

Intro to Semantic Web

Presented by:

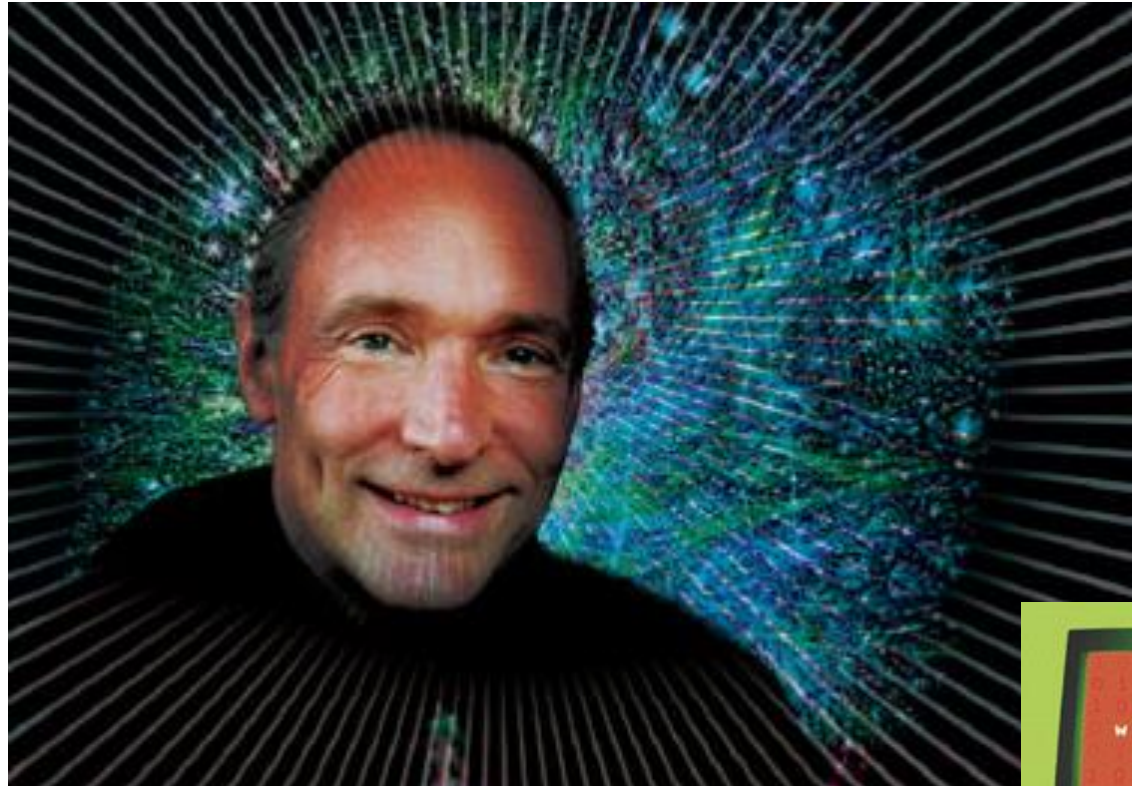
Aidan Hogan

Slides by EUCLID and Aidan

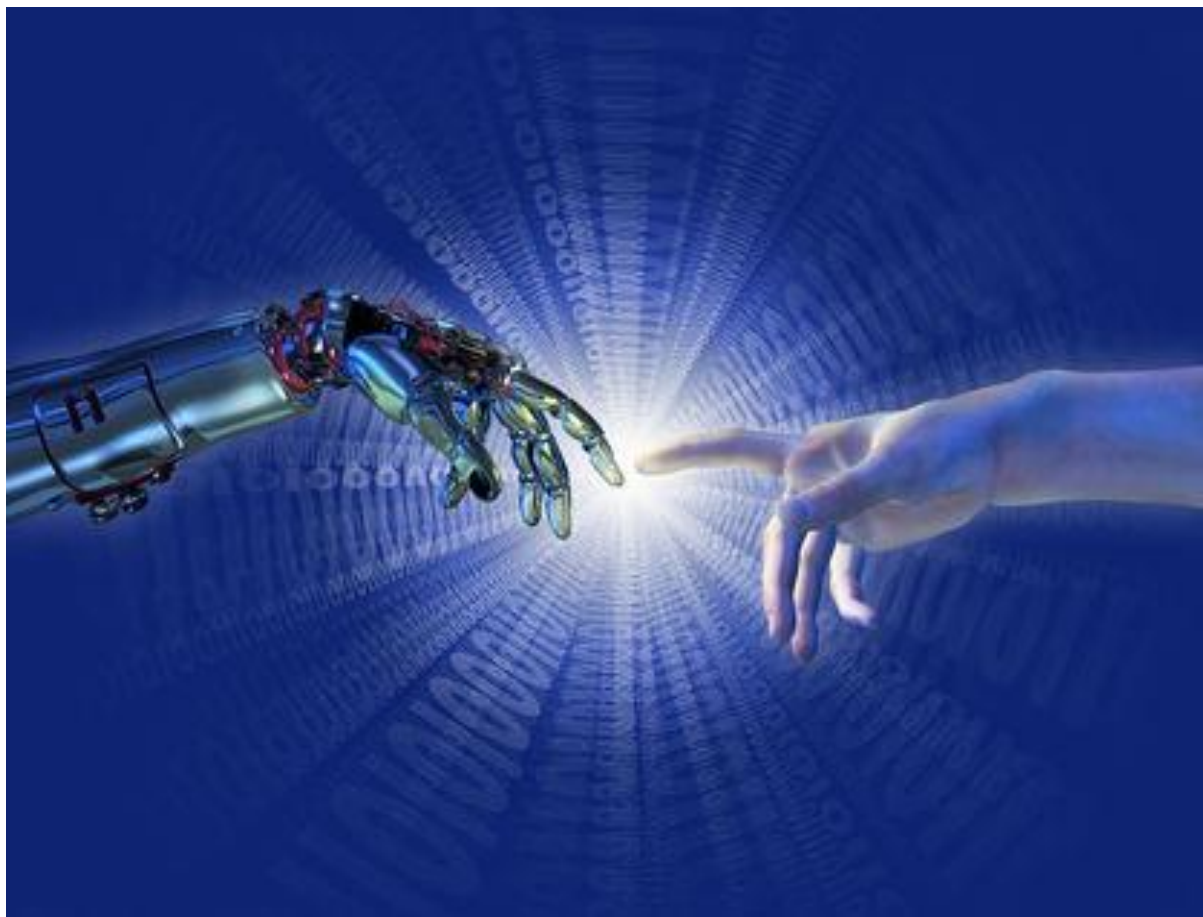


WHAT IS THE SEMANTIC WEB?

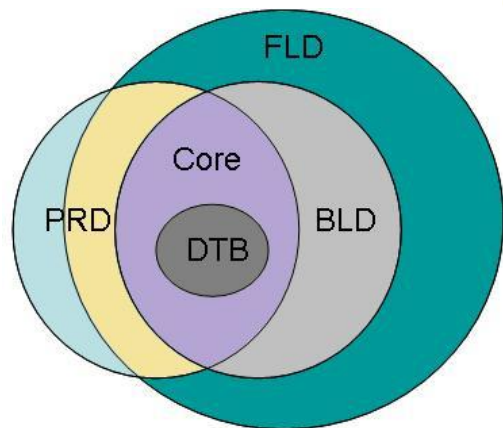
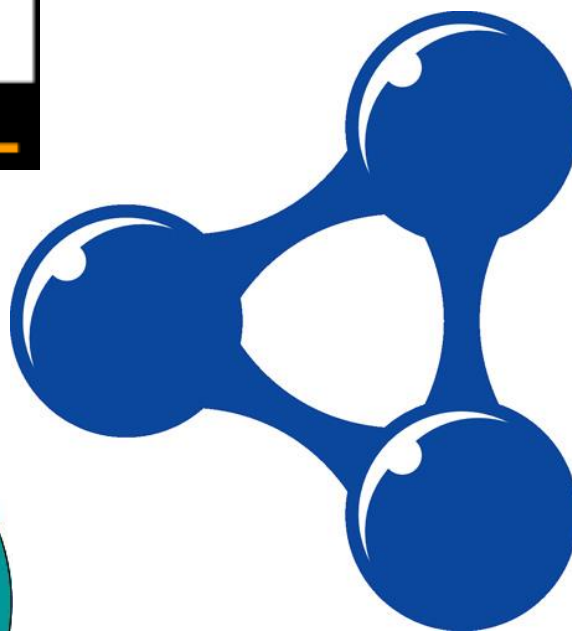
... machine readable Web?



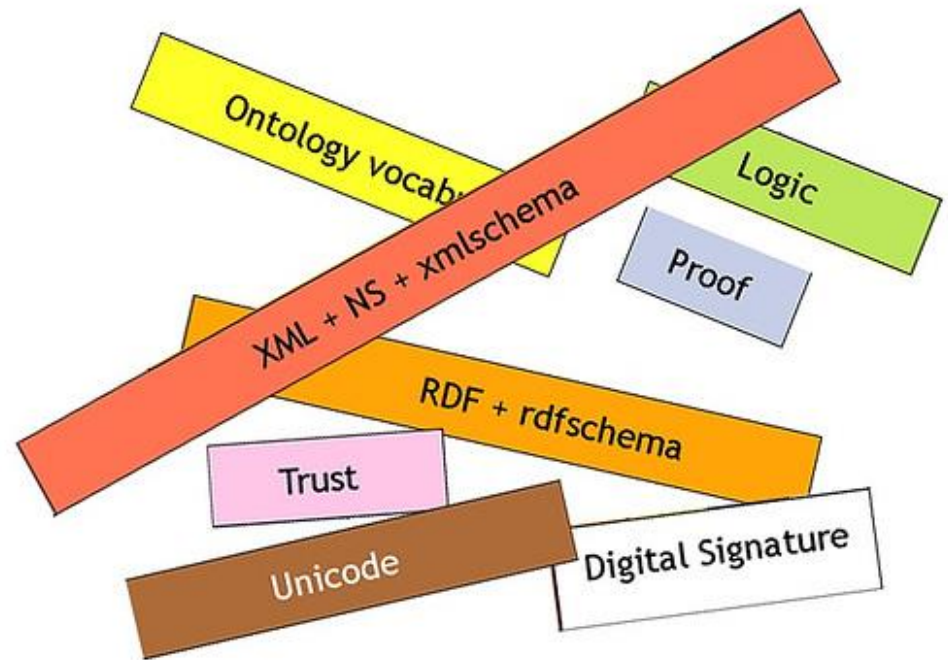
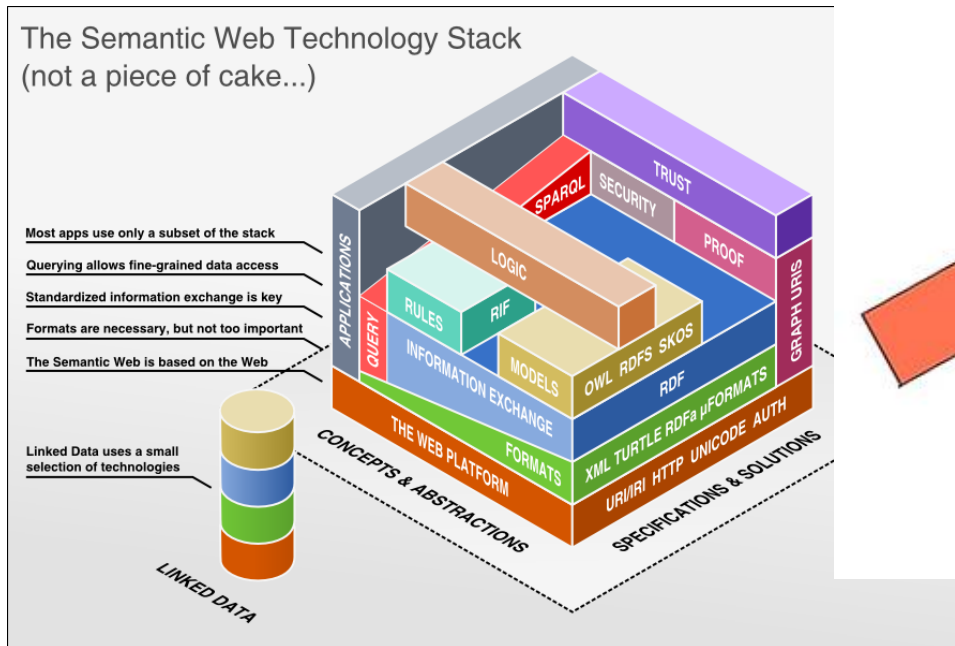
... benvolent machines?



... a bunch of standards?



... a cake?



Images by Hendler, Brickley Novack; <http://www.bnode.org/blog/tag/layer%20cake>

... complex ontologies?

