

Bases de datos de grafos



Hay casos en que interesa más modelar las **relaciones** entre cosas que las cosas!

facebook for developers

Products Docs Tools & Support News Success Stories

Search

My Apps

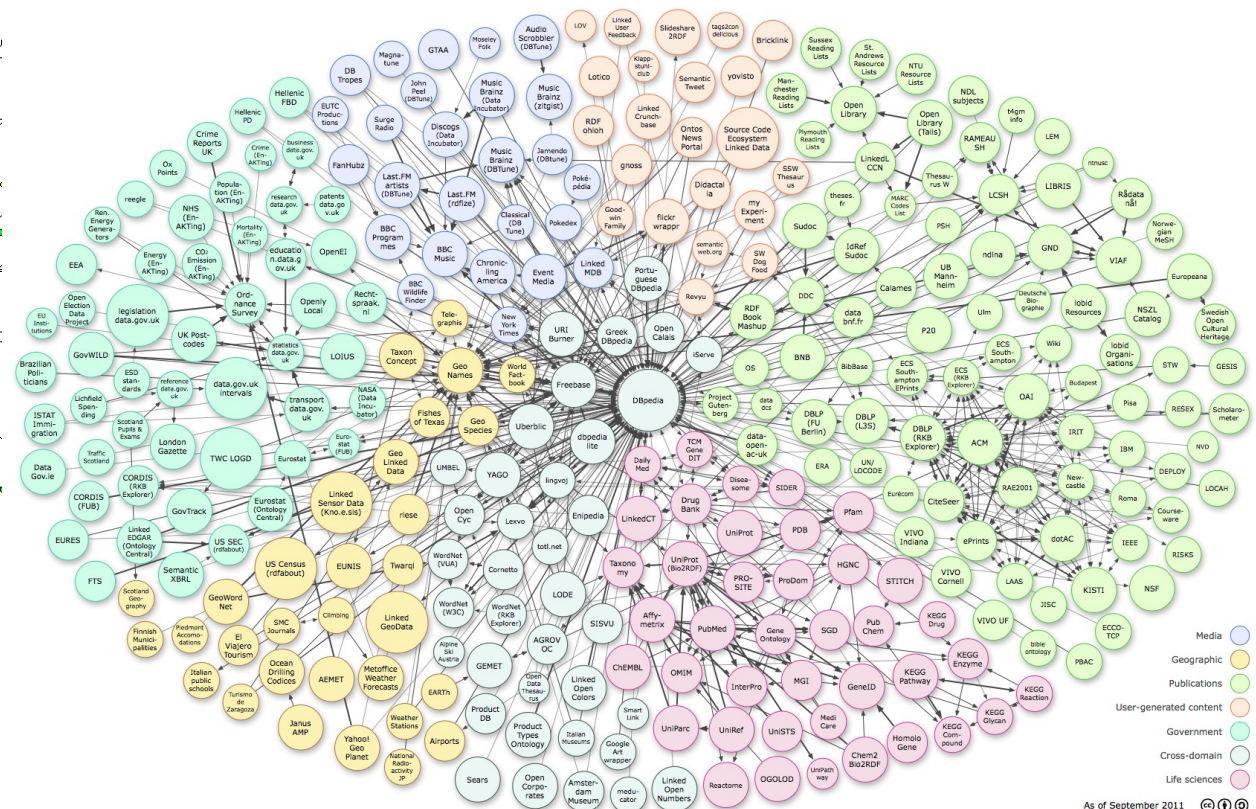
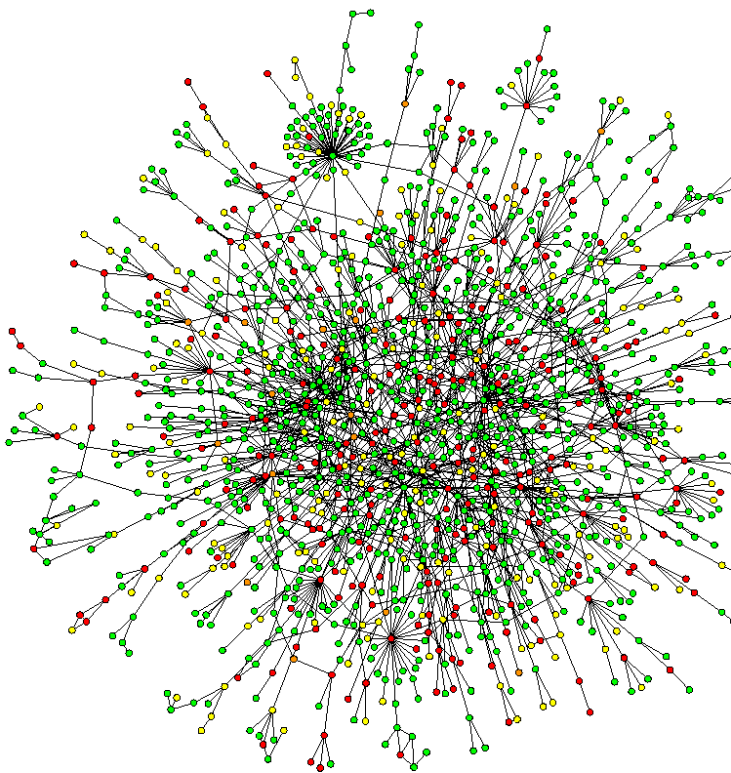
All Docs

