

# Semantics and Ontologies

## A Brief Introduction

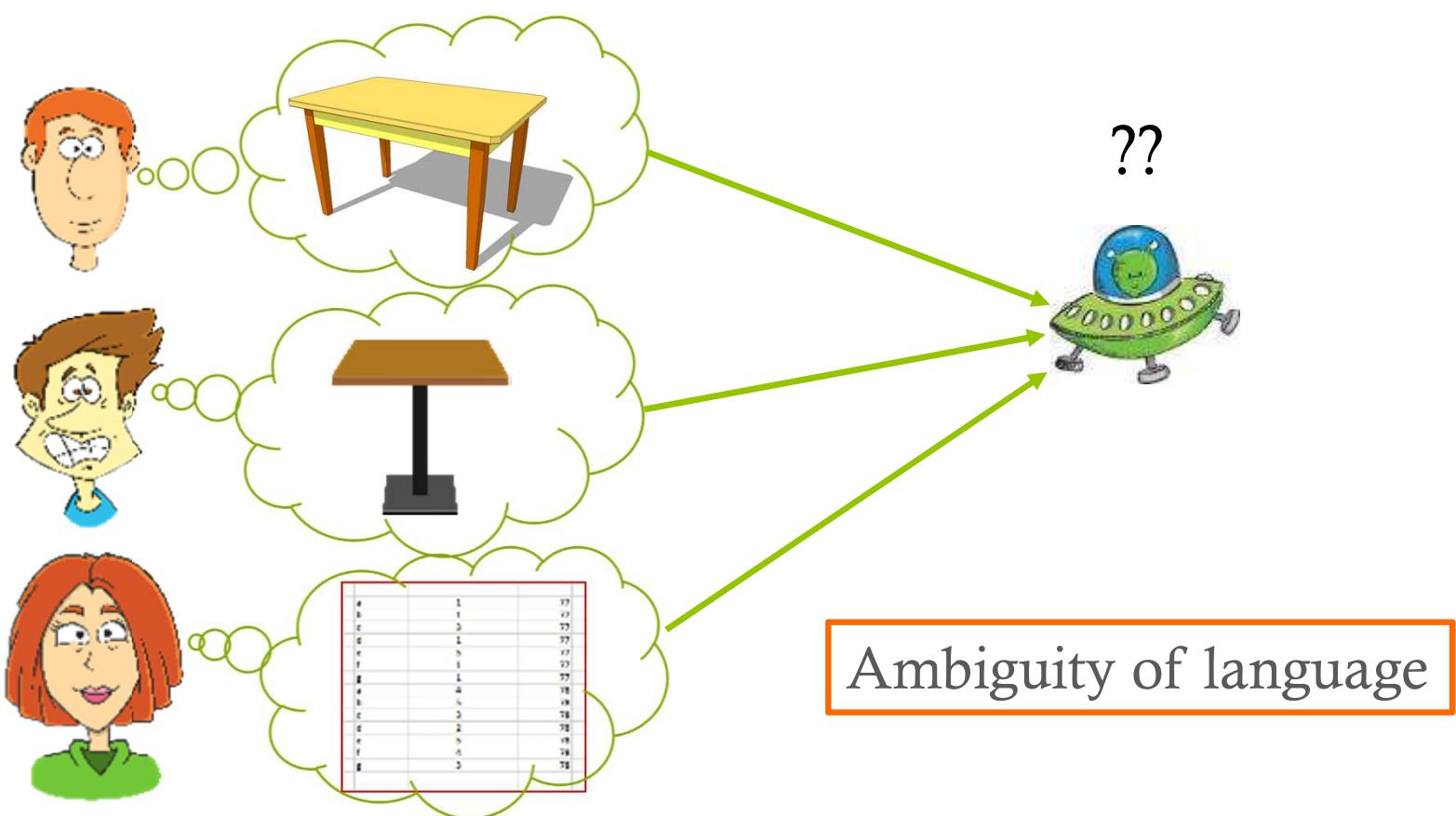
*TACC, March 2018*

**Daniel Garijo, Yolanda Gil**  
Information Sciences Institute  
and Department of Computer Science  
University of Southern California  
<http://w3id.org/people/dgarijo>  
[dgarijo@isi.edu](mailto:dgarijo@isi.edu)  
[@dgarijov](https://www.twitter.com/dgarijov)

<https://goo.gl/kBwu6f>

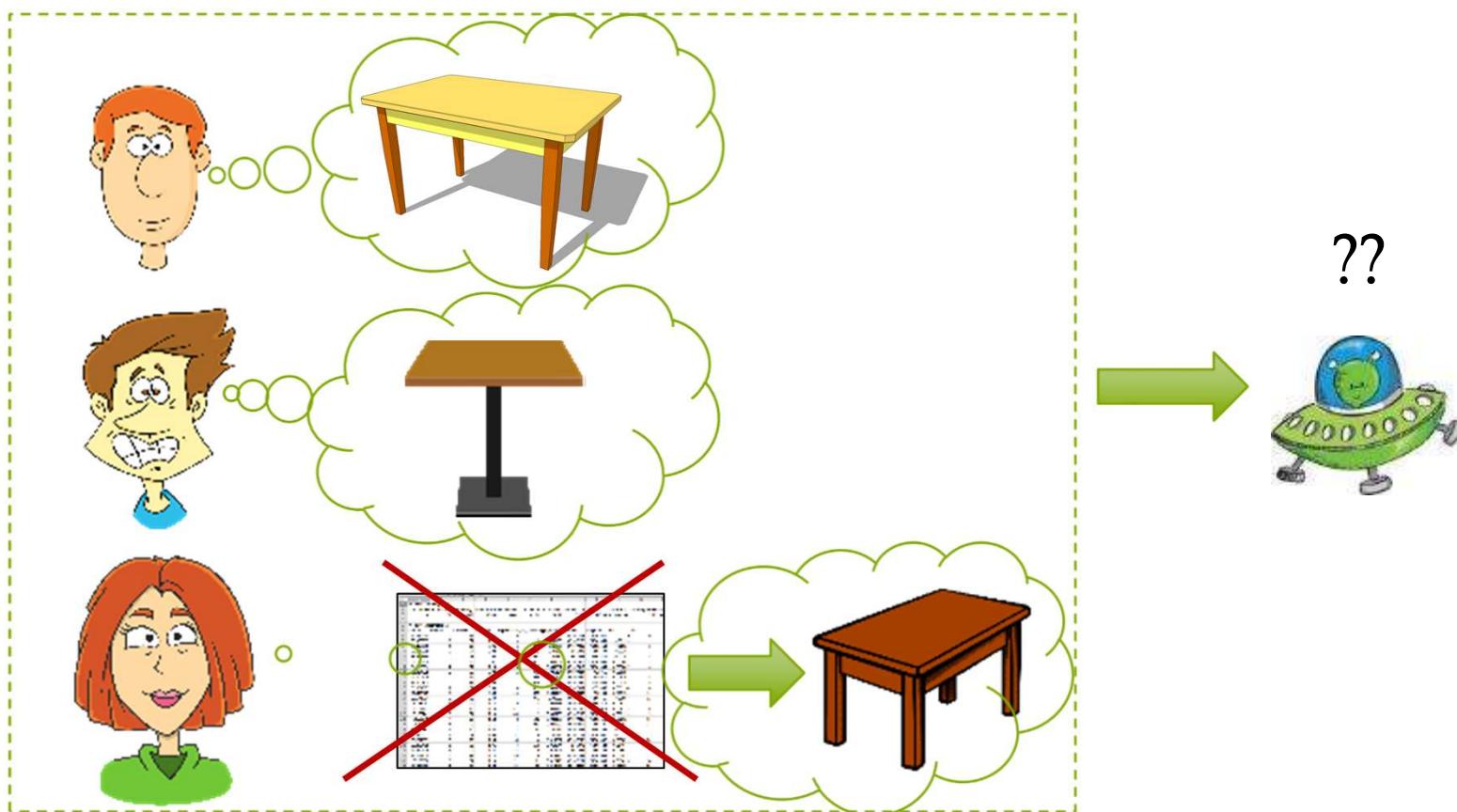
# The problem: Defining a concept

- ◆ Imagine if three people explained the concept of “table” to someone that has never seen one



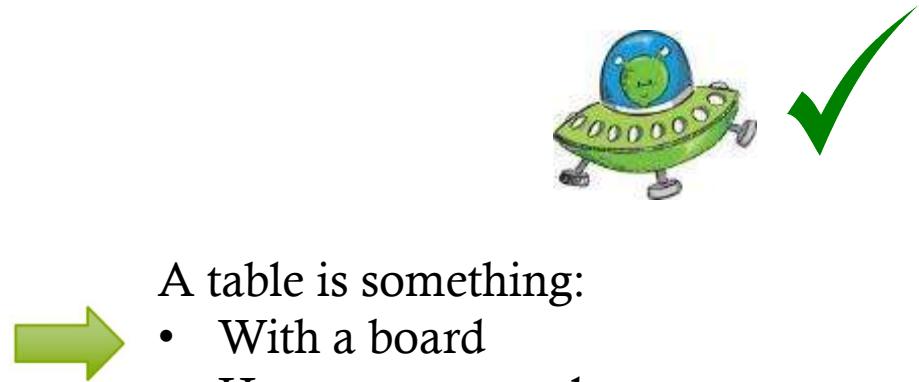
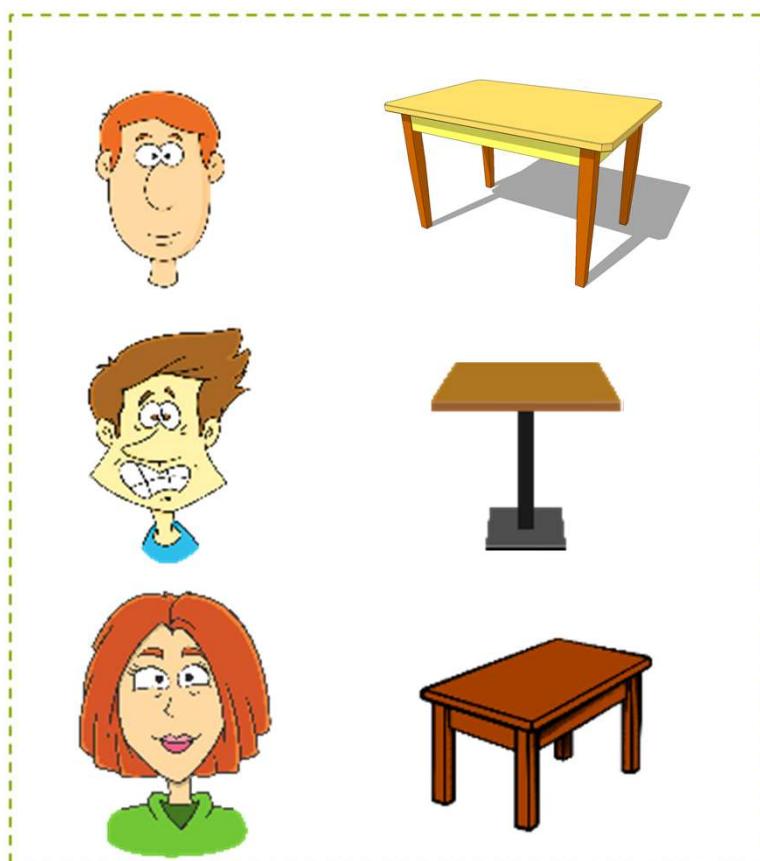
# The problem: Defining a concept

