Hes·so
Haute Ecole Spécialisée
de Suisse occidentale

# MAS – Web Sémantique

**Bases et Utilisations du Web Sémantique**

**Julien TSCHERRIG, Joël DUMOULIN, Omar ABOU KHALED**

**University of Applied Sciences of Western Switzerland
EIA-FR, Bd de Pérolles 80 - CP 32, CH-1705 Fribourg
julien.tscherrig@hefr.ch| Tél: +41 26 429 69 64
joel.dumoulin@hefr.ch| Tél: +41 26 429 69 60
omar.aboukhaled@hefr.ch |Tél: +41 26 429 65 89
http://humantech.eia-fr.ch | Fax: +41 26 429 66 00**

HumanTech
Technology for
Human Wellbeing Institute

# Plan de la présentation

- Organisation du cours

- Les Fondements du Web Sémantique
  - Historique (survole)
  - Web sémantique et Ontologie
  - RDF, (RDFa,) RDFS et OWL

- Utilisation du Web Sémantique
  - Linked Data et Open Data (survole)
  - Use Case : Parlement

HumanTech
Technology for
Human Wellbeing Institute

# Organisation du cours

MAS – Web Sémantique

# Organisation du cours

- Intervenants
  - Théorie
    - Julien Tscherrig (EIA-FR)
    - Maria Sokhn (HES-SO//Valais)
  - Pratique
    - Joël Dumoulin (EIA-FR)

- 3 Journées (jeudi)
  - 11.12.2014
  - 17.12.2014
  - 08.01.2015

- Répartition
  - Matin – Théorie/Pratique 09h-12h30
  - Après-midi – Pratique 13h30-16h30

- Evaluation du cours
  - Réalisation d'un projet

HumanTech
Technology for
Human Wellbeing Institute

# Organisation du cours *(11.12.2014)*

**Théorie**

- Les fondements
  - Historique (survole)
  - Web sémantique et Ontologie
  - RDF, (RDFa,) RDFS et OWL
- Usage du Web Sémantique
  - Linked Data et Open Data (survol)
  - Use case : le Parlement fédéral

**Pratique**

- Installation des outils (RDF4J, Tomcat, éclipse, etc.)
- Création d'une ontologie avec Protégé type « Hello world »
- ajout des individus dans l'ontologie

HumanTech
Technology for
Human Wellbeing Institute

# Organisation du cours *(18.12.2014)*

**Théorie**

- Approfondissement des concepts
  - Introdution au Big data
  - Modélisation d'une ontologie
  - Bases de données sémantiques

- Raisonnement et interrogation
  - Le langage SPARQL
    - introduction et approfondissement

**Pratique**

- Explorer FOAF

- Modéliser une ontologie

- Utilisation de SPARQL

# Organisation du cours *(08.01.2015)*

**Théorie**

- Les fondements
  - Les origines, du web 1.0 au web 2.0
  - Les composants du Web 2.0
  - Concepts clés.

- Usage collaboratif du Web 2.0
  - Les outils de publication (blog, wiki, podcast, etc.)
  - Les outils de partage (flux RSS, etc.)

**Pratique**

- Création d'un widget et mise en place d'un « Creative Common »

# Fundamentals
# of the Semantic Web

MAS – Web Sémantique

HumanTech
Technology for
Human Wellbeing Institute

# Development of the Web (1)

- ## Web 1.0
  - *"The World Wide Web ("WWW" or simply the "Web") is a system of interlinked, hypertext documents that runs over the Internet. With a Web browser, a user views Web pages that may contain text, images, and other multimedia and navigates between them using hyperlinks".*

    Source http://en.wikipedia.org/wiki/World_Wide_Web

- ## The success of Web1.0 is based on three simple principles
  - A simple and uniform addressing schema to identify information chunks
    - Uniform Resource Identifier (**URIs**)
  - A simple and uniform representation formalism to structure information chunks allowing browsers to render them i.e. **H**yper **T**ext **M**arkup **L**anguage
  - A simple and uniform protocol to access information chunks (**H**yper **T**ext **T**ransfer **P**rotocol)

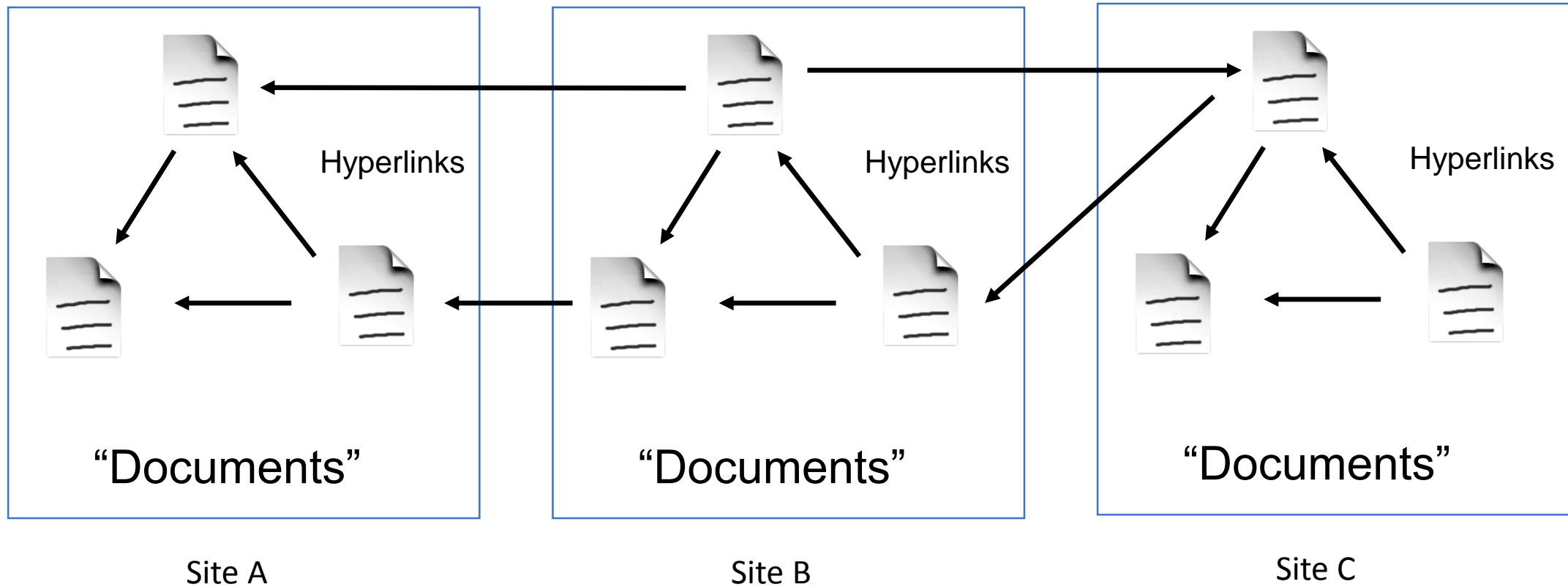HumanTech
Technology for
Human Wellbeing Institute

# Development of the Web (2)

- Web 2.0
  - "The term "Web 2.0" (2004–present) is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web"

Source http://en.wikipedia.org/wiki/Web_2.0

- The major breakthroughs of Web 2.0 are
  - Blurring the distinction between content consumers and content providers
  - Moving from media for individuals towards media for communities
  - Blurring the distinction between service consumers and service providers
  - Integrating human and machine computing in a new and innovative way

HumanTech
Technology for
Human Wellbeing Institute

# Current Web: Web of Documents



Hyperlinks                    Hyperlinks                    Hyperlinks

"Documents"          "Documents"          "Documents"

Site A                       Site B                       Site C

# Limitation of the curent Web (1)

- The current Web has its limitations when it comes to:

  - Finding relevant information

  - Extracting relevant information
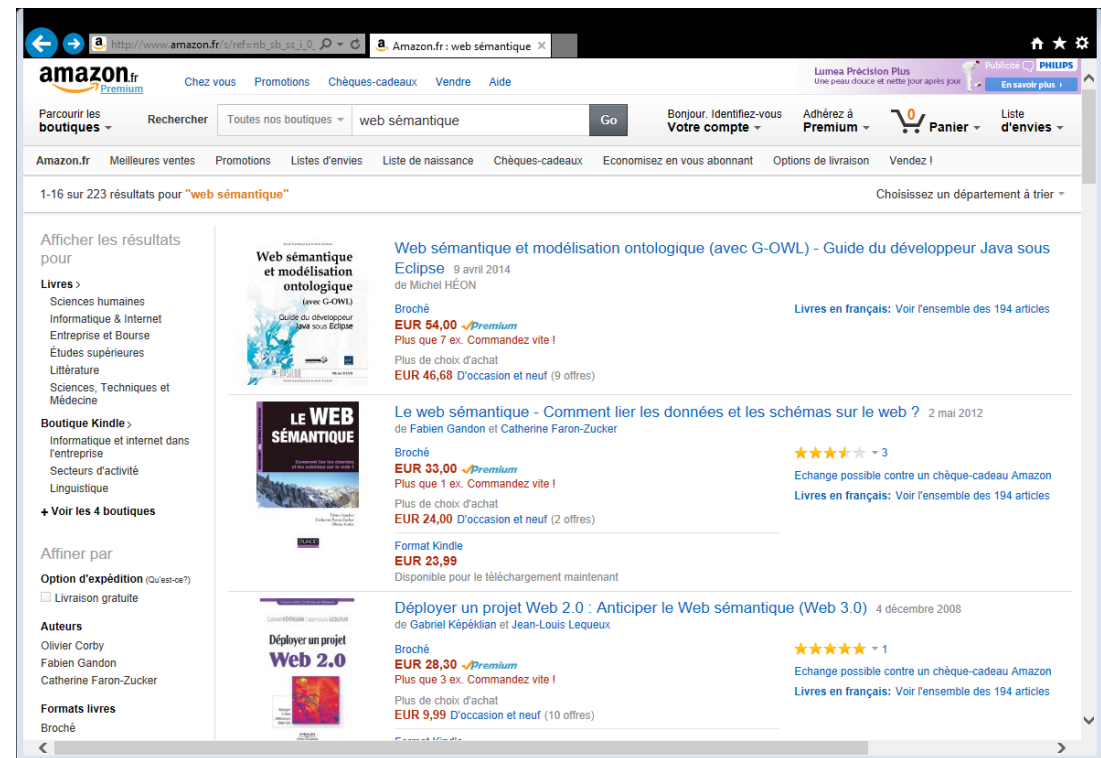
  - Combining and reusing information

