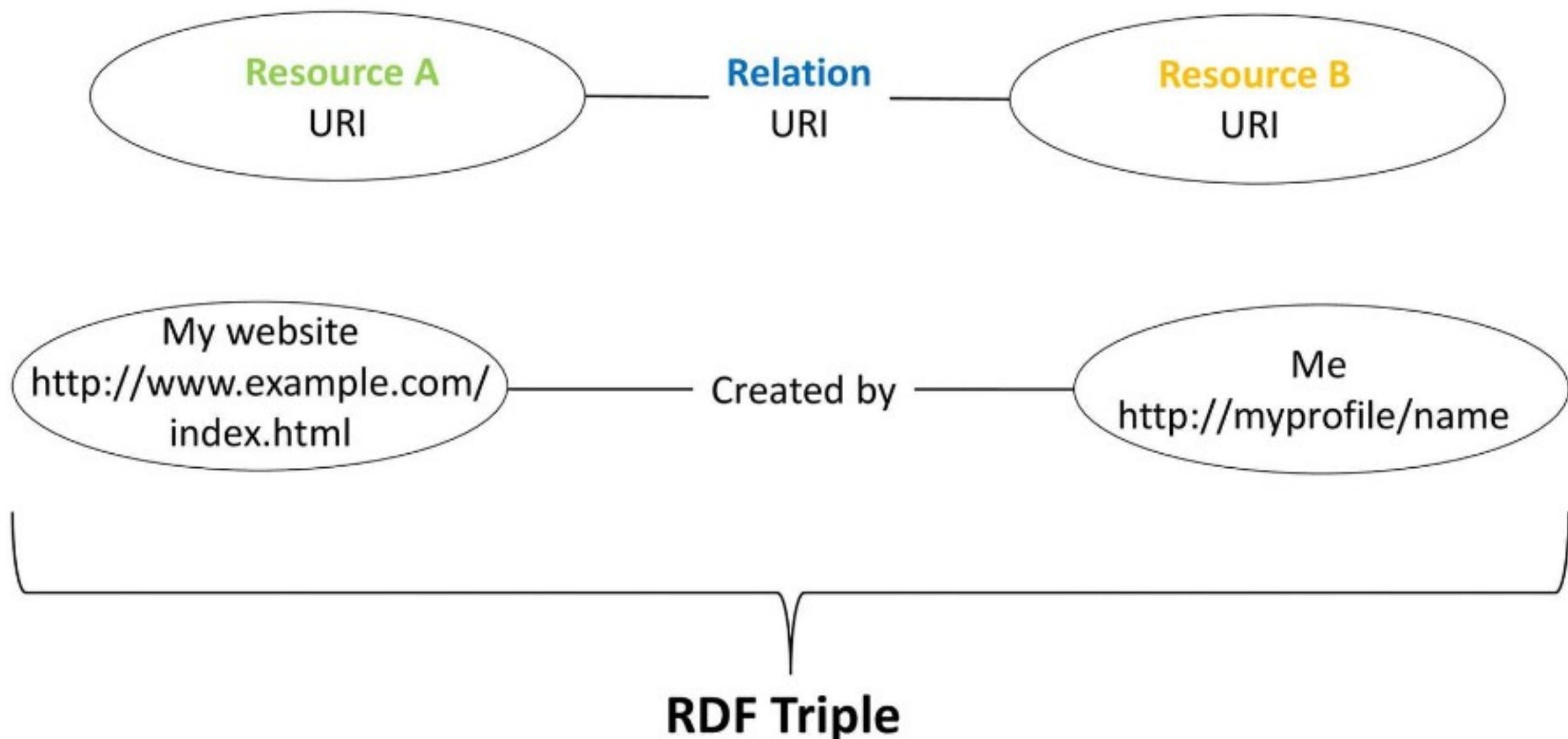
[http://www.esciencedatafactory.com/people/yann\\_le\\_fran.rdf](http://www.esciencedatafactory.com/people/yann_le_fran.rdf)

Description HTML (for humans)

[http://www.esciencedatafactory.com/people/yann\\_le\\_fran.html](http://www.esciencedatafactory.com/people/yann_le_fran.html)