- ◆ Reaching consensus can be difficult



# The problem: Defining a concept

- ◆ After discussion, a shared conceptualization of a table is defined

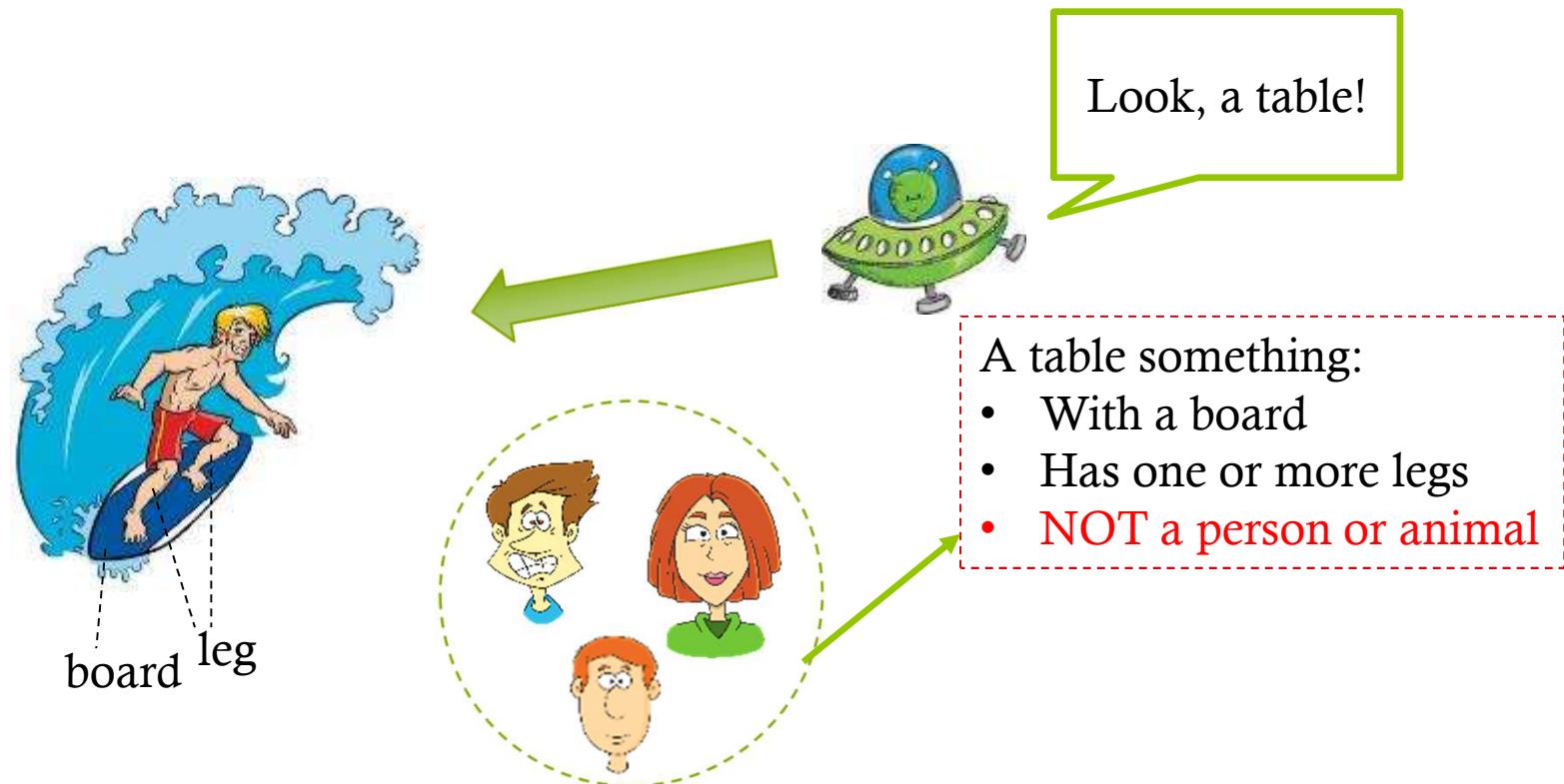


A table is something:

- With a board
- Has one or more legs

# The problem: Defining a concept

- ◆ But, is it enough?



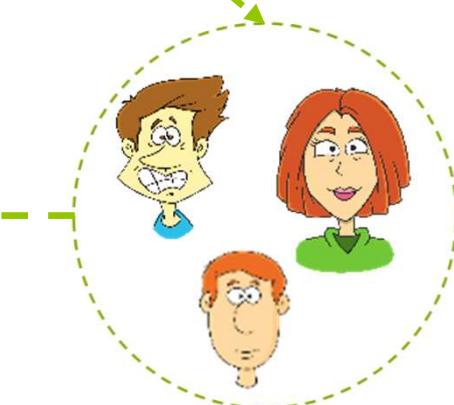
# The solution: Ontologies

- ◆ Formal specification of a shared conceptualization [1]
- ◆ Designed for machines (machine readable)



A table is something:

- With a board
- Has one or more legs
- **NOT a person or animal**



# Outline

- ◆ Background:
  - ◆ Biomedical data and the Gene Ontology
- ◆ How do ontologies work?
  - ◆ Structuring information: Ontology basics
  - ◆ The Semantic Web
- ◆ Ontologies and data on the Web
  - ◆ Linked Open Data
- ◆ Hands on: Pizza time!

# Background: Biomedical Knowledge and The Gene Ontology

# In the Beginning (circa 1998)



GenBank



# The Gene Ontology (GO)

<http://www.geneontology.org/>

## Three Organisms



## Three Focus Areas

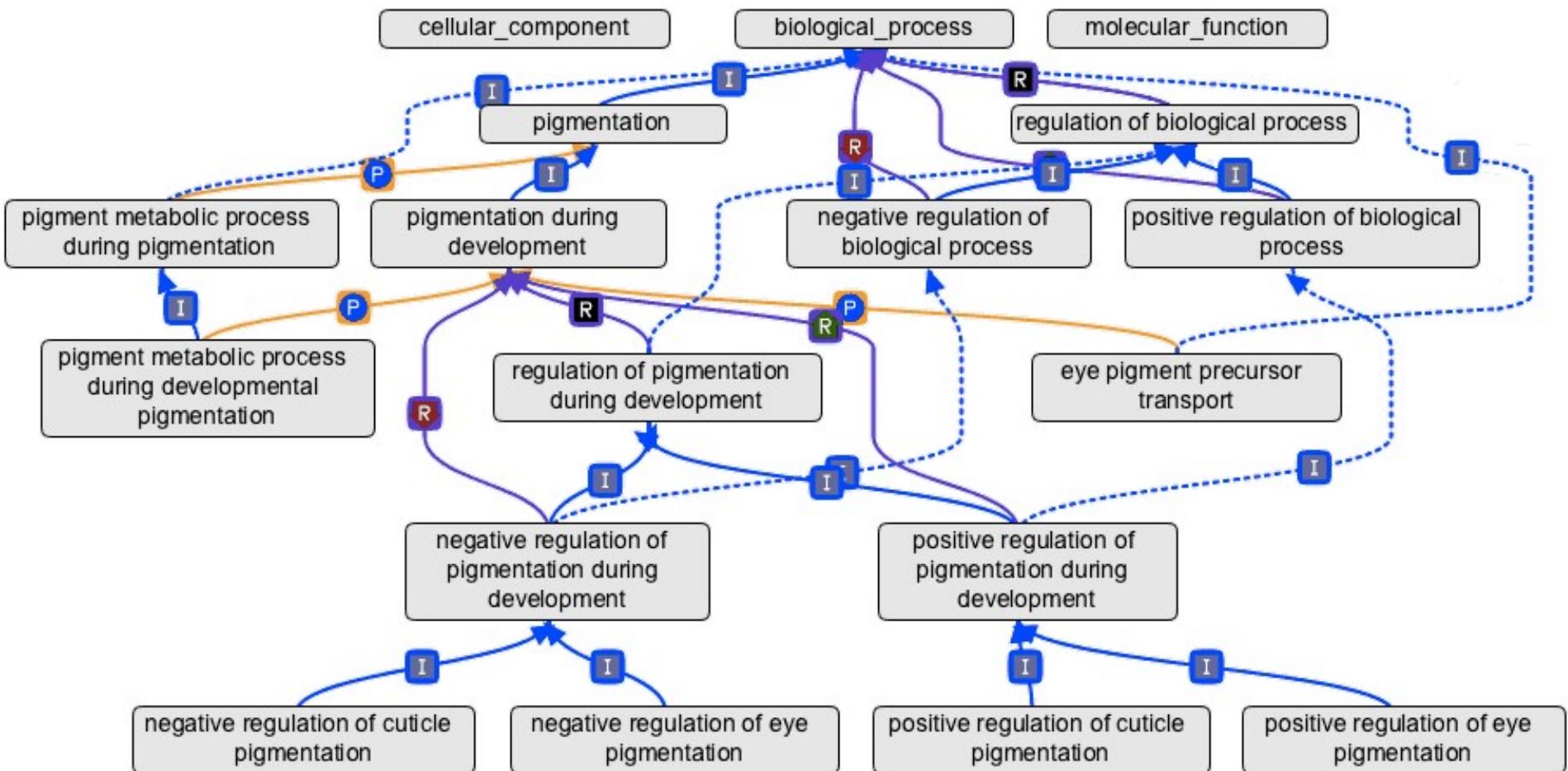
Cellular component

Biological process

Molecular function



# Concepts and the Relationships Among Them



# Sample GO Term

<http://geneontology.org/page/ontology-documentation>

## Sample GO Term

The following is a GO term taken from the OBO format file.

