CC7220-1
La Web de Datos
Primavera 2018

Lecture 9: Linked Data

Aidan Hogan
aidhog@gmail.com

Previously …

* More or less

Semantic Web: Data, Logic, Query

Data:

Logic: “$(b$,capital,$a) \rightarrow (a$,partOf,$b)$”

“$(a$,partOf,$b)$, $(b$,partOf,$c) \rightarrow (a$,partOf,$c)$”

Query: “$(x$,partOf,$y)$?”

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 ...

CC7220-1

La Web de Datos

Primavera 2018

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

* More or less

Semantic Web: Data, Logic, Query

Data:

What are we missing from here to build a Web of Data?

Logic: “$(b,\text{capital},a) \rightarrow (a,\text{partOf},b)$”

“$(a,\text{partOf},b),\ (b,\text{partOf},c) \rightarrow (a,\text{partOf},c)$”

Query: “(x,partOf,y)?”

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

* More or less

Semantic Web: Data, Logic, Query, Links

Data:

How can we add links to this picture?

Logic: "$(b$,capital,$a) \rightarrow (a$,partOf,$b)$"

"$(a$,partOf,$b)$, $(b$,partOf,$c) \rightarrow (a$,partOf,$c)$"

Query: "$(x$,partOf,$y)$?"

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

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!

Pre-Linked Data …

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!

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:

(The mug is explained 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.

Linked Data examples ...

The New York Times

7 Earth-Size Planets Orbit Dwarf Star, NASA and European Astronomers Say

Trappist-1, named after a robotic telescope in the Atacama Desert of Chile that the astronomers initially used to study the star, is what astronomers call an "ultracool dwarf," with only one-twelfth the mass of the sun and a surface temperature of 4,150 degrees Fahrenheit, much cooler than the 10,000 degrees radiating from the sun. Trappist is a shortening of Transiting Planets and Planetesimals Small Telescope.

TRAPPIST-1 (Q23986556)

ultra-cool dwarf star

2MASS J23062928-0502285 | Trappist 1

edit

In more languages Configure

Language	Label	Description	Also known as
English	TRAPPIST-1	ultra-cool dwarf star	2MASS J23062928-0502285 Trappist 1
Spanish	TRAPPIST-1	estrella enana ultra-fría	2MASS J23062928-0502285 Trappist 1
Mapuche	No label defined	No description defined	

All entered languages

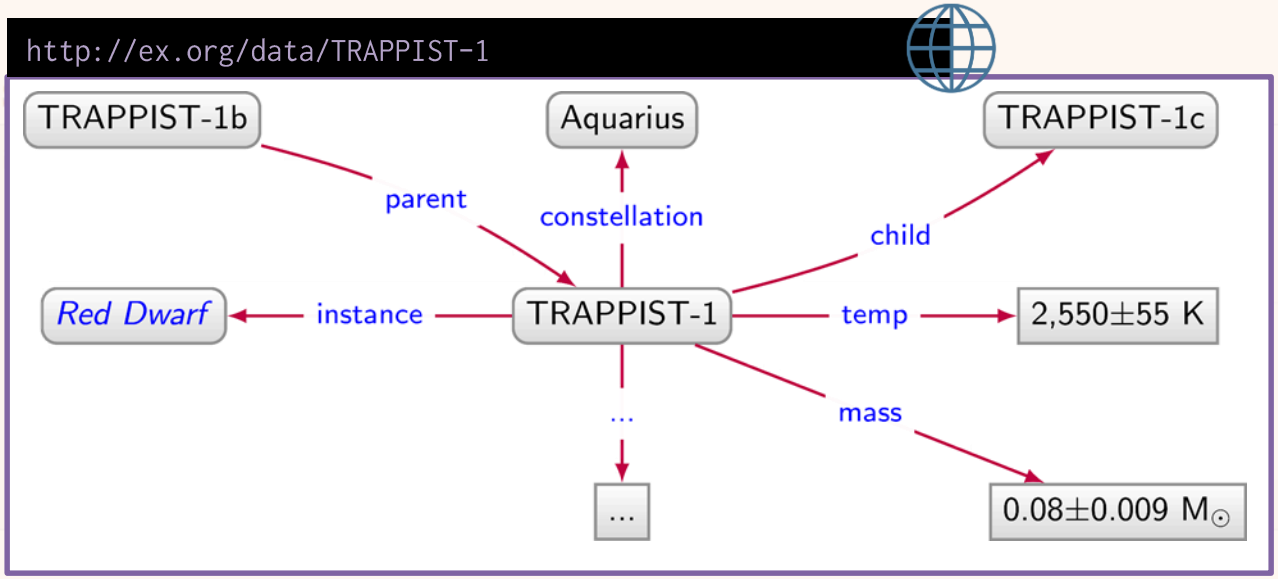
Statements

instance of	red dwarf 0 references add reference	edit
	ultra-cool dwarf 0 references add reference	edit
		add

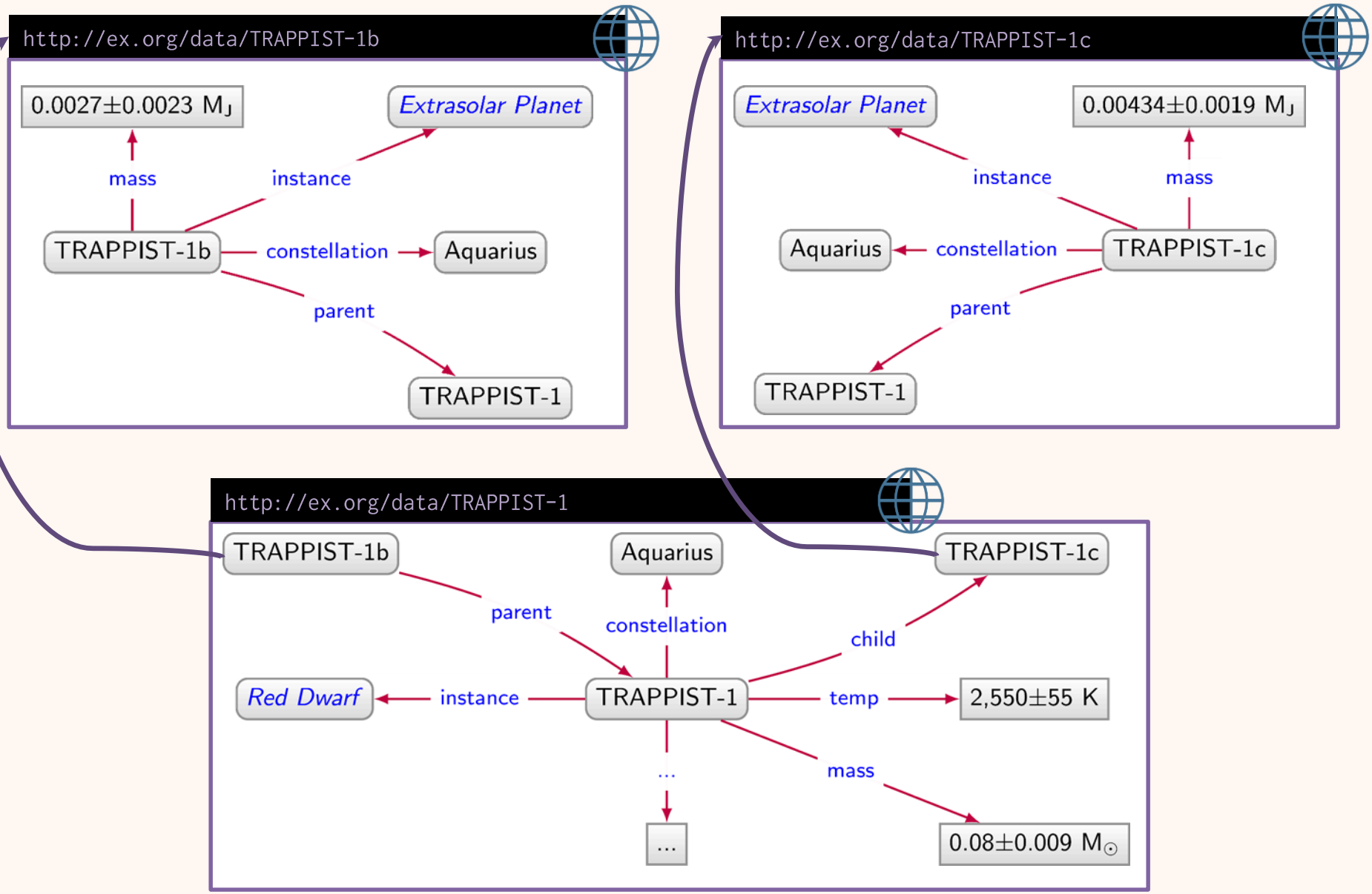
Main page
Community portal
Project chat
Create a new item
Item by title
Recent changes
Random item
Query Service
Nearby
Help
Donate

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

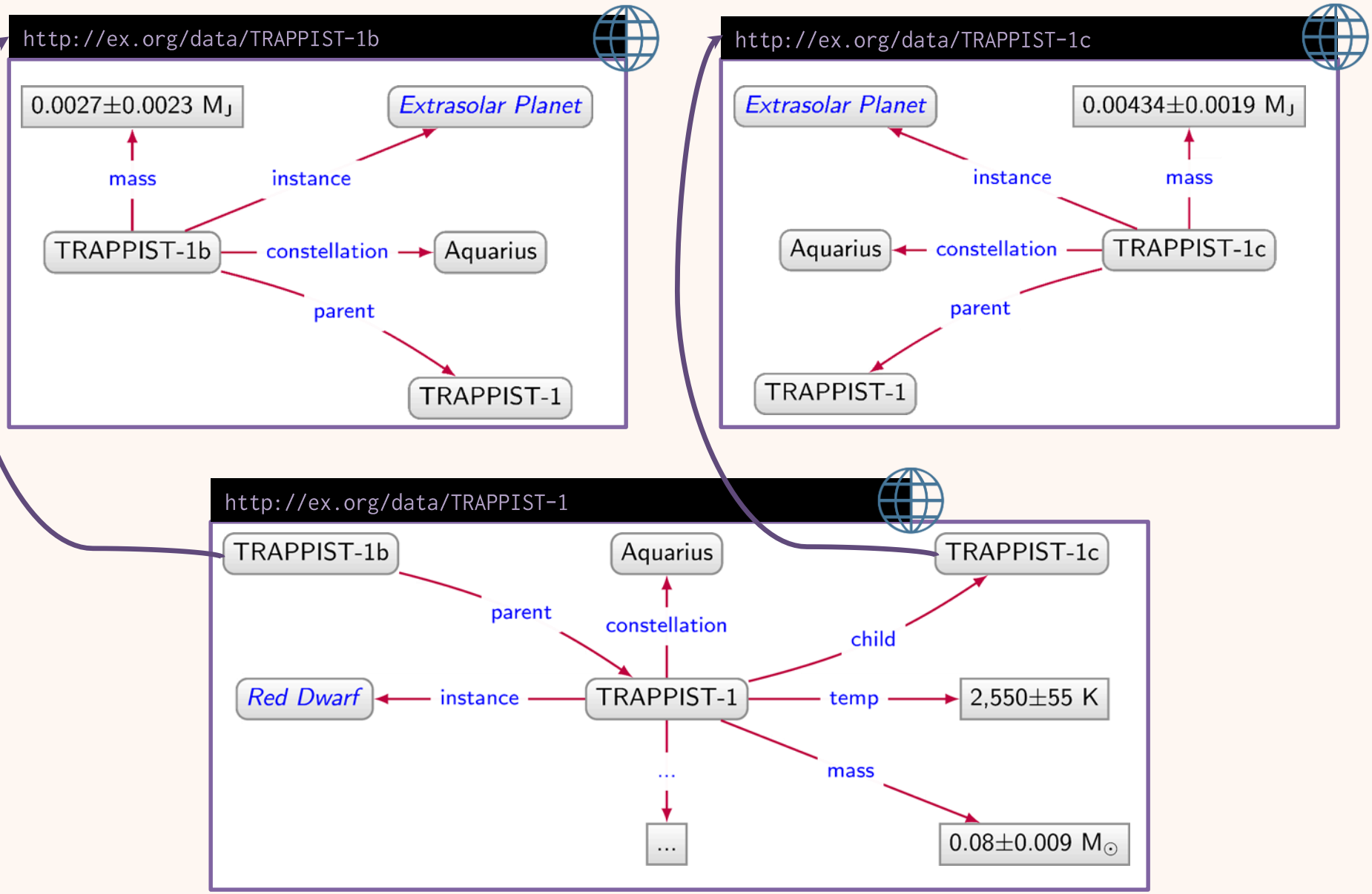
Linked Data Document



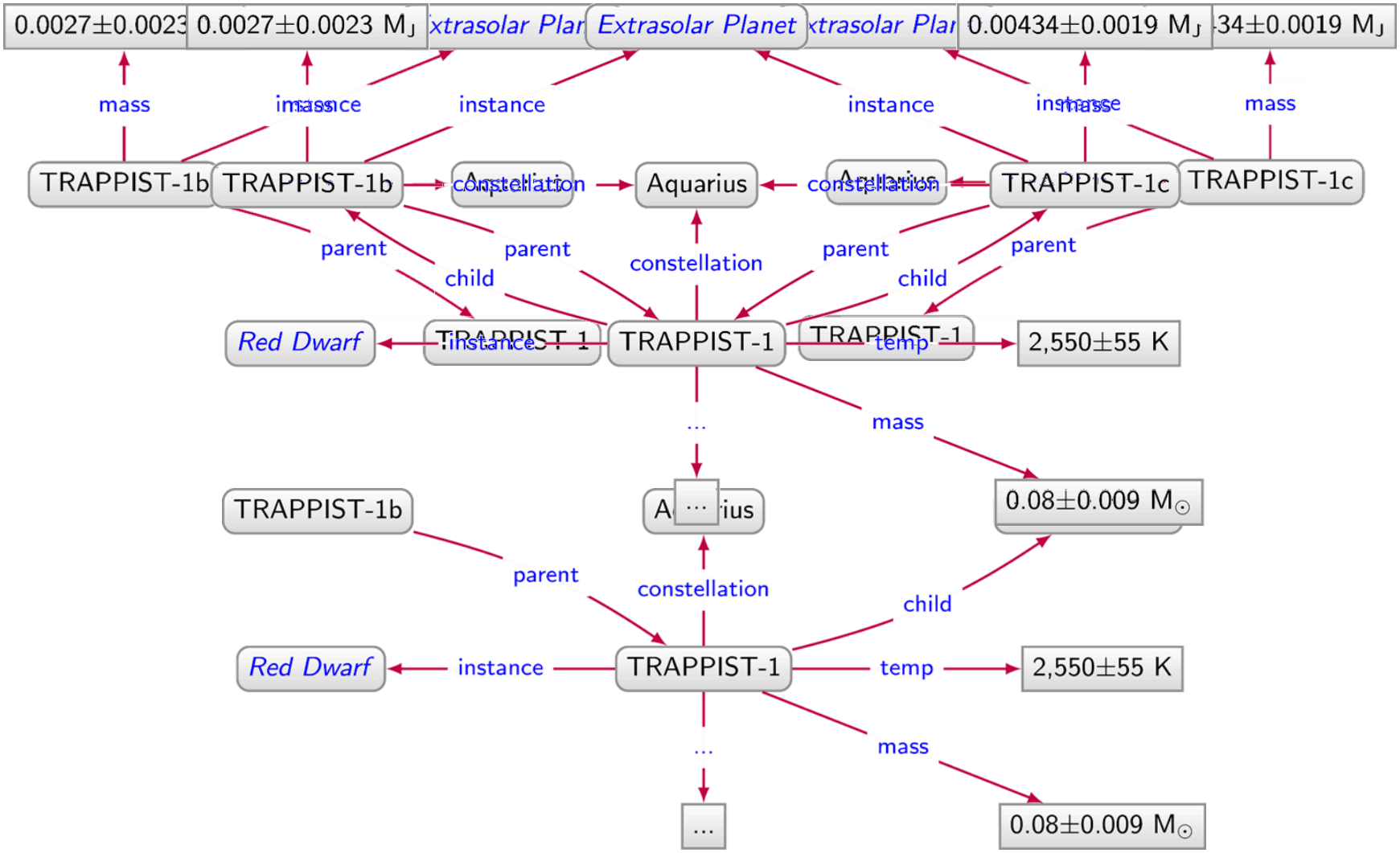
Linked Data Graph



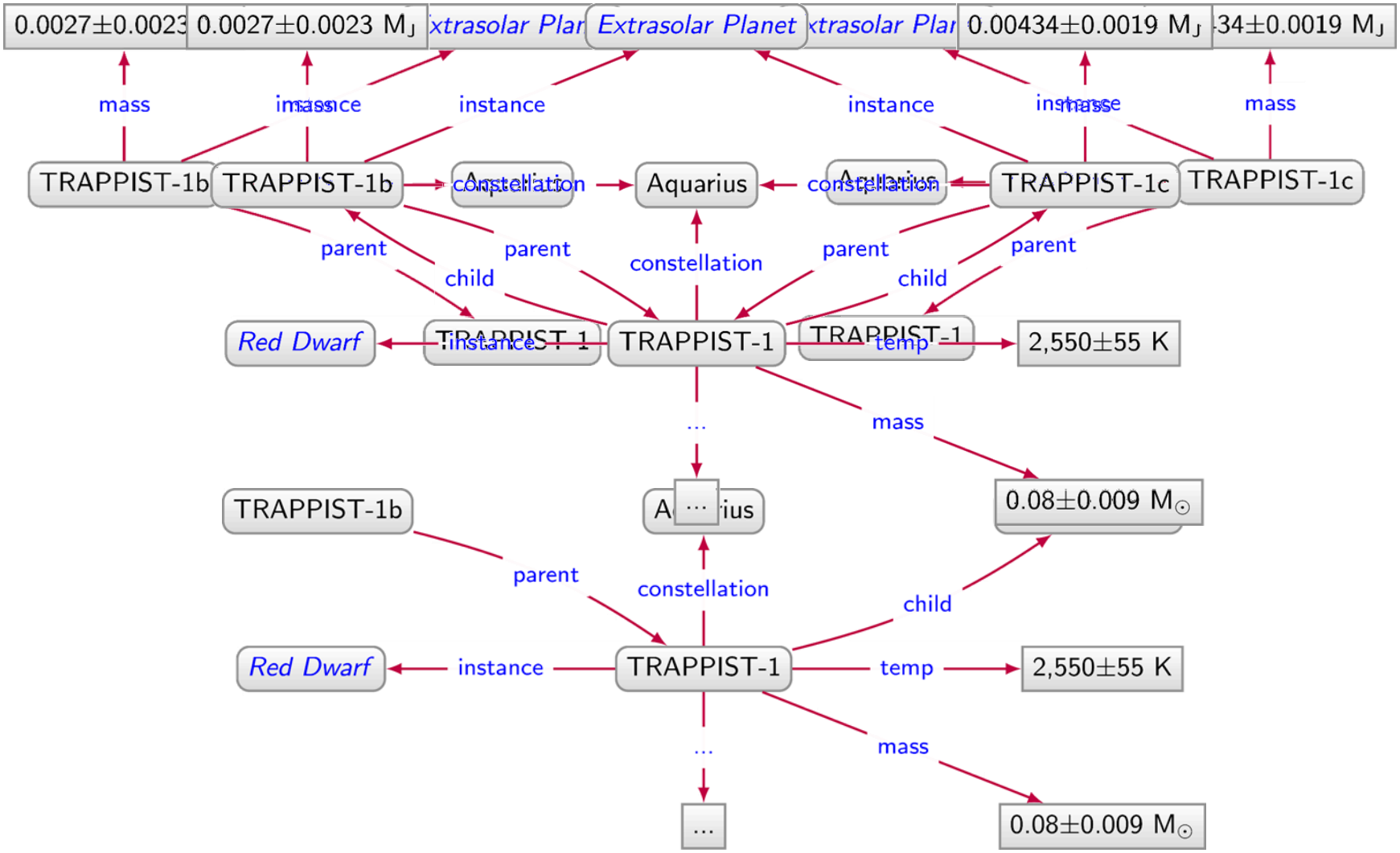
Linked Data Graph



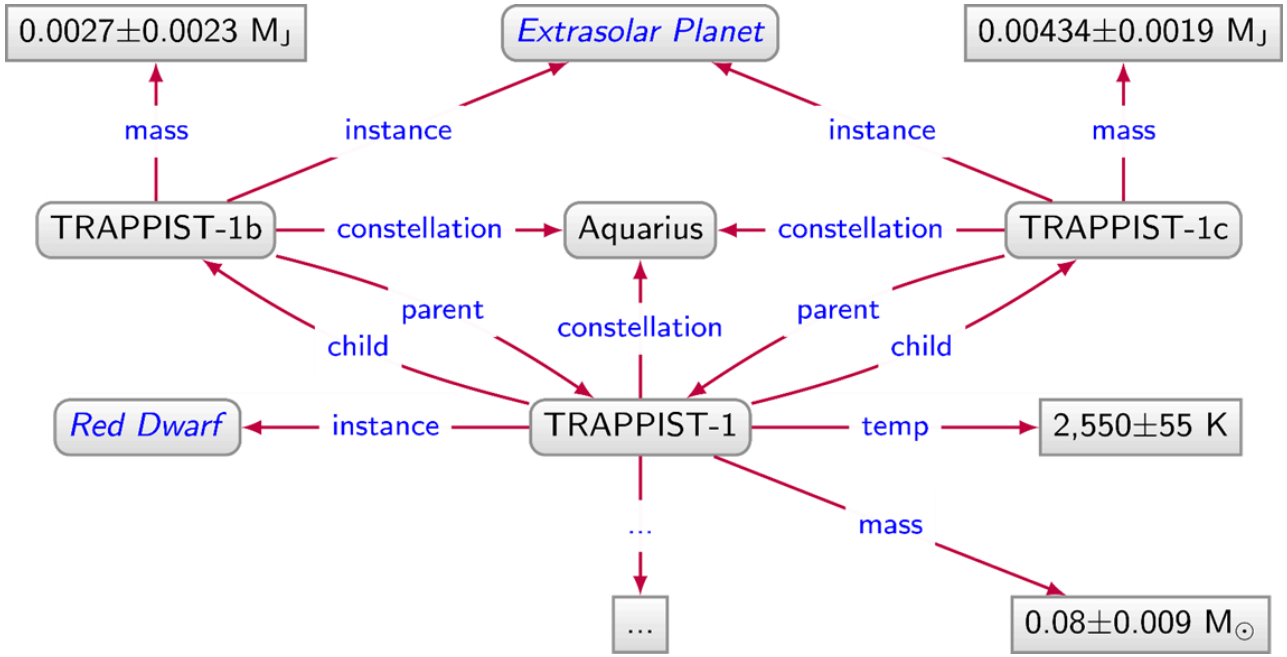
Linked Data Graph



Linked Data Graph

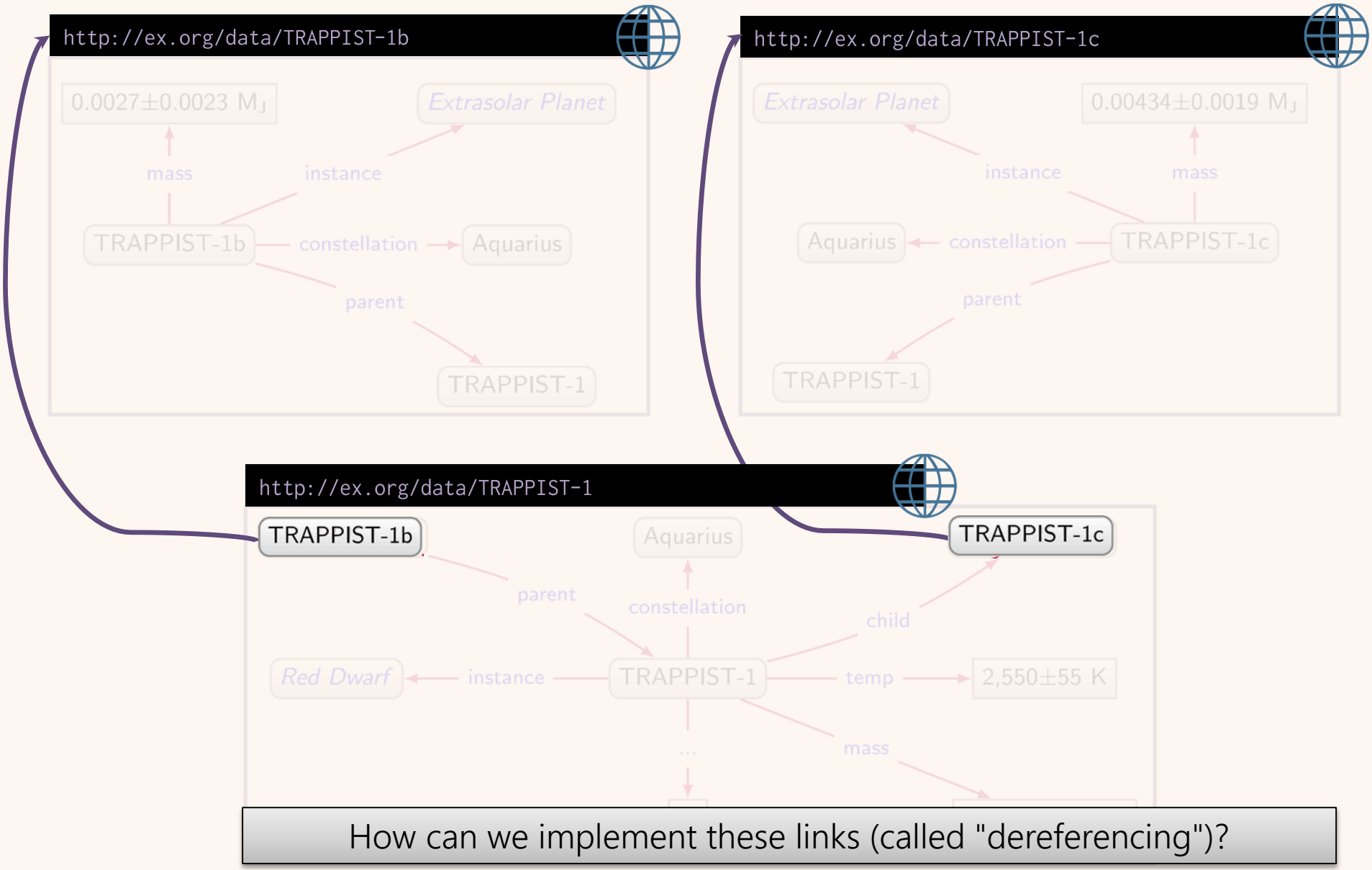


Answer queries over Linked Data using SPARQL




```
SELECT ?planet ?jupiterMass
WHERE {
  :TRAPPIST-1 :child ?planet .
  OPTIONAL { ?planet :mass ?jupiterMass }
}
ORDER BY DESC(?jupiterMass)
```

Linked Data Graph



Implementing Linked Data...

Three recipes for dereferencing

URL Recipe

Use document URLs to identify things

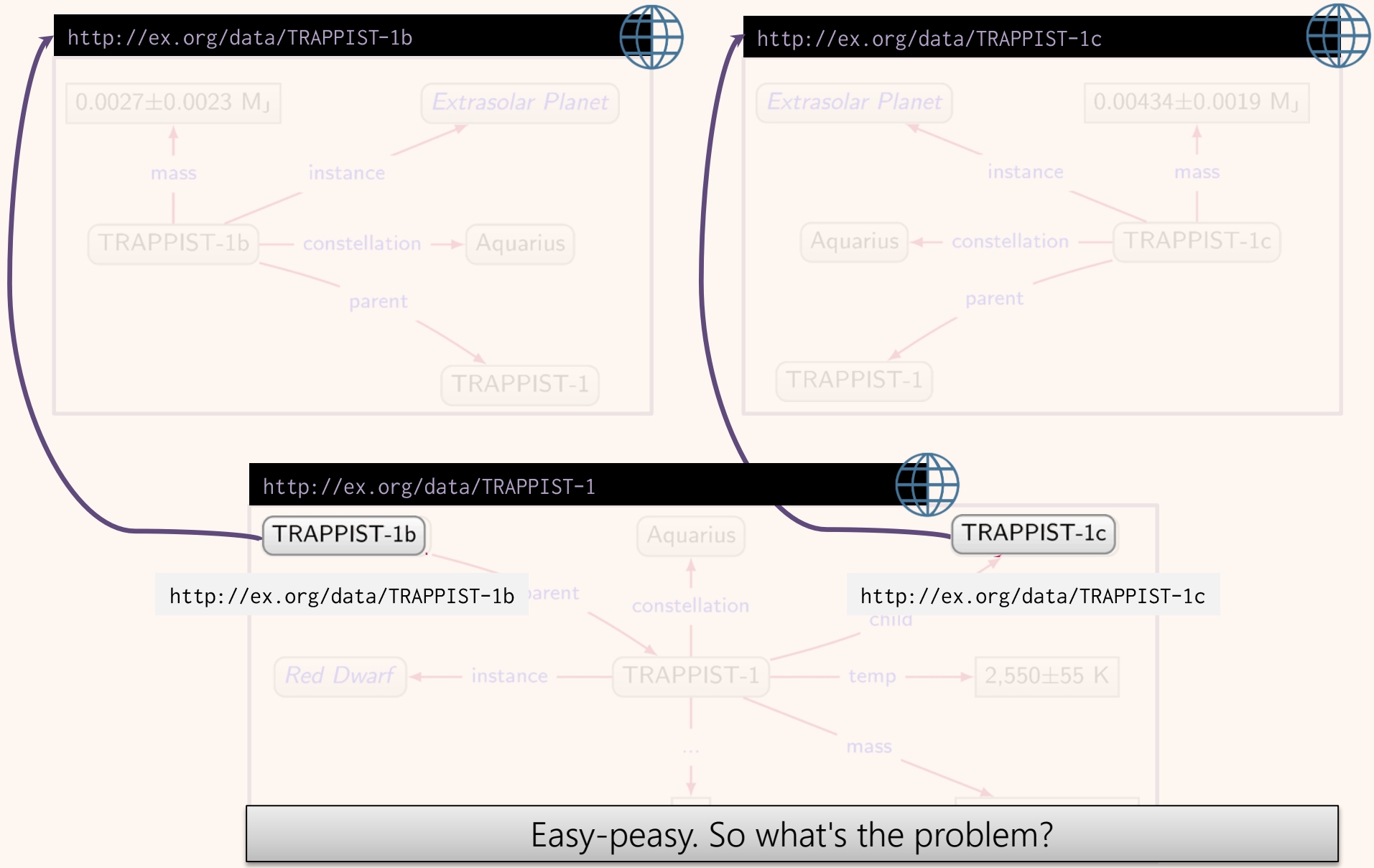
Hash Recipe

Use fragment identifiers to identify things

Slash Recipe

Use special redirects to identify things

URL Recipe

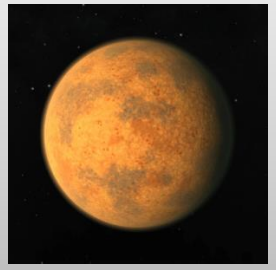


Easy-peasy. So what's the problem?

URL Recipe: The Problem

```
@prefix : <http://ex.org/data/> .
@prefix v: <http://ex.org/voc/> .

:TRAPPIST-1c v:parent :TRAPPIST-1 ;
   v:constellation :Aquarius .
```



```
@prefix : <http://ex.org/data/> .
@prefix v: <http://ex.org/voc/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:TRAPPIST-1c v:updated "2018-08-08"^^xsd:date ;
   v:creator :JaneSmith , :JohnSmith .
```

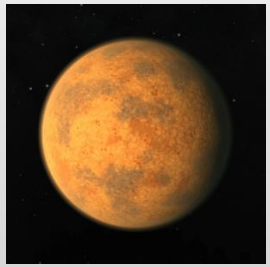
URL Recipe: The Problem

```
@prefix : <http://ex.org/data/> .
@prefix v: <http://ex.org/voc/> .

:TRAPPIST-1c v:parent :TRAPPIST-1 ;
   v:constellation :Aquarius .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:TRAPPIST-1c v:updated "2018-08-08"^^xsd:date ;
   v:creator :JaneSmith , :JohnSmith .
```



Document URLs should only identify documents!