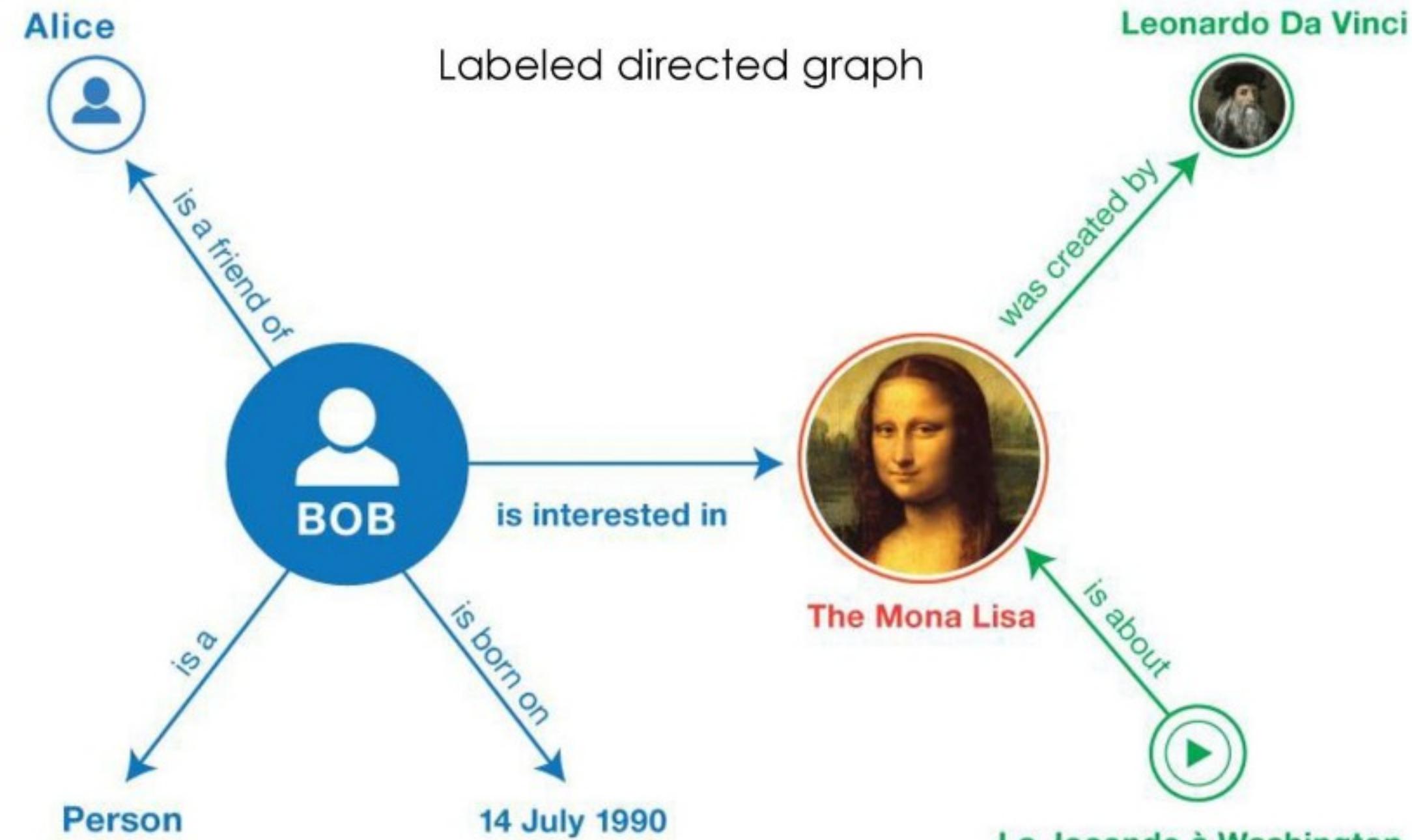
# The RDF Data Model

#CSIC



# RDF in action

#CSIC



From W3C RDF 1.1. Primer <https://www.w3.org/TR/rdf11-primer/>

# Mona Lisa is back!

#CSIC

- Vamos a definir una ontología sencilla para la Mona Lisa, teniendo en cuenta sus relaciones y tipos.
- Sujeto – Predicado – Objeto
- ¿Cómo funcionaría la Web Semántica?
  1. Definimos la Monalisa
  2. Seleccionamos un Objeto
  3. Definimos ese Objeto
  4. ¿Cómo consultaríamos información de ese objeto?

# Propiedades, relaciones - Mona Lisa

#CSIC

monalis, tipo, cuadro

La Mona Lisa - es una -  
pintura

MonaLisa pintada en el siglo  
XVI

Mona Lisa CreadaPor  
Leonardo Da Vinci

MonaLisa - se encuentra en -  
Museo del Louvre

monlisa -> representa -> mujer

Mona Lisa pintadaCon óleo

Mona Lisa está en Louvre

Mona Lisa - estilo de -  
Renacimiento

# Propiedades, relaciones - Mona Lisa

#CSIC

Mona Lisa - tamaño - 77 cm x  
53 cm

Mona Lisa tiene Coronavirus

<https://permid.org/onecalaisViewer>

“Branded the Bob Dylan of Brazil, Caetano Veloso co-founded Tropicalia, the progressive poetry, theater and music movement that helped define Latin America's psychedelic '60s. Alongside his fellow conspirator, Gilberto Gil, Veloso fused Bossa Nova, African rhythms, and acid-drenched acoustic guitar with a political consciousness that found him censored, banned, incarcerated and eventually exiled by the country's military dictatorship. The recipe was complex but simple: melodies as gorgeous as a Copacabana beach layered atop of a philosophical wit exposing his homeland's most gross imbalances.”

<http://discoveryhub.co/>

IMDB: og:...

- RDF no tiene un único formato para representar los recursos (basado en tripletas).
- Los más comunes
  - N-Triples
  - RDF/XML
  - Turtle
  - N-Quads
  - TriG

### N-Triples

```
<http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://xmlns.com/foaf/0.1/Person> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
<http://example.org/bob#me>
  <http://schema.org/birthDate> "1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/topic_interest> <http://www.wikidata.org/entity/Q12418> .
<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/title> "Mona Lisa" .
<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/creator>
  <http://dbpedia.org/resource/Leonardo_da_Vinci> .
<http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619>
  <http://purl.org/dc/terms/subject> <http://www.wikidata.org/entity/Q12418> .
```

### RDF/XML

```
<?xml version = "1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:foaf="http://xmlns.com/foaf/0.1">

    <rdf:Description rdf:about="http://www.esciencedatafactory.com/people/yann_le_franck">
        <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person">
            <foaf:name>Yann Le Franc</foaf:name>
    </rdf:Description>
```

### Turtle: (Terse RDF Triple Language) Formato para serializar RDF

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix foaf: <http://xmlns.com/foaf/0.1>

