

CC6202-1

LA WEB DE DATOS

PRIMAVERA 2016

Lecture 9: Linked Data

Aidan Hogan

aidhog@gmail.com

PREVIOUSLY ...



First SPARQL (1.0)
Then SPARQL 1.1

(1) Data, (2) Rules/Ontologies, (3) Query

INPUT: “ (x, partOf, y) ”

DATA:

<http://ex.org/Ireland>

Ireland



(Ireland, partOf, Europe)
(Ireland, a, Country)
(Ireland, capital, Dublin)

<http://ex.org/Dublin>

Dublin



(Dublin, population, 1000000)

RULES: $(a, \text{capital}, b) \rightarrow (b, \text{partOf}, a)$
 $(c, \text{partOf}, d), (d, \text{partOf}, e) \rightarrow (c, \text{partOf}, e)$

OUTPUT: $\{(x \mapsto \text{Ireland}, y \mapsto \text{Europe}), (x \mapsto \text{Dublin}, y \mapsto \text{Ireland})$
 $(x \mapsto \text{Dublin}, y \mapsto \text{Europe})\}$

But we have not spoken much about ...

CC6202-1

LA WEB DE DATOS

... how do we use RDF(S)/OWL/SPARQL
to build a “Web of Data”?

We've gotten from here ...

The current Web is document-centric

`http://ex.org/Ireland`



`http://ex.org/Dublin`



```
<html>
  <body>
    <h1>Ireland</h1>
    <div class="flag"></div>
    <p>Ireland is a European country.</p>
    <p>Its capital is <a href="http://ex.org/Dublin">Dublin</a>.</p>
    ...
  </body>
</html>
```

To here ...

(1) Data, (2) Rules/Ontologies, (3) Query

INPUT: (x, partOf, y)

DATA:

<http://ex.org/Ireland>



<http://ex.org/Dublin>



RULES: $(a, \text{capital}, b) \rightarrow (b, \text{partOf}, a)$
 $(c, \text{partOf}, d), (d, \text{partOf}, e) \rightarrow (c, \text{partOf}, e)$

OUTPUT: $\{(x \mapsto \text{Ireland}, y \mapsto \text{Europe}), (x \mapsto \text{Dublin}, y \mapsto \text{Ireland})$
 $(x \mapsto \text{Dublin}, y \mapsto \text{Europe})\}$

But what are we missing from here before we can have a “Web of Data”?

We need links!

(1) Data, (2) Rules/Ontologies, (3) Query

INPUT: (x, partOf, y)

DATA:

<http://ex.org/Ireland>



<http://ex.org/Dublin>



RULES: $(a, \text{capital}, b) \rightarrow (b, \text{partOf}, a)$
 $(c, \text{partOf}, d), (d, \text{partOf}, e) \rightarrow (c, \text{partOf}, e)$

OUTPUT: $\{(x \mapsto \text{Ireland}, y \mapsto \text{Europe}), (x \mapsto \text{Dublin}, y \mapsto \text{Ireland})$
 $(x \mapsto \text{Dublin}, y \mapsto \text{Europe})\}$

... can't have a Web without links!

We need links!

(1) Data, (2) Rules/Ontologies, (3) Query

INPUT: (x, partOf, y)

DATA:

<http://ex.org/Ireland>



<http://ex.org/Dublin>



RULES: $(a, \text{capital}, b) \rightarrow (b, \text{partOf}, a)$
 $(c, \text{partOf}, d), (d, \text{partOf}, e) \rightarrow (c, \text{partOf}, e)$

OUTPUT: $\{(x \mapsto \text{Ireland}, y \mapsto \text{Europe}), (x \mapsto \text{Dublin}, y \mapsto \text{Ireland})$
 $(x \mapsto \text{Dublin}, y \mapsto \text{Europe})\}$

So how can we include links in this picture?

RDF filled with IRIs!

```
@base <http://ex1.org/> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
@prefix ex1: <http://ex1.org/#> .  
<#Jen> a <http://ex1.org/#Person> , ex1:Female ;  
  rdfs:label "Jen"@en ; <#allergy> <#Citrus> ;  
  ex1:location [ ex1:lat 53.3 ; ex1:long -9.0 ] .
```

... any IRI could be a link!

We already saw this in Lab 1 ...

```
@base <http://anakena.dcc.uchile.cl/~calvarez/foaf.ttl> .

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

<#me> a foaf:Person;
      foaf:name "Camila Andrea Álvarez"@es;
      foaf:firstName "Camila"@es;
      foaf:surname "Álvarez"@es;
      foaf:birthday "02-24";
      foaf:gender "female";
      foaf:homepage <http://anakena.dcc.uchile.cl/~calvarez/cv.html>;
      foaf:based_near <http://sws.geonames.org/3871336/>;
      foaf:interest <http://dbpedia.org/resource/Tales_(series)>,
<http://dbpedia.org/resource/Embroidery>;
      foaf:img <http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;
      foaf:schoolHomepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;
      foaf:knows <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,
<http://anakena.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,
<http://anakena.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

These link to other RDF documents ...

```
@base <http://anakena.dcc.uchile.cl/~calvarez/foaf.ttl> .

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

<#me> a foaf:Person;
    foaf:name "Camila Andrea Álvarez"@es;
    foaf:firstName "Camila"@es;
    foaf:surname "Álvarez"@es;
    foaf:birthday "02-24";
    foaf:gender "female";
    foaf:homepage <http://anakena.dcc.uchile.cl/~calvarez/cv.html>;
    foaf:based_near <http://sws.geonames.org/3871336/>;
    foaf:interest <http://dbpedia.org/resource/Tales_(series)>,
<http://dbpedia.org/resource/Embroidery>;
    foaf:img <http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;
    foaf:schoolHomepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;
    foaf:knows <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,
<http://anakena.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,
<http://anakena.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

```
@name "Camila"@es,  
name "Álvarez"@es;  
birthday "02-24";  
gender "female";  
homepage <http://anakena.dcc.uchile.cl/~calvarez/cv.html>;  
based_near <http://sws.geonames.org/3871336/>;  
interest <http://dbpedia.org/resource/Tales_(series)>,  
<http://dbpedia.org/resource/Embroidery>;  
img <http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;  
foaf:homepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;  
foaf:knows <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,  
<http://www.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,  
<http://www.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

HTTP GET

```
@base <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl> .  
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
  
<#me> foaf:name "Jaime Salas"@es ;  
a foaf:Person ;  
foaf:homepage <http://anakena.dcc.uchile.cl/~jasalas> ;  
foaf:based_near <http://sws.geonames.org/3871336/> ;  
foaf:img <http://anakena.dcc.uchile.cl/~jasalas/web/cat_cage.jpg> ;  
foaf:knows <http://anakena.dcc.uchile.cl/~aantoine/web/foaf.ttl#yo> ;  
foaf:knows <http://anakena.dcc.uchile.cl/~calvarez/foaf.ttl#me> ;  
foaf:knows <http://anakena.dcc.uchile.cl/~jmoreno/foaf.ttl#myself> ;  
foaf:knows <http://users.dcc.uchile.cl/~sferrada/foaf.ttl#me> ;  
foaf:interest "Gaming, programming, lots of -ing." , <http://dbpedia.org/resource/Sharknado> ;  
foaf:nick "Ritto" ;  
foaf:birthday "25/08" .
```

```
name "Camila"@es,  
name "Álvarez"@es;  
birthday "02-24";  
gender "female";  
homepage <http://anakena.dcc.uchile.cl/~calvarez/cv.html>;  
located_near <http://sws.geonames.org/3871336/>;  
interests <http://dbpedia.org/resource/Tales_(series)>;  
<http://dbpedia.org/resource/Embroidery>;  
<http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;  
poolHomepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;  
foaf:is <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,  
<http://anakena.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,  
<http://anakena.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

HTTP GET

```
@prefix dbo:      <http://dbpedia.org/ontology/> .  
@prefix ns1:     <http://dbpedia.org/resource/Tales_of_Symphonia:> .  
ns1:_Dawn_of_the_New_World    dbo:series      <http://dbpedia.org/resource/Tales_(series)> .  
@prefix dbr:     <http://dbpedia.org/resource/> .  
dbr:Yoshito      dbo:wikiPageDisambiguates    <http://dbpedia.org/resource/Tales_(series)> .  
@prefix foaf:    <http://xmlns.com/foaf/0.1/> .  
<http://en.wikipedia.org/wiki/Tales_(series)>    foaf:primaryTopic  
<http://dbpedia.org/resource/Tales_(series)> .  
@prefix ns4:     <http://dbpedia.org/resource/Tales_of_the_Heroes:> .  
ns4:_Twin_Brave    dbo:series      <http://dbpedia.org/resource/Tales_(series)> .  
@prefix ns5:     <http://dbpedia.org/resource/Keroro_RPG:> .  
ns5:_Kishi_to_Musha_to_Densetsu_no_Kaizoku    dbo:series  
<http://dbpedia.org/resource/Tales_(series)> .  
@prefix dbp:     <http://dbpedia.org/property/> .  
dbr:Bandai_Namco_Games    dbp:products    <http://dbpedia.org/resource/Tales_(series)> ;  
                           dbo:product    <http://dbpedia.org/resource/Tales_(series)> .  
dbr:Tales_of_Symphonia_Chronicles    dbo:series      <http://dbpedia.org/resource/Tales_(series)> .  
dbr:Tales_of_Hearts        dbo:series      <http://dbpedia.org/resource/Tales_(series)> .  
dbr:Tales_of_Graces        dbo:series      <http://dbpedia.org/resource/Tales_(series)> .  
dbr:Tales_of_Vesperia      dbo:series      <http://dbpedia.org/resource/Tales_(series)> ;
```

These **also** link to other RDF documents ...

```
@base <http://anakena.dcc.uchile.cl/~calvarez/foaf.ttl> .

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

<#me> a foaf:Person;
    foaf:name "Camila Andrea Álvarez"@es;
    foaf:firstName "Camila"@es;
    foaf:surname "Álvarez"@es;
    foaf:birthday "02-24";
    foaf:gender "female";
    foaf:homepage <http://anakena.dcc.uchile.cl/~calvarez/cv.html>;
    foaf:based_near <http://sws.geonames.org/3871336/>;
    foaf:interest <http://dbpedia.org/resource/Tales_(series)>,
<http://dbpedia.org/resource/Embroidery>;
    foaf:img <http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;
    foaf:schoolHomepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;
    foaf:knows <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,
<http://anakena.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,
<http://anakena.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

```
foaf:based_near <http://sws.geonames.org/3871336/>;
foaf:interest <http://dbpedia.org/resource/Tales_(series)>,
p://dbpedia.org/resource/Embroidery>;
foaf:img <http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;
foaf:schoolHomepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;
foaf:knows <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,
p://anakena.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,
p://anakena.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

HTTP GET

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ns0: <http://www.w3.org/2003/06/sw-vocab-status/ns#> .
@prefix schema: <http://schema.org/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

foaf: a owl:Ontology ;
  dc11:title "Friend of a Friend (FOAF) vocabulary" ;
  dc11:description "The Friend of a Friend (FOAF) RDF vocabulary, des

foaf:Person
  a rdfs:Class, owl:Class ;
  rdfs:label "Person" ;
  rdfs:comment "A person." ;
  ns0:term_status "stable" ;
  owl:equivalentClass schema:Person, <http://www.w3.org/2000/10/swap/
  rdfs:subClassOf foaf:Agent, geo:SpatialThing ;
  rdfs:isDefinedBy foaf: ;
  owl:disjointWith foaf:Organization, foaf:Project .
```



PRE-LINKED DATA ...

Semantic Web, early days (pre-2006)

- Lots of dumps of RDF

Index of /pub/databases/uniprot/current_release/rdf/

Name	Size	Date Modified
 [parent directory]		
 README	3.2 kB	10/14/15, 2:00:00 PM
 RELEASE.meta4	10.3 kB	10/14/15, 2:00:00 PM
 RELEASE.metalink	11.5 kB	10/14/15, 2:00:00 PM
 citations.rdf.xz	590 MB	10/14/15, 2:00:00 PM
 core.owl	166 kB	10/14/15, 2:00:00 PM
 databases.rdf.xz	14.1 kB	10/14/15, 2:00:00 PM
 diseases.rdf.xz	576 kB	10/14/15, 2:00:00 PM
 enzyme.rdf.xz	690 kB	10/14/15, 2:00:00 PM
 go.owl.xz	2.9 MB	10/14/15, 2:00:00 PM
 journals.rdf.xz	274 kB	10/14/15, 2:00:00 PM
 keywords.rdf.xz	178 kB	10/14/15, 2:00:00 PM
 locations.rdf.xz	61.7 kB	10/14/15, 2:00:00 PM
 pathways.rdf.xz	49.9 kB	10/14/15, 2:00:00 PM
 proteomes.rdf.xz	3.4 MB	10/14/15, 2:00:00 PM
 taxonomy.rdf.xz	30.1 MB	10/14/15, 2:00:00 PM
 tissues.rdf.xz	34.1 kB	10/14/15, 2:00:00 PM
 uniparc.rdf.xz	52.8 GB	10/14/15, 2:00:00 PM
 uniprot.rdf.xz	61.9 GB	10/14/15, 2:00:00 PM
 uniref.rdf.xz	36.3 GB	10/14/15, 2:00:00 PM
 void.rdf	317 kB	10/14/15, 2:00:00 PM

Semantic Web, early days (pre-2006)

- Lots of dumps of RDF
- Big OWL ontologies (difficult to re-use)

DAML Ontology Library

Summaries

- [Ontologies by URI](#)
- [Ontologies by Submission Date](#)
- [Ontologies by Keyword](#)
- [Ontologies by Open Directory Category](#)
- [Ontologies by Class](#)
- [Ontologies by Property](#)
- [Ontologies by Namespace Used](#)
- [Ontologies by Funding Source](#)
- [Ontologies by Submitting Organization](#)

Queries

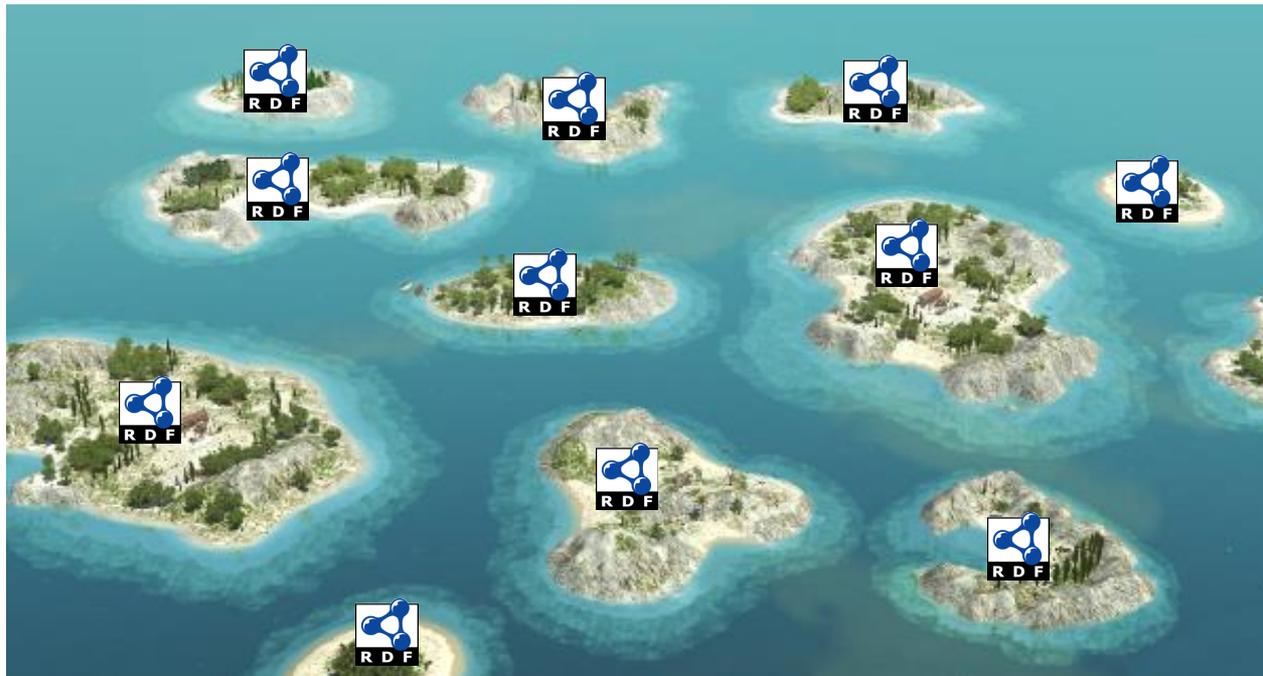
- [Classes by Name](#)
- [Properties by Name](#)

Statistics

- 282 ontologies
- [Number of Classes, Properties, and Instances Defined](#)
- [DAML Property \(Feature\) Use by Ontology](#)

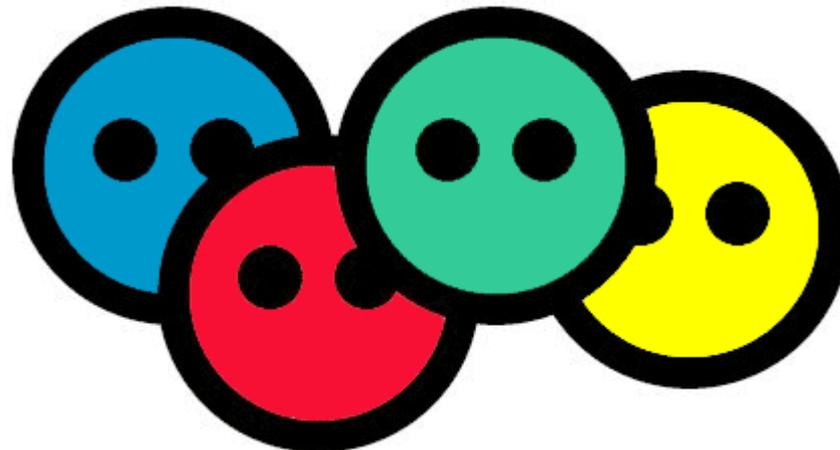
Semantic Web, early days (pre-2006)

- Lots of dumps of RDF
- Big OWL ontologies (difficult to re-use)
- No reuse of IRIs ... no links ... no Web!



Semantic Web, early days (pre-2006)

- Lots of dumps of RDF
- Big OWL ontologies (difficult to re-use)
- No reuse of IRIs ... no links ... no Web!
- Some exceptions, like FOAF
 - but still, blank nodes used more often than IRIs!



LINKED DATA ...

Linked Data ... 2006



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

Tim Berners-Lee

Date: 2006-07-27, last change: \$Date: 2009/06/18 18:24:33 \$

Status: personal view only. Editing status: imperfect but published.

[Up to Design Issues](#)

Linked Data

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF,. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow:



(I'll explain the mug later)

Four Principles of Linked Data

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



1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

This is Linked Data ...

```
@base <http://anakena.dcc.uchile.cl/~calvarez/foaf.ttl> .

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

<#me> a foaf:Person;
      foaf:name "Camila Andrea Álvarez"@es;
      foaf:firstName "Camila"@es;
      foaf:surname "Álvarez"@es;
      foaf:birthday "02-24";
      foaf:gender "female";
      foaf:homepage <http://anakena.dcc.uchile.cl/~calvarez/cv.html>;
      foaf:based_near <http://sws.geonames.org/3871336/>;
      foaf:interest <http://dbpedia.org/resource/Tales_(series)>,
<http://dbpedia.org/resource/Embroidery>;
      foaf:img <http://images.evisos.cl/2009/06/03/erizo-de-tierra_9de6128c_3.jpg>;
      foaf:schoolHomepage <http://www.uchile.cl/>, <http://www.dcc.uchile.cl/>;
      foaf:knows <http://anakena.dcc.uchile.cl/~jasalas/web/foaf.ttl#me>,
<http://anakena.dcc.uchile.cl/~jogarrid/web/foaf.ttl#yo>,
<http://anakena.dcc.uchile.cl/~ekauffma/foaf.ttl#eli> .
```

And so we have our “Web of Data”!!!!

(1) Data, (2) Rules/Ontologies, (3) Query

(4) Links

INPUT: “(x,partOf,y)”

DATA:

<http://ex.org/Ireland>



<http://ex.org/Dublin>



RULES: $(a, \text{capital}, b) \rightarrow (b, \text{partOf}, a)$
 $(c, \text{partOf}, d), (d, \text{partOf}, e) \rightarrow (c, \text{partOf}, e)$

OUTPUT: $\{(x \mapsto \text{Ireland}, y \mapsto \text{Europe}), (x \mapsto \text{Dublin}, y \mapsto \text{Ireland})$
 $(x \mapsto \text{Dublin}, y \mapsto \text{Europe})\}$

THIS SEEMS EASY



TOO EASY



So what's the catch?

ONE MINOR CATCH ...

... NAMING *THINGS* NOT DOCUMENTS

HTTP IRIs usually for documents, not pipes



My FOAF file is not me

<http://aidanhogan.com/foaf.rdf>

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:toilet="http://purl.org/az/Toilet#"
xmlns:rel="http://www.perceive.net/schemas/20031015/relationship/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#">

<foaf:Document rdf:about="http://aidanhogan.com/foaf.rdf">
<dc:title xml:lang="en">Aidan's FOAF profile</dc:title>
<dc:creator rdf:resource="http://aidanhogan.com/foaf.rdf#me" />
<dc:date xml:datatype="http://www.w3.org/2001/XMLSchema#date">2005-04-07</dc:date>
<foaf:primaryTopic rdf:resource="http://aidanhogan.com/foaf.rdf#me" />
</foaf:Document>

<foaf:Person rdf:about="http://aidanhogan.com/foaf.rdf#me">

<owl:sameAs>
<foaf:Person rdf:about="http://www.deri.ie/about/team/member/Aidan_Hogan#me">
<rdfs:seeAlso rdf:resource="http://www.deri.ie/fileadmin/scripts/foaf.php?id=231" />
</foaf:Person>
</owl:sameAs>

<owl:sameAs rdf:resource="http://sw.deri.org/~aidanh/foaf/foaf.rdf#Aidan_Hogan" />
<owl:sameAs rdf:resource="http://aidanhogan.com/foaf/foaf.rdf#Aidan_Hogan" />
<owl:sameAs rdf:resource="http://data.semanticweb.org/person/aidan-hogan" />
```

How to identify things not documents?

- We could invent a new protocol `real:///?`
`real://aidanhogan.com/foaf.rdf`

Problems?

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

Solutions?

How to identify things not documents

1. Hash (aka. fragment identifier)

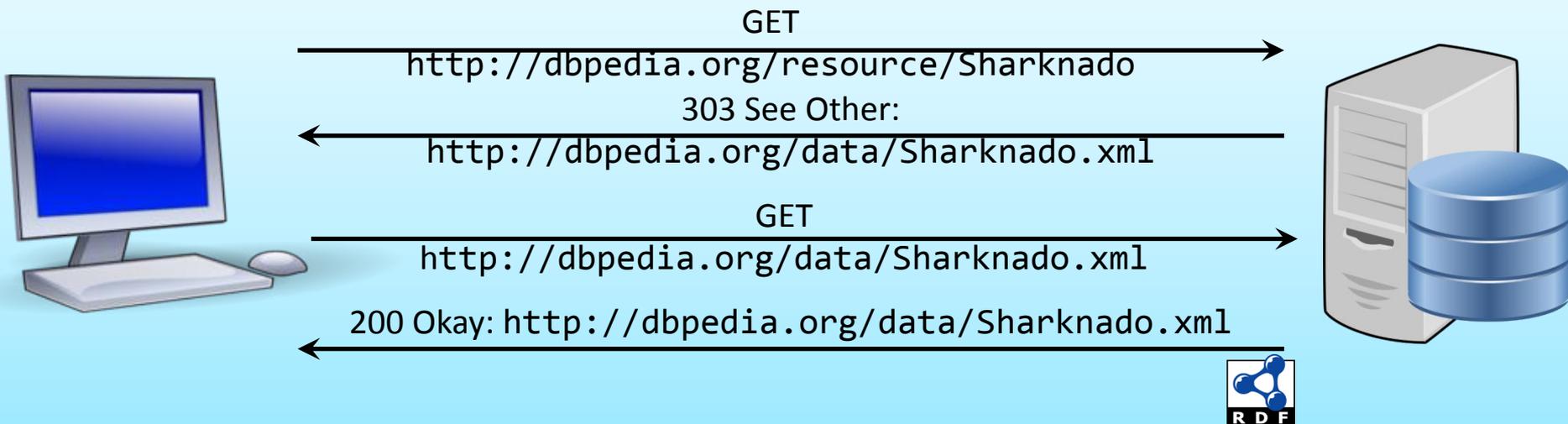
- `http://aidanhogan.com/foaf.rdf`
 - Identifies my FOAF file
- `http://aidanhogan.com/foaf.rdf#me`
 - Identifies me
 - Look it up with HTTP, you get my FOAF file



How to identify things not documents

2. Slash (with 303 redirect)

- `http://dbpedia.org/data/Sharknado.xml`
 - Identifies RDF document about Sharknado
- `http://dbpedia.org/resource/Sharknado`
 - Identifies the movie Sharknado
 - Look it up, it redirects with 303 to RDF doc. above

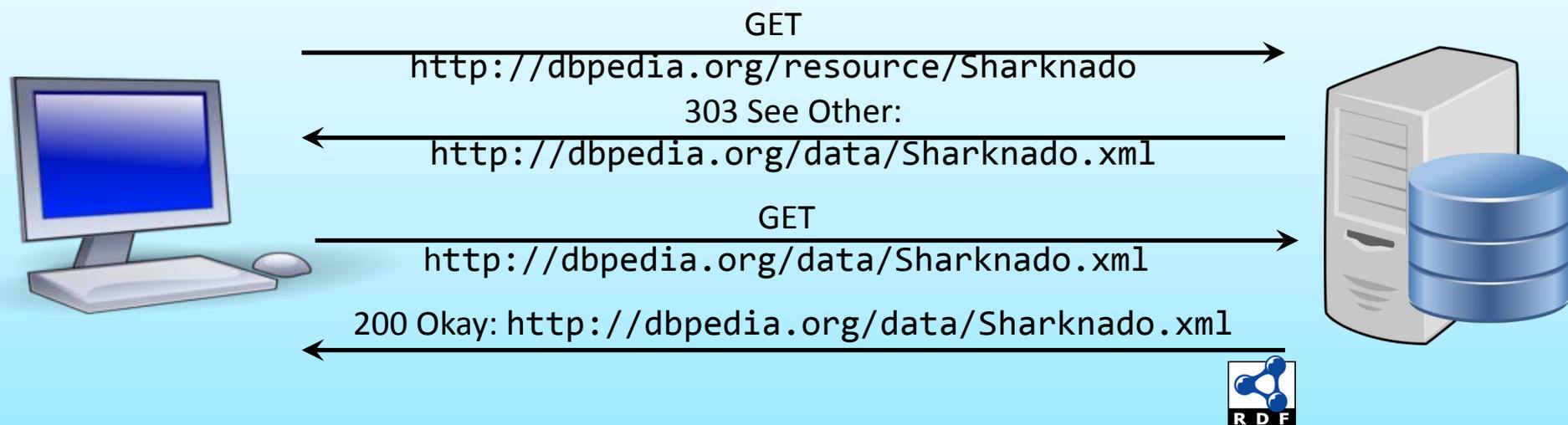


Hash vs. Slash



Which is better, **hash** or **slash**?