HTTP IRIs usually for documents, not pipes

Three recipes for dereferencing

URL Recipe

Use document URLs to identify **documents**

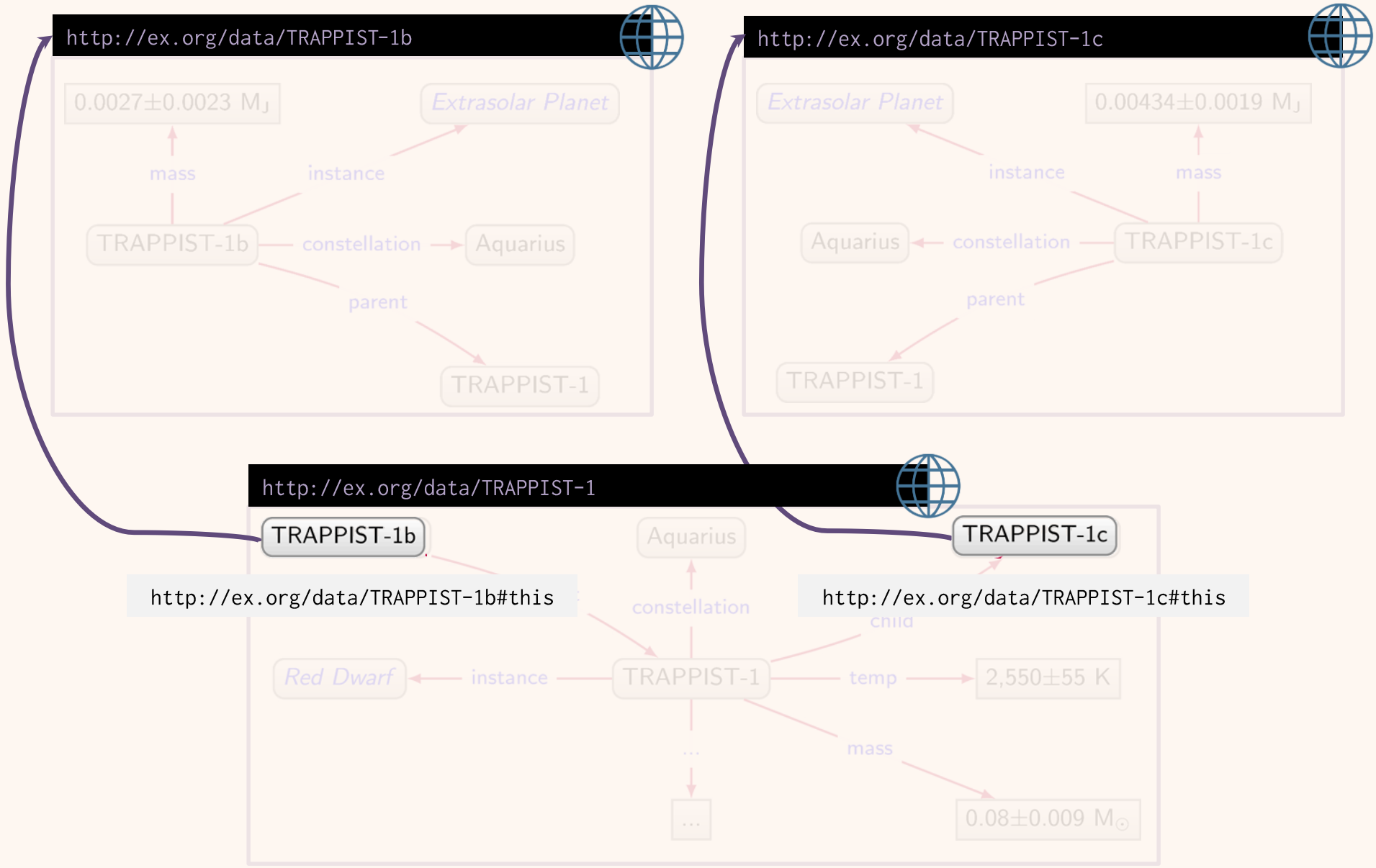
Hash Recipe

Use fragment identifiers to identify things

Slash Recipe

Use special redirects to identify things

Hash Recipe


http://ex.org/data/TRAPPIST-1b
0.0027±0.0023 M_J
Extrasolar Planet
mass
instance
TRAPPIST-1b
constellation
Aquarius
parent
TRAPPIST-1
http://ex.org/data/TRAPPIST-1c
Extrasolar Planet
0.00434±0.0019 M_J
instance
mass
Aquarius
constellation
TRAPPIST-1c
parent
TRAPPIST-1
http://ex.org/data/TRAPPIST-1
TRAPPIST-1b
Aquarius
TRAPPIST-1c
http://ex.org/data/TRAPPIST-1b#this
constellation
http://ex.org/data/TRAPPIST-1c#this
child
Red Dwarf
instance
TRAPPIST-1
temp
2,550±55 K
...
mass
...
0.08±0.009 M⊙

Hash Recipe

- `http://ex.org/data/TRAPPIST-1c`
  - Identifies the document
- `http://ex.org/data/TRAPPIST-1c#this`
  - Identifies the planet
  - Look it up, you get the document

Three recipes for dereferencing

URL Recipe

Use document URLs to identify ~~things~~ **documents**

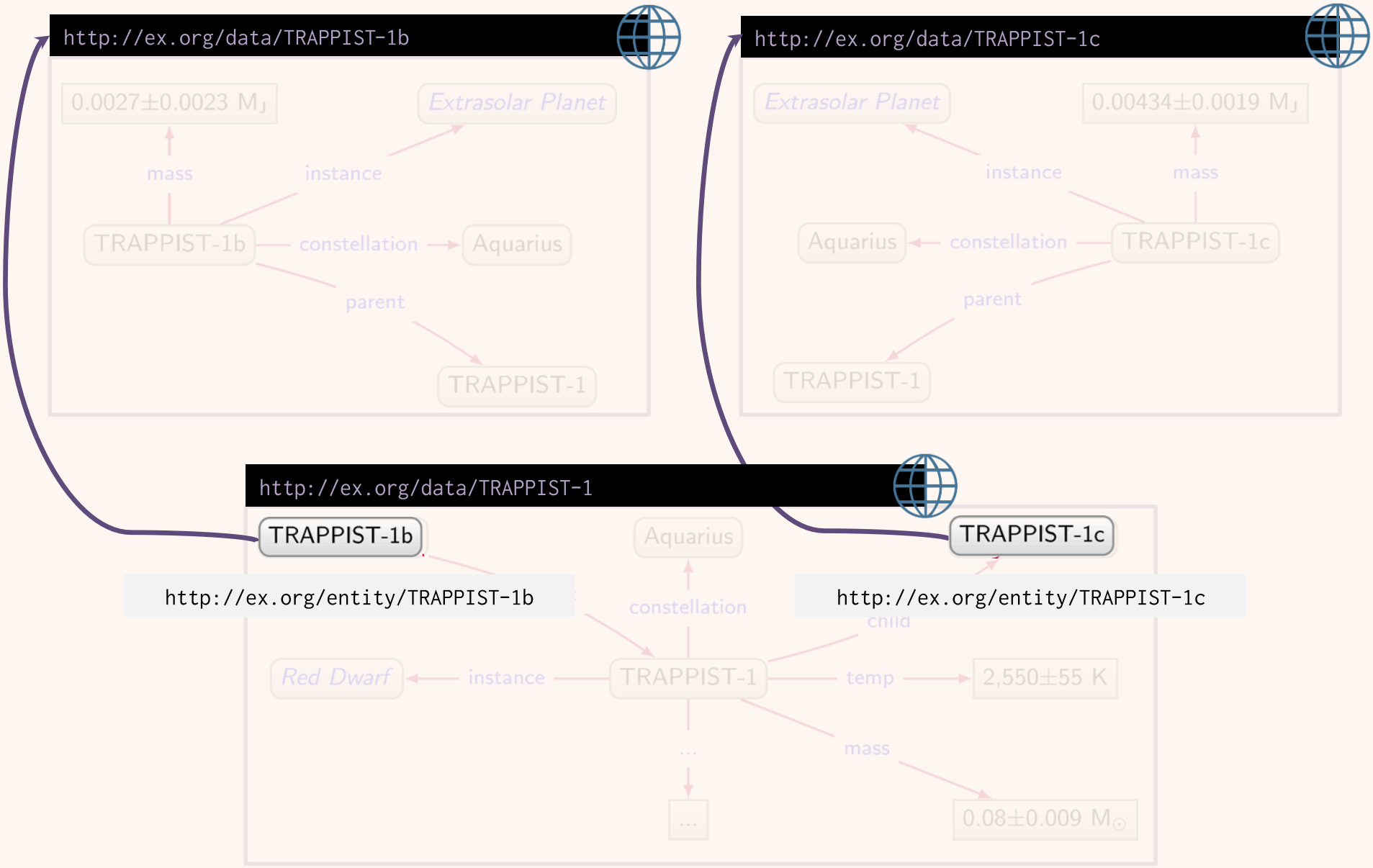
Hash Recipe

Use fragment identifiers to identify things

Slash Recipe

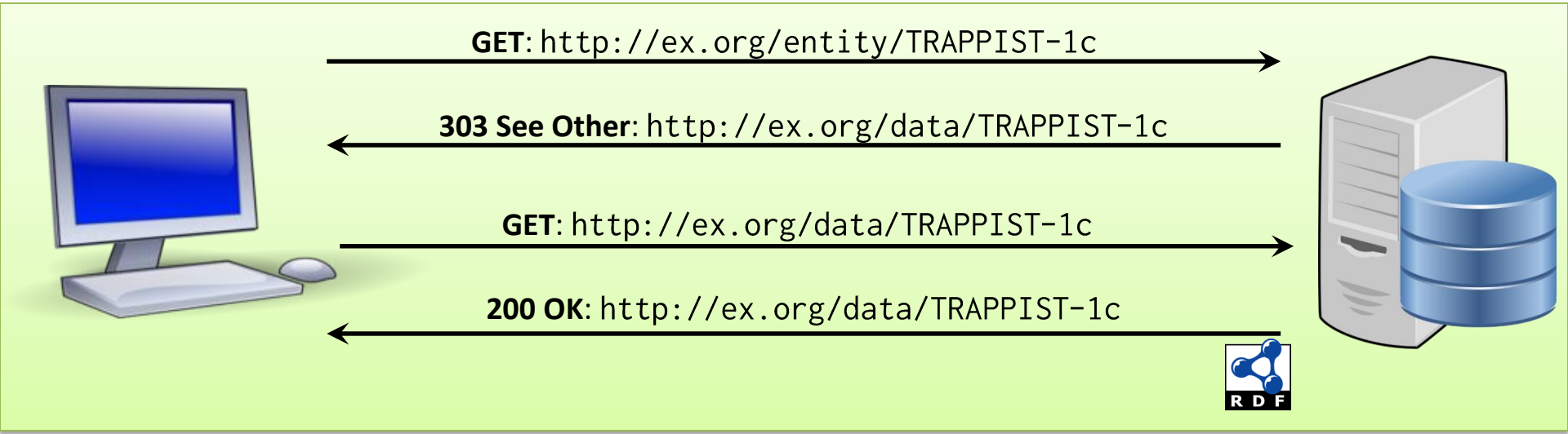
Use special redirects to identify things

Slash Recipe



Slash Recipe

- `http://ex.org/data/TRAPPIST-1c`
  - Identifies the document
- `http://ex.org/entity/TRAPPIST-1c`
  - Identifies the planet
  - Look it up, redirects to the document



Three recipes for dereferencing

URL Recipe

Use document URLs to identify ~~things~~ **documents**

Hash Recipe

Use fragment identifiers to identify things

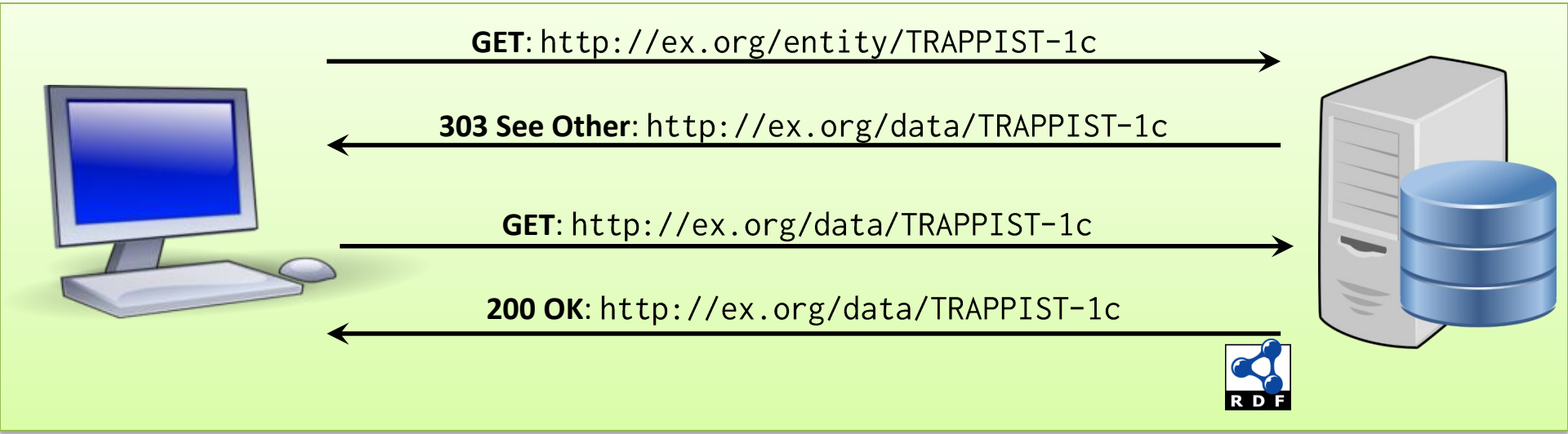
Slash Recipe

Use special redirects to identify things

Hash vs. Slash

Which is better, hash or slash?

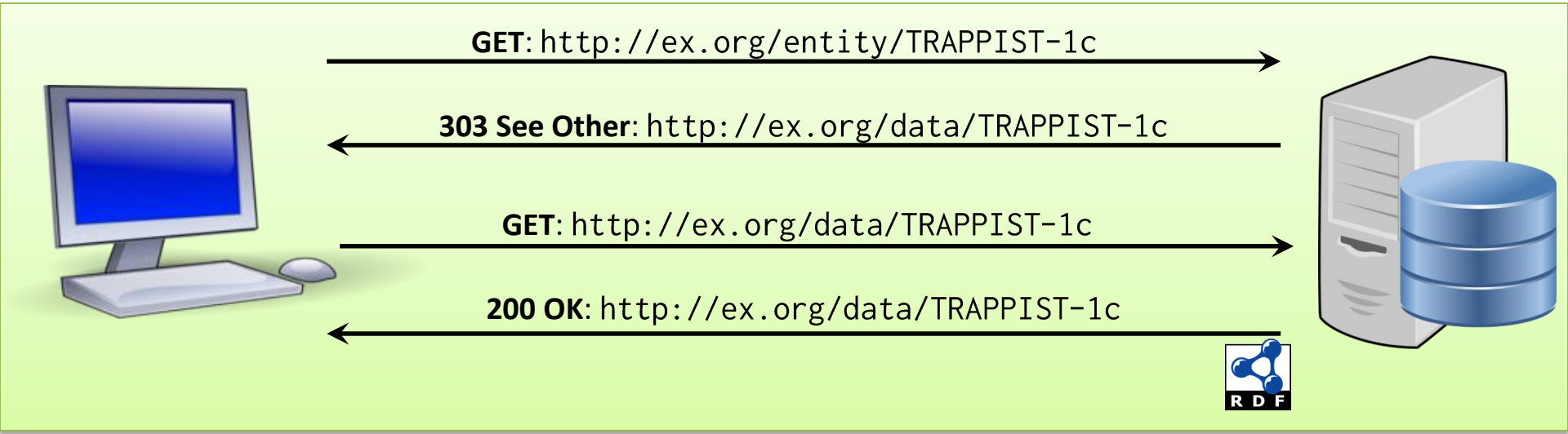
Well, hash has half the number of requests!



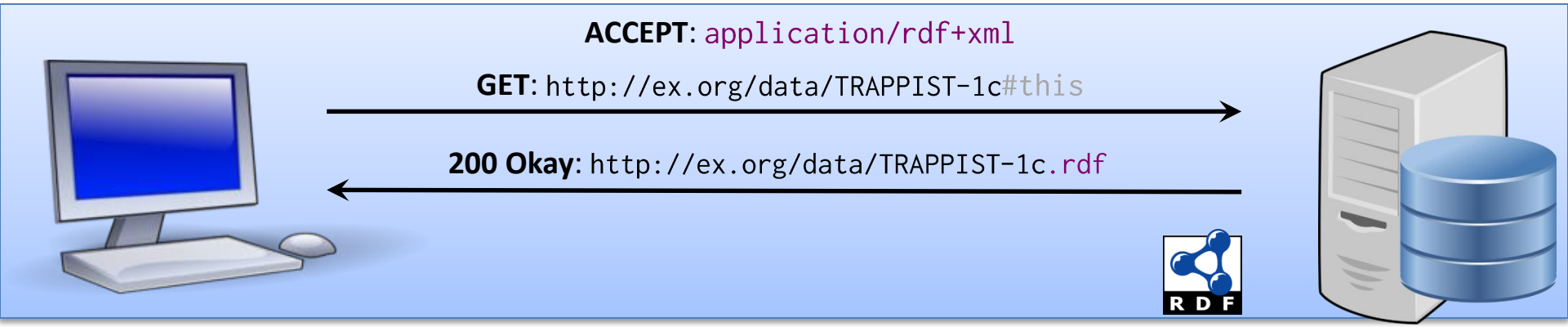
Hash vs. Slash

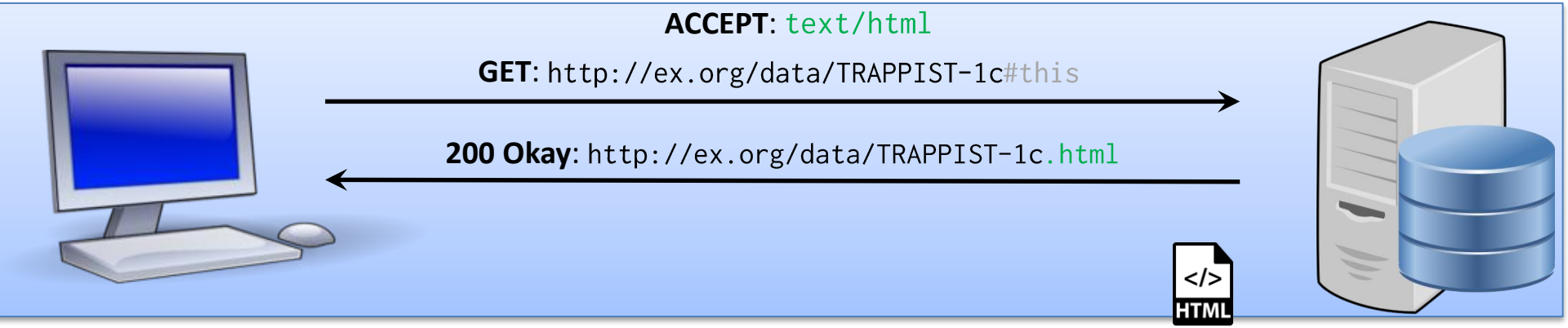
Which is better, hash or slash?

But slash decouples document URLs from entity IRIs



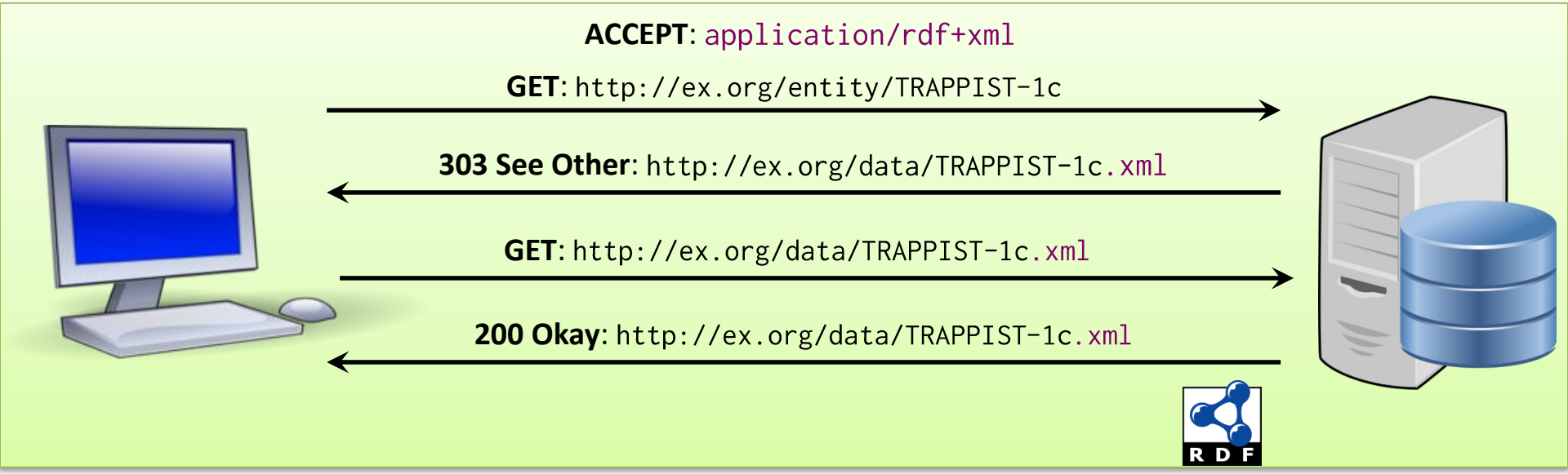
Content negotiation with hash

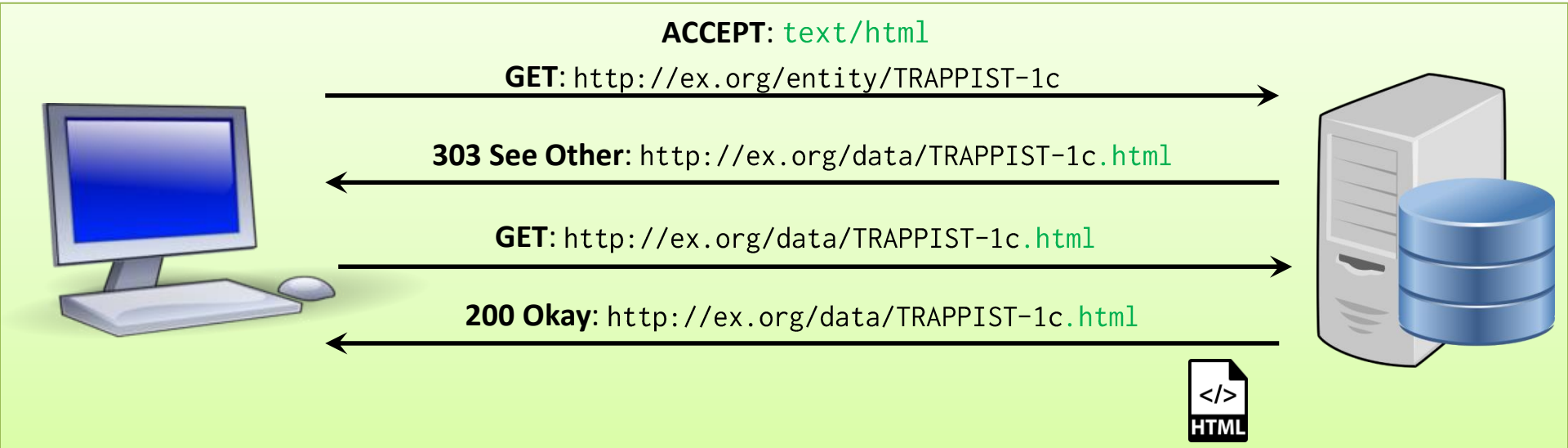




Can also choose from different RDF formats; e.g., Turtle, RDFa, etc.
(if supported by the server that is!)

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:

(The mug is explained now)

Open 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 …

… meets Linked Data

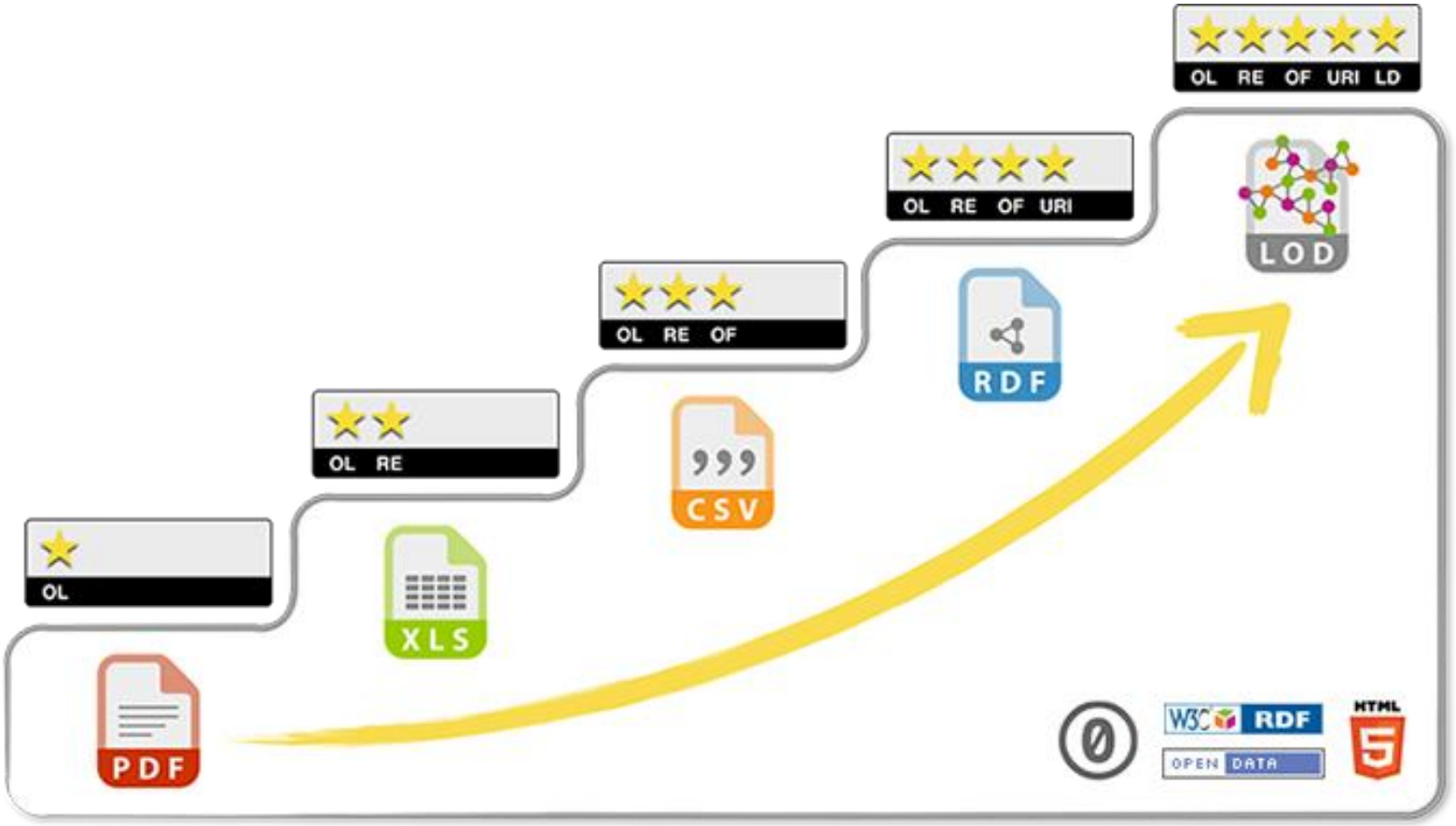
Linked Open Data

Linked Open Data

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 (hint: Linked Data)
  - so consumers can follow links to find out more

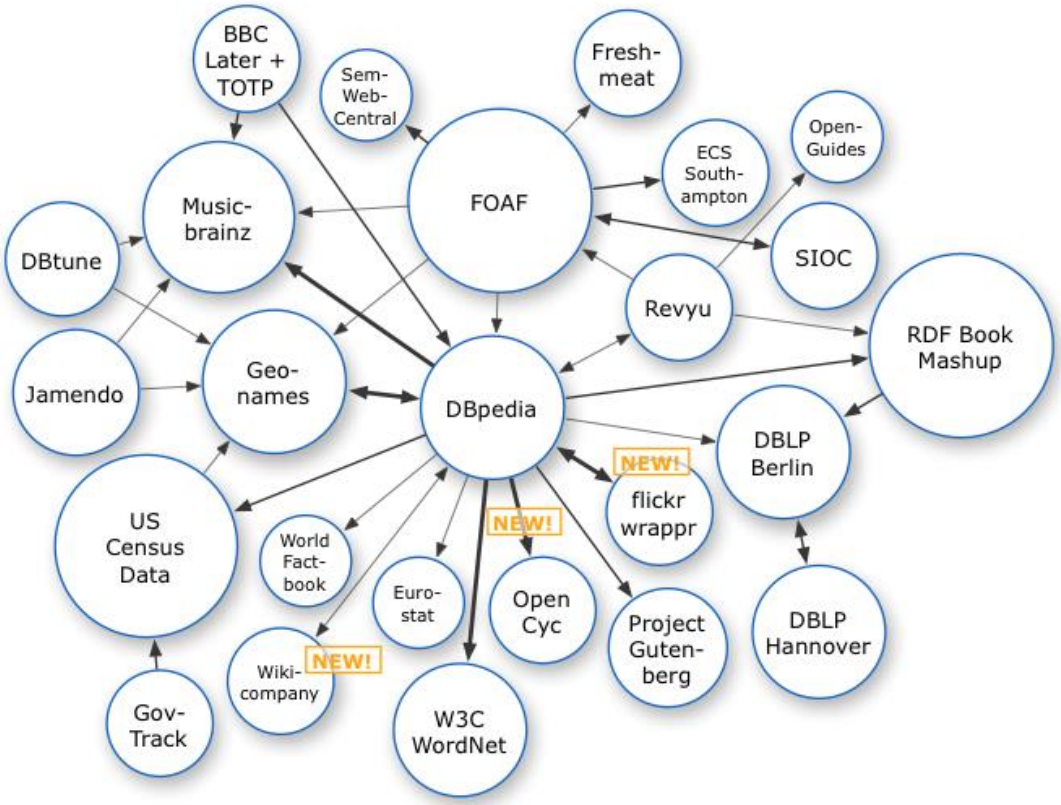
Each star improves interoperability of data