HumanTech

Technology for
Human Wellbeing Institute

# Limitation of the curent Web (2)

- Finding information on the current Web is based on keyword search
- Keyword search has a limited recall and precision due to:
  - Synonyms:
    - e.g. Searching information about "Cars" will ignore Web pages that contain the word "Automobiles" even though the information on these pages could be relevant
  - Homonyms:
    - e.g. Searching information about "Jaguar" will bring up pages containing information about both "Jaguar" (the car brand) and "Jaguar (the animal)

# Limitation of the curent Web (3)

- Extracting relevant information
  - One-fit-all automatic solution for extracting information from Web pages is not possible due to different formats, different syntaxes
  - Even from a single Web page is difficult to extract the relevant information

# Limitation of the curent Web (4)

- Combining and reusing information
  - Searches for the same information in different digital libraries
    - e.g. I want to travel from Fribourg to Florence
  - Information may come from different web sites and needs to be combined
    - e.g. I want to travel from Fribourg to Florence where I want to stay in a hotel and visit the city

HumanTech
Technology for
Human Wellbeing Institute
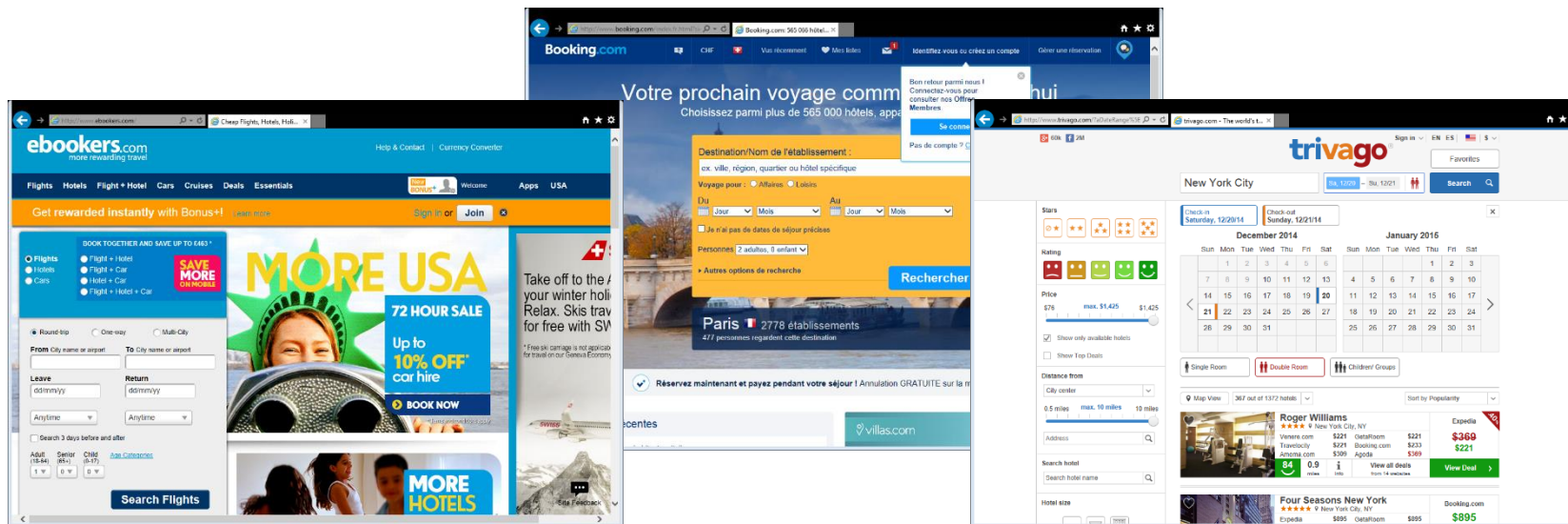
# Limitation of the curent Web (5)

- Try these queries with Google:
  - "Distance between Paris and Madrid"
    - https://www.google.ch/?gws_rd=ssl#q=Distance+between+Paris+and+Madrid : 1053.4 km

  - "Largest city of France"
    - https://www.google.ch/?gws_rd=ssl#q=Largest+city+of+France : Paris

  - "Largest city of Spain"
    - https://www.google.ch/?gws_rd=ssl#q=Largest+city+of+Spain : Madrid

- Now, try these with Google:
  - "Distance between largest city of France and largest city of Spain"
    - *No result returned by Google!*

# How do we improve the current Web ?

- Increasing automatic linking among data

- Increasing recall and precision in search

- Increasing automation in data integration

- Increasing automation in the service life cycle

Adding semantics to data
and services is the solution!

Serious Problems in information
- finding
- extracting
- representing
- interpreting
- and maintaining

| WWW | | Semantic Web |
|---|---|---|
| URI, HTML, HTTP | → | RDF, RDF(S), OWL |
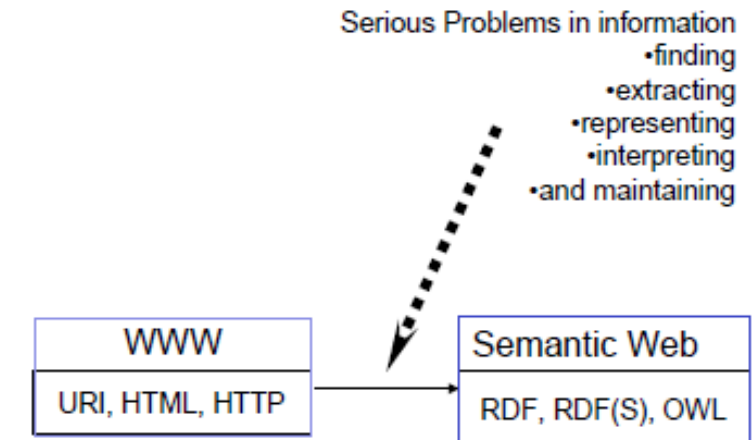
# Web of Documents to Web of Data

Hyperlinks

Typed Links

"Documents"
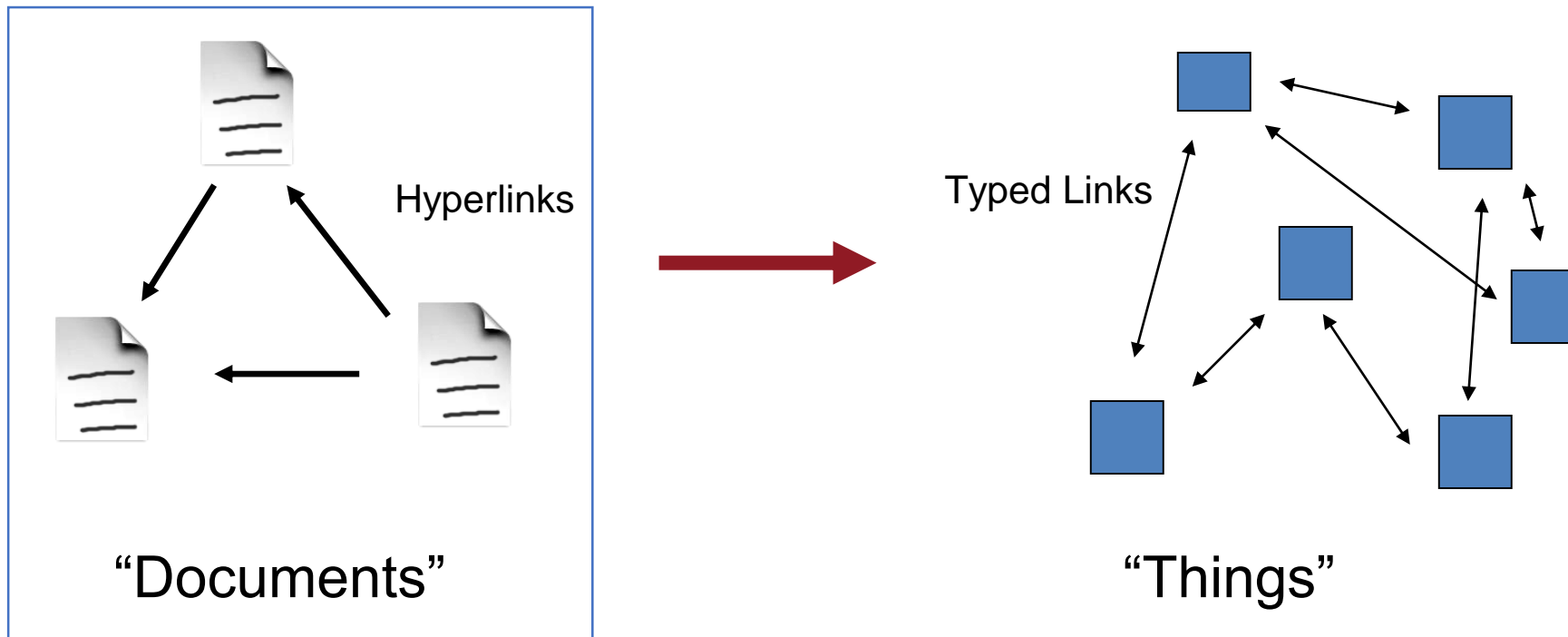
"Things"