< http://www.esciencedatafactory.com/people/yann\_le\_franck
```

**RDFa:** conjunto de extensiones de XHTML propuestas por W3C para introducir semántica en los documentos

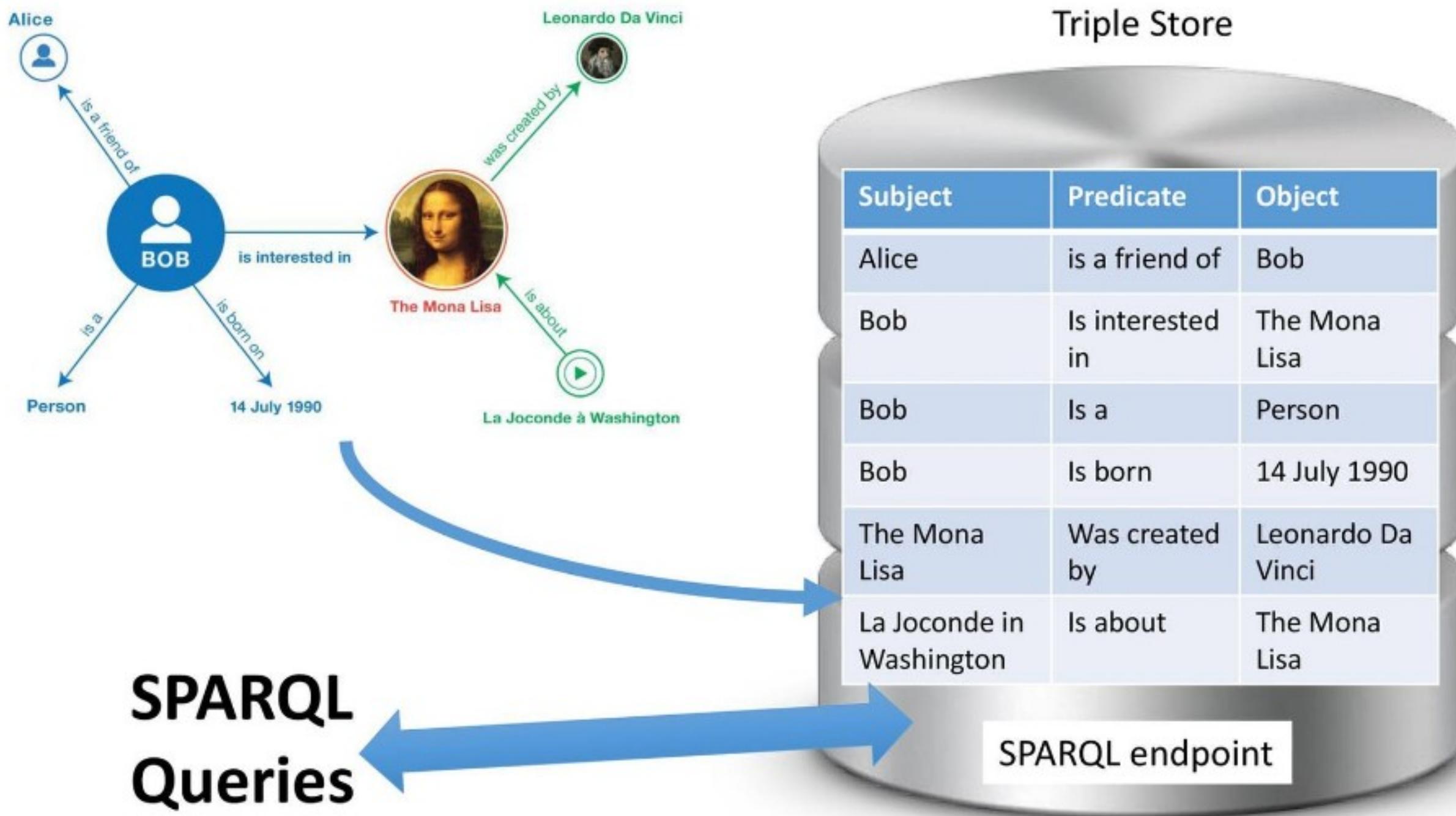
```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
  "http://www.w3c.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3c.org/1999/xhtml"
  xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

<head>
  <meta http-equiv="Content-Type" content="application/xhtml+xml; charset=UTF-8"/>
  <title>Profile page for Yann Le Franc</title>
</head>