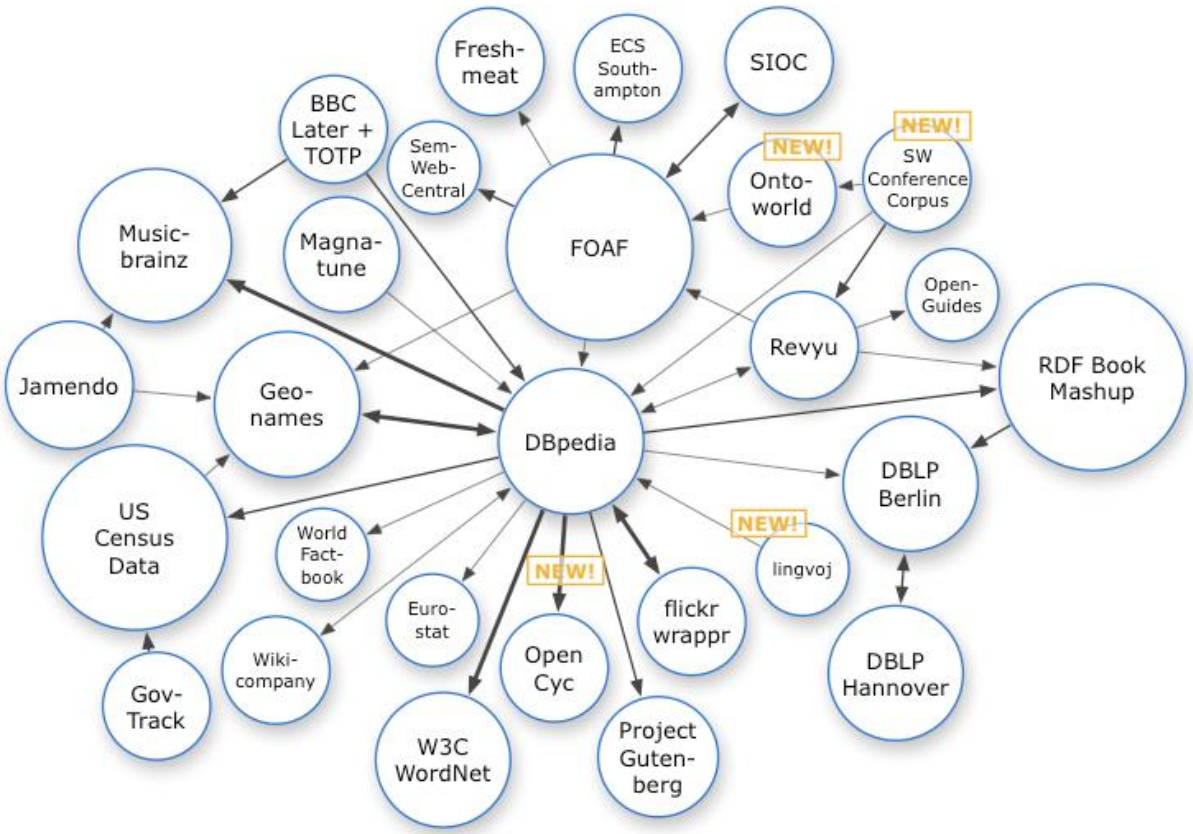
Linked Open Datasets

The LOD cloud

Oct. 2007

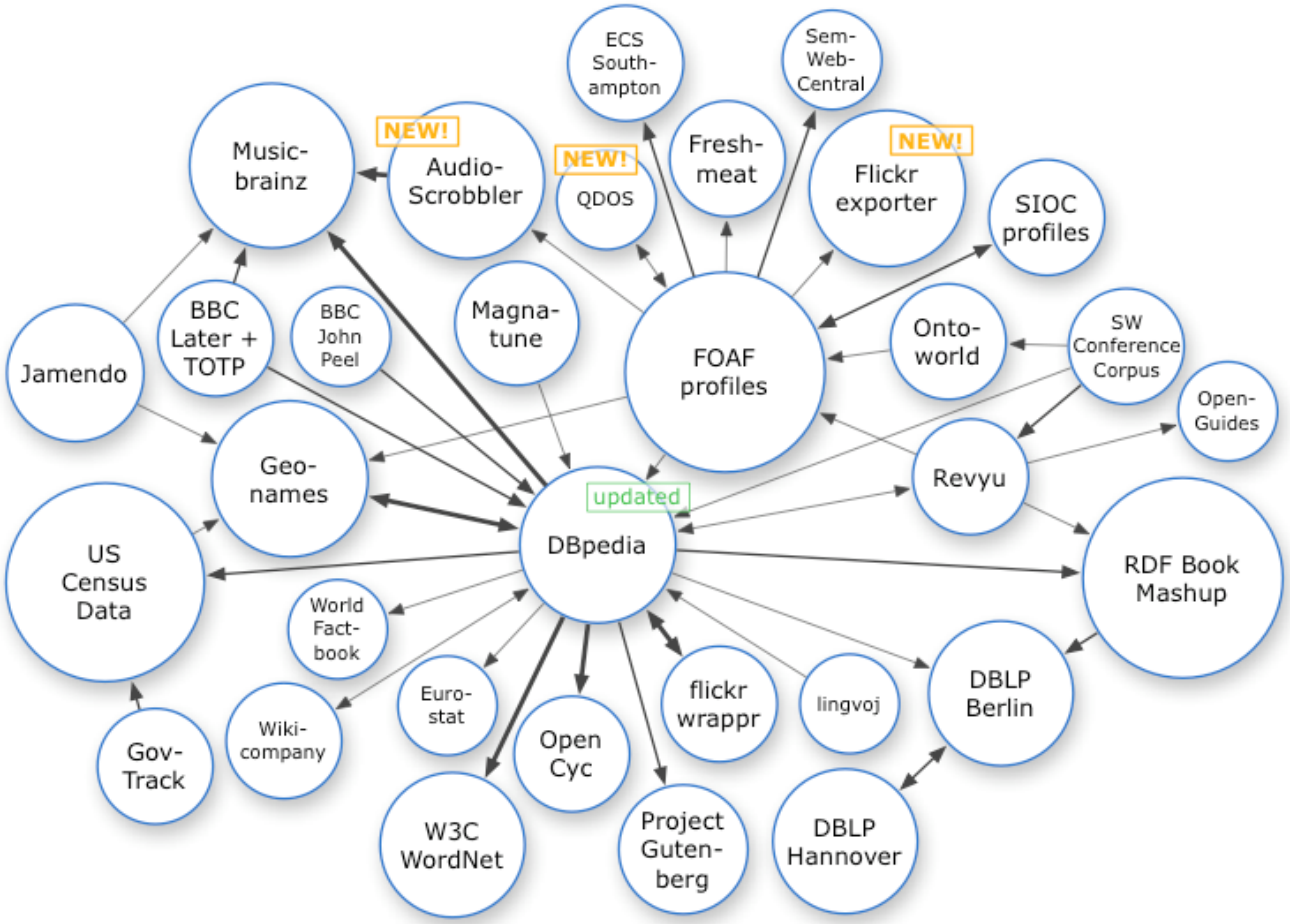


The LOD cloud



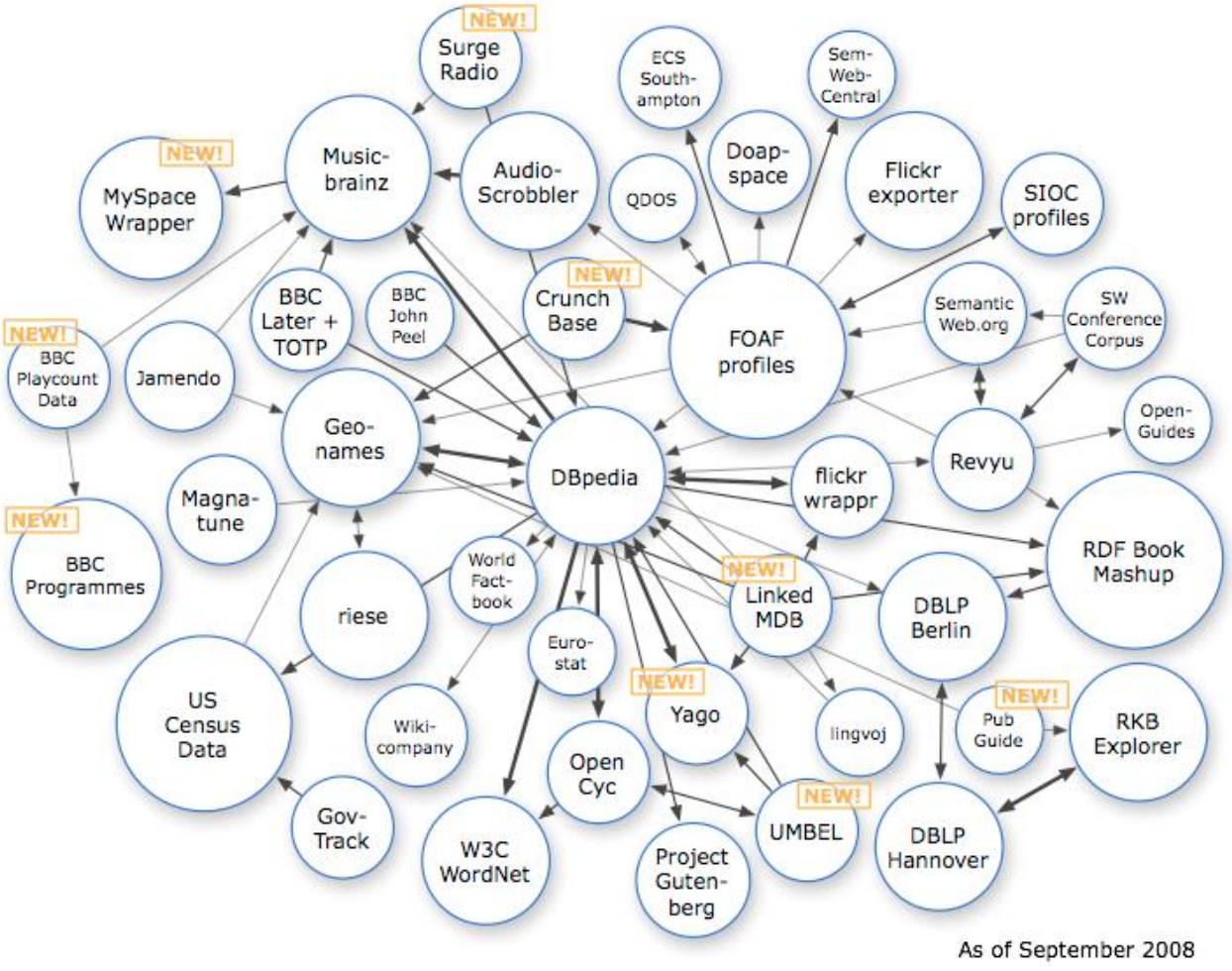
Oct. 2007
Nov. 2007

The LOD cloud



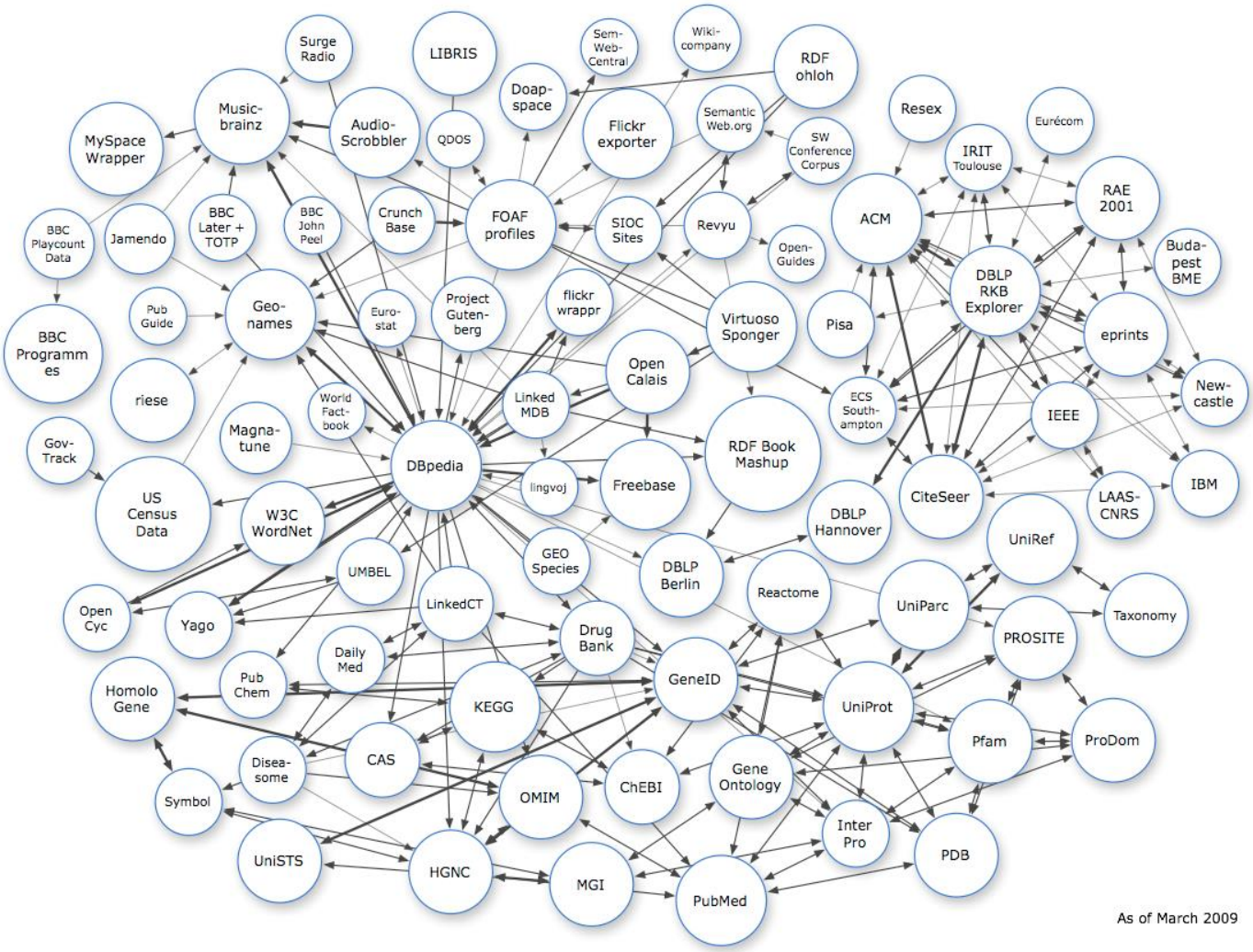
Oct. 2007
Nov. 2007
Feb. 2008

The LOD cloud



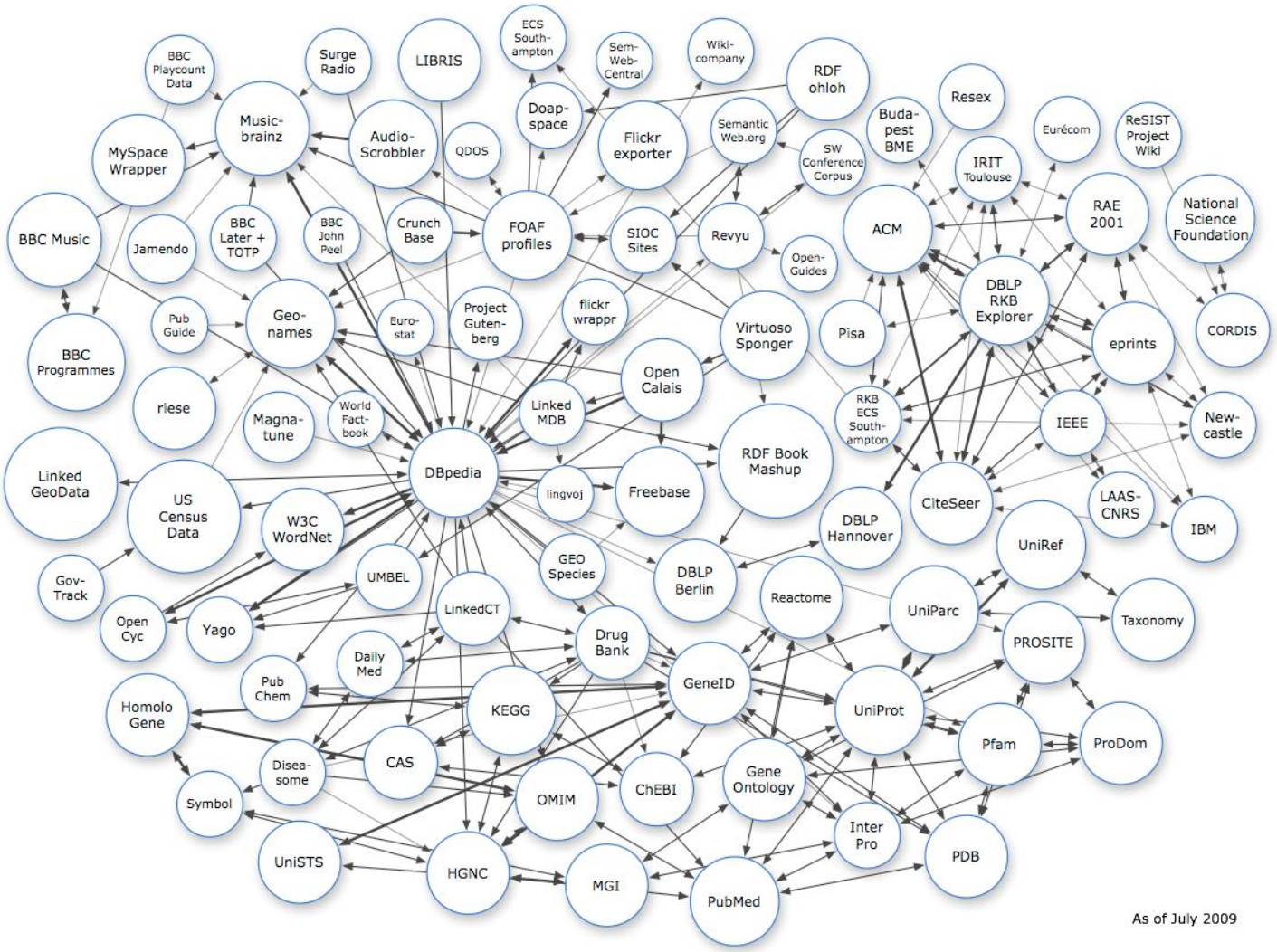
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008

The LOD cloud



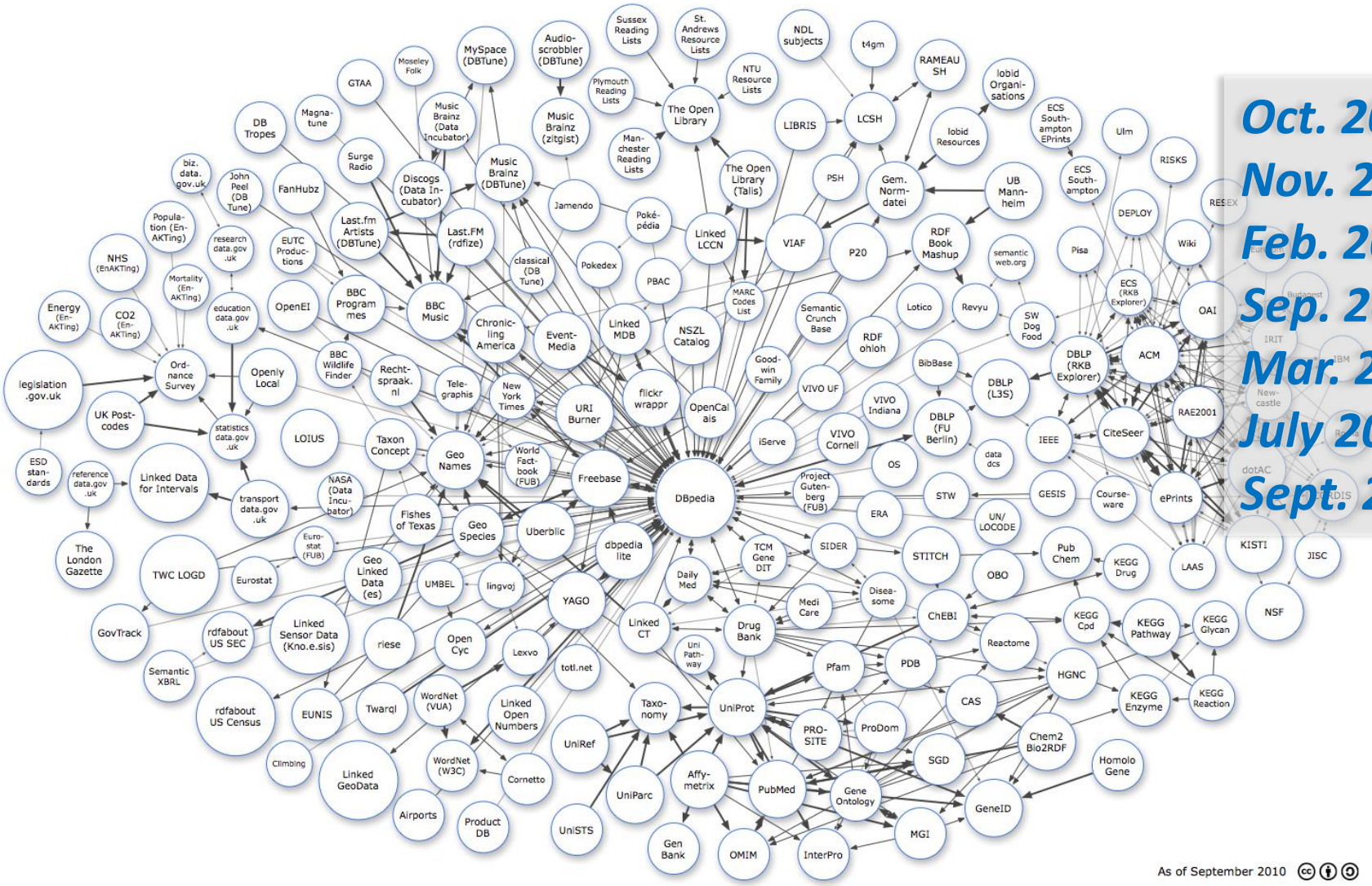
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009

The LOD cloud



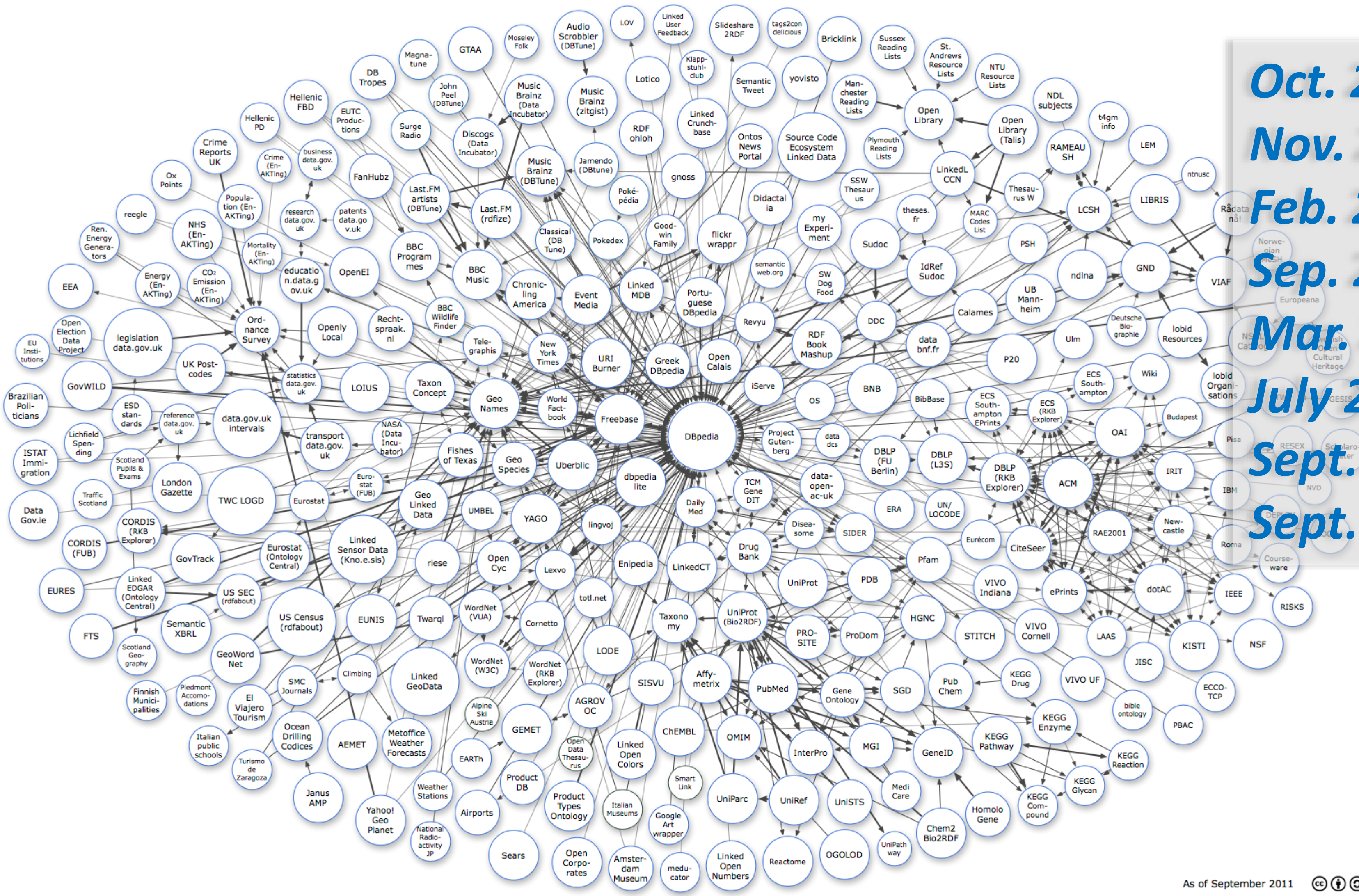
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009

The LOD cloud



As of September 2010

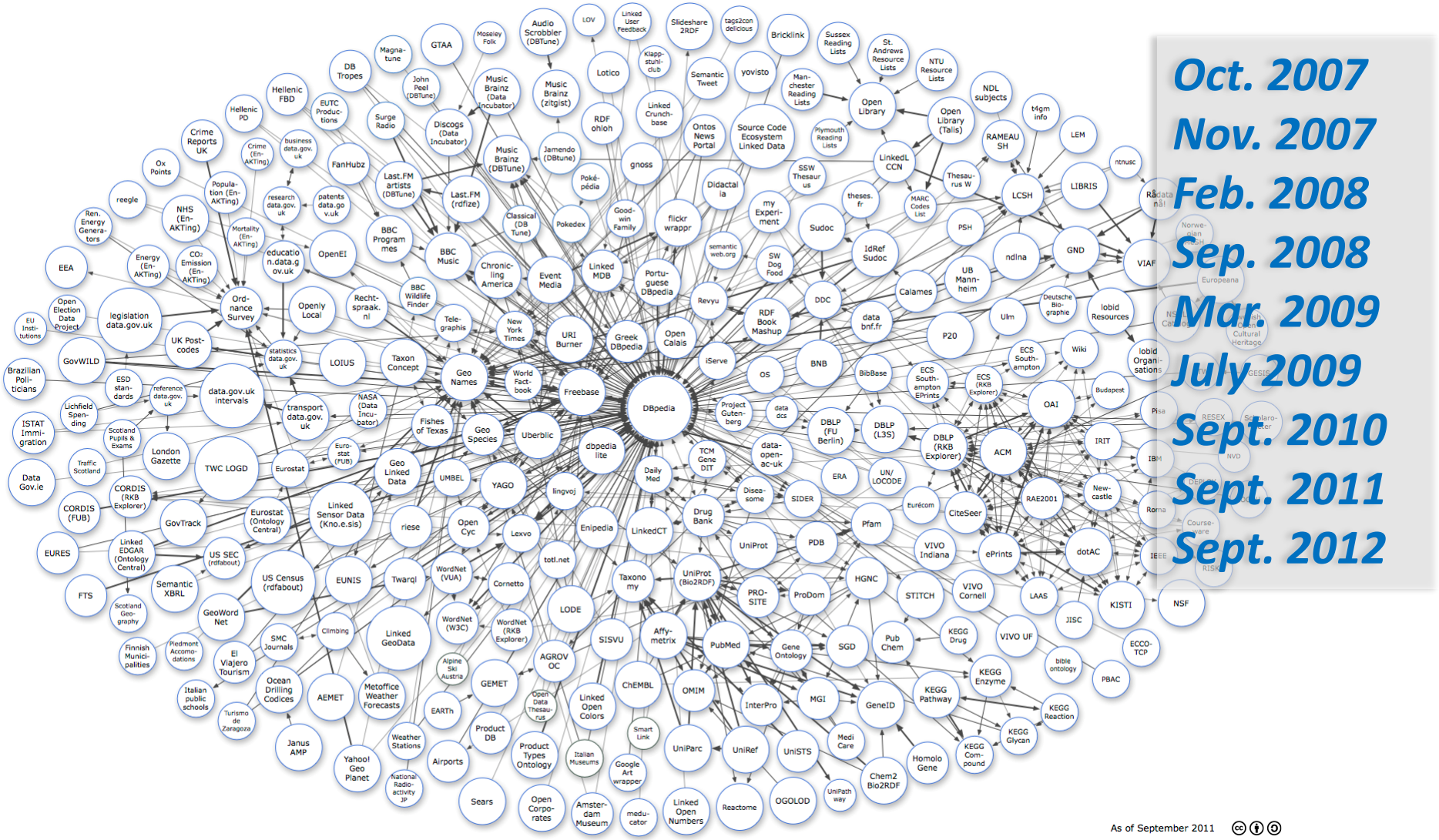
THE LOD CLOUD



Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011

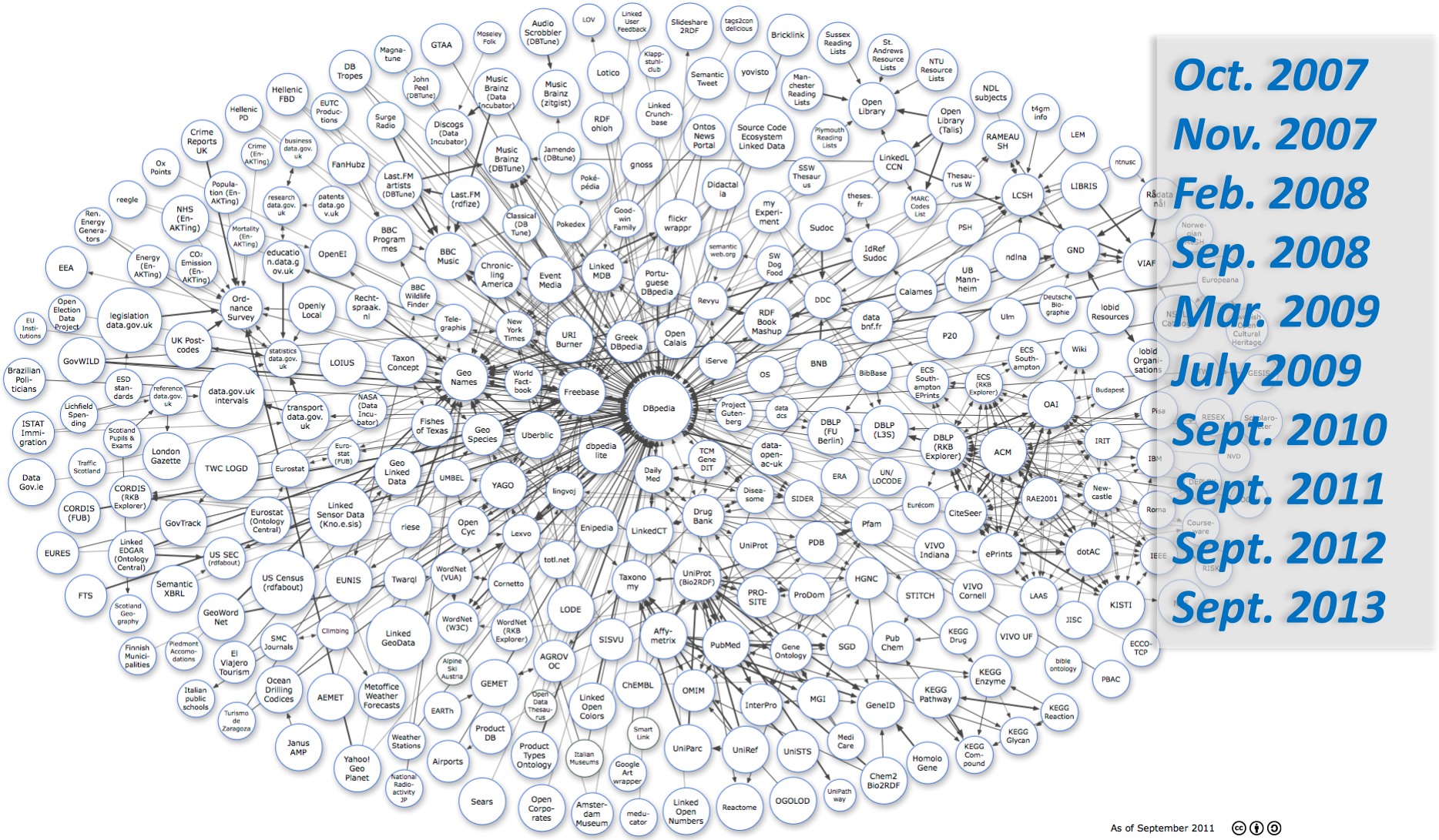
As of September 2011

The LOD cloud



As of September 2011

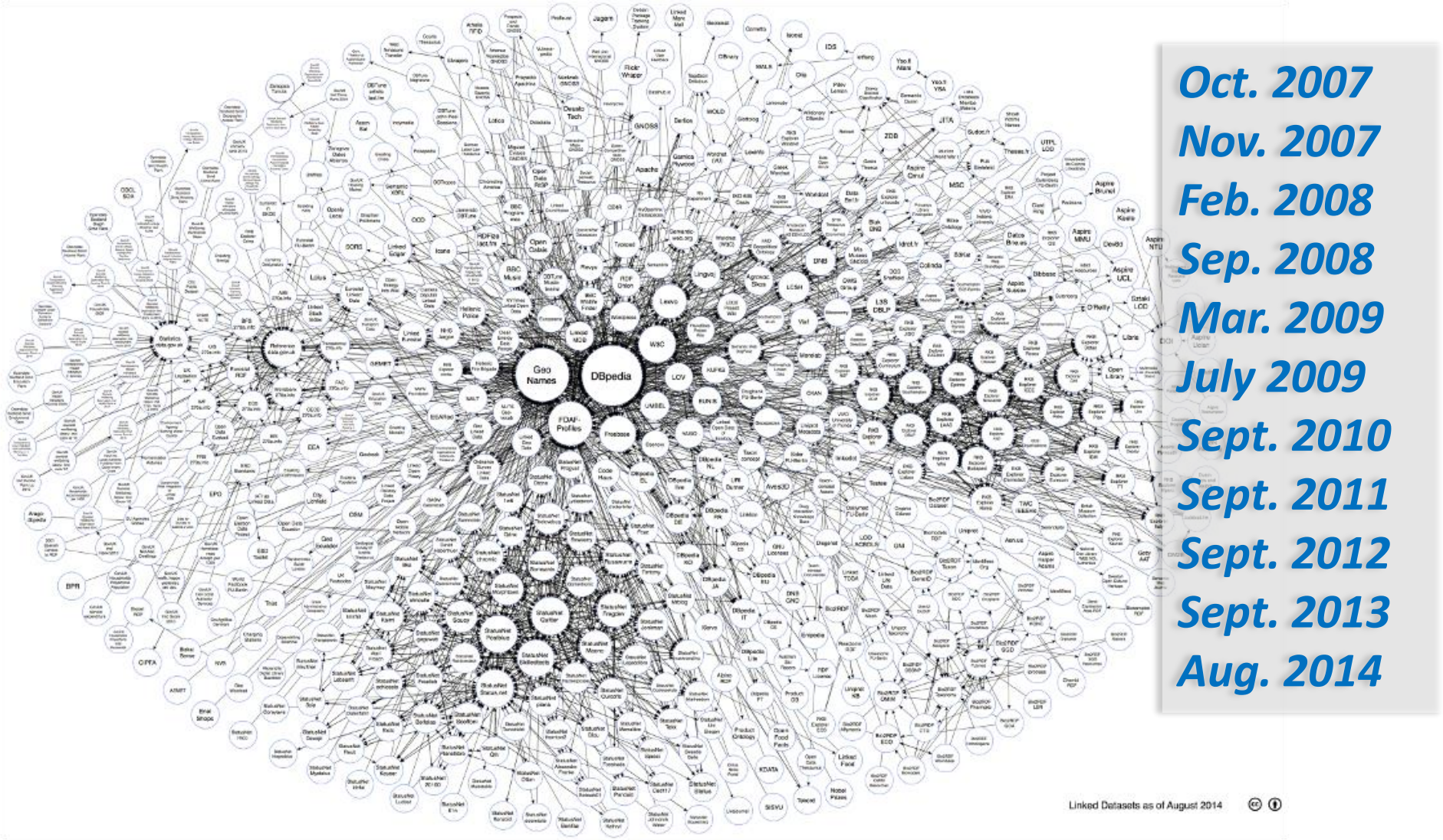
The LOD cloud



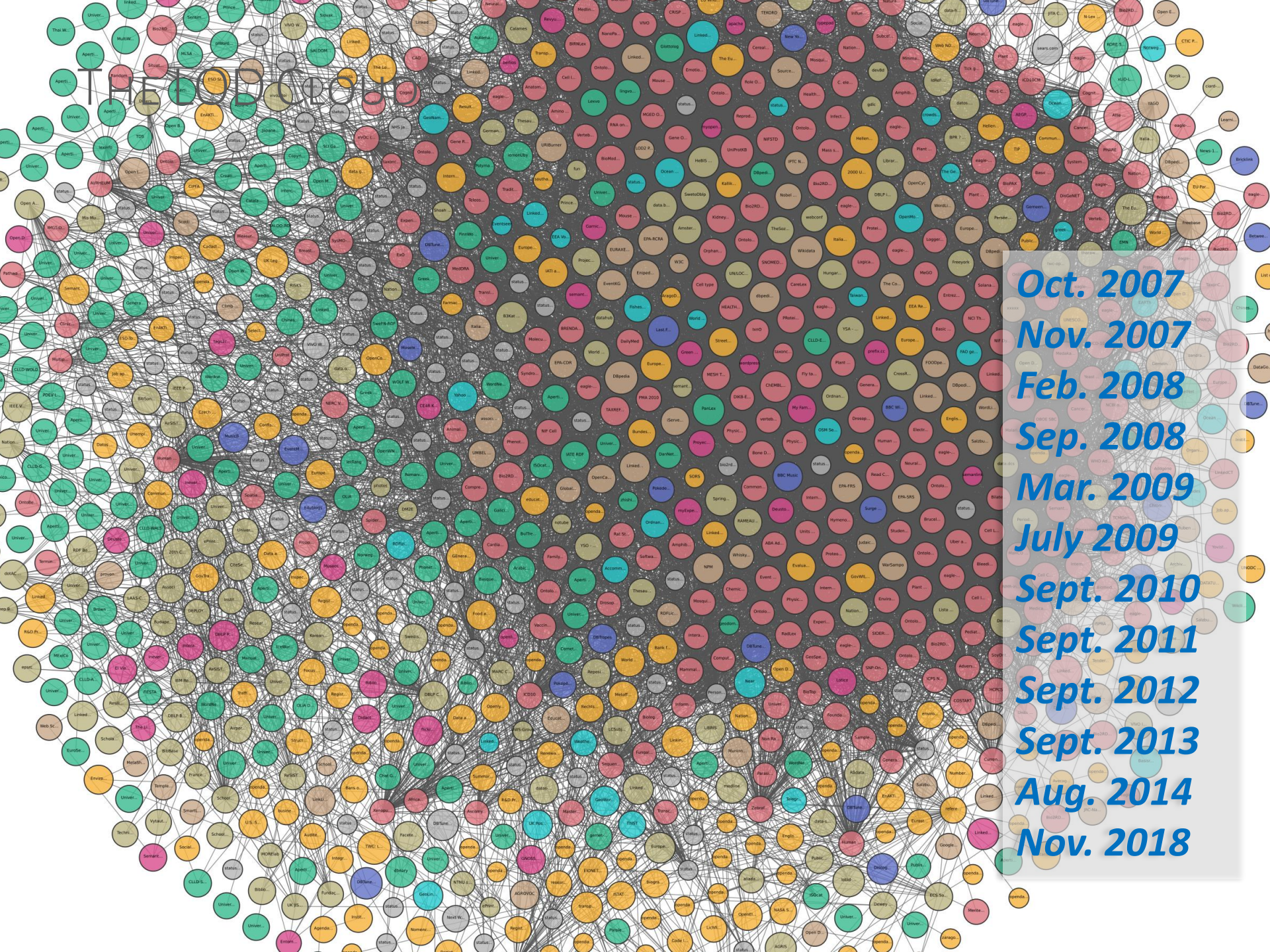
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011
Sept. 2012
Sept. 2013

As of September 2011

The LOD cloud



Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011
Sept. 2012
Sept. 2013
Aug. 2014


THE LOD CLOUD
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011
Sept. 2012
Sept. 2013
Aug. 2014
Nov. 2018

Cross-Domain

Apply

Explore current DBpedia projects and applications. Learn about interlinking and accessing DBpedia.

Develop

Develop amazing things with our DBpedia datasets and our API. Need support? Ask the DBpedia community, they will find a solution to your problem!

Research

Find current research papers, journal articles and conference contributions in this section.

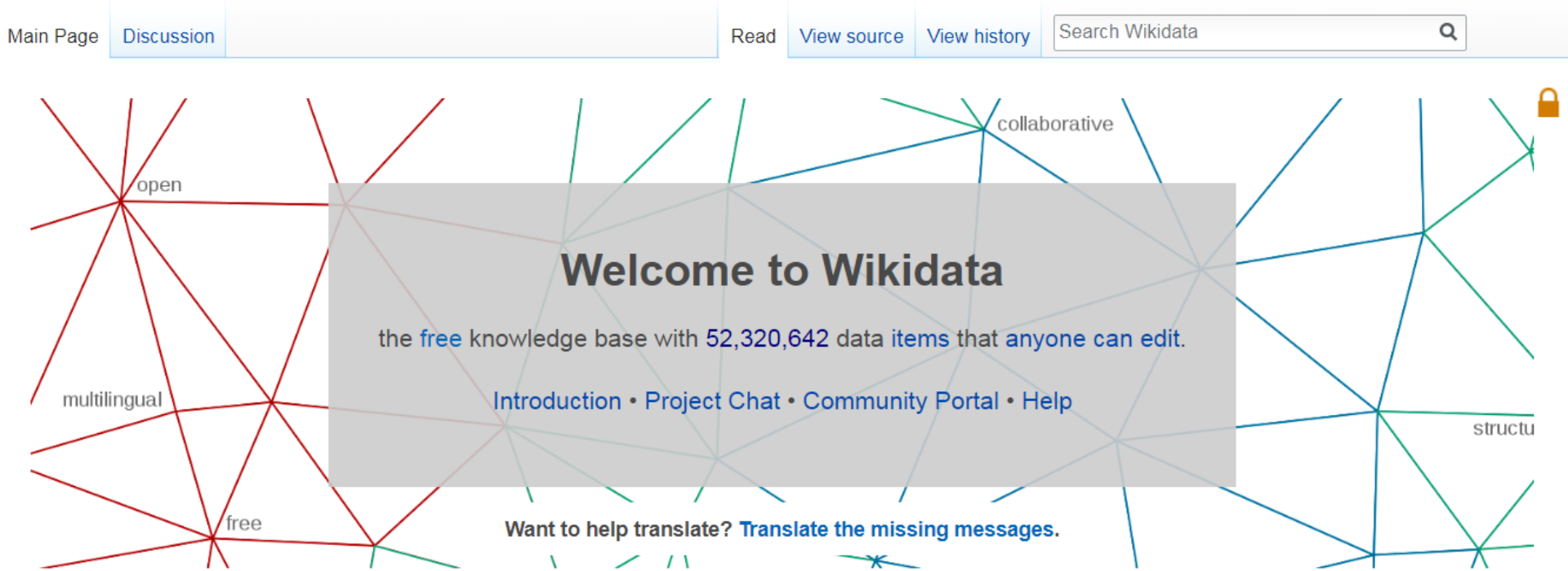
Join

Get in touch with us. Become a member of the DBpedia Association, join our various committees or help to develop a language chapter.