2.2 Interpretations

Given a datatype map D and a vocabulary V over D , an *interpretation* $I = (\Delta_I, \Delta_D, \cdot^C, \cdot^{OP}, \cdot^{DP}, \cdot^I, \cdot^{DT}, \cdot^{LT}, \cdot^{FA})$ for D and V is a 9-tuple with the following structure:

- Δ_I is a nonempty set called the *object domain*.
- Δ_D is a nonempty set disjoint with Δ_I called the *data domain* such that $(DT)^{DT} \subseteq \Delta_D$ for each datatype $DT \in V_{DT}$.
- \cdot^C is the *class interpretation function* that assigns to each class $C \in V_C$ a subset $(C)^C \subseteq \Delta_I$ such that
 - $(owl:Thing)^C = \Delta_I$ and
 - $(owl:Nothing)^C = \emptyset$.
- \cdot^{OP} is the *object property interpretation function* that assigns to each object property $OP \in V_{OP}$ a subset $(OP)^{OP} \subseteq \Delta_I \times \Delta_I$ such that
 - $(owl:topObjectProperty)^{OP} = \Delta_I \times \Delta_I$ and
 - $(owl:bottomObjectProperty)^{OP} = \emptyset$.
- \cdot^{DP} is the *data property interpretation function* that assigns to each data property $DP \in V_{DP}$ a subset $(DP)^{DP} \subseteq \Delta_I \times \Delta_D$ such that

$(owl:topDataProperty)^{DP} = \Delta_I \times \Delta_D$ and

$(owl:bottomDataProperty)^{DP} = \emptyset$.

\cdot^I is the *individual interpretation function* that assigns to each individual $a \in V_I$ an element $(a)^I \in \Delta_I$.

\cdot^{DT} is the *datatype interpretation function* that assigns to each datatype $DT \in V_{DT}$ a subset $(DT)^{DT} \subseteq \Delta_D$ such that

$(DT)^{DT}$ is as in D for each datatype $DT \in N_{DT}$, and

$(DT)^{DT} = \Delta_D$.

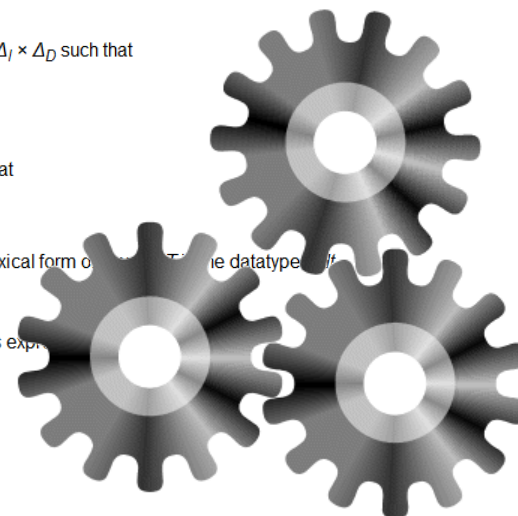
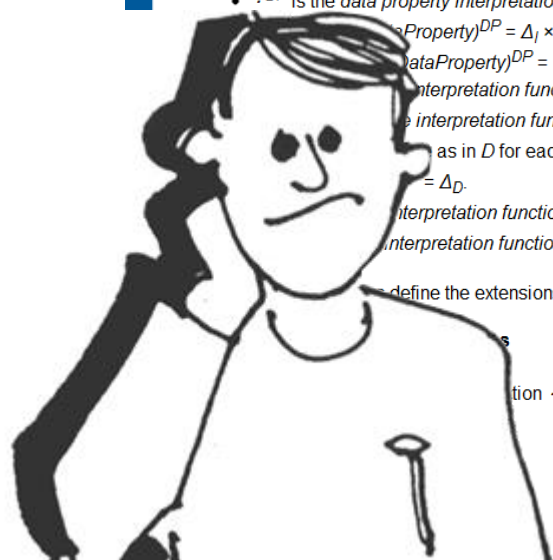
\cdot^{LT} is the *lexical form interpretation function* that is defined as $(It)^{LT} = (LV, DT)^{LS}$ for each $It \in V_{LT}$, where LV is the lexical form of It and DT is the datatype of It .

\cdot^{FA} is the *functional array interpretation function* that is defined as $(F, It)^{FA} = (F, (It)^{LT})^{FS}$ for each $(F, It) \in V_{FA}$.

\cdot^C , \cdot^{OP} , \cdot^{DP} , and \cdot^I are extended to object property expressions, data ranges, and class expressions as follows:

\cdot^C is extended to class expressions as shown in Table 1.

\cdot^{OP} is extended to object property expressions as shown in Table 1.



... free/open data?

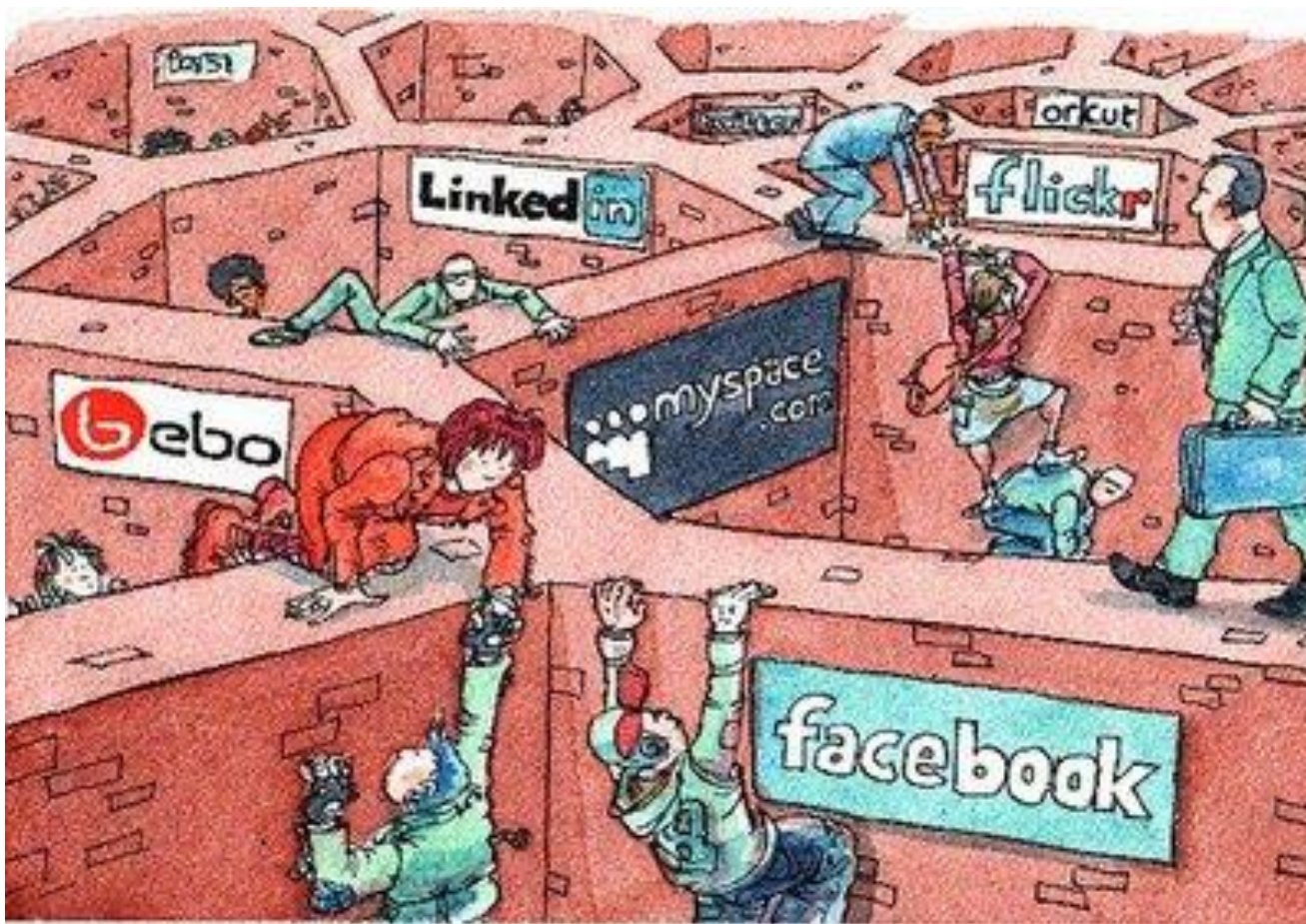


Image from <http://www.digitaltimes.ie>

... naming everything?



"Now! ... *That* should clear up a few things around here!"

... Web 3.0 ?

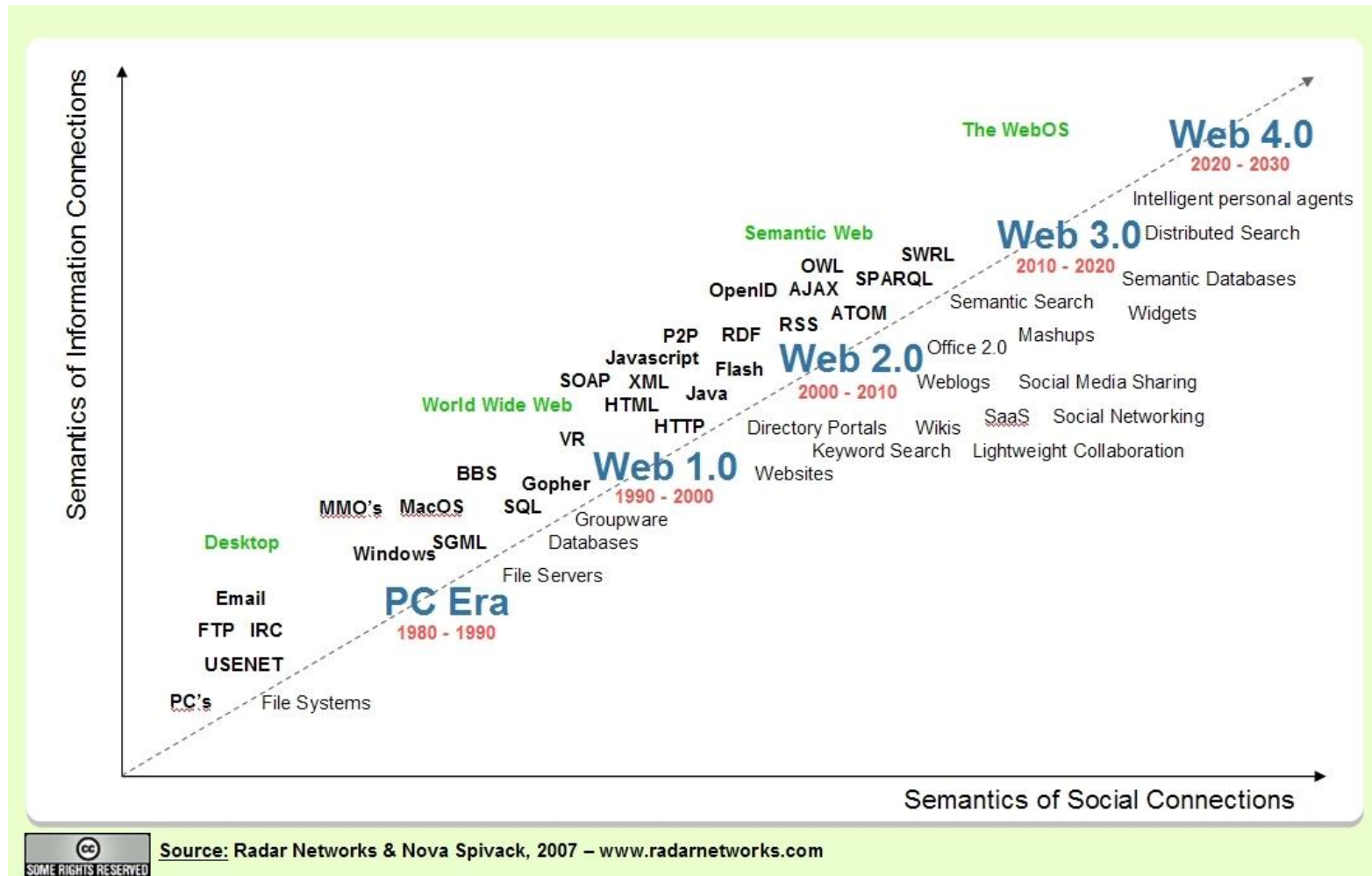



Image by Radar Networks; Nova Spivak; <http://memebox.com/futureblogger/show/824>

... a buzzword?!



 [Log in / create account](#)

[Article](#) [Talk](#) [Read](#) [Edit](#) [View history](#)



Semantic Web

From Wikipedia, the free encyclopedia



Categories: [Buzzwords](#) | [Semantic Web](#) | [Web services](#)

... the research area
your supervisor
assigned you to but
that you don't really
understand?