- id: GO:0016049
- name: cell growth
- namespace: biological\_process
- def: "The process in which a cell irreversibly increases in size over time by accretion and biosynthetic
- subset: goslim\_generic
- subset: goslim\_plant
- subset: gosubset\_prok
- synonym: "cell expansion" RELATED []
- synonym: "cellular growth" EXACT []
- synonym: "growth of cell" EXACT []
- is\_a: GO:0009987 ! cellular process
- is\_a: GO:0040007 ! growth
- relationship: part\_of GO:0008361 ! regulation of cell size

## Found entities

Total: 282; showing 1-10

Results count 10

<input type="checkbox"/>	Acc	Name	Taxon	PANTHER family	Type	Source	Synonyms	
<input type="checkbox"/>	<b>BRCA1</b>	AT4G21070	Arabidopsis thaliana		protein	TAIR	AT4G21070 ATBRCA1	
<input type="checkbox"/>	<b>Brca1</b>	breast cancer 1, early onset	Mus musculus	breast cancer type 1 susceptibility protein <b>brca1</b> pthr13763	protein	MGI	<b>Brca1</b> <b>brca1</b>	Breast cancer type 1 susceptibility protein Felis catus
<input type="checkbox"/>	<b>Brca1</b>	breast cancer 1	Rattus norvegicus	breast cancer type 1 susceptibility protein <b>brca1</b> pthr13763	gene	RGD	<b>BRCA1</b> <b>Brca1</b> <b>brca1</b>	Breast cancer type 1 susceptibility protein Homo sapiens
<input type="checkbox"/>	<b>Brca1</b>	Breast cancer type 1 susceptibility protein homolog	Rattus norvegicus		protein	UniProtKB	<b>BRCA1_RAT</b>	Canis lupus familiaris
<input type="checkbox"/>	<b>BRCA1</b>	Breast and ovarian cancer susceptibility-like protein	Gallus gallus		protein	UniProtKB	<b>Q90Z51_CHICK</b>	Equus caballus
<input type="checkbox"/>	<b>BRCA1</b>	Uncharacterized protein	Takifugu rubripes	breast cancer type 1 susceptibility protein <b>brca1</b> pthr13763	protein	UniProtKB	<b>H2TDW1_TAKRU</b>	Xenopus tropicalis



Gene Ontology widely adopted

## FlyBase



UNIVERSITY OF EDINBURGH INSTITUTE OF CELLULAR BIOLOGY  
UNIVERSITY OF EDINBURGH INSTITUTE OF GENETICS  
UNIVERSITY OF EDINBURGH INSTITUTE OF MEDICAL GENETICS  
UNIVERSITY OF EDINBURGH INSTITUTE OF MOLECULAR MEDICINE  
**UniProt**  
the universal protein resource



InterPro



A Comparative mapping resource  
**GRAMENE**



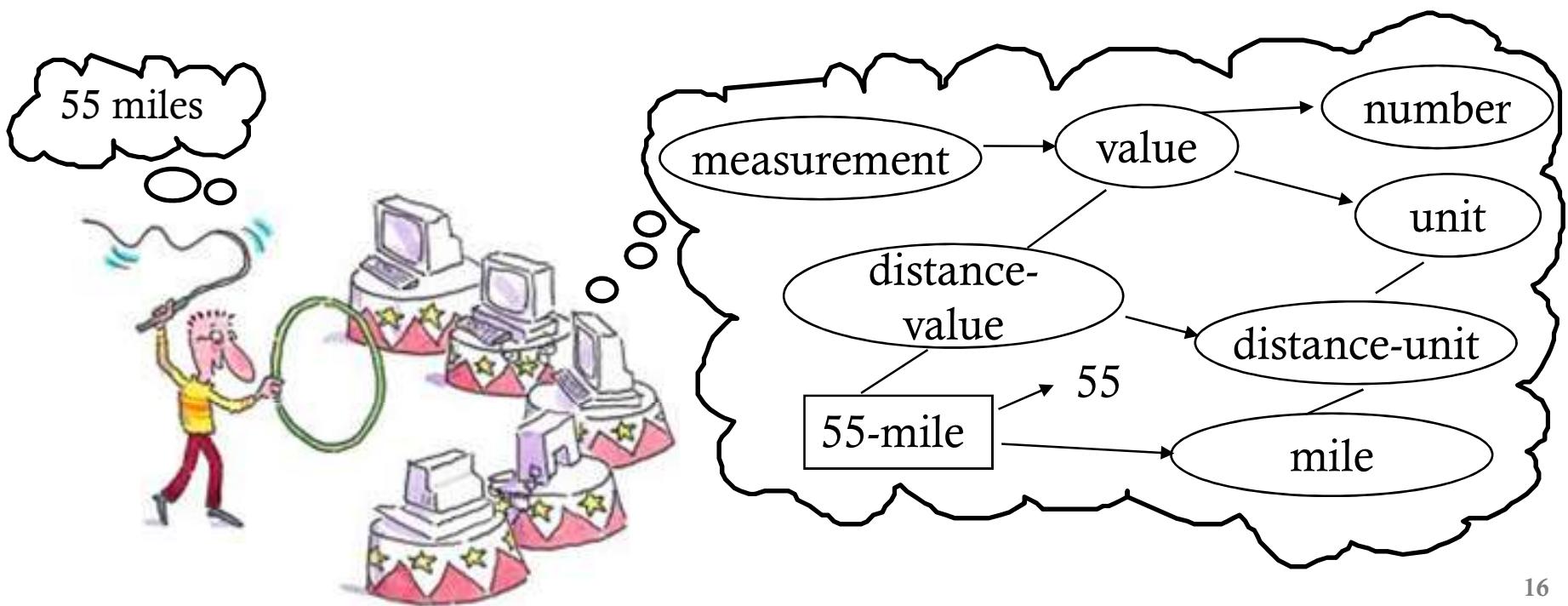
The Wellcome Trust  
Sanger Institute



# Ontology Basics

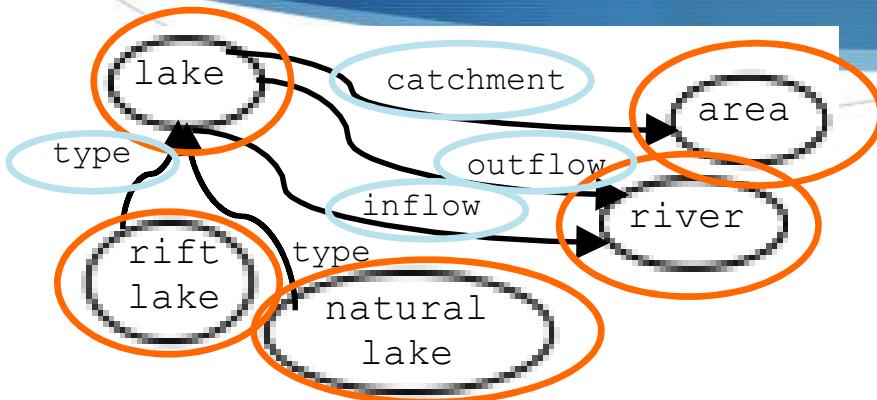
# Knowledge Representation

- ◆ A subfield of artificial intelligence concerned with capturing formal (machine readable) information about the world
- ◆ Ontologies are one approach to knowledge representation
  - ◆ Focused on descriptions of objects
  - ◆ Based on mathematical logic

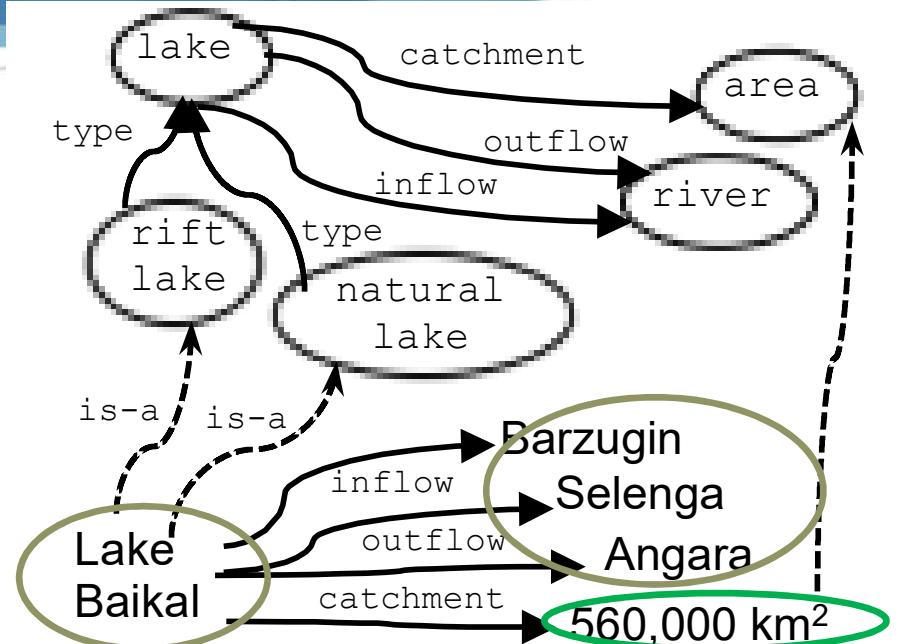


# Ontology Basics

## Ontology



## Assertions (Data)



- ◆ Concepts (or classes)
- ◆ Properties
  - ◆ Domain -- applicability
  - ◆ Range – possible values