Well, **hash** has half the number of requests!



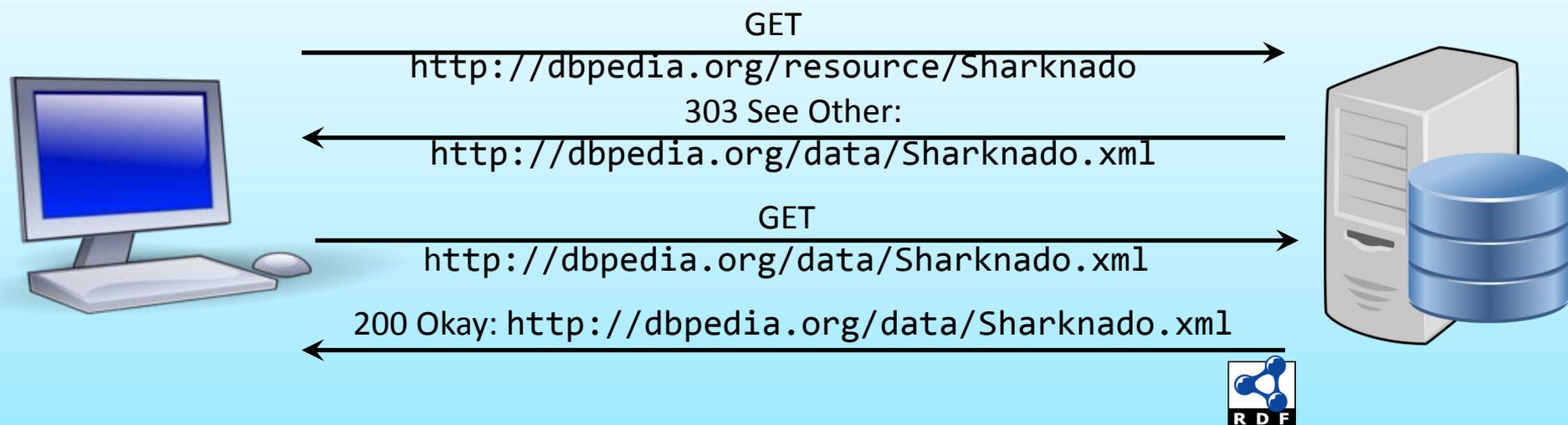
Why does **slash** exist then?

- **Hash** often used to describe multiple resources in one document
- **Slash** often used when there are too many resources to fit in one document
 - But nothing would stop, e.g.:
 - `http://dbpedia.org/data/Sharknado.xml#this`

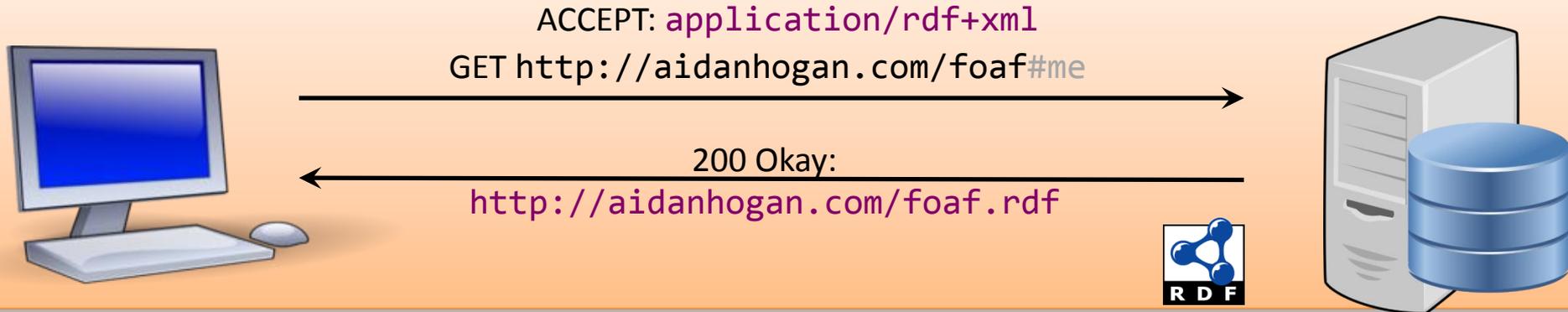
So ... erm ... why does slash exist then?



Well, server has more flexibility. It can change which document is returned. In the hash scenario, the document is fixed in the resource IRI.

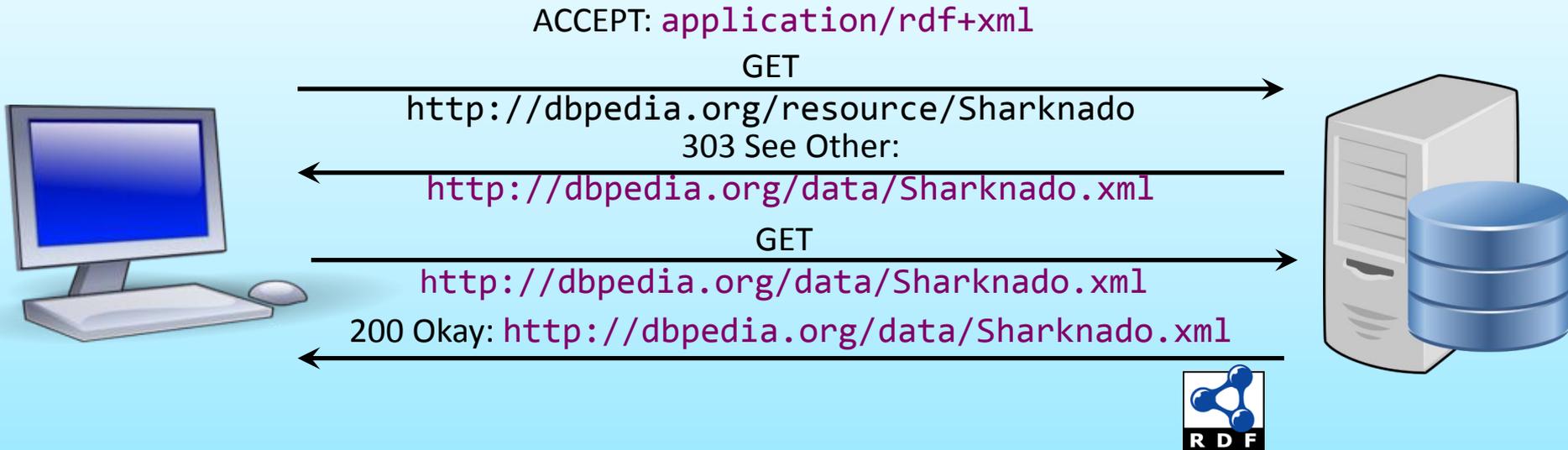


Content negotiation with hash



Can also choose from different RDF formats; e.g., Turtle, RDFa, etc.
(if supported by the server that is!)
Can also specify preferences using q values ...

Content negotiation with slash



LINKING OPEN DATA

Linked Data ... 2006



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

Tim Berners-Lee

Date: 2006-07-27, last change: \$Date: 2009/06/18 18:24:33 \$

Status: personal view only. Editing status: imperfect but published.

[Up to Design Issues](#)

Linked Data

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF,. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow:



(I'll explain the mug later)

Open Data, meet Linked Data



We've got all these people who want to publish Open Data but how should they publish it on the Web?

We've got this new way of publishing Linked Data on the Web but no data to publish ...



Linked Open Data



Are you thinking what I'm thinking?

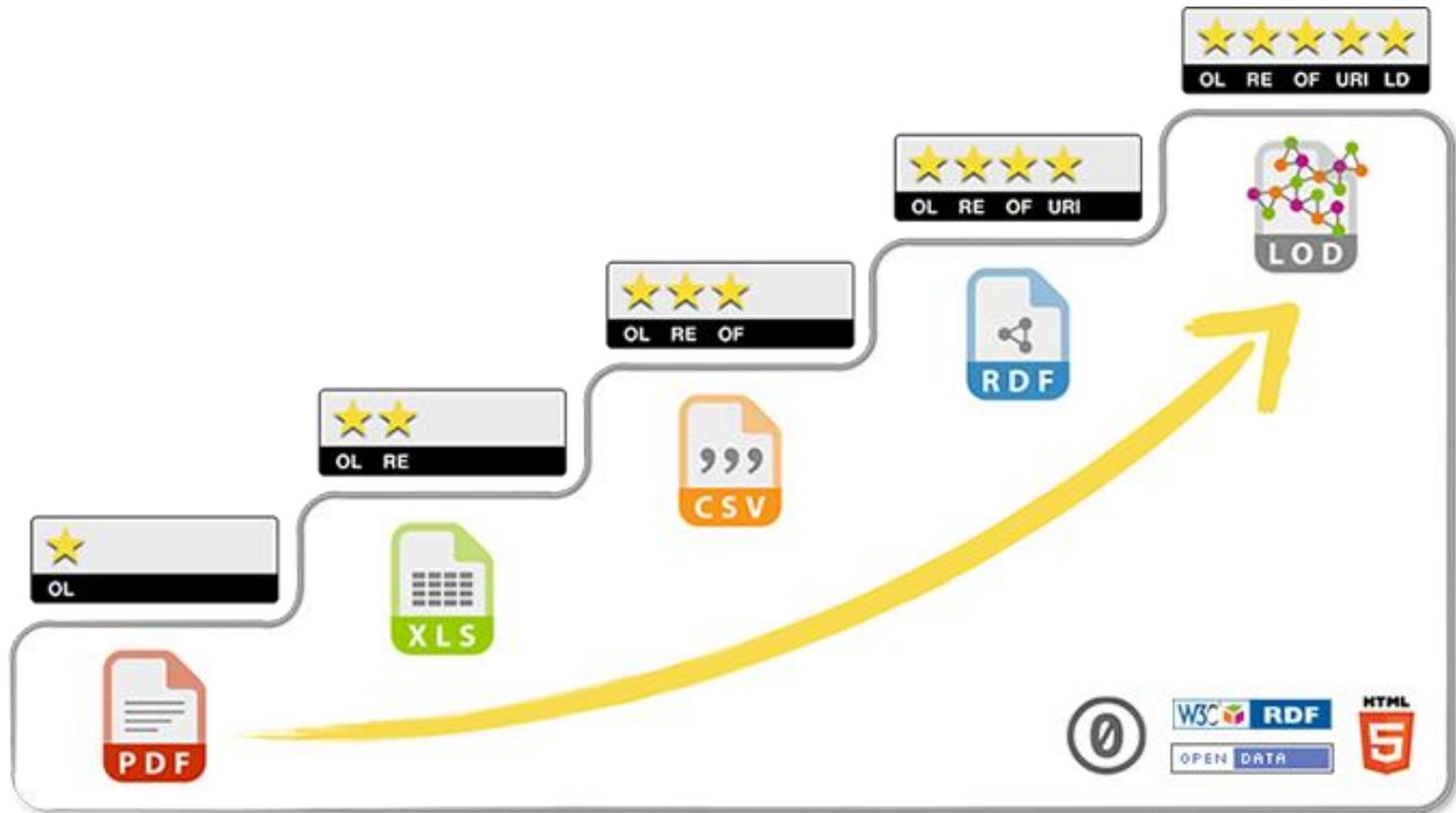


The 5 ★'s of Linked Open Data

- ★ Publish data under open licence
- ★★ Make the data “machine readable”
 - e.g., a Spreadsheet better than a PDF table
- ★★★ Use non-proprietary formats
 - e.g., a CSV text file better than Excel
- ★★★★ Use URIs to name your stuff ([hint: RDF](#))
 - use unambiguous identifiers that can be linked/looked up
- ★★★★★ Provide links to other content
 - so consumers can follow links to find out more