WHY IS THE SEMANTIC WEB?



Great Wave of Data



WIKIPEDIA
The Free Encyclopedia

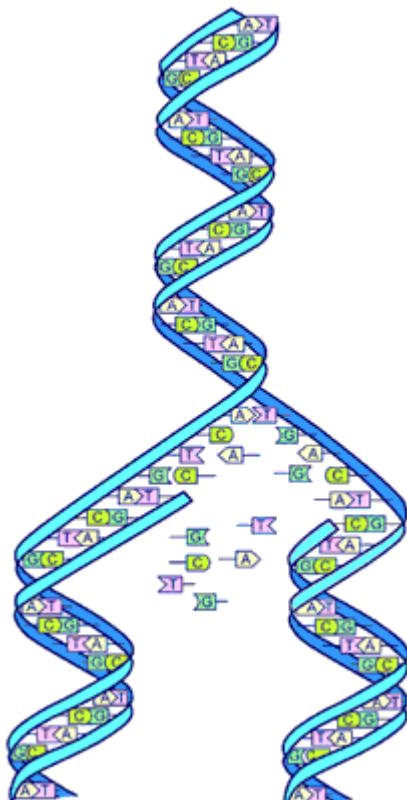
Wikipedia

≈ 5.9 TB of data

(Jan. 2010 Dump)

1 Wiki = 1 Wikipedia

Great Wave of Data



Human Genome

≈ 4 GB/person

≈ 0.0006 Wiki/person

Great Wave of Data

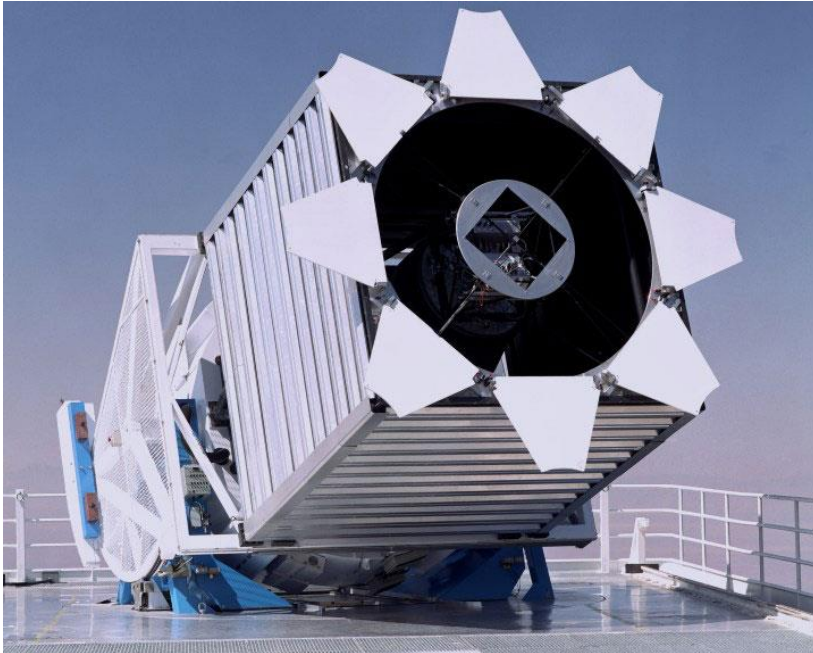


US Library of Congress

≈ 235 TB archived

≈ 40 Wiki

Great Wave of Data



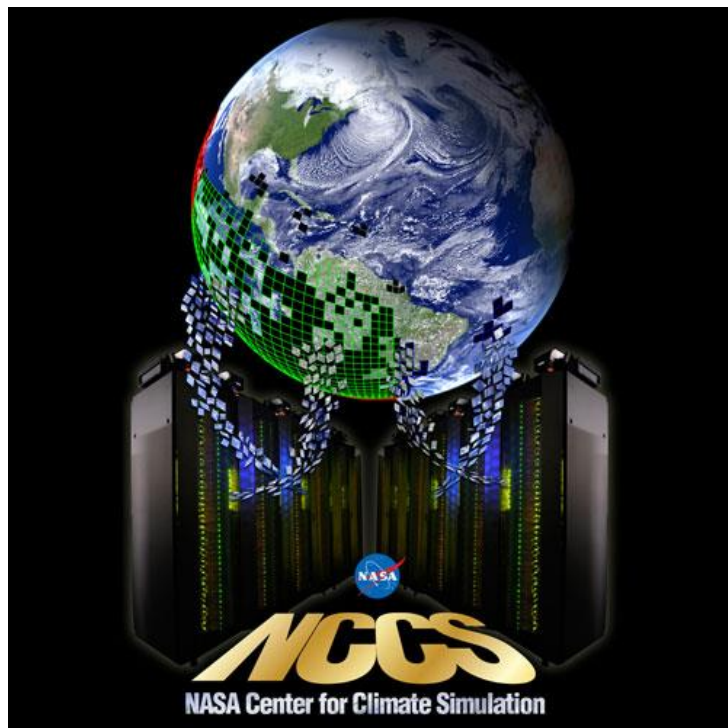
Sloan Digital Sky Survey

≈ 200 GB/day

≈ 73 TB/year

≈ 12 Wiki/year

Great Wave of Data



NASA Center for Climate Simulation

≈ 32 PB archived

≈ 5,614 Wiki

Great Wave of Data



Facebook

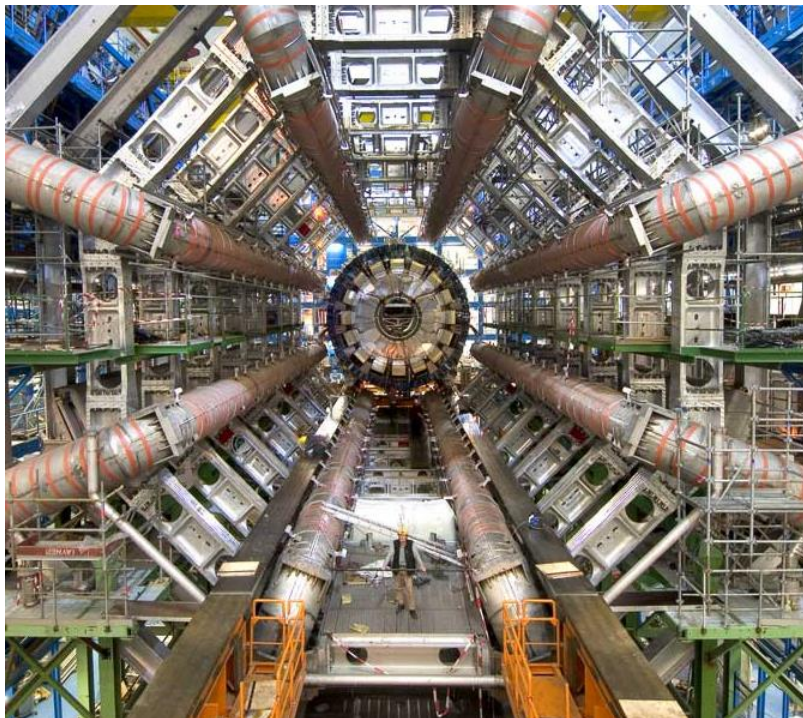
≈ 12 TB/day added

≈ 2 Wiki/day

≈ 782 Wiki/year

(as of Mar. 2010)

Great Wave of Data



Large Hadron Collider

≈ 15 PB/year

≈ 2,542 Wikipedias/year

Great Wave of Data



Google

≈ 20 PB/day processed

≈ 3,389 Wiki/day

≈ 7,300,000 Wiki/year

(Jan. 2010)

Great Wave of Data



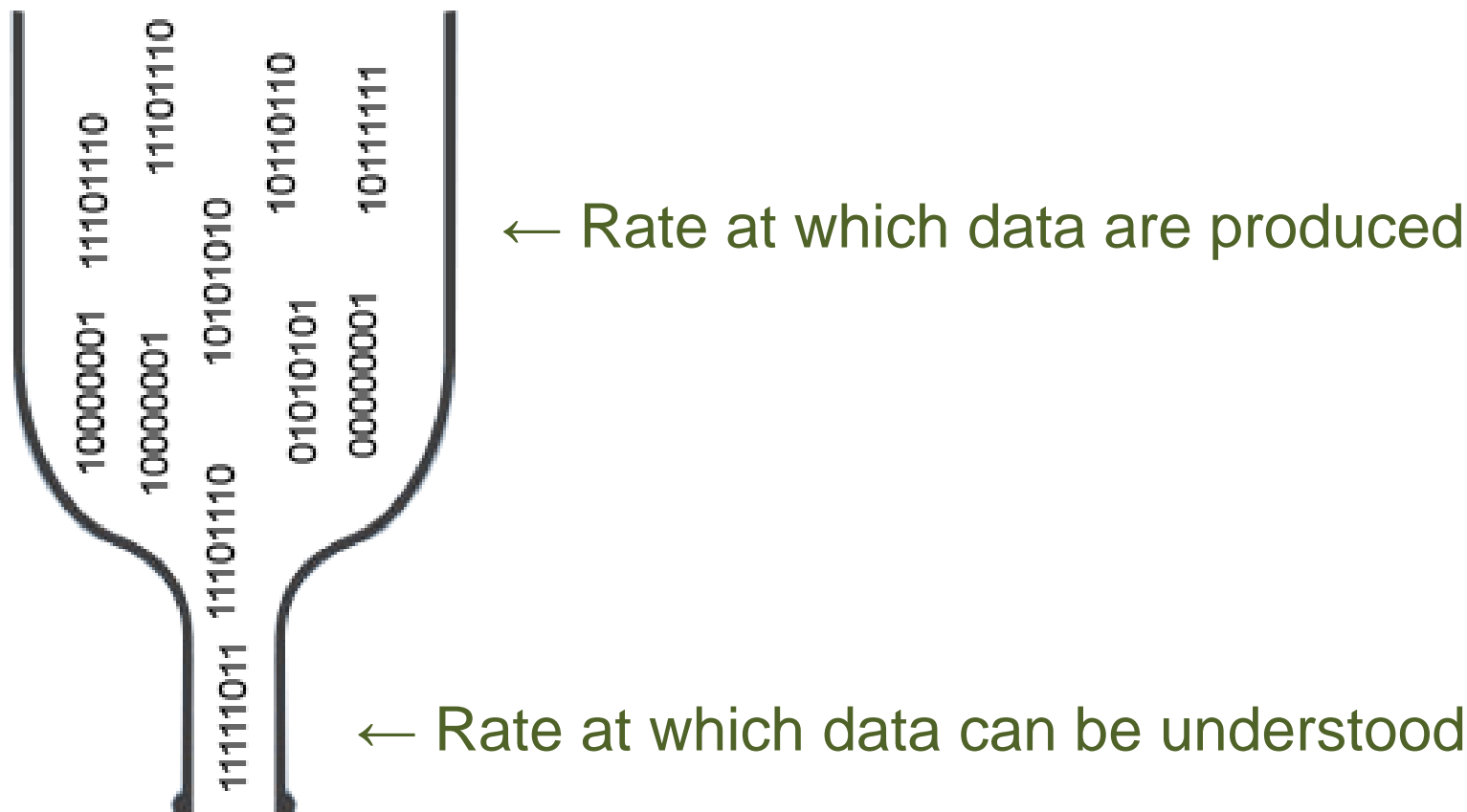
Internet (2016)