<body>
  <div about="http://www.esciencefactory.com/people/yann_le_fran" typeof="foaf:Person">
    <span property="foaf:name">Yann Le Franc</span>
  </div>
</body>
</html>
```

# Publicación RDF

#CSIC



#CSIC

## RDF Triple store



Apache Jena



## Frameworks



<https://github.com/vivo-project/Vitro>

## Graph database



M. Junghanns and A. Petermann, "Management and Analysis of Big Graph Data: Current Systems and Open Challenges," ... (eds: S Sakr, 2017).

B. Haslhofer, E. Momeni Roochi, B. Schandler, and S. Zander, "Europeana RDF Store Report," Mar. 2011.

Z. Kaoudi and G. Weikum, *RDF in the clouds: a survey* In *The VLDB Journal*. 2014.

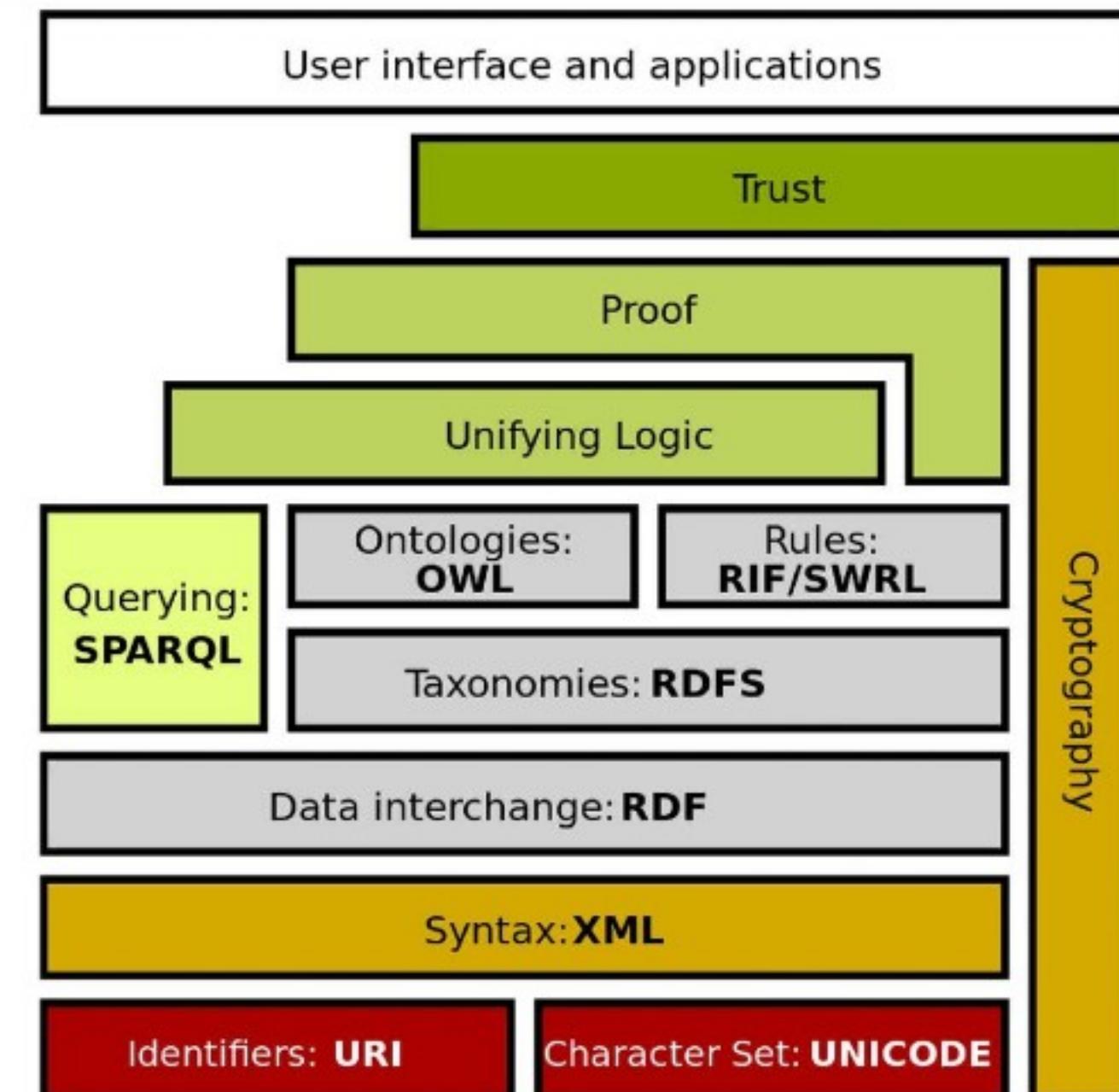
# Para lanzarnos a desarrollar/utilizar web semántica

#CSIC

- Crear o entender “controlled vocabularies” u ontologías que pueden ser usados como vocabulario estándar para describir e identificar tus recursos.
- **¿Qué es una ontología?**
- En ciencia de datos y tecnologías de la información, una ontología representa formalmente un área de conocimiento como un conjunto de conceptos dentro de un dominio, utilizando un vocabulario compartido que define tipos, propiedades e interrelaciones de los conceptos.
- **¿Cómo lo aplicamos en el contexto “web”?**
- W3C semantic web standards:
  - ✓ RDF Schema
  - ✓ OWL (Web Ontology Language)
  - ✓ SKOS (Simple Knowledge Organization System)

# The semantic web stack

#CSIC



By user:Marobi1 [CC0], via Wikimedia Commons

[https://en.wikipedia.org/wiki/Semantic\\_Web\\_Stack](https://en.wikipedia.org/wiki/Semantic_Web_Stack)

- JSON-LD, o JavaScript Object Notation for Linked Data, es un método de codificación de datos enlazados usando JSON.
- Uno de sus objetivos era requerir el menor esfuerzo posible de los desarrolladores para transformar su JSON existente en JSON-LD.
- Esto permite que los datos sean serializados de una manera que es similar al JSON tradicional.

```
{  
  "@context": {  
    "name": "http://xmlns.com/foaf/0.1/name",  
    "homepage": {  
      "@id": "http://xmlns.com/foaf/0.1/workplaceHomepage",  
      "@type": "@id"  
    },  
    "Person": "http://xmlns.com/foaf/0.1/Person"  
  },  
  "@id": "http://me.example.com",  
  "@type": "Person",  
  "name": "John Smith",  
  "homepage": "http://www.example.com/"  
}
```

# LD JSON embebido en HTML

#CSIC

```
<script type="application/ld+json">
```

```
{  
    "@context": "http://schema.org",  
    "@id": "http://www.apple.com/#organization",  
    "@type": "Organization",  
    "name": "Apple",  
    "url": "http://www.apple.com/",  
    "logo": "https://www.apple.com/ac/structured-data/  
        /images/knowledge_graph_logo.png?201610281557",  
    "contactPoint": [  
        {  
            "@type": "ContactPoint",  
            "telephone": "+1-800-692-7753",  
            "contactType": "sales",  
            "areaServed": [ "US" ]  
        }  
    ],  
    "sameAs": [  
        "http://www.wikidata.org/entity/Q312",  
        "https://www.youtube.com/user/Apple",  
        "https://www.linkedin.com/company/apple"  
    ]  
}
```