# Vision of the Web of Data

- The Web today
  - Consists of data silos which can be accessed via specialized search engines in an isolated fashion.
  - One site (data silo) has movies, the other reviews, again another actors.
  - Many common things are represented in multiple data sets
  - Linking identifiers link these data sets
- The Web of Data is envisioned as a global database
  - consisting of objects and their descriptions
  - in which objects are linked with each other
  - with a high degree of object structure
  - with explicit semantics for links and content
  - which is designed for humans and machines
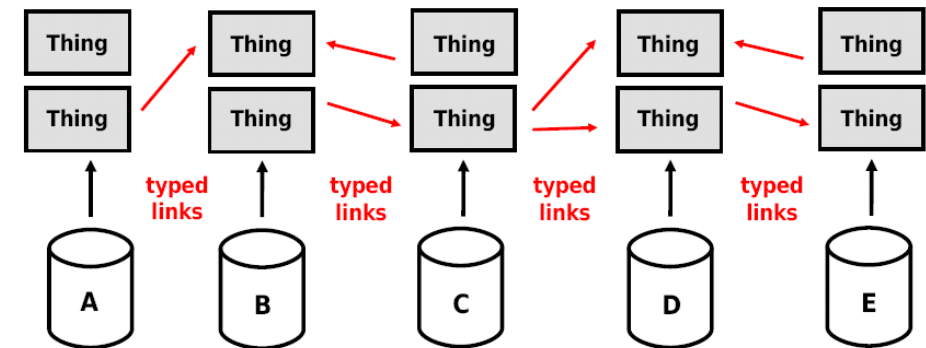


*Content on this slide by Chris Bizer, Tom Heath and Tim Berners-Lee*

HumanTech
Technology for
Human Wellbeing Institute

# What is the Semantic Web?

- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", Scientific American, May 2001

- The next generation of the WWW
- Information has machine processable and machine understandable semantics
- Not a separate Web but an augmentation of the current one
- The backbone of Semantic Web are **ontologies**

HumanTech
Technology for
Human Wellbeing Institute

# What is the Semantic Web?

- Ontology Definition
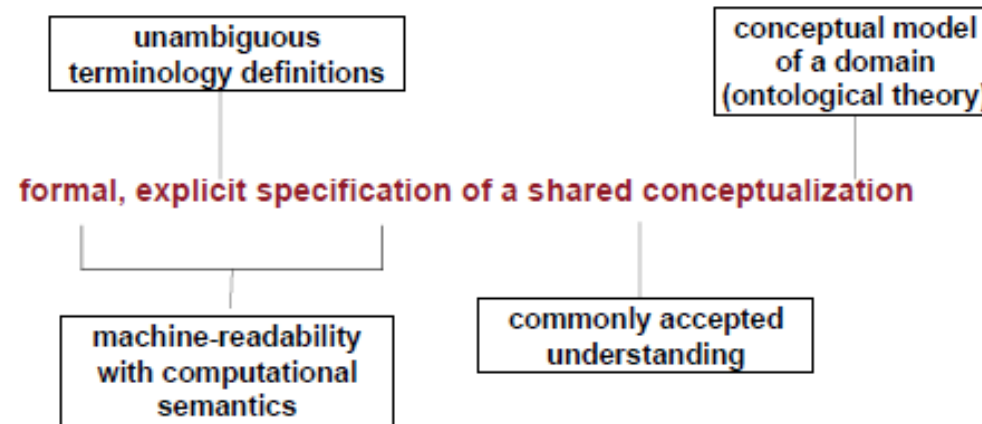


Gruber, Toward principles for the design of ontologies used or knowledge sharing?, Int. J. Hum.-Comput. Stud., vol. 43, no. 5-6,1995

- Ontologies are the modeling foundations to Semantic Web
  - They provide the *well-defined meaning* for information

# Ontology definition

- Concept
  - Conceptual entity of the domain

- Property
  - Attribute describing a concept
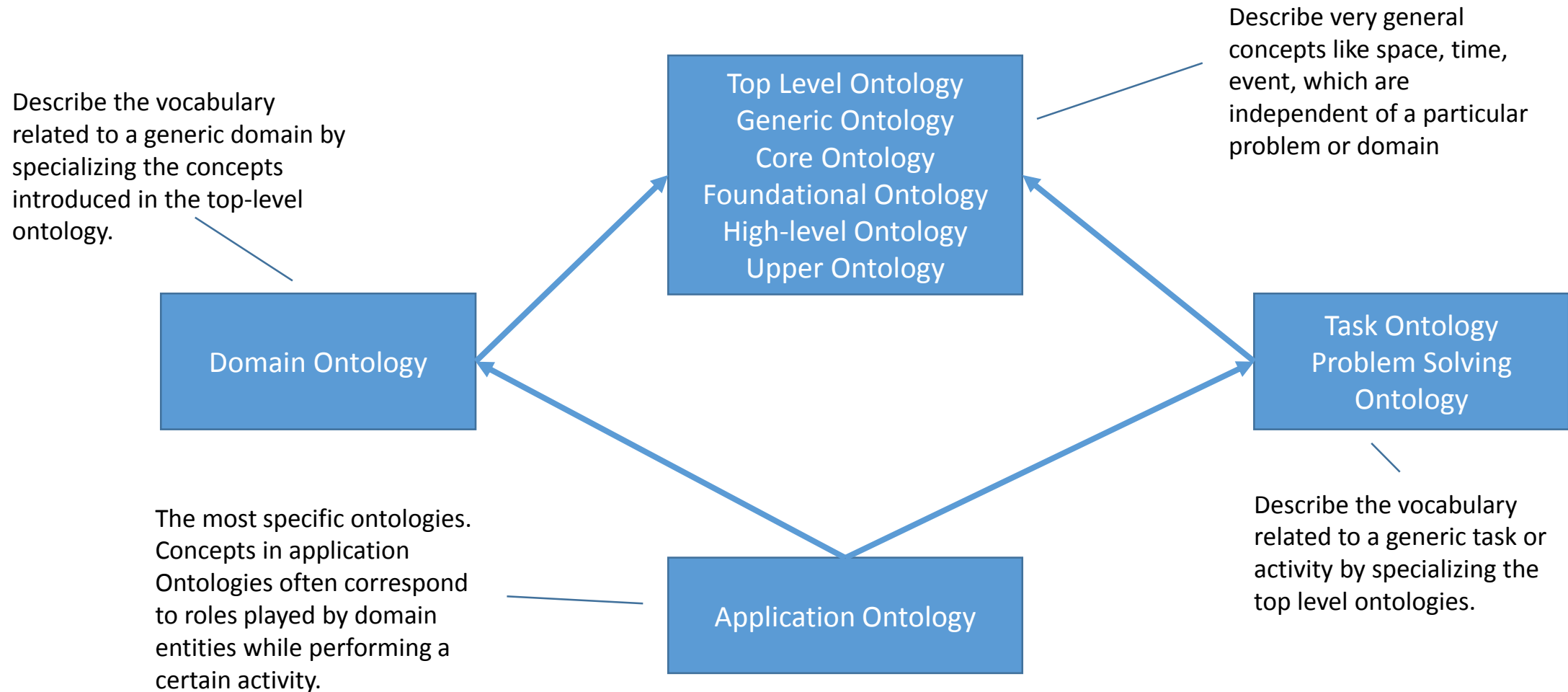
- Relation
  - Relationship between concepts or properties
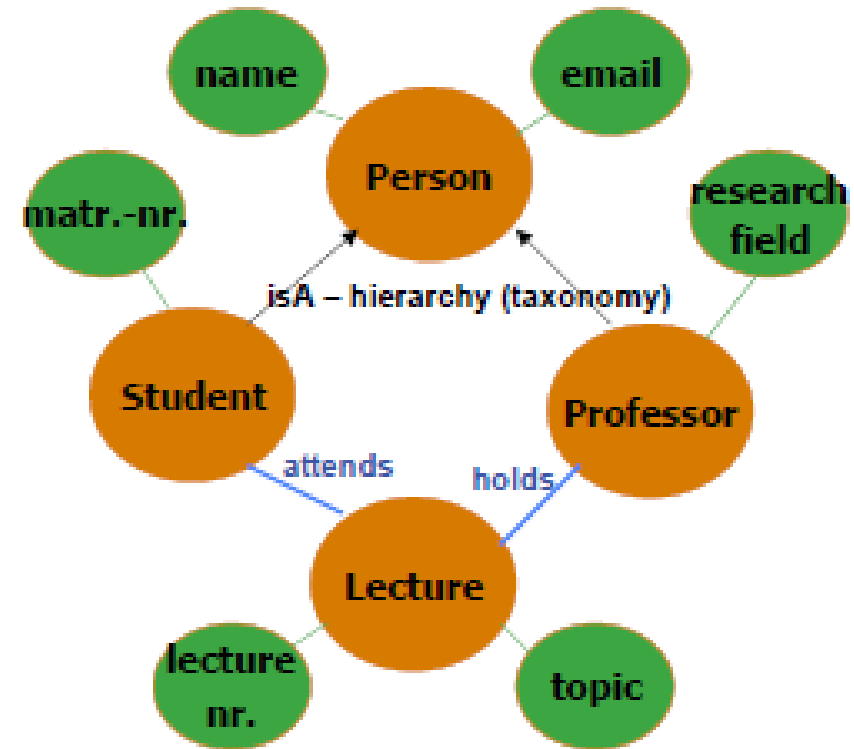
# Ontology is…

- **A conceptualization:** An ontology is a model of the most relevant concepts of a phenomenon from the real world

- **Explicit:** The model explicitly states the type of the concepts, the relationships between them and the constraints on their use