Docs / Graph API / Overview / On This Page

Graph API Overview

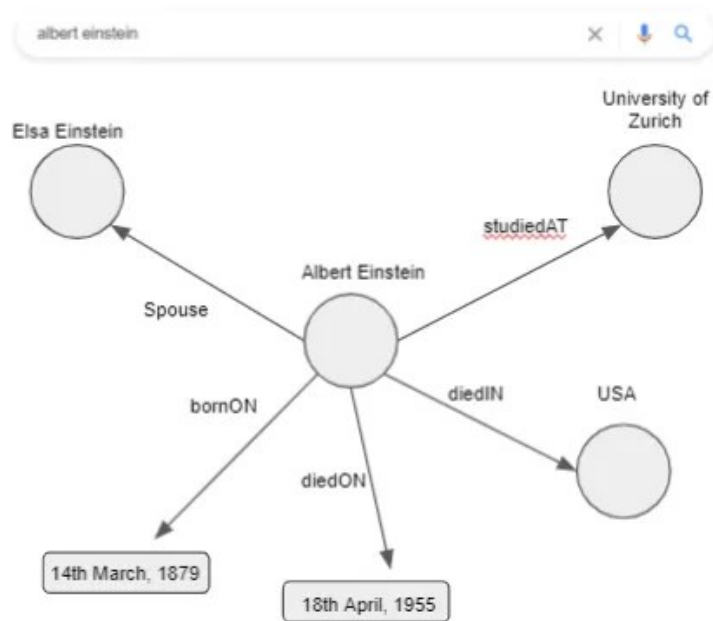
Google Inside Search

Home How Search Works Tips & Tricks Features Search Stories Playground Blog Help



Knowledge Graphs

Google Knowledge Panel



Albert Einstein

Theoretical physicist

Albert Einstein was a German-born theoretical physicist, widely acknowledged to be one of the greatest physicists of all time. Einstein is known widely for developing the theory of relativity, but he also made important contributions to the development of the theory of quantum mechanics. [Wikipedia](#)

Born: 14 March 1879, Ulm, Germany

Died: 18 April 1955, Penn Medicine Princeton Medical Center, New Jersey, United States

Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)

Education: University of Zurich (1905), ETH Zürich (1896–1900), MORE

Books

View 35+ more



Relativity :
the special
a...
1916

The World
As I see It
1934

Out of My
Later Years
1950

The
Evolution
of Physics
1938

Quotes

View 7+ more

Imagination is more important than knowledge.

If you can't explain it simply, you don't understand it well enough.

Life is like riding a bicycle. To keep your balance you must keep moving.

People also search for

View 15+ more



Eduard
Einstein

Isaac
Newton

Elsa
Einstein

Stephen
Hawking

¿en qué escenarios son adecuadas las BD de grafos?

- Las entidades están muy conectadas a través de relaciones descriptivas.
- Existen relaciones cíclicas o entidades autorreferenciadas.
 - Esto suele ser un reto cuando se utilizan bases de datos relacionales o documentales.
- Las relaciones entre entidades evolucionan dinámicamente.
 - especialmente aplicable a datos jerárquicos o estructurados en árbol con muchos niveles.
- Existen relaciones de muchos a muchos entre las entidades.
- Existen requisitos de escritura y lectura tanto en las entidades como en las relaciones.

Agenda

- Modelos en bases de datos de grafos
 - El modelo *Property graph*
 - RDF y modelos en la web semántica
 - Diseño de bases de datos de grafos
- Motores de bases de datos de grafos
 - Estrategias de almacenamiento
 - Lenguajes de consulta
 - Modelos de procesamiento distribuído

Modelos en bases de datos de grafos

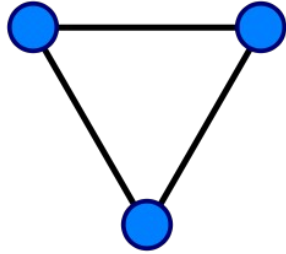
¿Qué es un grafo?

Un grafo G consiste en un conjunto de nodos o vértices V , y un conjunto de aristas E . Las aristas conectan nodos entre si.

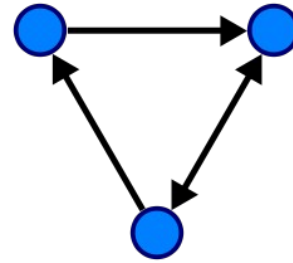
Las bases de datos de grafos implementan diferentes variantes¹.

¹ *Survey of Graph Database Models*, Angles and Gutierrez, ACM Computing Surveys, 2008

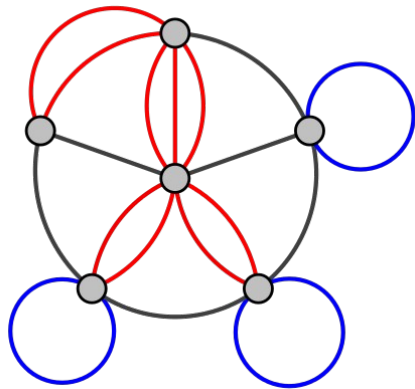
Grafos según su estructura



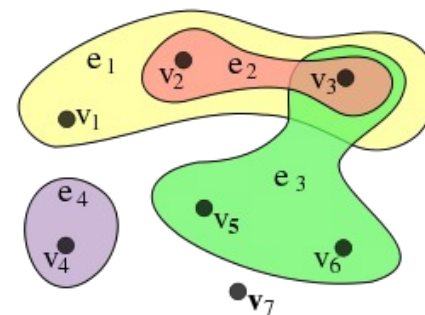
Grafo no dirigido



Grafo dirigido

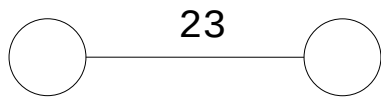


Multigrafos y pseudografos



Hipergrafos

Grafos según los datos asociados



Etiquetas en las aristas



Etiquetas en los nodos



Atributos en los nodos

¿qué tipos de grafos se utilizan en las bases de datos de grafos?

Existen muchos modelos

(ver *Survey of Graph Database Models*, Angles and Gutierrez, ACM Computing Surveys, 2008)

Los sistemas más populares actualmente implementan el modelo ***property graph*** o **RDF**

Property Graph Model (PGM)

Pseudografos dirigidos.