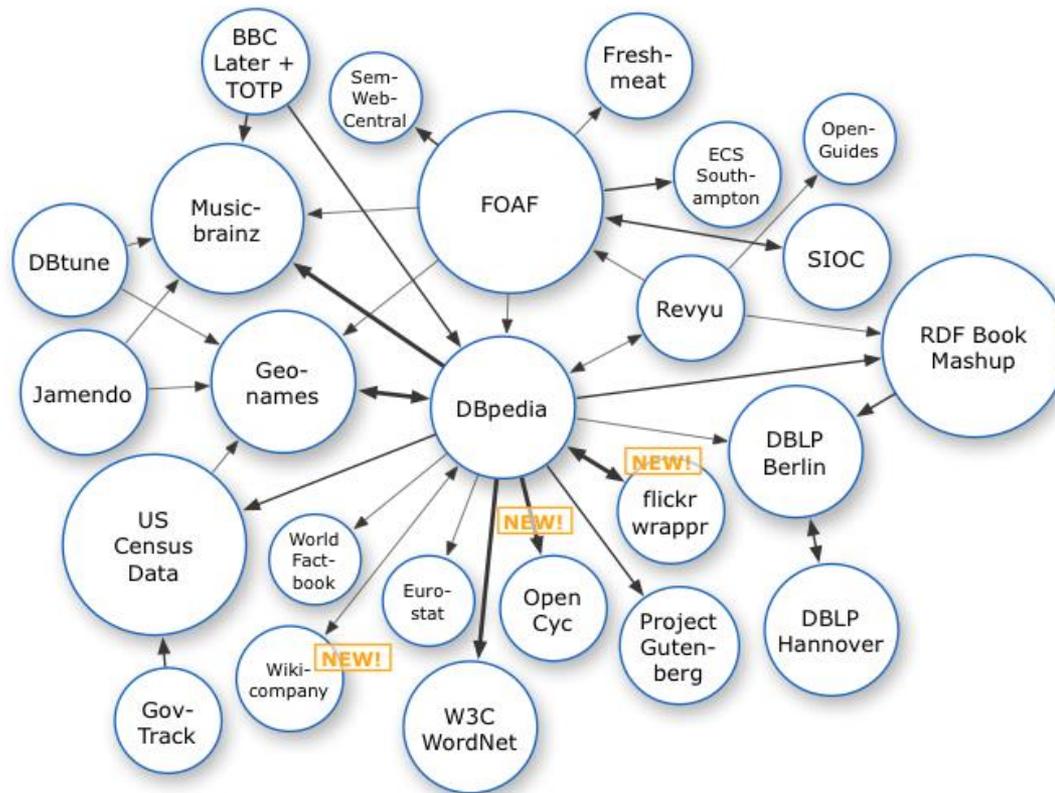


Each star improves interoperability of data



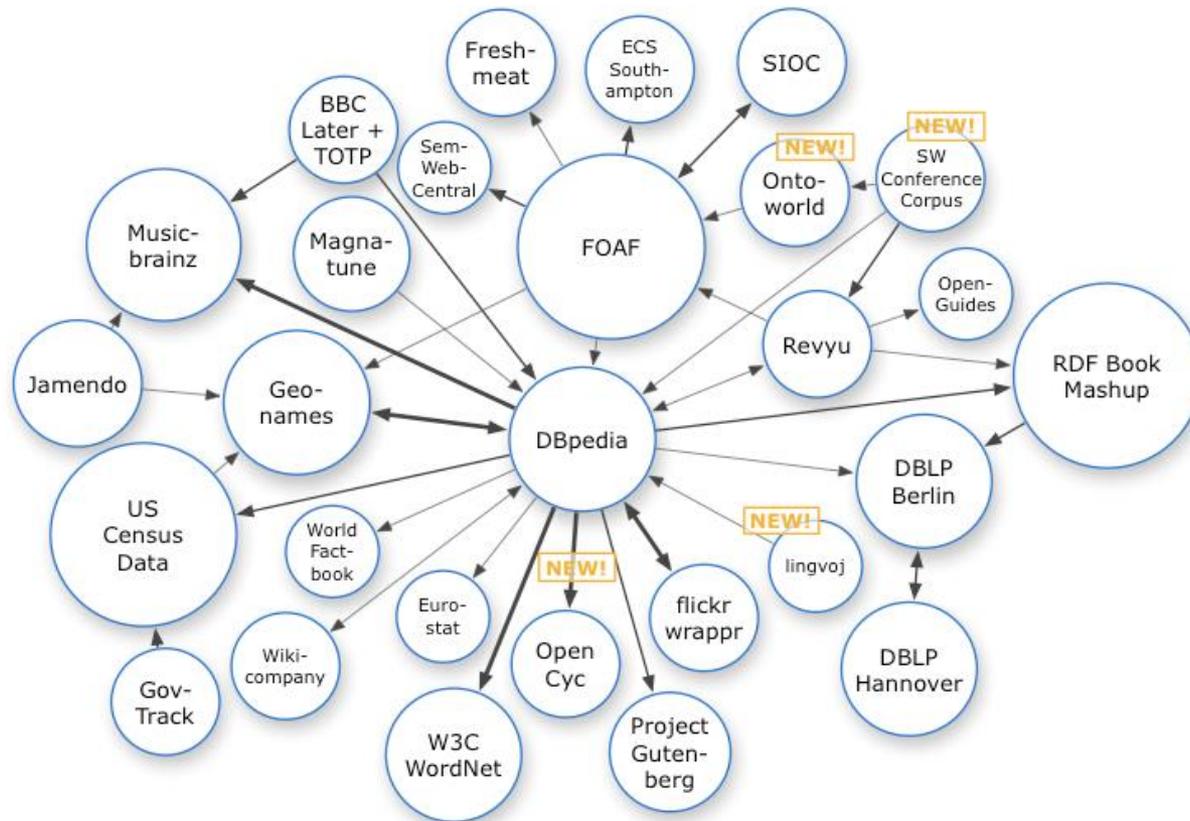
Growth of the Linked Data Cloud

Oct. 2007

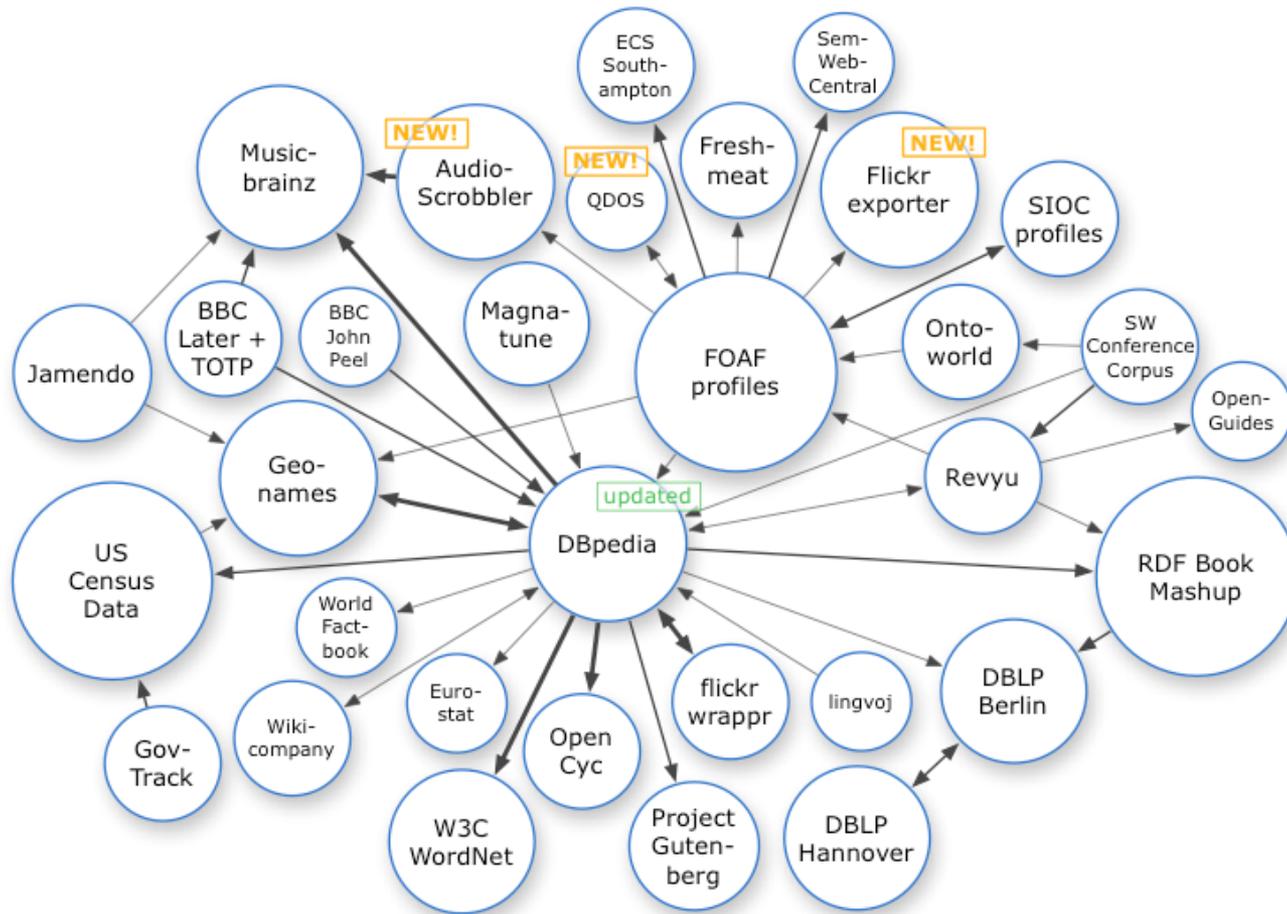


Growth of the Linked Data Cloud

Oct. 2007
Nov. 2007

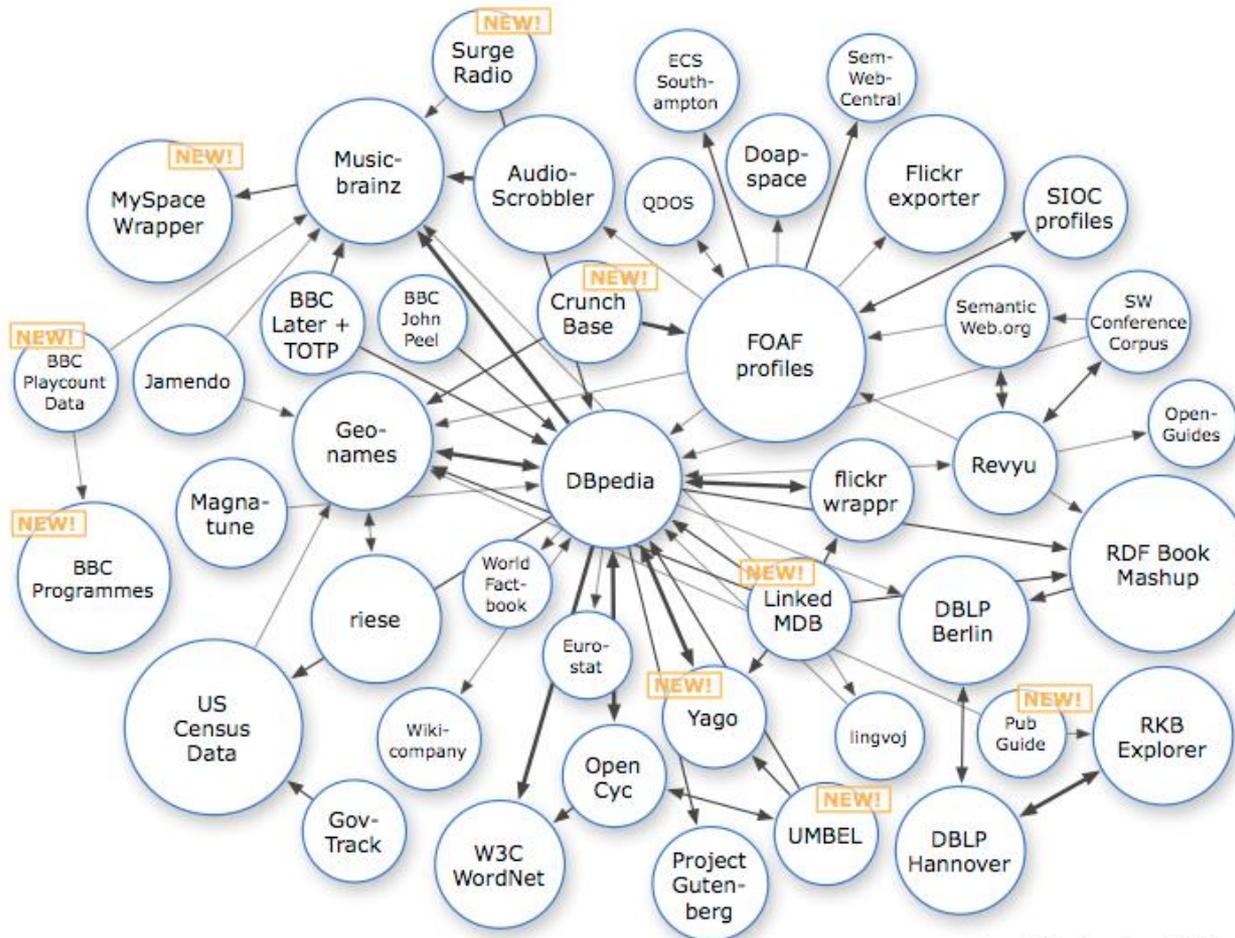


Growth of the Linked Data Cloud



Oct. 2007
Nov. 2007
Feb. 2008

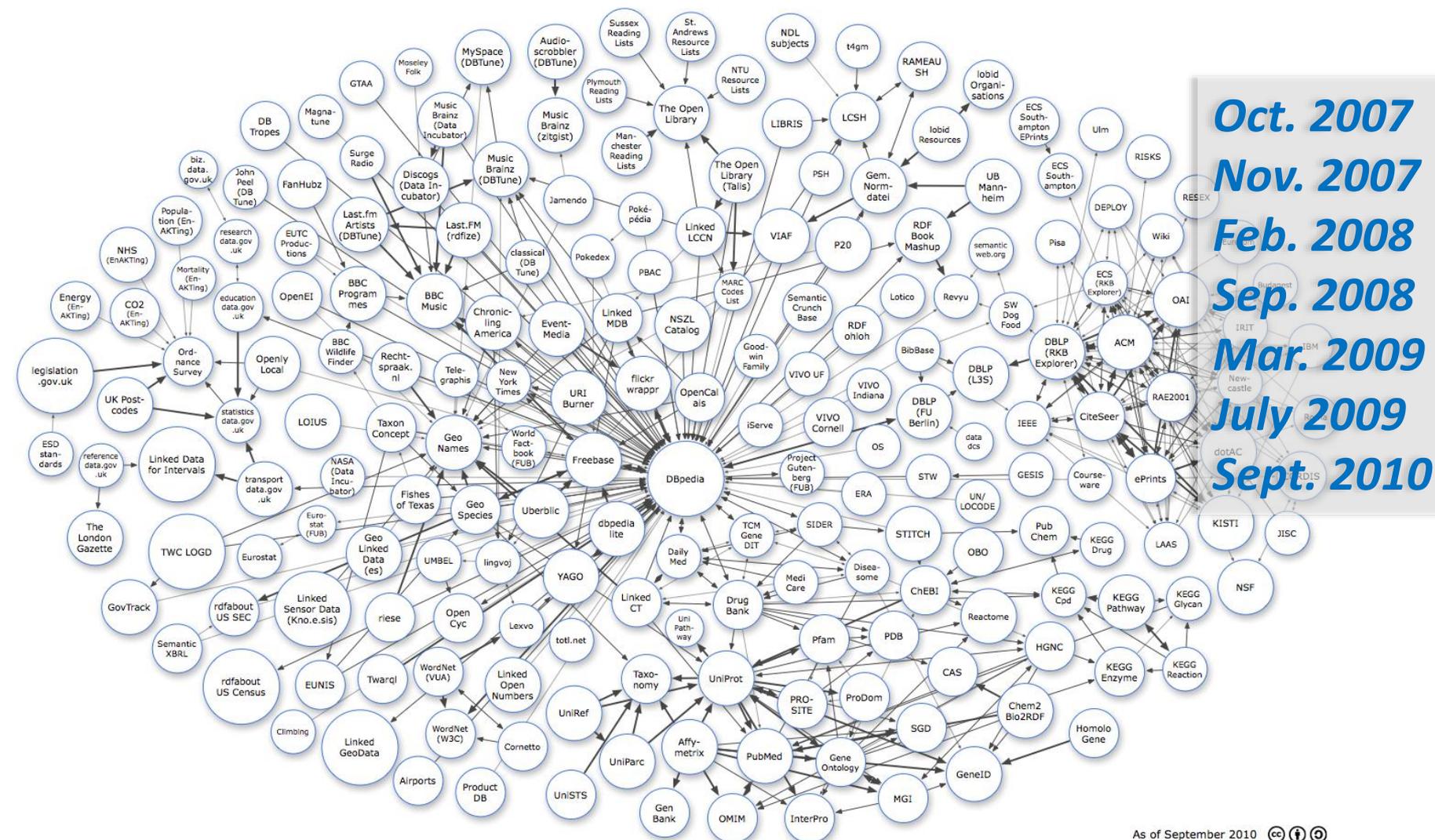
Growth of the Linked Data Cloud



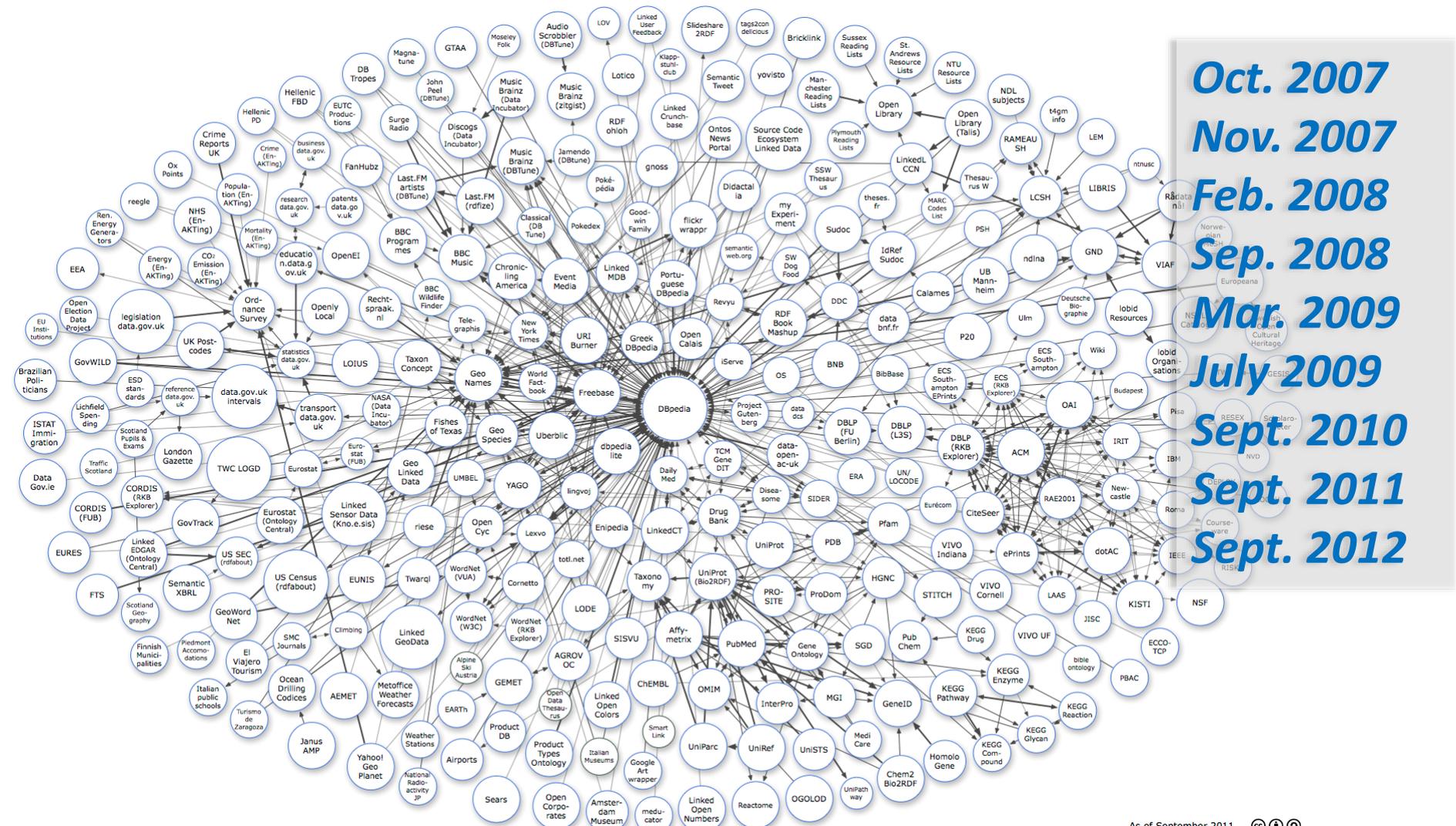
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008

As of September 2008

Growth of the Linked Data Cloud

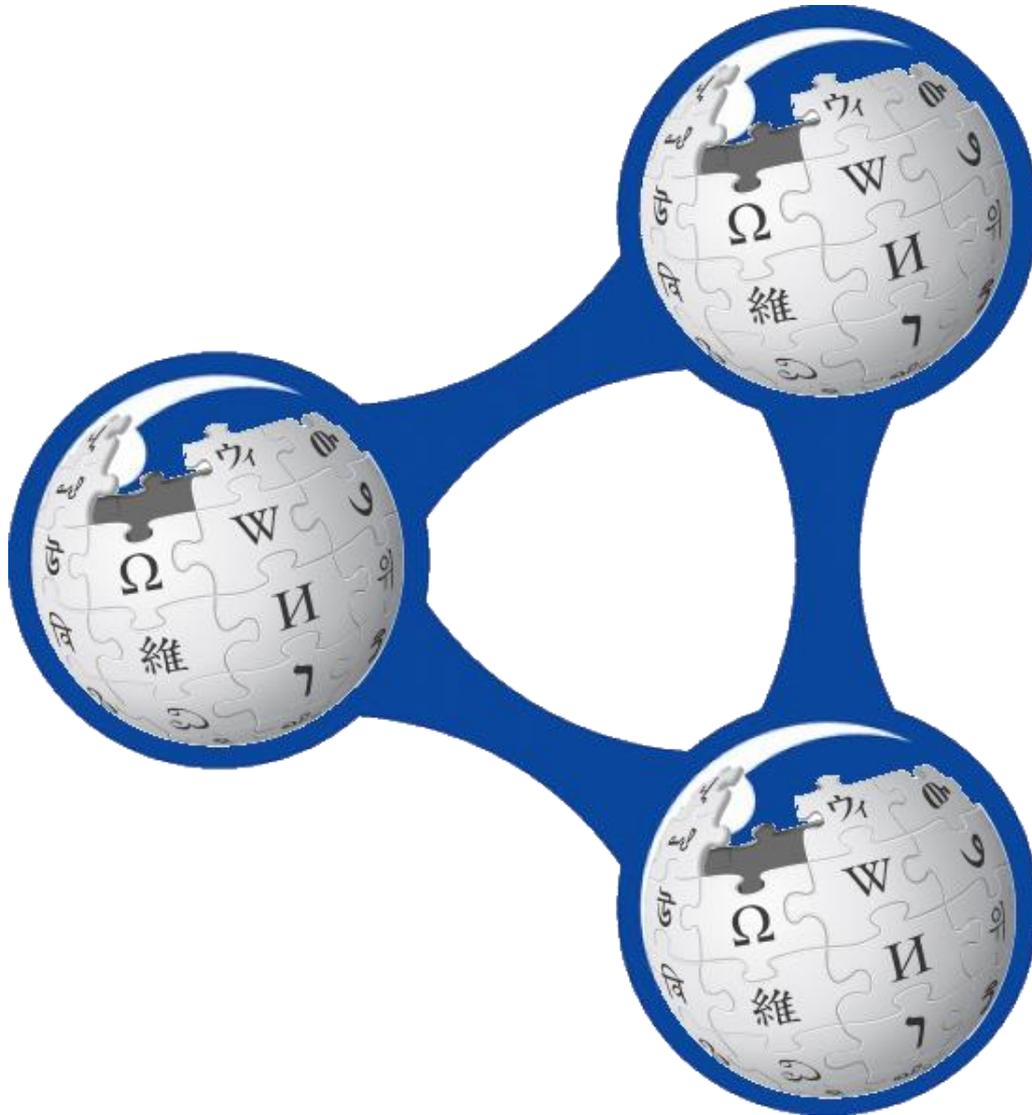


Growth of the Linked Data Cloud

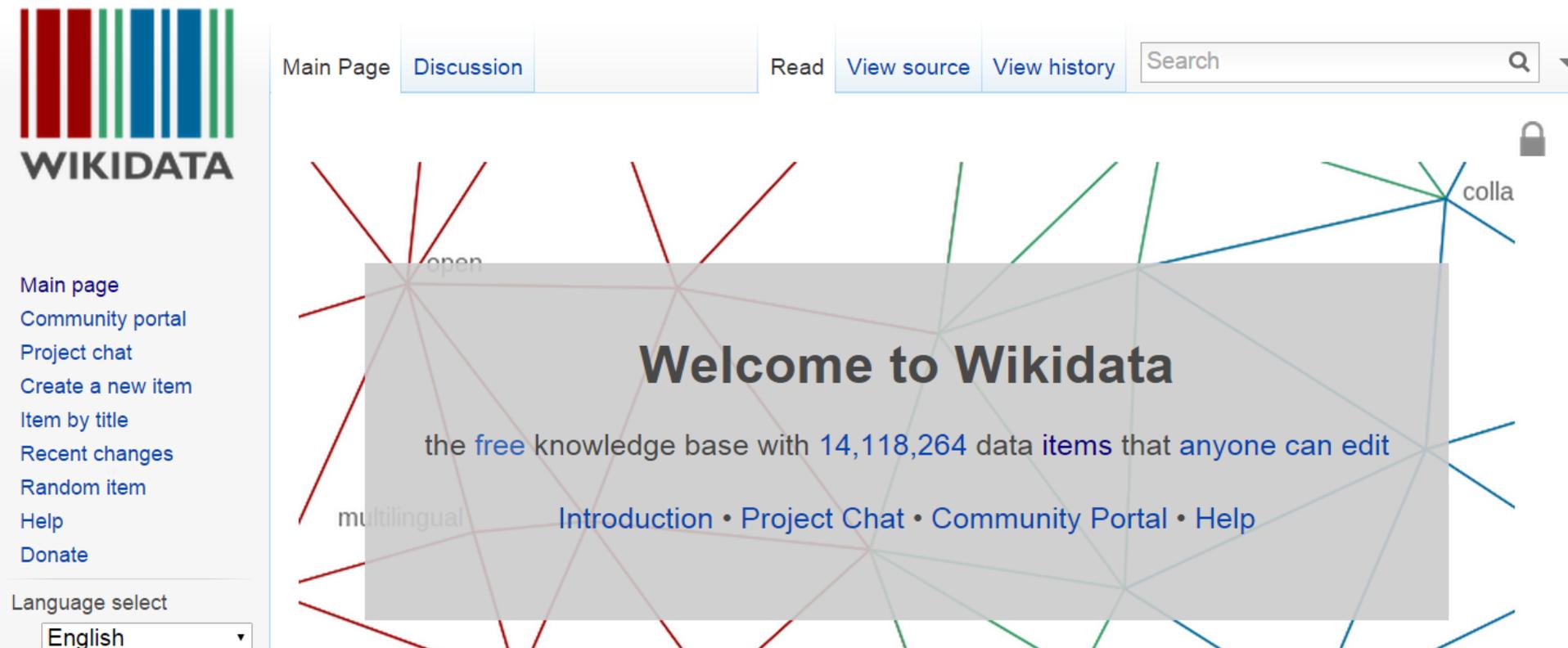


**THERE IS A LOT OF LINKED DATA
OUT THERE NOW ...**

DBpedia: An RDF Encyclopaedia



Wikidata: Another RDF Encyclopaedia



The image shows the Wikidata main page with a network diagram overlay. The network consists of nodes and edges. Nodes are represented by colored lines radiating from a central grey box. The colors include red, green, and blue. Some nodes are labeled with text: 'open' (top left), 'colla' (top right), and 'multilingual' (bottom left). The central grey box contains the following text:

Welcome to Wikidata
the [free](#) knowledge base with [14,118,264](#) data items that anyone can edit
[Introduction](#) • [Project Chat](#) • [Community Portal](#) • [Help](#)

The page layout includes the Wikidata logo (vertical bars) and the word 'WIKIDATA' on the left. The top navigation bar contains 'Main Page', 'Discussion', 'Read', 'View source', and 'View history'. A search box is on the right. A sidebar on the left lists navigation options: 'Main page', 'Community portal', 'Project chat', 'Create a new item', 'Item by title', 'Recent changes', 'Random item', 'Help', and 'Donate'. A language select dropdown is at the bottom left, set to 'English'. A lock icon is visible in the top right corner.

Geographical Data: Geonames

GeoNames About Browse Download API Help Paris, Mount Everest, New York anonymous

Found 35 items in this area

Kalamaki
P PPL populated place 7874338
Greece GR » Crete ESVE43 » Irákleion AS » Faistos 0908
35.0281, 24.76009 N 35°01'41" E 24°45'36"
geotree .kml .rdf

Ormos Mesaras

Map Satellite

Map data ©2013 Google Imagery ©2013 Cnes/Spot Image, DigitalGlobe, European Space Imaging, Landsat 1 km Terms of Use Report a map error

Linked Government Data: data.gov

An Official Web Site of the United States Government Tuesday, May 24, 2011 Text: A+ A- A Share

DATA.GOV
EMPOWERING PEOPLE

HOME DATA APPS COMMUNITY METRICS OPEN DATA SITES GALLERY WHAT'S NEW

Earthquake and Tsunami Datasets and Information

- Worldwide M1+ Earthquakes, Past 7 Days
- RadNet Map Interface for Near-Real-Time Radiation Monitoring Data
- Search other related datasets
- World Earthquake Interactive Map Demo

SEARCH OUR CATALOGS

Search our catalogs.. SEARCH

WORLDWIDE M1+ EARTHQUAKES, PAST 7 DAYS

Real-time, worldwide earthquake list for the past 7 days

DATA AND APPS

- 389,714 raw and geospatial datasets
- 977 government apps
- 236 citizen-developed apps

COMMUNITIES

Come explore, discuss, meet others in the same field, and develop the data and apps in the community that you care about. Join in the

OPEN GOVERNMENT

Latest News: Japanese Earthquake and Radiation Data

Linked Government Data: data.gov.uk

The screenshot shows the data.gov.uk website interface. At the top, there is a black header with the HM Government logo and a 'Log in or sign up' link. Below this is the 'data.gov.uk BETA' logo and the tagline 'Opening up government'. A search bar with the placeholder text 'What are you looking for?' and a 'Search' button is positioned to the right. A green navigation bar contains links for Home, Data, Apps, Ideas, Forum, Wiki, Blogs, Linked Data, Resources, and About. The main content area features a large background image of a data table. On the left, there are four vertical tiles: 'Call for dataset requests', 'Instructions for data publishers', 'Public Data Corporation', and 'Met office data'. A large green banner reads 'Instructions for data publishers' with the subtext 'Calling all data publishers - new guide to publishing to data.gov.uk'. To the right, a section titled 'Over 6,900 datasets to view' includes a sub-section 'Inside Government Data' with the text 'Who's who in Government and where does the money go? Follow these links to find the data that opens it all up.' Below this are three links: 'Government spend over £25,000, by department', 'Who does what in Whitehall - and how much are they paid?', and 'Hospitality, gifts and expenses'. A 'Share this' button with social media icons is also present. At the bottom, a dark grey bar contains the text 'Facts, figures, apps and more'. Below this are three columns: 'Find data of interest' with a description and a green image, 'Apps' with a description and a blue image, and 'Tags' with a list of tags: 'health (2,328)', 'care (1,646)', 'transparency (1,594)', and 'communities (1,318)'.

HM Government Log in or sign up

data.gov.uk ^{BETA}
Opening up government

What are you looking for?

Home Data Apps Ideas Forum Wiki Blogs Linked Data Resources About

Call for dataset requests

Instructions for data publishers

Public Data Corporation

Met office data

Instructions for data publishers
Calling all data publishers - new guide to publishing to data.gov.uk

Over 6,900 datasets to view

Inside Government Data

Who's who in Government and where does the money go? Follow these links to find the data that opens it all up.

- Government spend over £25,000, by department
- Who does what in Whitehall - and how much are they paid?
- Hospitality, gifts and expenses

Share this

Facts, figures, apps and more

Find data of interest

Looking for something specific, or just want to know more about how Government spends your money? You'll find datasets here to help you get answers.

Apps

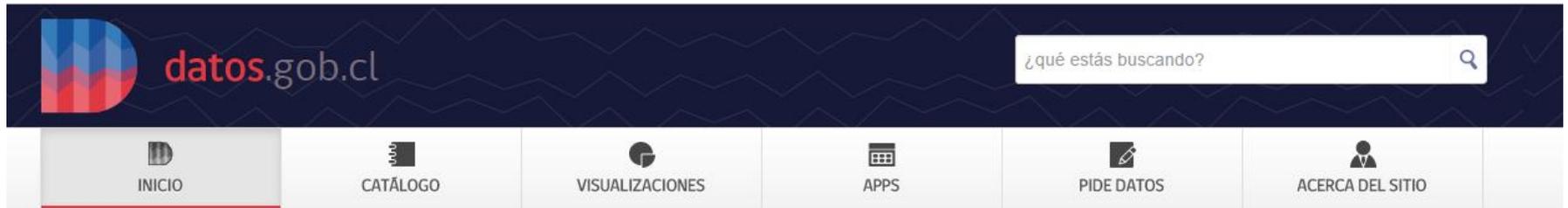
Want your phone to wake you when you get to your train station? Discover over 100 apps harnessing public data to make your life easier.

Tags

Can't decide where to start? You can browse the data by clicking on popular topics. Try one of the tags here to find what you're most interested in.

health (2,328) care (1,646)
transparency (1,594) communities (1,318)

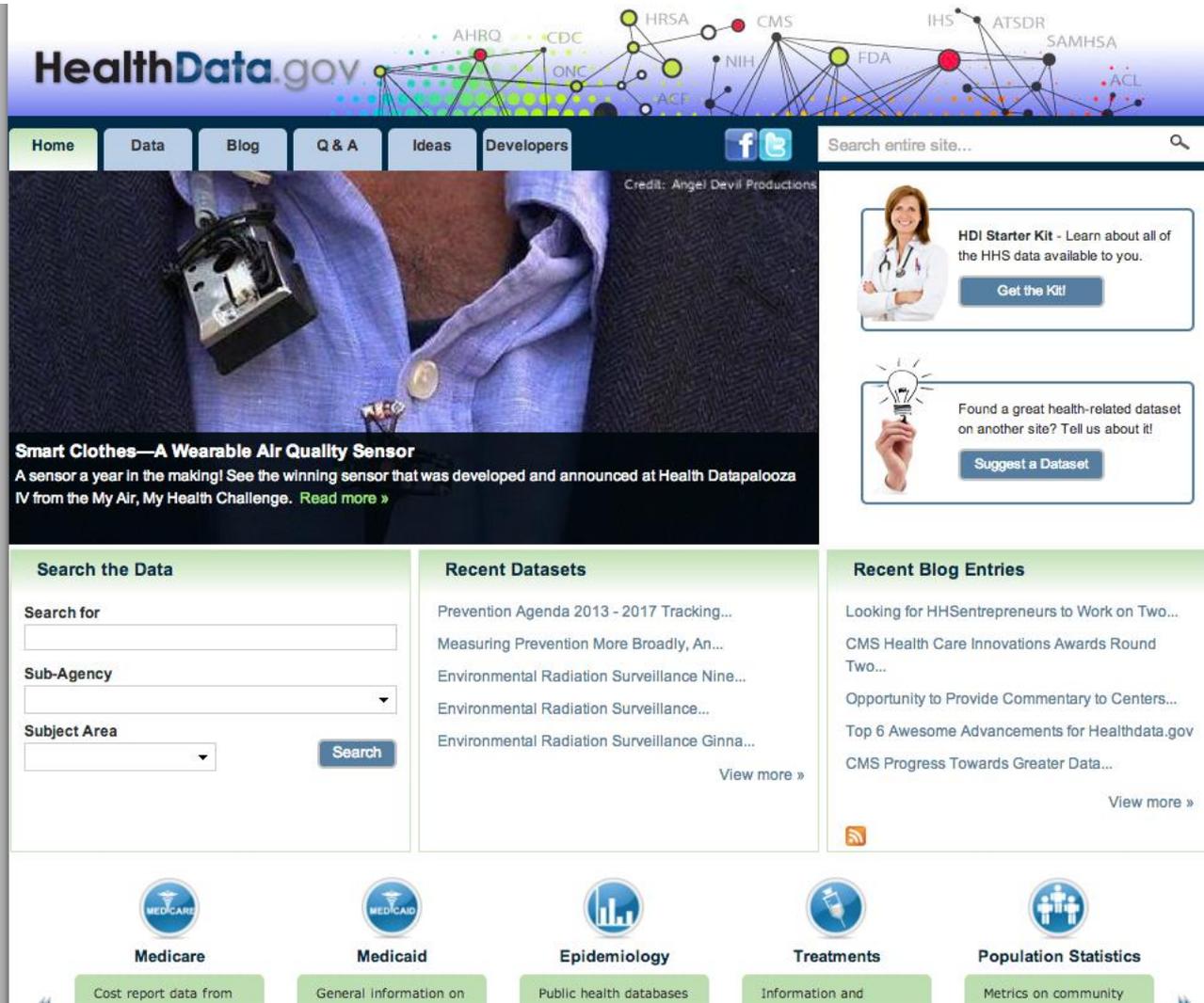
Linked Government Data: datos.gob.cl



 CATÁLOGO

DATASETS PUBLICADOS: 1.232

Life Sciences



The screenshot shows the HealthData.gov website interface. At the top, there is a network diagram with nodes representing various agencies: AHRQ, CDC, HRSA, CMS, IHS, ATSDR, SAMHSA, FDA, NIH, ONC, ACF, and ACL. Below the diagram is the 'HealthData.gov' logo. A navigation bar includes links for Home, Data, Blog, Q & A, Ideas, and Developers, along with social media icons for Facebook and Twitter. A search bar is located on the right side of the navigation bar.

Home | **Data** | **Blog** | **Q & A** | **Ideas** | **Developers** | Search entire site...

Smart Clothes—A Wearable Air Quality Sensor
A sensor a year in the making! See the winning sensor that was developed and announced at Health Datapalooza IV from the My Air, My Health Challenge. [Read more »](#)

Search the Data

Search for:

Sub-Agency:

Subject Area:

Recent Datasets

- Prevention Agenda 2013 - 2017 Tracking...
- Measuring Prevention More Broadly, An...
- Environmental Radiation Surveillance Nine...
- Environmental Radiation Surveillance...
- Environmental Radiation Surveillance Ginna...

[View more »](#)

Recent Blog Entries

- Looking for HHSentreprenuers to Work on Two...
- CMS Health Care Innovations Awards Round Two...
- Opportunity to Provide Commentary to Centers...
- Top 6 Awesome Advancements for Healthdata.gov
- CMS Progress Towards Greater Data...

[View more »](#)

Medicare | **Medicaid** | **Epidemiology** | **Treatments** | **Population Statistics**

Cost report data from | General information on | Public health databases | Information and | Metrics on community

Life Sciences

ClinicalTrials.gov

A service of the U.S. National Institutes of Health

ClinicalTrials.gov is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world. [Learn more about clinical studies](#) and [about this site](#), including relevant history, policies, and laws.

[Find Studies](#) [About Clinical Studies](#) [Submit Studies](#) [Resources](#) [About This Site](#)

ClinicalTrials.gov currently lists **151,056 studies** with locations in all 50 states and in **184 countries**.

[Text Size](#)

Search for Studies

Example: "Heart attack" AND "Los Angeles"

[Advanced Search](#) | [See Studies by Topic](#)
[See Studies on a Map](#)

Search Help

- [How to search](#)
- [How to find results of studies](#)
- [How to read a study record](#)

Locations of Recruiting Studies



Total N = 30,860 studies
Data as of August 27, 2013

- [See more trends, charts, and maps](#)

Learn More

- [ClinicalTrials.gov Online Training](#)
- [Glossary of common site terms](#)

[For the Press](#)

[Using our RSS Feeds](#)

For Patients & Families

- [How to find studies](#)
- [See studies by topic](#)
- [Learn about clinical studies](#)
- [Learn more...](#)

For Researchers

- [How to submit studies](#)
- [Download content for analysis](#)
- [About the results database](#)
- [Learn more...](#)

For Study Record Managers

- [Why register?](#)
- [How to register study records](#)
- [FDAAA 801 Requirements](#)
- [Learn more...](#)

[HOME](#) [RSS FEEDS](#) [SITE MAP](#) [TERMS AND CONDITIONS](#) [DISCLAIMER](#) [CONTACT NLM HELP DESK](#)

Life Sciences



Home Browse Search Downloads About Help Tools Contact Us

Search: [Help / Advanced](#)

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 6811 drug entries including 1528 FDA-approved small molecule drugs, 150 FDA-approved biotech (protein/peptide) drugs, 87 nutraceuticals and 5080 experimental drugs. Additionally, 4294 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 150 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

DrugBank is supported by [David Wishart](#), Departments of [Computing Science & Biological Sciences](#), [University of Alberta](#).