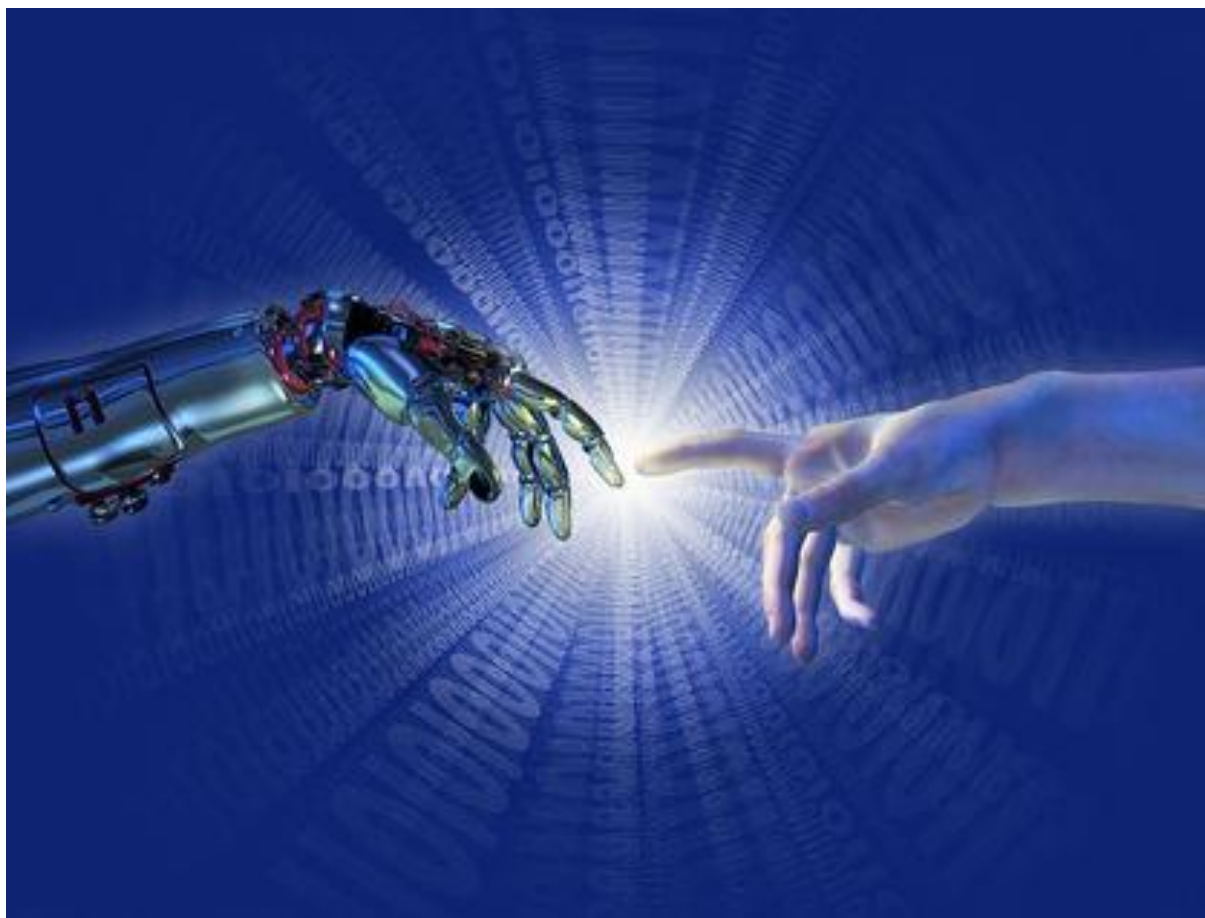
≈ 1.3 ZB/year

≈ 220,338,983 Wiki/year

(2016 IP traffic; Cisco est.)

Data Bottleneck





NUI Galway
OÉ Gaillimh



DERI

WHAT PROBLEM DOES IT SOLVE?

... a personal answer



OÉ Gaillimh
NUI Galway



DERI

ca. 2005



... the research area
your supervisor
assigned you to but
that you don't really
understand?



The Dessert Problem

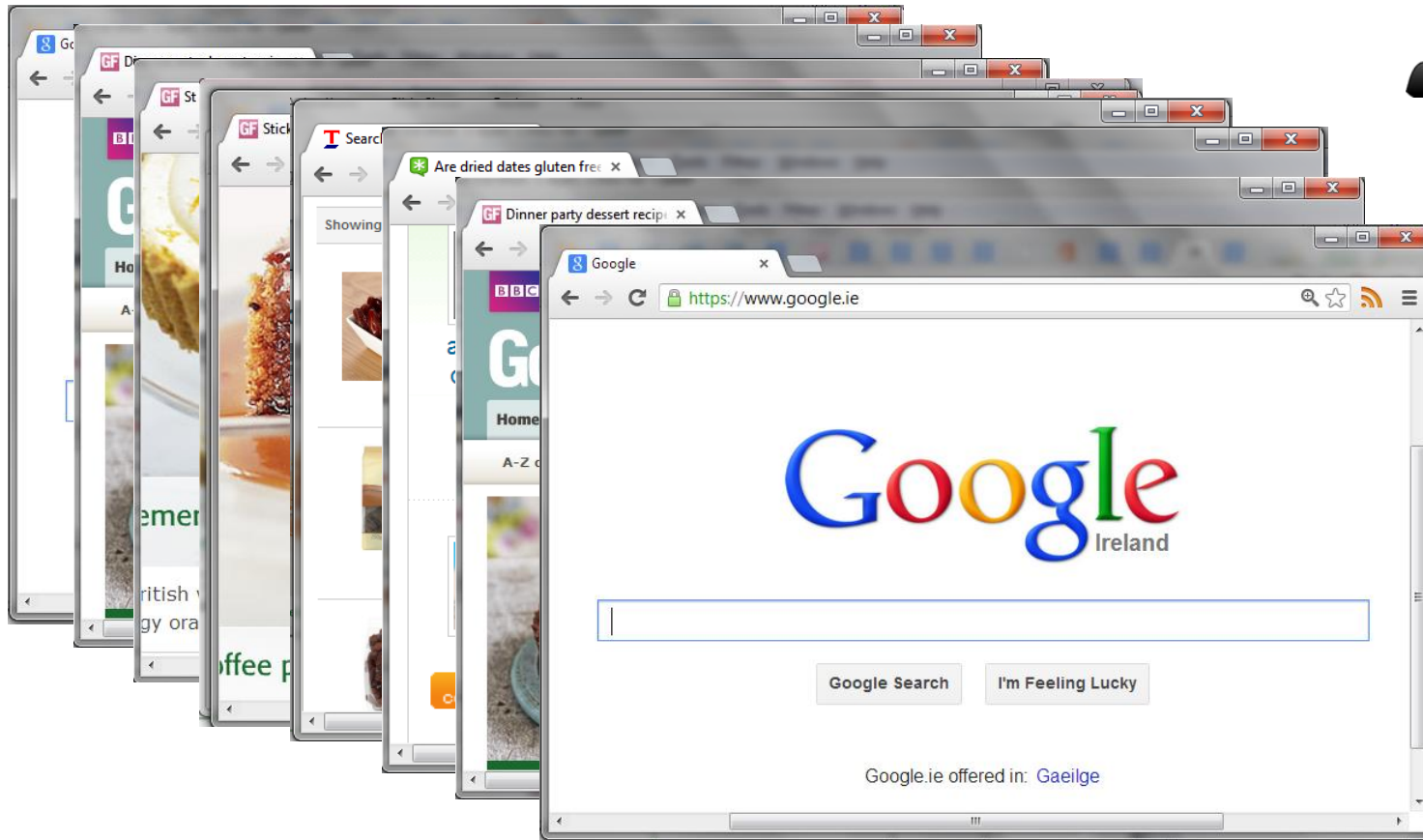


The Dessert Problem



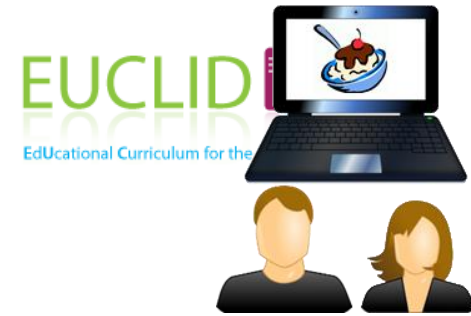
- Requirements:
 - Must be a dessert
 - Must be citrus-free
 - Must be gluten-free
 - Ingredients available in local supermarket(s)
 - Cheap (students)
 - Must be delicious

The Dessert Problem



Dessert Algorithm: Naive

```
candidates := ∅
for all recipe-site in google-results
  for all dessert-recipe in recipe-site
    if dessert-recipe type looks-delicious
      suitable := true
      for all ingredient in dessert-recipe
        if searchNutrition(ingredient) type wheat or lemon or lime ...
          suitable := false
        else if searchShops(ingredient) type null or expensive
          suitable := false
        end
      end
      if suitable add dessert-recipe to candidates end
    end
  end
end
return candidates
```



Dessert Algorithm: Utopia



```
candidates := ∅
```

```
for all safe-cheap-local-dessert-recipe in magical-sem-web-results()
```

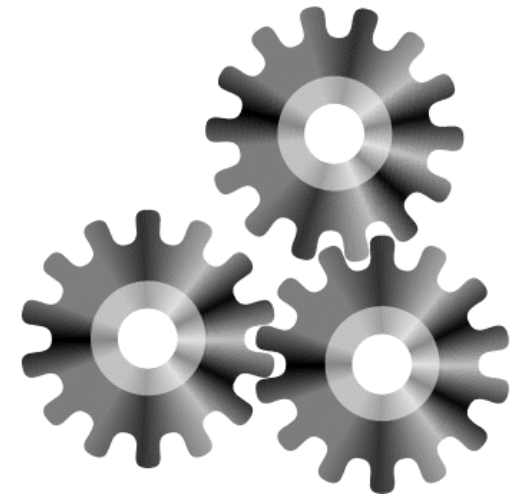
```
  if safe-cheap-local-dessert-recipe type looks-delicious
```

```
    add safe-cheap-local-dessert-recipe to candidates end
```

```
  end
```

```
end
```

```
return candidates
```



The Dessert Solution?



TESCO

Groceries Favourites Special Offers **Happy Easter**

Fresh Food | Bakery | Food Cupboard | Frozen Food | Baby | Health & Beauty

Groceries > Eat Me Glovebox Dates 227G

LIFE 21+

Eat Me Glovebox Dates 227G

€2.49 (€10.97/kg)

Quantity: 1

Product Details Ingredients Nutrition Storage Instructions

A serving contains the following of your guideline daily amount

Calories	Sugar	Fat	Saturates	Salt
300	61g	1g	Trace	Trace
15%	67%	1%	<1%	<1%

Ready to eat dried Dates

Other Information

Warnings: Caution : contains stone

Print this page

Sticky toffee pudding recipe

www.bbcgoodfood.com/recipes/4257/sticky-toffee-pudding

get our newsletter.

Ingredients

- 200g dried dates, stoned and chopped (buy Medjool if you can)
- 250ml black tea (not too strong)
- 1/2 tsp bicarbonate of soda
- 85g unsalted butter, softened
- 175g self-raising flour
- 1 tsp mixed spice

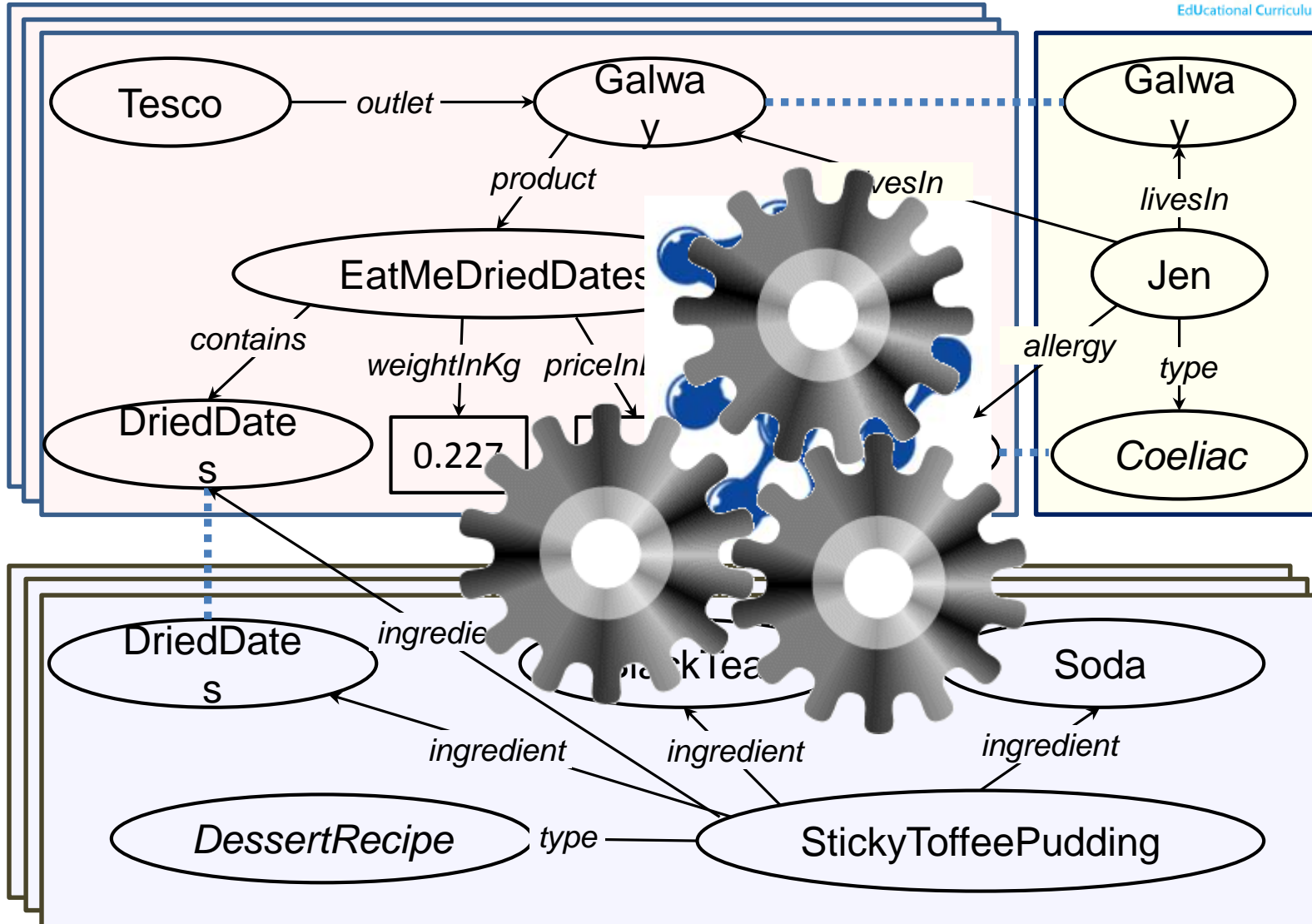
[magical-sem-web-results\(\)](#)



NUI Galway
OÉ Gaillimh



The Dessert Solution?