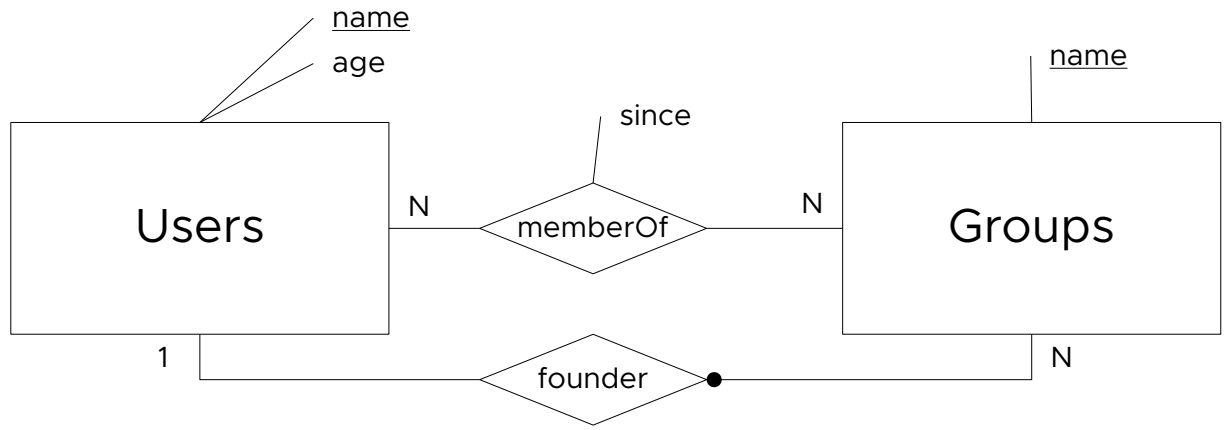
Parejas (clave,valor) llamadas *propiedades* asociadas a nodos y aristas.

Además puedo etiquetar nodos y aristas.

Cada nodo o arista puede tener más de una etiqueta

Neo4j y Titan son ejemplos de sistemas de bases de datos que soportan PGM.