DrugBank is also supported by [The Metabolomics Innovation Centre](#), a Genome Canada-funded core facility serving the scientific community and industry with world-class expertise and cutting-edge technologies in metabolomics.

[More about DrugBank](#)

Tweets



 **Wishart Lab**
@WishartLab

16 Jun

You can now browse Reactions ([drugbank.ca/reactions](#)) in DrugBank, data is in BETA. Feedback appreciated, more data coming soon!
[#drugbank](#)

Expand

[Load More](#)

Compose new Tweet...

Recent Comments

PEOPLE RECENT POPULAR

Recent Comments



[Brenton Home](#)

In Australia it is a Schedule 8 (controlled) drug.

E-Commerce

Feedback

Automotive

Manufacturers and dealers can publish rich details of new and used vehicles so that search engines and browser extensions present them all to your potential customers.

[Read more](#)



<<Prev Next>>

The most powerful Web vocabulary for e-commerce
A paradigm shift for e-commerce. Since 2008.

Only 5% of all potential visitors of your site will actually see your offers in their original beauty. 95% will never get beyond a reduced preview of your great products and services as provided by a Web search engine.

GoodRelations is the most powerful vocabulary for publishing all of the details of your products and services in a way friendly to search engines, mobile applications, and browser extensions. By adding a bit of extra code to your Web content, you make sure that potential customers realize all the great features and services and the benefits of doing business with you, because their computers can extract and present this information with ease.

Video



Contact

Univ.-Prof. Dr. Martin Hepp

Who uses GoodRelations?

Google
Yahoo!
BestBuy
sears.com
kmart.com

... and 10,000 more

See [here](#) for additional references.

License

The GoodRelations ontology is  **creative**

News from Twitter

Acknowledgments

Many organizations and individuals have supported

New York Times Meta-data

The New York Times

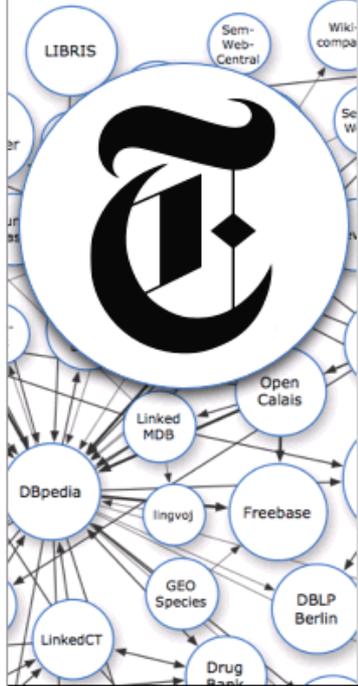
Linked Open Data BETA

[View Application Source](#)

Alumni In The News

Enter a school name below and see our coverage of that school's alumni.

San Francisco State University



George Miller
Attorney
Born: May 17, 1945

- [Congress Considers Concussion Protections](#) - September 24, 2010
- [EDITORIAL; Fairness for Older Workers](#) - September 14, 2010
- [EDITORIAL; Saving the Teachers](#) - May 06, 2010
- [House Bill Would Assure Workers Paid Sick Days](#) - November 04, 2009
- [EDITORIAL; Preventing Age Discrimination](#) - October 13, 2009
- [OP-ED COLUMNIST; Someday, a Bill Will Pass](#) - September 17, 2009
- [Obama Plan to End Role of Banks in Federal Student Loans Wins Support](#) - July 11, 2009
- [House Unveils Health Bill, Minus Key Details](#) - June 20, 2009
- [Democrats Nearing Consensus on Health](#) - June 10, 2009
- [U.S. Charges 7 Accused of Ties To Bonannos](#) - August 29, 2008

Please note that portions of this application rely on user generated data from external sources.
It is hoped but not guaranteed that this data is accurate.

[About This Page](#) | [Copyright 2009 The New York Times Company](#) | [Terms of Service](#) | [Privacy Policy](#) | [Work for Us](#)

<http://data.nytimes.com/schools/schools.html>

BBC Music

BBC Sign in News Sport Weather iPlayer TV Radio More... London 2012 Search

MUSIC

HOME | SHOWCASE | REVIEWS | GENRES

The Beatles

Formed 1957. Disbanded 10 April 1970.

PLAYED MOST ON **BBC RADIO 2**

Share This Page 15 so far

Share facebook twitter

BBC Music Showcase

BBC MUSIC SHOWCASE
Watch and listen to exclusive music clips

Latest Tracks Played On The BBC

- Twist & Shout**
BBC RADIO 2 | ZOE BALL AND RICHARD BACON GO FOR GOLD 01/08/2012
- Tomorrow Never Knows**
BBC 6 MUSIC | RADCLIFFE AND MACONIE WEDNESDAY - JOHNNY SHARP
- She Loves You**
BBC RADIO 2 | KEN BRUCE CLAUDIA WINKLEMAN SITS IN
- Sgt Pepper's Lonely Hearts/With A Little Help**
BBC RADIO 2 | STEVE WRIGHT IN THE AFTERNOON PATRICK KIELTY SITS IN

BBC TWO



Brian Epstein Finds The Beatles

ARENA | BBC TWO

Brian Epstein, Paul McCartney and John Lennon talk about the early days of The Beatles in the Arena documentary, The Brian Epstein Story (1998)

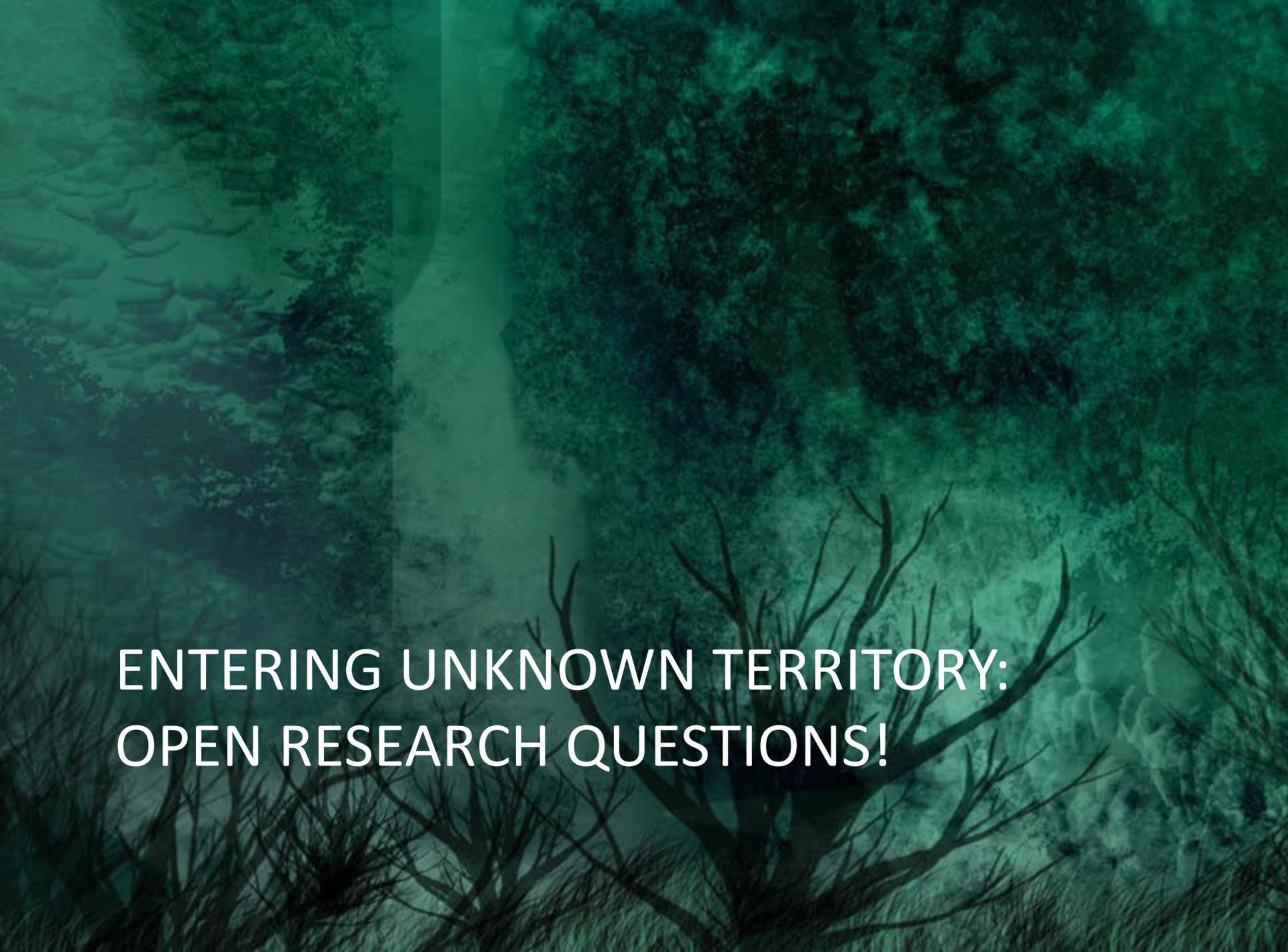
THE BEATLES 1-4 OF 15

Featured on BBC MUSIC SHOWCASE

- RINGO STARR ON THE BEATLES...
- RINGO STARR ON PLAYING WITH THE...
- THE BEATLES AND BRIAN...
- BRIAN EPSTEIN FINDS THE...



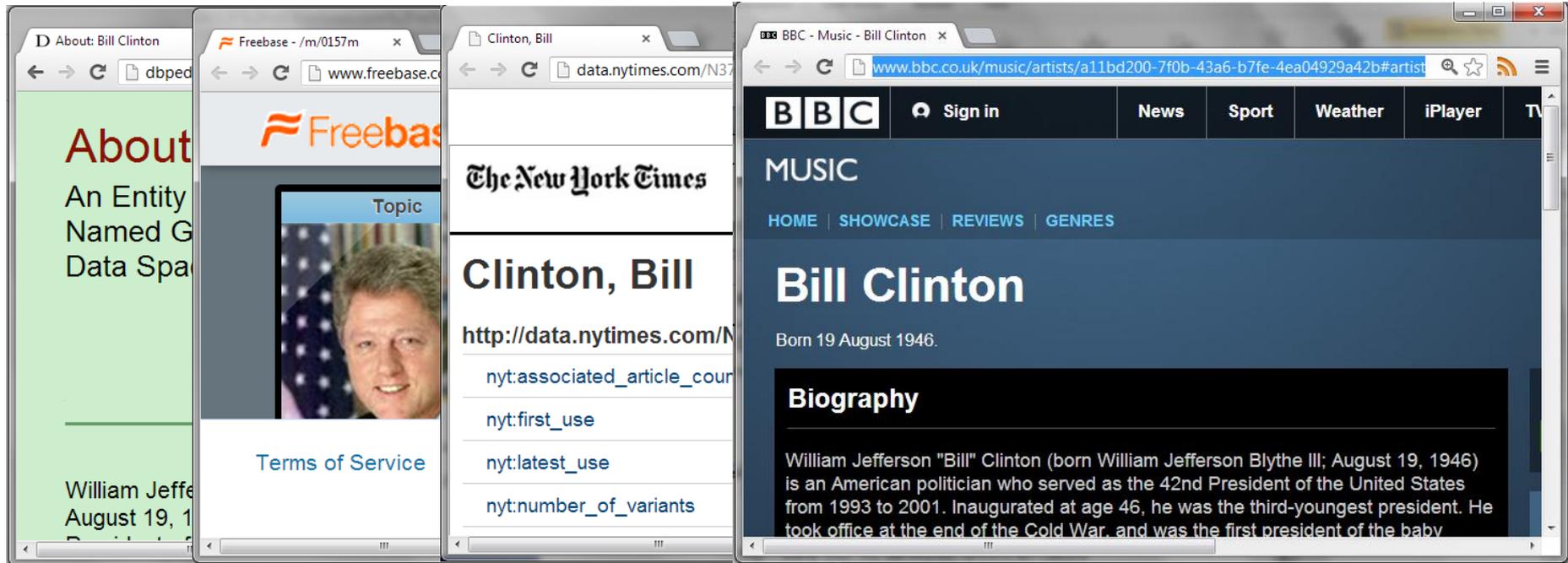
**ONLY PROBLEM IS FEW PEOPLE ARE
USING THE DATA ...**

A dark, teal-tinted background image of a dense forest. A path or road leads from the foreground into the distance, disappearing into the trees. The overall mood is mysterious and unknown.

ENTERING UNKNOWN TERRITORY:
OPEN RESEARCH QUESTIONS!

CATCH NUMBER 1: INTEGRATION

Need for Integration



http://dbpedia.org/resource/Bill_Clinton

http://rdf.freebase.com/ns/en.bill_clinton

http://data.nytimes.com/clinton_bill_per

<http://www.bbc.co.uk/music/artists/...>

How could OWL help here?

`owl:sameAs`

A (Hypothetical) Integration Example



```
SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED ...
WHERE {
    ?p1 ex:cites ?p2 .
    GRAPH :dblp { ?p1 a :Paper . ?p2 a :Paper }
    ?p1 ex:writtenBy ?a . ?a ex:basedIn wiki:Chile .
    NOT EXISTS {
        ?p1 ex:writtenBy ?b . ?p2 ex:writtenBy ?b .
    }
} GROUP BY ?a ORDER BY DESC(?c)
```

What is this query asking?

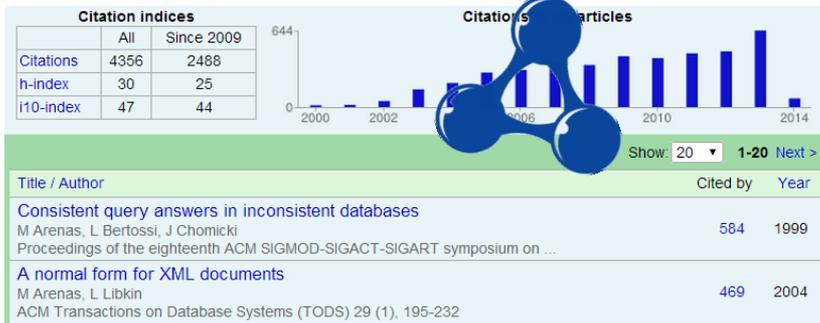
Find the top cited authors based in Chile, only including papers from DBLP, excluding self-citations from the count.

A (Hypothetical) Integration Example



Marcelo Arenas

Professor of Computer Science, PUC Chile
 Database theory - applications of logic to computer science - semantic Web
 Verified email at ing.puc.cl
[Homepage](#)



Marcelo Arenas

List of publications from the [DBLP Bibliography Server](#) - [FAQ](#)
 Other views: [by type](#) - [by year \(modern\)](#) - [classic-C](#) [Facets and more with CompleteSearch](#)

Ask others: [ACM DL](#) [Guide](#) - [CSB](#) - [MetaPress](#) - [Google](#) - [Bing](#) - [Yahoo](#)

Year	Publications
2013	<p>j30 Marcelo Arenas, Pablo Barceló, Ronald Fagin, Leonid Libkin: Solutions and query rewriting in data exchange. <i>Inf. Comput.</i> 228: 28-61 (2013)</p> <p>j29 Marcelo Arenas, Jorge Pérez, Juan L. Reutter: Data exchange beyond complete data. <i>J. ACM</i> 60(4): 28 (2013)</p>

Refine by AUTHOR

- Marcelo Arenas (101)
- Jorge Pérez (25)
- Leonid Libkin (22)
- Pablo Barceló (15)
- [\[top 4\]](#) [\[top 50\]](#) [\[all 60\]](#)

Refine by VENUE

- PODS (10)
- CoRR (7)
- Encyclopedia of Database Systems (6)
- Description Logics (4)
- [\[top 4\]](#) [\[top 50\]](#) [\[all 52\]](#)

Academic > Authors > Marcelo Arenas

Co-authors (93)

- Leonid Libkin
- Leopoldo Bertossi
- Claudio Gutierrez (Claudio Gutiérrez)
- Pablo Barcelo (Pablo Barceló)
- Jan Chomicki

Publications: 102 | Citations: 2074

Fields: Databases, Algorithms & Theory, Collaborative

Collaborated with 93 co-authors from 1994

1742 (1994-2014)

[Share this on](#) [f](#) [t](#) [200](#)

WIKIPEDIA
The Free Encyclopedia

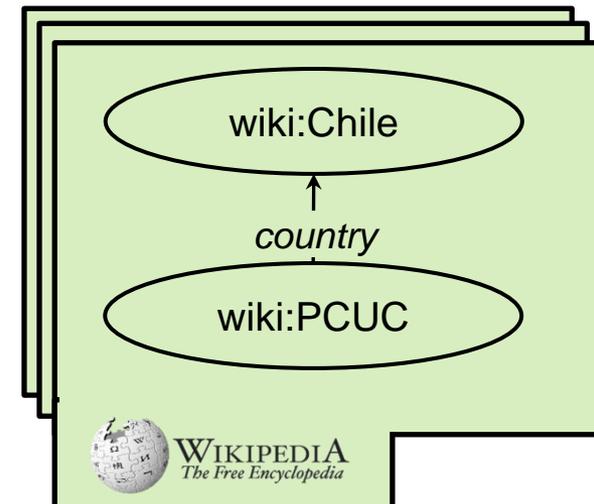
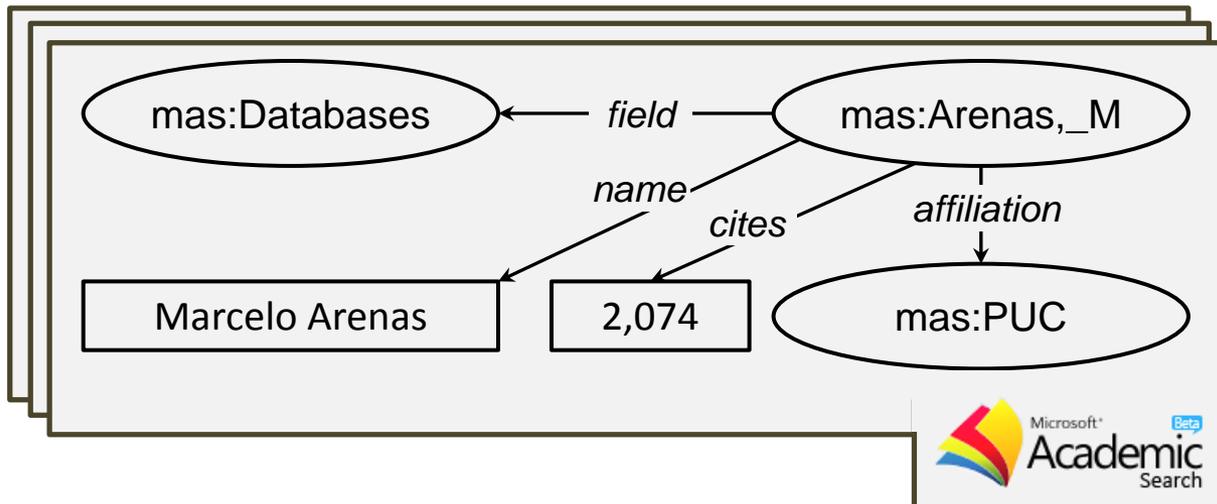
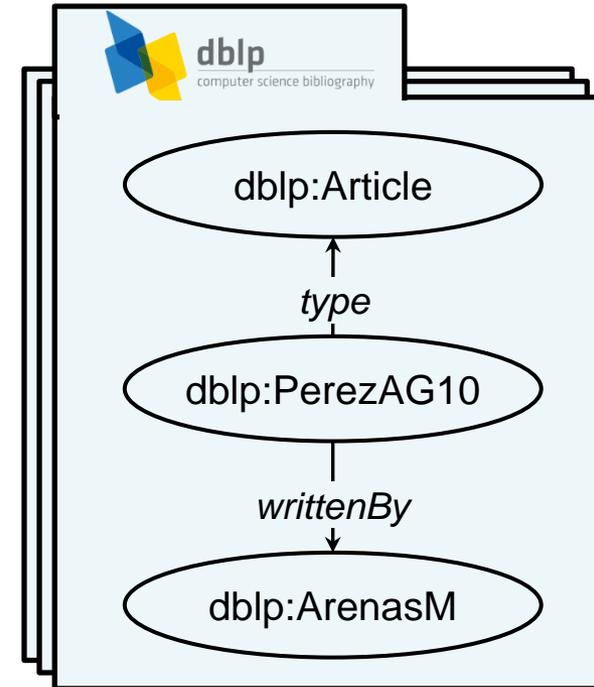
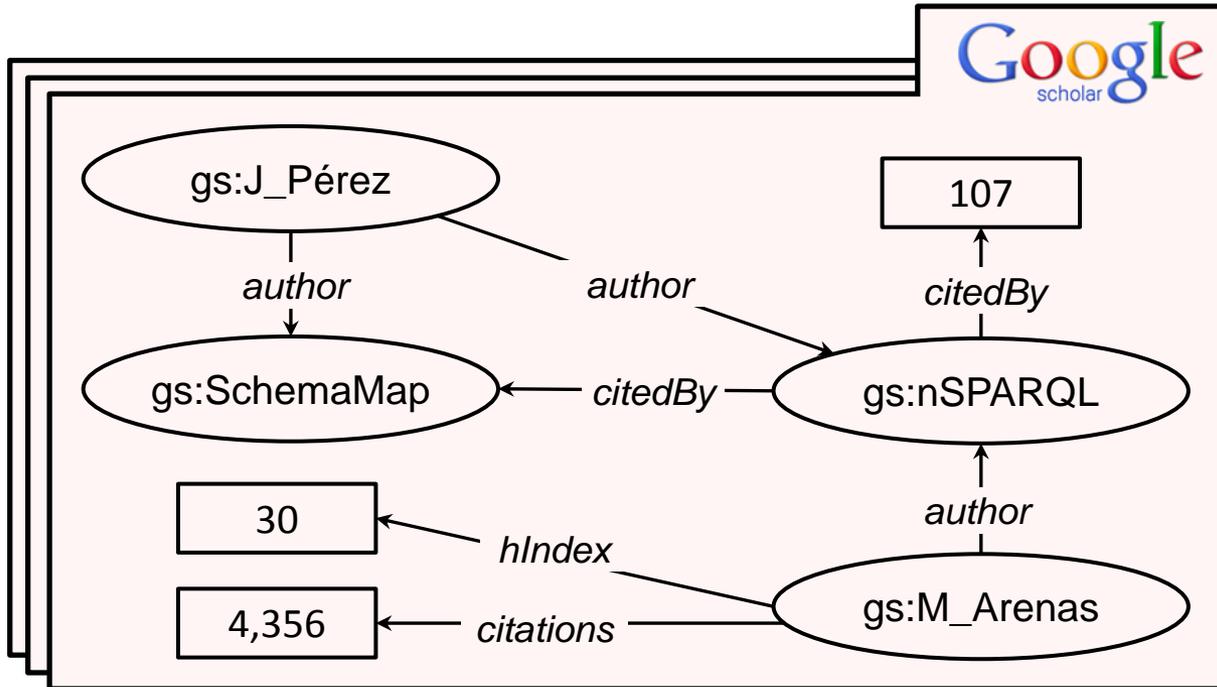
Article | Talk | Read |

Pontifical Catholic University of Chile

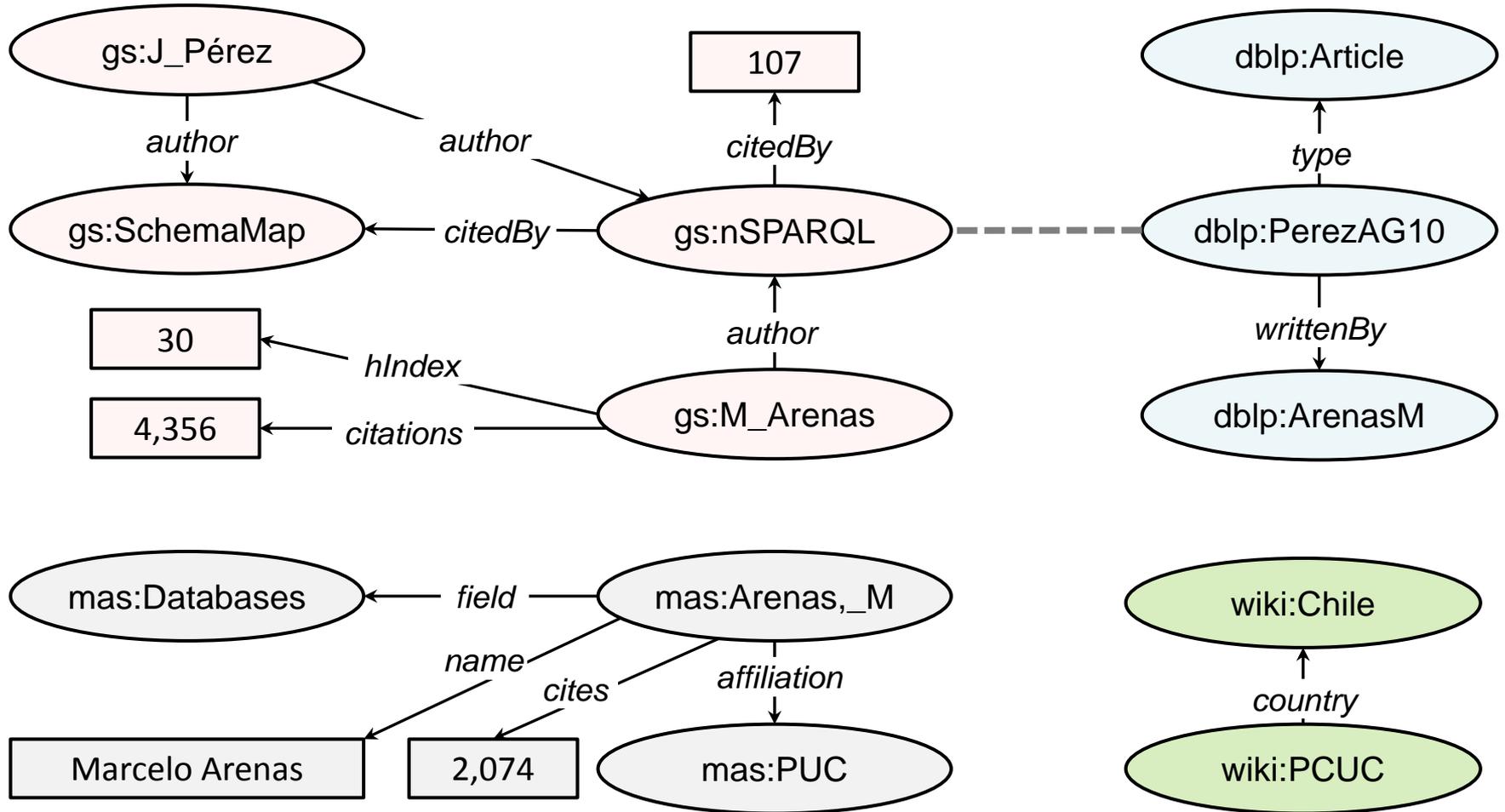
From Wikipedia, the free encyclopedia Coordinates: 33.4411°S 70.6408°W﻿ / ﻿33.4411°S 70.6408°W﻿ / -33.4411; -70.6408