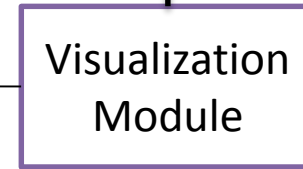
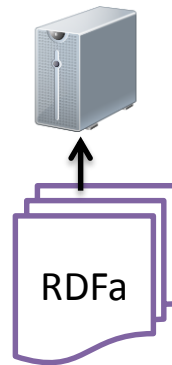
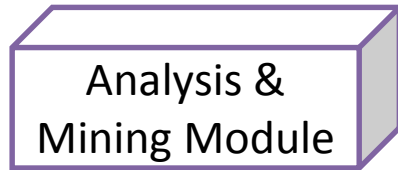
BEYOND DESSERT ...

Music!

- Provision of a music-based portal.
- Bring together a number of disparate components of data-oriented content:
 1. **Musical content** (streaming data & downloads)
 2. **Music and artist metadata**
 3. **Review content**
 4. **Visual content** (pictures of artists & albums)

Music!

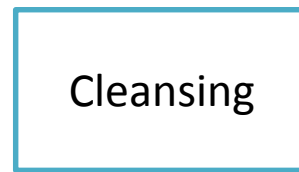
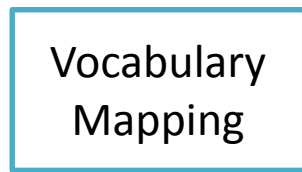
Application



Access

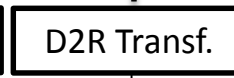
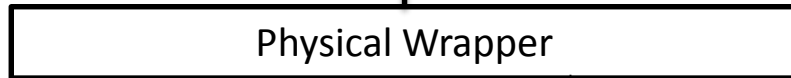


LD Dataset



Publishing

Data acquisition

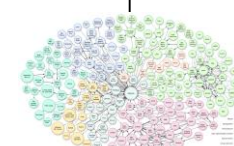


Streaming providers

Downloads



Metadata

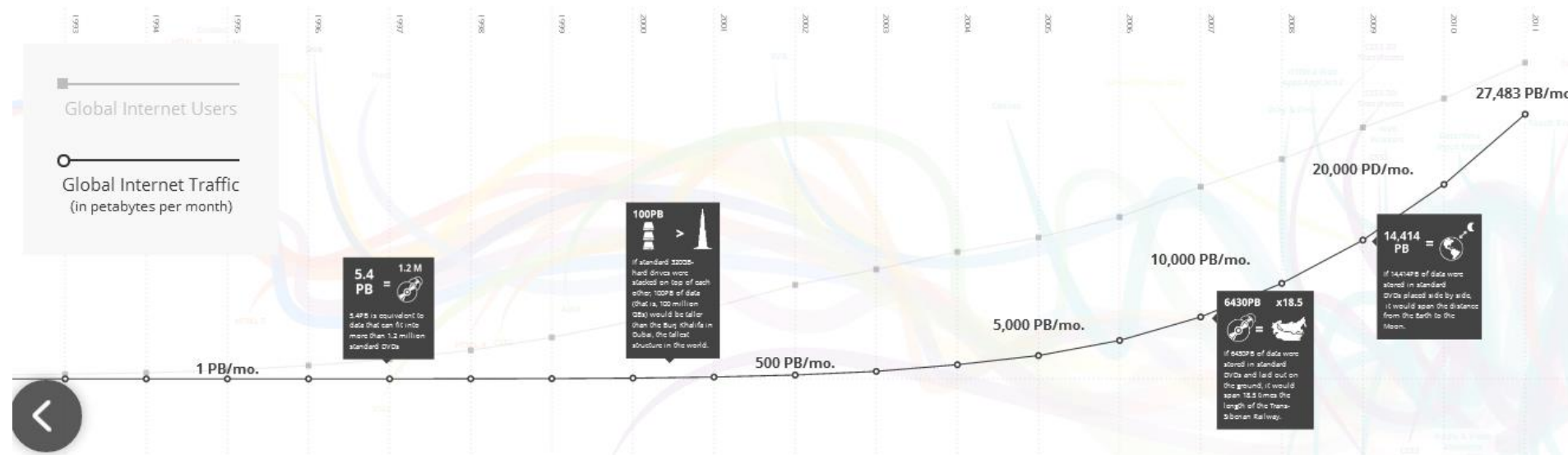
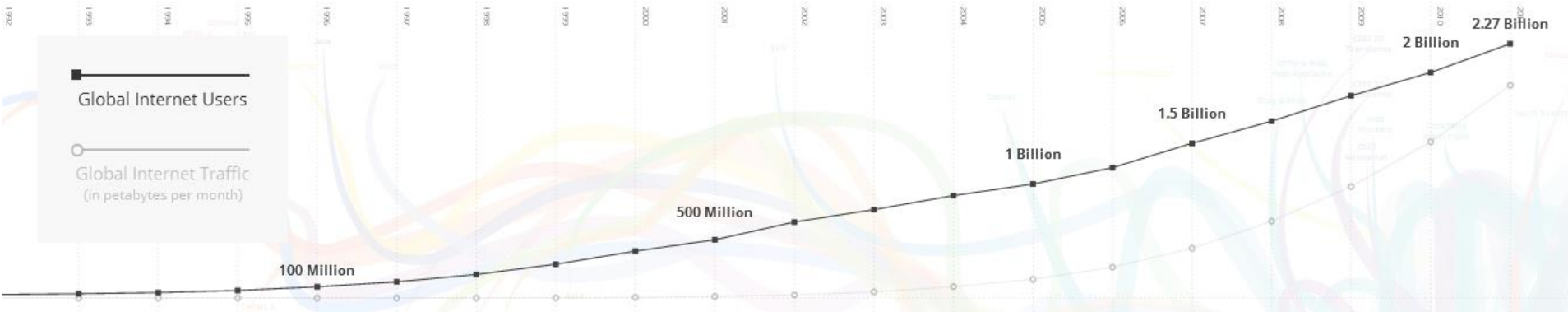


Other content

SEMANTIC WEB FOUNDATIONS

Internet

The growth of the Internet



Source: <http://www.evolutionoftheweb.com>

The Web

- There is a wealth of information on the Web.
- It is aimed mostly towards consumption by **humans as end-users**:
 - Recognize the meaning behind content and draw conclusions,
 - Infer new knowledge using context and
 - Understand background information.



The image shows a screenshot of the Wikipedia article for Germany. At the top left is the Wikipedia logo, a globe with letters. Below it is the word "WIKIPEDIA". To the right is a search bar with "Germany" entered. Below the search bar is the title "Germany" and the text "From Wikipedia, the free encyclopedia". The main text of the article is visible, starting with "Germany (4 /ˈdʒɜːməni/), officially the **Federal Republic of Germany** (German: *Bundesrepublik Deutschland*, pronounced [ˈbʊndəsʁepuˌbliːk ˈdɔʏtʃlant]) is a federal parliamentary republic in west-central Europe. The country consists of 16 states, and its capital and largest city is Berlin. Germany covers an area of 357,021 square kilometres (137,847 sq mi) and has a largely temperate seasonal climate. With 81.8 million inhabitants, it is the most populous member state in the European Union. It is one of the major political powers of the European continent and a technological leader in many fields." To the right of the main text is a sidebar with the title "Federal Republic of Germany" and "Bundesrepublik Deutschland". It contains the German flag and the German coat of arms. Below the flag and coat of arms is the text "Anthem: The third stanza of Lied der Deutschen (Song of the Germans)".



[Germany – Wikipedia](#)
[de.wikipedia.org/wiki/Germany](#)
Germany ist: die englische Bezeichnung für Deutschland. **Germany** (Mondkrater), ein Mondkrater; **Germany** (Rapper), deutscher Rapper. Zudem steht der Name ...

[Germany - Wikipedia, the free encyclopedia](#)
[en.wikipedia.org/wiki/Germany](#) - Diese Seite übersetzen
Song of the Germans. Location of **Germany** (dark green)– in Europe (green & dark grey). Location of **Germany** (dark green). – in Europe (green & dark grey) ...
[Flag of Germany - History of Germany - Geography of Germany - German cuisine](#)

[Tourism in Germany – travel, breaks, holidays](#)
[www.germany.travel/](#) - Diese Seite übersetzen
Tourism in **Germany** – travel, breaks, holidays. ... BMWi Logo **Germany**, the travel destination. Towns, cities & culture Towns, cities & culture; Leisure and ...

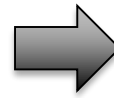
[Startseite: Das Deutschland-Portal](#)
[www.deutschland.de/](#)
deutschland.de ist das offizielle und unabhängige Portal der Bundesrepublik Deutschland im Internet. Es bietet in fünf Sprachen eine Sammlung wichtiger ...

The Web

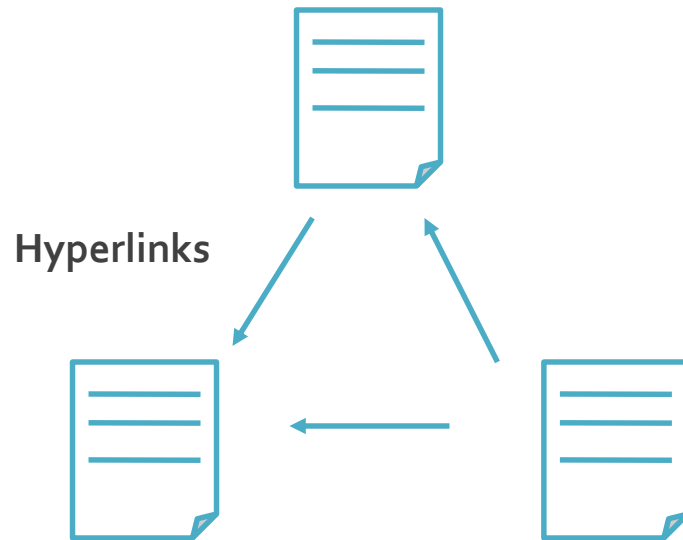
- Billions of diverse documents online, but it is not easily possible to automatically:
 - Retrieve relevant documents.
 - Extract information.
 - Combine information in a meaningful way.
- Idea:
 - Also publish machine processible data on the web.
 - Formulate questions in terms understandable for a machine.
 - Do this in a standardized way so machines can interoperate.
- The Web becomes a **Web of Data**
 - This provides a common framework to share knowledge on the Web across application boundaries.