- **Formal:** The ontology has to be machine readable

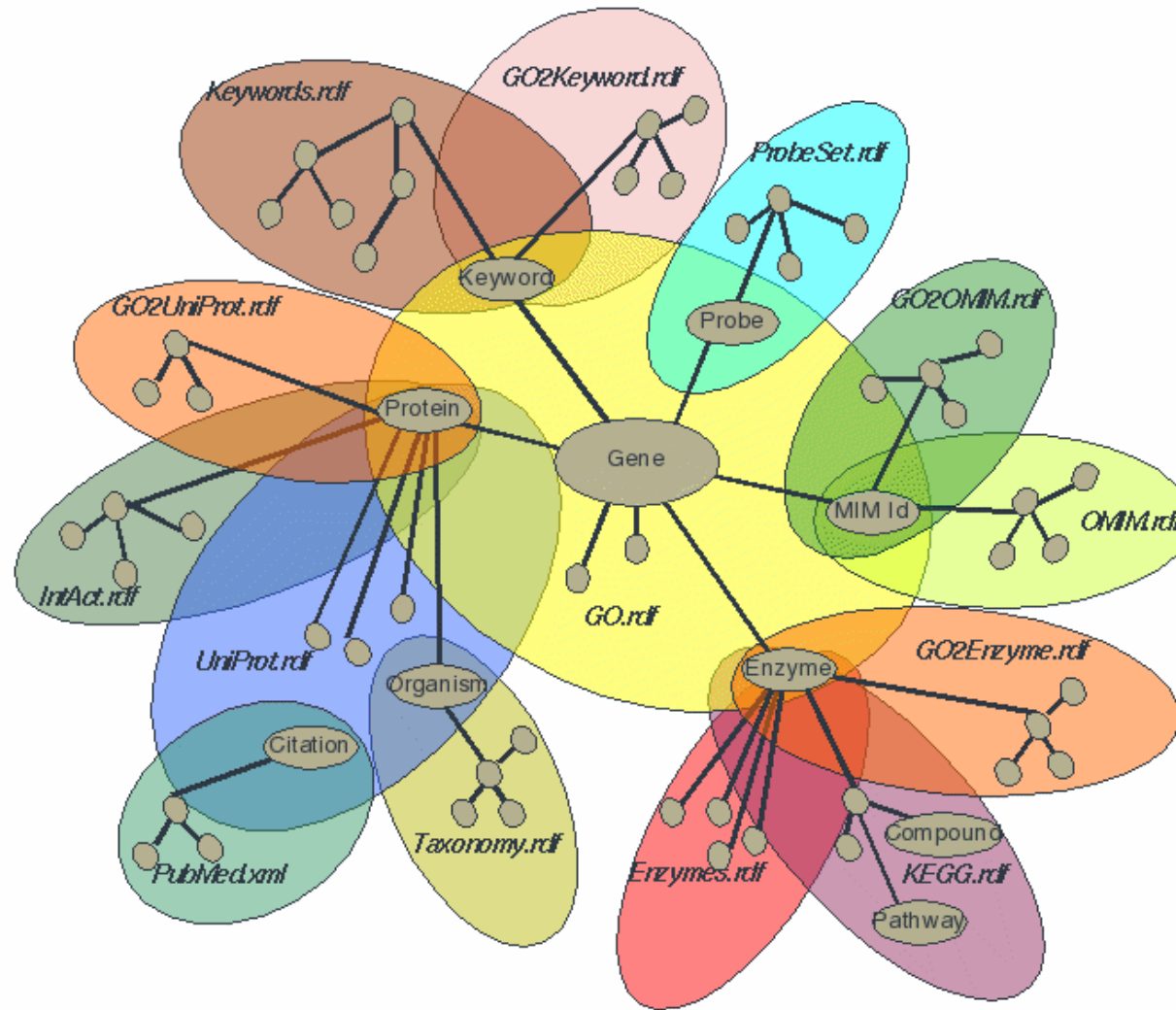- **Shared :**The knowledge contained in the ontology is consensual, i.e. it has been accepted by a group of people.

Studer, Benjamins, D. Fensel, Knowledge engineering: Principles and methods, Data Knowledge Engineering, vol. 25, no. 1-2, 1998.

HumanTech
Technology for
Human Wellbeing Institute

# Ontology Types

Describe the vocabulary related to a generic domain by specializing the concepts introduced in the top-level ontology.

Describe very general concepts like space, time, event, which are independent of a particular problem or domain

**Top Level Ontology**
**Generic Ontology**
**Core Ontology**
**Foundational Ontology**
**High-level Ontology**
**Upper Ontology**

**Domain Ontology**

**Task Ontology**
**Problem Solving Ontology**

The most specific ontologies. Concepts in application Ontologies often correspond to roles played by domain entities while performing a certain activity.

**Application Ontology**

Describe the vocabulary related to a generic task or activity by specializing the top level ontologies.

Formal Ontology in Information Systems, http://www.loa-cnr.it/Papers/FOIS98.pdf

HumanTech
Technology for
Human Wellbeing Institute

# Life Sciences Ontology Example

HumanTech
Technology for
Human Wellbeing Institute

# Semantic Web is about…

- Web Data Annotation
  - connecting (syntactic) Web objects, like text chunks, images, … to their semantic notion
  - e.g., this image is about Innsbruck, Dieter Fensel is a professor

- Data Linking on the Web (Web of Data)
  - global networking of knowledge through URI, RDF, and SPARQL
  - e.g., connecting my calendar with my rss feeds, my pictures, ...

- Data Integration over the Web
  - seamless integration of data based on different conceptual models
  - e.g., integrating data coming from my two favorite book sellers

- Example: DBpedia, BBC Music

# Semantic Web and….

- **The cloud computing**
  - Flexible, dynamically scalable and virtualized data layer as part of the cloud
  - Accurate search and acquire various data from the Internet

- **The internet of things (IoT)**
  - Interoperability at the level of sensors data and protocols
  - More precise search for mobile capabilities and sensors with desired capability

- **The mobile applications**
  - Exploiting information coming formdifferent datasets, e.g. DBpedia Mobile

# Semantic Web - Language Tower
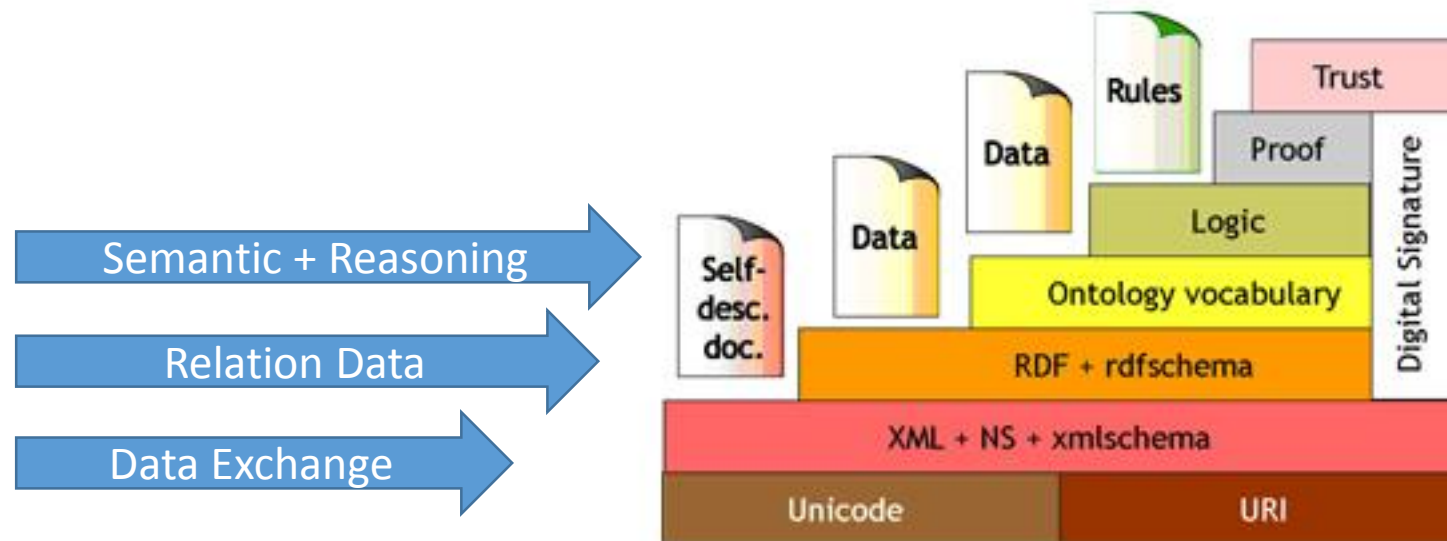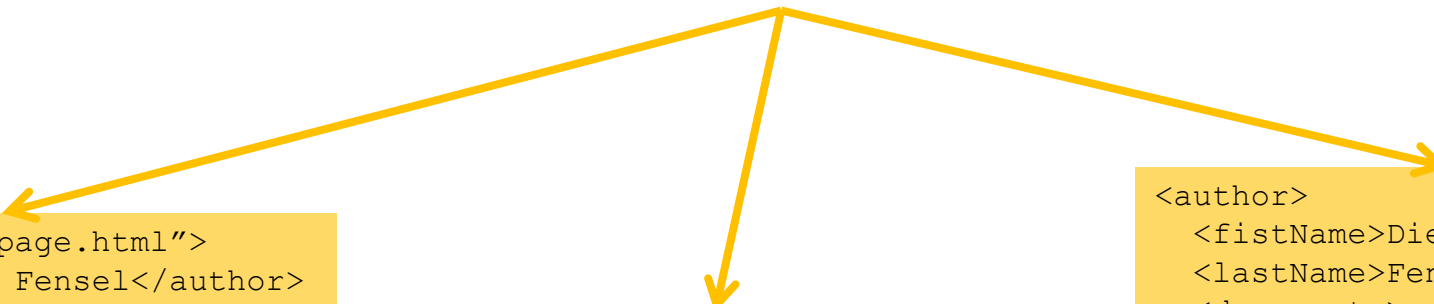
# Web Semantic Technologies - RDF

- The Resource Description Framework is an language to describe resources.
  - Resource (e.g. webpage, book, person, photo, etc.)
  - While XML documents attach meta data to parts of a document, one use of RDF is to create meta data about the document as a standalone entity.