The **Pontifical Catholic University of Chile (UC or PUC)** (Spanish: *Pontificia Universidad Católica de Chile*) is one of the six **Catholic Universities** existing in the Chilean university system and one of the two **Pontifical Universities** in the country, along with the **Pontifical Catholic University of Valparaiso**. It is also one of Chile's oldest

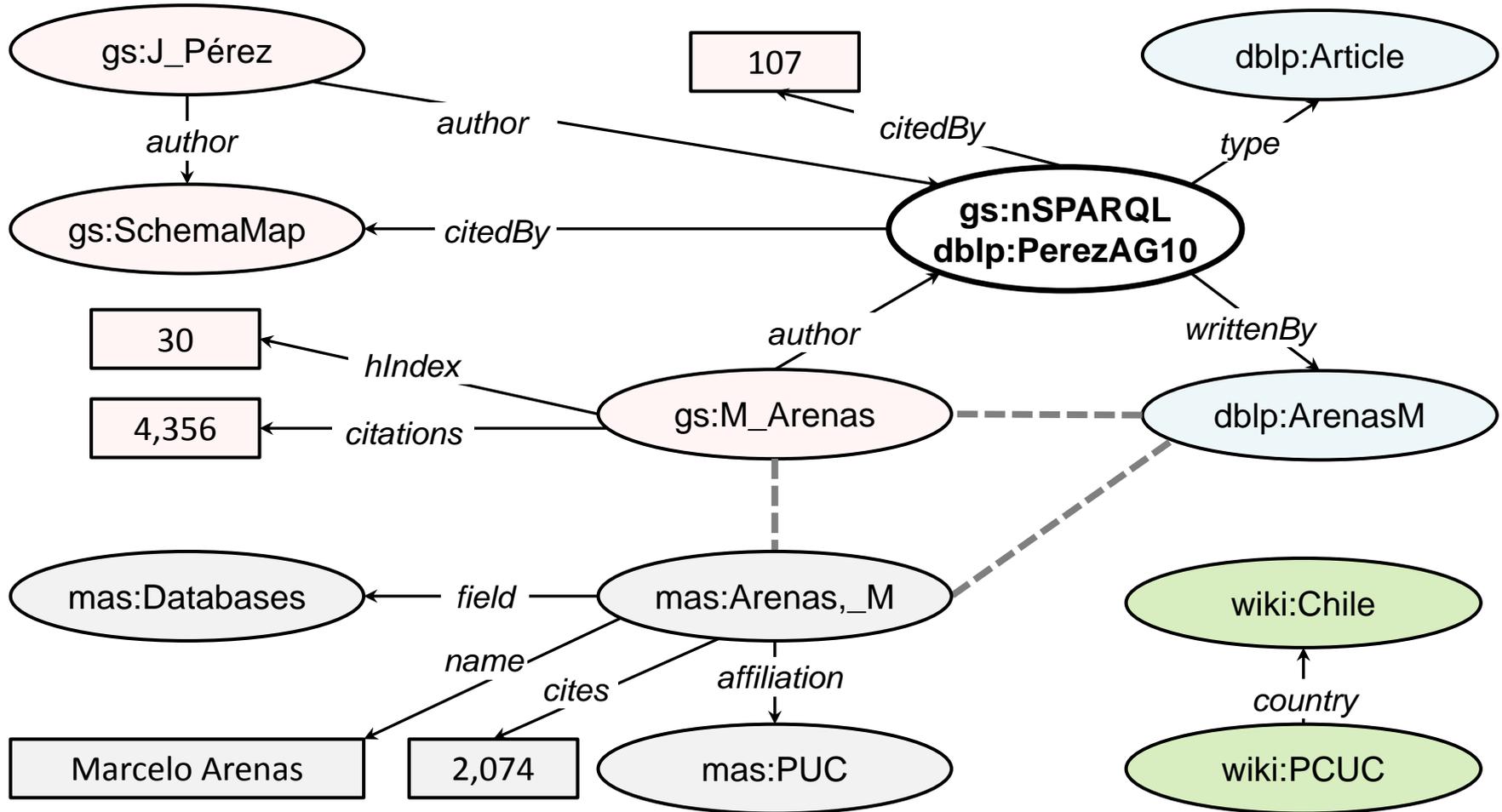
A (Hypothetical) Integration Example



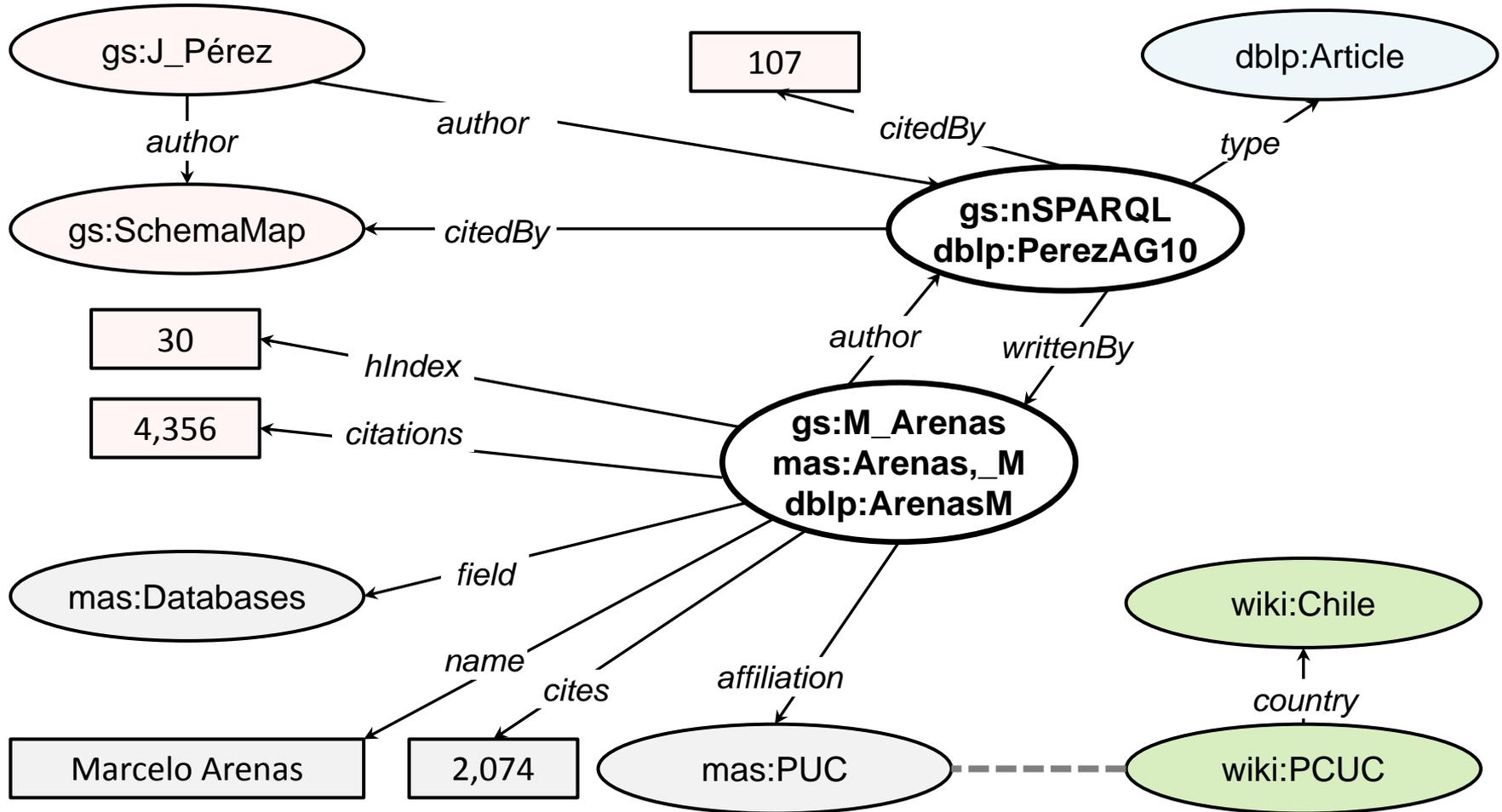
A (Hypothetical) Integration Example



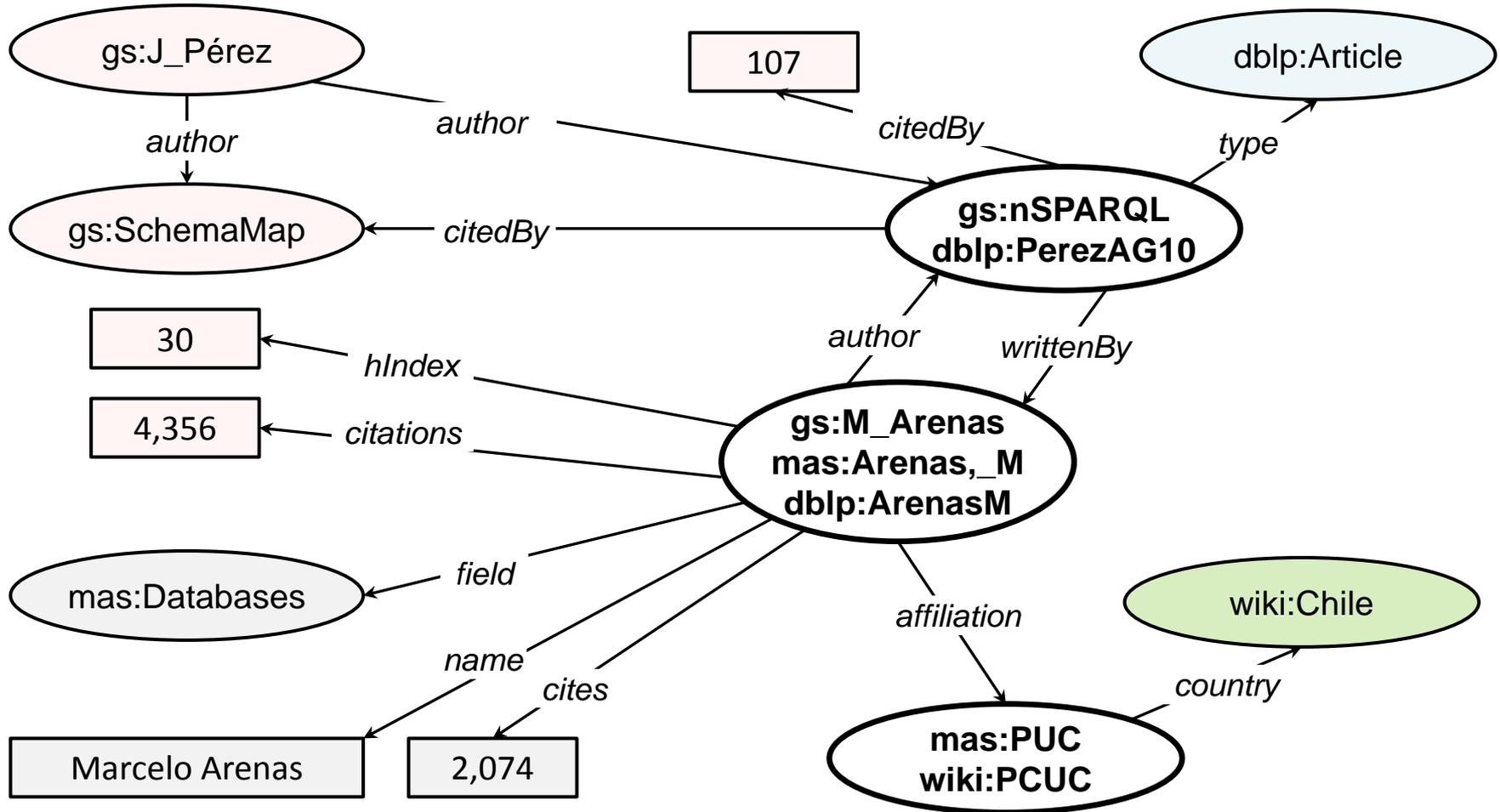
A (Hypothetical) Integration Example



Semantic Web: Tackling Heterogeneity

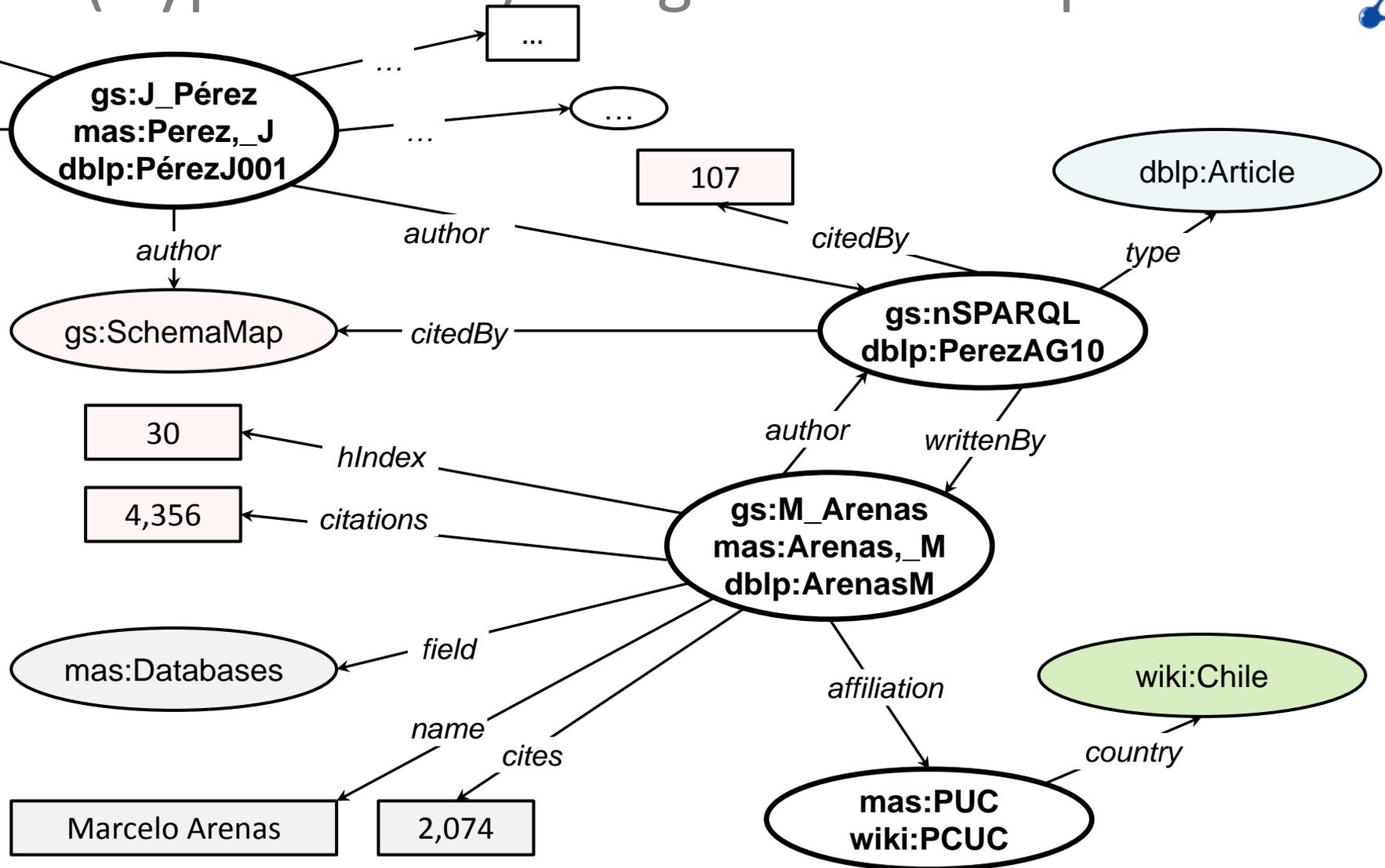


A (Hypothetical) Integration Example



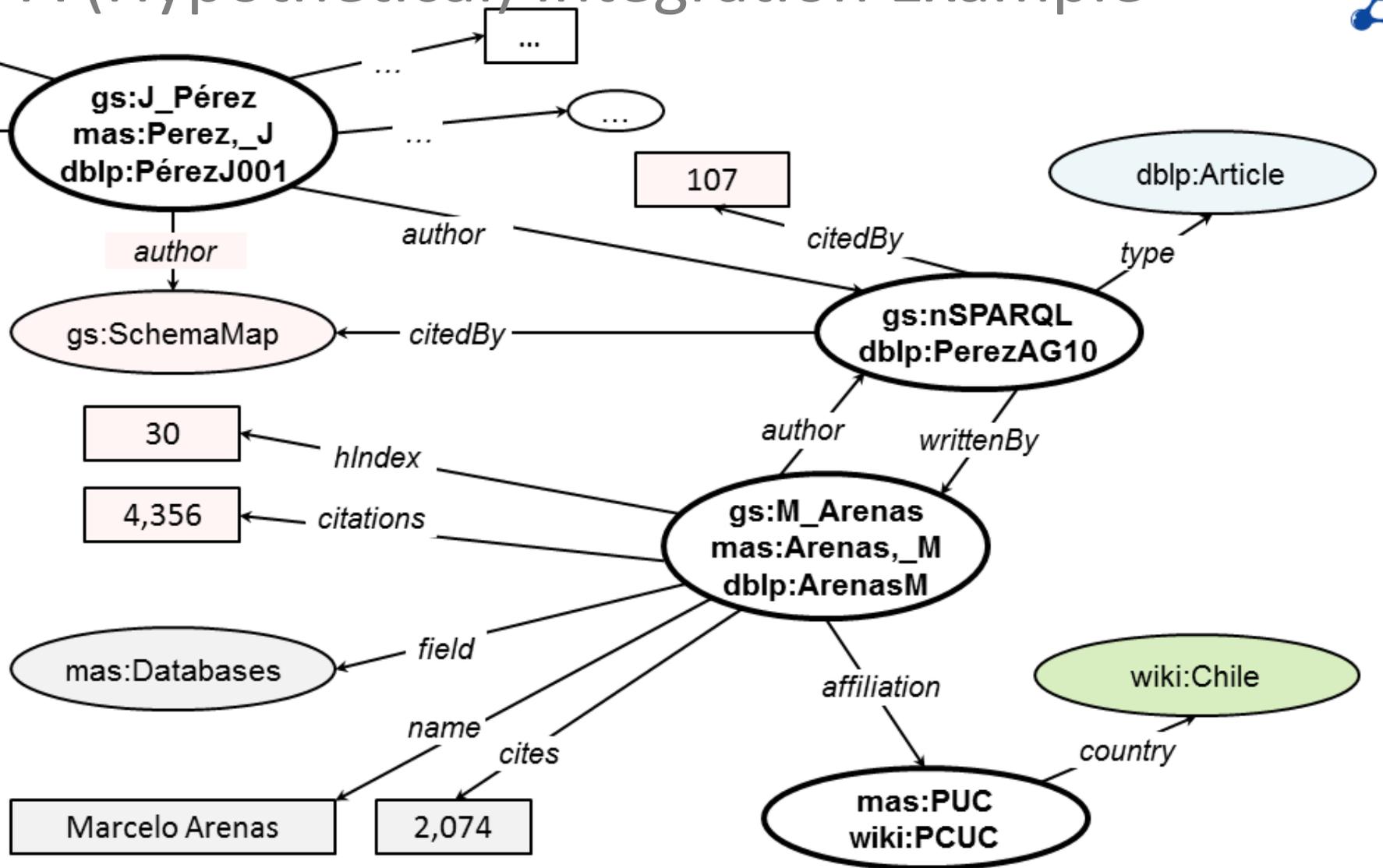


A (Hypothetical) Integration Example





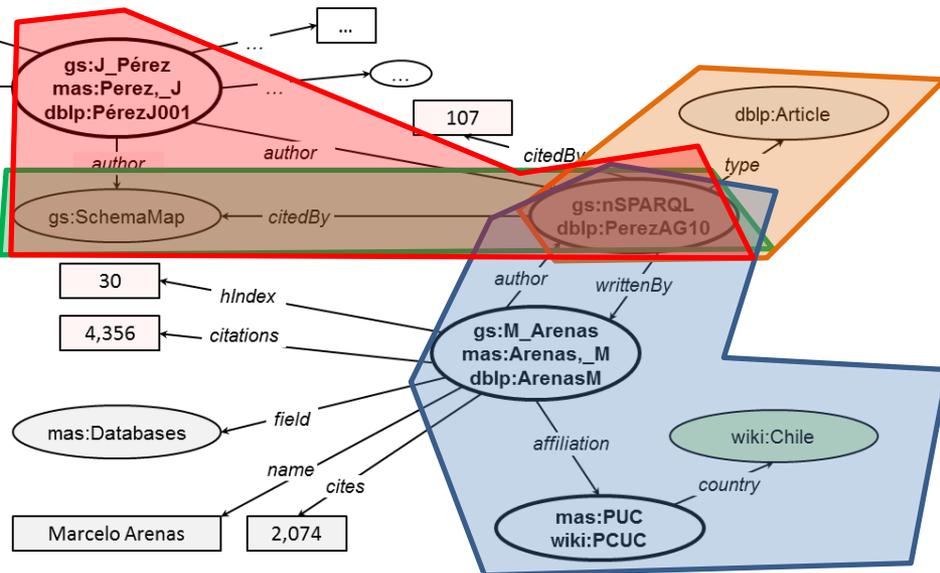
A (Hypothetical) Integration Example





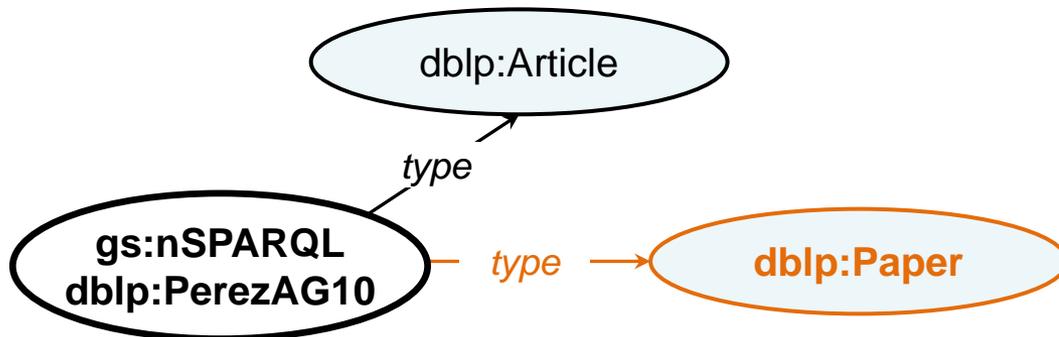
A (Hypothetical) Integration Example

```
SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED ...
WHERE {
    ?p1 ex:citedBy ?p2 .
    GRAPH :dblp { ?p1 a :Paper . ?p2 a :Paper }
    ?p1 ex:writtenBy ?a . ?a ex:basedIn wiki:Chile .
    NOT EXISTS {
        ?p1 ex:writtenBy ?b . ?p2 ex:writtenBy ?b .
    }
} GROUP BY ?a ORDER BY DESC(?c)
```



A (Hypothetical) Integration Example

```
SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED ...
WHERE {
  ?p1 ex:citedBy ?p2 .
  GRAPH :dblp { ?p1 a :Paper . ?p2 a :Paper }
  ?p1 ex:writtenBy ?a . ?a ex:basedIn wiki:Chile .
  NOT EXISTS {
    ?p1 ex:writtenBy ?b . ?p2 ex:writtenBy ?b .
  }
} GROUP BY ?a ORDER BY DESC(?c)
```

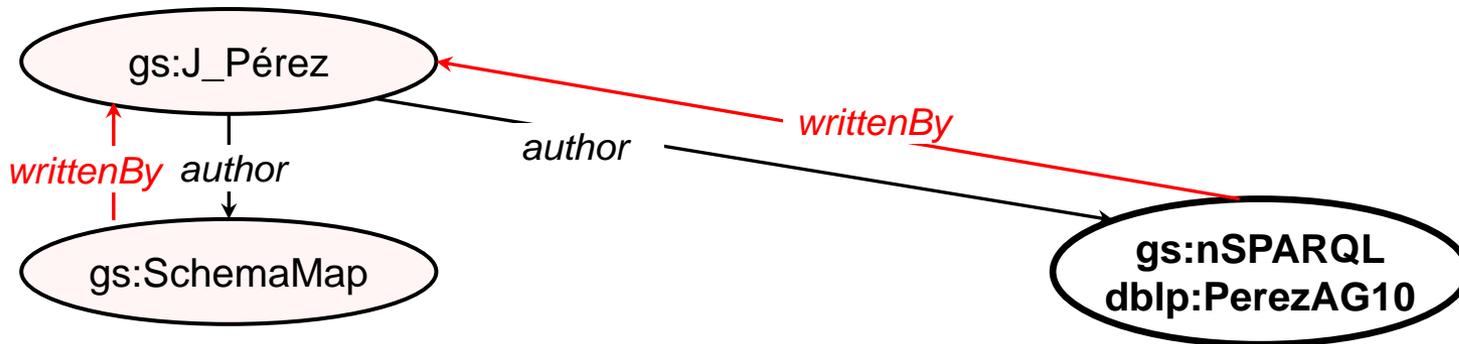


Article \sqsubseteq Paper

dblp:Article rdfs:subClassOf dblp:Paper .