- ◆ Instances
- ◆ Property values

# Ontologies vs Taxonomies vs Vocabularies

- ◆ **Vocabularies:** a set of terms
- ◆ **Taxonomies:** a hierarchical organization of terms
- ◆ **Ontologies:** formal taxonomies with logic constraints
  1. Mathematical logic provides formal semantics for what the hierarchies mean
    - ◆ Subclasses imply containment of instances
    - ◆ Instances of a class have all the properties of the class
  2. Logic constraints
    - ◆ Example: all humans have **exactly one** biological mother

# Benefits – Take 1

1. Provide standards for data integration
  - ◆ eg the Gene Ontology
2. Data validation
  - ◆ Ensure that the data represent what was intended
3. Programmatic access to data
  - ◆ Enable structured queries

# The Semantic Web

# The Semantic Web

Ontology about  
core samples

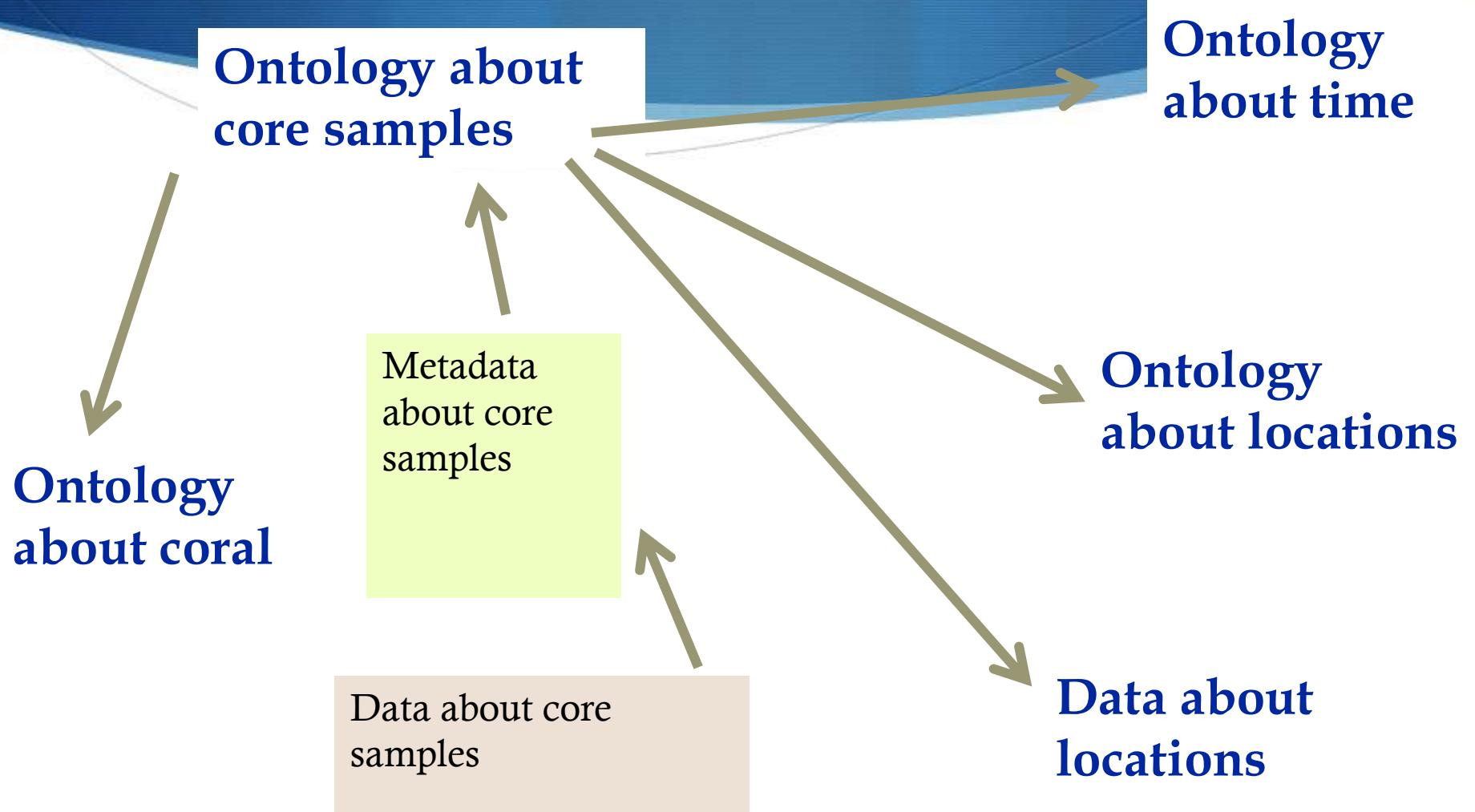


Metadata  
about core  
samples

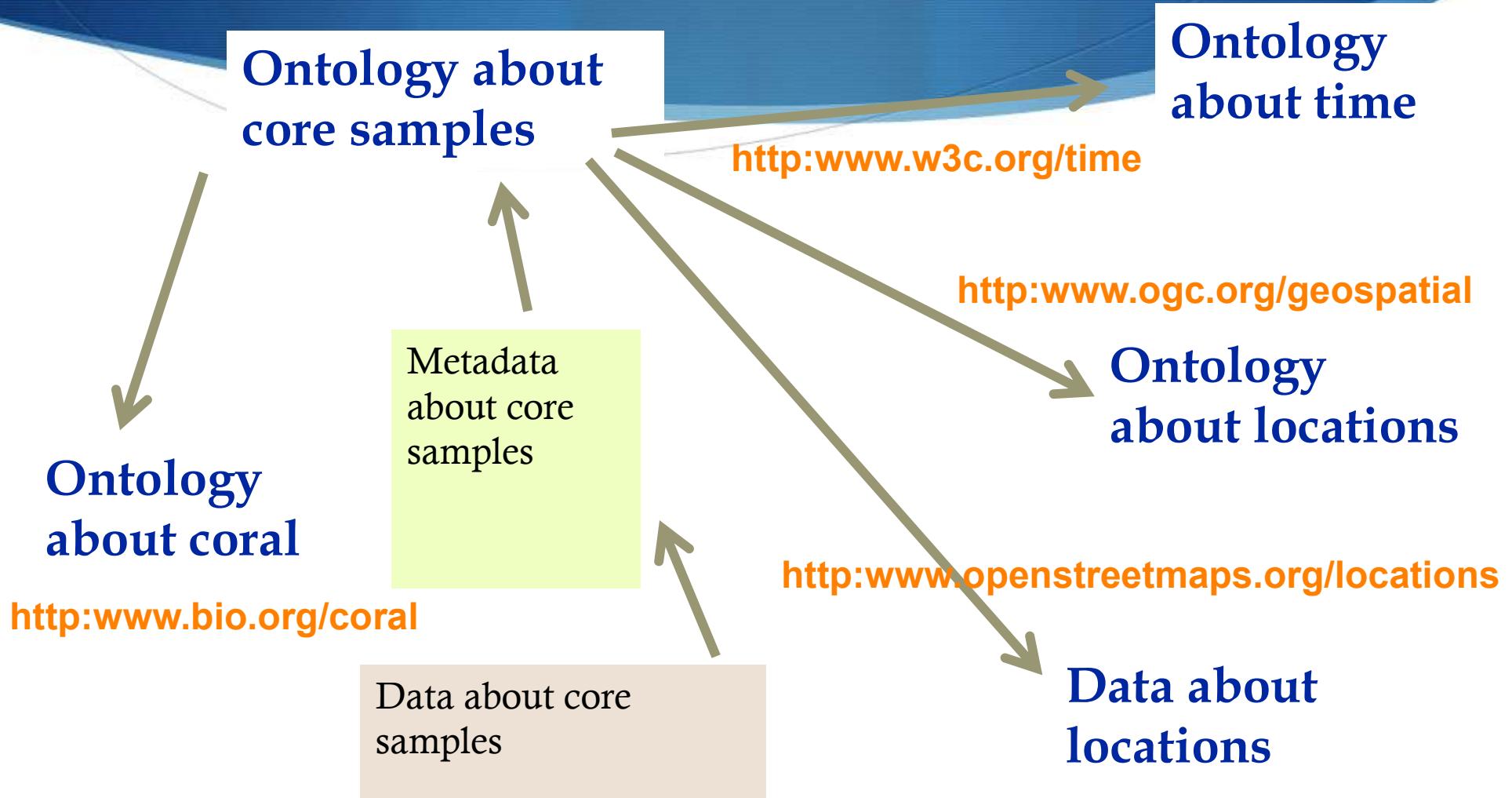


Data about core  
samples

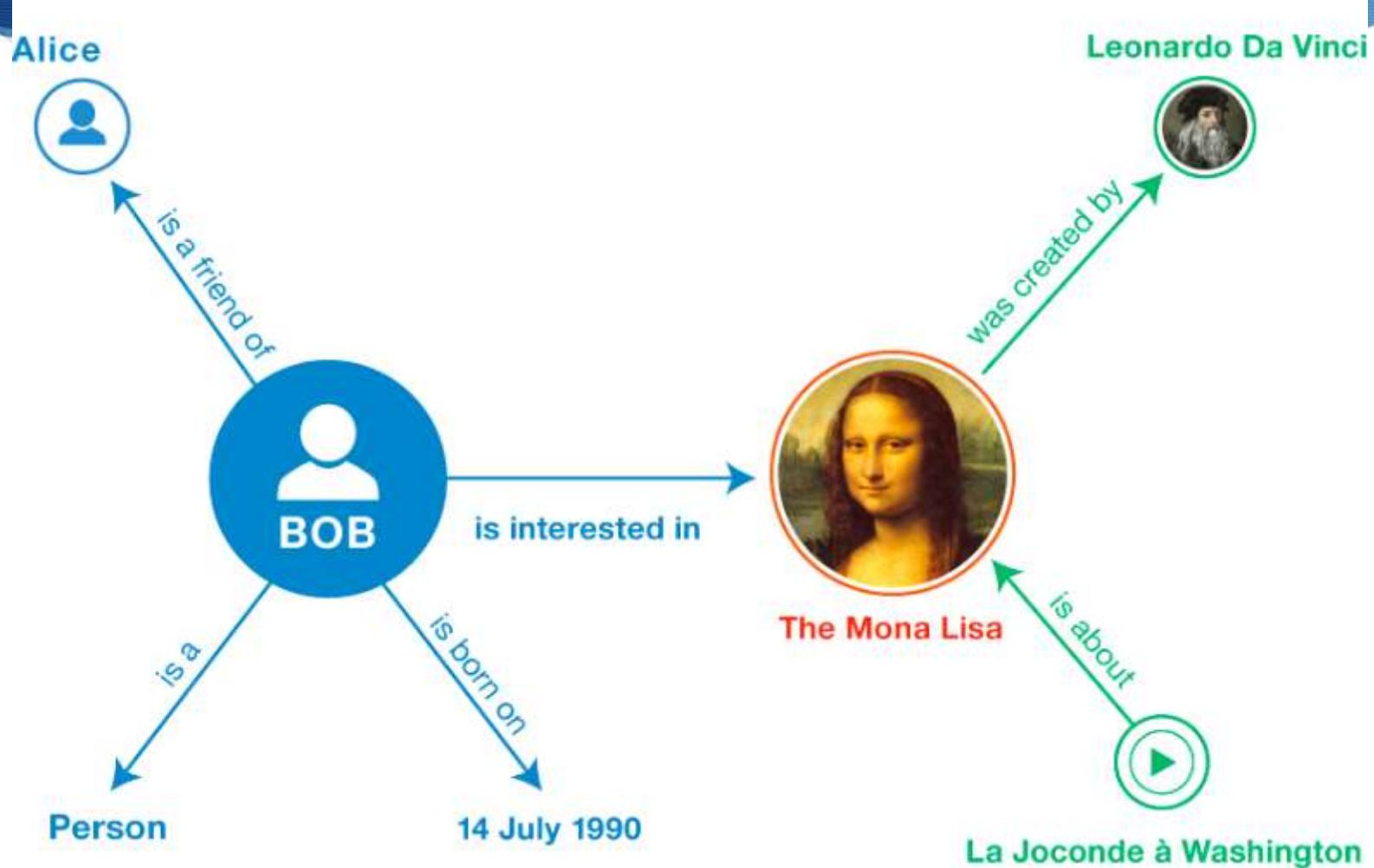
# The Semantic Web



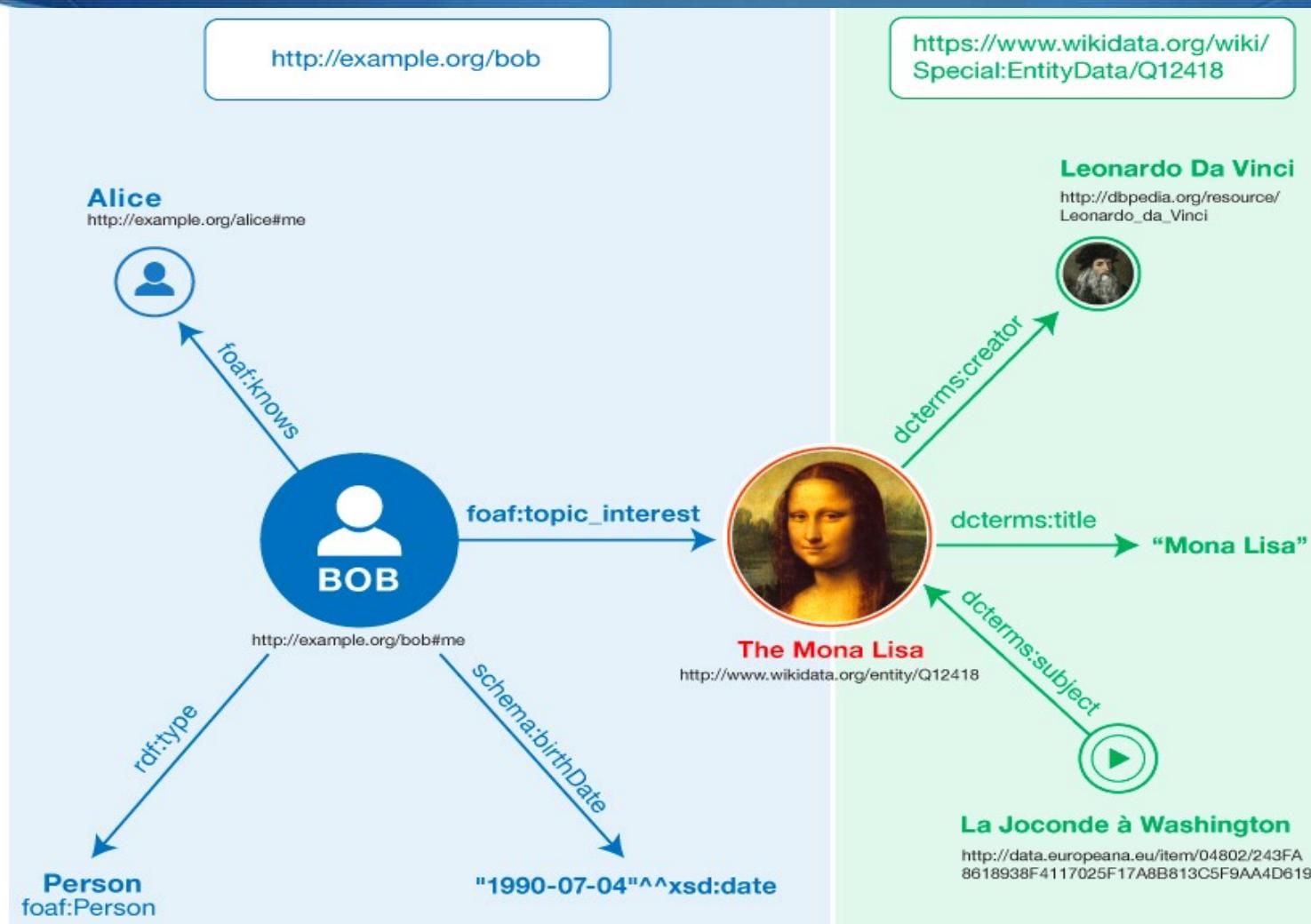
# The Semantic Web



# RDF: The Resource Description Framework



# Giving Meanings to Hyperlinks on the Web



<http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>



# Connecting to Other Ontologies/Data

**Palmyra Atoll**

Structured Properties

main type (GND)	geographical feature	[edit]
is in the administrative unit	United States Minor Outlying Islands	[edit]

NCBITaxon ontology:  
porites <http://bioportal1.bioontology.org/ontologies/NCBITAXON?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FNCBITAXON%2F46719>

**Porites**

Structured Properties

Property Name	Topic: Finger Coral	[edit]
• [x] <a href="http://dbpedia.org/resource/Porites">http://dbpedia.org/resource/Porites</a>		
• [edit source]		

Wikipedia Entry [go to original Wikipedia article](#)

Porites is a genus of stony coral; they are SPS (Small Polyp Stony) corals. They are characterised by a finger-like morphology. Members of this genus have widely spaced

**Palmyra coral 20C**

Data

- DOWNLOAD

From: <http://www.ncdc.noaa.gov/paleo/metadata/noaa-coral-1865.html>

Structured Properties

[x] SiteName	Palmyra	(By Julien)
[x] Archive	Coral	(By Julien)
[x] Domain(s)	Climate,geochemistry	(By Julien)
[x] Forward model	10.1029/2011GL048224	(By Julien)
[x] Genus	Porites	(By Julien)
[x] Interpretation	SST,SSS	(By Nick)
[x] Measurement	Delta18O	(By Julien)
[x] MeasurementMaterial	Skeletal aragonite	(By Julien)
[x] MeasurementStandard	VPDB	(By Nick)
[x] MeasurementUnits	Permil	(By Julien)