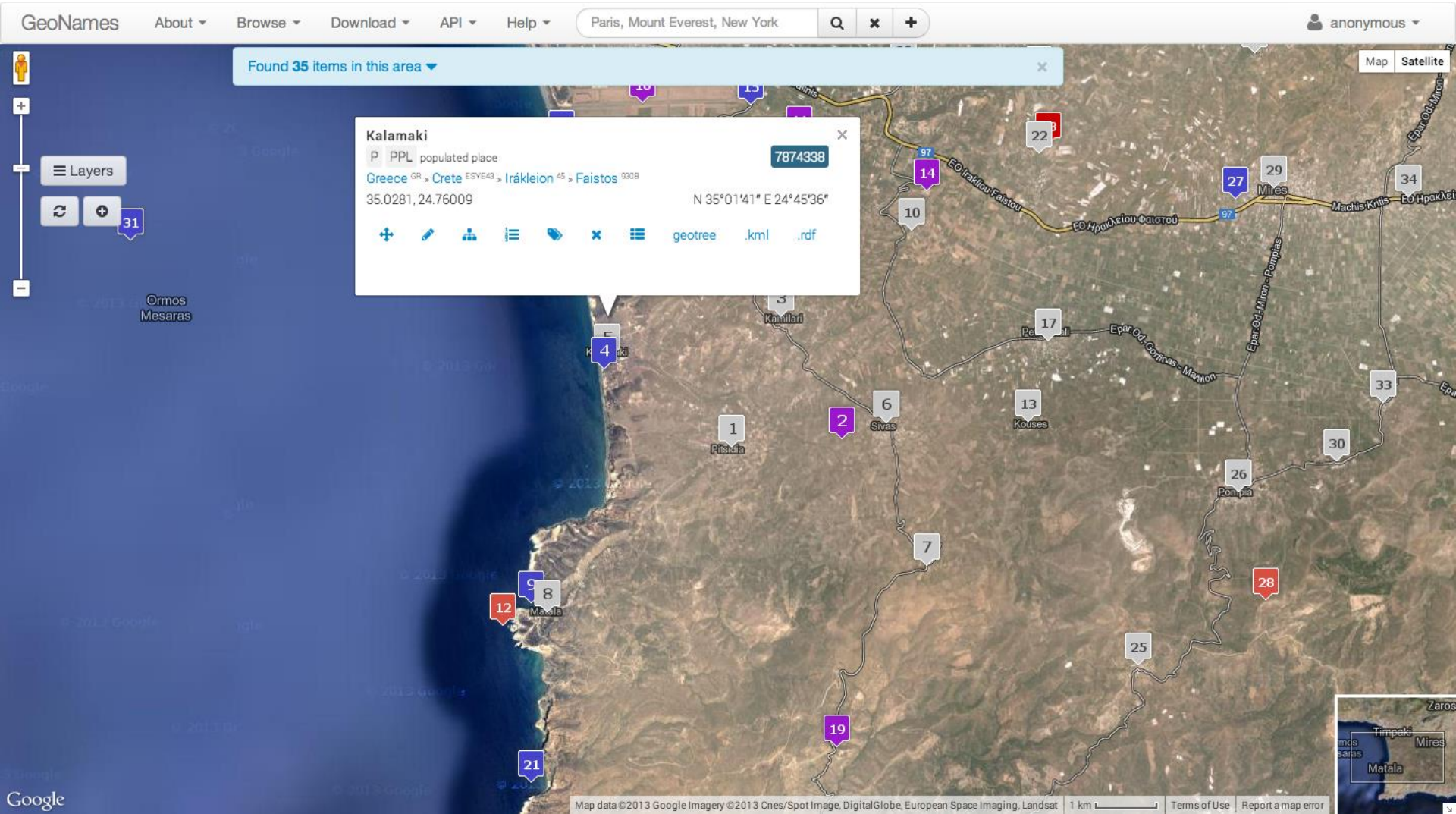
Contribute

Check out DBpedia's career opportunities, our challenges, website contributions, Donation options and many more.

Cross-Domain



Geographic


GeoNames
About
Browse
Download
API
Help
Paris, Mount Everest, New York
anonymous
Found 35 items in this area
Map
Satellite
Layers
Kalamaki
P PPL populated place
7874338
Greece GR » Crete ESYE43 » Irákleion 45 » Faistos 9908
35.0281, 24.76009
N 35°01'41" E 24°45'36"
geotree
.kml
.rdf
Ormos Mesaras
Kamilari
Ptisidia
Sivas
Kouses
Pompia
Mires
Matala
Timpaki
Zaros
Map data ©2013 Google Imagery ©2013 Cnes/Spot Image, DigitalGlobe, European Space Imaging, Landsat
1 km
Terms of Use
Report a map error
Google

Geographic

Linked GeoData

LinkedGeoData.org

Adding a spatial dimension to the Web of Data.

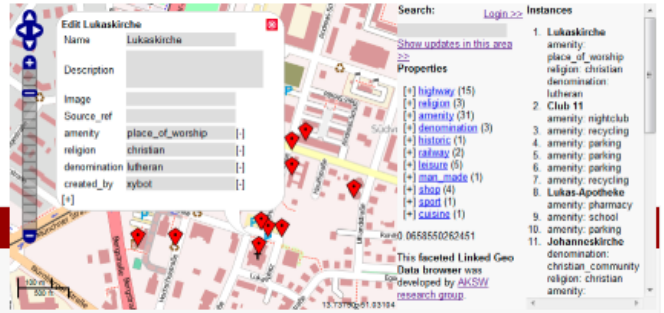
Project created by: AKSW

- About / News
- Downloads
- Online Access
- RDF Mapping
- Use Cases
- LGD Browser
- Publications
- Community
- Blog
- Contact / Imprint

2018 May 7: Linked Data interface operation back to normal

Quick Links: Downloads – SPARQL – Virtual-SPARQL by Sparqlify – HTML interface – Example Queries

LinkedGeoData is an effort to add a spatial dimension to the Web of Data / Semantic Web. LinkedGeoData uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. It interlinks this data with other knowledge bases in the Linking Open Data initiative.



News

LinkedGeoData: New RDF versions of OpenStreetMap datasets available

The AKSW research group is happy to announce that a new LinkedGeoData maintenance release with more than 1.2 billion triples based on the OpenStreetMap planet file from 2015-11-02 is now online. Enjoy! Quick Links Project Website Downloads SPARQL Endpoint Virtual … Continue reading →

AKSW at #ISWC2014. Come and join, talk and discuss with us!

Hello AKSW Follower! We are very pleased to announce that nine of our papers were accepted for presentation at ISWC 2014. In the main track of the conference we will present the following papers: AGDISTIS – Graph-Based Disambiguation of Named … Continue reading →

AKSW at TU Dresden PLT

On June 8, I (Jens) visited the process control engineering research group (PLT) of Leon Urbas at the Dresden University of Technology. We first met on the Leipziger Semantic Web Day where Leon Urbas presented interactive Linked Data applications and … Continue reading →

The LinkedGeoData Knowledge Base

In order to employ the Web as a medium for data and information integration, comprehensive datasets and vocabularies are required as they enable the disambiguation and alignment of other data and information. Many real-life information integration and aggregation tasks are impossible without comprehensive background knowledge related to spatial features of the ways, structures and landscapes surrounding us.

Governmental

This section has been archived and will not be updated any more.

UKGovLD

Submitted by David Buck on Thu, 27/09/2012 - 12:22 | Updated on Tue, 12/03/2013 - 15:37

On the 28th June the Government made a commitment in the Open Data White Paper to establish a new cross-government linked data working group.
The outline was for the UK Government Linked Data Working Group *'to lead the creation and maintenance of the underpinning technologies within the public sector and promote the benefits across the public sector. A key role for the group will be to work with data owners, data users and bodies such as the W3C Government Linked Data Working Group, to promote and set standards for the adoption of common URIs across government. This provision of a core of authoritative identifiers (for instance for businesses, contracts, postcodes and geo-spatial entities such as roads and bus stops) will be the key to connecting data across the information economy and allowing businesses to add value and to exchange information reliably in the digital world.'*

To establish the group a Quick Start Team was formed the first step being to draft terms of reference. This was ratified by vote at the first working group event.

UK GOVERNMENT LINKED DATA WORKING GROUP

The terms of reference for the UK Government Linked Data Working Group (UKGovLD) are available in a number of formats.

pdf - UKGovLD Terms of Reference http://data.gov.uk/sites/default/files/UKGovLDDraftTermsofReference.pdf
odt - UKGovLD Terms of Reference http://data.gov.uk/sites/default/files/UKGovLDDraftTermsofReference.odt
doc - UKGovLD Terms of Reference http://data.gov.uk/sites/default/files/UKGovLDDraftTermsofReference.doc

Membership

10 SECOND TOUR

Overview of Linked Data
Across government over the last ten years there has been a growing realisation to the power of linked data for exposing, sharing, and connecting pieces of data and information using uniform resource identifiers (URIs).

What is Linked Data?
Linked Data is data in which real-world things are given addresses on the web (URIs), and data is published about them in machine-readable formats.

List of Linked Datasets & Vocabularies
Linked data to explore, use and build other data on.

Governmental

datos.gob.es
reutiliza la información pública

Español

INICIO | INICIATIVA APORTA | CATÁLOGO DE DATOS | IMPACTO | INTERACTÚA | ACTUALIDAD

Inicio | linked data

linked data

La Biblioteca Nacional de España pone en marcha una nueva versión de su portal de datos abiertos

02-08-2018

La Biblioteca Nacional Española (BNE) continúa impulsando la difusión y reutilización de sus fondos documentales. Además de contribuir a la conservación del patrimonio cultural que custodia, a través de la digitalización y la preservación digital de sus...

El derecho a la tierra y el movimiento abierto: la Fundación Land Portal

24-01-2018

La alta diplomacia desempeña tradicionalmente un papel en ayudar a culturas y naciones a dialogar entre sí. Pero cuando se trata de fortalecer el derecho a la tierra, son las propias comunidades locales quienes tienen que involucrarse. Ésta es la...

Pubby y LODI, abriendo los datos enlazados a los humanos

31-01-2018

Una parte importante de los datos que están publicados bajo las premisas de la Web Semántica, donde los recursos están identificados por

Life Sciences

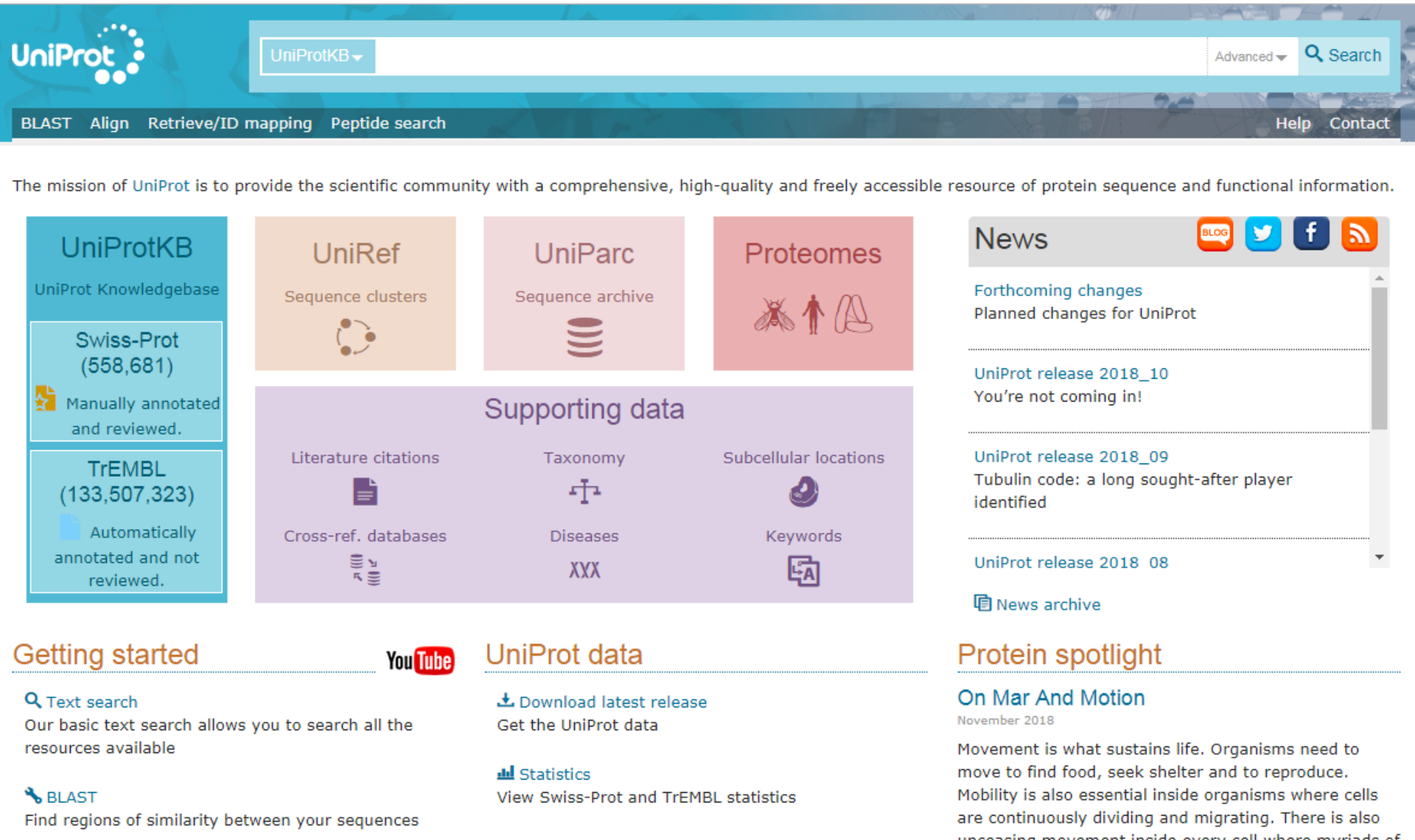
Select an ID example ▾ Submit

about datasets download API SPARQL query repository mailing list

Powered by Docker, PHP, and Virtuoso Open-Source Edition

Maastricht University UNIVERSITÉ LAVAL

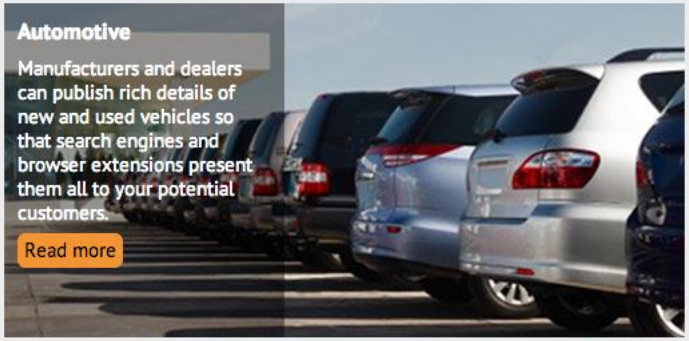
Life Sciences



E-Commerce

GoodRelations
The Web Vocabulary for E-Commerce

Main | Quickstart | Cookbook | User's Guide | Specification | Tools | Community



<<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

Who uses GoodRelations?

Google
Yahoo!
BestBuy
sears.com
kmart.com

... and 10,000 more

See here for additional references.

News from Twitter

Contact

Univ.-Prof. Dr. Martin Hepp

License

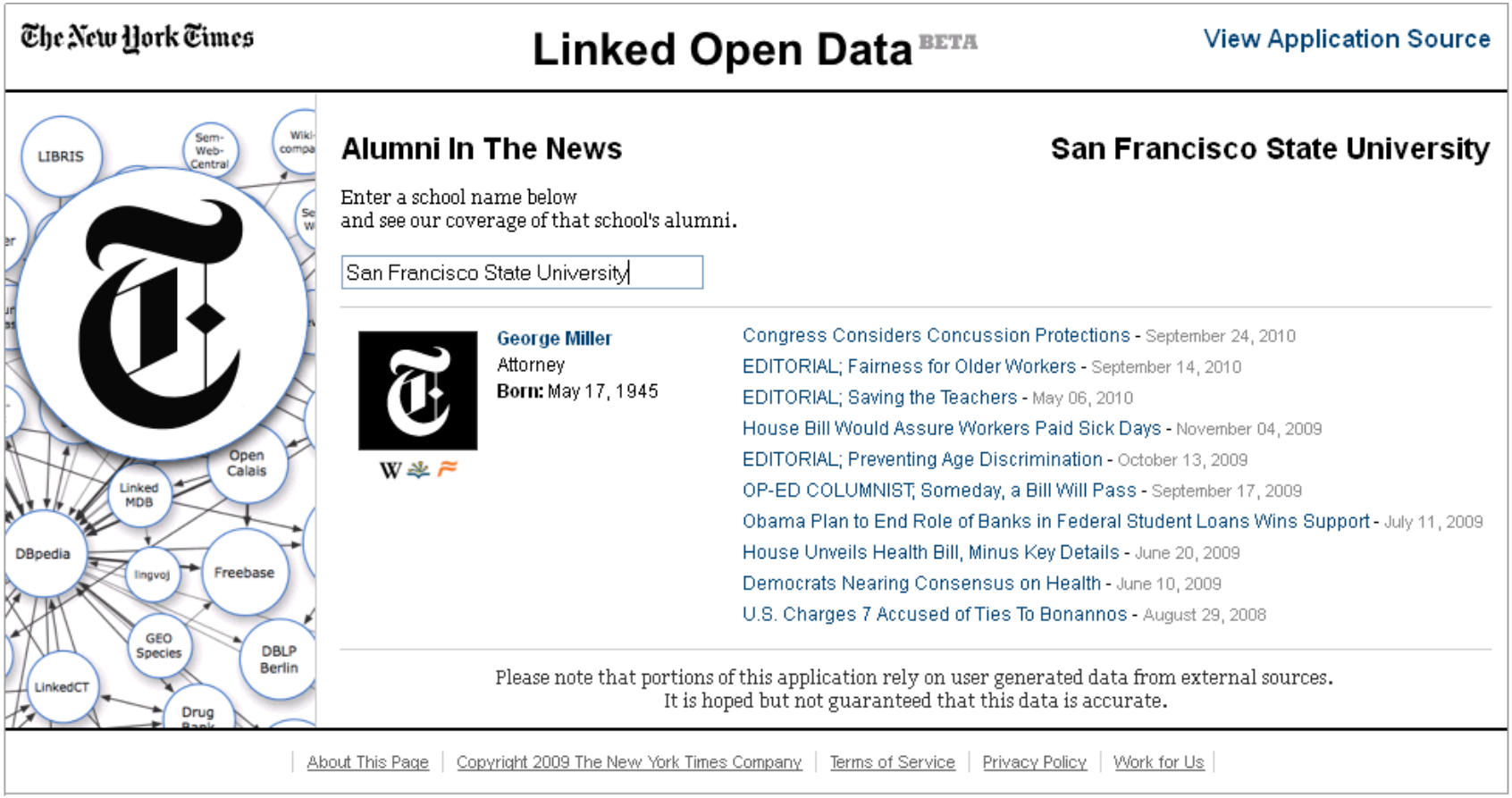
The GoodRelations ontology is creative

Acknowledgments

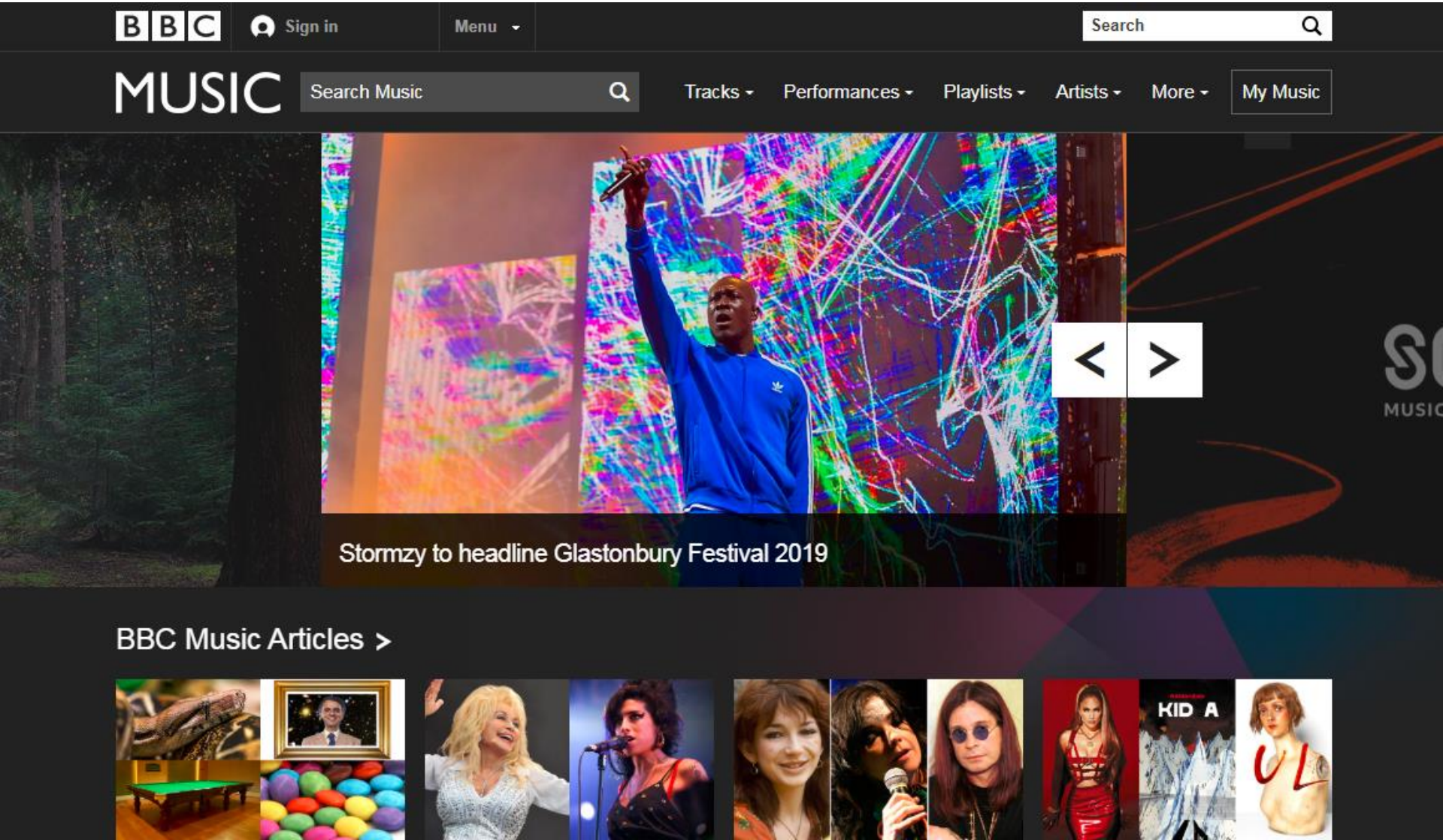
Many organizations and individuals have supported

Feedback

Media


The New York Times
Linked Open Data BETA
View Application Source
LIBRIS
Sem-Web-Central
Open Calais
Linked MDB
DBpedia
lingvoj
Freebase
GEO Species
DBLP Berlin
LinkedCT
Drug
Alumni In The News
San Francisco State University
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

Media


BBC
Sign in
Menu
Search
MUSIC
Search Music
Tracks
Performances
Playlists
Artists
More
My Music
Stormzy to headline Glastonbury Festival 2019
BBC Music Articles >
KID A

Pokémon

Bienvenue sur Poképédia

L'encyclopédie Pokémon à laquelle tout le monde peut participer !

[Deutsch | English | Español | Italiano | 日本語 | 中文]

22 683 ARTICLES EN FRANÇAIS*

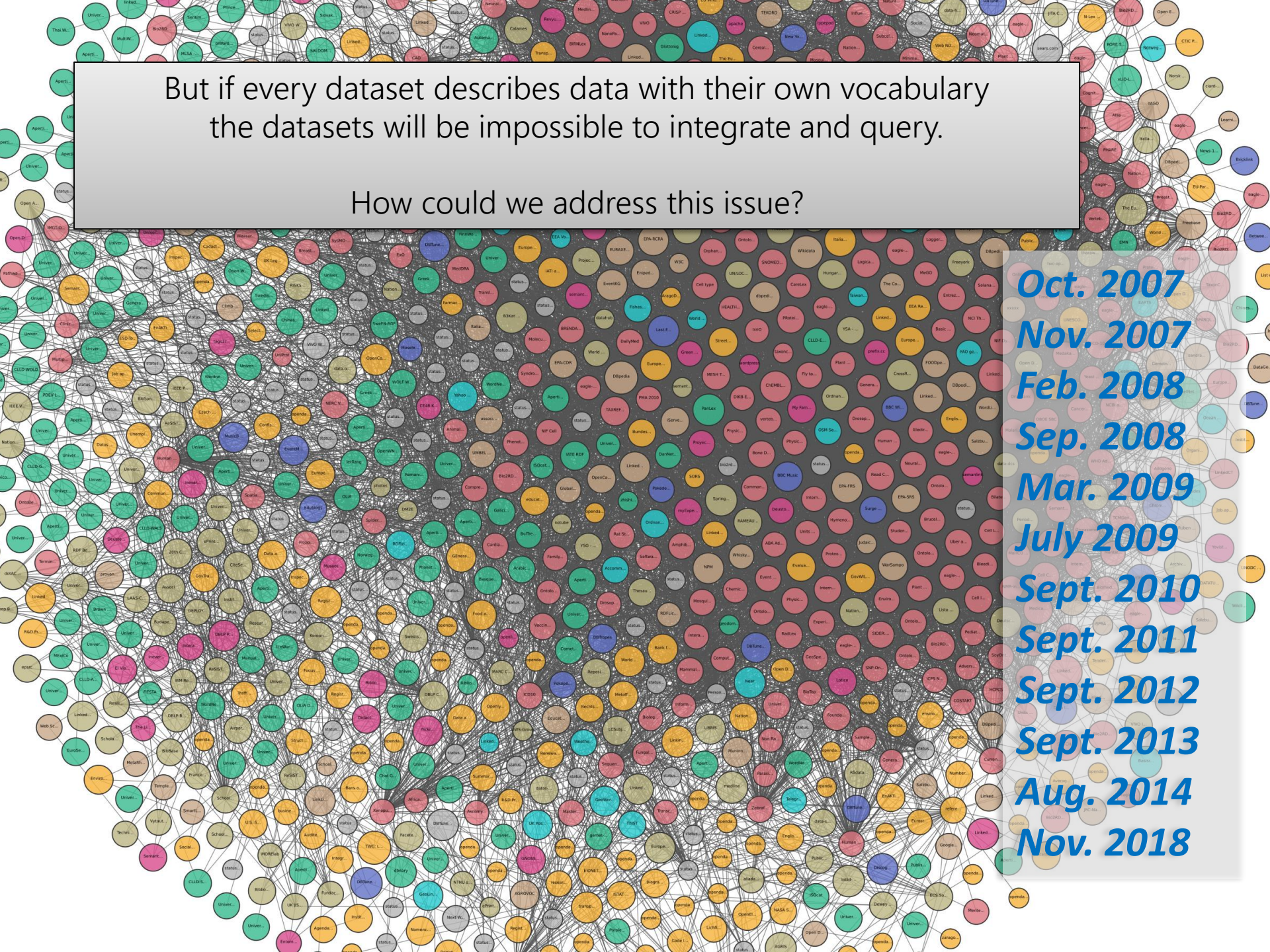
INDEX DES ARTICLES DANS L'ORDRE ALPHABÉTIQUE

INDEX DES CATÉGORIES

LUNDI 19 NOVEMBRE 2018

AIDE · PREMIERS PAS · À PROPOS

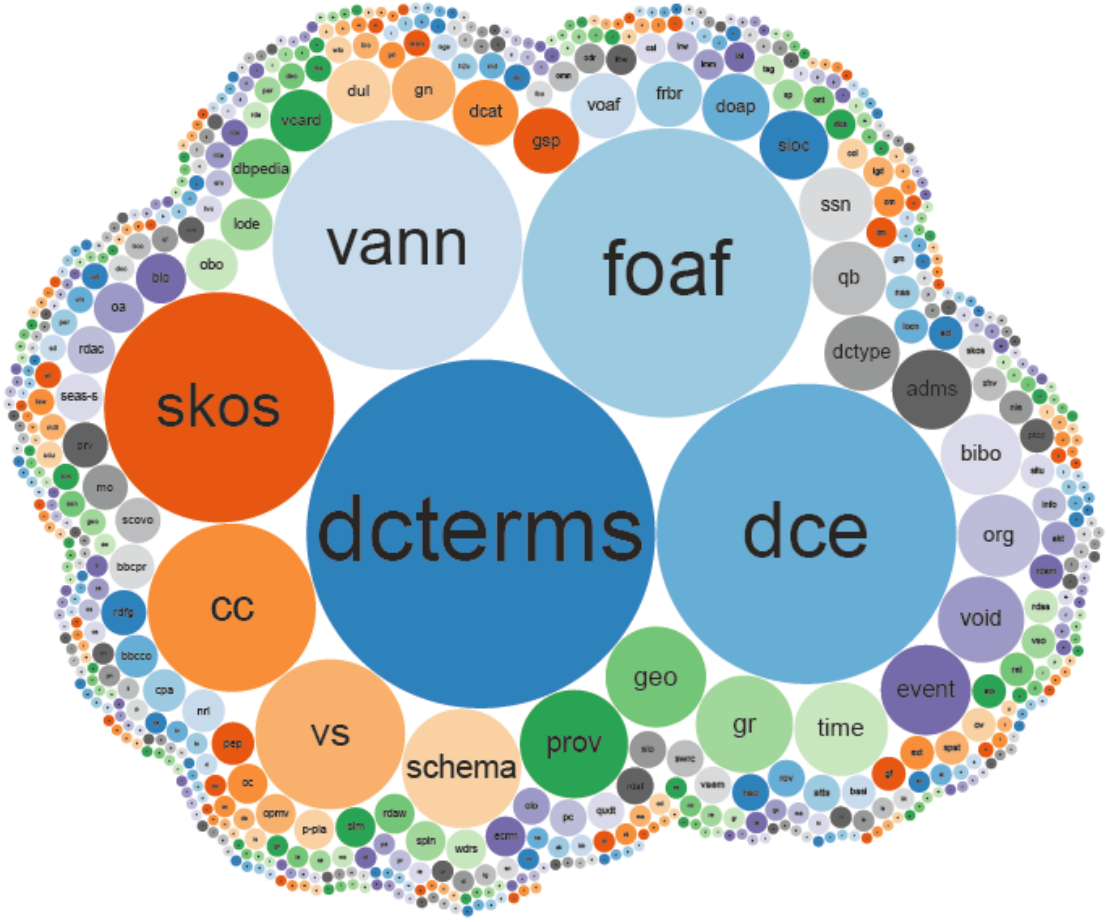
PRINCIPES FONDATEURS · RÈGLES · CONVENTIONS


But if every dataset describes data with their own vocabulary
the datasets will be impossible to integrate and query.
How could we address this issue?
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011
Sept. 2012
Sept. 2013
Aug. 2014
Nov. 2018

Linked Open Vocabularies

Linked Open Vocabularies

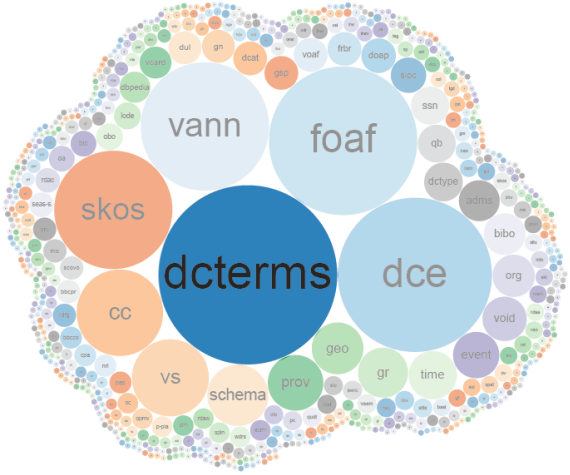
- Indexes vocabularies for re-use



From https://lov.linkeddata.es/

DCTERMS

- Describes documents



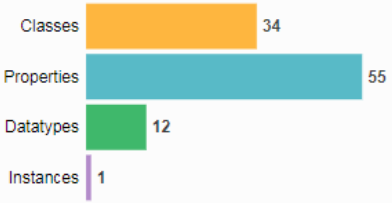
DCMI Metadata Terms (dcterms)

Metadata

URI	http://purl.org/dc/terms/
Namespace	http://purl.org/dc/terms/
homepage	http://dublincore.org/documents/dcmi-terms
Description	an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes. @en
Language	English en
Creator	Dublin Core Metadata Initiative http://purl.org/dc/aboutdcmi#DCMI
Publisher	Dublin Core Metadata Initiative http://purl.org/dc/aboutdcmi#DCMI

n3

Statistics



Expressivity

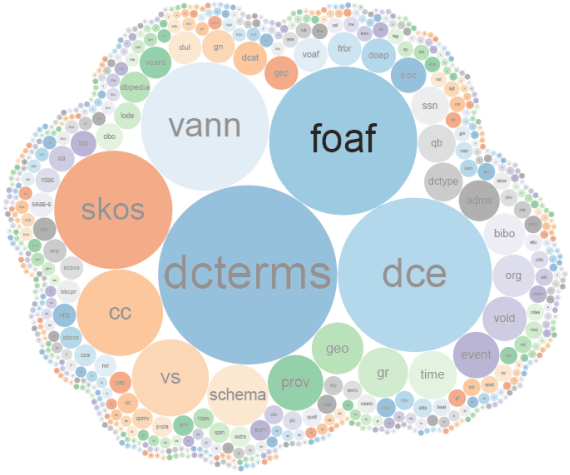
RDF RDFS OWL

Tags

Metadata

FOAF

- Describes people



Friend of a Friend vocabulary (foaf)

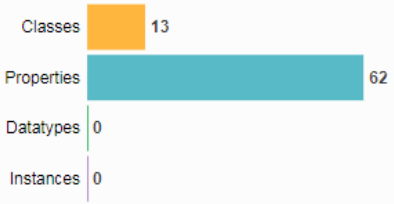
Metadata

URI	http://xmlns.com/foaf/0.1/
Namespace	http://xmlns.com/foaf/0.1/
homepage	http://www.foaf-project.org/
Description	FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. @en
Language	
Creator	Libby Miller http://data.semanticweb.org/person/libby-miller; Dan Brickley http://google.com/+DanBrickley
Publisher	Dan Brickley http://google.com/+DanBrickley

(2013-06-04) Bernard Vatant: From the specification : "FOAF has been evolving gradually since its creation in mid-

n3

Statistics



Expressivity

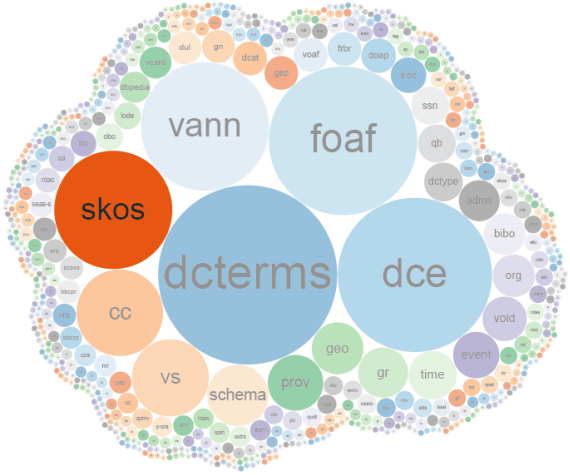
RDF RDFS

Tags

People

SKOS

- Describes taxonomies



Simple Knowledge Organization System (skos)

Metadata

URI	http://www.w3.org/2004/02/skos/core
Namespace	http://www.w3.org/2004/02/skos/core#
isDefinedBy	http://www.w3.org/2009/08/skos-reference/skos.rdf
homepage	http://www.w3.org/2009/08/skos-reference/skos.html
Description	The Simple Knowledge Organization System (SKOS) is a common data model for sharing and linking knowledge organization systems via the Semantic Web. @en
Language	English en
Creator	Alistair Miles http://purl.net/aliman Sean Bechhofer https://plus.google.com/117822622810723317855

n3

Statistics

Classes	4
Properties	28
Datatypes	0
Instances	0

Expressivity

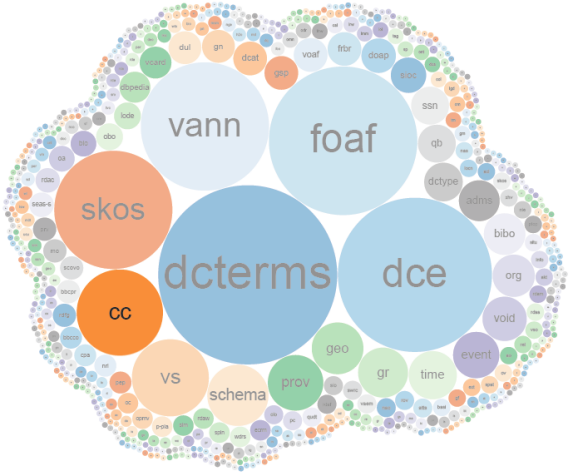
RDF RDFS

Tags

W3C Rec

CC

- Describes licencing



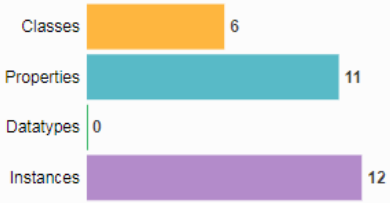
Creative Commons Rights Expression Language (cc)