A (Hypothetical) Integration Example

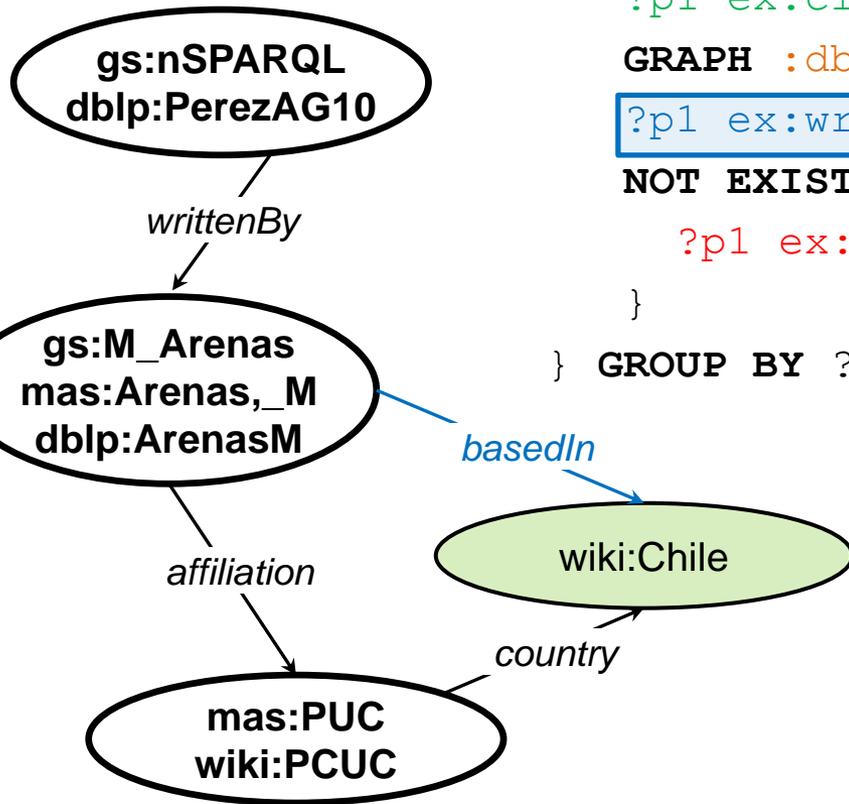
```
SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED ...
WHERE {
  ?p1 ex:citedBy ?p2 .
  GRAPH :dblp { ?p1 a :Paper . ?p2 a :Paper }
  ?p1 ex:writtenBy ?a . ?a ex:basedIn wiki:Chile .
  NOT EXISTS {
    ?p1 ex:writtenBy ?b . ?p2 ex:writtenBy ?b .
  }
} GROUP BY ?a ORDER BY DESC(?c)
```



```
author ≡ writtenBy -
ex:author owl:inverseOf ex:writtenBy .
```

A (Hypothetical) Integration Example

```
SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED ...
WHERE {
  ?p1 ex:citedBy ?p2 .
  GRAPH :dblp { ?p1 a :Paper . ?p2 a :Paper }
  ?p1 ex:writtenBy ?a . ?a ex:basedIn wiki:Chile .
  NOT EXISTS {
    ?p1 ex:writtenBy ?b . ?p2 ex:writtenBy ?b .
  }
} GROUP BY ?a ORDER BY DESC(?c)
```



affiliation o country \sqsubseteq basedIn

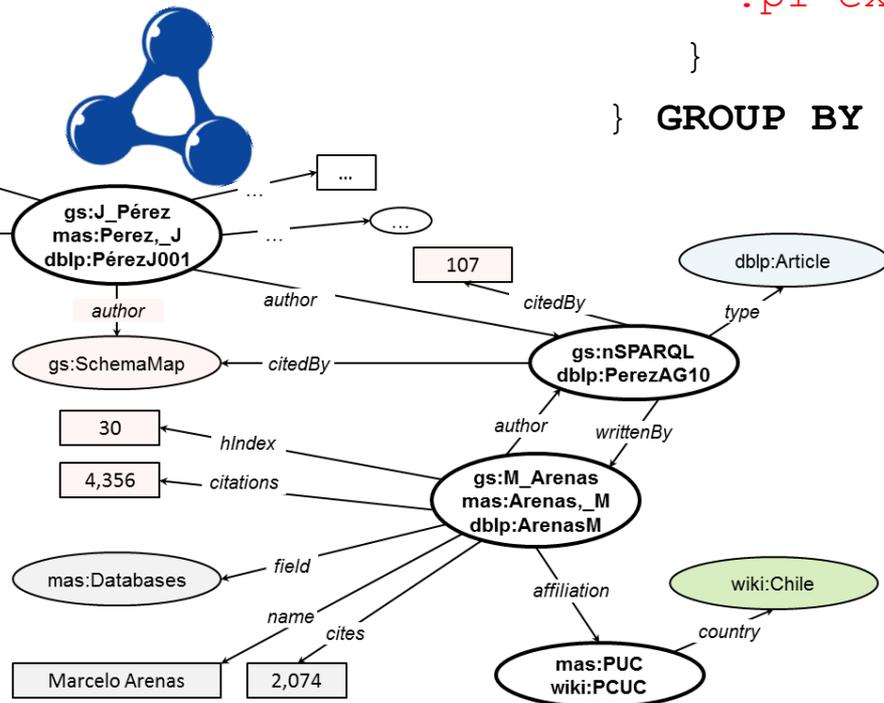
ex:basedIn owl:propertyChainAxiom(ex:affiliation ex:country)

A (Hypothetical) Integration Example



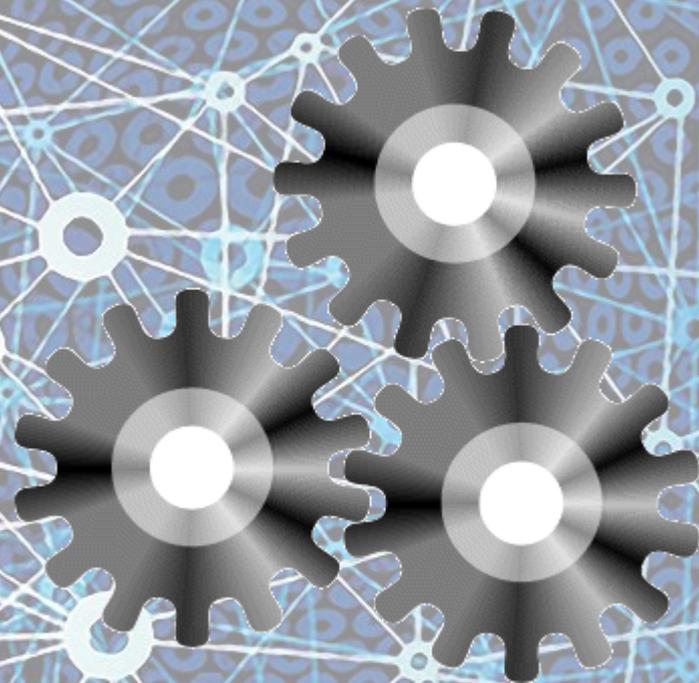
```

SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED ...
WHERE {
    ?p1 ex:citedBy ?p2 .
    GRAPH :dblp { ?p1 a :Paper ; ?p2 a :Paper }
    ?p1 ex:writtenBy ?a .
    NOT EXISTS {
        ?p1 ex:writtenBy ?k .
    }
} GROUP BY ?a
    
```



Article \sqsubseteq Paper
 author \equiv writtenBy
 affiliation o country \sqsubseteq basedIn

Not clear yet how to do reasoning on the Web!



Side Note: Fact or Fiction?



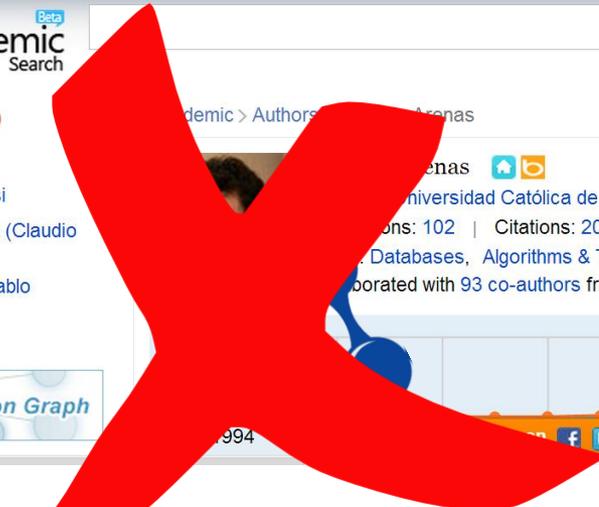
Marcelo Arenas
Professor of Computer Science, PUC Chile
Database theory, Logic, Knowledge Representation - semantic Web
Verified email at puc.cl
[Homepage](#)

Citation indices

	All	Since 2009
Citations	4356	2488
h-index	30	25
i10-index	47	44

Publications

Title / Author	Year
Consistent query answers in provenance databases M Arenas, L Bertossi, J Chomicki, J Gehrke, J Gottman, J Han, J Li, J Liu, J Minker, J Neumann, J Reagin, J Riedinger, J Sussman, J Wang, J Wu, J Yang, J Zhou, J Zhu, J Zou	584 1999
A normal form for XML documents M Arenas, L Libkin	469 2004

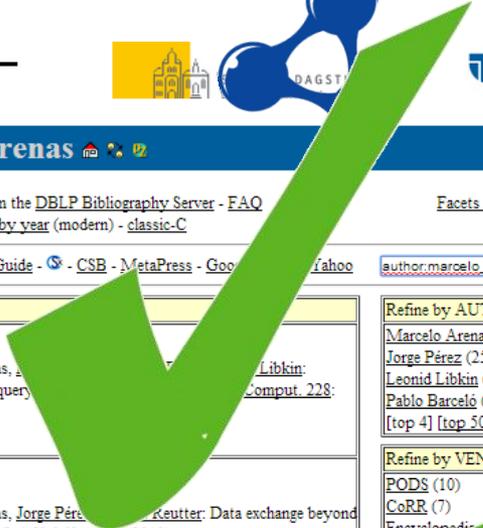


Microsoft Academic Search

Co-authors (93)

- Leonid Libkin
- Leopoldo Bertossi
- Claudio Gutierrez (Claudio Gutiérrez)
- Pablo Barcelo (Pablo Barceló)
- Jan Chomicki

Citation Graph



dblp.uni-trier.de
Computer Science Bibliography

Universität Trier

Marcelo Arenas

List of publications from the [DBLP Bibliography Server](#) - [FAQ](#)
Other views: [by type](#) - [by year \(modern\)](#) - [classic-C](#)

Ask others: [ACM DL](#) - [Guide](#) - [CSB](#) - [MetaPress](#) - [Google Scholar](#) - [Yahoo](#)

author:marcelo_arenas

Refine by AUTHOR

- Marcelo Arenas (101)
- Jorge Pérez (25)
- Leonid Libkin (22)
- Pablo Barceló (15)
- [top 4] [top 50] [all 60]

Refine by VENUE

- PODS (10)
- CoRR (7)
- Encyclopedia of Database Systems (6)
- Descriptive Logics (4)
- [top 50] [all 52]



WIKIPEDIA
The Free Encyclopedia

Pontifical Catholic University of Chile

Coordinates: 33.4411°S 70.6408°W

Pontificia Universidad Católica de Chile

Pontificia Universidad Católica de Chile

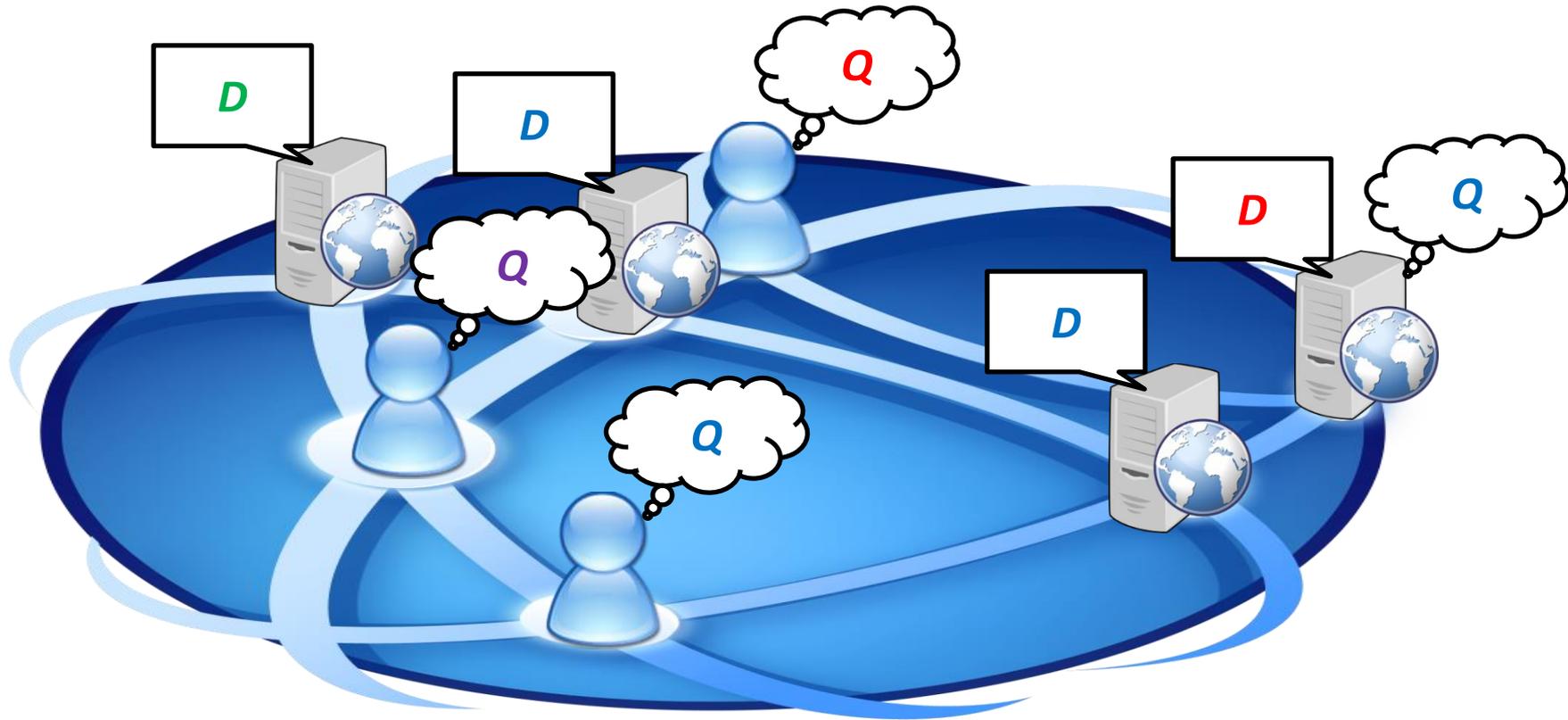
BIG CATCH NUMBER 2: DATA ACCESS

Access Methods

- Client has a request/query Q
- Server has a dataset D
- Client issues Q to server
- Server computes and returns response $Q(D)$



Access Methods



- Multiple clients / multiple servers (blurred)
- Remote, decentralised, uncoordinated
- Web scale

Linked Data Access Methods

1. Dereferencing:

- Look up a URI, get an RDF document

2. Dumps:

- Get all data in an archive

3. SPARQL Queries:

- Send a query, get the answers

Dereferencing (what is it?)

Q = “<http://dbpedia.org/resource/Columbia>”

Q(D) =

```
▼<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:dbpprop="http://dbpedia.org/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:ns5="http://dbpedia.org/ontology/PopulatedPlace/" xmlns:dbpedia-
  owl="http://dbpedia.org/ontology/" xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" xmlns:ns9="http://www.w3
  xmlns:grs="http://www.georss.org/georss/">
  ▼<rdf:Description rdf:about="http://dbpedia.org/resource/Afro-Latin_American
    <dbpedia-owl:populationPlace rdf:resource="http://dbpedia.org/resource/Col
    </rdf:Description>
  ▼<rdf:Description rdf:about="http://dbpedia.org/resource/Alb%C3%A1n">
    <dbpedia-owl:country rdf:resource="http://dbpedia.org/resource/Colombia"/>
    </rdf:Description>
  ▼<rdf:Description rdf:about="http://dbpedia.org/resource/Andr%C3%A9s_Pastran
    <dbpedia-owl:birthPlace rdf:resource="http://dbpedia.org/resource/Colombia
    <dbpedia-owl:country rdf:resource="http://dbpedia.org/resource/Colombia"/>
    <dbpprop:birthPlace rdf:resource="http://dbpedia.org/resource/Colombia"/>
    <dbpprop:placeOfBirth rdf:resource="http://dbpedia.org/resource/Colombia"/
    </rdf:Description>
  ▼<rdf:Description rdf:about="http://dbpedia.org/resource/Antanas_Mockus">
    <dbpedia-owl:birthPlace rdf:resource="http://dbpedia.org/resource/Colombia
    <dbpprop:birthPlace rdf:resource="http://dbpedia.org/resource/Colombia"/>
    </rdf:Description>
```

Dereferencing (what's wrong with it?)

- **Responses vary from server to server**
 - local triples where URI is **subject** (83%) vs.
 - local triples where URI is **subject** or **object** (55%)

WELL-DEFINED: For a given Q and D , clients and servers agree on what $Q(D)$ should be.



Dereferencing (what's wrong with it?)

- **Very coarse:**

- Give me all capitals of South American countries.

- Dereference documents for all country URIs
- See which ones are in South America, throw away rest
- Throw away triples other than capitals

```
User-agent: *  
Crawl-delay: 10
```

GRANULAR: The language for Q allows the client to be specific so as to avoid wasting bandwidth

WELL-DEFINED

GRANULAR

Dereferencing



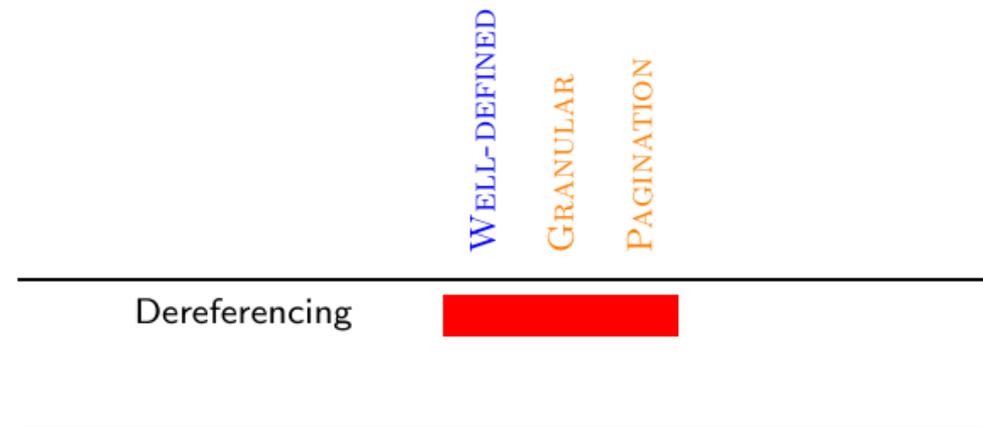
Dereferencing (what's wrong with it?)

- **No pagination:**

- Give me some information about Italy.

- Load document with 100,000 triples
- Throw away 99,900 triples the user won't read

PAGINATION: A large response $Q(D)$ can be split into chunks



Dumps (what are they?)



[DBpedia Blog](#) | [Get Involved](#) | [Get Help](#)

- [About / News](#)
- [Applications](#)
- [Use Cases](#)
- [Datasets](#)
- [Online Access](#)
- [DBpedia Live](#)
- [Downloads](#)
- [Interlinking](#)

DBpedia 3.9 Downloads

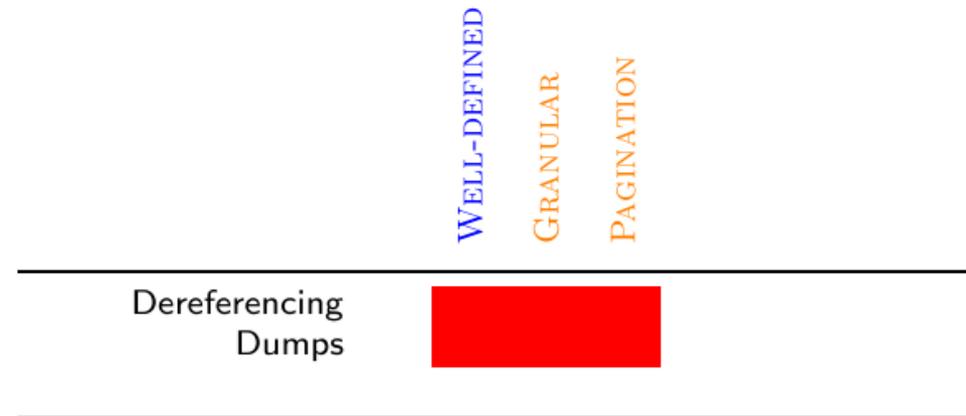
This page provides downloads of the DBpedia datasets. The datasets are licensed under the terms of the [Creative Commons Attribution](#) and the [GNU Free Documentation License](#).  The datasets are available as N-Triples and N-Quads, where the N-Quads version contains more information for each statement. All files are [bzip2](#) ¹ packed.

In addition to the RDF version of the data, we also provide a tabular version of the core DBpedia data sets as CSV and JSON files. See [DBpedia](#)

Older Versions: [DBpedia 3.8](#), [DBpedia 3.7](#), [DBpedia 3.6](#), [DBpedia 3.5](#), [DBpedia 3.4](#), [DBpedia 3.3](#), [DBpedia 3.2](#), [DBpedia 3.1](#), [DBpedia 2.0](#)

Dumps (what's wrong with them?)

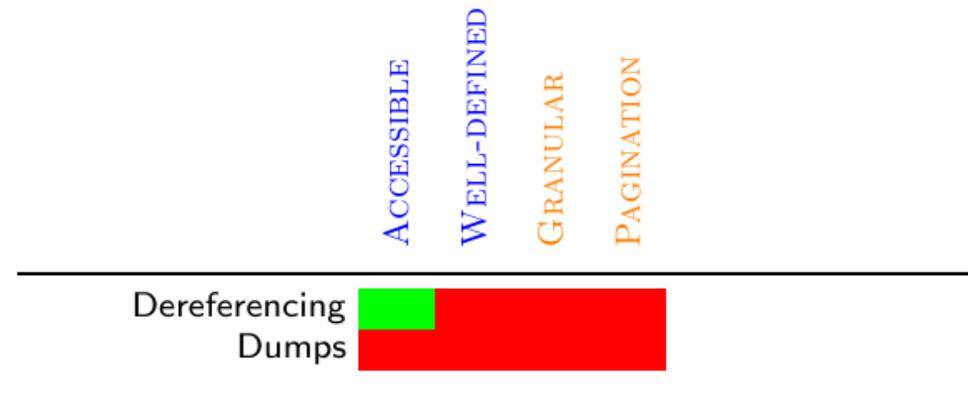
- 15× compression for RDF achievable
- But **same weaknesses as for deref. still apply**



Dumps (what's wrong with them?)

- 15× compression for RDF achievable
- But **same weaknesses as for deref. still apply**
- **Also, no standard access methods:**
 - Various compressions and formats
 - Linked through generic HTML

ACCESSIBLE: The protocol and formats are defined for automatic access by software agents



SPARQL (what is it?)

$Q =$

```
PREFIX dbo: <http://dbpedia.org/ontology/>
...

SELECT ?capital
WHERE {
  ?s a dbo:Country ; dbp:capital ?c ;
    dcterms:subject category:Countries_in_South_America .
  ?c rdfs:label ?capital . FILTER (lang(?capital)="en")
}
```

$Q(D) =$

capital
"Caracas"@en
"Buenos Aires"@en
"Asunción"@en
"Brasília"@en
"Georgetown, Guyana"@en
"Montevideo"@en
"Paramaribo"@en
"Bogotá"@en
"Lima"@en
"Quito"@en
"Santiago"@en

SPARQL (to the rescue?)

ACCESSIBLE

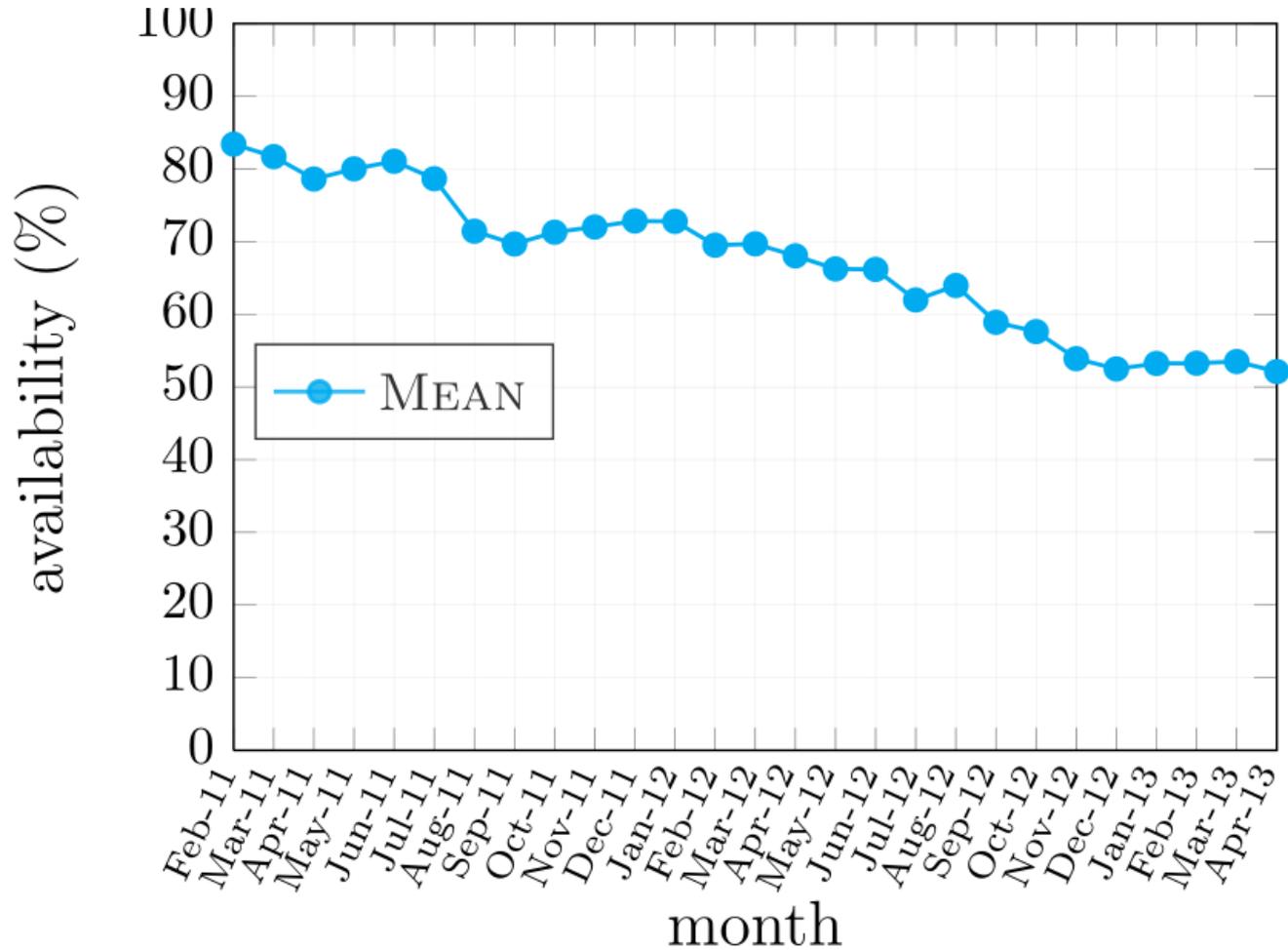
WELL-DEFINED

GRANULAR

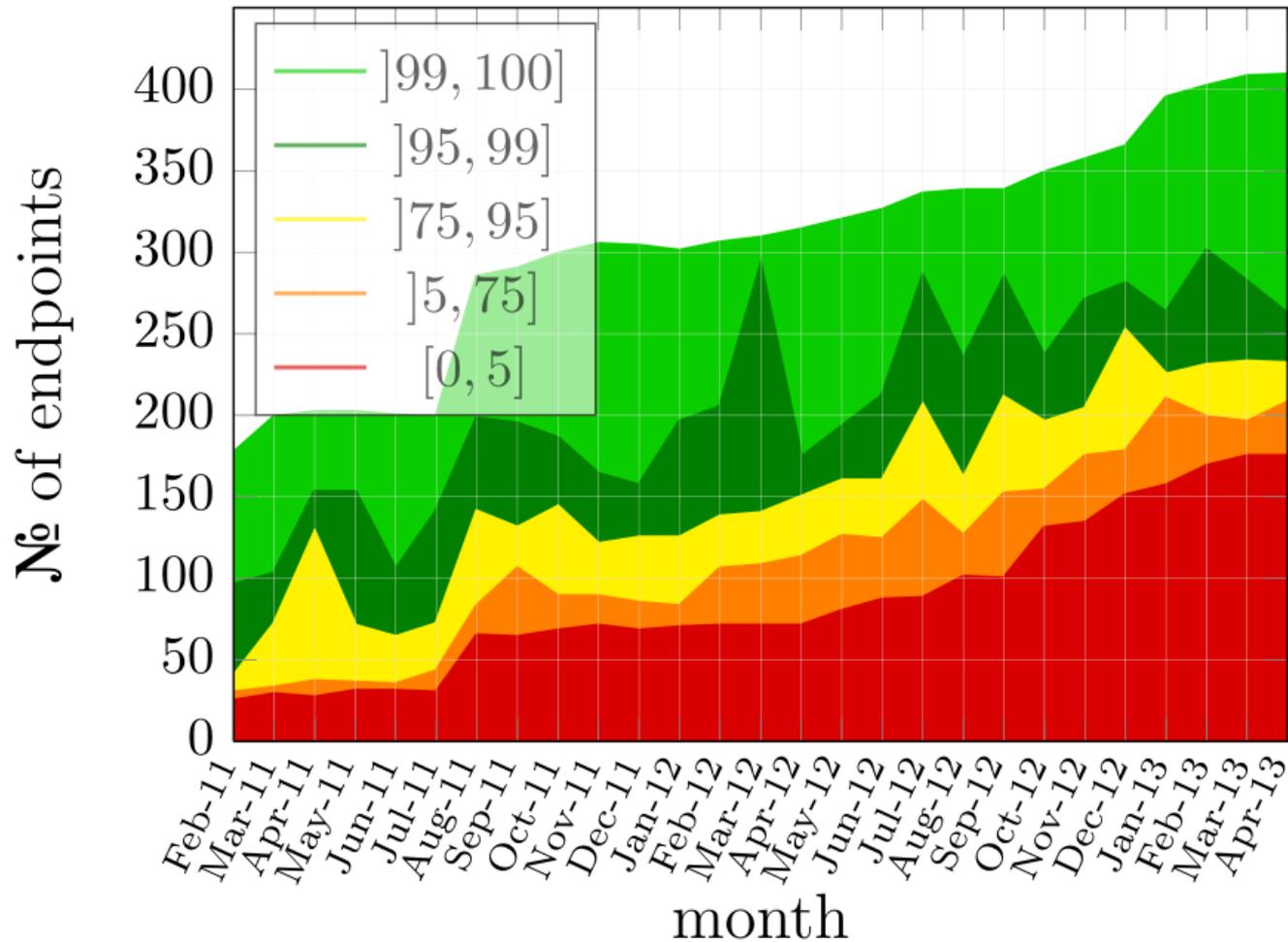
PAGINATION

Dereferencing		
Dumps		
SPARQL endpoints		

SPARQL (what's wrong with it?)