The Web: Evolution

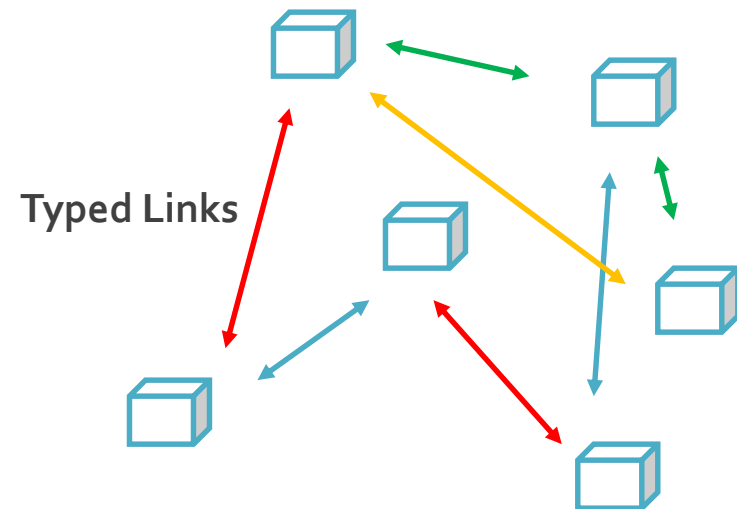
Web of Documents



Web of Data



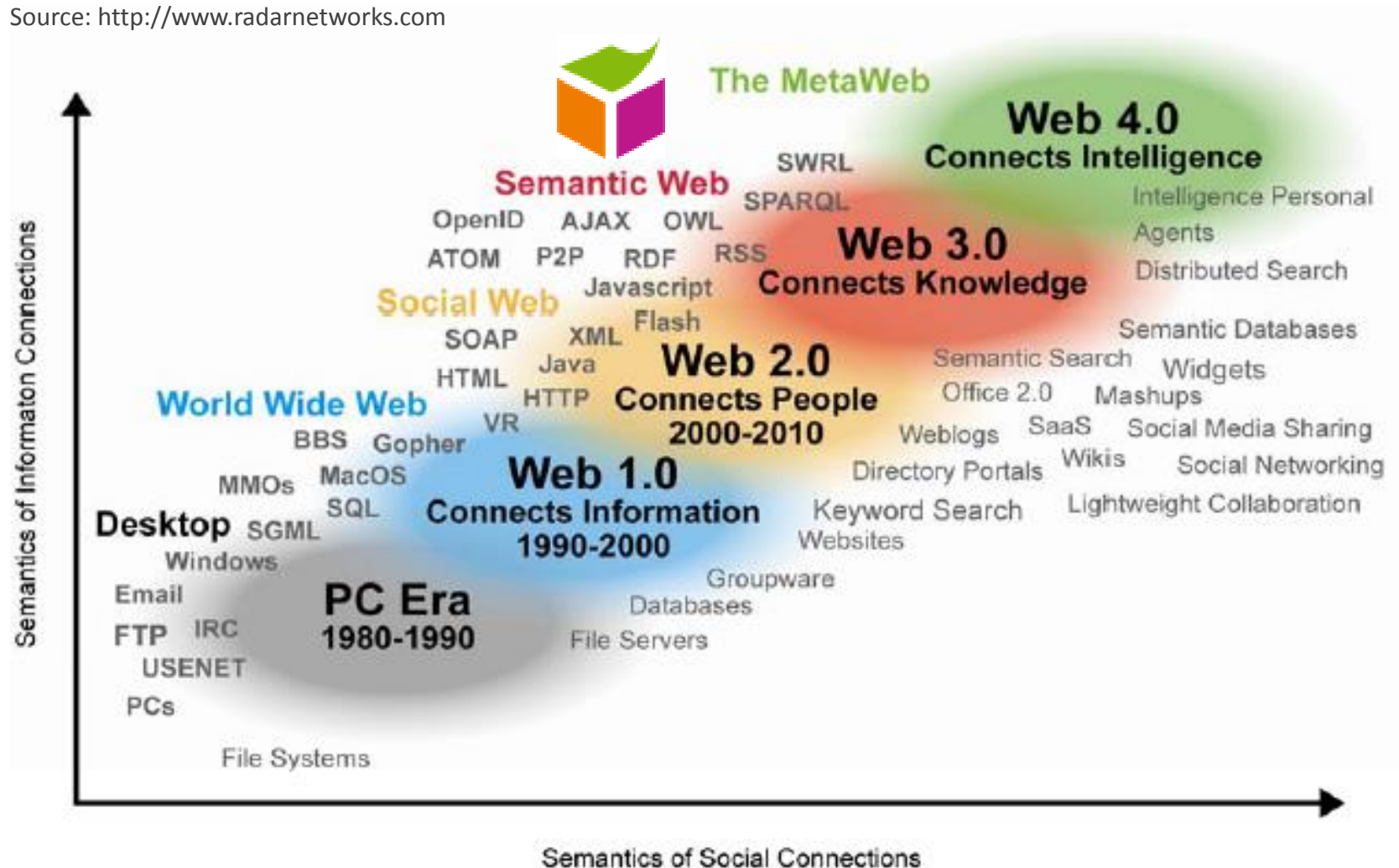
"Documents"



"Things"

The Web: Evolution

Source: <http://www.radarnetworks.com>



HTML – HyperText Markup Language

- Language for displaying web pages and other information in a web browser.
- HTML elements consist of **tags** (enclosed in angle brackets), **attributes** and **content**.

HTTP – HyperText Transfer Protocol

- Foundation of data communication for the WWW.
- Client-server protocol.
- Every interaction is based on: **request** and **response**.

Uniform Resource Identifier (URI)

- Compact sequence of characters that identifies an abstract or physical resource.

- **Examples:**

`ldap://[2001:db8::7]/c=GB?objectClass?one`

`mailto:John.Doe@example.com`

`news:comp.infosystems.www.servers.unix`

`tel:+1-816-555-1212`

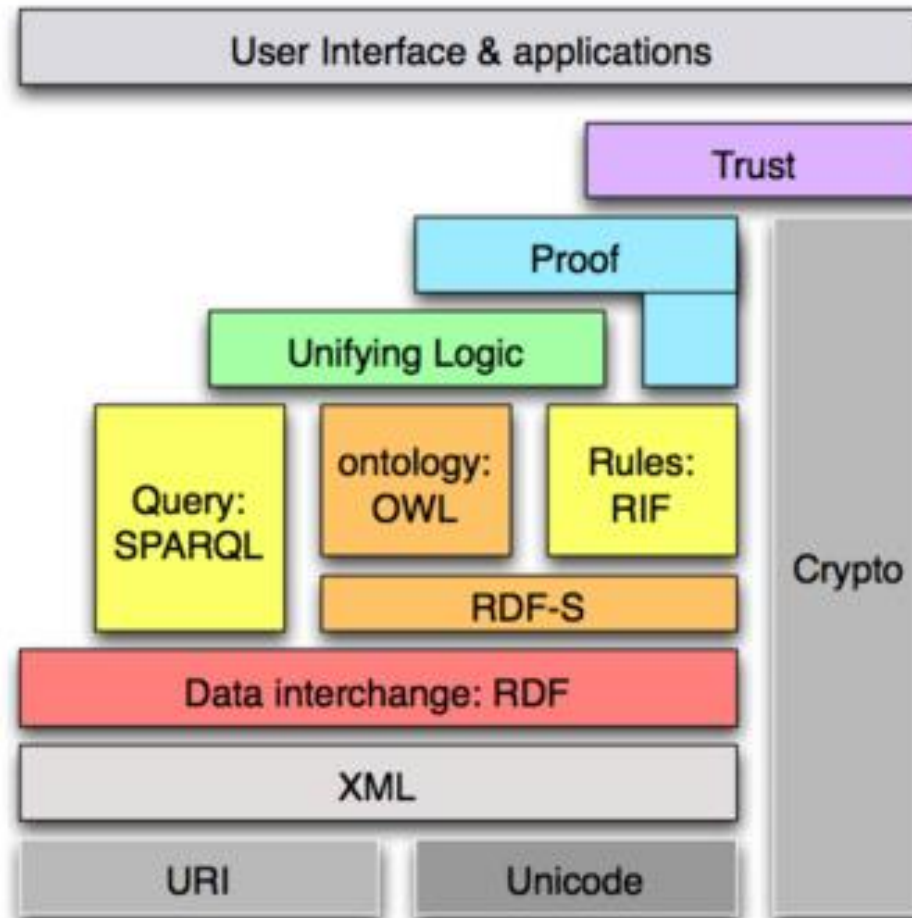
`telnet://192.0.2.16:80/`

`urn:oasis:names:specification:docbook:dtd:xml:4.1.2`

`http://dbpedia.org/resource/Karlsruhe`

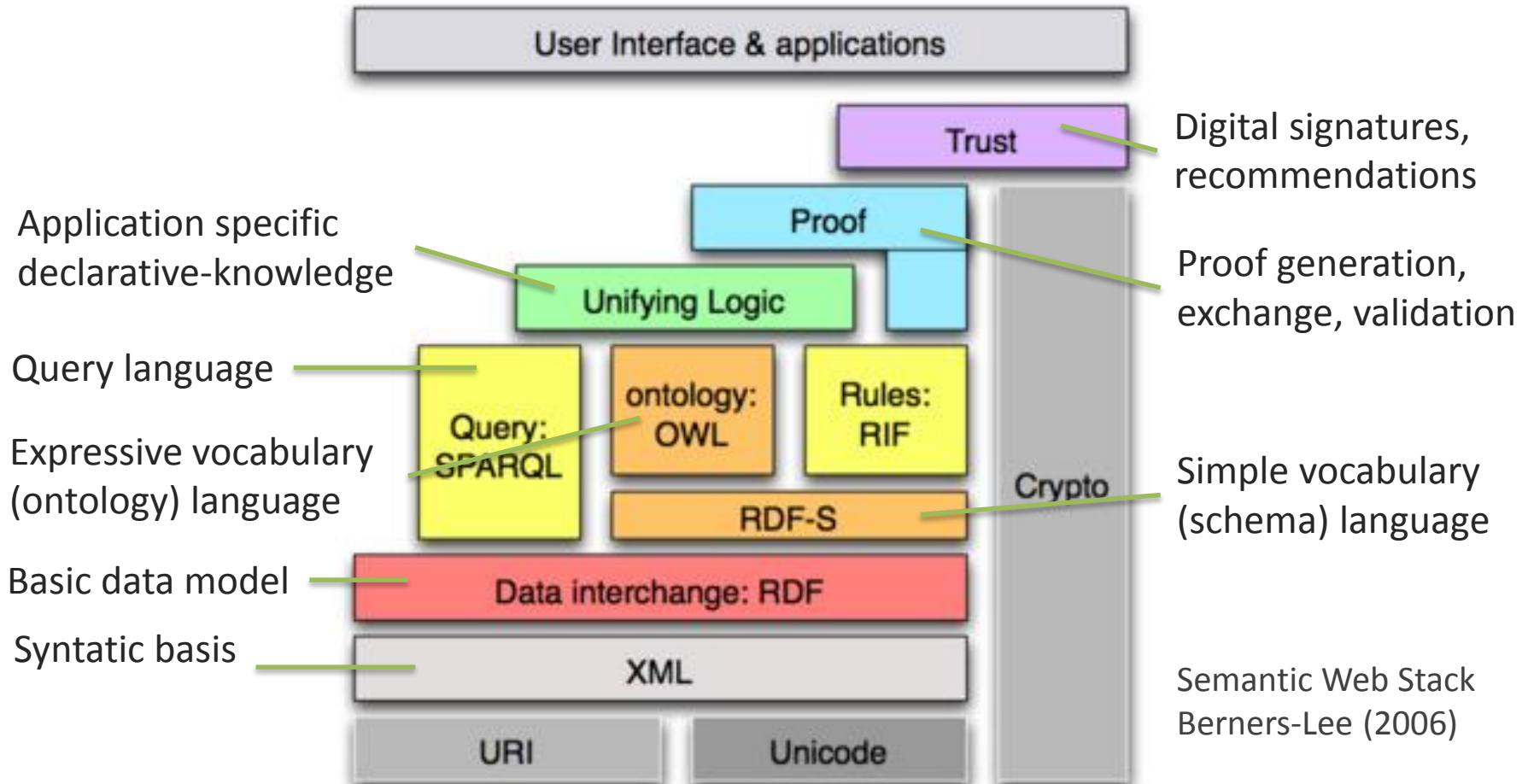
CORE OF THE SEMANTIC WEB

Semantics on the Web

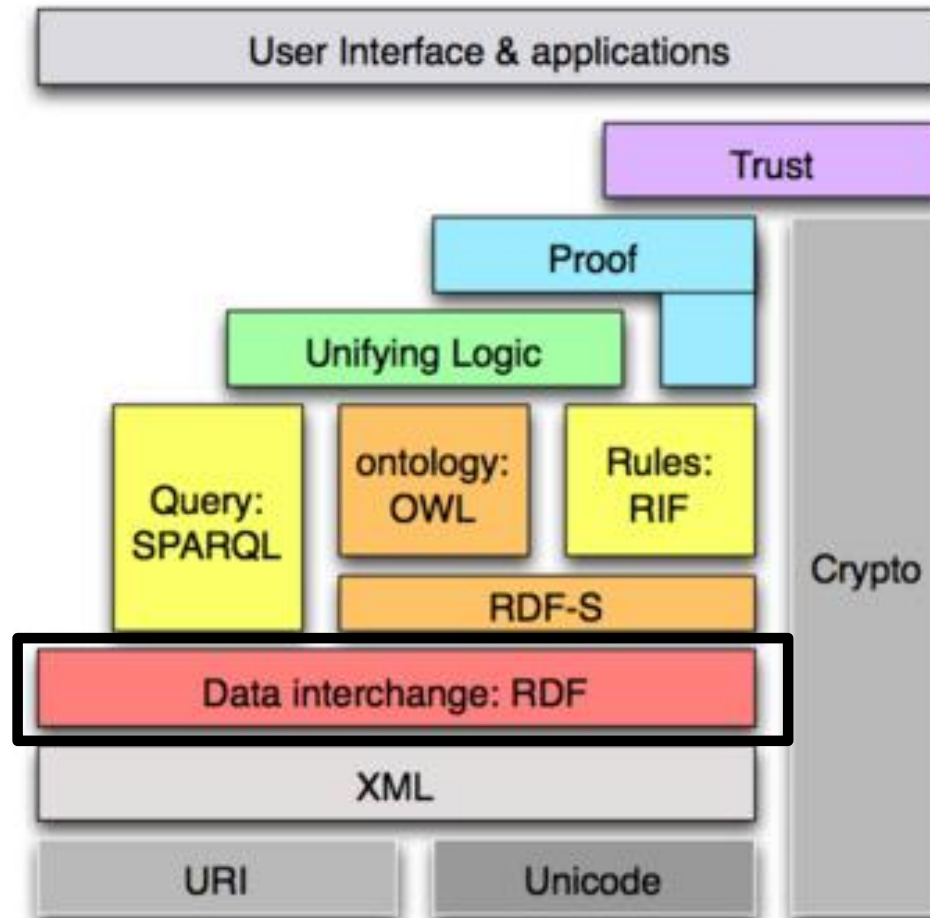


Semantic Web Stack
Berners-Lee (2006)