- In other words, instead of marking up the internals of a document, RDF captures meta data about the "externals" of a document, like the author, the creation date, and type.

- A particularly good use of RDF is to describe resources, which are "opaque" like images or audio files.

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - RDF

**Why do we need RDF?**

The author of 'page.html' is Dieter Fensel.

```
<document href="page.html">
  <author>Dieter Fensel</author>
</document>
```

```
<author>
  <name>Dieter Fensel</name>
  <uri>page.html</uri>
</author>
```

```
<author>
  <fistName>Dieter</fistName>
  <lastName>Fensel</lastName>
  <documents>
    <uri>page.html</uri>
  </documents>
</author>
```
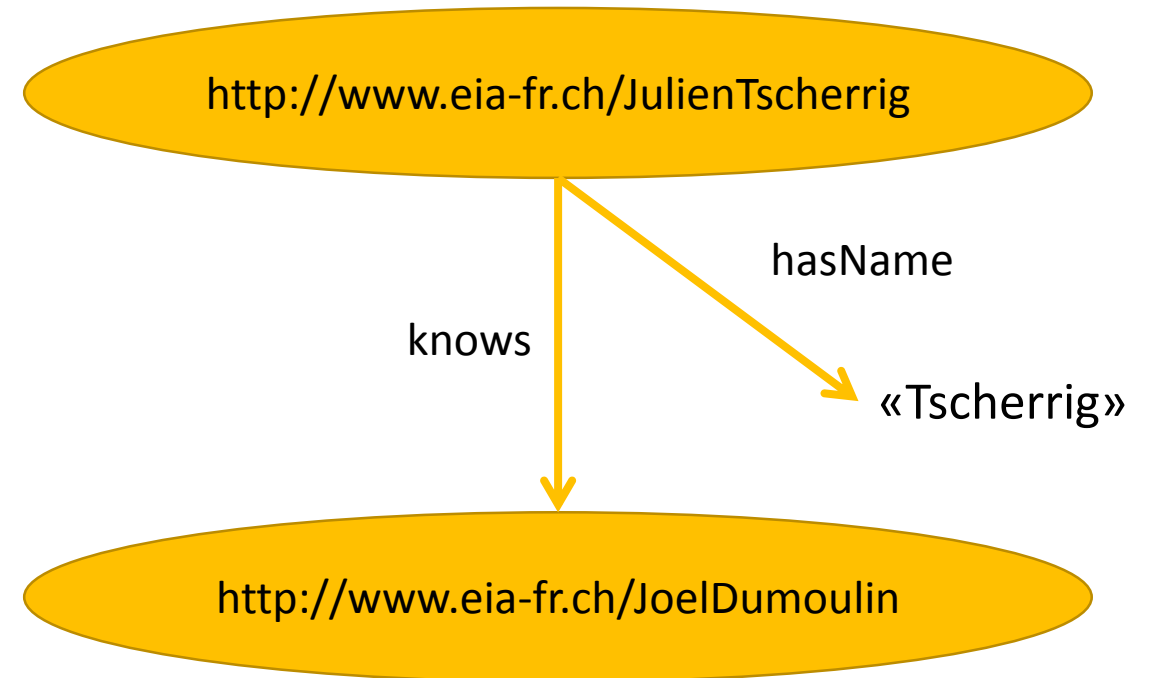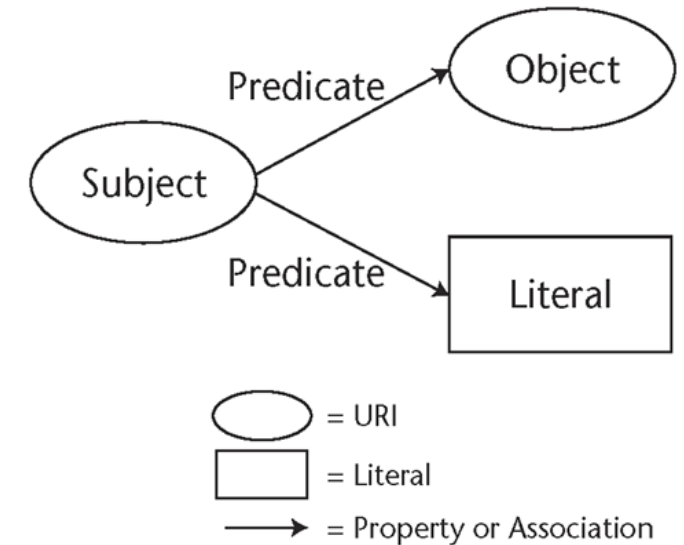
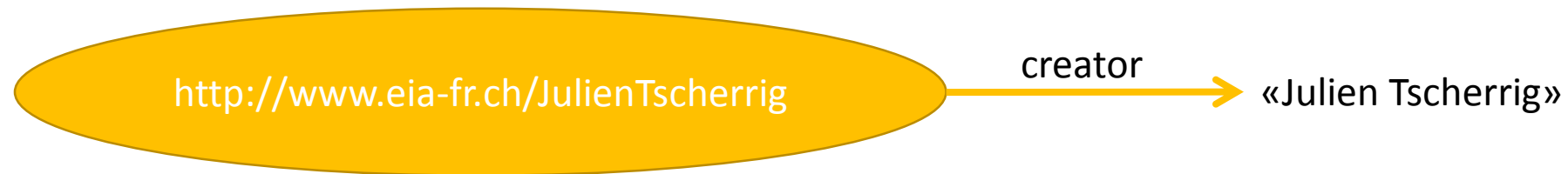What is the correct way of expressing it?
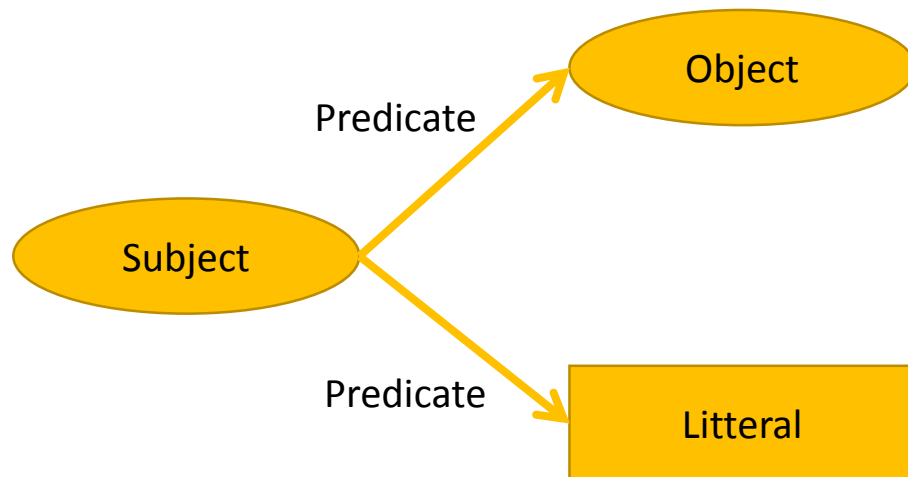
# Web Semantic Technologies - RDF

- **Subject:**
  - In grammar, this is the noun or noun phrase that is the doer of the action
  - In RDF, this is the resource that is being described by the succession of predicates and objects
- **Predicate:**
  - In grammar, this is the part of a sentence that modifies the subject and includes the verb phrase
  - In RDF, a predicate is a relation between the subject and the object
- **Object:**
  - In grammar this is a noun that is acted upon by the verb.
  - In RDF, an object is either a resource referred to by the predicate or a literal value.
- **Statement**: In RDF, the combination of the preceding three elements, subject, predicate, and object, as a single unit.



= URI

= Literal

= Property or Association

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - RDF

**RDF Data Model**



```
http://www.eia-fr.ch/JulienTscherrig          creator      «Julien Tscherrig»
```

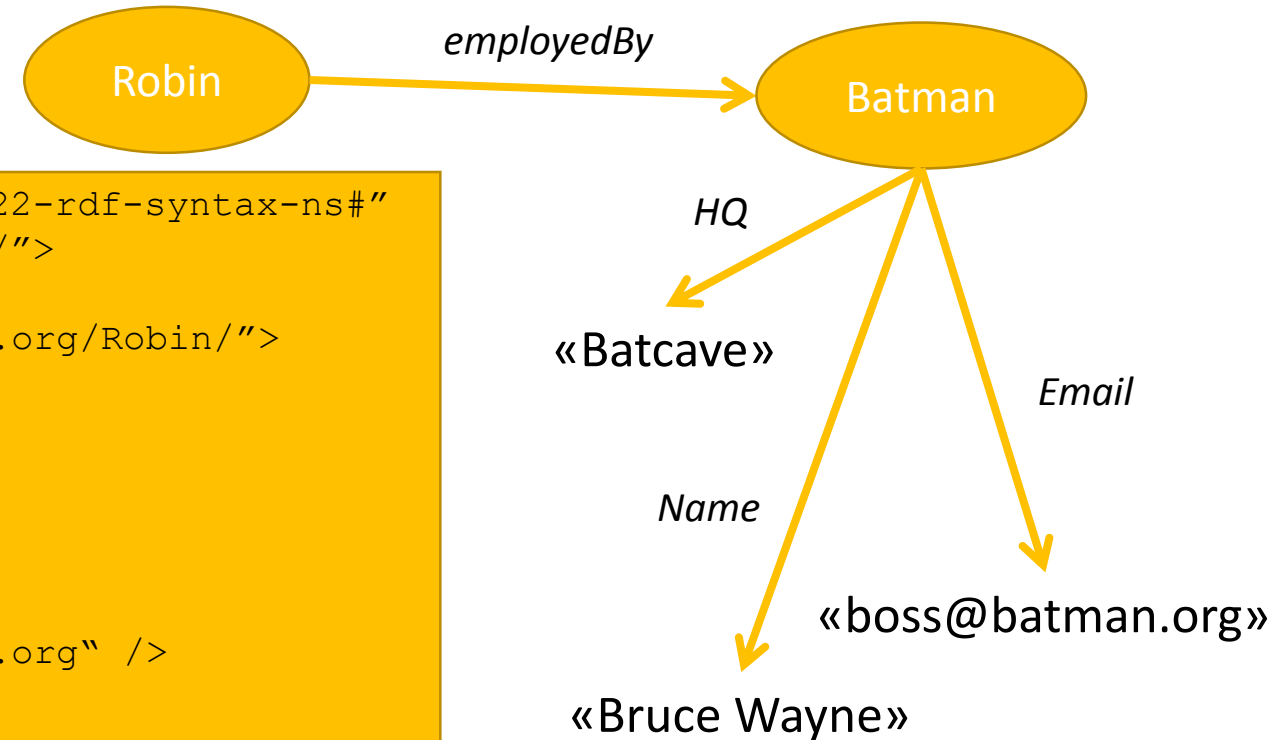Format RDF/XML

```
<rdf:rdf xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc = "http://purl.oclc.org/DC">
  <rdf:description rdf:href = "http://www.eia-fr.ch/JulienTscherrig">
    <dc:creator>Julien Tscherrig</dc:creator>
  </rdf:description>
</rdf:RDF>
```