CONCEPTUAL



LÓGICO

Relacional

Users

<u>name</u>	<u>age</u>
Bob	25
Alice	null

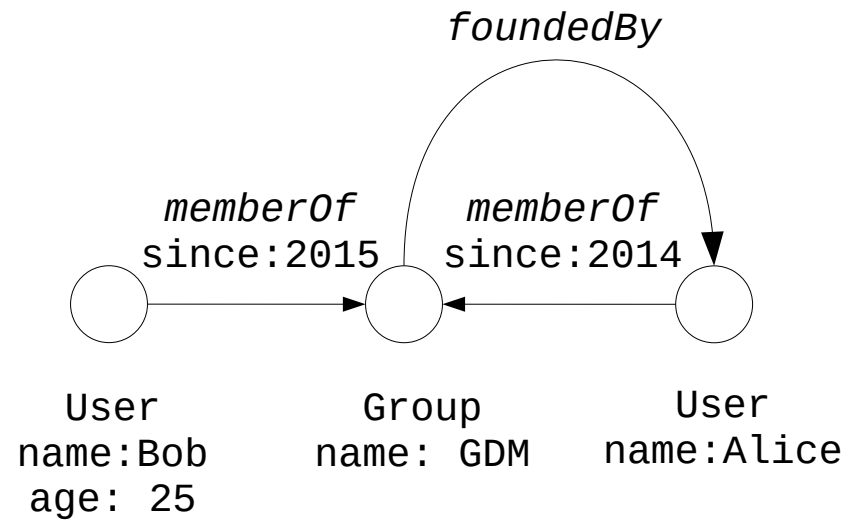
Groups

<u>name</u>	<u>founder</u>
GDM	Alice

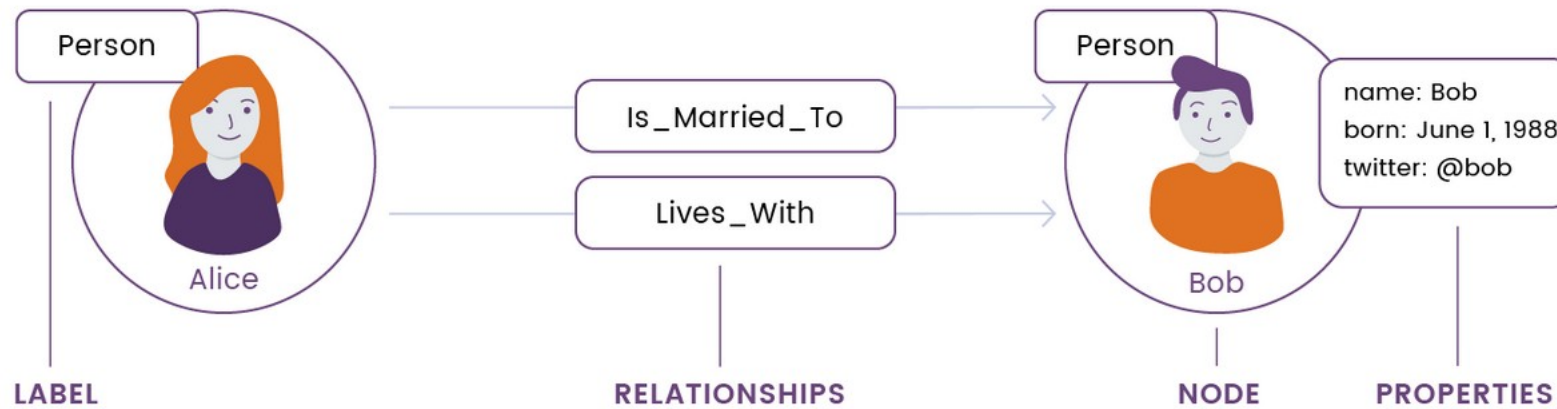
GroupMembers

<u>userName</u>	<u>groupName</u>	<u>since</u>
Bob	GDM	2015
Alice	GDM	2014

Property Graphs



Otro ejemplo



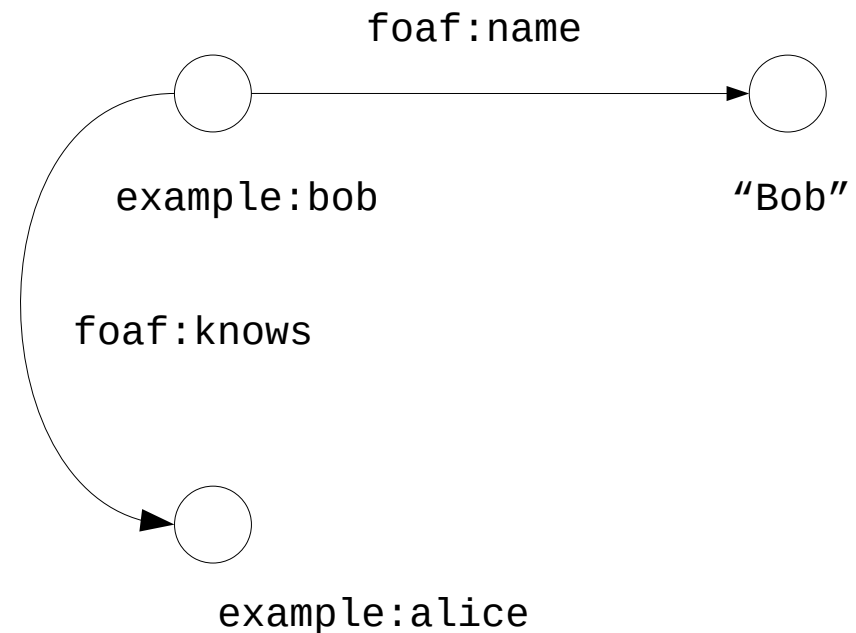
Resource Description Framework (RDF)

Pseudografos dirigidos.

Etiquetas en nodos y aristas.

Las etiquetas pueden ser:
IRIs, literales, o nodos blancos.

Cada nodo o arista tiene una sola etiqueta



Virtuoso y Stardog son ejemplos de sistemas de bases de datos que soportan RDF.

Restricciones y esquemas

- En ambos casos es posible definir nociones de esquema más o menos fuertes.
- En la familia de lenguajes y modelos vinculados a RDF puedo hacer cosas muy expresivas y potentes.
- Sobre el PGM puedo definir algunas restricciones pero dependen del vendedor
 - Ejemplo [Neo4j constraints](#)

Grafos de conocimiento y la web semántica

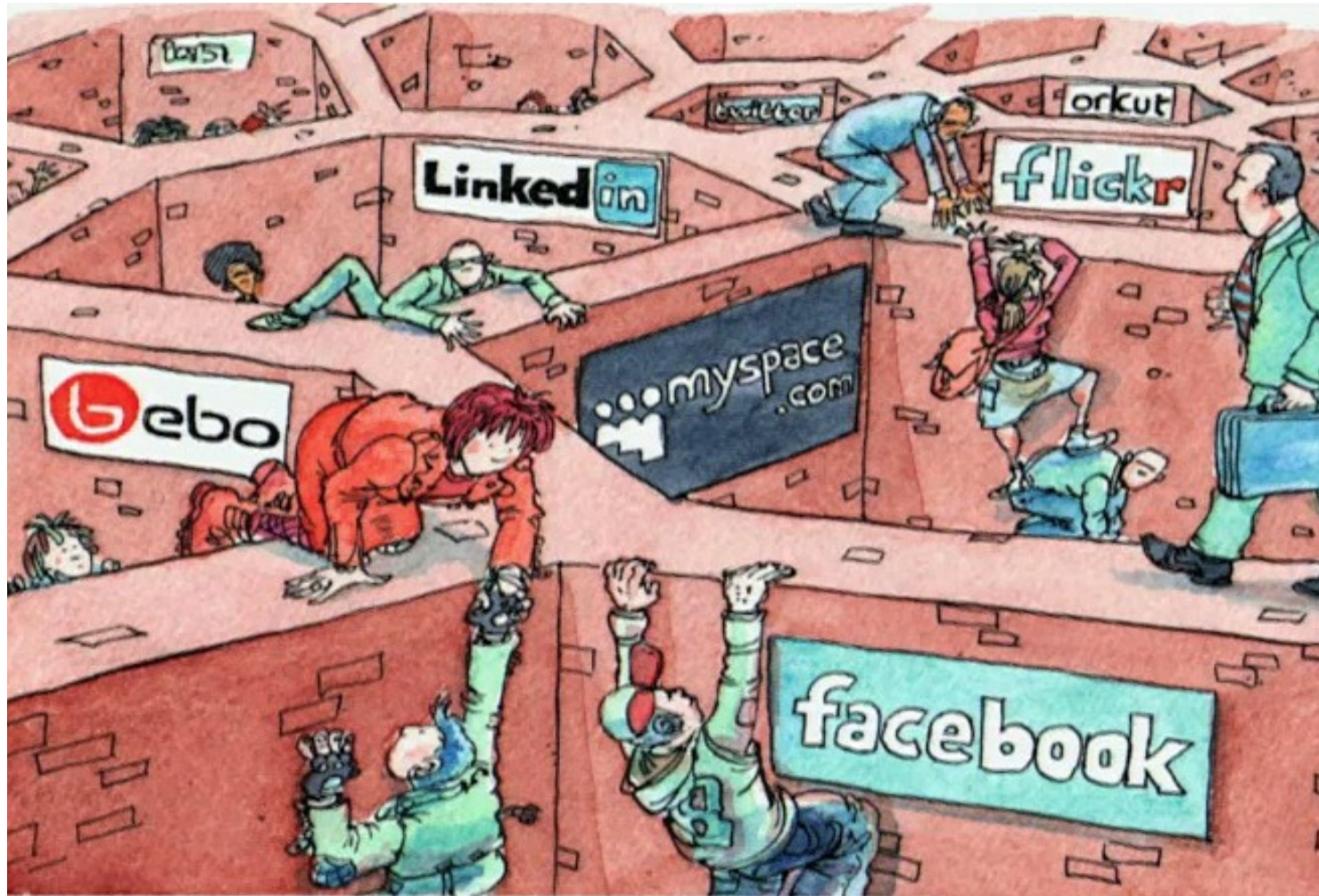
La web clásica es una red de **documentos**,

interpretable por **humanos**,

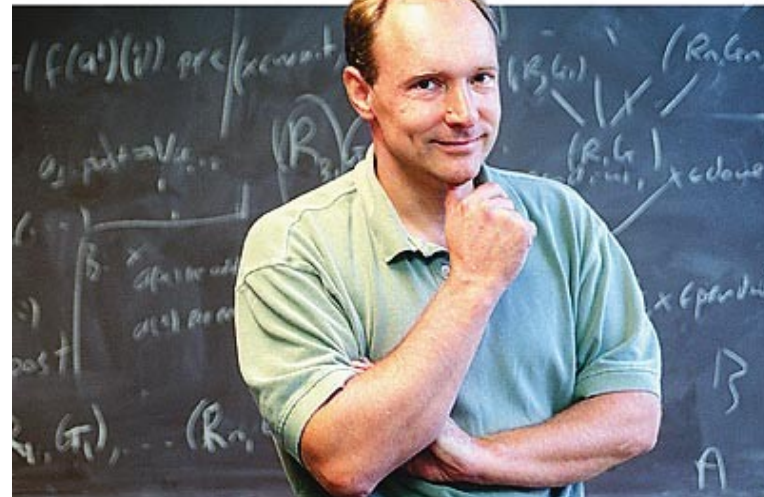
donde las relaciones entre documentos no tienen un

significado.