[x] Reference	Cobb et al, 2001	(By Julien)
[x] Species	Lutea	(By Julien)

Credits

Users who have contributed to this Page:

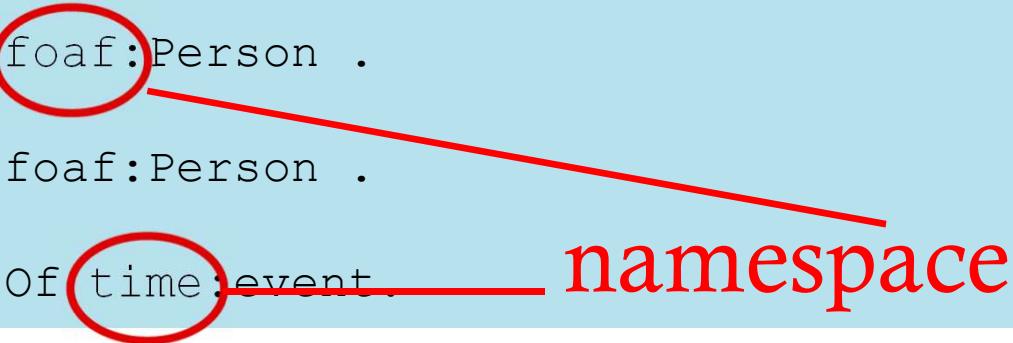
- Julien (43 Edits)
- Nick (34 Edits)






# Reusing Ontologies Through URLs/URIs

```
@prefix : <http://www.usc.edu/people#> .  
@prefix foaf: <http://www.w3.org/2012/foaf#> .  
@prefix time: <http://www.w3.org/2009/time#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix owl: <http://www.w3.org/2002/07/owl#> .  
  
:faculty rdfs:subClassOf foaf:Person .  
  
:student rdfs:subClassOf foaf:Person .  
  
:graduation rdfs:subClassOf time:event .
```



namespace

# Semantic Search Engines

<http://watson.kmi.open.ac.uk/WatsonWUI/>



[What is it?](#) - [Submit URI](#) - [Website](#) - [APIs](#)

person

[Search Watson](#)

[Search Options](#)

<http://www.w3.org/2000/10/swap/pim/contact.rdf> +

- ◆ <http://www.w3.org/2000/10/swap/pim/contact#Person>
- © <http://www.w3.org/2000/10/swap/pim/contact#ContactLocation>
- ¶ <http://www.w3.org/2000/10/swap/pim/contact#knownAs>
- © <http://www.w3.org/2000/10/swap/pim/contact#SocialEntity>

<http://sw.cyc.com/2006/07/27/P/e/Person.owl> +

- © <http://sw.cyc.com/2006/07/27/cyc/#Person>

<http://www.xml.com/2004/09/01/examples/Politics.rdf> +

- © <http://hackingcongress.org/ns/Politics#Person>
- ¶ <http://hackingcongress.org/ns/Politics#Gender>
- ¶ <http://hackingcongress.org/ns/Politics#BirthPlace>
- ¶ <http://hackingcongress.org/ns/Politics#HomeTown>
- ¶ <http://hackingcongress.org/ns/Politics#Birthday>
- ¶ <http://hackingcongress.org/ns/Politics#FullName>

# schema.org

schema.org

## Review

Thing > CreativeWork > Review

A review of an item – for example, of a restaurant, n

Usage: Between 250,000 and 500,000 domains

Property	On Types
<code>resultReview</code>	ReviewAction
<code>review</code>	Organization or Place or Event or Service or CreativeWork or Product or Brand or Offer

The screenshot shows a search results page for "cannon rebel" on a search engine interface featuring Bing, Yahoo!, Google, and Yandex. The results are filtered under the "All" tab. The first result is for EOS DSLR Cameras | EOS Rebel T5 18-55 IS II Kit, with a green arrow pointing to it. The second result is for Canon EOS Rebel T5 DSLR Camera 2 Lens Bundle - Costco, also with a green arrow. The third result is for Canon T5 EOS Rebel DSLR Camera with EF-S 18-55mm IS ..., and the fourth result is for Canon EOS Rebel T5 DSLR Camera with 18-55mm and ... Both of these last two have green arrows pointing to them.