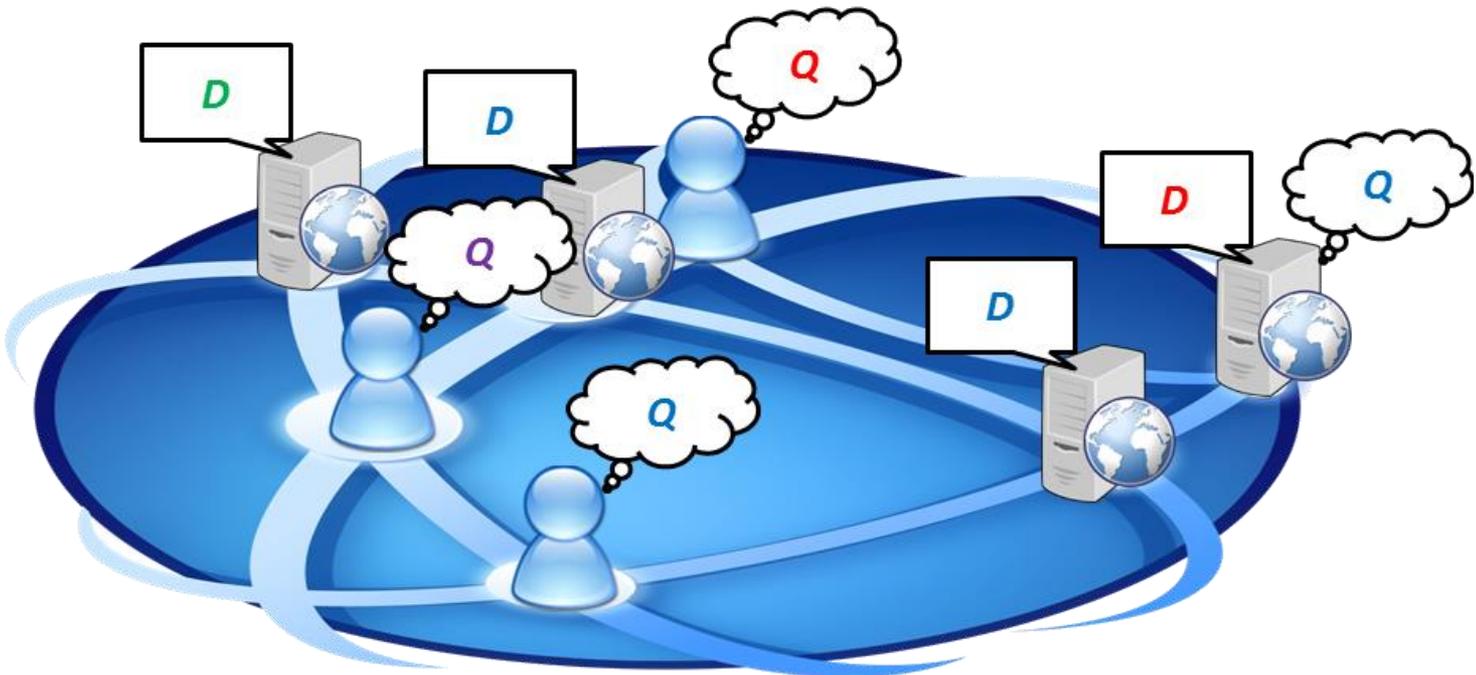


SPARQL (what's wrong with it?)



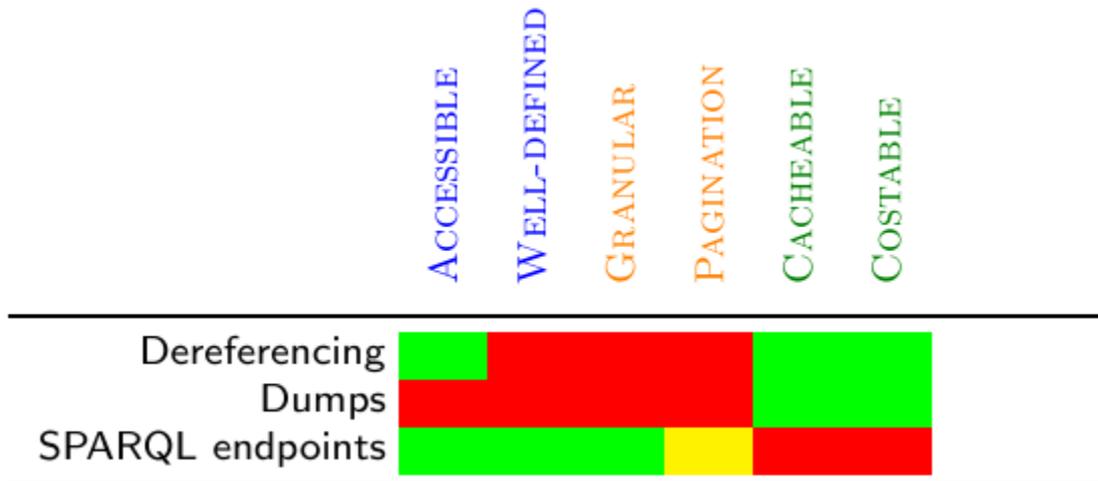
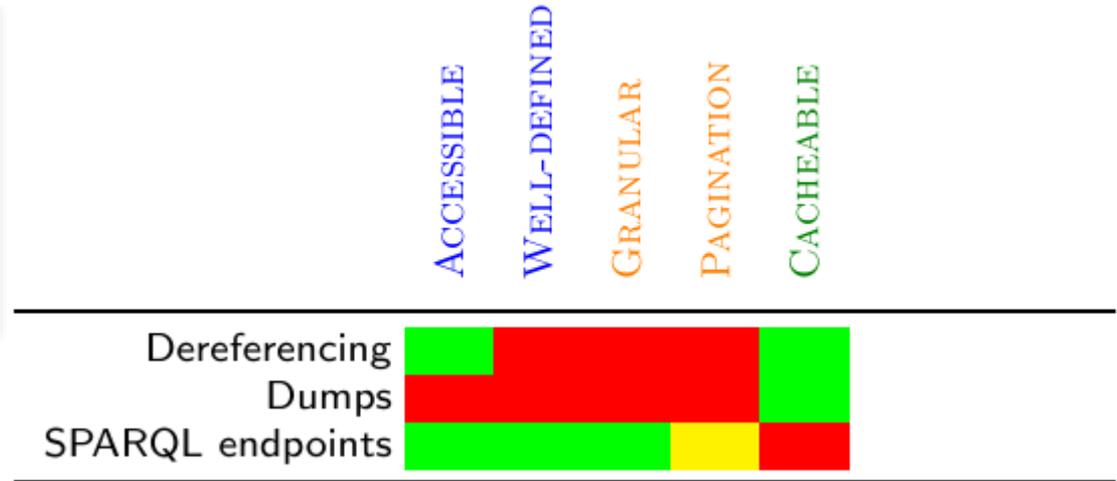
SPARQL (what's wrong with it?)

- ~~Simple~~ Protocol and **RDF Query Language**
- SPARQL evaluation: **PSPACE-complete**



SPARQL (what's wrong with it?)

CACHEABLE: Common requests can be cached and re-used. Queries can be anticipated.



COSTABLE: The cost of processing a query can be anticipated before actual processing.

SPARQL (what's wrong with it?)

<u>Nº of Results</u>	<u>Nº of Endpoints</u>
500	1
1,000	3
1,500	1
5,000	1
10,000	49
20,000	2
40,000	1
50,000	3
100,000	7
TOTAL:	68

```
SELECT * WHERE { ?s ?p ?o } LIMIT 100002
```

SPARQL (what's wrong with it?)

```
Virtuoso 42000 Error The estimated execution time 0  
(sec) exceeds the limit of 3000 (sec).
```

```
SPARQL query:
```

```
PREFIX dbo: <http://dbpedia.org/ontology/>
```

```
PREFIX dbp: <http://dbpedia.org/ontology/>
```

```
SELECT ?capital
```

```
WHERE {
```

```
  ?s a dbo:Country ;
```

```
    dbp:capital ?c ;
```

```
    dcterms:subject
```

```
category:Countries_in_South_America .
```

```
  ?c rdfs:label ?capital .
```

```
  FILTER (langMatches(?c,"en"))
```

```
}
```

SPARQL (what's wrong with it?)



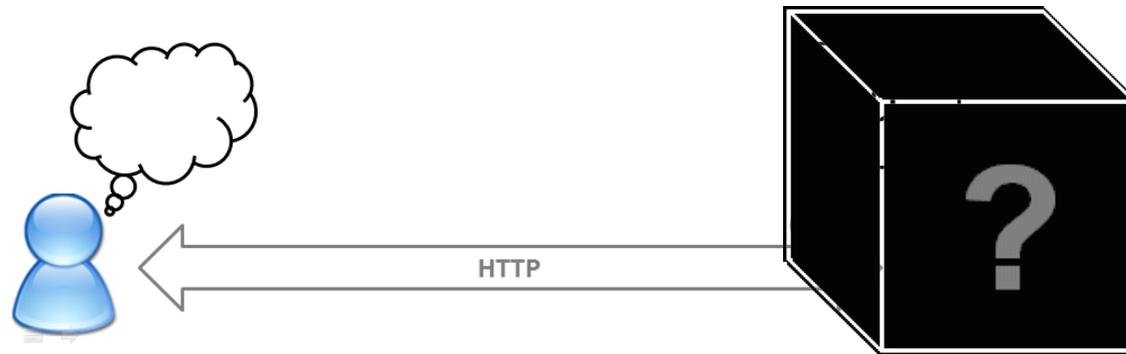
• **Simple Protocol** And RDF Query Language

- **Protocol always expects a perfect answer**
 - No support for partial results, timeouts, exception handling, pagination ...

ROBUST: Access method can gracefully handle error cases and provide Quality of Service

	ACCESSIBLE	WELL-DEFINED	GRANULAR	PAGINATION	CACHEABLE	COSTABLE	ROBUST
Dereferencing	Green	Red	Red	Red	Green	Green	Green
Dumps	Red	Red	Red	Red	Green	Green	Green
SPARQL endpoints	Green	Green	Green	Yellow	Red	Red	Red

SPARQL (what's wrong with it?)



- *D* is a **black-box** for the user

SPARQL (what's wrong with it?)

Q =

```
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT (COUNT(?c) as ?count)
WHERE {
  ?c a dbo:Country .
}
```

Q(D) =

count
7182

TRANSPARENT: The client can determine if a dataset D is relevant and the service sufficient.

	ACCESSIBLE	WELL-DEFINED	GRANULAR	PAGINATION	CACHEABLE	COSTABLE	ROBUST	TRANSPARENT
Dereferencing	Green	Red	Red	Red	Green	Green	Green	Green
Dumps	Red	Red	Red	Red	Green	Green	Green	Green
SPARQL endpoints	Green	Green	Yellow	Red	Red	Red	Red	Red

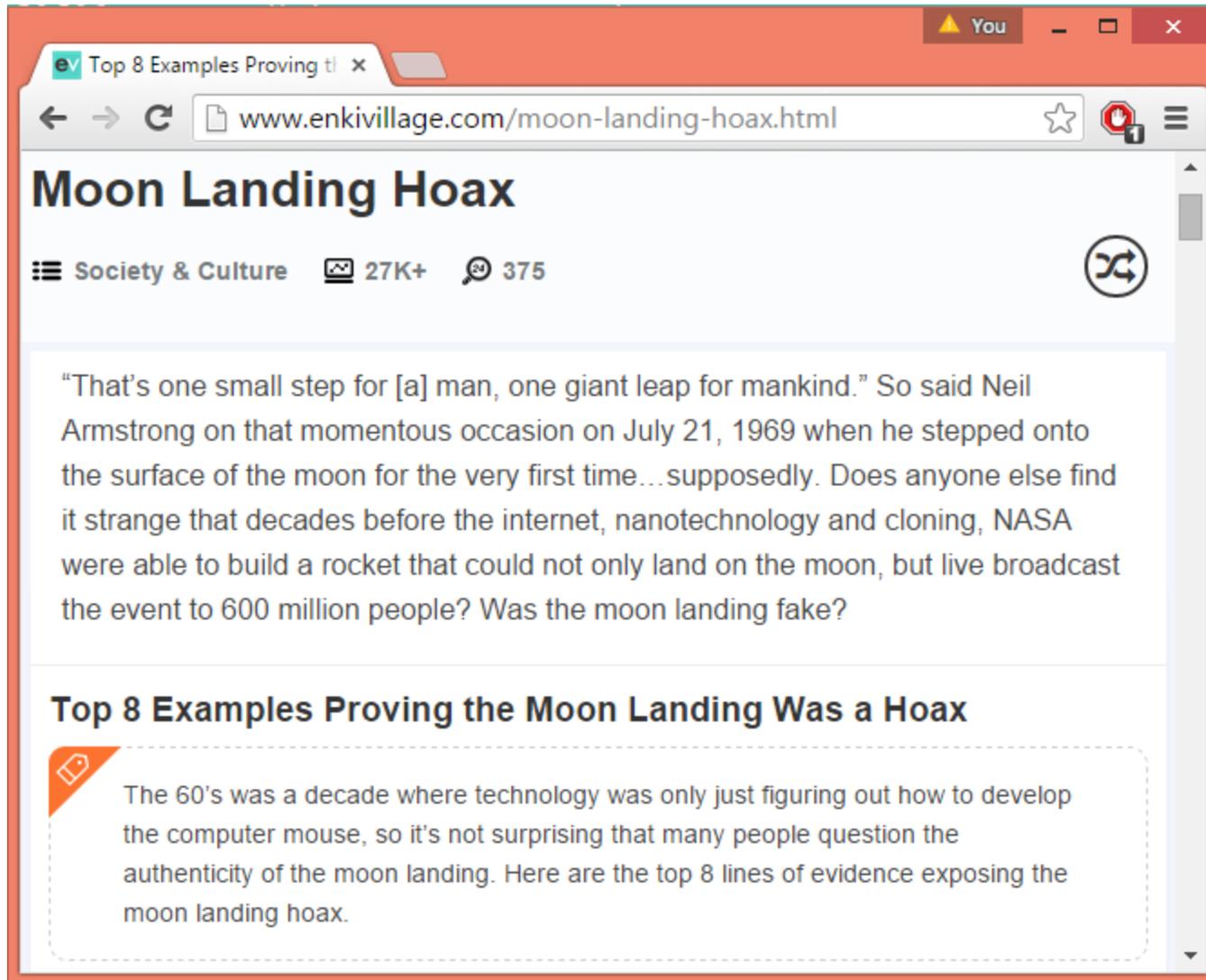
Problem Categories

1. Standardised
2. Bandwidth-efficient
3. Server-processing-efficient
4. Usable by client

	ACCESSIBLE	WELL-DEFINED	GRANULAR	PAGINATION	CACHEABLE	COSTABLE	ROBUST	TRANSPARENT
Dereferencing	Green	Red	Red	Red	Green	Green	Green	Green
Dumps	Red	Red	Red	Red	Green	Green	Green	Green
SPARQL endpoints	Green	Green	Green	Yellow	Red	Red	Red	Red

BIG CATCH NUMBER 3: DATA QUALITY

Can't trust everything you read on the Web

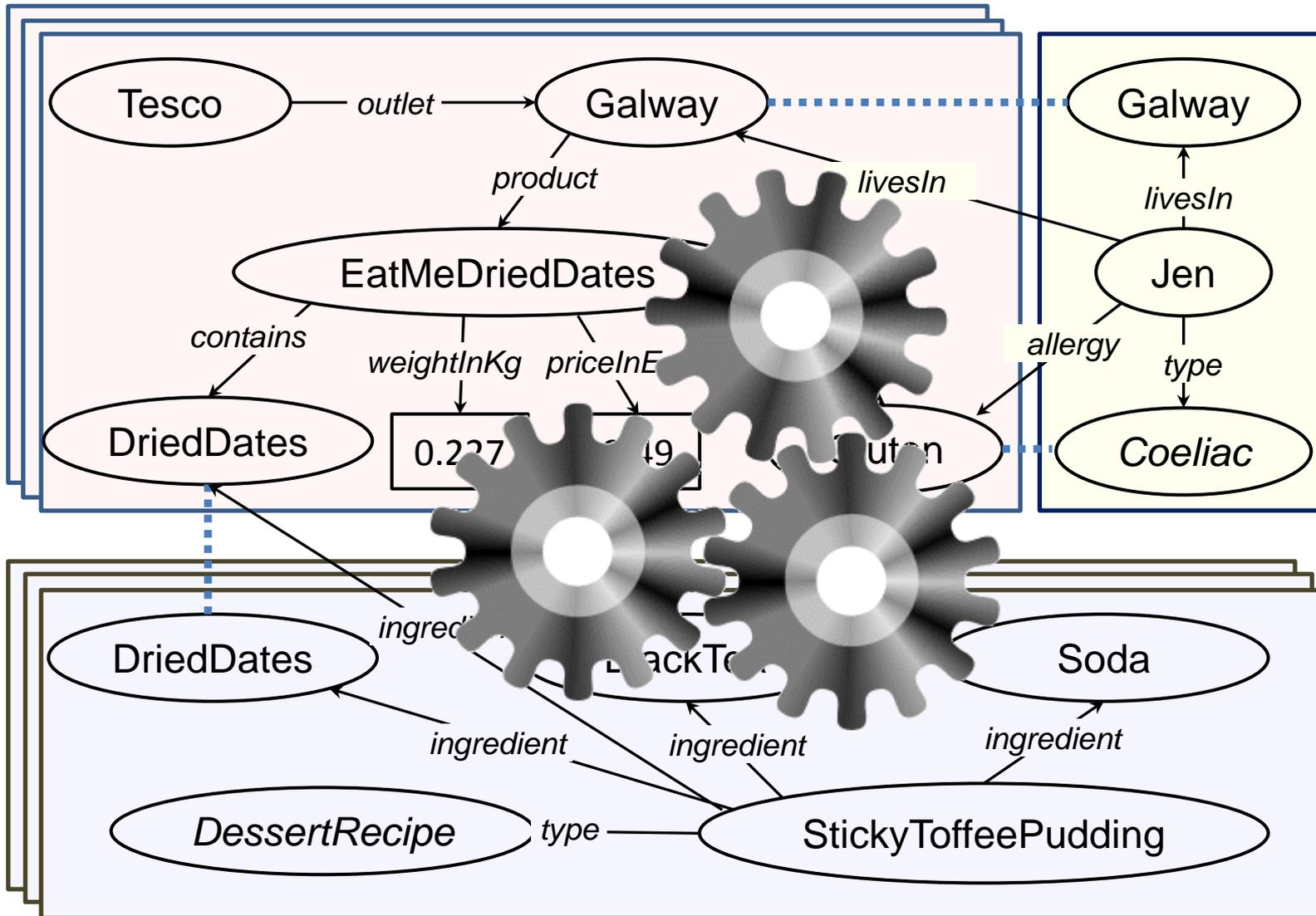


The screenshot shows a web browser window with a red title bar. The address bar contains the URL www.enkivillage.com/moon-landing-hoax.html. The page title is "Moon Landing Hoax". Below the title, there are navigation icons, a category label "Society & Culture", a view count "27K+", and a comment count "375". The main content area features a paragraph of text: "That's one small step for [a] man, one giant leap for mankind." So said Neil Armstrong on that momentous occasion on July 21, 1969 when he stepped onto the surface of the moon for the very first time...supposedly. Does anyone else find it strange that decades before the internet, nanotechnology and cloning, NASA were able to build a rocket that could not only land on the moon, but live broadcast the event to 600 million people? Was the moon landing fake? Below this is a section titled "Top 8 Examples Proving the Moon Landing Was a Hoax" with a dashed border and a small icon in the top left corner. The text inside this section reads: "The 60's was a decade where technology was only just figuring out how to develop the computer mouse, so it's not surprising that many people question the authenticity of the moon landing. Here are the top 8 lines of evidence exposing the moon landing hoax."

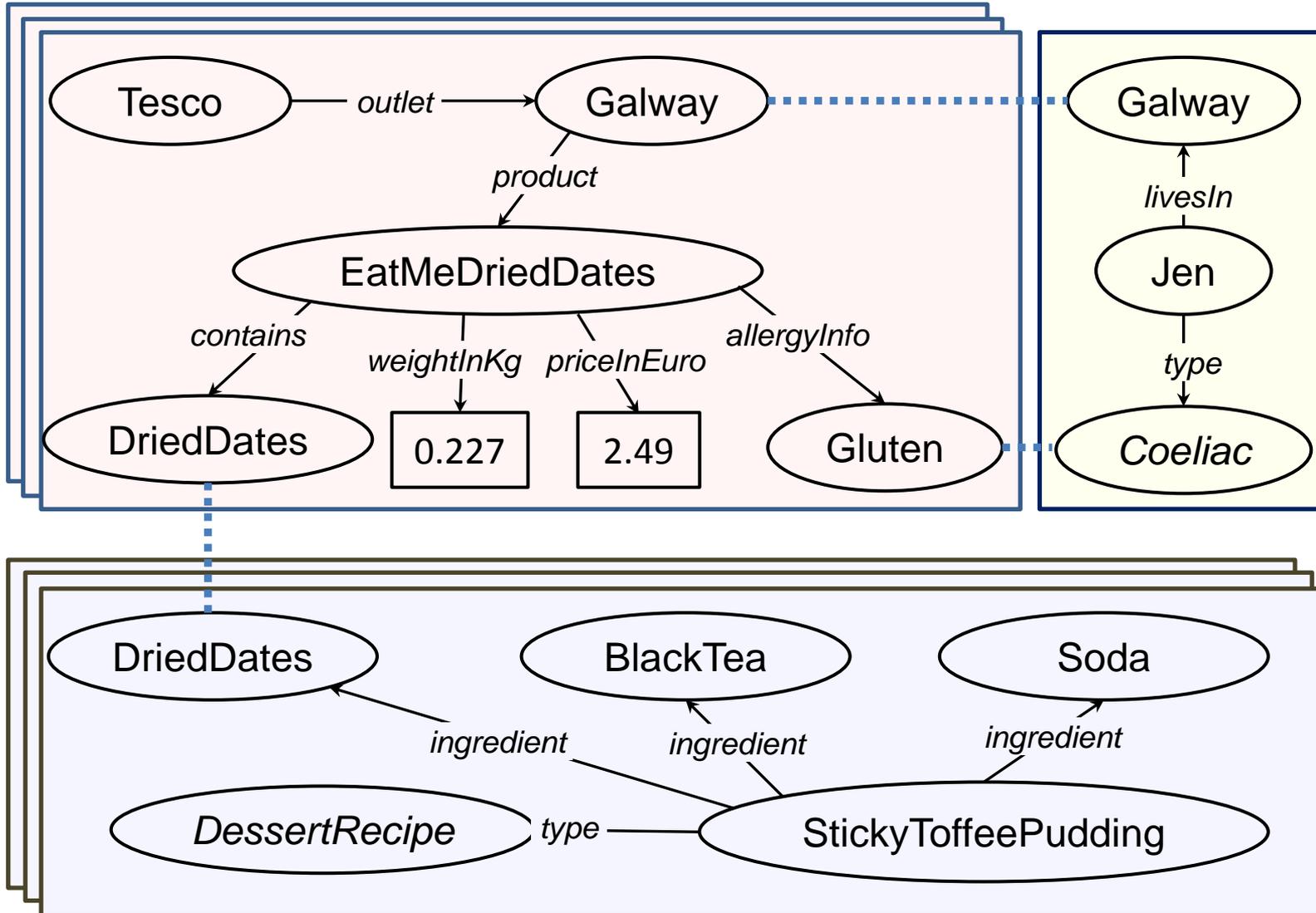
Top 8 Examples Proving the Moon Landing Was a Hoax

The 60's was a decade where technology was only just figuring out how to develop the computer mouse, so it's not surprising that many people question the authenticity of the moon landing. Here are the top 8 lines of evidence exposing the moon landing hoax.