Además, la mayoría de los datos en la web están
aislados



La web de datos es una red de **afirmaciones**, interpretable por humanos y **máquinas**, donde las afirmaciones se relacionan y tienen un **significado**.



RDF: un **modelo de datos** basado en **triplas** que permite representar relaciones.

SPARQL: el lenguaje de **consultas** sobre RDF.

RDF-S y OWL, para representar metadatos y darles **significado** (ontologías).



<http://www.w3.org/2001/sw/wiki/RDF>



<http://www.w3.org/2001/sw/wiki/SPARQL>

afirmación o tripla

<sujeito, predicado, objeto>



Ing. Eladio Dieste

esAutor



MVD Shopping
Center

afirmación o **tripla**

<sujeito, predicado, objeto>



En RDF todo es una tripla.
Los **recursos** y las **propiedades** se
identifican por URIs

Linked Data

es un conjunto de
buenas prácticas para
publicar y *relacionar*
datos en la web,
usando
tecnologías de la
web semántica

Principios de Linked Data

- 1) Usar **URIs** para nombrar cosas
- 2) Usar **URIs HTTP** que sean consultables por humanos y máquinas
- 3) Proveer información **útil** acerca de cada URI en RDF
- 4) Crear **links** entre URIs

Tim Berners-Lee – Linked Data (2006)

<http://www.w3.org/DesignIssues/LinkedData>

#TED Talk Tim Berners-Lee on the next Web (2009)

http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html

1) Usar **URIs** para nombrar cosas

URI's *minting*: the act of establishing the association between the URI and the resource it denotes. A URI **MUST** only be minted by the URI's owner or delegate. Minting a URI from someone else's URI space is known as URI *squatting*

The URI Lifecycle in Semantic Web Architecture, David Booth,
<http://dbooth.org/2009/lifecycle/>

1) Usar **URIs** para nombrar cosas

- Evitar el uso de *namespaces* que no están bajo nuestro control

<http://www.imdb.com/title/tt0057012/#film>

- No incluir detalles de implementación

<http://tiger.biglynx.co.uk/people.php?id=dave-smith&format=rdf>

- Evitar claves “artificiales”

<http://data.nytimes.com/66209802438676211043>

2) Usar **URIs HTTP** que sean consultables por humanos y máquinas

HTTP URI: mecanismo universal de acceso a una URI

Las URIs HTTP deberían ser
desreferenciables

“HTTP **clients** can look up the URI using the HTTP protocol and retrieve a **description** of the resource that is identified by the URI”

Redireccionamiento y negociación de contenido

text/html



dbpedia.org/page/Uruguay

About: Uruguay

An Entity of Type : [populated place](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Uruguay Listen/'juərəgwaɪ/, officially the Oriental Republic of Uruguay and sometimes called the Eastern Republic of Uruguay, is a country in the southeastern part of South America. It is home to 3.3 million people, of whom 1.8 million live in the capital Montevideo and its metropolitan area. An estimated 88% of the population is of European descent. With an area of approximately 176,000 square kilometres, Uruguay is the second-smallest nation in South America by area, after Suriname.

Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	<ul style="list-style-type: none">176215.0176215.0210629304

URI que identifica el documento **HTML** que describe al concepto Uruguay

dbpedia.org/data/Uruguay.rdf

This XML file does not appear to have any style information associated with it.

```
-<rdf:RDF>
- <rdf:Description rdf:about="http://dbpedia.org/resource/Uruguay"
  <dbpedia-owl:birthPlace rdf:resource="http://dbpedia.org/resource/Montevideo">
```

URI que identifica el documento **RDF** que describe al concepto Uruguay

URI que identifica al concepto Uruguay



application/rdf+xml

HTTP GET request(text/html)
`http://dbpedia.org/resource/Uruguay`



HTTP 303 See Other response
<http://dbpedia.org/page/Uruguay>



HTTP GET request
`http://dbpedia.org/page/Uruguay`



HTTP 200 OK response
Código HTML



```
curl -i -L -H "Accept: text/html" http://dbpedia.org/resource/Uruguay
```

```
HTTP/1.1 303 See Other
```

```
Date: Wed, 02 Oct 2013 22:21:01 GMT
```

```
Content-Type: text/html; charset=UTF-8
```

```
Server: Virtuoso/07.00.3203 (Linux) i686-generic-linux-glibc212-64 VDB
```

```
Location: http://dbpedia.org/page/Uruguay
```

```
...
```

```
HTTP/1.1 200 OK
```

```
Date: Wed, 02 Oct 2013 21:59:40 GMT
```

```
Content-Type: text/html; charset=UTF-8
```

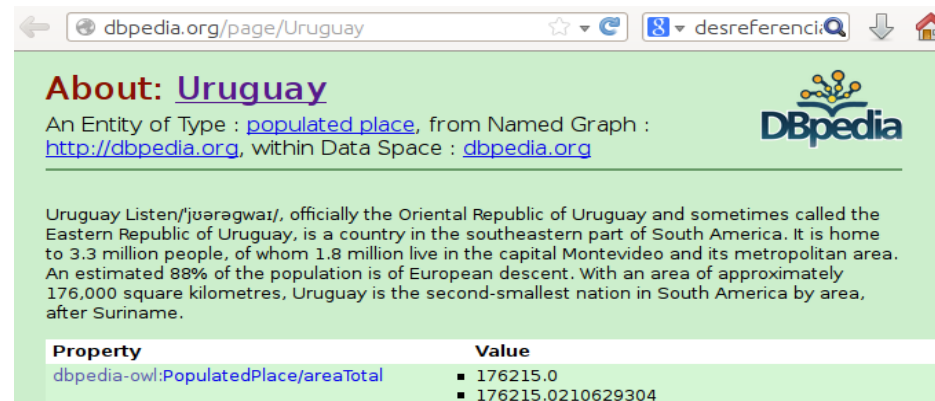
```
Content-Length: 608380
```

```
Server: Virtuoso/07.00.3203 (Linux) i686-generic-linux-glibc212-64 VDB
```

```
...
```