Metadata

URI	http://creativecommons.org/ns
Namespace	http://creativecommons.org/ns#
isDefinedBy	http://creativecommons.org/schema.rdf
Description	The Creative Commons Rights Expression Language (CC REL) lets you describe copyright licenses in RDF @en
Language	English en
Publisher	Creative Commons http://dbpedia.org/resource/Creative_Commons
Comment	(2014-11-05) Ghislain Atemezing: Annual review OK. (2012-01-18) Bernard Vatant: This vocabulary is still referenced by many other vocabularies through its historical URI http://web.resource.org/cc/. This URI is redirected, but should not be used anymore. (2013-10-10) Bernard Vatant: A de facto standard for representation of rights. Used so far in the metadata of less than

n3

Statistics



Expressivity

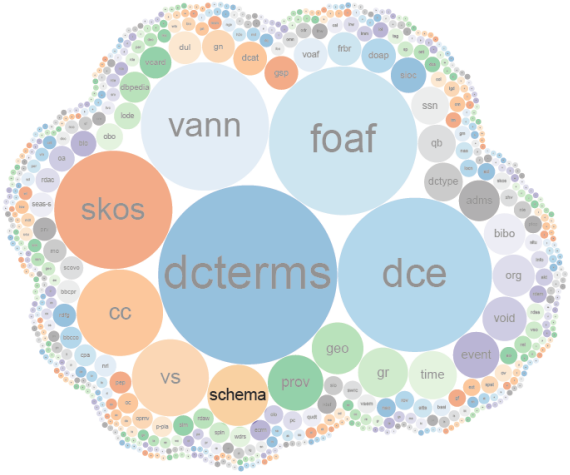
RDF RDFS OWL

Tags

Metadata

SCHEMA

- Describes everything



Schema.org vocabulary (schema)

Metadata

URI	http://schema.org/
Namespace	http://schema.org/
isDefinedBy	http://www.w3.org/2012/pyRdfa/extract?uri=http%3A%2F%2Fschema.org%2Fdocs%2Fschema_org_rdfa.html&format=n3
homepage	https://schema.org/docs/about.html
Description	Search engines including Bing, Google, Yahoo! and Yandex rely on schema.org markup to improve the display of search results, making it easier for people to find the right web pages. @en
Language	
Contributor	Google http://dbpedia.org/resource/Google; Microsoft Corporation http://dbpedia.org/resource/Microsoft; Yahoo!, Inc. http://dbpedia.org/resource/Yahoo!; Dan Brickley http://google.com/+DanBrickley

n3

Statistics

Classes	593
Properties	846
Datatypes	0
Instances	123

Expressivity

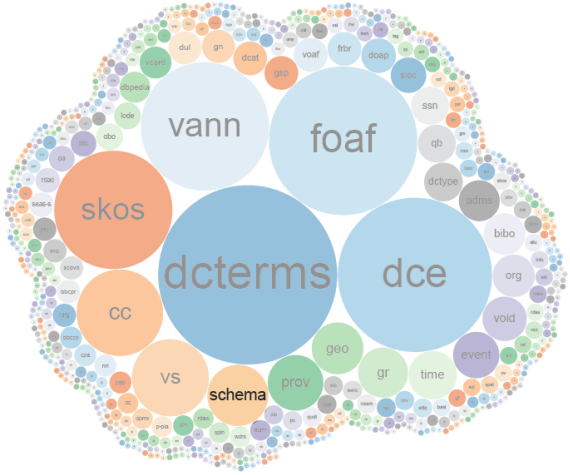
RDF RDFS

Tags

General & Upper

SCHEMA

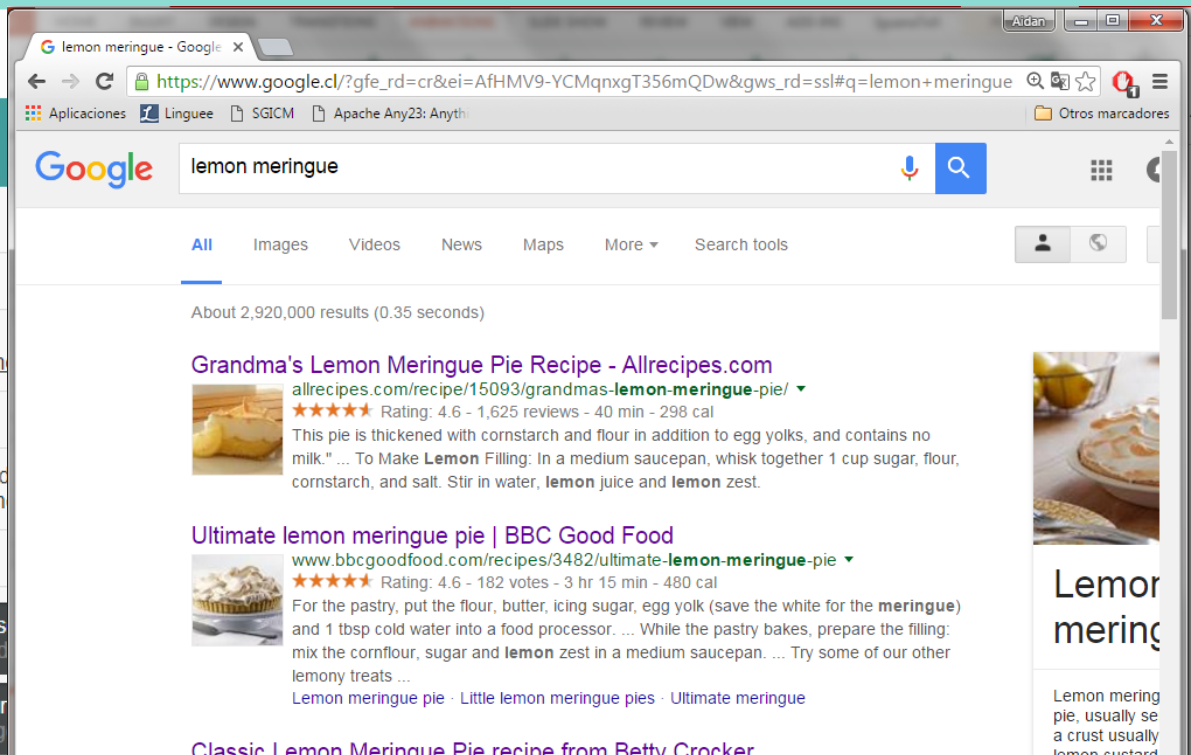
- Describes everything



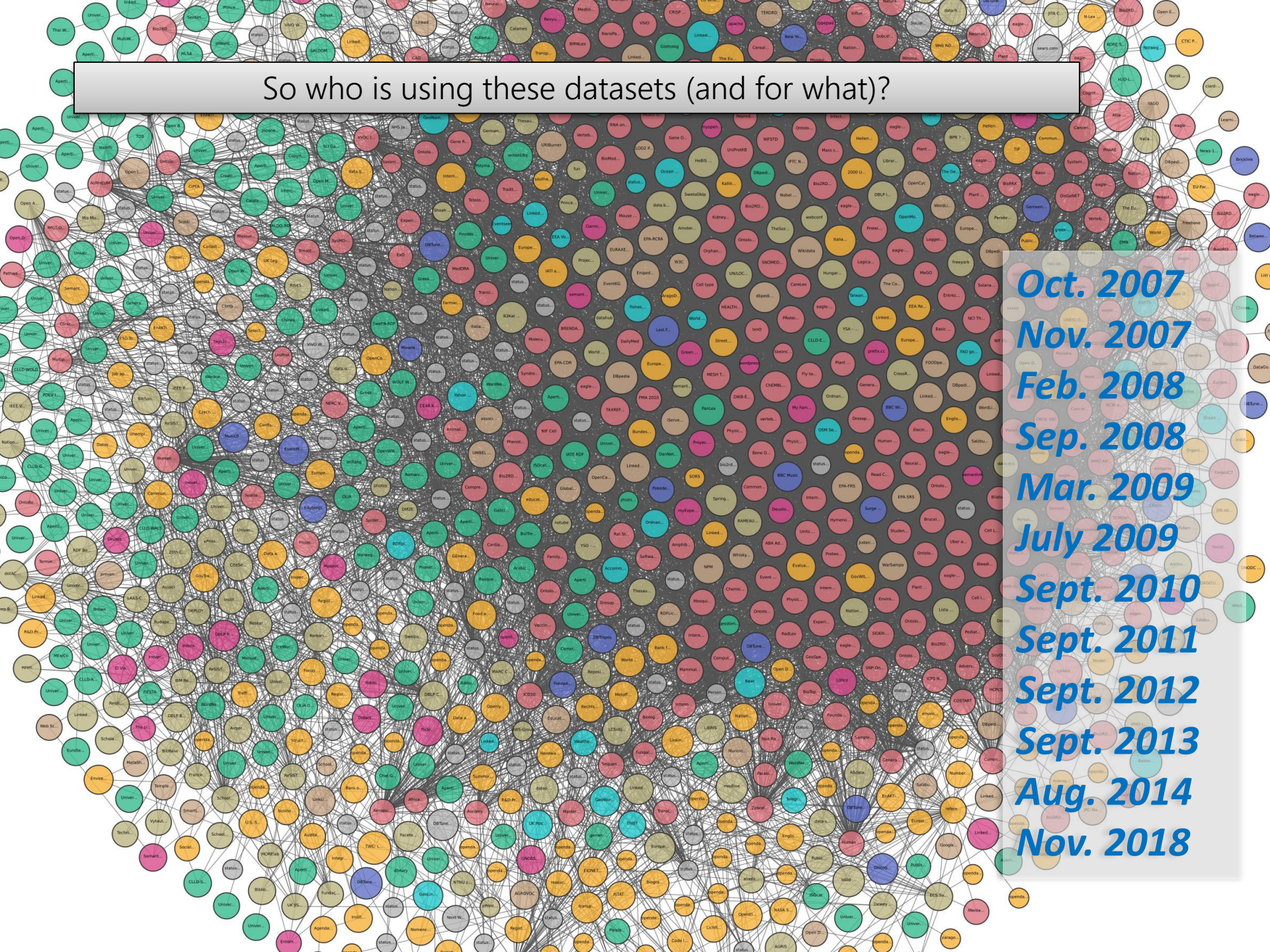
Schema.org vocabulary (schema)

Metadata

URI	http://schema.org/
Namespace	http://schema.org/
isDefinedBy	http://www.w3.org/2012/pyRdfa/extract?uri=http%3A%2F%2Fschema.org%2Fdocs%2Fsch
homepage	https://schema.org/docs/about.html
Description	Search engines including Bing, Google, Yahoo! and search results, making it easier for people to find th
Language	
Contributor	Google http://dbpedia.org/resource/Google Micros http://d Yahoo!, Inc. http://dbpedia.org/resource/Yahoo! Dan Br http://g

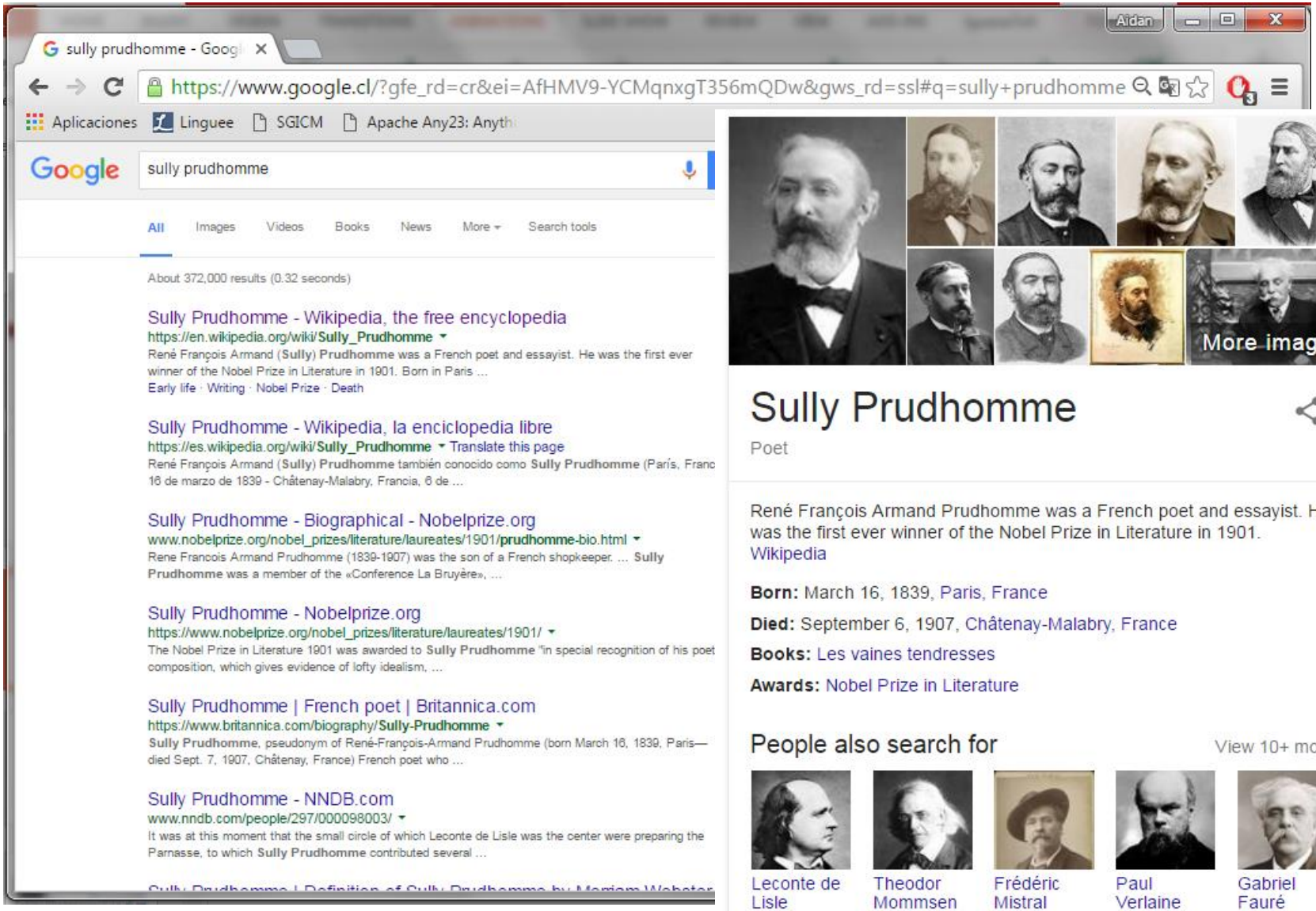


Linked Data Applications


So who is using these datasets (and for what)?
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011
Sept. 2012
Sept. 2013
Aug. 2014
Nov. 2018

Google's Knowledge Graph

WIKIDATA



Apple's Siri

WIKIDATA

Siri Erroneously Told People Stan Lee Was Dead

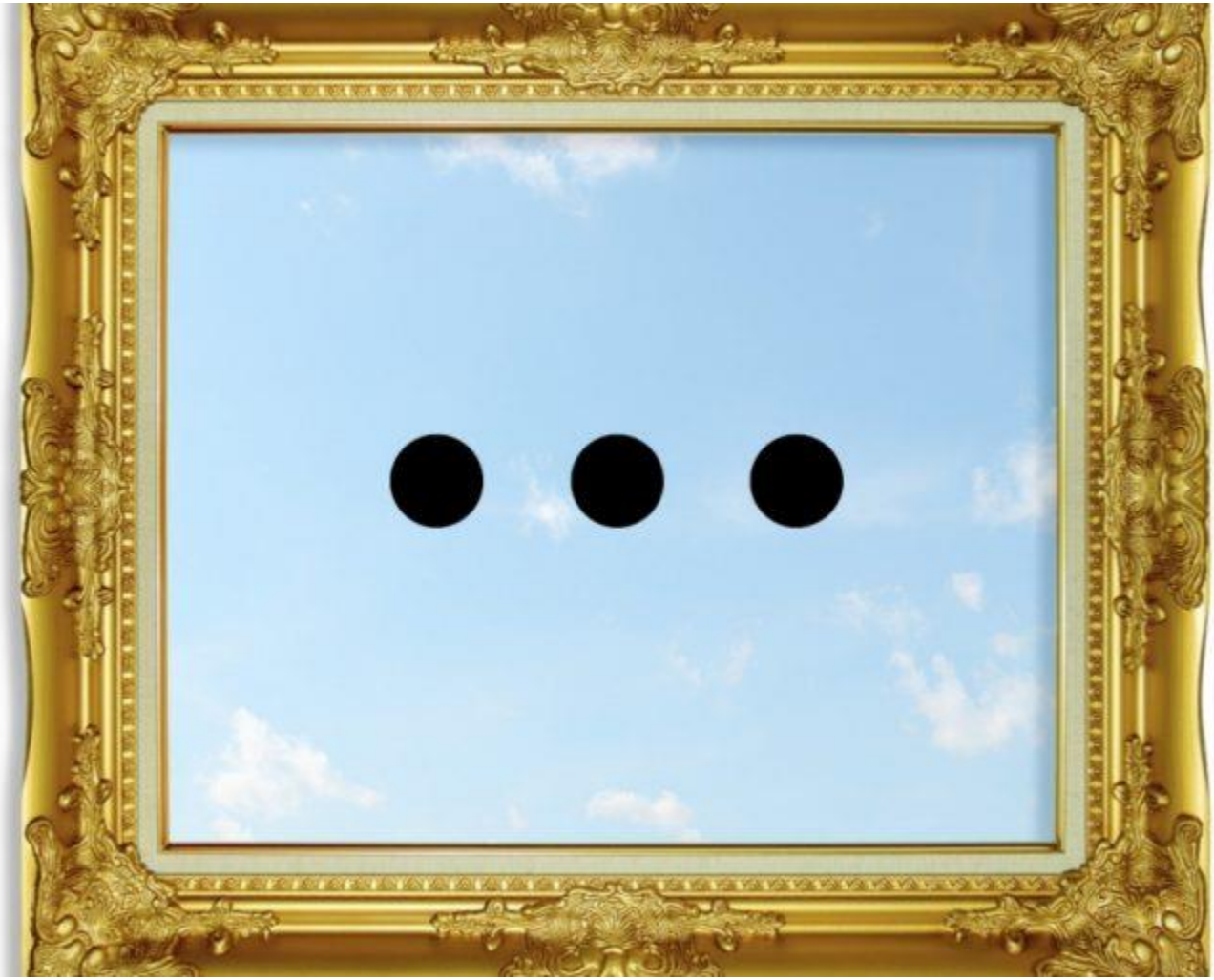
Beth Elderkin
7/03/18 2:45pm • Filed to: STAN LEE

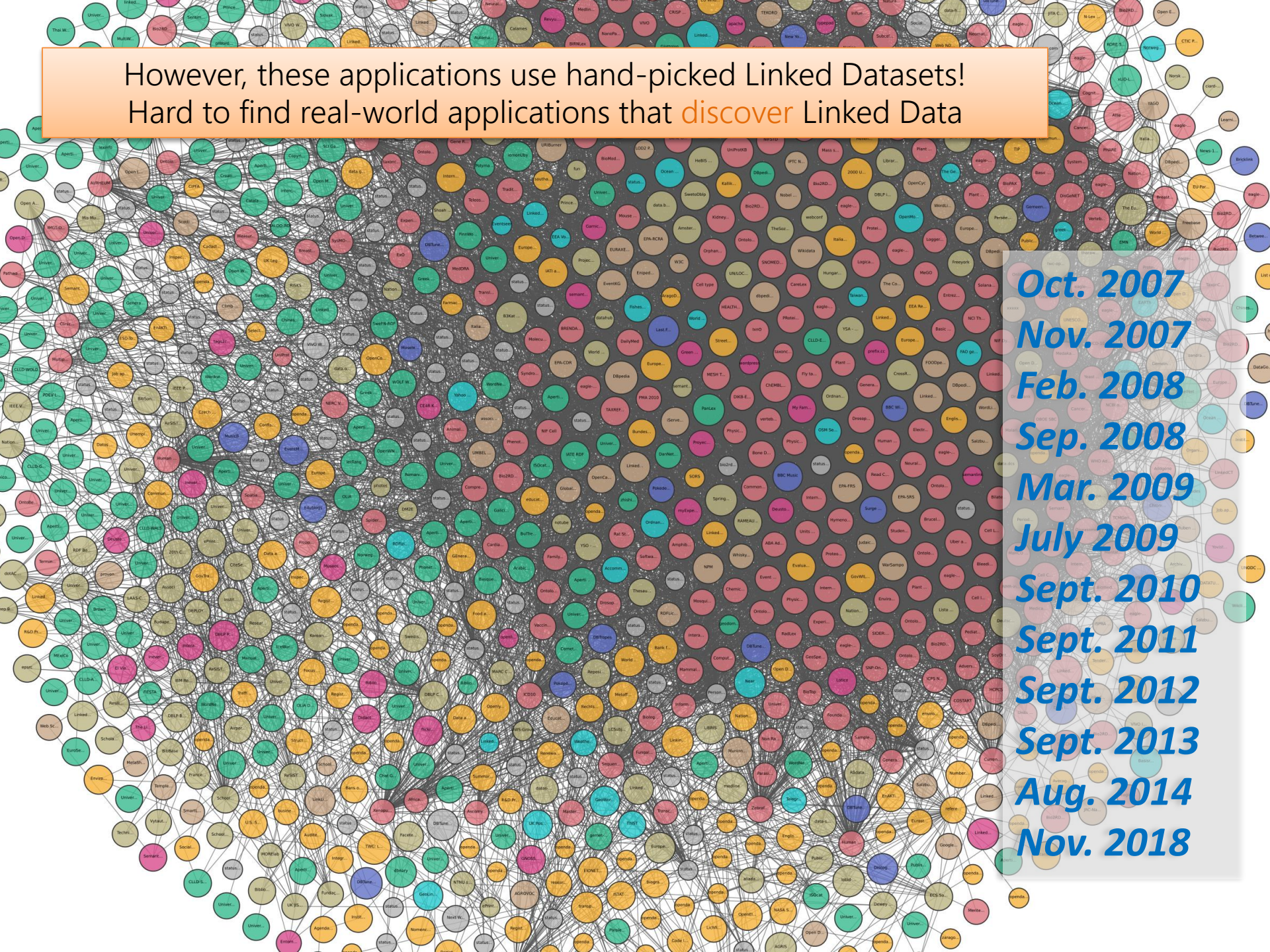
16.0K 48 5

IBM's Watson

Freebase™

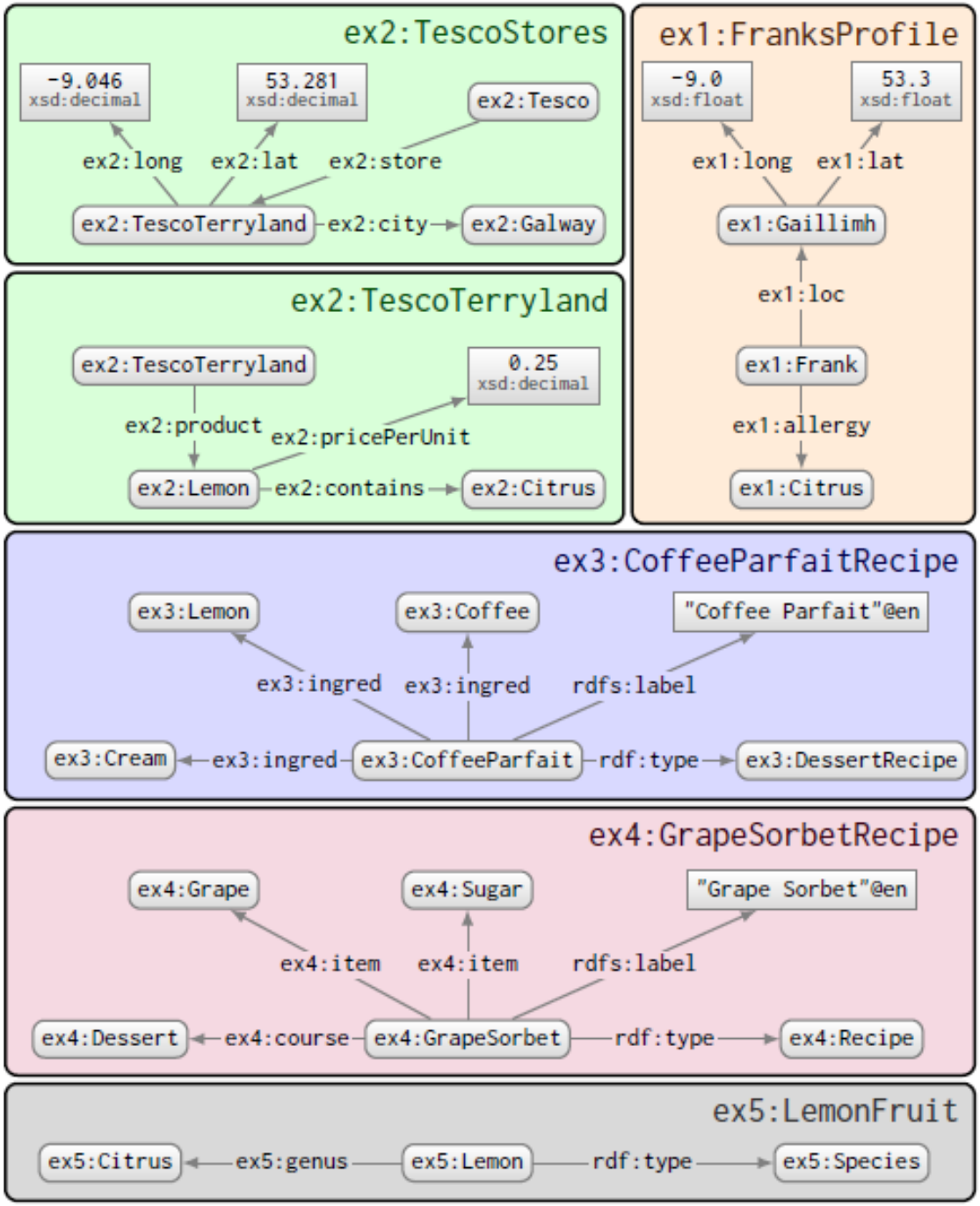
DBpedia




However, these applications use hand-picked Linked Datasets!
Hard to find real-world applications that discover Linked Data
Oct. 2007
Nov. 2007
Feb. 2008
Sep. 2008
Mar. 2009
July 2009
Sept. 2010
Sept. 2011
Sept. 2012
Sept. 2013
Aug. 2014
Nov. 2018

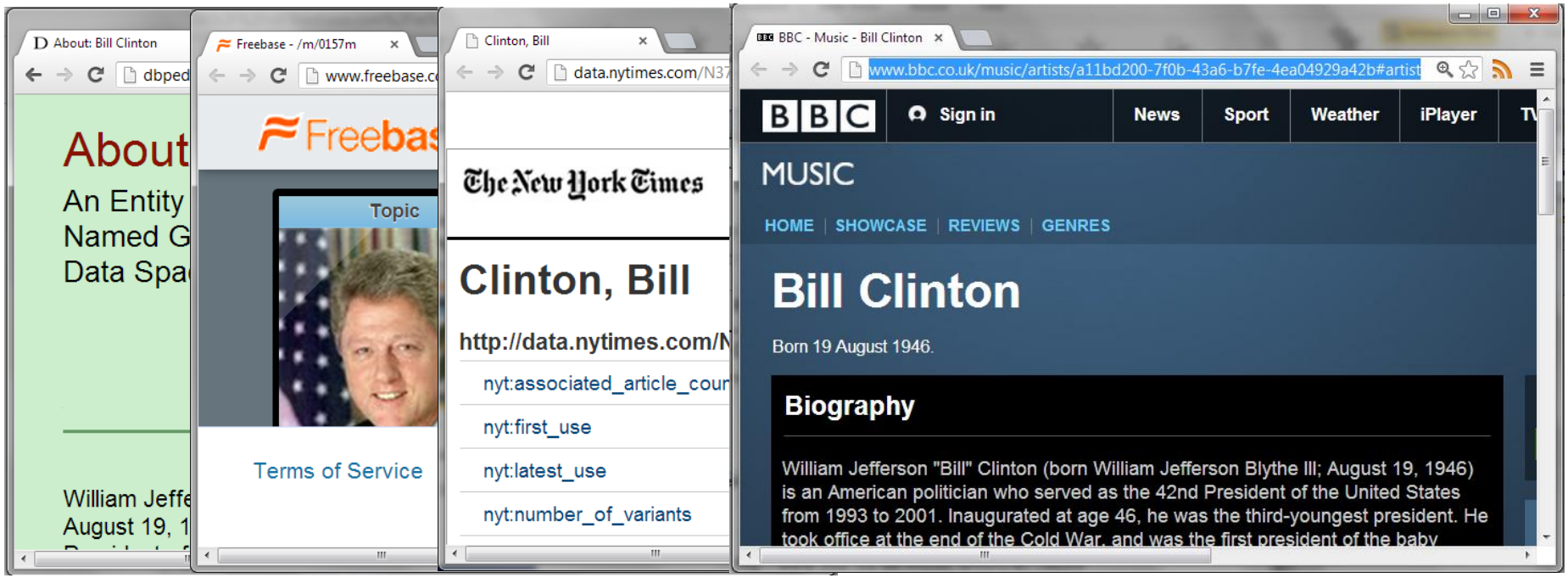

Entering Unknown Territory:
Open Research Questions!

Diversity ...


ex2:TescoStores
-9.046
xsd:decimal
53.281
xsd:decimal
ex2:Tesco
ex2:long
ex2:lat
ex2:store
ex2:TescoTerryland
ex2:city
ex2:Galway
ex1:FranksProfile
-9.0
xsd:float
53.3
xsd:float
ex1:long
ex1:lat
ex1:Gaillimh
ex1:loc
ex1:Frank
ex1:allergy
ex1:Citrus
ex2:TescoTerryland
ex2:TescoTerryland
0.25
xsd:decimal
ex2:product
ex2:pricePerUnit
ex2:Lemon
ex2:contains
ex2:Citrus
ex3:CoffeeParfaitRecipe
ex3:Lemon
ex3:Coffee
"Coffee Parfait"@en
ex3:ingred
ex3:ingred
rdfs:label
ex3:Cream
ex3:ingred
ex3:CoffeeParfait
rdf:type
ex3:DessertRecipe
ex4:GrapeSorbetRecipe
ex4:Grape
ex4:Sugar
"Grape Sorbet"@en
ex4:item
ex4:item
rdfs:label
ex4:Dessert
ex4:course
ex4:GrapeSorbet
rdf:type
ex4:Recipe
ex5:LemonFruit
ex5:Citrus
ex5:genus
ex5:Lemon
rdf:type
ex5:Species

Open Issue:
Linked Data 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.