Bing YAHOO! Yandex

cannon rebel

All Shopping Images Videos News More Search tools

About 16,400,000 results (1.08 seconds)

Showing results for **canon rebel**  
Search instead for [cannon rebel](#)

**EOS DSLR Cameras | EOS Rebel T5 18-55 IS II Kit | Canon ...**  
<https://www.usa.canon.com/.../eos-rebel-t5-ef-s-18-55-is-ii-kit> ▾ Canon Inc. ▾  
★★★★★ Rating: 4.5 - 34 reviews  
Perfect for families, budding photo enthusiasts and first-time SLR users alike, the EOS Rebel T5 makes it easy to capture movies and photos that are nothing ...

**Canon EOS Rebel T5 DSLR Camera 2 Lens Bundle - Costco**  
[www.costco.com/Canon-EOS-Rebel-T5-DSLR-Camera-2-Lens-Bu...](http://www.costco.com/Canon-EOS-Rebel-T5-DSLR-Camera-2-Lens-Bu...) ▾ Costco ▾  
★★★★★ Rating: 4.8 - 10 reviews  
Canon EOS Rebel T5 DSLR Camera 2 Lens Bundle EF-S 18-55mm f/3.5-5.6 IS II and EF75/300 III Lens, 16GB SD Card, Camera Bag, EOS Digital Solution Disc ...

**Canon T5 EOS Rebel DSLR Camera with EF-S 18-55mm IS ...**  
[www.bhphotovideo.com/.../DSLR-Cameras](http://www.bhphotovideo.com/.../DSLR-Cameras) ▾ B & H Photo Video ▾  
★★★★★ Rating: 4.5 - 117 reviews - \$399.00 - In stock  
Canon EOS T5 Rebel DSLR Camera with EF-S 18-55mm IS II Lens, Canon EOS 1200D features 18MP APS-C CMOS Sensor, EF-S 18-55mm IS II Lens ...

**Canon EOS Rebel T5 DSLR Camera with 18-55mm and ...**  
[www.bhphotovideo.com/.../DSLR-Cameras](http://www.bhphotovideo.com/.../DSLR-Cameras) ▾ B & H Photo Video ▾  
★★★★★ Rating: 5 - 74 reviews - \$499.00 - In stock  
Buy Canon EOS Rebel T5 DSLR Camera with 18-55mm and 75-300mm Lenses Bundle features 18MP APS-C CMOS Sensor, DIGIC 4 Image Processor.

# Microdata

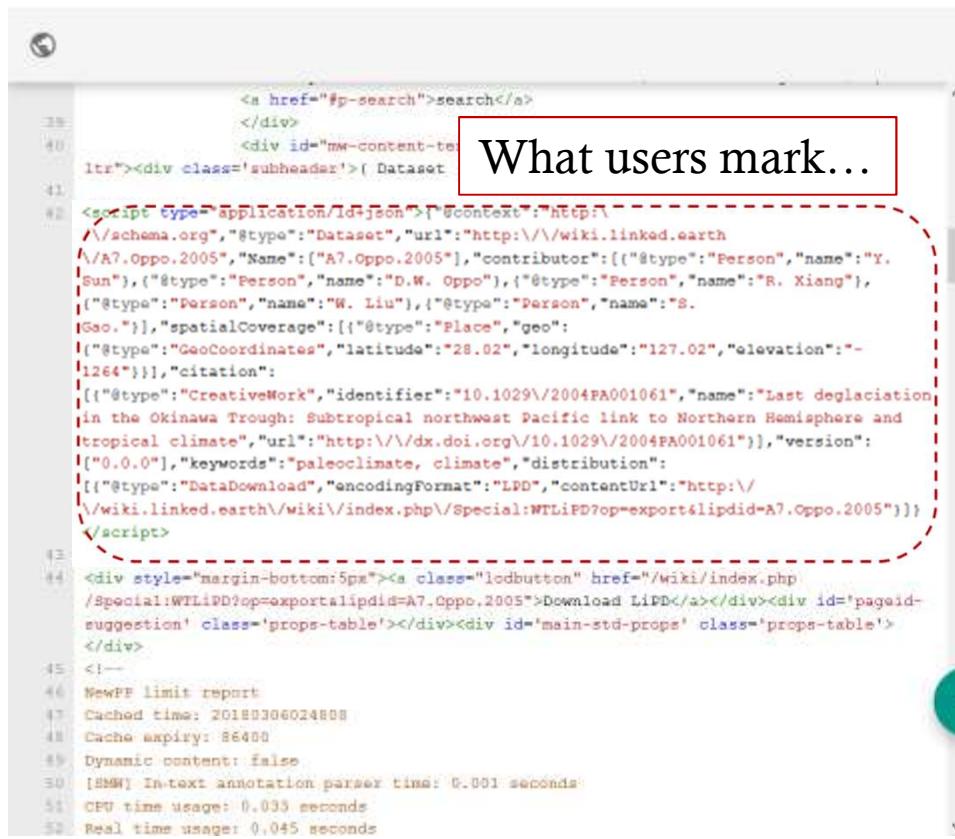
- ◆ Microdata is markup added to HTML files that search engines use

```
<section> Hello, my name is John Doe, I am a graduate research assistant at  
the University of Dreams.  
My friends call me Johnny.  
You can visit my homepage at <a href="http://www.JohnnyD.com">www.JohnnyD.com</a>.  
I live at 1234 Peach Drive, Warner Robins, Georgia.</section>
```

```
<section itemscope itemtype="http://schema.org/Person">  
Hello, my name is  
<span itemprop="name">John Doe</span>,  
I am a  
<span itemprop="jobTitle">graduate research assistant</span>  
at the  
<span itemprop="affiliation">University of Dreams</span>.  
My friends call me  
<span itemprop="additionalName">Johnny</span>.  
You can visit my homepage at  
<a href="http://www.JohnnyD.com" itemprop="url">www.JohnnyD.com</a>.  
<section itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">  
I live at  
<span itemprop="streetAddress">1234 Peach Drive</span>,  
<span itemprop="addressLocality">Warner Robins</span>,  
<span itemprop="addressRegion">Georgia</span>.  
</section>  
</section>
```

# JSON-LD snippets

- ◆ Markup added to HTML files that search engines use (alternative to microdata)



The screenshot shows a browser's developer tools with the element inspector open. A red box highlights a portion of the page source code containing JSON-LD markup. Below this, another red box highlights the text "What users mark...". A dashed red arrow points from the "What users mark..." box to the JSON-LD code in the source.

What users mark...

```
43 <a href="#p-search">search</a>
44 </div>
45 <div id="mw-content-te
46 ltr"><div class='subheader'> Dataset
47
48 <script type="application/ld+json">{"@context": "http://
49 //schema.org", "@type": "Dataset", "url": "http://wiki.linked.earth
50 /A7.Oppo.2005", "Name": ["A7.Oppo.2005"], "contributor": [{"@type": "Person", "name": "Y.
51 Sun"}, {"@type": "Person", "name": "D.W. Oppo"}, {"@type": "Person", "name": "R. Xiang"}, {"@type": "Person", "name": "W. Liu"}, {"@type": "Person", "name": "S.
52 Gao."}], "spatialCoverage": [{"@type": "Place", "geo": {"@type": "GeoCoordinates", "latitude": "28.02", "longitude": "127.02", "elevation": "-1264"}]}, "citation": [{"@type": "CreativeWork", "identifier": "10.1029/2004PA001061", "name": "Last deglaciation in the Okinawa Trough: Subtropical northwest Pacific link to Northern Hemisphere and tropical climate", "url": "http://dx.doi.org/10.1029/2004PA001061"}], "version": ["0.0.0"], "keywords": "paleoclimate, climate", "distribution": [{"@type": "DataDownload", "encodingFormat": "LPP", "contentUrl": "http://
53 /wiki.linked.earth/wikiv/index.php?Special:WTLiPD?op=export&lipid=A7.Oppo.2005"}]}
54 </script>
55
56 <div style="margin-bottom:5px"><a class="lcbbutton" href="/wiki/index.php?
57 /Special:WTLiPD?op=export&lipid=A7.Oppo.2005">Download LiPD</a></div><div id="pageid-
58 suggestion" class='props-table'></div><div id="main-std-props" class='props-table'>
59 </div>
60 <!--
61 NewPF limit report
62 Cached time: 20180306024808
63 Cache expiry: 86400
64 Dynamic content: false
65 [BBMF] In-text annotation parser time: 0.001 seconds
66 CPU time usage: 0.033 seconds
67 Real time usage: 0.045 seconds
68
```

What engines see...