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" ...
```



dbpedia.org/page/Uruguay

About: Uruguay

An Entity of Type : [populated place](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)


Uruguay Listen//ˈjuərəɡwaɪ/, officially the Oriental Republic of Uruguay and sometimes called the Eastern Republic of Uruguay, is a country in the southeastern part of South America. It is home to 3.3 million people, of whom 1.8 million live in the capital Montevideo and its metropolitan area. An estimated 88% of the population is of European descent. With an area of approximately 176,000 square kilometres, Uruguay is the second-smallest nation in South America by area, after Suriname.

Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	<ul style="list-style-type: none">176215.0176215.0210629304

HTTP GET request(application/rdf+xml)
`http://dbpedia.org/resource/Uruguay`



HTTP 303 See Other response
`http://dbpedia.org/data/Uruguay.xml`



HTTP GET request
`http://dbpedia.org/data/Uruguay.xml`



HTTP 200 OK response
Datos en RDF



```
curl -i -L -H "Accept: application/rdf+xml" http://dbpedia.org/resource/Uruguay
```

HTTP/1.1 303 See Other

Date: Wed, 02 Oct 2013 22:28:35 GMT

Content-Type: application/rdf+xml; qs=0.95

Server: Virtuoso/07.00.3203 (Linux) i686-generic-linux-glibc212-64 VDB

Location: http://dbpedia.org/data/Uruguay.xml

...

HTTP/1.1 200 OK

Date: Wed, 02 Oct 2013 22:28:35 GMT

Content-Type: application/rdf+xml; charset=UTF-8

Content-Length: 603067

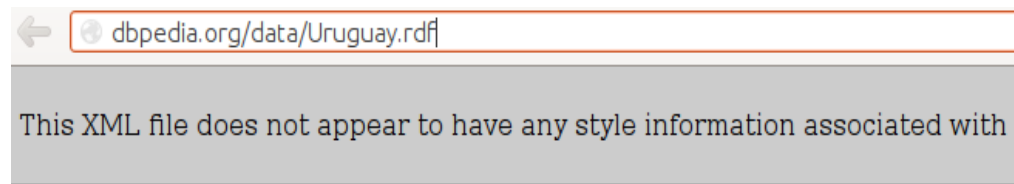
...

```
<?xml version="1.0" encoding="utf-8" ?>
```

```
<rdf:RDF
```

```
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dcterms="http://purl.org/dc/terms/"
```

...



```
-<rdf:RDF>
```

```
  -<rdf:Description rdf:about="http://dbpedia.org/resource/%C3%81lvaro
```

```
    <dbpedia-owl:birthPlace rdf:resource="http://dbpedia.org/resource,
```

```
    <dbpprop:birthPlace rdf:resource="http://dbpedia.org/resource/Uru
```

```
    <dbpprop:placeOfBirth rdf:resource="http://dbpedia.org/resource/I
```



```
curl -i -L -H "Accept: text/rdf+n3" http://dbpedia.org/resource/Uruguay
```

```
HTTP/1.1 303 See Other
```

```
Date: Wed, 02 Oct 2013 22:29:03 GMT
```

```
Content-Type: text/rdf+n3; qs=0.8
```

```
Server: Virtuoso/07.00.3203 (Linux) i686-generic-linux-glibc212-64 VDB
```

```
Location: http://dbpedia.org/data/Uruguay.nt
```

```
...
```

```
HTTP/1.1 200 OK
```

```
Date: Wed, 02 Oct 2013 22:29:04 GMT
```

```
Content-Type: text/rdf+n3; charset=UTF-8
```

```
Content-Length: 296903
```

```
...
```

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
```

```
@prefix ns1: <http://cs.dbpedia.org/resource/> .
```

```
@prefix dbpedia: <http://dbpedia.org/resource/> .
```

```
ns1:Uruguay owl:sameAs dbpedia:Uruguay .
```

```
@prefix ns3: <http://pl.dbpedia.org/resource/> .
```

```
ns3:Urugwaj owl:sameAs dbpedia:Uruguay .
```

Utilización de Hash URIs

```
curl -i http://www.w3.org/2006/07/SWD/recipes/examples-20080421/  
example1#ClassA
```

```
HTTP/1.1 200 OK
```

```
Date: Thu, 03 Oct 2013 13:08:38 GMT
```

```
Server: Apache/2
```

```
Content-Type: application/rdf+xml;
```

```
...
```

```
<?xml version="1.0"?>
```

```
<rdf:RDF
```

```
xml:base="http://www.w3.org/2006/07/SWD/recipes/examples20080421/example1">
```

```
<owl:Ontology
```

```
rdf:about="http://www.w3.org/2006/07/SWD/recipes/examples20080421/example1">
```

```
  <rdfs:comment>An example vocab for minimal hash setup.</rdfs:comment>
```

```
  <rdfs:label>Example 1 Ontology</rdfs:label>
```

```
</owl:Ontology>
```

```
<owl:Class rdf:about="http://www.w3.org/2006/07/SWD/recipes/examples-  
20080421/example1#ClassA">
```

```
</owl:Class> ...
```

```
</rdf:RDF>
```

3) Proveer información **útil** acerca de cada URI en RDF

```
dbpedia:Eladio_Dieste dbpedia-owl:birthPlace dbpedia:Uruguay ;  
  dbpprop:birthPlacedbpedia:Uruguay ;  
  dbpprop:deathPlacedbpedia:Uruguay ;  
  dbpedia-owl:deathPlace dbpedia:Uruguay .
```