Format turtle, n3 ou notation 3

```
http://www.eia-fr.ch/JulienTscherrig dc:creator "Julien Tscherrig"
```

# Web Semantic Technologies - RDF

**RDF Data Model (graph)**

Format RDF/XML

```
<rdf:rdf xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:mySchema = "http://www.Batman.org/mySchema/">

  <rdf:description rdf:href = "http://www.Batman.org/Robin/">
    <mySchema:employedBy rdf:href = "#Batman"/>
  </rdf:description>

  <rdf:description id = "Batman">
    <mySchema:HQ>Batcave</mySchema:HQ>
    <mySchema:Name>Bruce Wayne</mySchema:Name>
    <mySchema:Email rdf:href="mailto:boss@batman.org" />
  </rdf:description>

</rdf:rdf>
```

Robin  —*employedBy*→  Batman

*HQ*  «Batcave»

*Email*  «boss@batman.org»

*Name*  «Bruce Wayne»

Format turtle, n3 ou notation 3

```
http://www.Batman.org/Robin/ mySchema:employedBy http://www.Batman.org/mySchema/Batman
http://www.Batman.org/mySchema/Batman mySchema:HQ Batcave
http://www.Batman.org/mySchema/Batman mySchema:Name Bruce Wayne
...
```

# Web Semantic Technologies - RDF**a**

- RDFa is a serialization syntax for embedding an RDF graph into XHTML

- RDFa is domain-independent.

- Benefits inherited from RDF
    - Independence
    - Modularity
    - Evolvability
    - Reusability

# Web Semantic Technologies - RDFa

**Library** **http://schema.org**

- Details
  - Provides a collection of schemas that webmasters can use to markup their pages in ways recognized by major search providers
  - To improve the display of search results, making it easier for people to find the right web pages
- Search engines including
  - Bing
  - Google
  - Yahoo!
- Available Schemas
  - Creative works: CreativeWork, Book, Movie, MusicRecording, Recipe, etc.
  - Embedded non-text objects: AudioObject, ImageObject, VideoObject , etc.
  - Others: Event, Organization, Person, etc.

# Web Semantic Technologies - RDF**a**

- **Visualization**
  - https://support.google.com/webmasters/answer/146898?hl=en
  - http://rdfa.info/play/

- **Example**
  - https://support.google.com/webmasters/answer/146750

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - RDF**a**



Page contains RDFa tags based on http://schema.org/Recipe
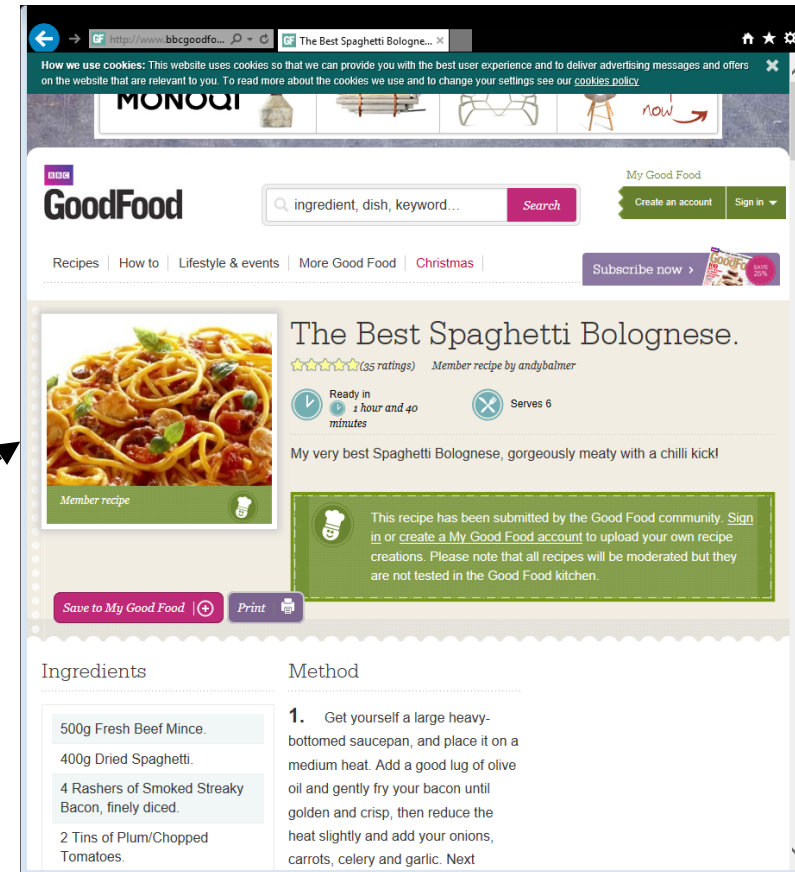
# Web Semantic Technologies - RDF**a**

- Before RDFa -> Microformats…
  - Adding of simple semantics with Microformats (small pieces)
  - Adding semantics to the today's Web instead of creating a new one (evolutionary not revolutionary)
  - Designed for humans first and machines second (user centric design
  - https://support.google.com/webmasters/answer/146897?hl=en

- Drawbacks of Microformats
  - Only a fixed set of Microformats exists
  - No way to connect data elements
  - Fixed vocabulary, not extendable and customizable
  - Separate parsing rules for each microformat needed

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - RDF**S**

- RDFS = RDF Schema

- RDF-schema is an extension of RDF

- RDFs provides a simple and basic modeling language for ontologies
  - Classes
  - Properties
    - *Is-a hierarchy and simple domain and range restrictions*

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - RDF**S**

**RDFS Model**

**RDF Instance**
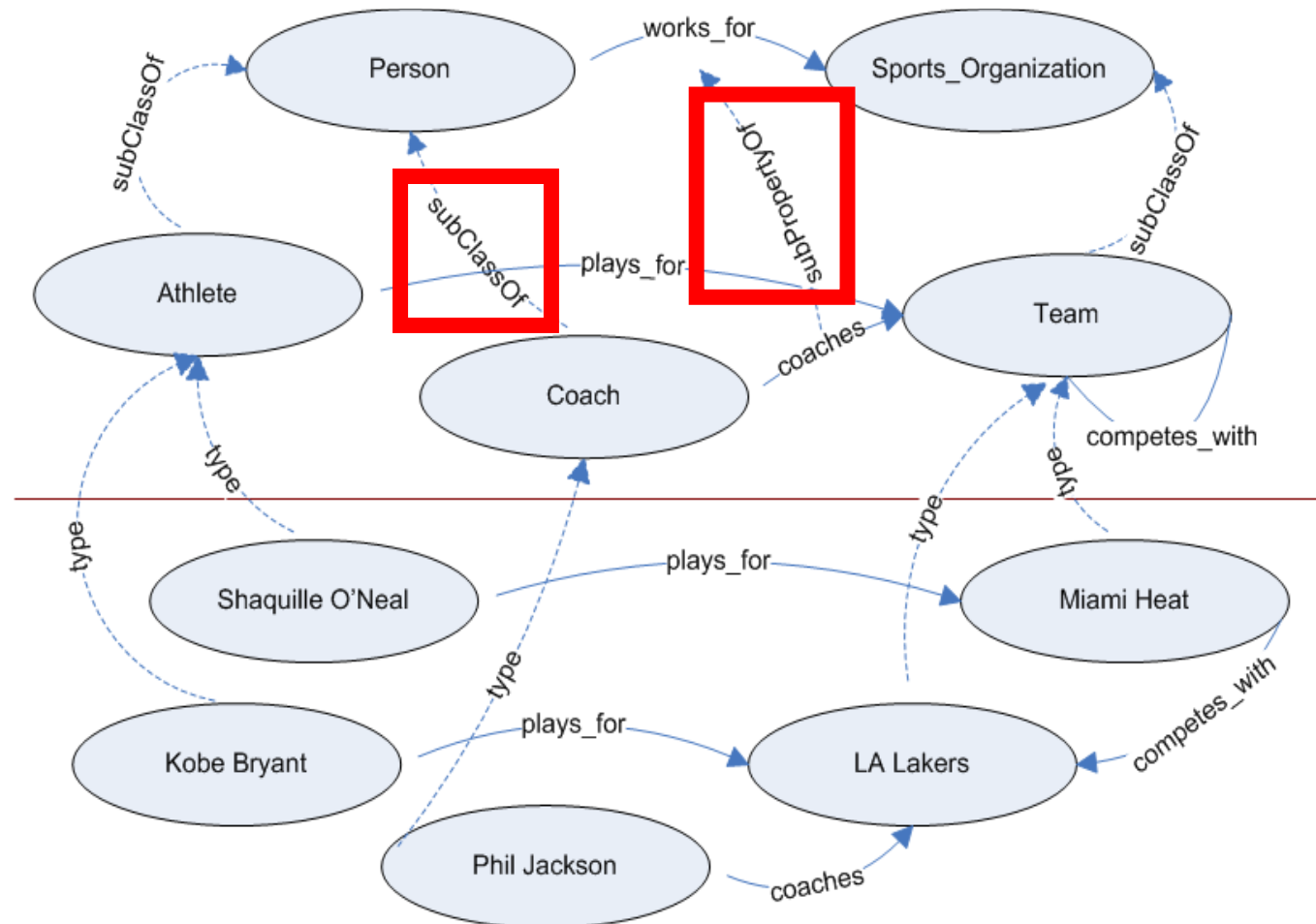
# Web Semantic Technologies - RDF**S**

```xml
<?xml version='1.0' encoding='ISO-8859-1'?> <!DOCTYPE rdf:RDF [
<!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
<!ENTITY example_chp5 'http://protege.stanford.edu/example-chp5#'> <!ENTITY rdfs
'http://www.w3.org/TR/1999/PR-rdf-schema-19990303#'> ]>

<rdf:RDF xmlns:rdf="&rdf;" xmlns:example_chp5="&example_chp5; "xmlns:rdfs="&rdfs;">

  <rdfs:Class rdf:about="&example_chp5;Artifacts" rdfs:label="Artifacts">
    <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
  </rdfs:Class>

  <rdfs:Class rdf:about="&example_chp5;DesignDocument" rdfs:label="DesignDocument">
    <rdfs:subClassOf rdf:resource="&example_chp5;Artifacts"/>
  </rdfs:Class>

  <rdfs:Class rdf:about="&example_chp5;Employee" rdfs:label="Employee">
    <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
  </rdfs:Class>

</rdf:RDF>
```