Dataset	0 ERRORES - 0 ADVERTENCIAS
@type	Dataset
url	http://wiki.linked.earth/A7.Oppo.2005
version	0.0.0
keywords	paleoclimate, climate
name	A7.Oppo.2005
contributor	
@type	Person
name	Y. Sun
contributor	
@type	Person
name	D.W. Oppo
contributor	
@type	Person
name	R. Xiang
contributor	
@type	Person
name	W. Liu
contributor	
@type	Person
name	S. Gao..
spatialCoverage	Place
@type	

# The Google Knowledge Graph



Google john lennon

All News Images Videos Books More Search tools

About 62,900,000 results (0.50 seconds)

**John Lennon Official Website**  
[www.johnlennon.com/](http://www.johnlennon.com/) ▾  
An official website that features news, biography, videos, historical timeline, downloads, and drawings.

**John Lennon - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/John\\_Lennon](https://en.wikipedia.org/wiki/John_Lennon) ▾ Wikipedia ▾  
John Winston Ono Lennon, MBE (born John Winston Lennon; 9 October 1940 – 8 December 1980) was an English singer and songwriter who rose to worldwide ...  
Death of John Lennon · Mark David Chapman · Yoko Ono · Cynthia Lennon

**Death of John Lennon - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/Death\\_of\\_John\\_Lennon](https://en.wikipedia.org/wiki/Death_of_John_Lennon) ▾ Wikipedia ▾  
John Lennon was an English musician who gained worldwide fame as one of ...  
Deaths: 1 Location: The Dakota, New York, New York  
Date: 10:50 pm, 8 December 1980 (US ...) Weapon: Charter Arms Undercover.38 ...

## John Lennon

Singer-songwriter · [johnlennon.com](http://johnlennon.com)

John Winston Ono Lennon, MBE was an English singer and songwriter who rose to worldwide fame as a co-founder of the band the Beatles, the most commercially successful band in the history of popular music.  
[Wikipedia](#)

**Born:** October 9, 1940, Liverpool, United Kingdom

**Assassinated:** December 8, 1980, New York City, NY

**Spouse:** Yoko Ono (m. 1969–1980), Cynthia Lennon (m. 1962–1968)

**Children:** Sean Lennon, Julian Lennon

## Songs

Jealous Guy	Imagine
Beautiful Boy	Double Fantasy
Starting Over	Double Fantasy
Stand By Me	Rock 'n' Roll
Watching the Wheels	Double Fantasy

## Quotes

[View 7+ more](#)

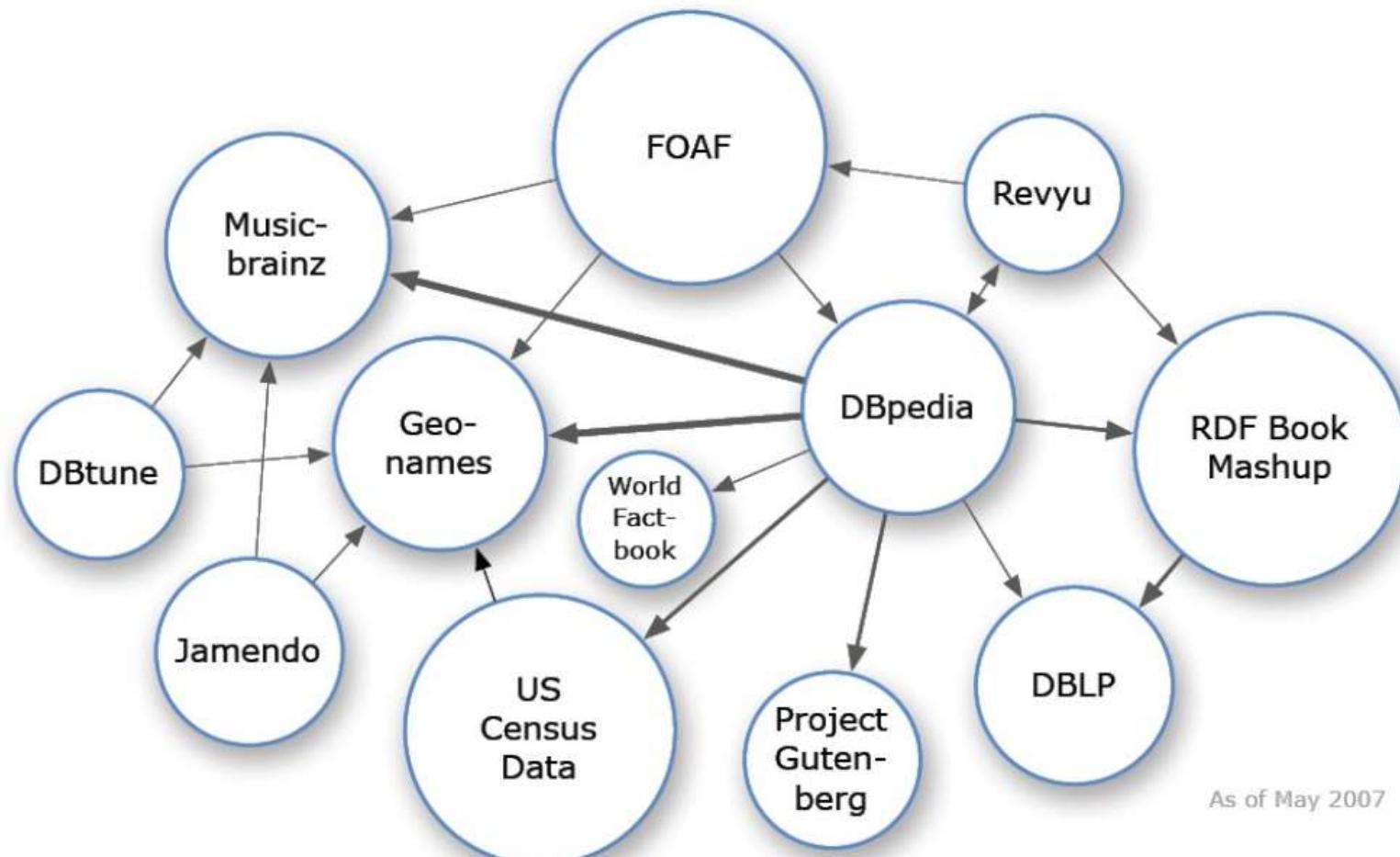
*If everyone demanded peace instead of another television set, then there'd be peace.*

*Reality leaves a lot to the imagination.*

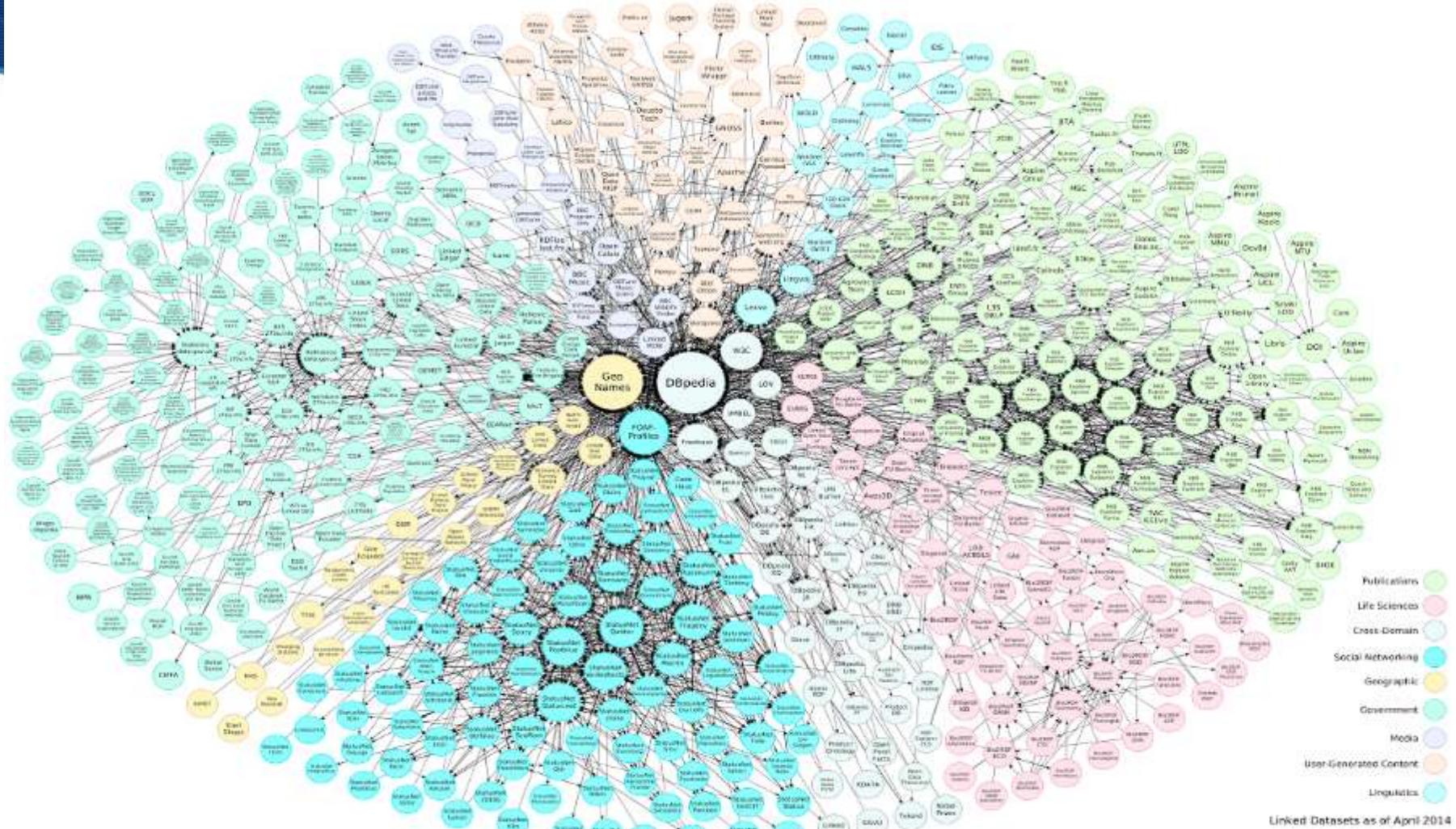
*Time you enjoy wasting, was not wasted.*