```
</script>
```

- Google Knowledge Graph
  - <https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html>
  - <https://developers.google.com/knowledge-graph/>
- Facebook graph:
  - <https://developers.facebook.com/docs/graph-api/overview/>
- Wikidata: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)
- Freebase
- Dbpedia
  - <https://wiki.dbpedia.org/>
- EBI RDF store - Biology
  - <https://www.ebi.ac.uk/rdf/services/sparql>

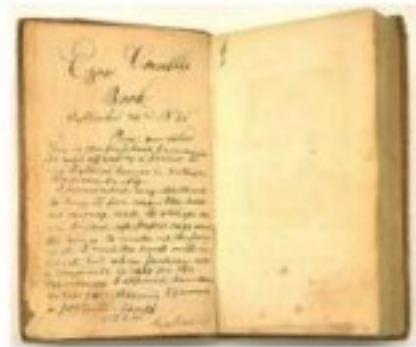
# Ejemplo: ¿Cómo funciona?



[http://dublincore.org/resources/training/dc-2011/Tutorial\\_Hermans.pdf](http://dublincore.org/resources/training/dc-2011/Tutorial_Hermans.pdf)

# Ejemplo: ¿Cómo funciona?

①



<http://www.rubenianum.be/book/B2195>

②

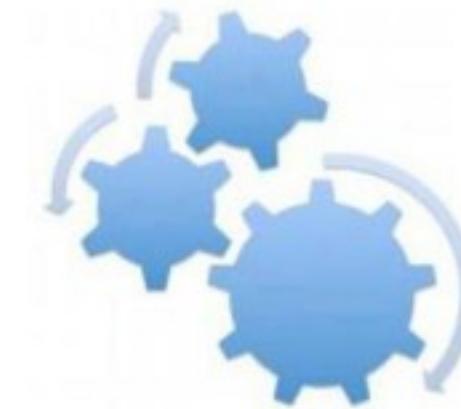


<http://www.rubenianum.be/book/B2195.html>

③

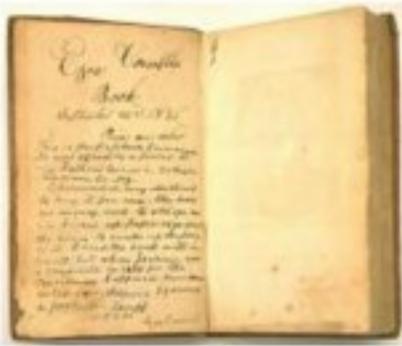


<http://www.rubenianum.be/book/B2195.rdf>



# Ejemplo: ¿Cómo funciona?

①



<http://id.rubenianum.be/book/B2195>

②



<http://page.rubenianum.be/book/B2195>

③

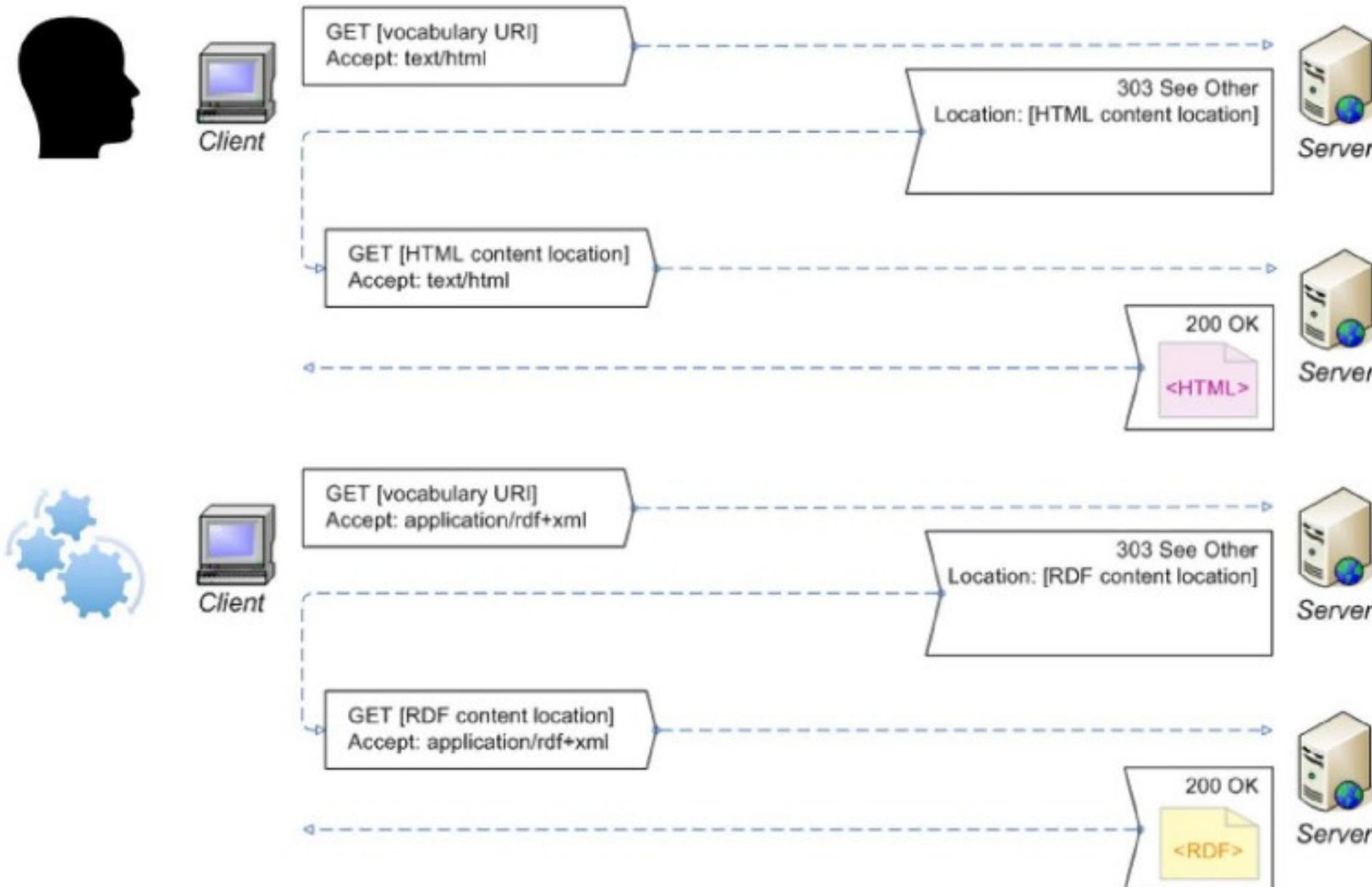


<http://data.rubenianum.be/book/B2195>



# Ejemplo: ¿Cómo funciona?

#CSIC



# Ejemplo: Dublin Core

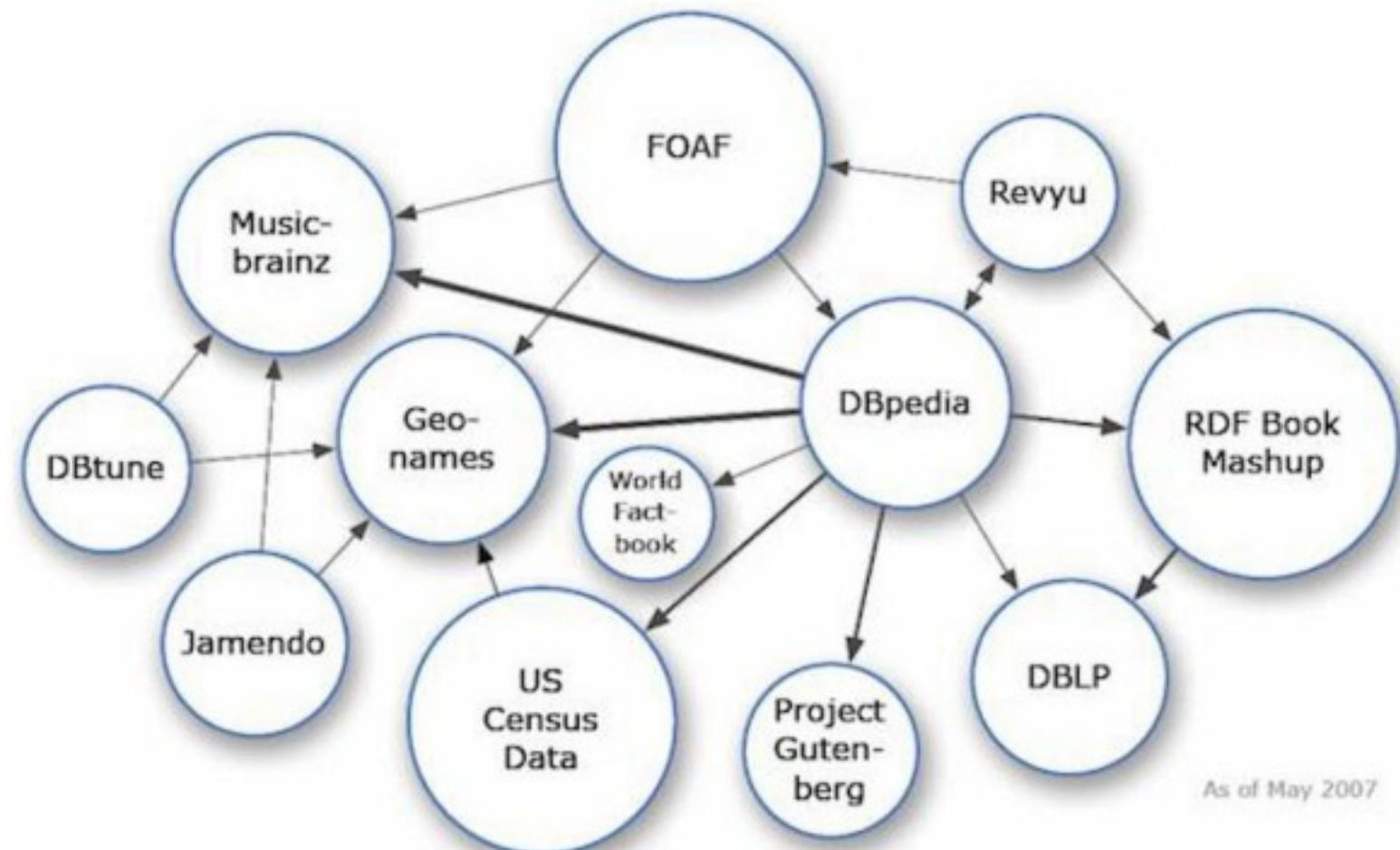
### Standaarden Overheid.nl

Home	OWMS
U bent hier: Standaarden > OWMS > (	- <rdf:RDF xml:base="http://standaarden.overheid.nl/owms/terms/Schoorl_(gemeente).rdf">
OWMS	- <rdf:Description rdf:about="http://standaarden.overheid.nl/owms/terms/Schoorl_(gemeente)">
OWMS 4.0	<rdf:type rdf:resource="http://standaarden.overheid.nl/owms/terms/Agent"/>
OWMS 3.5	<rdfs:label>Schoorl</rdfs:label>
Beheer	<dcterms:date rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2000-12-31</dcterms:date>
FAQ	<overheid:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2000-12-31</overheid:endDate>

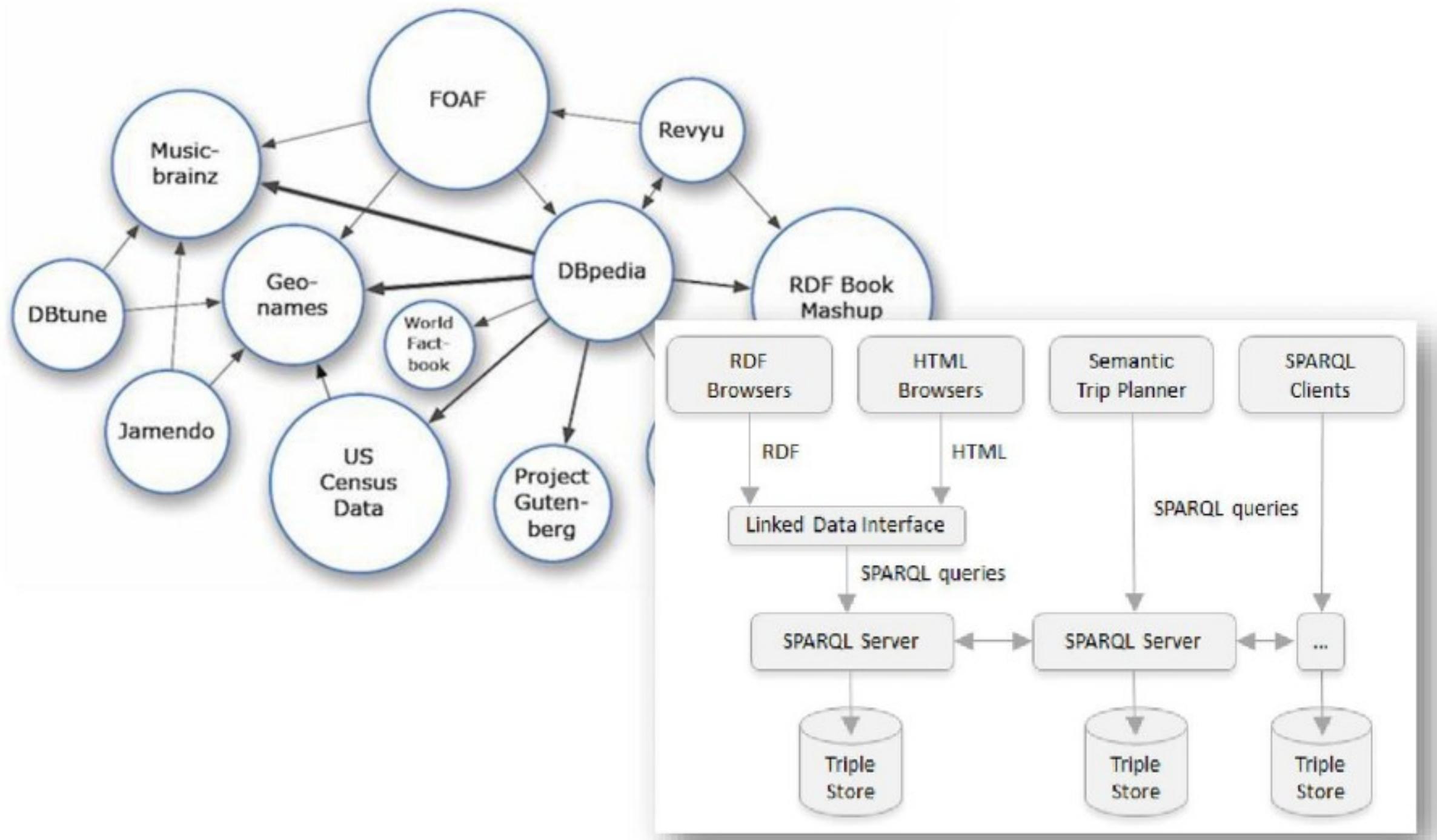
```
- <rdf:type rdf:resource="http://standaarden.overheid.nl/owms/terms/Gemeente"/>
<overheid:successor rdf:resource="http://standaarden.overheid.nl/owms/terms/Bergen_(NH)"/>
<skos:prefLabel>Schoorl</skos:prefLabel>
<dcterms:isPartOf rdf:resource="http://standaarden.overheid.nl/owms/terms/OWMSdataset"/>
<rdf:type rdf:resource="http://purl.org/dc/terms/Agent"/>