# Web Semantic Technologies - RDF**S**

```
<rdfs:Class rdf:about="&example_chp5;Software-Engineer" rdfs:label="Software-Engineer">
  <rdfs:subClassOf rdf:resource="&example_chp5;Employee"/>
</rdfs:Class>

<rdf:Property rdf:about="&example_chp5;knows" rdfs:label="knows">
  <rdfs:domain rdf:resource="&example_chp5;Employee"/>
  <rdfs:range rdf:resource="&example_chp5;Topic"/>
</rdf:Property>

<rdf:Property rdf:about="&example_chp5;writes" rdfs:label="writes">
  <rdfs:range rdf:resource="&example_chp5;Artifacts"/>
  <rdfs:domain rdf:resource="&example_chp5;Employee"/>
</rdf:Property> </rdf:RDF>
```
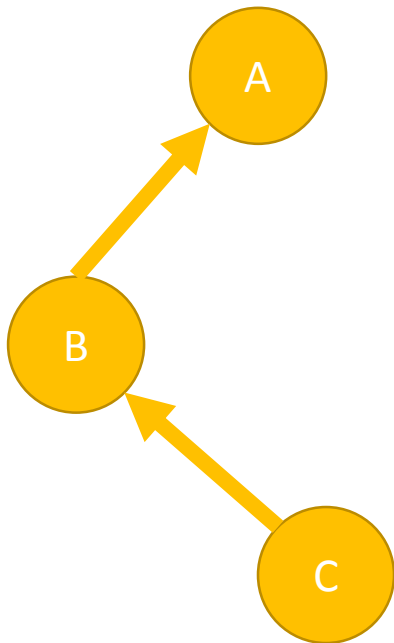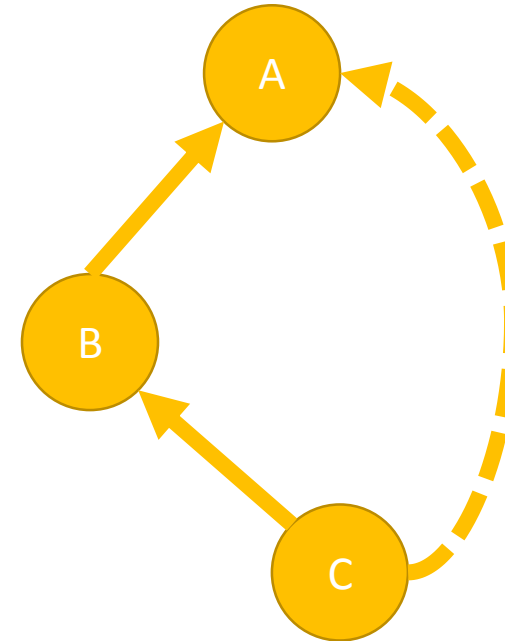
# Web Semantic Technologies - Reasoner



C is SubClassOf A ?

Yes !

*Reasoner will add some triples to the graph !*

SubClassOf      SubClassOf (inferred)

HumanTech
Technology for
Human Wellbeing Institute
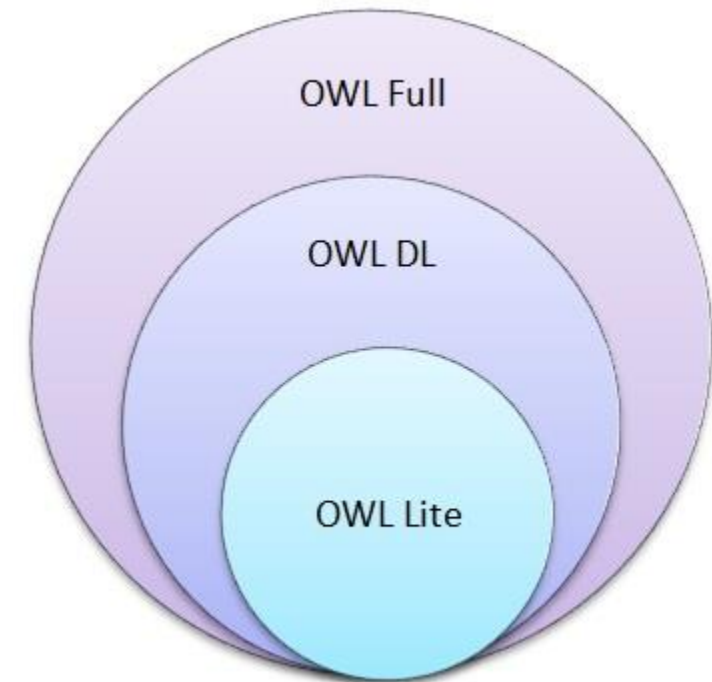
# Web Semantic Technologies - Reasoner

- A semantic reasoned is a piece of software able to infer logical consequences from a set of asserted facts or axioms. The notion of a semantic reasoner generalizes that of an inference engine, by providing a richer set of mechanisms to work with.

- The inference rules are commonly specified on the ontology

# Web Semantic Technologies - OWL

- OWL = Web Ontology language.
- OWL builds on RDF and RDF Schema
  - adds more vocabulary for describing properties and classes
- OWL provides a richer modeling language for ontologies
  - classes
  - properties
  - more complex restrictions and properties (with respect to RDF/RDFS)
    - transitivity property, symmetric property, etc.
- Offers: More accurate web searches, Intelligent agents, Knowledge management

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - OWL

- Classes

- Properties
  - Object properties which relate objects to other objects
    - *e.g. is-taught-by, supervises, etc.*
  - Data properties which relate objects to datatype values
    - *e.g. phone, title, age, etc.*
  - Inverse properties, equivalent properties, property restrictions, etc.

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - OWL

- Transitivity Property
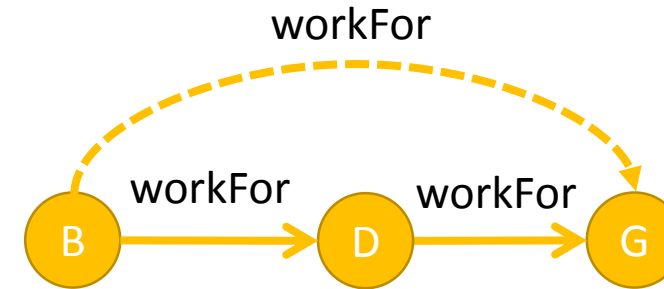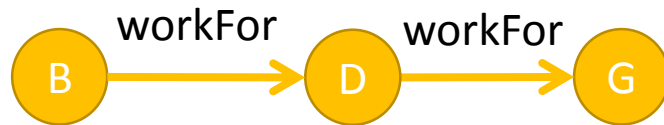
```
<Person rdf:ID='G'/>

<Person rfd:ID='D'>
  <worksFor resource='#G'/>
</Person>

<Person rfd:ID='B'>
  <worksFor resource='#D'/>
</Person>
```



```
<owl:ObjectProperty rdf:ID='worksFor'>
  <rdf:type rdf:resource="&owl;TransitiveProperty" />
  <rdf:range rdf:resource='#Person'/>
</owl:ObjectProperty>

<Person rdf:ID='G'/>

<Person rfd:ID='D'>
  <worksFor resource='#G'/>
</Person>

<Person rfd:ID='B'>
  <worksFor resource='#D'/>
</Person>
```

HumanTech
Technology for
Human Wellbeing Institute

# Web Semantic Technologies - OWL

- ## Symmetric Property
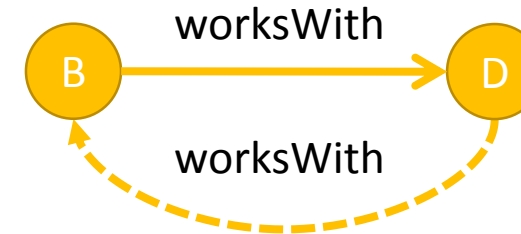


```
<Person rdf:ID='A'>
  <worksWith rdf:ID='#B'>
</Person>
```

```
<owl:ObjectProperty rfd:ID='worksWith>
  <rdf:type rdf:resource="&owl;SymmetricProperty" />
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="#Person" />
</owl: ObjectProperty>

<Person rdf:ID='A'>
  <worksWith rdf:ID='#B'>
</Person>
```

HumanTech
Technology for
Human Wellbeing Institute

# Semantic Web Use

MAS – Web Sémantique

HumanTech
Technology for
Human Wellbeing Institute

# Linked Data - Introduction

- Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful.