# Linked Open Data

# Connecting Data on the Web

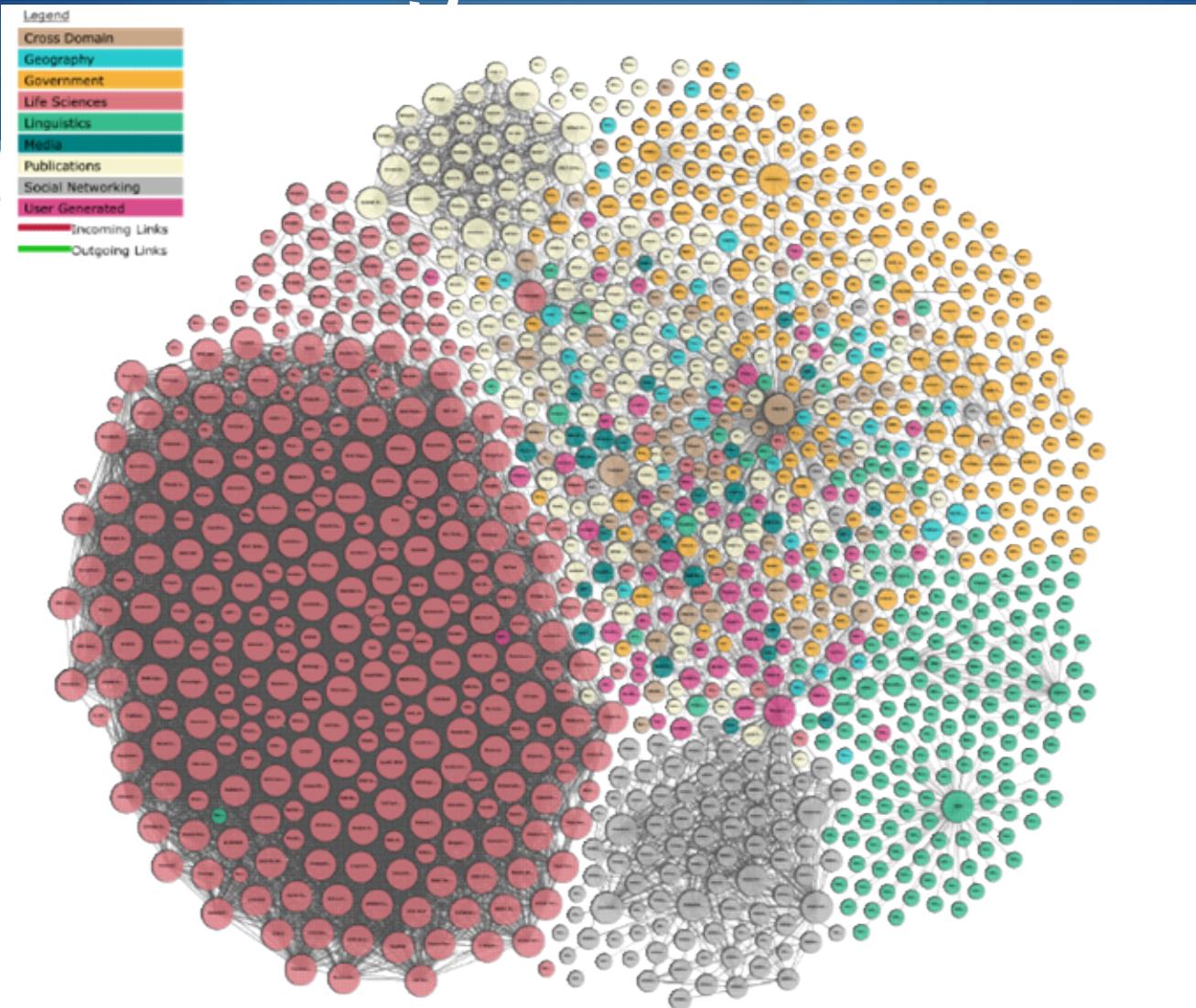


# Interlinked Data and Ontologies on the Web



"Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>"

# Interlinked Data and Ontologies on the Web



# Interlinked Data and Ontologies on the Web



BBC  
ONTOLOGIES



	2007	2011	2015
Datasets	294	571	3426
Triples	2B	31B	85B
Cross-refs	2M	500 M	

74% of datasets in a weakly connected component

FOAF: from 27% to 59%

DC: from 31% to 56%

<http://lod-cloud.net>  
<http://stats.lod2.eu>

# Linked Open Data Principles: URIs, HTTP, RDF

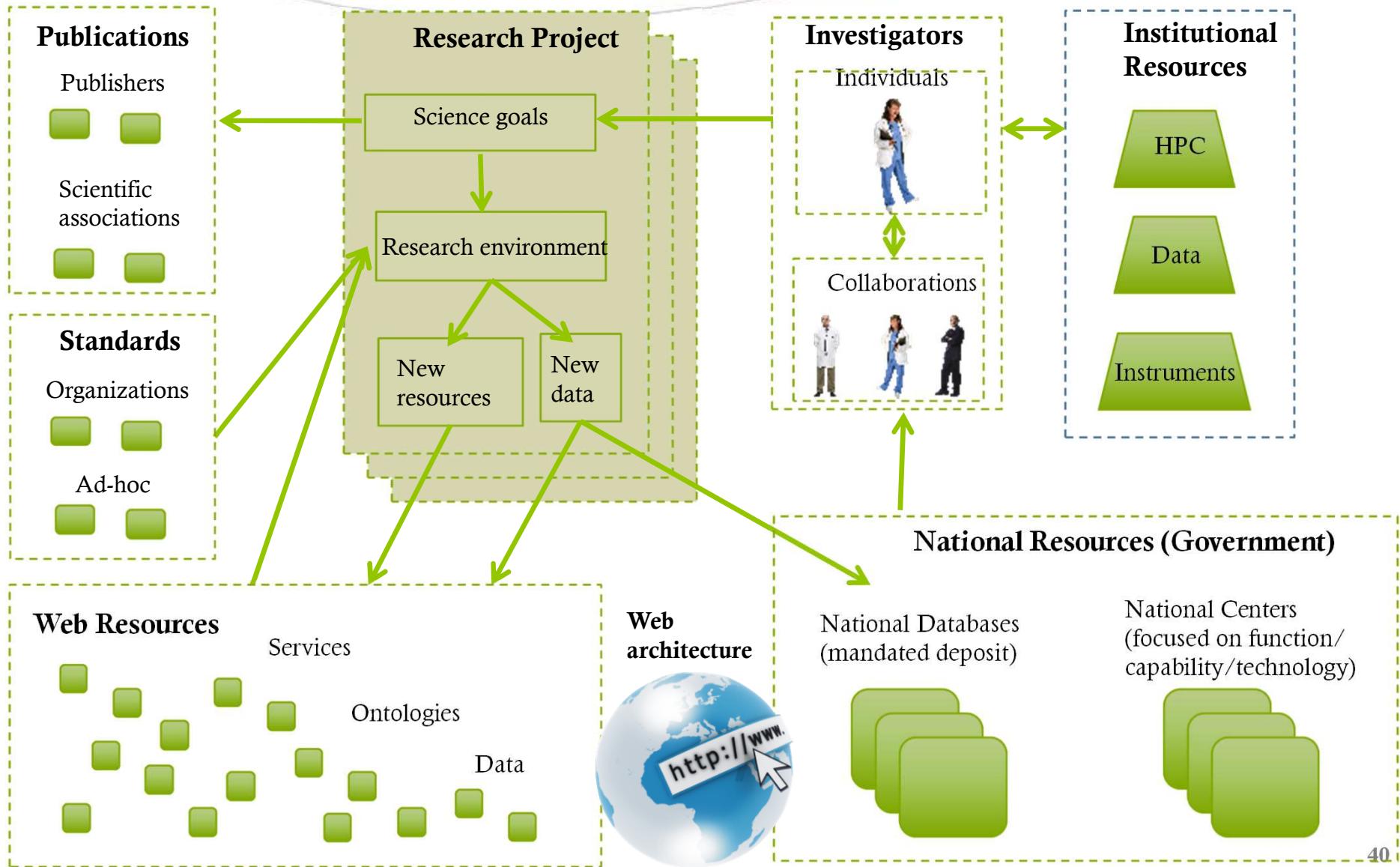
1. URIs are used to access any entity
2. HTTP can be used to access information about entities through their URIs
3. Information about entities is available in RDF, SPARQL, etc.
4. References to other entities are done through relationships that point to their URIs



# LinkedEarth: Linked Data



# Web Architecture



# Benefits – Take 2

1. Provide standards for data integration
  - ◆ eg the Gene Ontology
2. Data validation
  - ◆ Ensure that the data represent what was intended
3. Programmatic access to data
  - ◆ Enable structured queries
4. Linking and reusing other ontologies and data
  - ◆ Connecting across the Web
5. Publishing data in open web-accessible formats

Hands on:  
Time for Pizza

# Time for Pizza

- ◆ Divide in 2 groups
  - ◆ A group will aim to describe the pizzas in a restaurant menu
  - ◆ A group will aim to describe pizzas in a supermarket
- ◆ How do you define a Pizza Margherita?
- ◆ How do you define a 4 cheese pizza?
- ◆ How do you define a vegetarian pizza?
- ◆ Can I order a pizza with no tomato?
- ◆ Can I order a plain pizza?

# Time for Pizza

- ◆ **What are your Classes? Are there Dubclasses?**
- ◆ **What are your Properties? Domain and Range?**
- ◆ **What are your Data Properties?**
- ◆ **What are your axioms?**
  
- ◆ **Finished?**
- ◆ **Do a diagram of your ontology:**  
[http://visualdataweb.de/webowl\\_editor/](http://visualdataweb.de/webowl_editor/) or
- ◆ **Download Protégé:** <https://protege.stanford.edu/>

# Introductory Readings on Semantics and Ontologies

- ◆ “Ontology Development 101: A Guide to Creating Your First Ontology.” Natalya F. Noy and Deborah McGuinness.  
[http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf)
- ◆ “A short study on the success of the Gene Ontology.” Bada, M.; Stevens, R.; Goble, C. A.; Gil, Y.; Ashburner, M.; Blake, J. A.; Cherry, J. M.; Harris, M. A.; and Lewis, S. Journal of Web Semantics, 1(2): 235-240. 2004. DOI: 10.1016/j.websem.2003.12.003
- ◆ “Social Knowledge Collection.” Yolanda Gil. Handbook of Human Computation. Michelucci P. (Ed), MIT Press 2014.