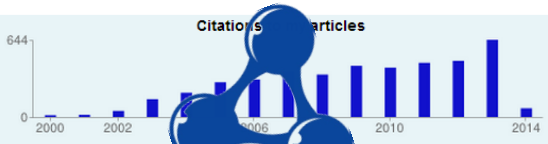
A (Hypothetical) Integration Example

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

Citation indices

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



Show: 20 1-20 Next >

Title / Author	Cited by	Year
Consistent query answers in inconsistent databases M Arenas, L Bertossi, J Chomicki Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on ...	584	1999
A normal form for XML documents M Arenas, L Libkin ACM Transactions on Database Systems (TODS) 29 (1), 195-232	469	2004

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

Facets and more with CompleteSearch

Ask others: ACM DL/Guide - CSB - MetaPress - Google - Bing - Yahoo

author:marcelo_arenas:

2013

j30 Marcelo Arenas, Pablo Barceló, Ronald Fagin, Leonid Libkin: Solutions and query rewriting in data exchange. Inf. Comput. 228: 28-61 (2013)

j29 Marcelo Arenas, Jorge Pérez, Juan L. Reutter: Data exchange beyond complete data. J. ACM 60(4): 28 (2013)

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]

Microsoft Academic Search Beta

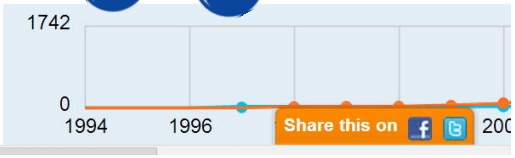
Co-authors (93)

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

Citation Graph

Academic > Authors > Marcelo Arenas

celo Arenas
icia Universidad Católica de Chile
Publications: 102 | Citations: 2074
Fields: Databases, Algorithms & Theory,
Collaborated with 93 co-authors from 199



Create account Log in

Article Talk Read

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop
Interaction

Pontifical Catholic University of Chile

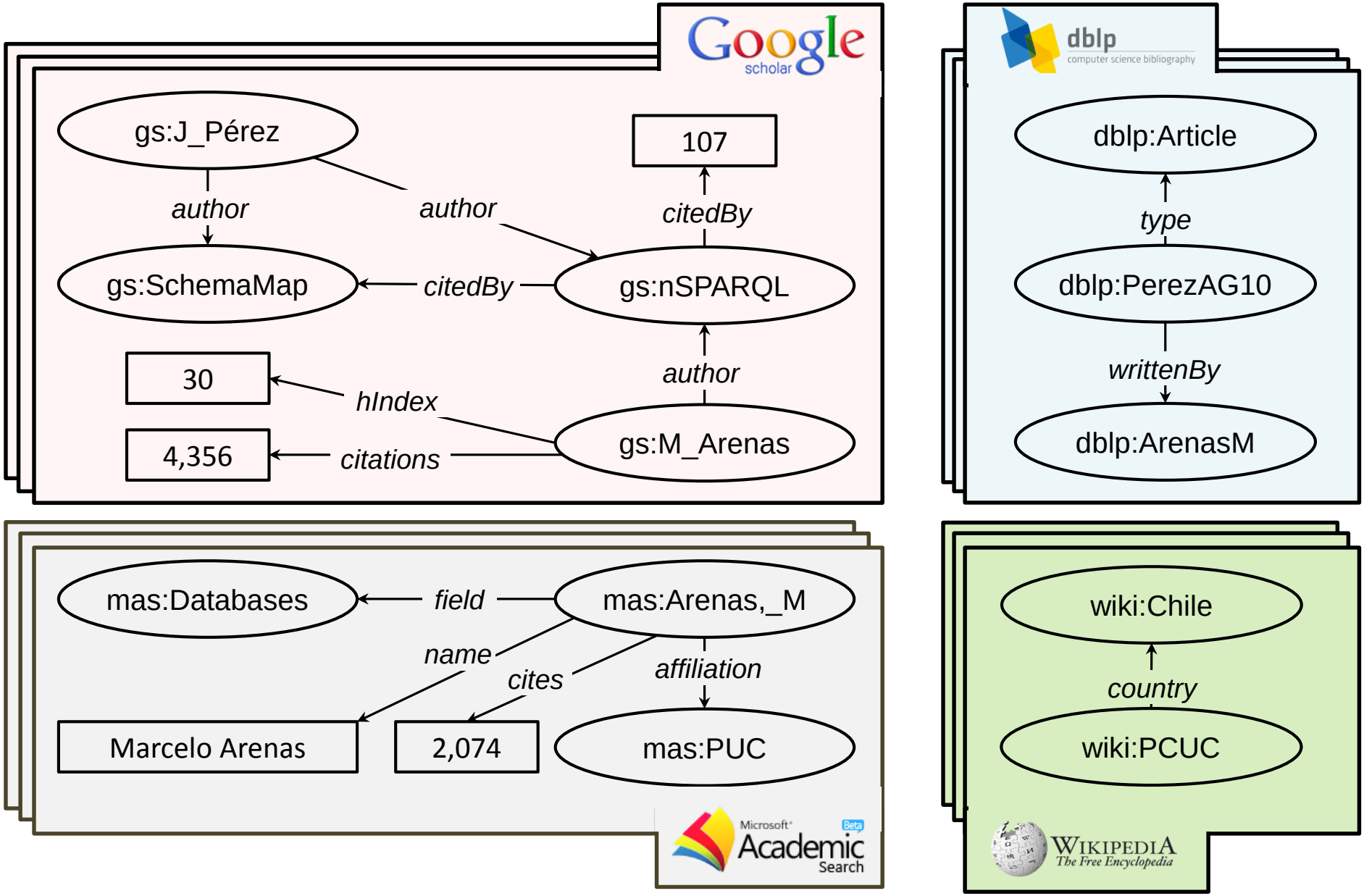
From Wikipedia, the free encyclopedia

Coordinates: 33.4411°S 70.6408°W

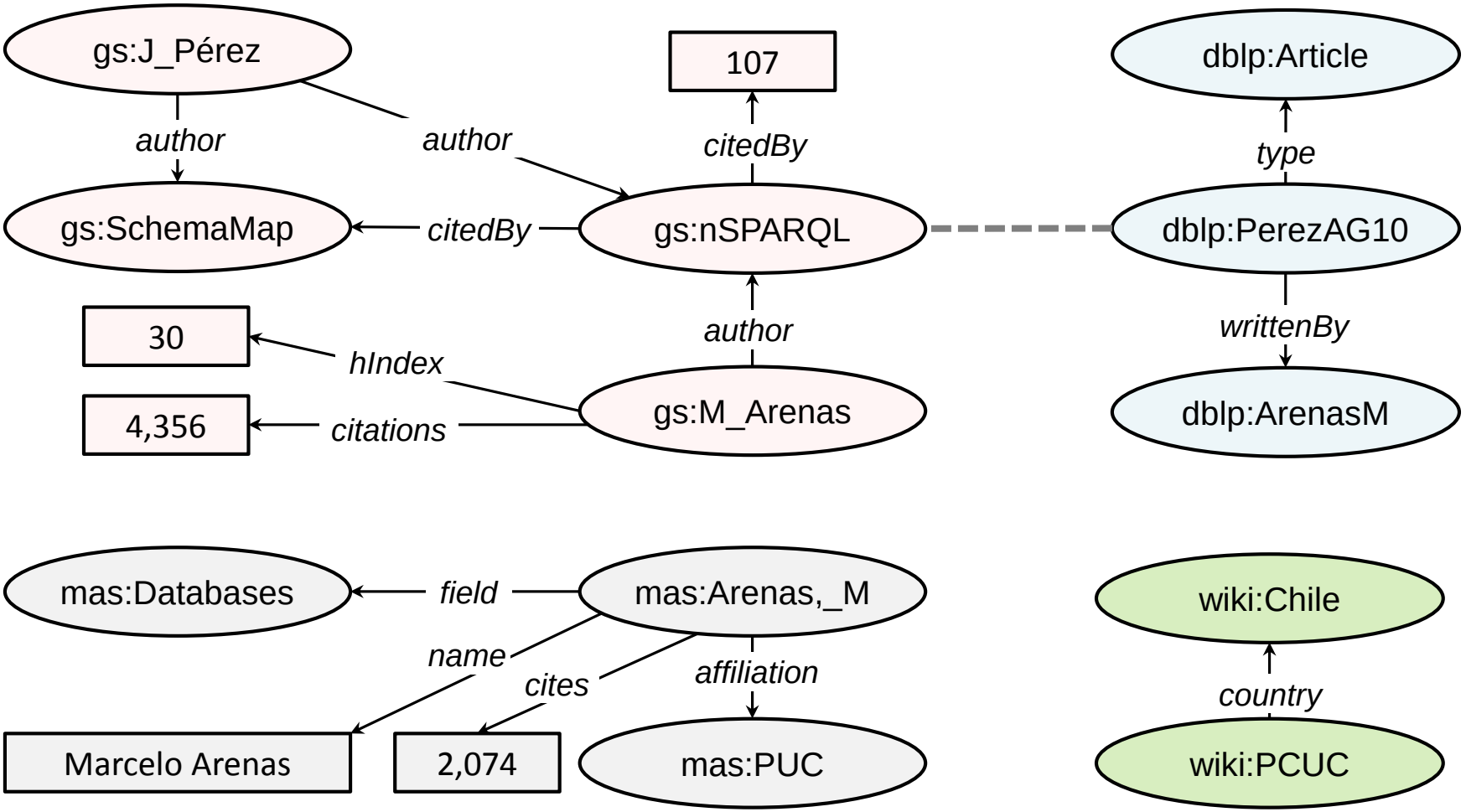
The **Pontifical Catholic University of Chile (UC or PUC)** (Spanish: *Pontificia Universidad Católica de Chile*) is one of the six Catholic Universities existing in the Chilean university system and one of the two Pontifical Universities in the country, along with the Pontifical Catholic University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

Pontifical Catholic University of Chile
Pontificia Universidad Católica de Chile

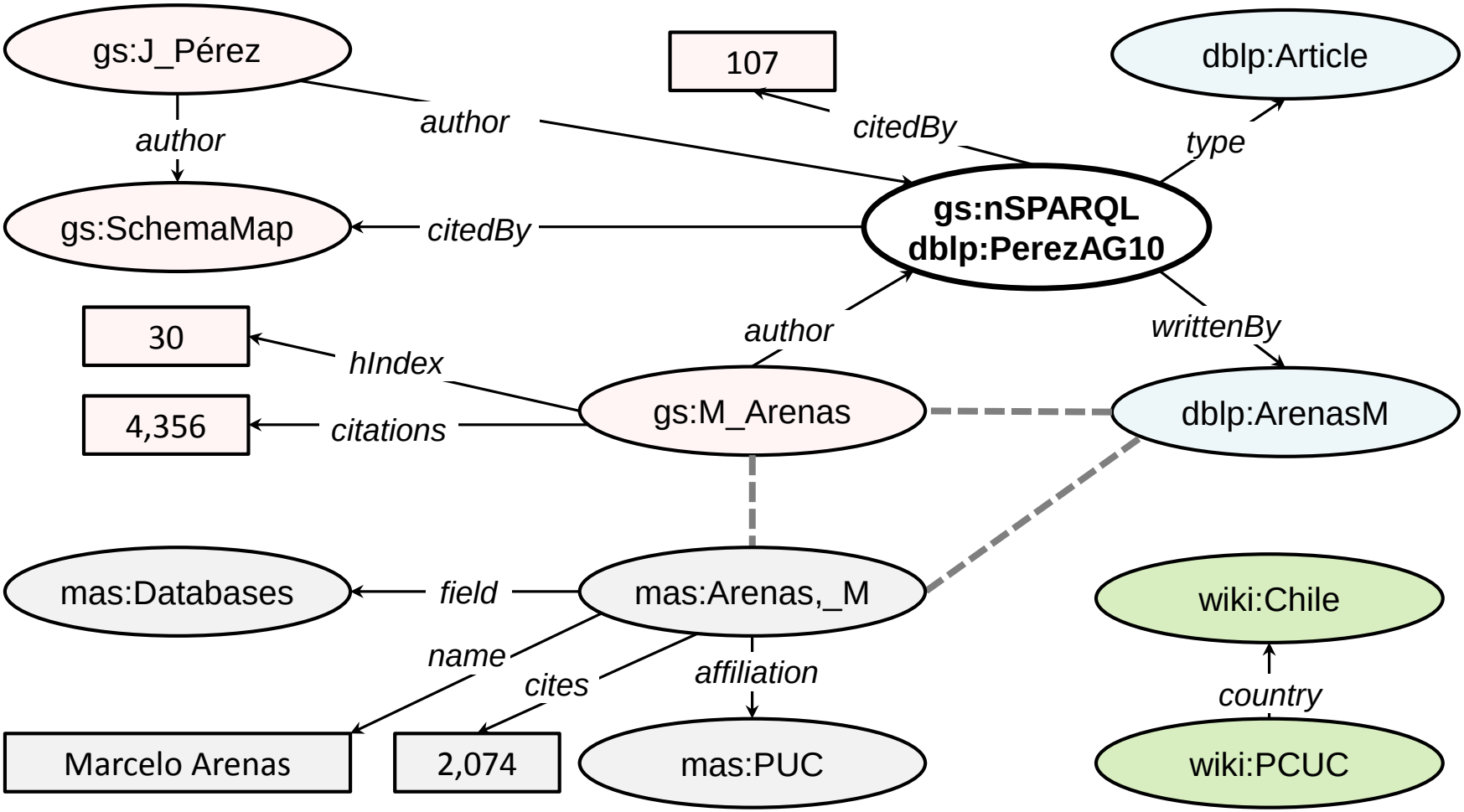
A (Hypothetical) Integration Example



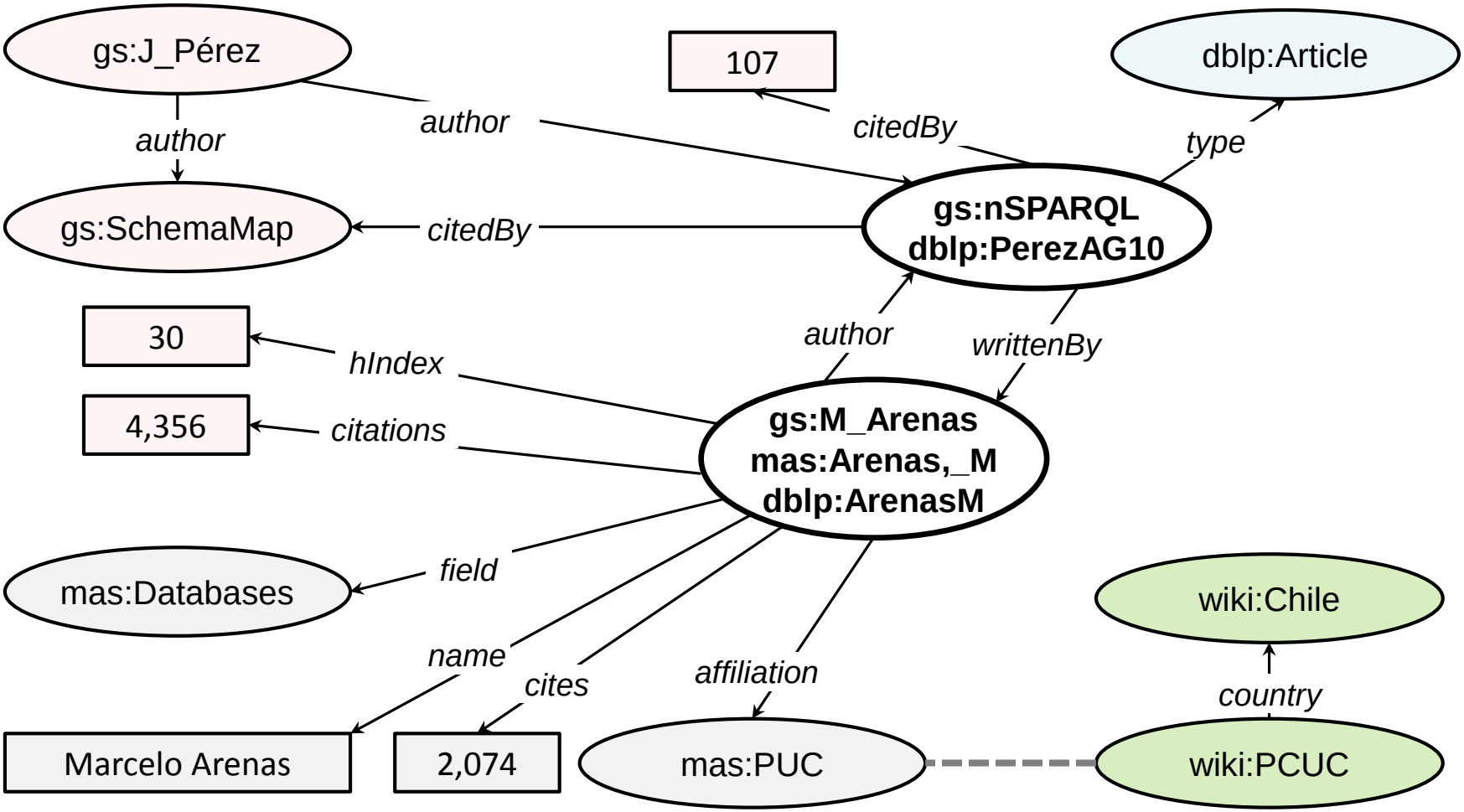
A (Hypothetical) Integration Example



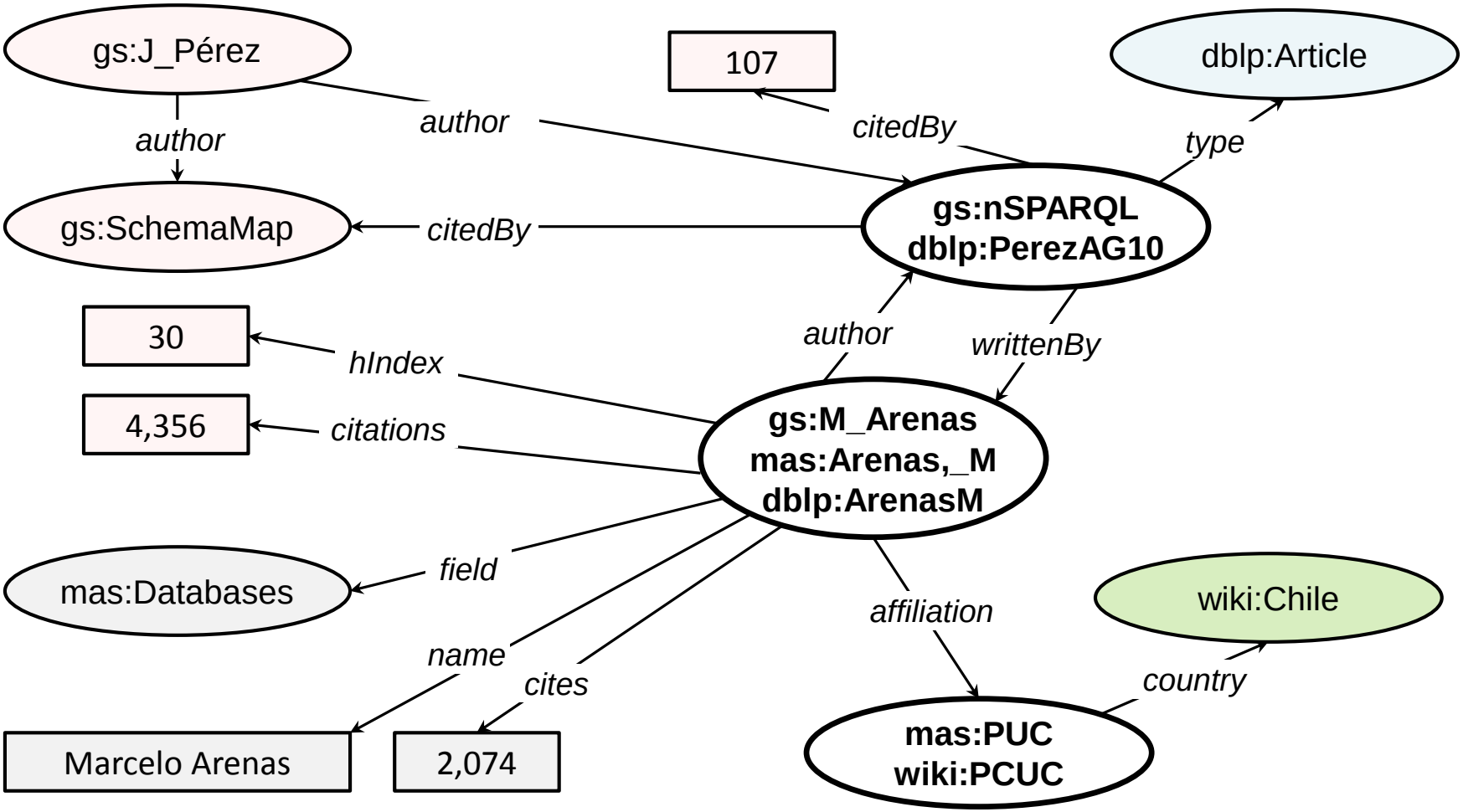
A (Hypothetical) Integration Example



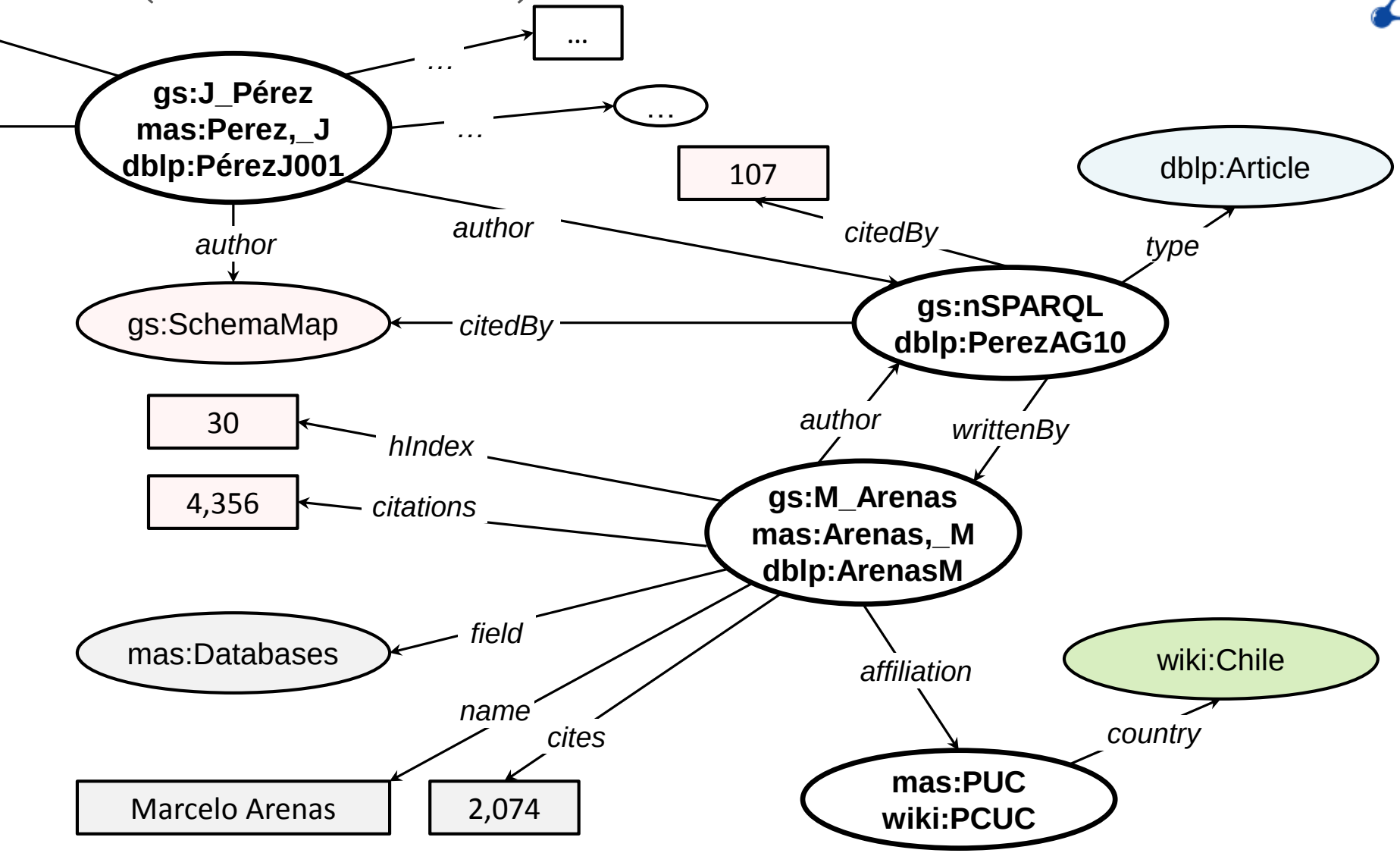
Semantic Web: Tackling Heterogeneity



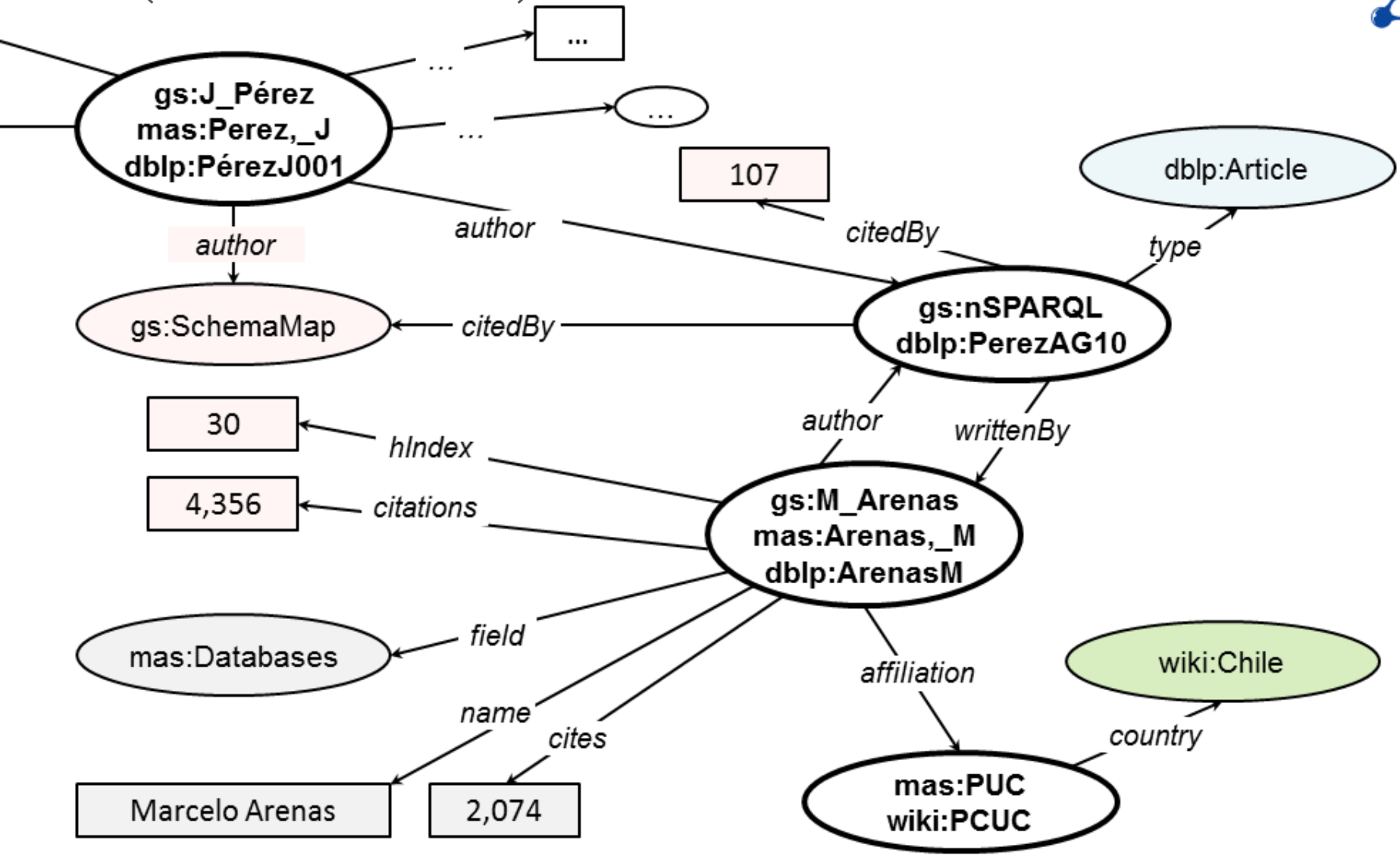
A (Hypothetical) Integration Example



A (Hypothetical) Integration Example

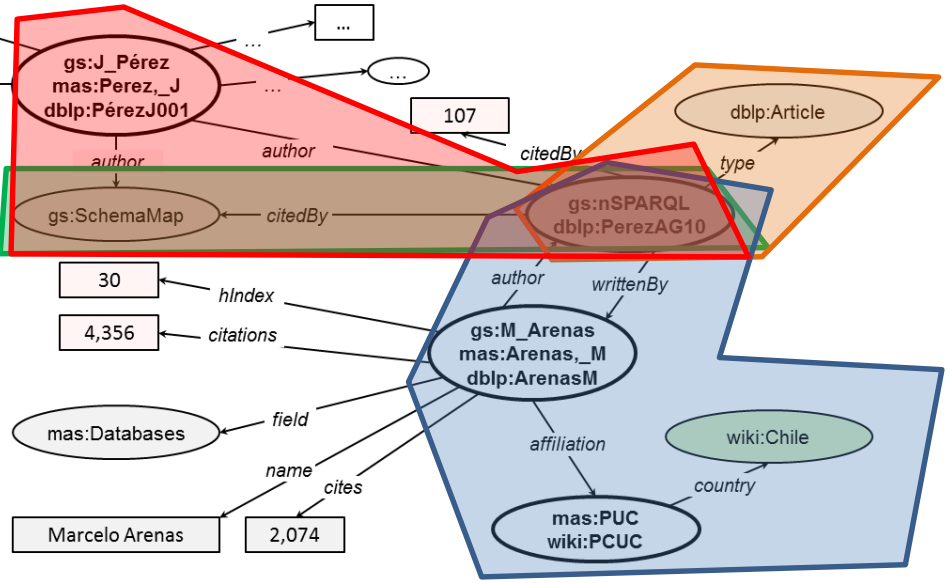


A (Hypothetical) Integration Example



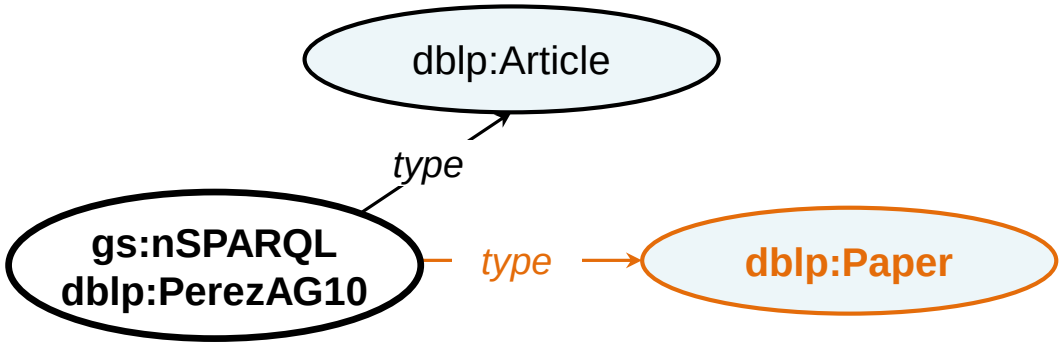
A (Hypothetical) Integration Example

```
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)
```

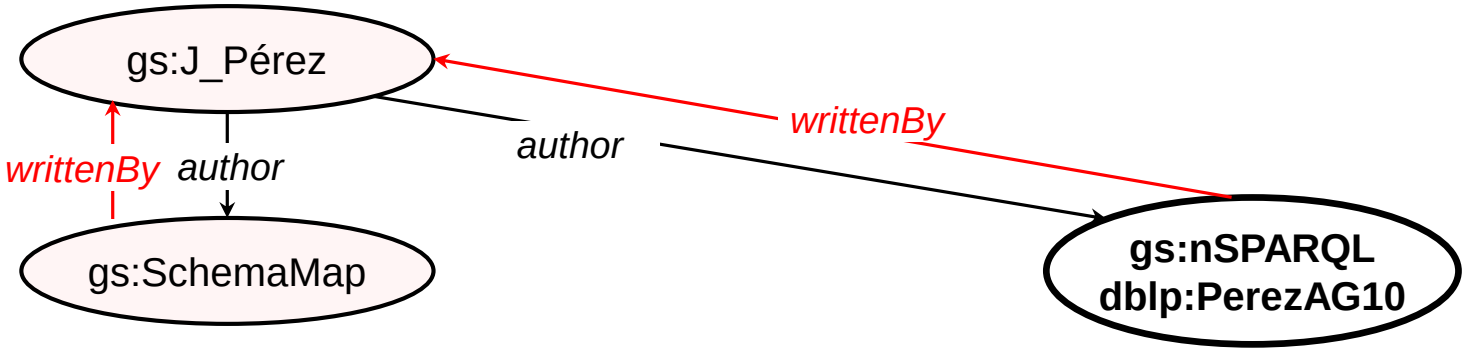


Article ⊑ Paper

```
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)
```

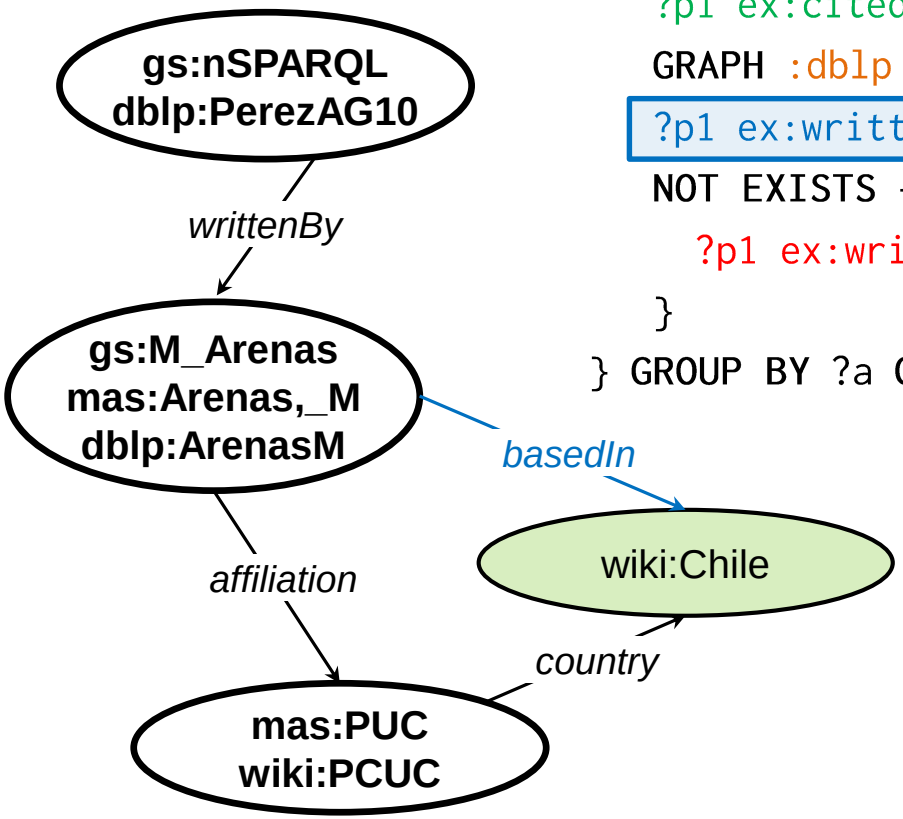


author ≡ writtenBy $^-$

```
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)
```

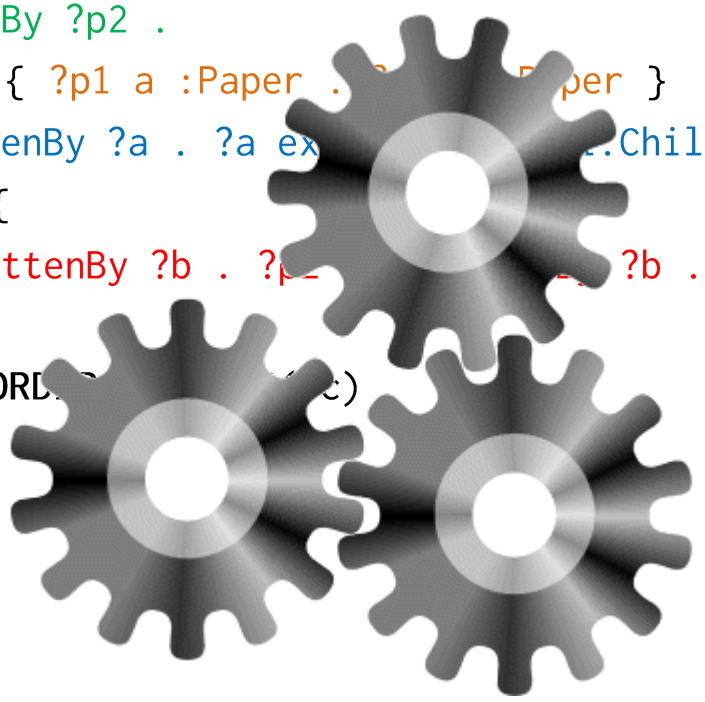


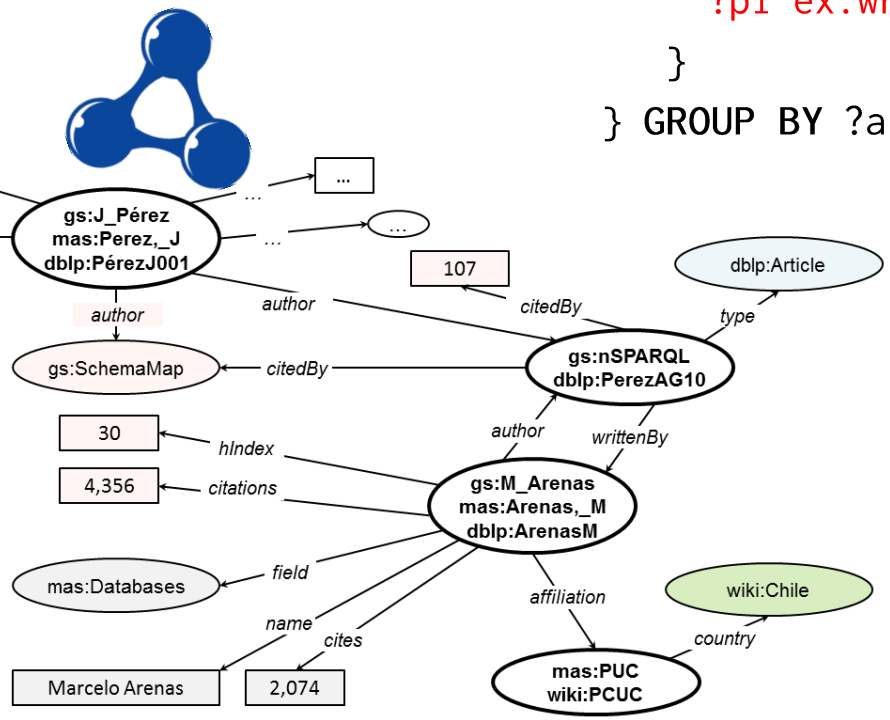
affiliation ∘ country ⊑ basedIn

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

A (Hypothetical) Integration Example

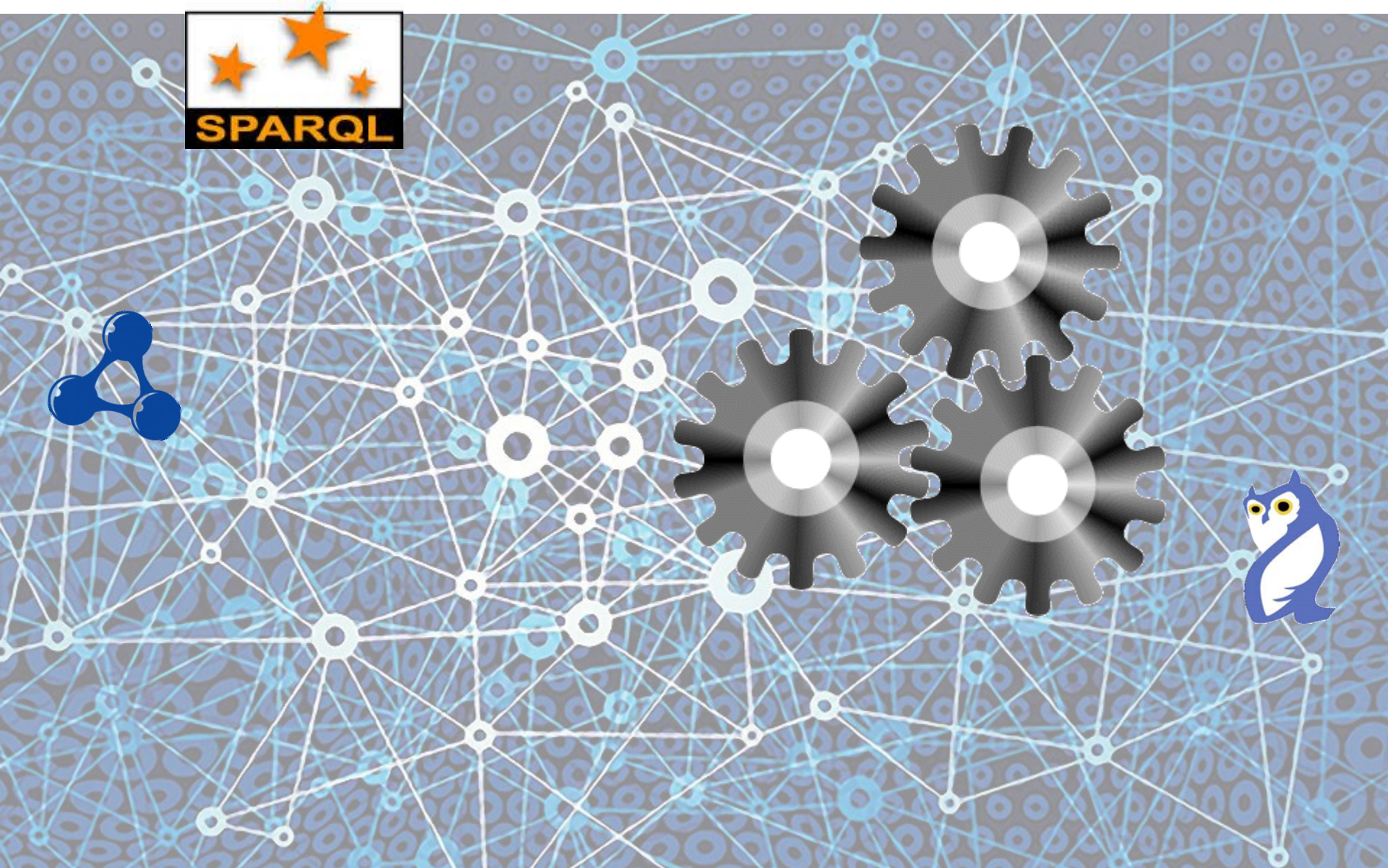

```
SELECT ?a (COUNT(DISTINCT ?p2) AS ?c)
FROM NAMED …
WHERE {
   ?p1 ex:citedBy ?p2 .
   GRAPH :dblp { ?p1 a :Paper .  per }
   ?p1 ex:writtenBy ?a . ?a ex  .Chile .
   NOT EXISTS {
     ?p1 ex:writtenBy ?b . ?p  ?b .
   }
} GROUP BY ?a ORD  c)
```





Article ⊑ Paper
author ≡ writtenBy ⁻
affiliation ∘ country ⊑ basedIn

Not clear yet how to do reasoning on the Web!