Semantics on the Web



RDF – Resource Description Framework



Semantic Web Stack
Berners-Lee (2006)

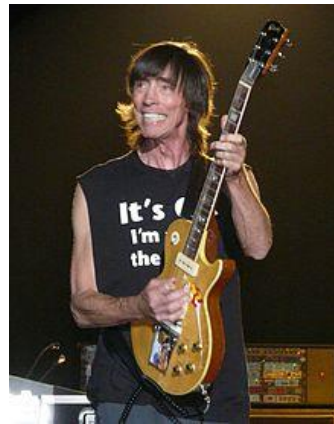
RDF – Resource Description Framework

- RDF is the basis layer of the Semantic Web stack ‘layer cake’.
- Basic building block: RDF triple.
 - **Subject** – a thing (identified by URI)
 - **Predicate** – a relationship (identified by URI)
 - **Object** – another thing (identified by URI) or a literal
- **Subject** has relationship **Predicate** to **Object**
- (**Tom_Scholz** has relationship **lead_singer** to **Boston**)

Why URIs?

RDF – Resource Description Framework

- **Subject** has relationship **Predicate** to **Object**
- (**Tom_Scholz** has relationship **lead_singer** to **Boston**)



- URIs:
 - Avoid ambiguity!
 - Enable linking (discussed later)
 - Enable dereferencing (discussed later)

RDF – Resource Description Framework (Example)

`<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_>`

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

`<http://dbpedia.org/resource/The_Beatles>`.

URIs are given in angle brackets in N-Triples.

`<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_>`

`<http://xmlns.com/foaf/0.1/name>`

`"The Beatles"` .

In N-Triples every statement is terminated with a full stop.

Literals are given in quotes in N-Triples.

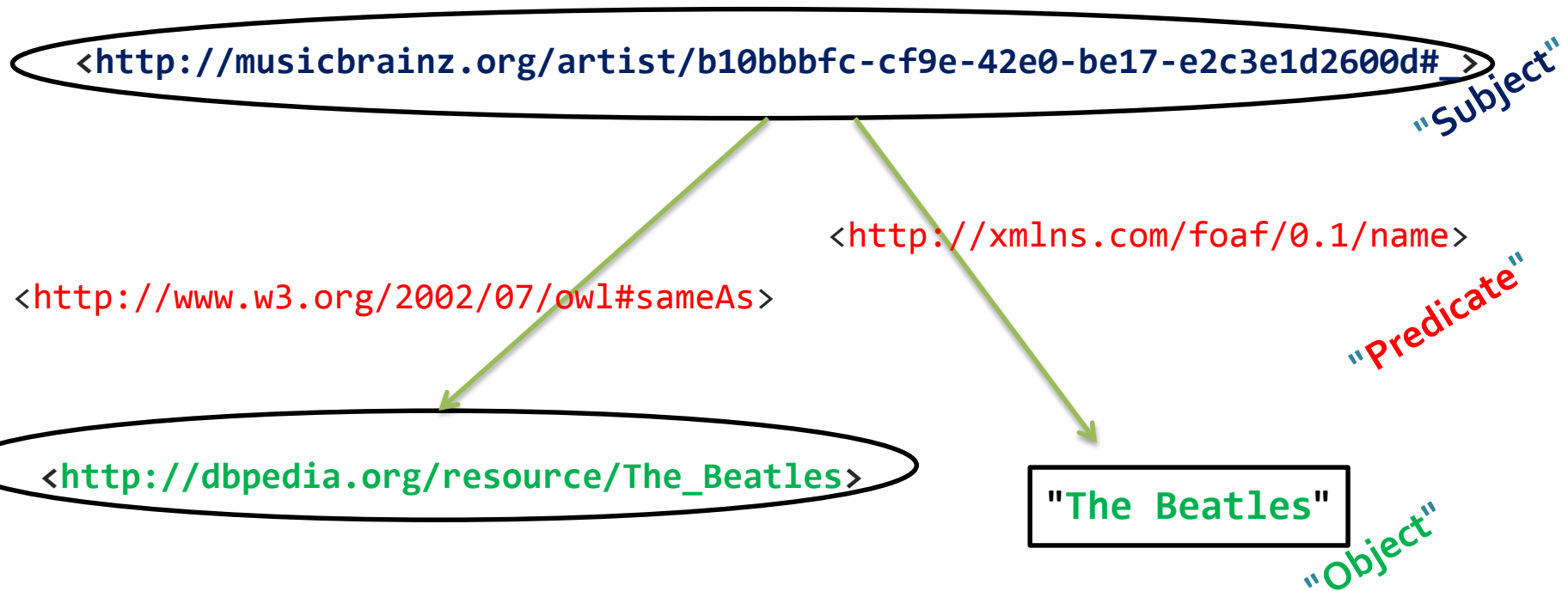
RDF Graphs

- Every set of RDF assertions can then be drawn and manipulated as a (directed labelled) graph:
 - **Resources** – the subjects and objects are nodes of the graph.
 - **Predicates** – each predicate use becomes a label for an arc, connecting the subject to the object.



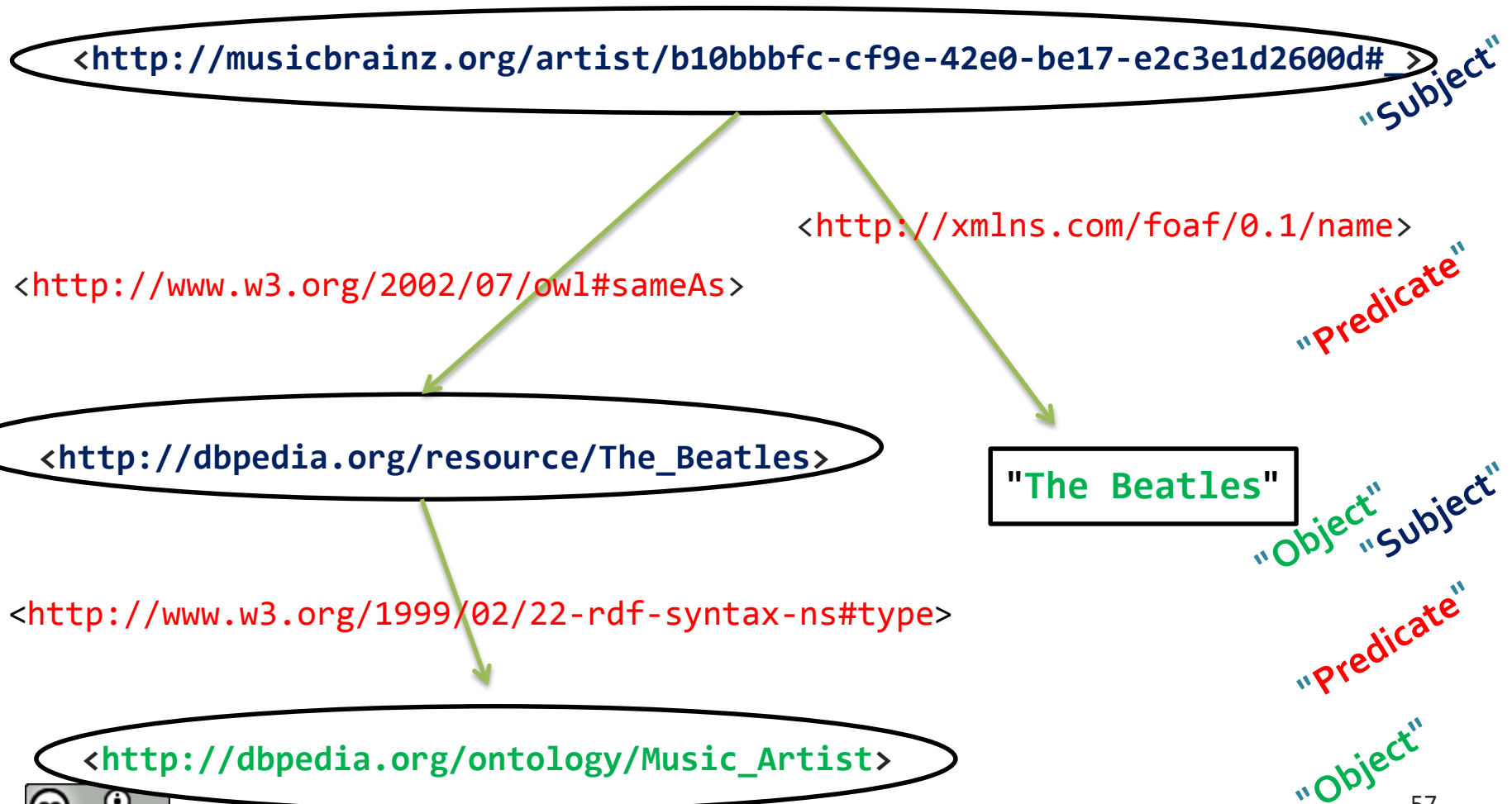
Semantics on the Web

RDF Graphs (Example)



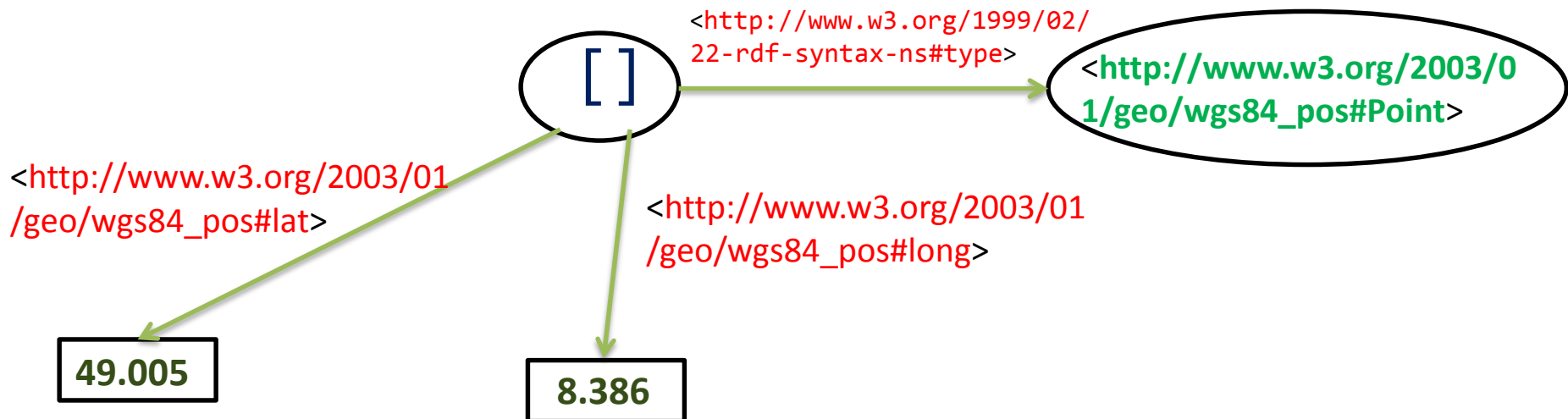
Semantics on the Web

RDF Graphs (Example)



RDF Blank Nodes

- RDF graphs can also contain unidentified resources, called *blank nodes*:



- Blank nodes can group related information, but their use is often discouraged

RDF Turtle

- Turtle is a syntax for RDF more readable.
- Since many URIs share same basis we use **prefixes**:

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

```
@prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#>.
```

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

```
@prefix mo:<http://purl.org/ontology/mo/>.
```

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

And (sometimes) a unique base:

```
@base <http://musicbrainz.org/>.
```

RDF Turtle

- Also has a simple *shorthand* for class membership:

```
@base <http://musicbrainz.org/>.
```

```
@prefix mo:<http://purl.org/ontology/mo/>.
```

```
<artist/b10bbbfccf9e-42e0-be17-e2c3e1d2600d#_> a mo:MusicGroup.
```