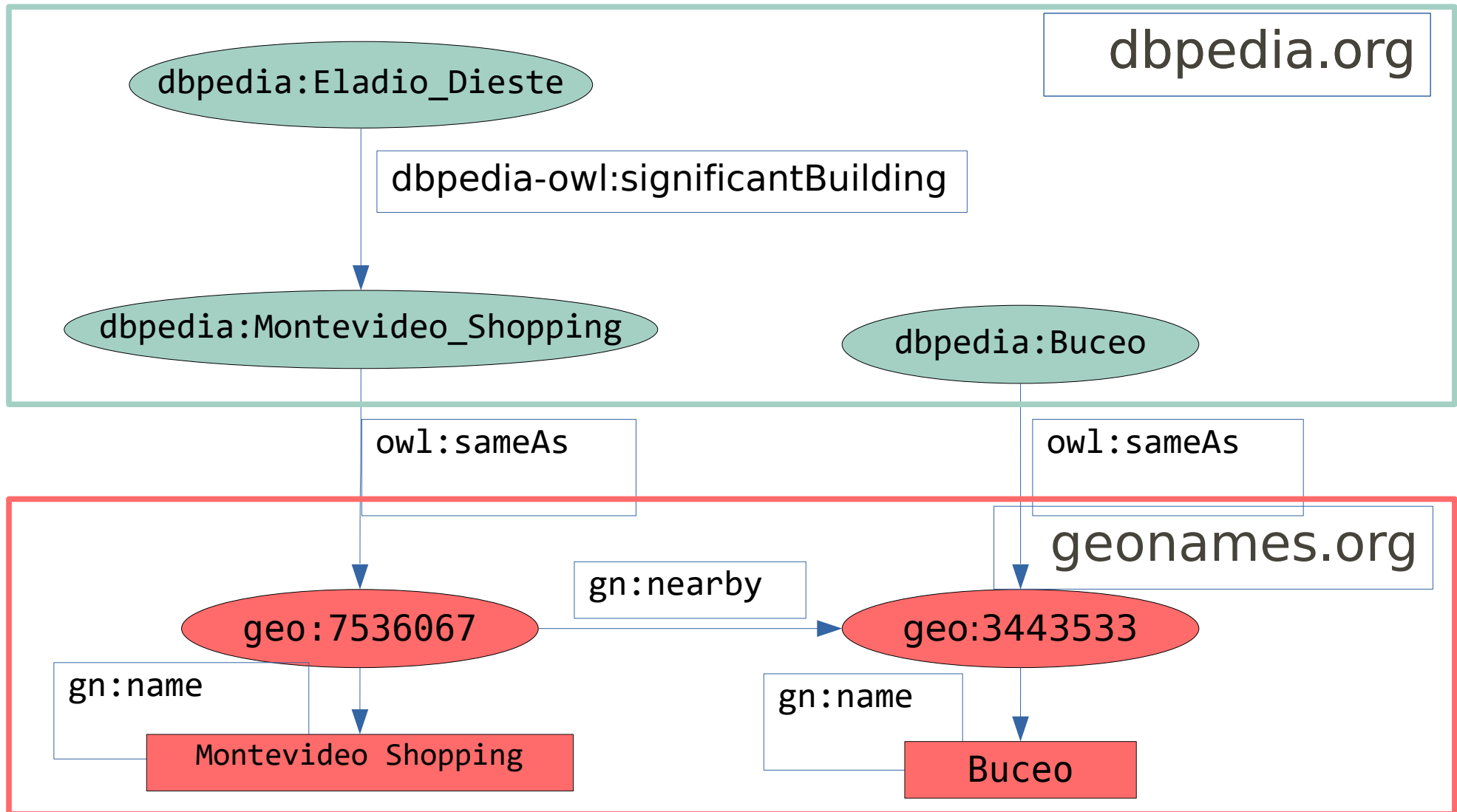
4) Crear **links** entre URIs

<http://dbpedia.org/resource/Uruguay>

```
@prefix ns192: <http://www.wikidata.org/entity/> .  
@prefix ns17: <http://nl.dbpedia.org/resource/> .  
@prefix ns18: <http://eu.dbpedia.org/resource/> .
```

```
dbpedia:Uruguay owl:sameAs ns192:Q77 ,  
    <http://gadm.geovocab.org/id/0_233> ,  
    <http://rdf.freebase.com/ns/m.07twz> ,  
    ns18:Uruguai ,  
    ns17:Uruguay .
```

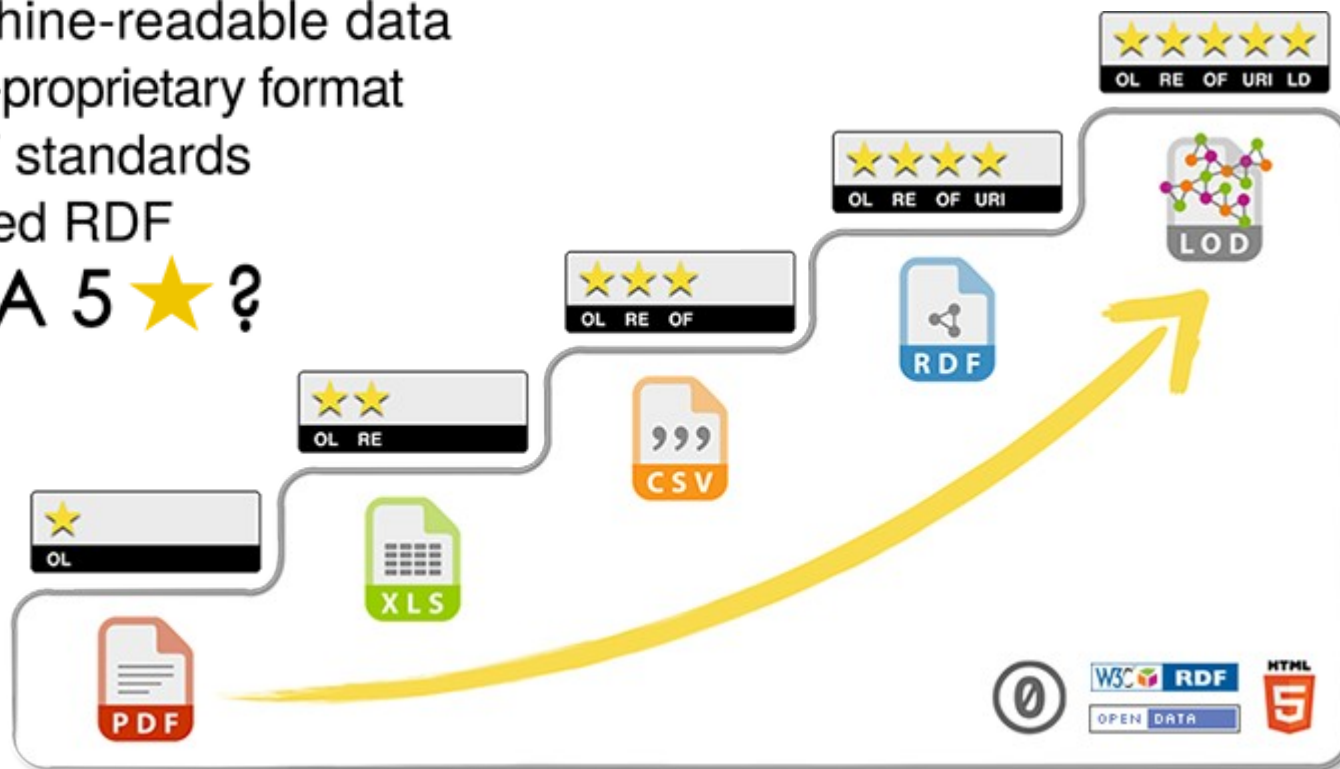
@prefix gn: http://www.geonames.org/ontology
@prefix dbpedia: http://dbpedia.org/resource/
@prefix dbpedia-owl: http://dbpedia.org/ontology/
@prefix geo: http://sws.geonames.org/
@prefix owl: http://www.w3.org/2002/07/owl#