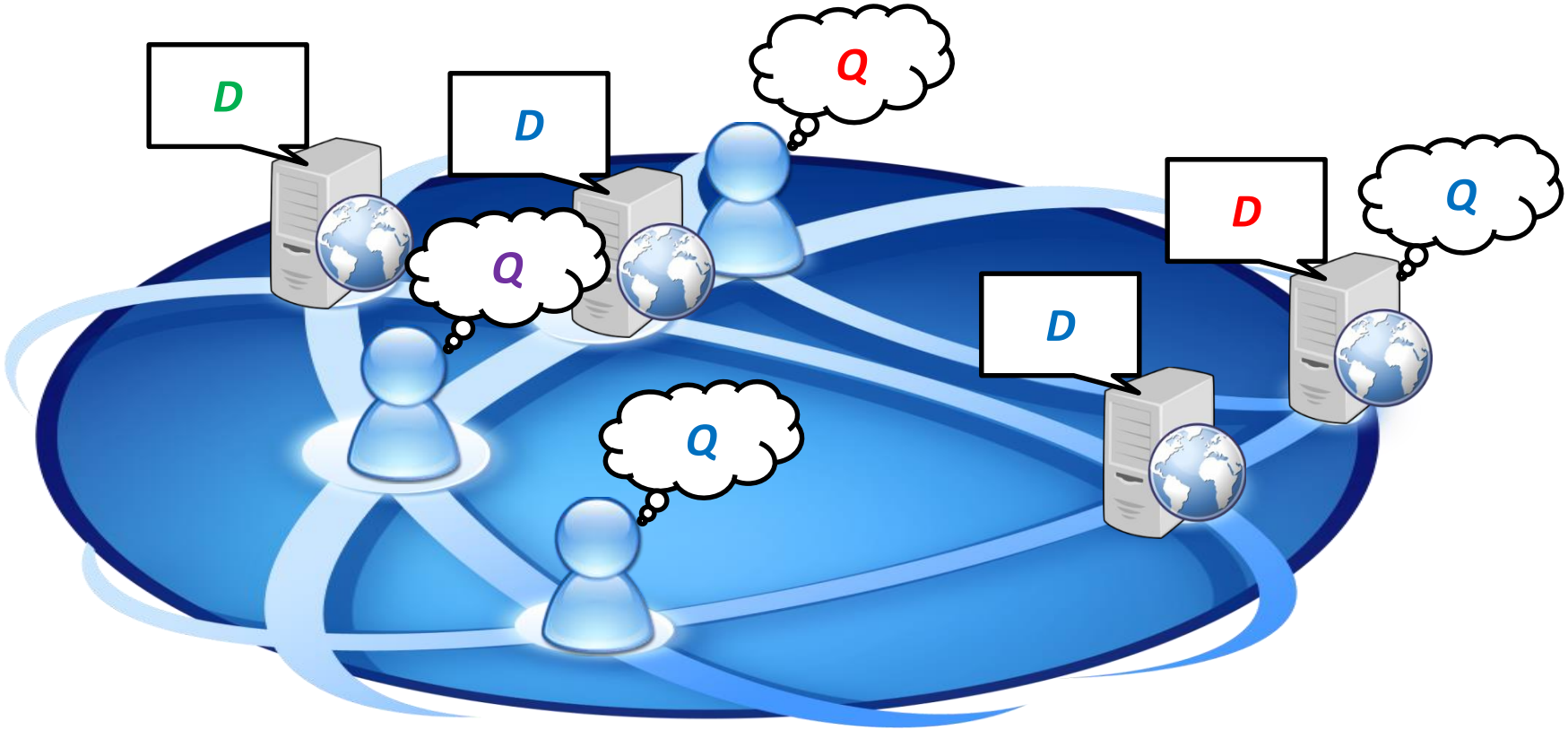
Side Note: Fact or Fiction?

Open Issue:
Linked 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/Colombia”

$Q(D)$ =

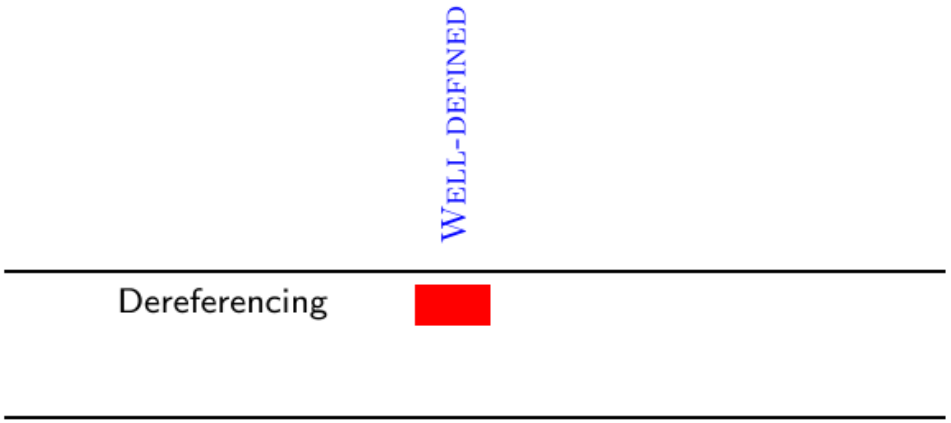
```
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix dbr: <http://dbpedia.org/resource/> .
@prefix ns3: <http://purl.org/linguistics/gold/> .
@prefix dbp: <http://dbpedia.org/property/> .

dbr:Maria_Full_of_Grace dbo:country dbr:Colombia .
dbr:John_Leguizamo dbo:birthPlace dbr:Colombia .
dbr:Tres_Coronas dbo:hometown dbr:Colombia .
dbr:Victoriana_Mejía_Marulanda dbo:birthPlace dbr:Colombia ;
 dbo:country dbr:Colombia .
dbr:Kristin_Amparo dbo:birthPlace dbr:Colombia .
dbr:Per-Olov_Kindgren dbo:birthPlace dbr:Colombia .
<http://dbpedia.org/resource/2010_Davis_Cup_World_Group_Play-offs> dbp:venue dbr:Colombia .
dbr:Edwin_Cassiani dbo:nationality dbr:Colombia .
dbr:Edwin_Mosquera dbo:birthPlace dbr:Colombia .
dbr:We_are_Colombia ns3:hypernym dbr:Colombia .
<http://dbpedia.org/resource/Alejandro_Gonz\u00E1lez_(tennis)> dbo:country dbr:Colombia .
dbr:Pan_de_coco dbo:region dbr:Colombia .
dbr:Elida_Campodónico dbo:birthPlace dbr:Colombia .
dbr:Latin_American_Federation_of_the_Society_of_Jesus dbo:location dbr:Colombia .
dbr:Productora_de_Software dbo:locationCountry dbr:Colombia .
dbr:Rudolf_Rettberg dbp:deathPlace dbr:Colombia .
dbr:Luis_Felipe_Restrepo dbo:birthPlace dbr:Colombia .
dbr:William_Braucher_Wood dbo:country dbr:Colombia .
dbr:Nueva_Granada_Military_University dbo:country dbr:Colombia .
dbr:University_of_Boyacá dbo:country dbr:Colombia .
...
```

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

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

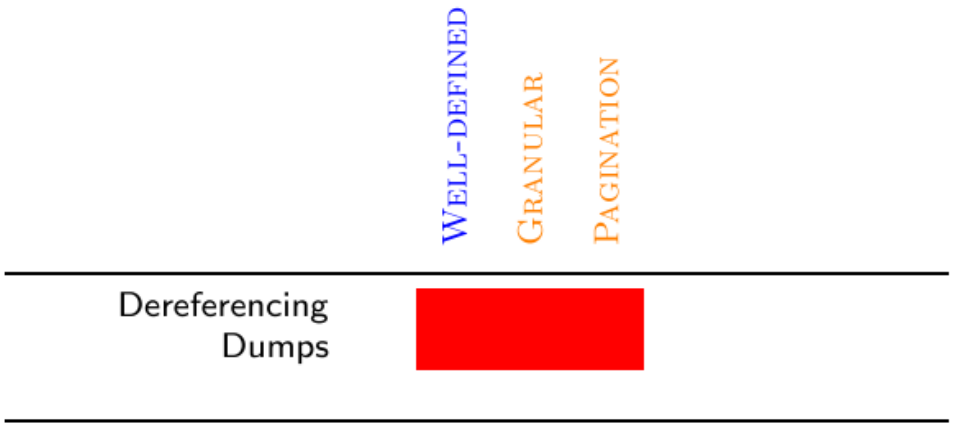
DBpedia 3.9 Downloads

This pages provides downloads of the DBpedia datasets. The DB
are licensed under the terms of the Creative Commons Attributio
and the GNU Free Documentation License. OPEN DATA The dow
as N-Triples and N-Quads, where the N-Quads version contains
information for each statement. All files are bzip2 [1] packed.
In addition to the RDF version of the data, we also provide a tabu
of the core DBpedia data sets as CSV and JSON files. See DBp

Older Versions: DBpedia 3.8, DBpedia 3.7, DBpedia 3.6, DBped
DBpedia 3.4, DBpedia 3.3, DBpedia 3.2, DBpedia 3.1, DBpedia
DBpedia 2.0

Dumps (what's wrong with them?)

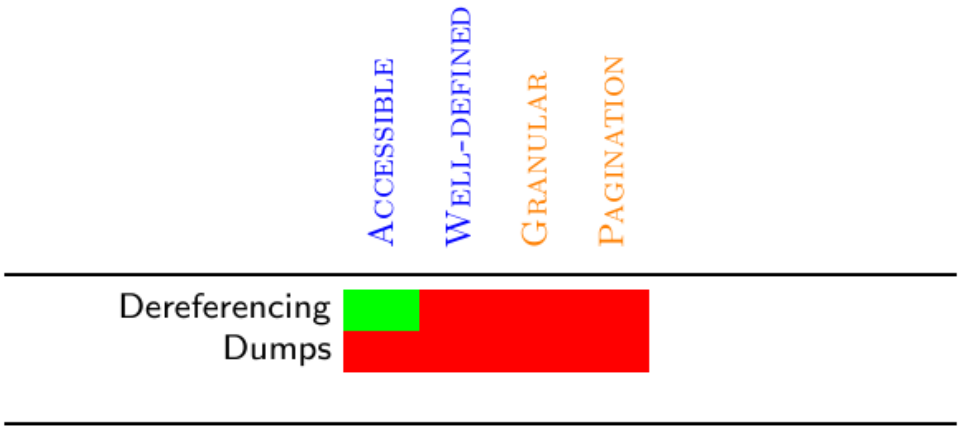
- 15× compression for RDF achievable
- But same weaknesses as for deref. still apply



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 =

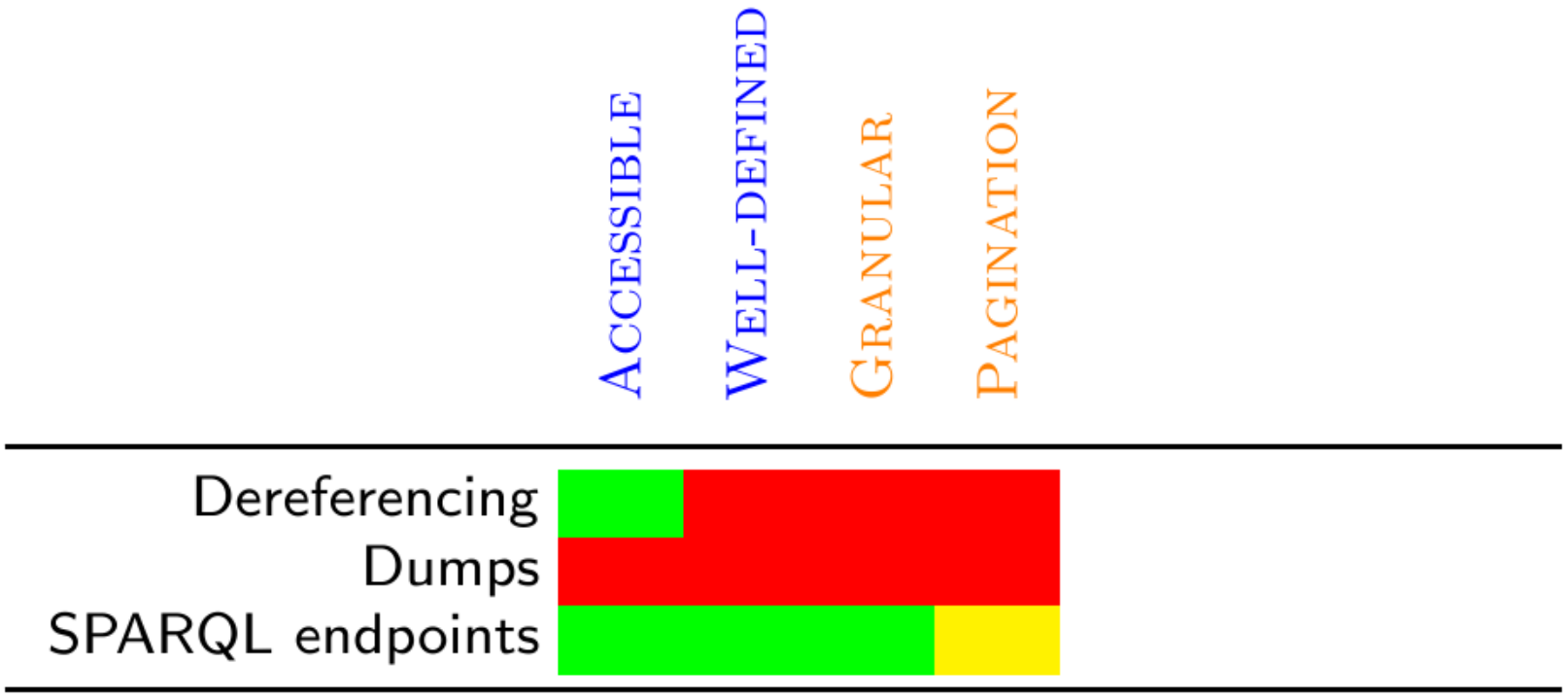
```
PREFIX dbo: <http://dbpedia.org/ontology/>
...

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")
}
```

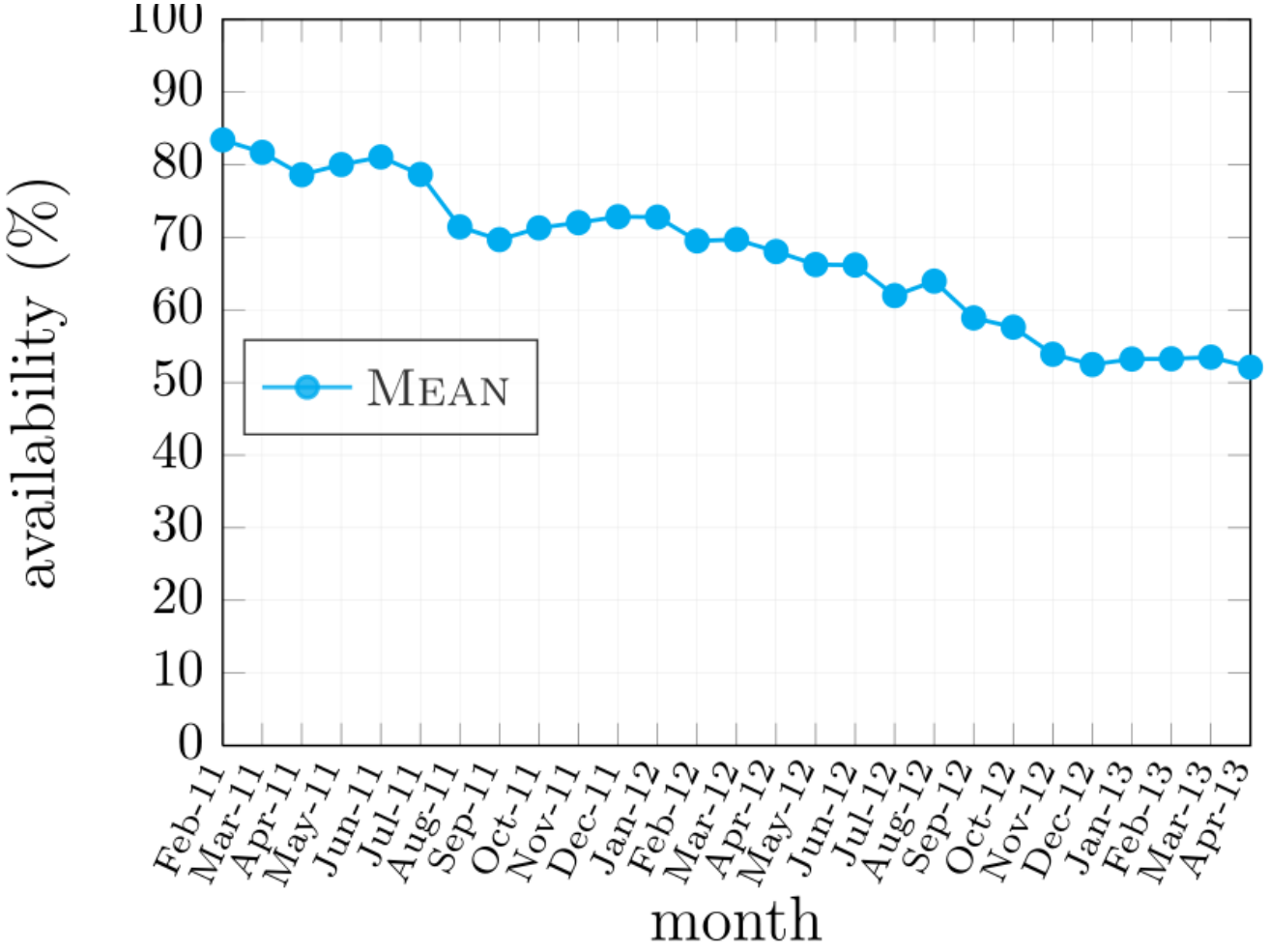
$Q(D)$ =

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

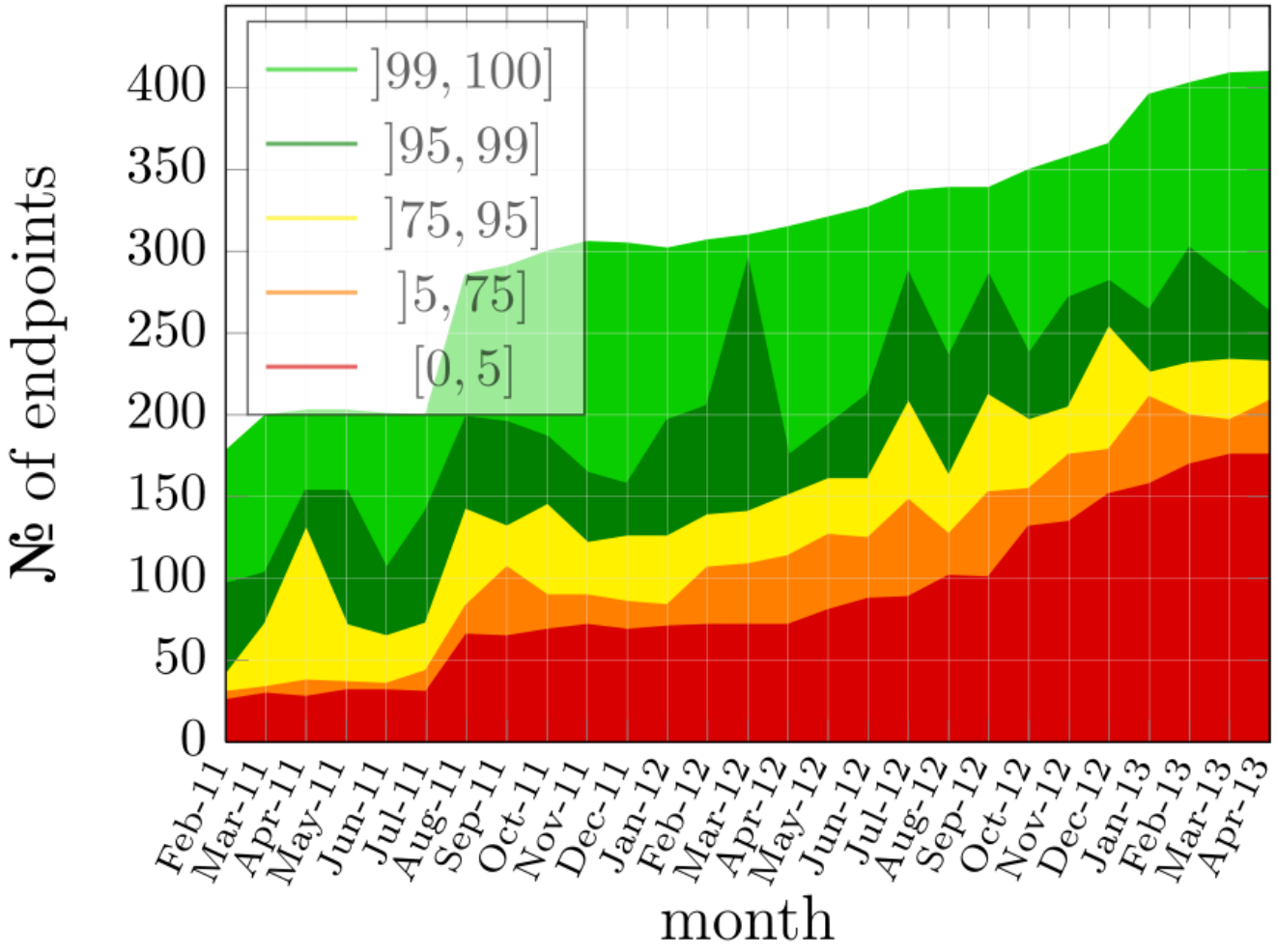
SPARQL (TO THE RESCUE?)



SPARQL (WHAT'S WRONG WITH IT?)

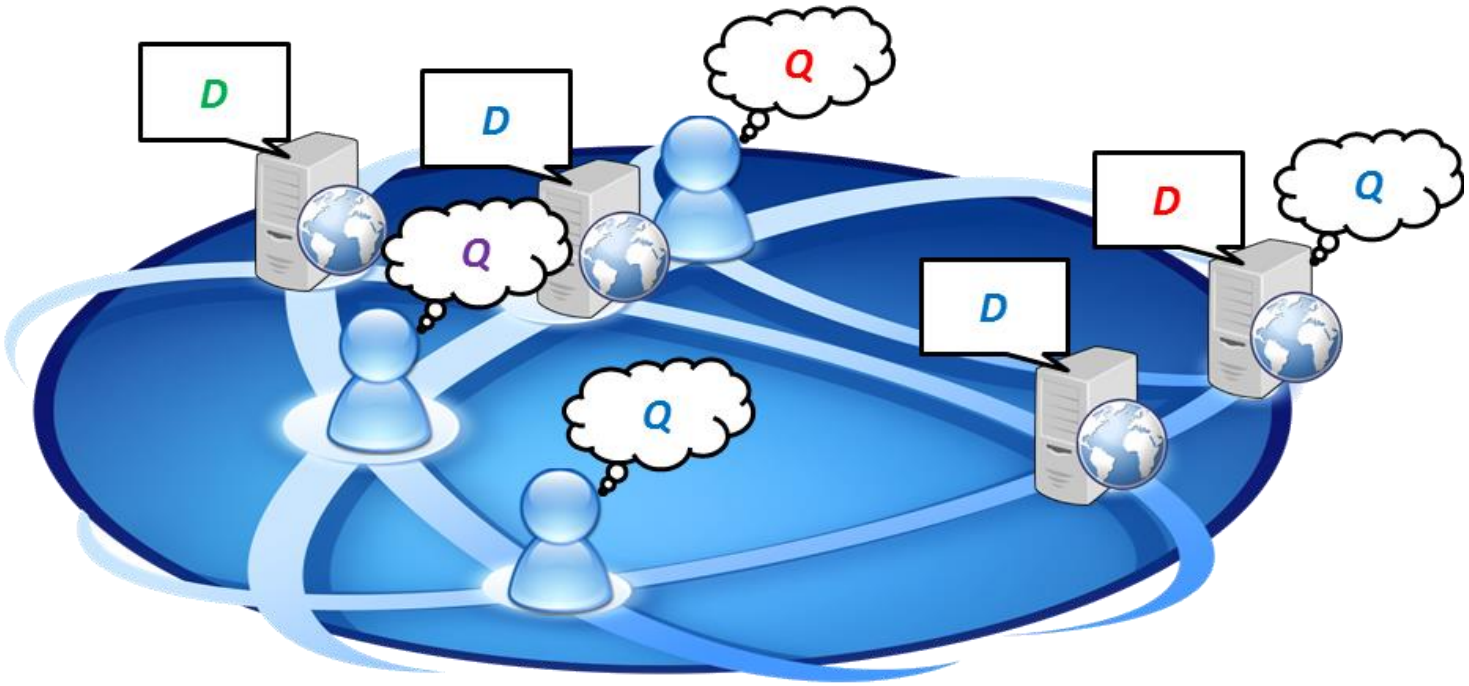


SPARQL (WHAT'S WRONG WITH IT?)



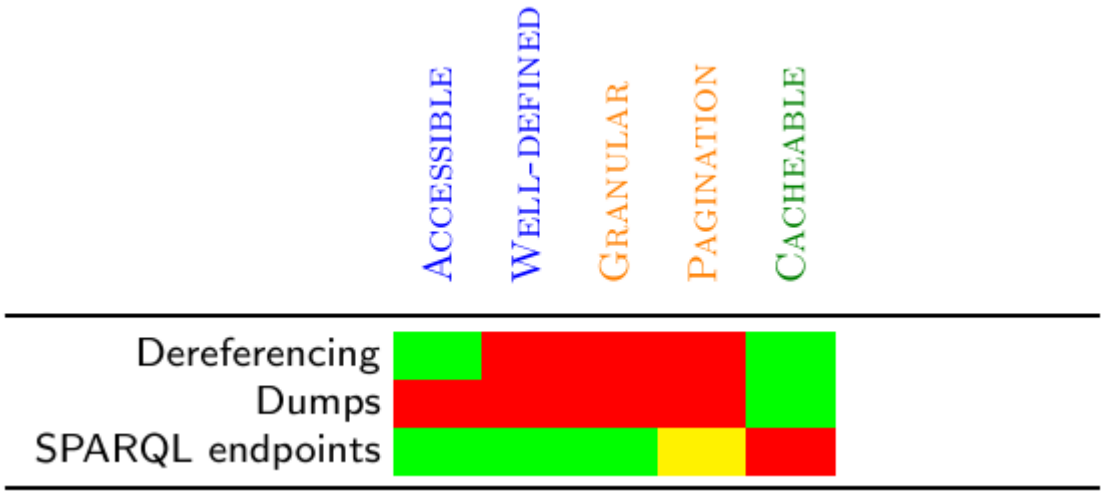
SPARQL (WHAT'S WRONG WITH IT?)

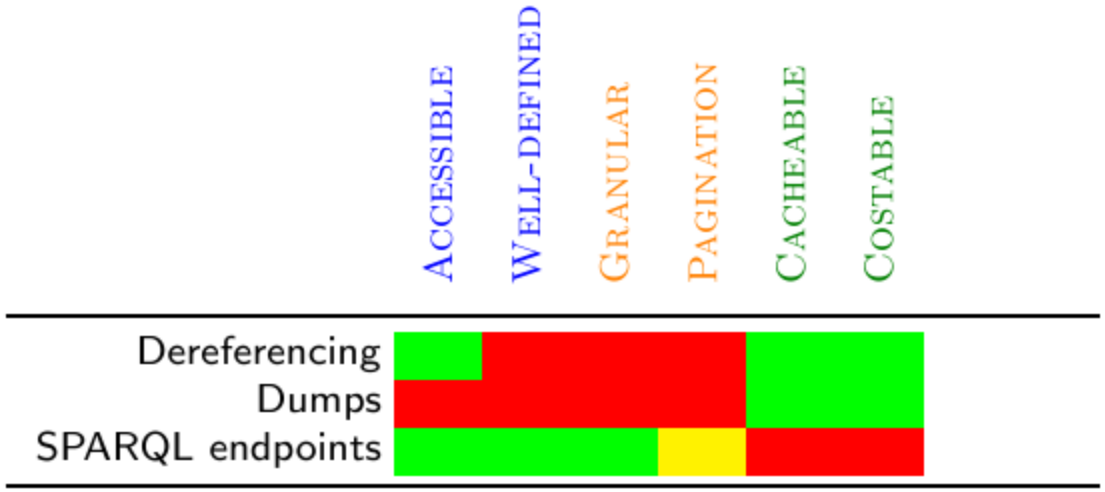
- ~~Simple~~ Protocol and RDF Query Language
- SPARQL evaluation: PSPACE-complete



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?)

№ of Results	№ of Endpoints
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?)