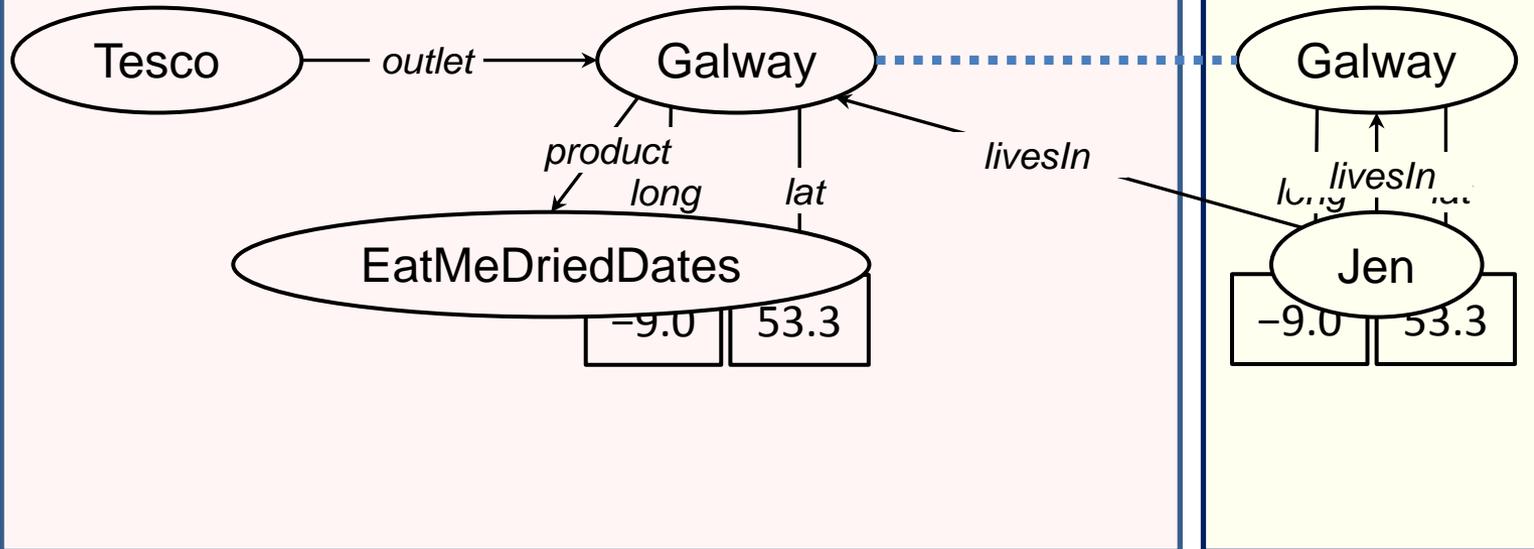
Automatic Integration possible ...



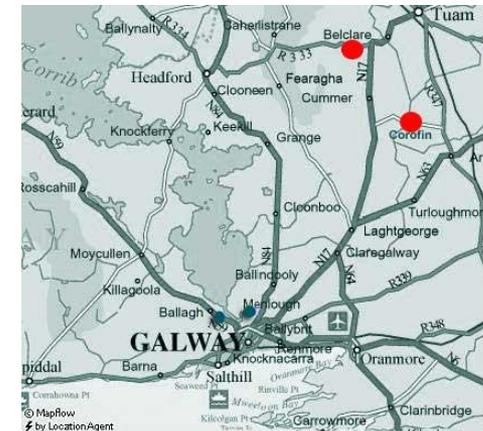
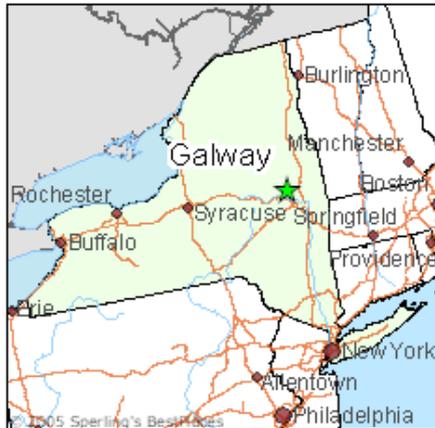
Automatic Integration possible ...



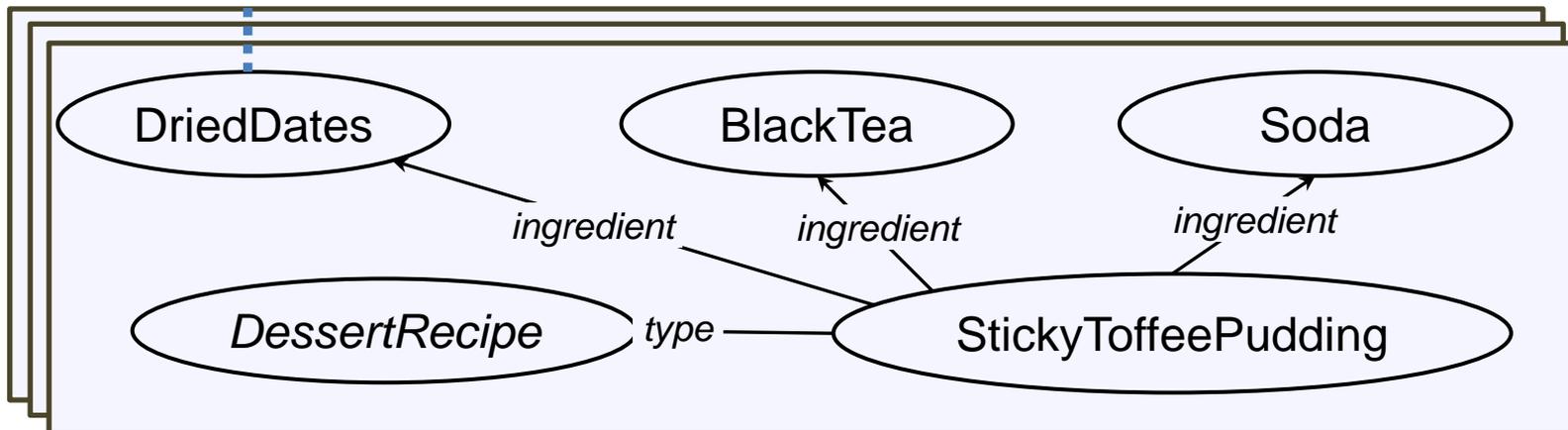
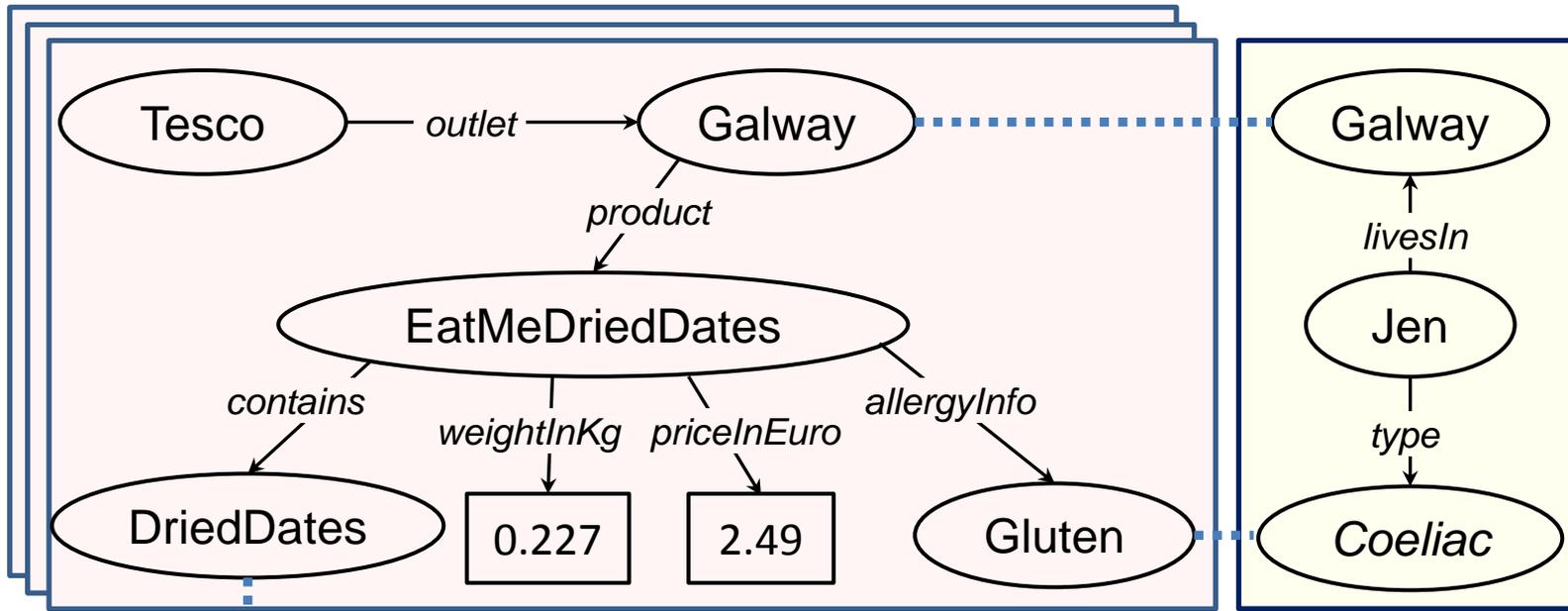
Automatic Integration possible ...



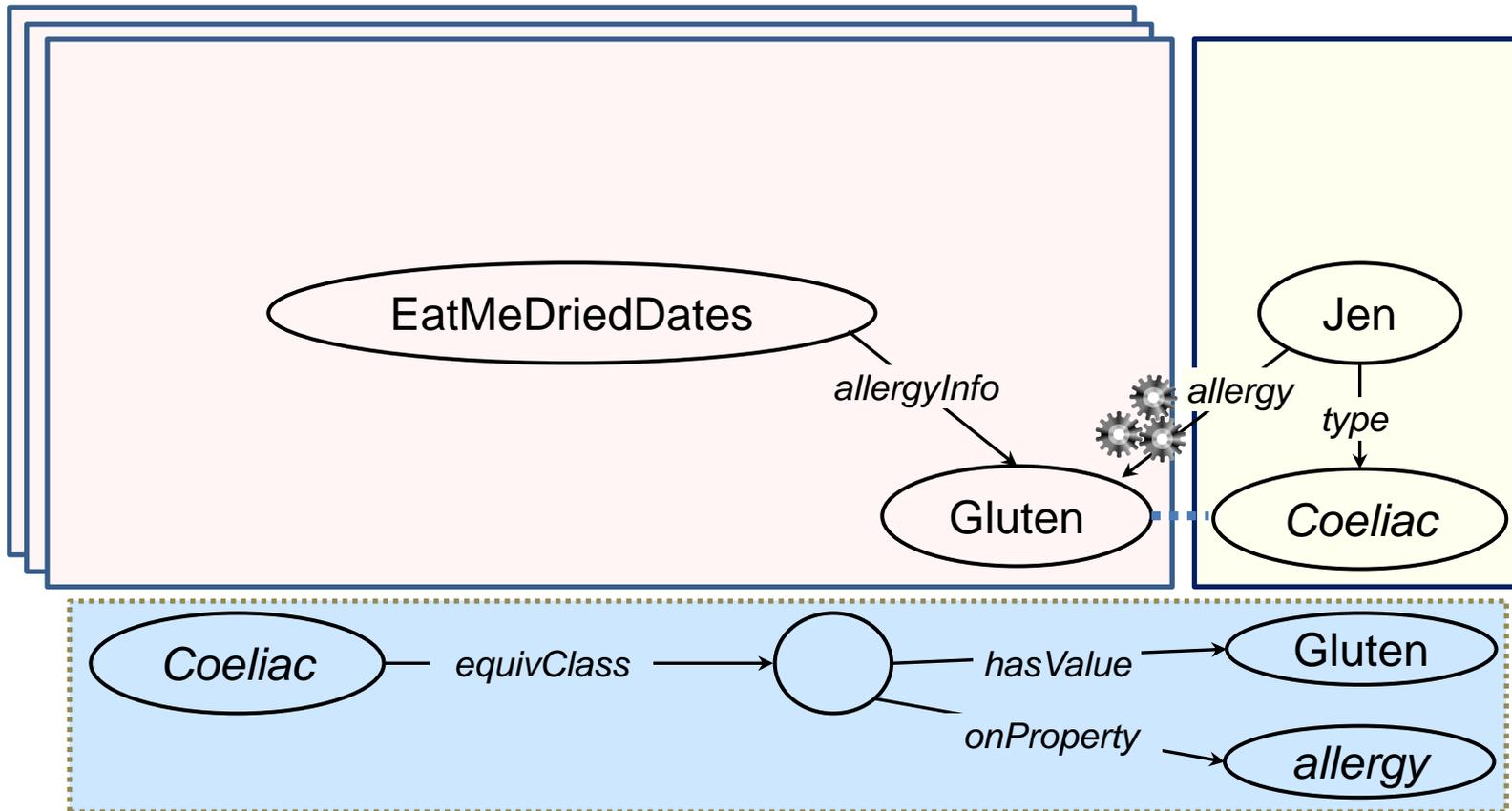
New York



Automatic Integration possible ...



Automatic Integration possible ...

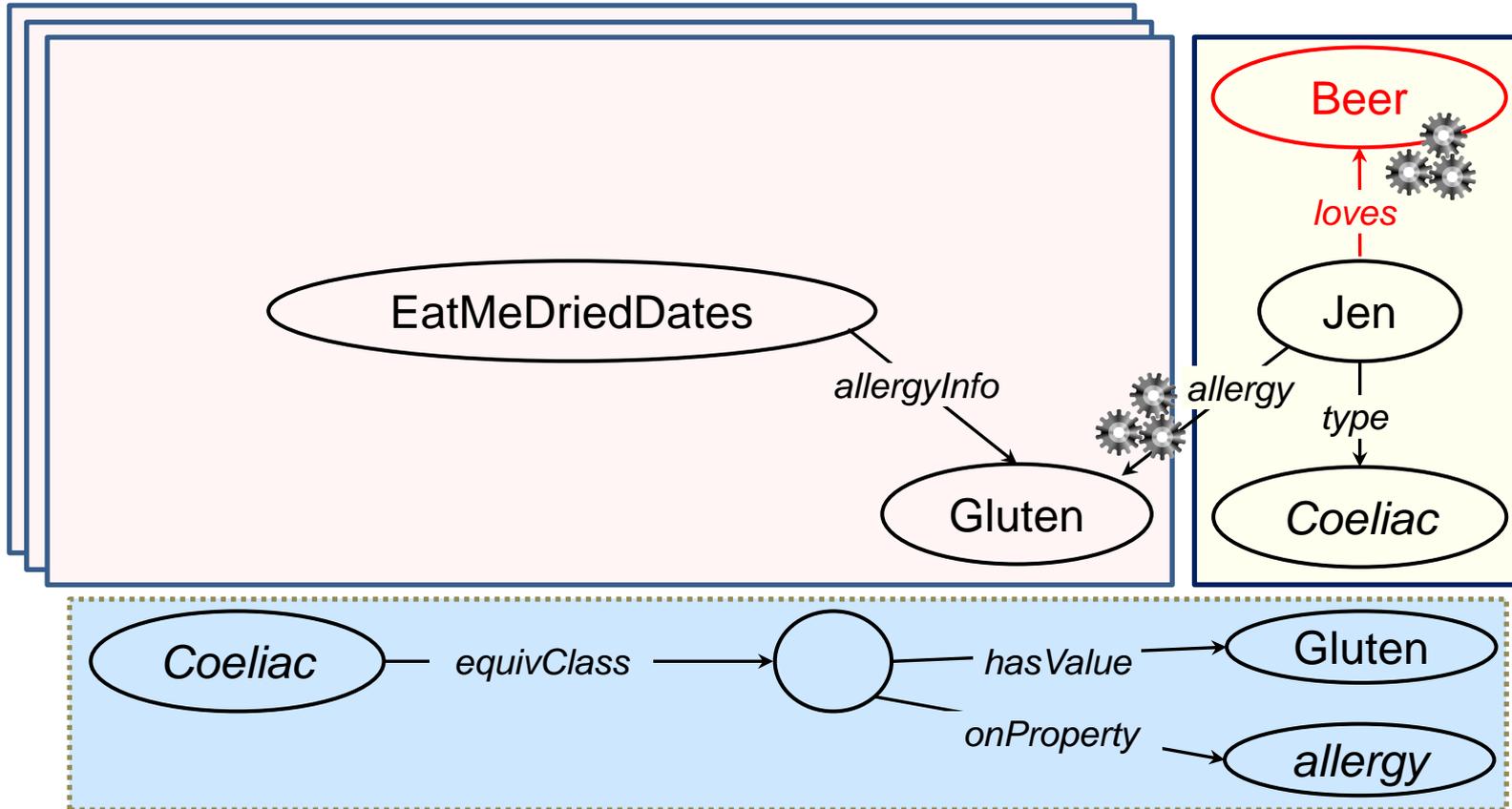


Relationship between Coeliac and Gluten?

- Coeliac(Jen) [Jen is a Coeliac]
- Coeliac $\equiv \exists$ allergy.{Gluten} [a Coeliac has allergy to Gluten]
- \therefore allergy(Jen,Gluten) [Jen has allergy to Gluten]



But not everything on the Web is true



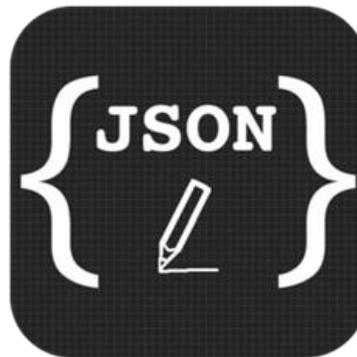
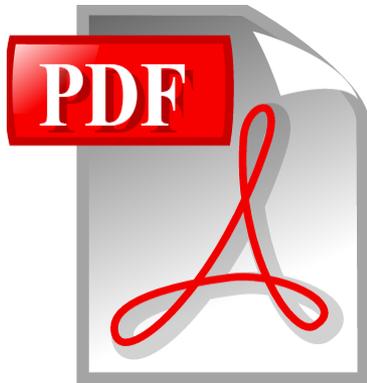
- Coeliac(Jen)
- Coeliac $\equiv \exists$ allergy.{Gluten} *loves beer*
- *loves* \therefore allergy(Jen, Gluten) *loves beer*

- [Jen is a Coeliac] *loves beer*
- [a Coeliac has allergy to Gluten] *loves beer*
- [Jen has allergy to Gluten] *loves beer*



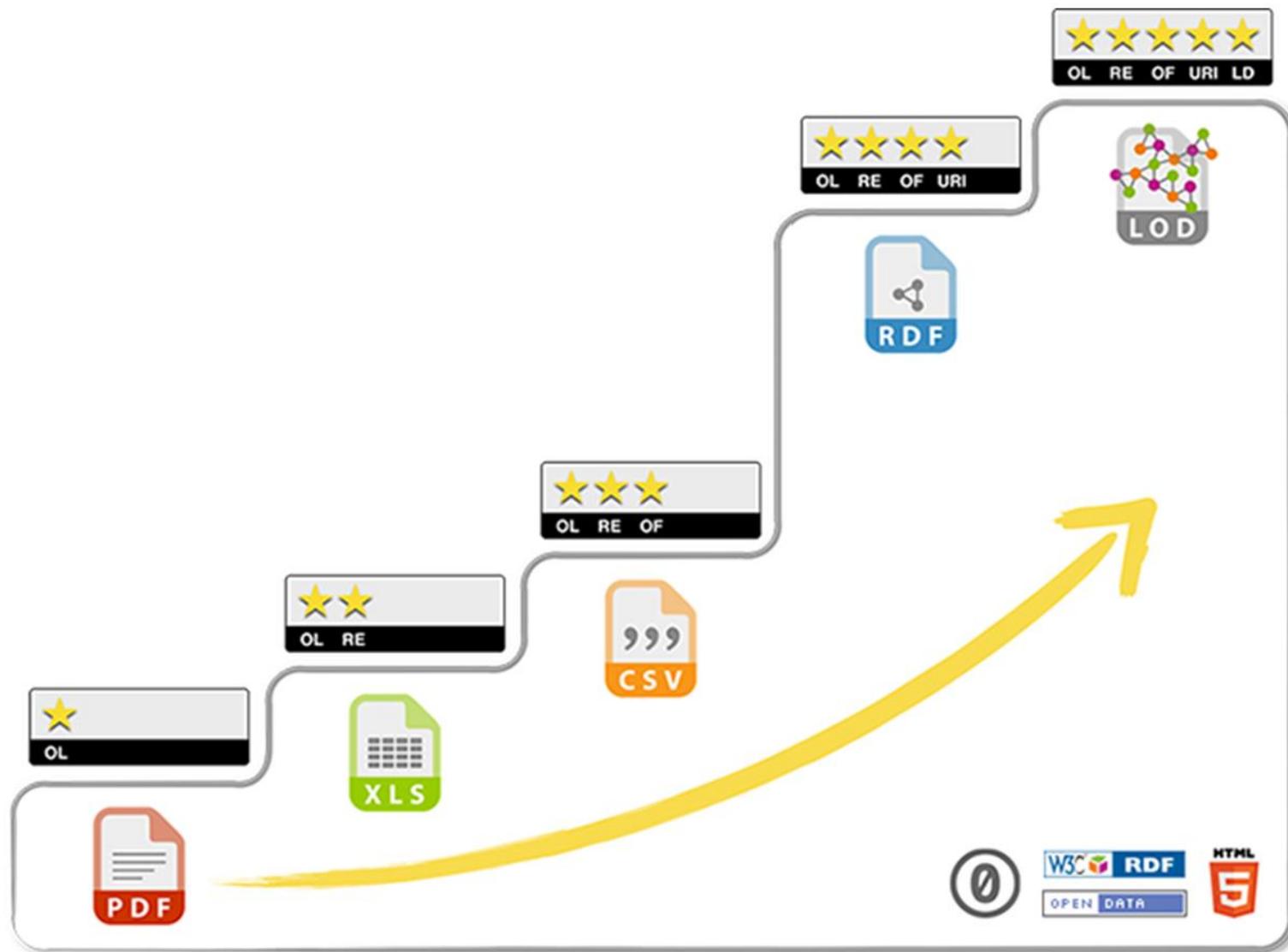
**BIG CATCH NUMBER 4:
LEGACY DATA NOT RDF**

Most Web (meta-)data in ...



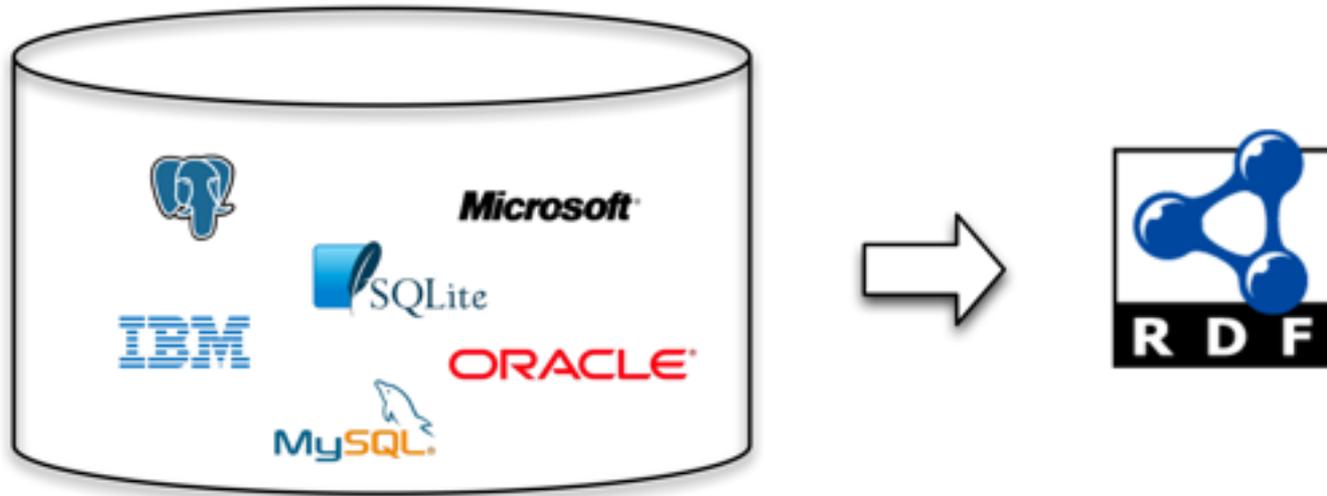
... and so on ...

From ★★ to ★★★★★ is a big step!!



Next week: RDB2RDF

- From relational databases (RDB) to RDF ...



**ACTUALLY THERE ARE LOTS OF
OPEN ISSUES**

Open Issues / Research Questions

- How to efficiently access Linked Data?
- How to automatically link datasets?
- How to reason over Web data?
- How to verify/measure quality?
- How to deal with deceit?
- How to make it all “easy to use”?
- How best to model vocabularies for re-use?
- How to convert legacy data to Linked Data?
- ...

None of these problems is a deal-breaker.
All of these problems are subject to research!
Solutions are being proposed!

RECAP

Linked Data

- Four principles:
 - Use URIs (IRIs) to name things
 - Use HTTP URIs that can be looked up
 - Return RDF about that thing when looked up
 - Provide links in the RDF to other RDF documents
- Identifying things, not documents:
 - Hash: use a fragment identifier
 - Slash: use a 303 See Other redirect

Linked Data

- **Linked Open Data:**
 - Publishing Open Data as Linked Data
- **Many open issues:**
 - How to integrate/link data
 - How to apply reasoning
 - How to access data efficiently
 - Dealing with data quality
 - ...



Structure of the course

- Marking structure:
 - 45% labs
 - 15% project
 - 10% reading group
 - 30% exam

Speculative Schedule 😊 (NOT DEFINITE!)

- Now: Week 11(1)
- 11(2): Lab
- 12(1): RDB2RDF
- 12(2): Lab
- 13(1): Project
- 13(2): Project
- 14(1): Project
- 14(2): Project Presentations
- 15(1): Lecture (Recap/Revision)
- 15(2): Auxiliar / Q&A / No class?

Questions?