<rdfs:isDefinedBy rdf:resource="http://standaarden.overheid.nl/owms/terms/Schoorl_(gemeente).n3"/>
<rdf:type rdf:resource="http://standaarden.overheid.nl/owms/terms/Organisatie"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
<rdf:type rdf:resource="http://standaarden.overheid.nl/owms/terms/Overheidsorganisatie"/>
<rdfs:isDefinedBy rdf:resource=""/>
<foaf:page rdf:resource="http://standaarden.overheid.nl/owms/terms/Schoorl_(gemeente).html"/>
</rdf:Description>
- <rdf:Description rdf:about="">
<dcterms:rights rdf:resource="http://en.wikipedia.org/wiki/WP:GFDL"/>
<dcterms:publisher rdf:resource="http://standaarden.overheid.nl/owms/terms/ICTU"/>
<dcterms:available rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-07-28</dcterms:available>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Ontology"/>
</rdf:Description>
</rdf:RDF>
<!-- Created with TopBraid -->
```

# Interconexiones

#CSIC



# Interconexiones



# Ejemplo: Extracción de Información Semántica

- Podemos encontrar gran cantidad de información estructurada (Incluso sin adoptar un formato semántico).
- Ejemplo: Wikipedia
- [https://es.wikipedia.org/wiki/Santander\\_\(Espa%C3%B1a\)](https://es.wikipedia.org/wiki/Santander_(Espa%C3%B1a))
- Tipo, tabla de información.

The screenshot shows a table from the Wikipedia page for Santander (Spain). The table contains various facts about the city, such as its location, area, population, and history. An orange arrow points from the table to a snippet of the underlying HTML code, illustrating how structured data can be extracted from a web page.

País	España
• Com. autónoma	Cantabria
• Provincia	Cantabria
• Comarca	Santander
• Partido judicial	Santander
• Capital	Santander
Ubicación	43°28'00"N 3°48'00"O
• Altitud	61 msnm (mín.: 0, máx.: 139 <sup>2</sup> )
Superficie	34,76 km <sup>2</sup>
Fundación	26 a. C., como Portus Victoriae Iulioorigenium 11 de julio de 1187, con la concesión del Fuero de Santander 9 de enero de 1755, concesión del título de ciudad
Población	172 044 hab. (2018)
• Densidad	4946,81 hab./km <sup>2</sup>
Gentilicio	santanderino/a
Código postal	39001-39012
Alcalde (2016-)	Gema Igual
Patrona	Nuestra Señora del Mar
Sitio web	Ayto. de Santander