Is equivalent to:

```
<http://musicbrainz.org/artist/b10bbbfccf9e-42e0-be17-e2c3e1d2600d#_>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://purl.org/ontology/mo/MusicGroup>.
```

RDF Turtle

- When multiple statements apply to **same subject** they can be abbreviated as follows:

```
<artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_>
```

```
rdfs:label "The Beatles"; _____ Same subject
```

```
owl:sameAs dbpedia:The_Beatles , _____ Same subject &  
<http://www.bbc.co.uk/music/artists/ predicate  
b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#artist> .
```

RDF Turtle

- Turtle also provides a simple syntax for **datatypes** and **language tags** for literals, respectively:

```
<recording/5098d0a8-d3c3-424e-9367-1f2610724410#_> a mo:Signal;  
    rdfs:label "All You Need Is Love" ;  
    mo:duration "PT3M48S"^^xsd:duration .
```

```
dbpedia:The_Beatles dbpedia-owl:abstract  
    "The Beatles were an English rock band formed (...) "@en,  
    "The Beatles waren eine britische Rockband in den (...) "@de .
```

RDF/XML

- Common misconception: RDF is not exotic XML!
 - RDF is triples!
- RDF/XML widespread, but not very nice

```
<mo:MusicArtist>  
  rdf:about="http://musicbrainz.org/artist/b10bbbf-cf9e-42e0-be17-e2c3e1d2600d#_">  
  <foaf:name>The Beatles</foaf:name>  
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/The_Beatles"/>  
</mo:MusicArtist>
```

Describing Data

Vocabularies

- Collections of defined **properties** and **classes** of resources.
 - Classes group together similar resources.
 - **mo:MusicArtist, foaf:Person, geo:Point**
 - Properties denote relationships
 - **mo:member, foaf:name, rdf:type**
- Terms from well-known vocabularies should be **reused** wherever possible.
- New terms should only be defined when you cannot find required terms in existing vocabularies.

Describing Data

Vocabularies

A set of well-known vocabularies has evolved in the Semantic Web community. *Some* of them are:

Vocabulary	Description	Classes and Relationships
Friend-of-a-Friend (FOAF)	Vocabulary for describing people.	foaf:Person, foaf:Agent, foaf:name, foaf:knows, foaf:member
Dublin Core (DC)	Defines general metadata attributes.	dc:FileFormat, dc:MediaType, dc:creator, dc:description
Semantically-Interlinked Online Communities (SIOC)	Vocabulary for representing online communities.	sioc:Community, sioc:Forum, sioc:Post, sioc:follows, sioc:topic
Music Ontology (MO)	Provides terms for describing artists, albums and tracks.	mo:MusicArtist, mo:MusicGroup, mo:Signal, mo:member, mo:record
Simple Knowledge Organization System (SKOS)	Vocabulary for representing taxonomies and loosely structured knowledge.	skos:Concept, skos:inScheme, skos:definition, skos:example

Describing Data

Vocabularies

More extensive lists of well-known vocabularies are maintained by:

- W3C SWEO Linking Open Data community project
<http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies>
- Mondeca: Linked Open Vocabularies
<http://labs.mondeca.com/dataset/lov>
- Library Linked Data Incubator Group: Vocabularies in the library domain
<http://www.w3.org/2005/Incubator/Ild/XGR-Ild-vocabdataset-20111025>

Describing Data

Defining Vocabularies:

1. Informal human-readable documentation
2. Formal machine-readable documentation

Class: mo:Festival - stable - level 2

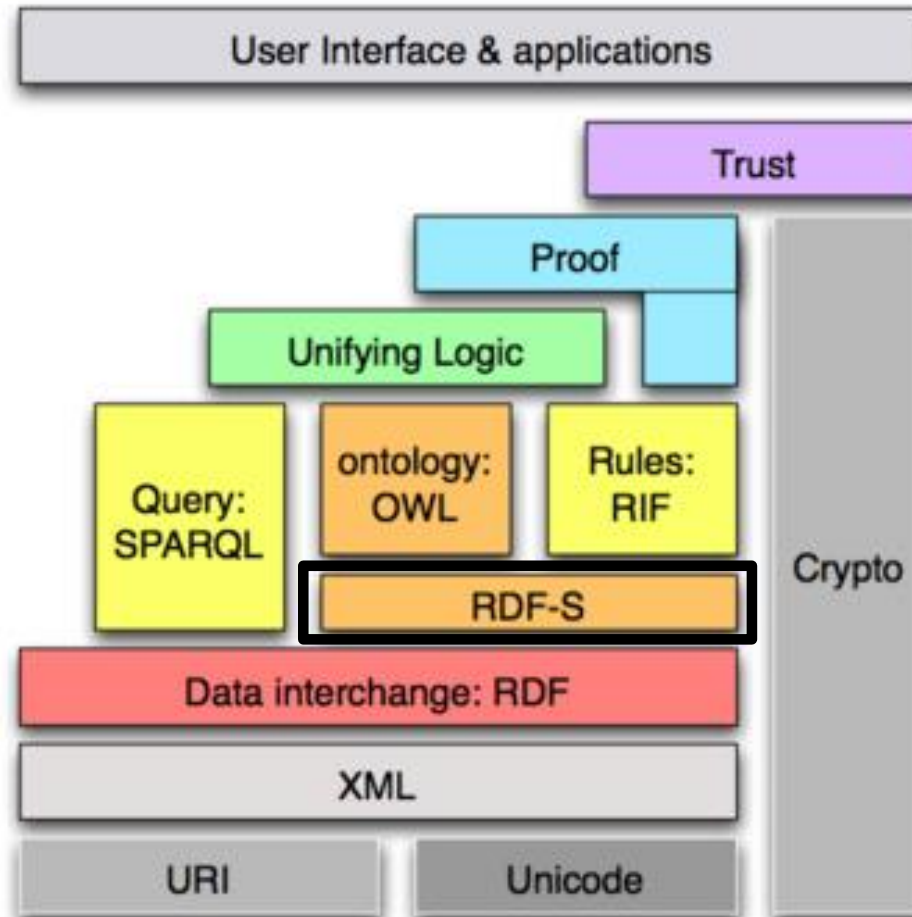
Festival - A festival - musical/artistic event lasting several days, like Glastonbury, Rock Am Ring...

We might decompose this event (which is in fact just a classification of the space/time region related to a particular festival) using `hasSubEvent` in several performances at different space/time.

sub-class-of: [event:Event](#)

Later on ...

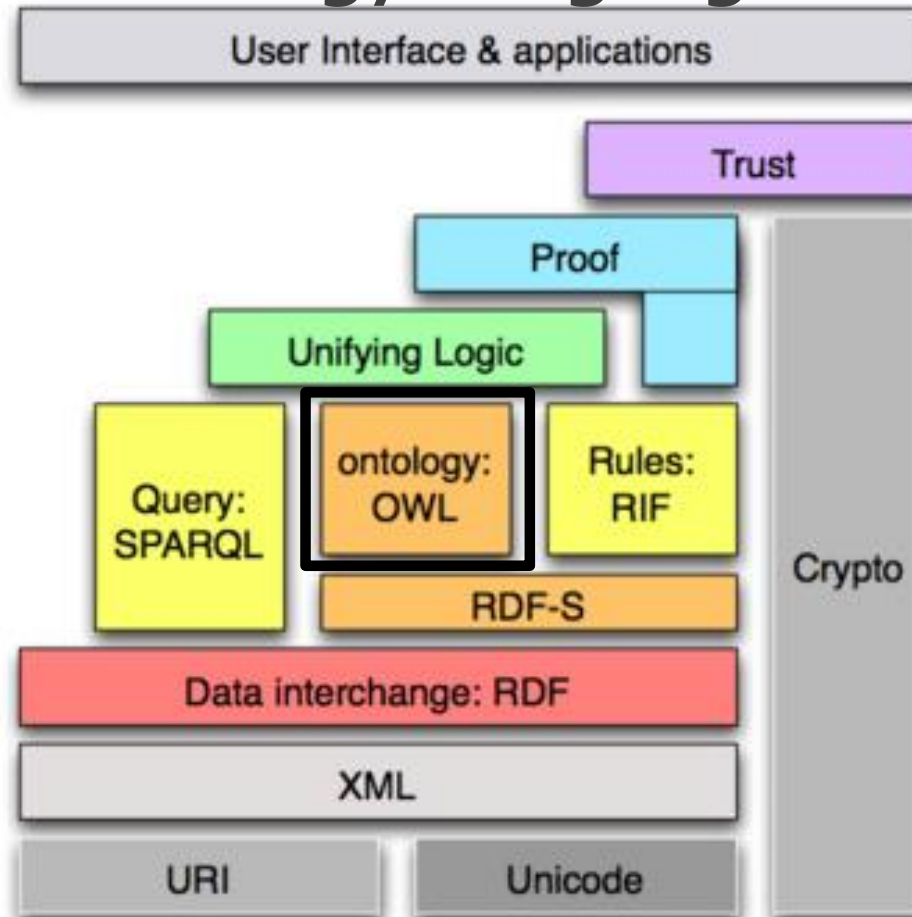
RDF-S – RDF Schema



Semantic Web Stack
Berners-Lee (2006)

Later on ...

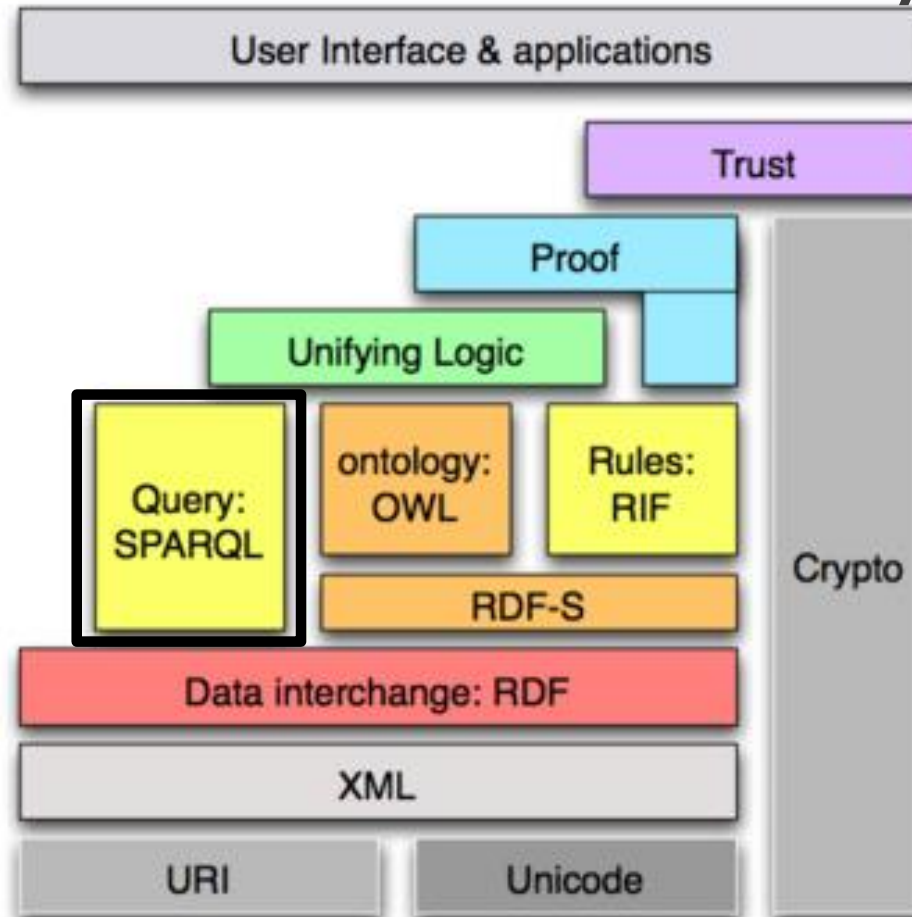
OWL – Web Ontology Language



Semantic Web Stack
Berners-Lee (2006)

Later on ...

SPARQL – * Protocol and RDF Query Language



Semantic Web Stack
Berners-Lee (2006)

Acknowledgements

- Alexander Mikroyannidis
- Alice Carpentier
- Andreas Harth
- Andreas Wagner
- Andriy Nikolov
- Barry Norton
- Daniel M. Herzig
- Elena Simperl
- Günter Ladwig
- Inga Shamkhalov
- Jacek Kopecky
- John Domingue
- Juan Sequeda
- Kalina Bontcheva
- Maria Maleshkova
- Maria-Esther Vidal
- Maribel Acosta
- Michael Meier
- Ning Li
- Paul Mulholland
- Peter Haase
- Richard Power
- Steffen Stadtmüller

INTRODUCTION TO: LINKED DATA