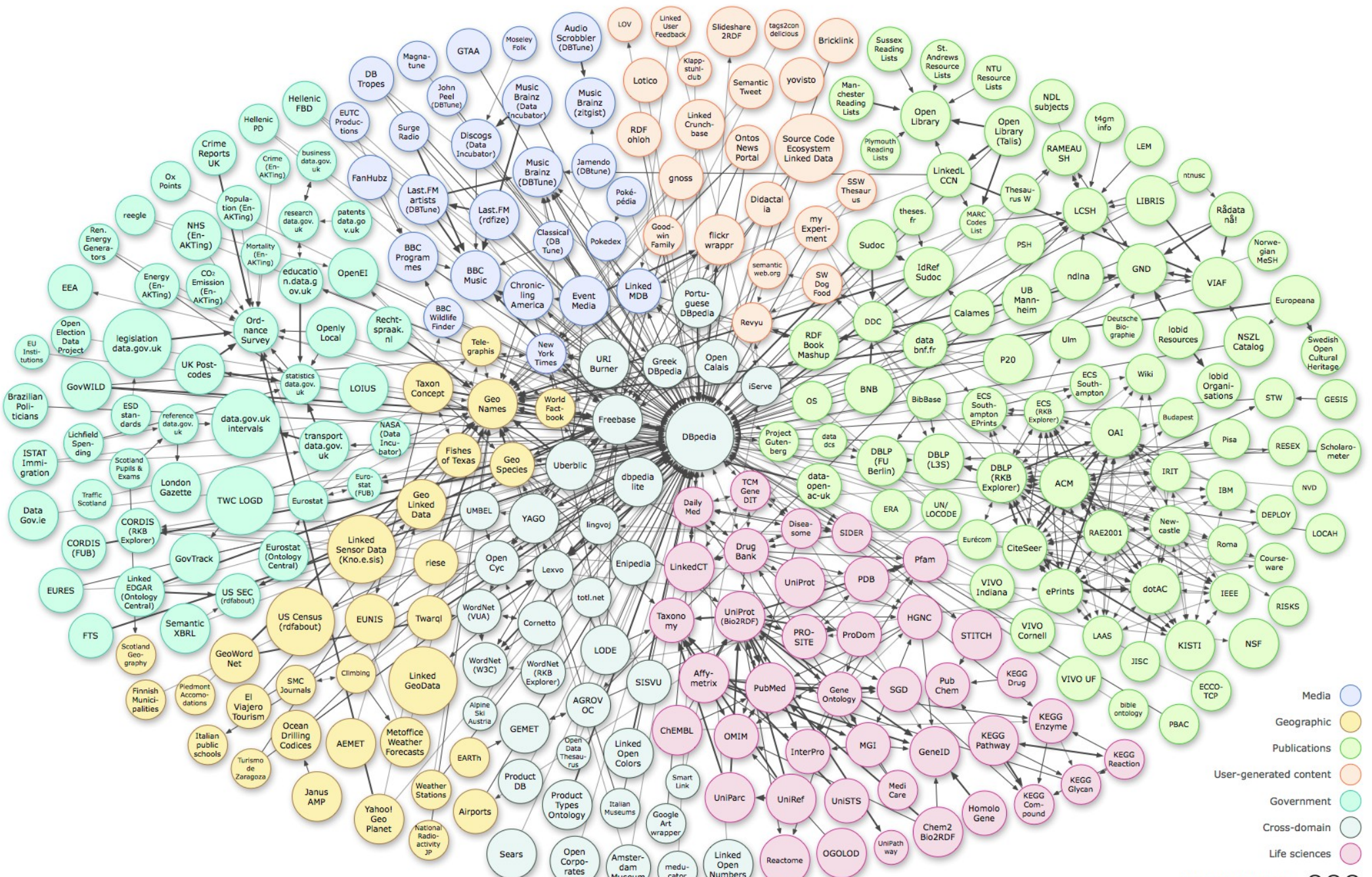
Linked Data + Open Data = LOD

LINKED DATA

- ★ On the web, open license
 - ★ ★ Machine-readable data
 - ★ ★ ★ Non-proprietary format
 - ★ ★ ★ ★ RDF standards
 - ★ ★ ★ ★ ★ Linked RDF
- IS YOUR DATA 5 ★ ?



<http://5stardata.info/en/>



- Media ●
- Geographic ●
- Publications ●
- User-generated content ●
- Government ●
- Cross-domain ●
- Life sciences ●

As of September 2011

<http://lod-cloud.net/> 2011