```
<tr>
<th scope="row" style="text-align:left;padding:1px 12px; ;"> ==>
<a href="/wiki/%C3%81rea" title="Área">Superficie</a>
</th>
<td colspan="2" style="padding:1px 12px 1px 1px;">
<br>
34,76 <br>
<a href="/wiki/Kil%C3%B3metro_cuadrado" title="Kilómetro cuadrado">km2</a>
</td>
</tr>
```

- Schema.org
- APIs semánticas redes sociales
  - <https://developer.twitter.com/en/docs/tweets/optimize-with-cards/guides/getting-started>
- Extracción semántica (ejemplo: IMDB)
  - <https://www.w3.org/2012/pyRdfa/Overview.html>

#CSIC

# SPARQL

Fernando Aguilar

- Acrónimo recursivo: SPARQL Protocol and RDF Query Language
- Lenguaje estandarizado para la consulta de grafos RDF.
- Normalizado por el RDF Data Access Working Group (DAWG) del World Wide Web Consortium (W3C).
- Como SQL, lenguaje de consulta. Necesario un motor de búsqueda y almacenamiento de datos.
- SQL (Lenguaje) – MySQL (Motor)
- SPARQL (Lenguaje) – Virtuoso (Motor)
- <http://vos.openlinksw.com/owiki/wiki/VOS/VOSSparqlProtocol>
- Los resultados pueden ser conjuntos (tablas) o grafos RDF.

# Consejos para las consultas

#CSIC

- Entender qué es una ontología
- Conocer las relaciones existentes
- Conocer los esquemas utilizados
- Distinguir entre los datos en sí y las clases
- Buscar Clases e instancias para conocer las relaciones de un objeto.

**PREFIX** – Espacios de nombres

**SELECT** – Atributos a consultar

**WHERE** – Condiciones a cumplir

# Listado Películas – Intento 1

1. Buscar clases (Película, Movie, Film...) :  
<http://mappings.dbpedia.org/server/ontology/classes/>
2. Ver sus propiedades (y las de sus super clases, que puede heredar). Un ejemplo puede ayudar también (12\_Monkeys)
3. Seleccionar las propiedades que nos interesan
4. Entender el grafo de relaciones entre unas propiedades y otras, para obtener la información.
5. Ejemplo: Título de películas
6. Elaborar la consulta en el endpoint SPARQL: <https://dbpedia.org/sparql>