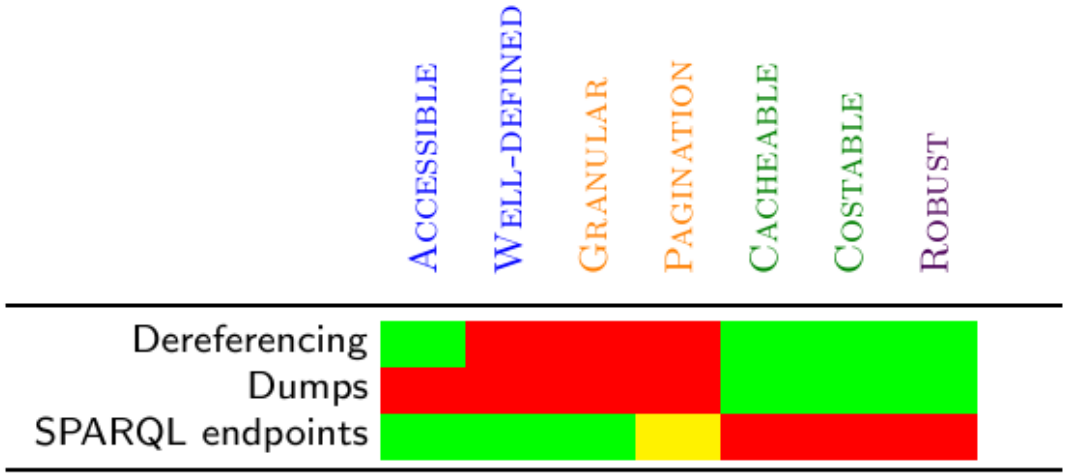
```
Virtuoso 42000 Error The estimated execution time 0
(sec) exceeds the limit of 3000 (sec).

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



SPARQL (WHAT'S WRONG WITH IT?)

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

SPARQL (WHAT'S WRONG WITH IT?)

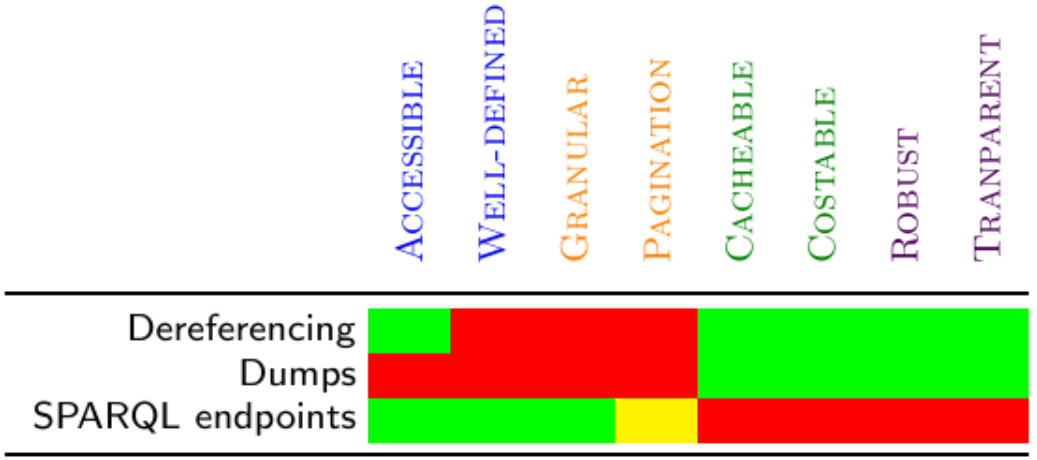
Q =

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

$Q(D)$ =

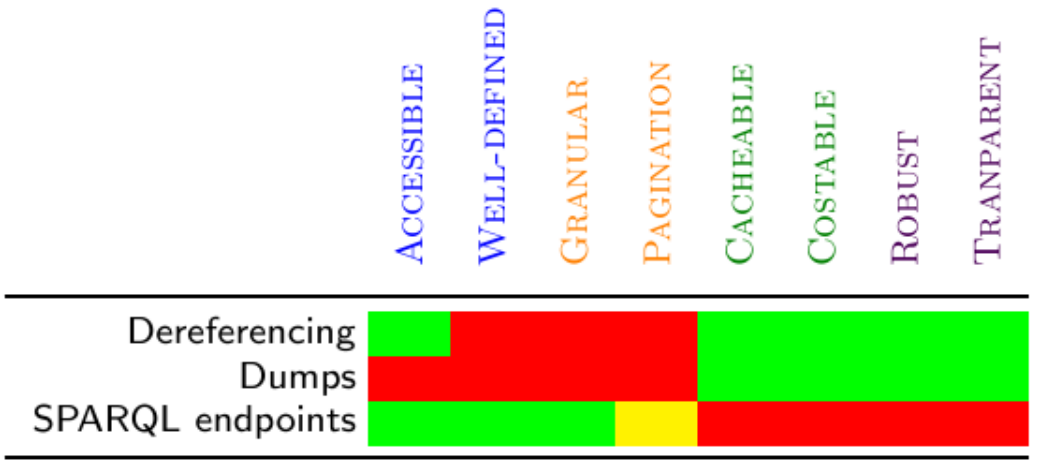
count
7182

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

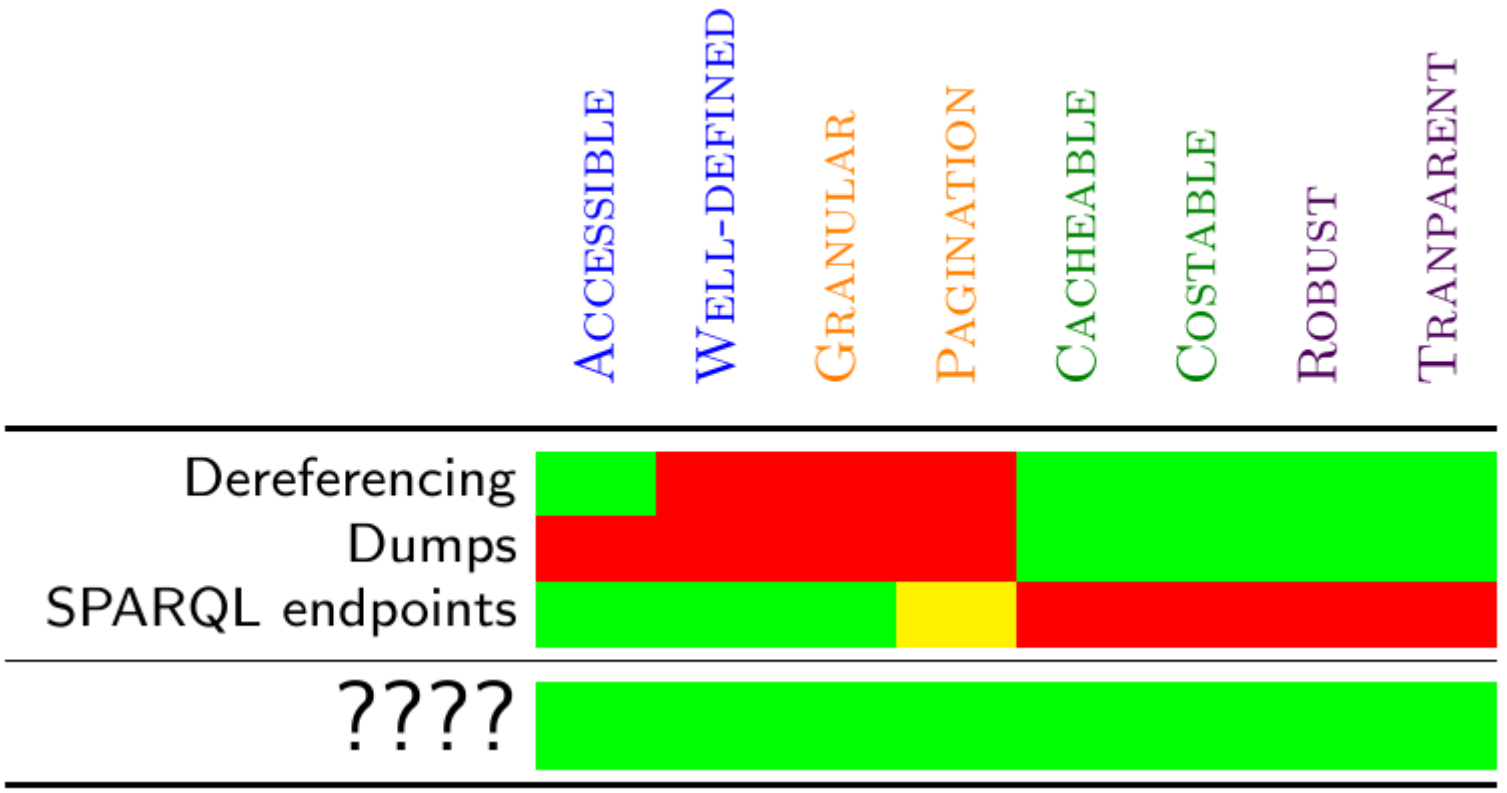


Problem Categories

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

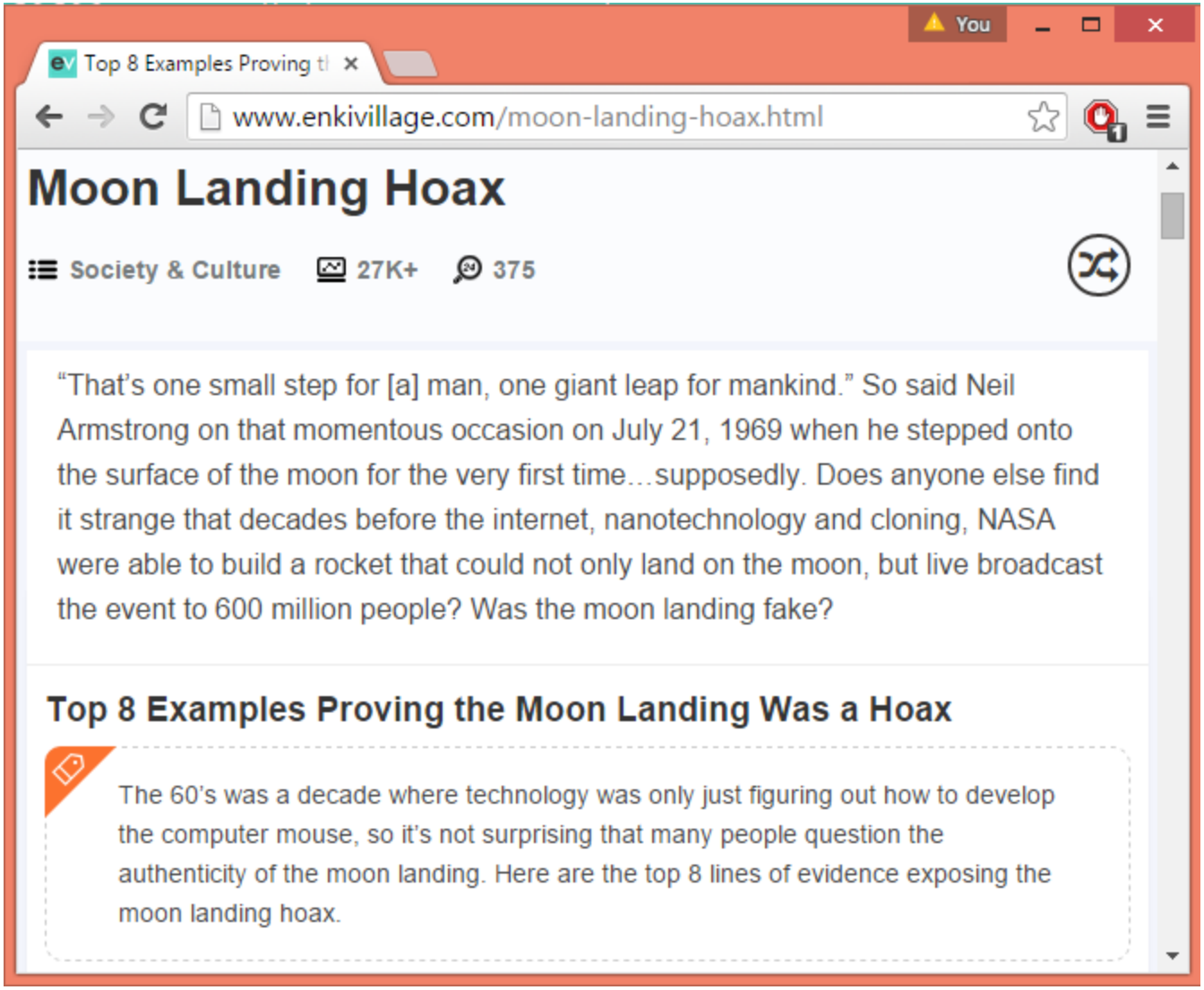


Data Access: Open Problem

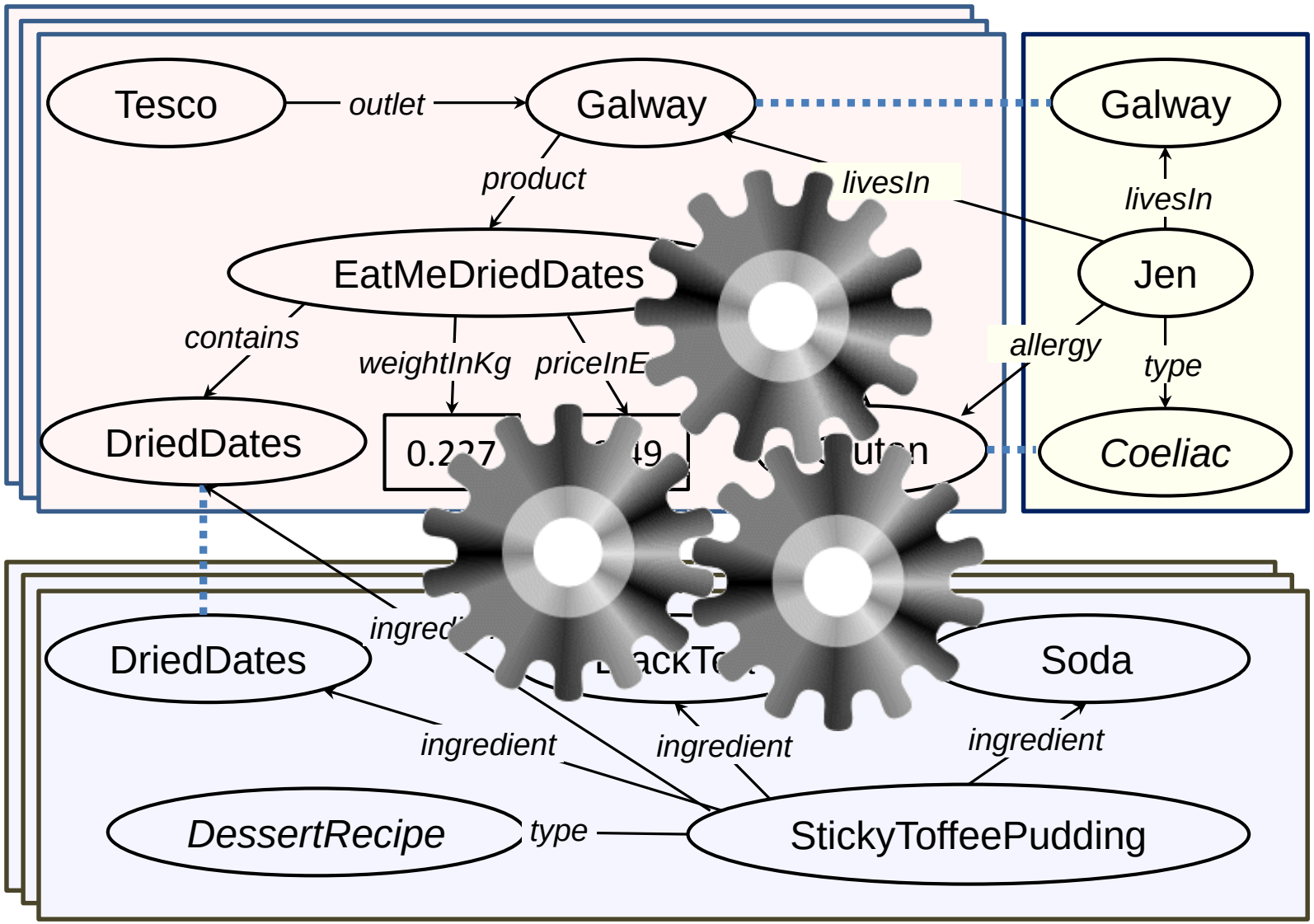


Open Issue:
Linked Data quality

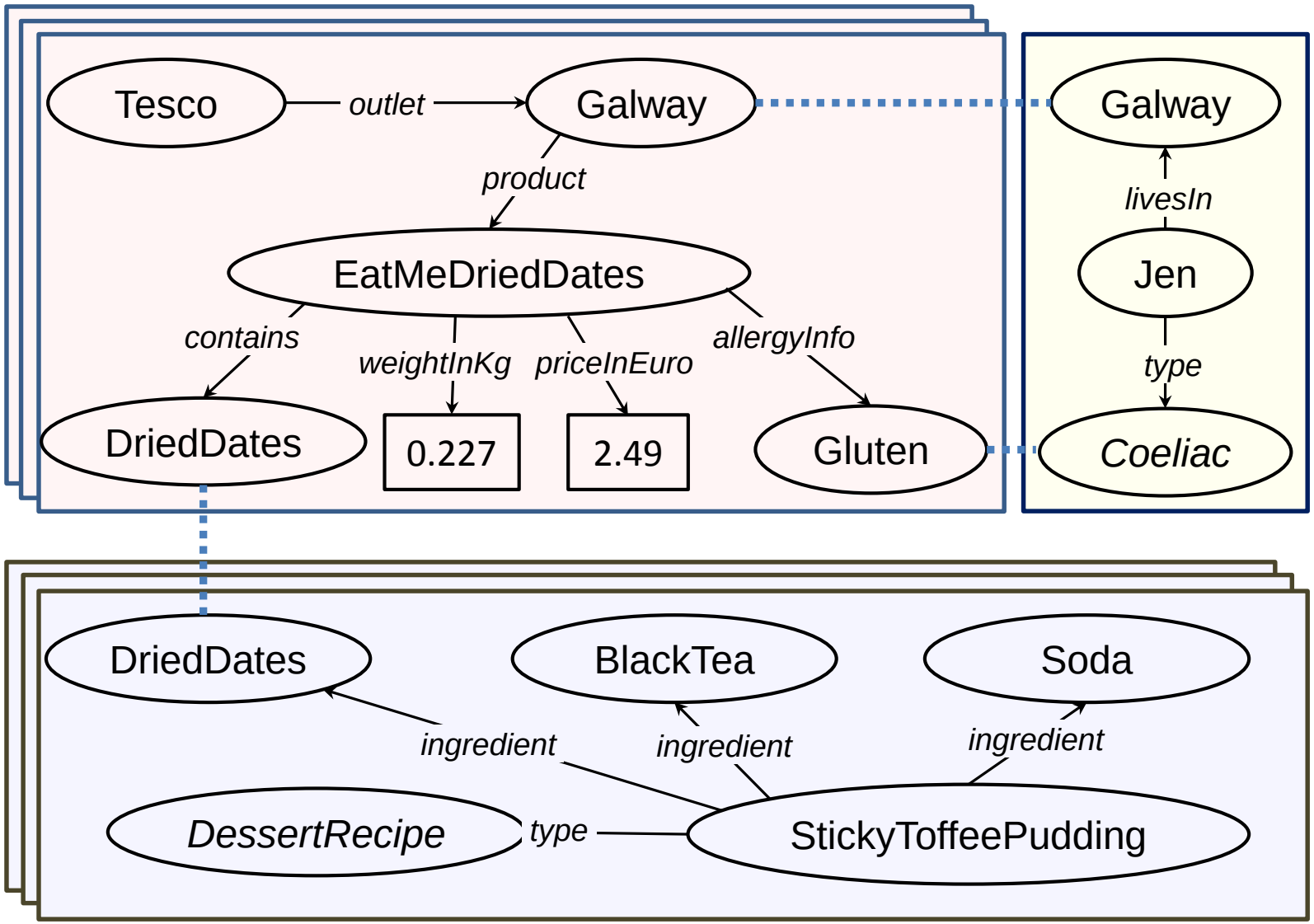
Can't trust everything you read on the Web



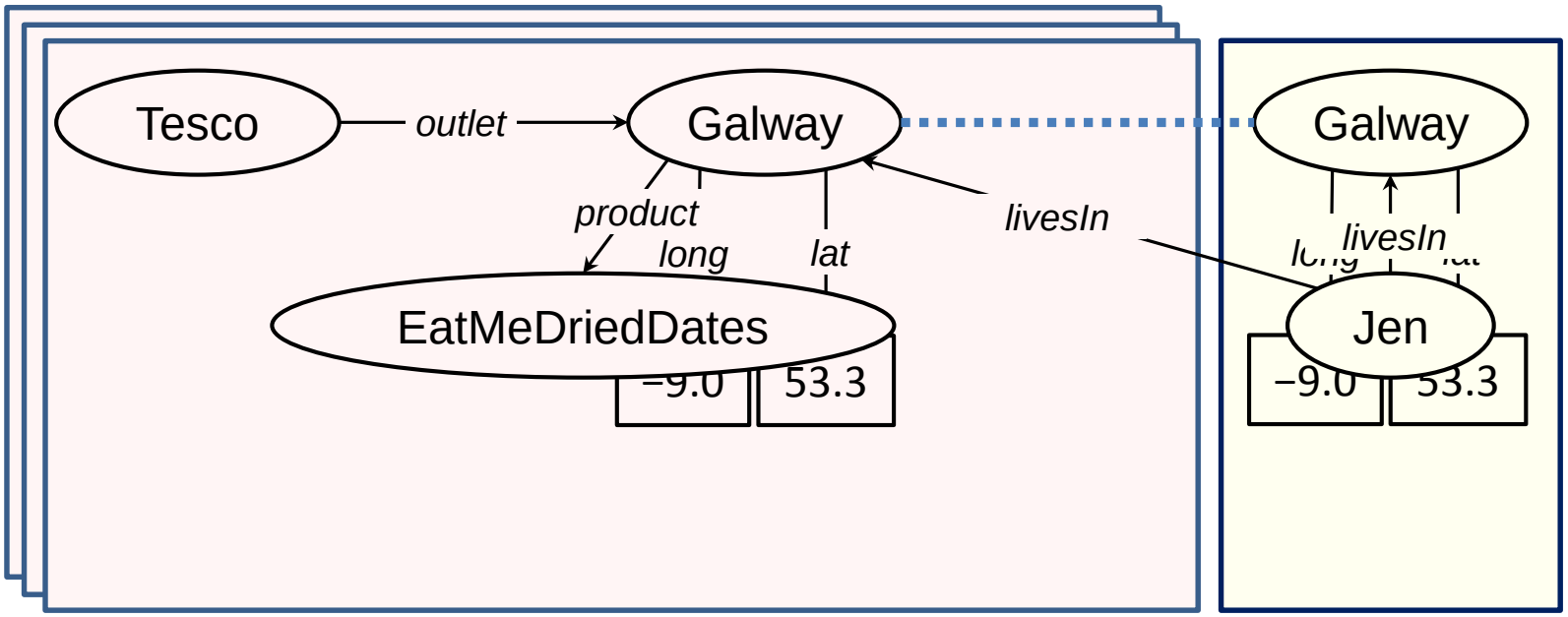
Automatic Integration possible …

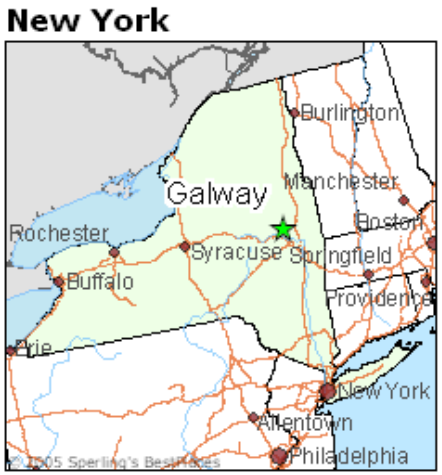


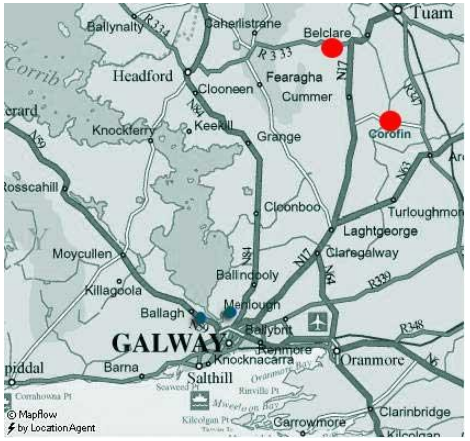
Automatic Integration possible …



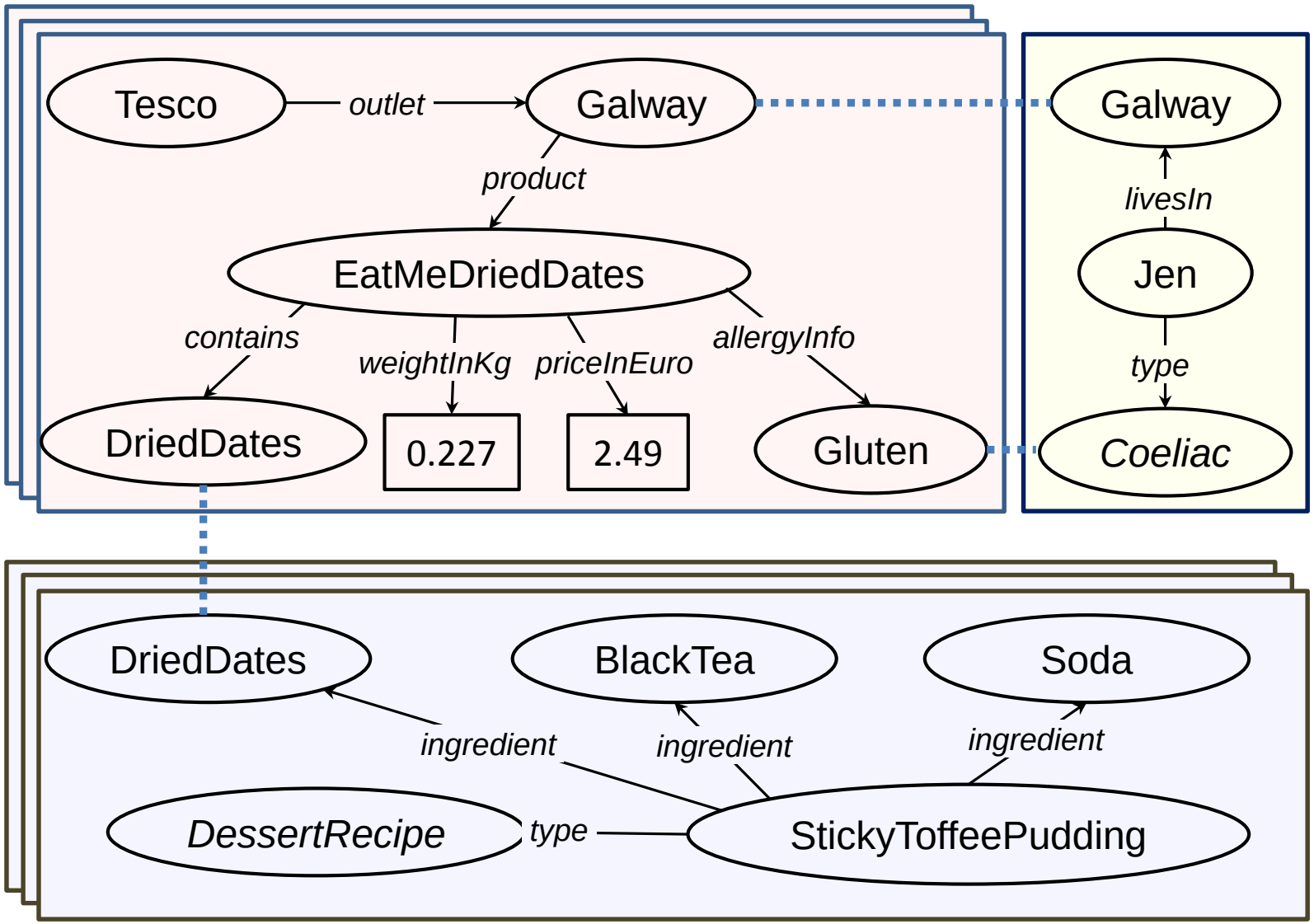
Automatic Integration possible …



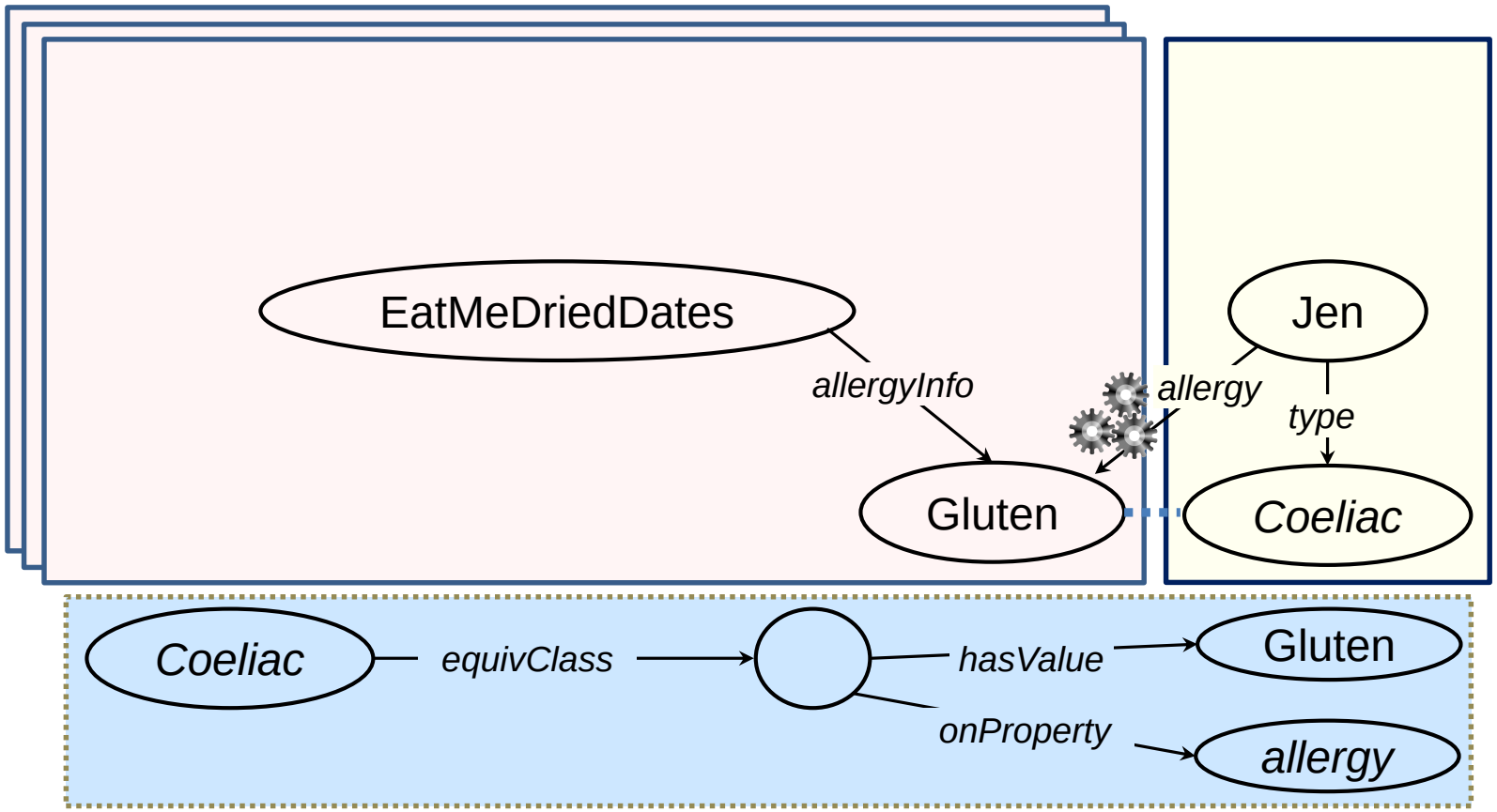




Automatic Integration possible …



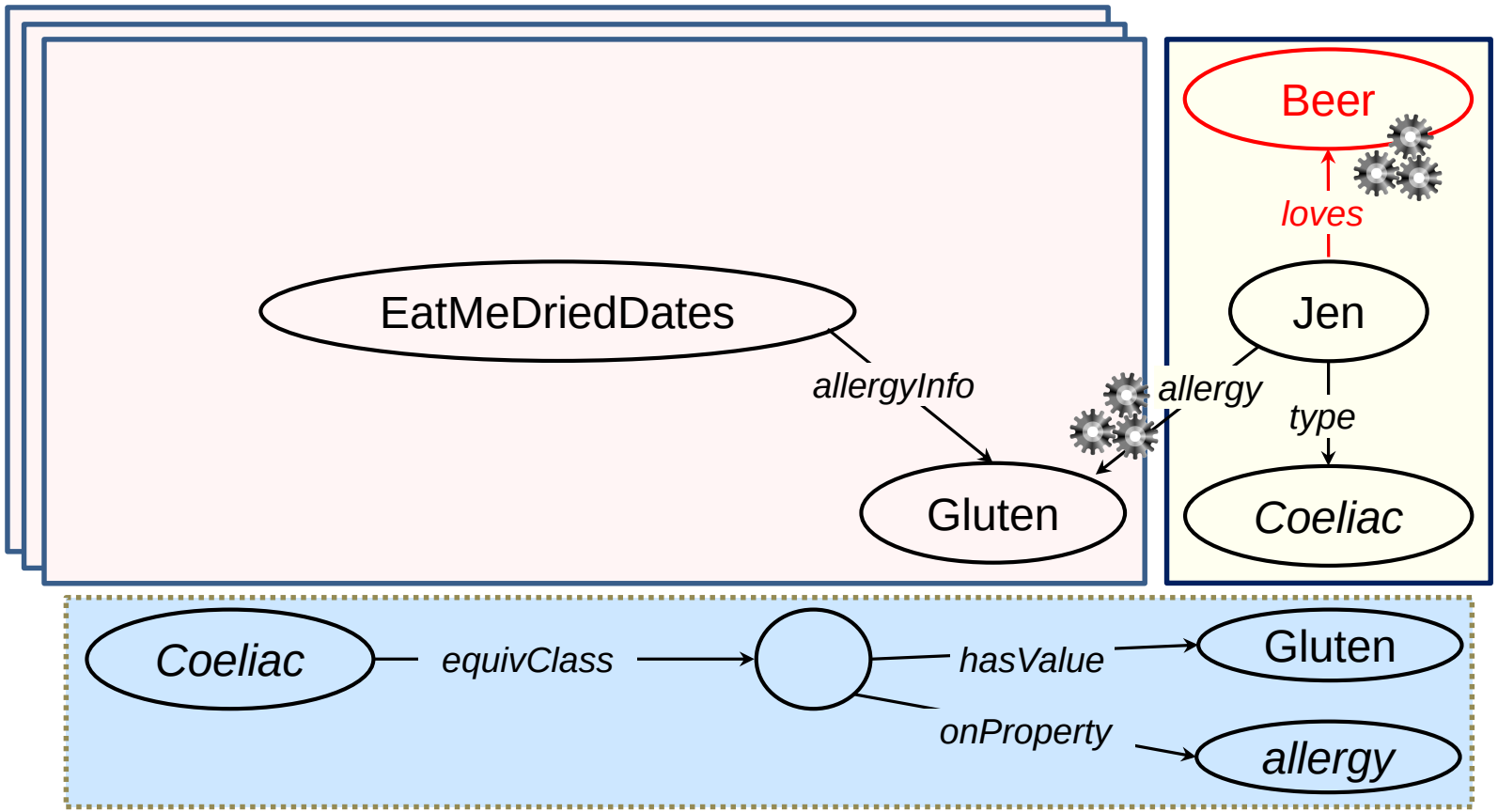
Automatic Integration possible …



Relationship between Coeliac and Gluten?

- Coeliac(Jen) [Jen is a Coeliac]
- Coeliac ≡ ∃allergy.{Gluten} [a Coeliac has allergy to Gluten]
- ∴ allergy(Jen,Gluten) [Jen has allergy to Gluten]

But not everything on the Web is true



- Coeliac(Jen)
- Coeliac ≡ ∃allergy.{Gluten}
- ∴ allergy(Jen,Gluten)

[Jen is a Coeliac]
[a Coeliac has allergy to Gluten]
[Jen has allergy to Gluten]

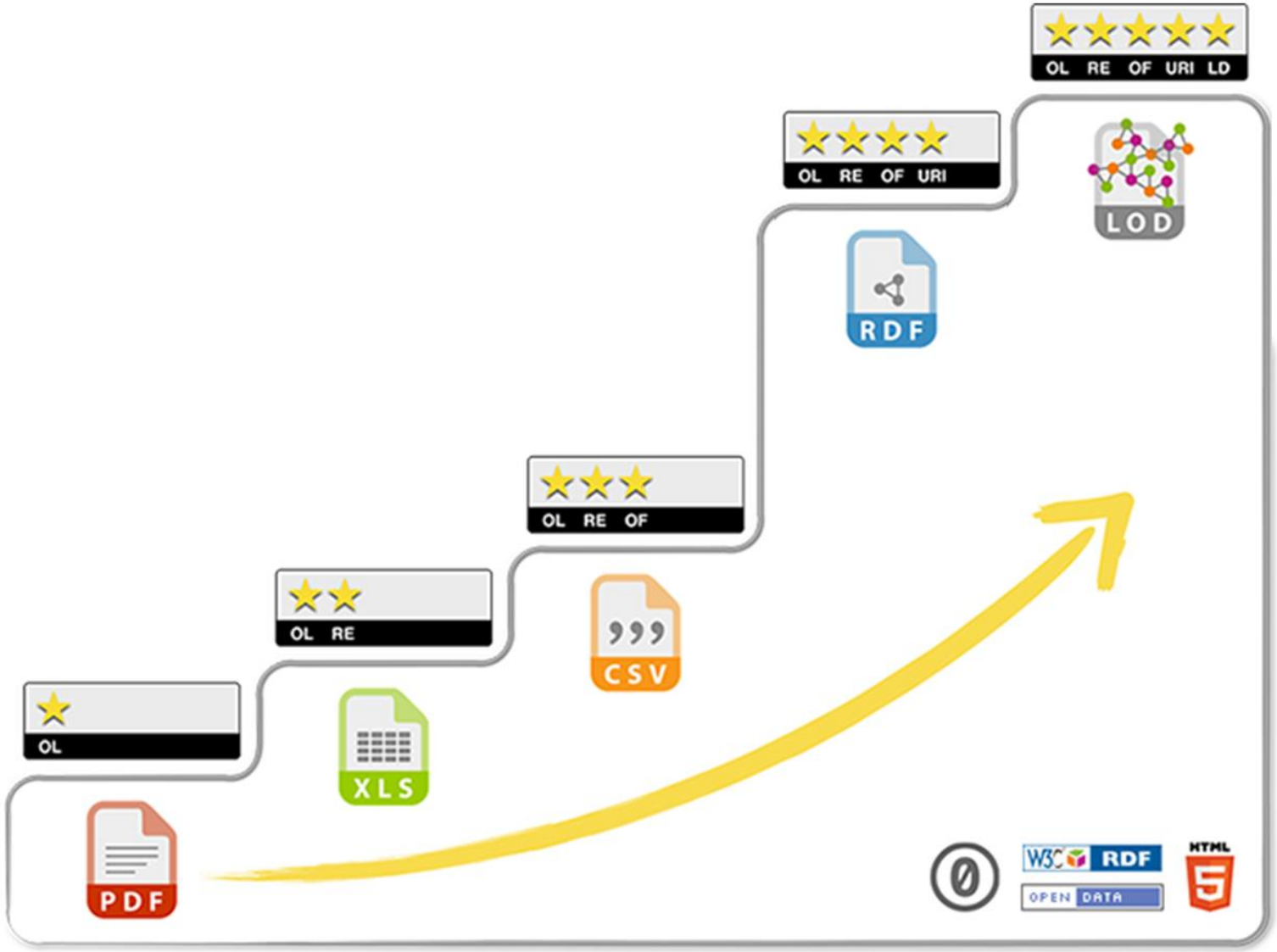
loves beer
loves beer
loves beer

Open Issue:
Legacy data

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 fundamental.
All of these problems are subject to research!
Solutions are being proposed!

Questions?