```
SELECT DISTINCT ?title
WHERE {
?peli <http://dbpedia.org/ontology/Film> ?title
} LIMIT 100
```

[http://dbpedia.org/page/12\\_Monkeys](http://dbpedia.org/page/12_Monkeys)

# Listado Películas – Intento 2

#CSIC

1. ¿Por qué no ha funcionado?
2. Faltan relaciones...

```
SELECT DISTINCT ?title
WHERE{
    ?peli ?es <http://dbpedia.org/ontology/Film> .
    ?peli <http://xmlns.com/foaf/0.1/name> ?title
} LIMIT 100
```



El punto para separar

# Listado Películas – Intento 3

1. Los espacios de nombres pueden ser engorrosos para consultas complejas.
2. Se pueden declarar en PREFIX
3. El resultado es el mismo

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?title
WHERE{
    ?peli ?es dbp:Film .
    ?peli foaf:name ?title
} LIMIT 100
```

# Fecha Nacimiento directores

- FILTER – Cumplir un criterio dentro del WHERE

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?director ?bd
WHERE{
    ?peli ?es dbp:Film .
    ?peli dbp:director ?director .
    ?director dbp:birthDate ?bd .
    FILTER (datatype (?bd) IN (<http://www.w3.org/2001/XMLSchema#date>))
} LIMIT 100
```

# Directores nacidos en ciudades pequeñas

#CSIC

- FILTER – Ciudad < 100000 habitantes

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?director ?bp
WHERE{
    ?peli ?es dbp:Film .
    ?peli dbp:director ?director .
    ?director dbp:birthPlace ?bp .
    ?bp dbp:populationTotal ?population .
    FILTER (?population < 10000)
} LIMIT 100
```

¿Quieres ver el número de habitantes?  
Problemas con falta de información.  
¿Más de un filtro? &&, ||

- Concordancia literal de texto (igual con números)

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?artista ?city
WHERE{
    ?a dbp:artist ?artista .
    ?artista dbp:birthPlace ?bp .
    ?bp foaf:name ?city .
    FILTER (?city = "London"@en)
} LIMIT 100
```

# Compositores nacidos en Londres

#CSIC

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?artist ?city
WHERE{
    ?a dbp:artist ?artist .
    ?artist dbp:birthPlace ?bp .
    ?obra dbp:composer ?artist .
    ?bp foaf:name ?city .
    FILTER (?city = "London"@en)
} LIMIT 100
```

# Artistas cuyo nombre empieza por “Ma”

#CSIC

- Restricciones sobre cadenas de texto

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?artist ?name
WHERE{
    ?a dbp:artist ?artist .
    ?artist dbp:birthName ?name .
    FILTER regex(?name, "^\u00c1Ma", "1")
} LIMIT 100
```

# Otros operadores de SPARQL

- UNION – Combina grafos, por ejemplo, especificados de forma diferente (ej. títulos de películas)
- ORDER BY – Ordenar
- DISTINCT – Ya utilizado. Evitar repetidos.
- Operadores con fecha – Igual que en SQL

```
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?artist ?city ?bd
WHERE{
    ?a dbp:artist ?artist .
    ?artist dbp:birthPlace ?bp .
    ?artist dbp:birthDate ?bd .
    ?algo dbp:composer ?artist .
    ?bp foaf:name ?city .
    FILTER (?city = "London"@en && ?bd>= "1900-8-
15"^^<http://www.w3.org/2001/XMLSchema#date> &&
?bd<= "1979-8-
15"^^<http://www.w3.org/2001/XMLSchema#date>)
} LIMIT 100
```