- Build upon Web technologies (e.g. HTTP, RDF and URIs)

- Sharing information that can be read automatically by computers

# Linked Data - Principles

- Use URIs to denote things

- Use HTTP URIs so that these things can be referred to and looked up by people […]

- Include links to other related things (using their URIs) when publishing data on the Web

HumanTech
Technology for
Human Wellbeing Institute

# Linked Data - Community

- Linking Open Data community project

- Extend the Web with a data commons

- by publishing various open datasets as RDF on the Web
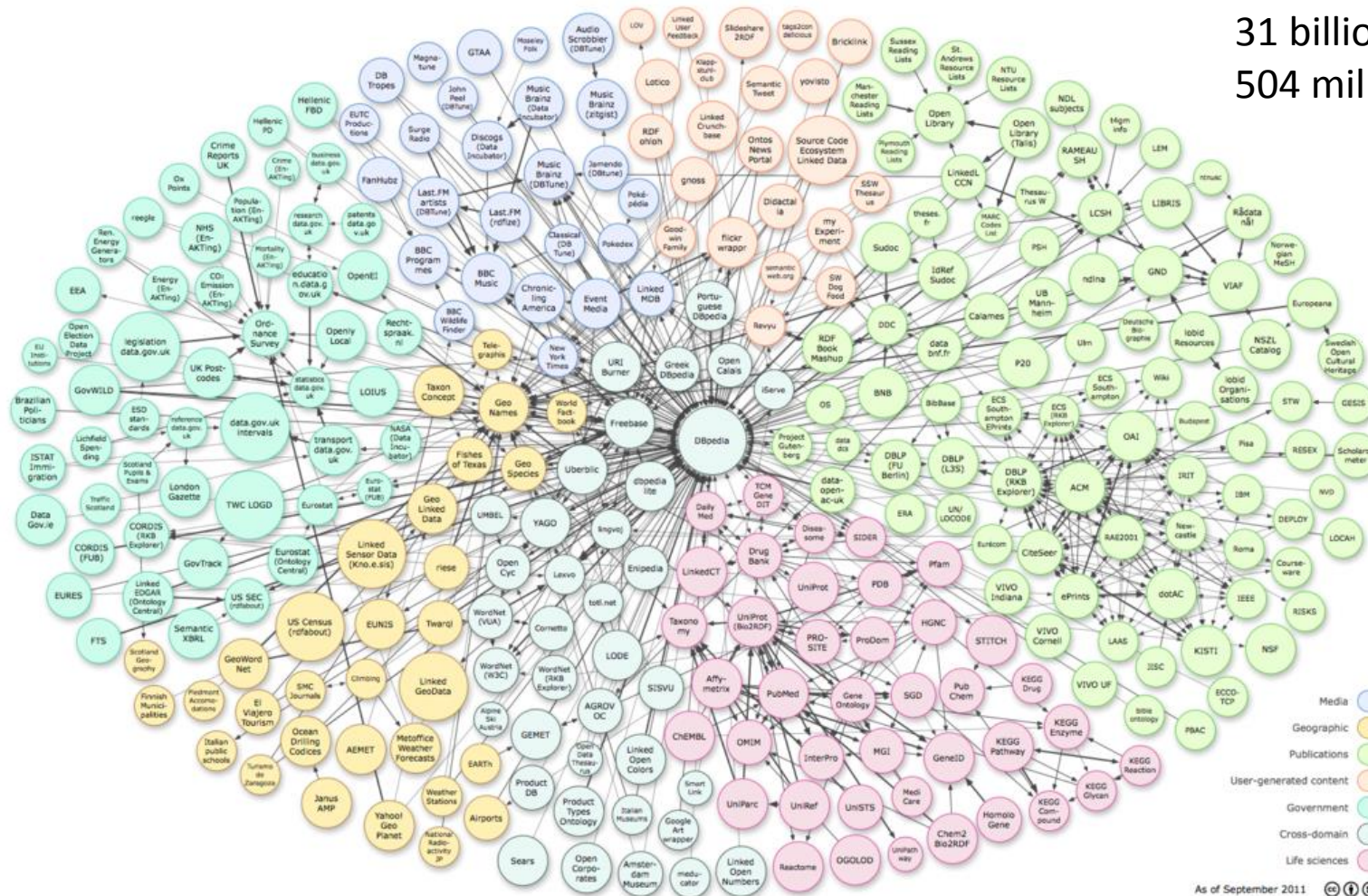- by setting RDF links between data items from different data sources

HumanTech
Technology for
Human Wellbeing Institute

# Linked Data - Community in 2007
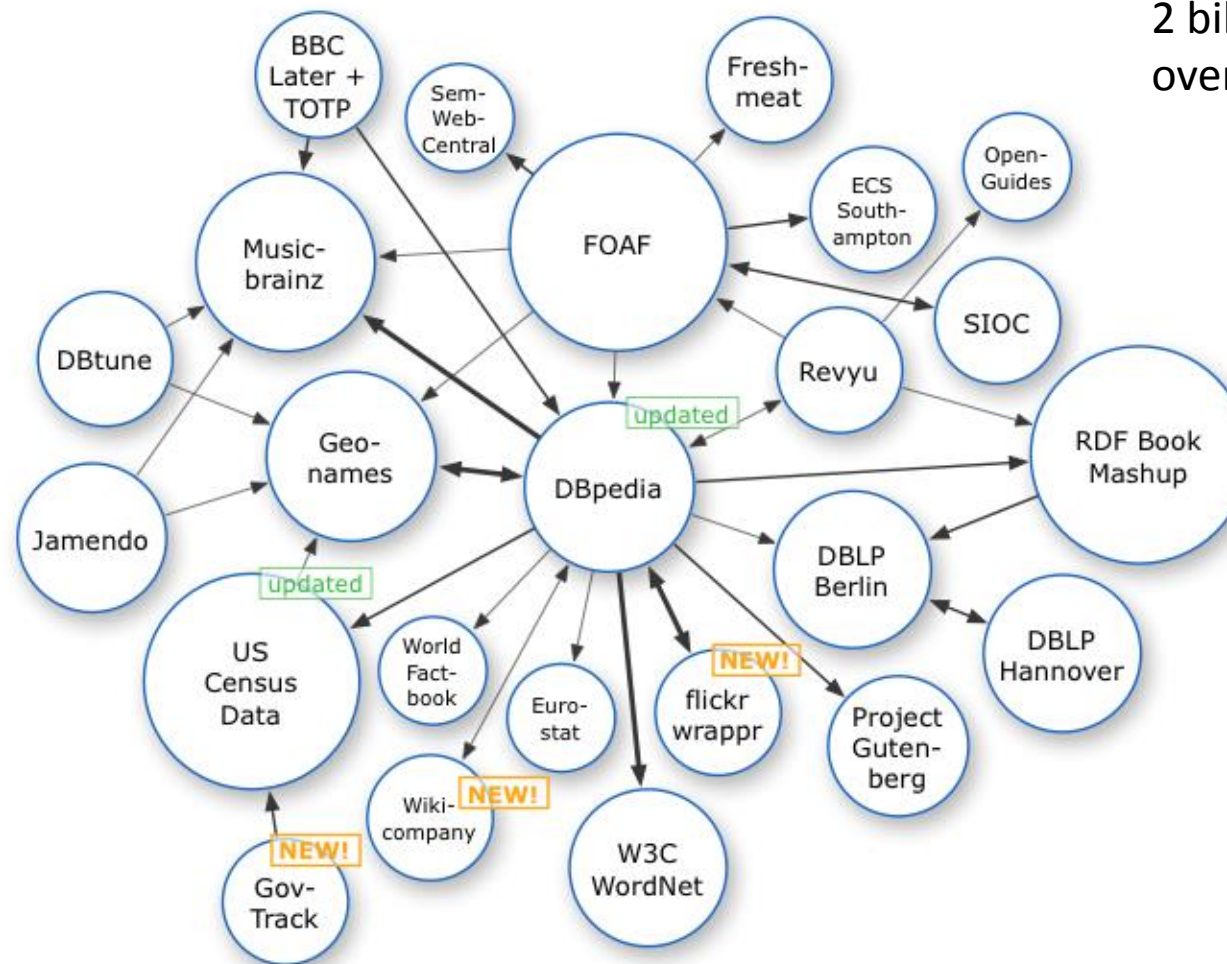
2 billion RDF triples
over 2 million RDF links

HumanTech
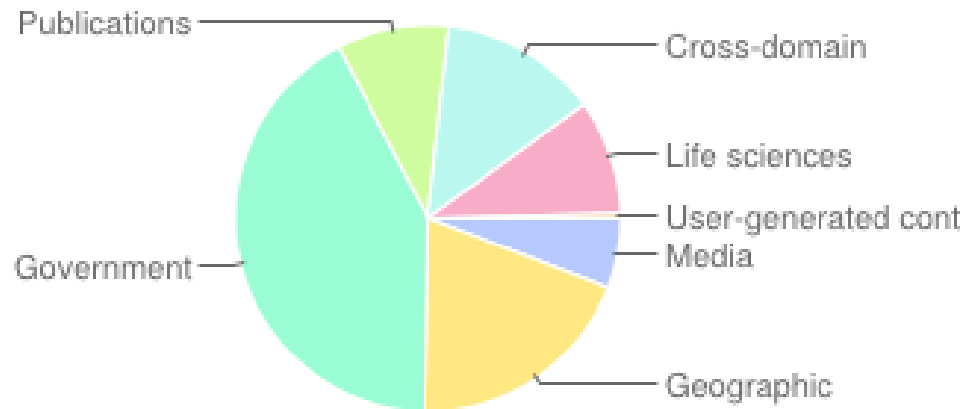Technology for
Human Wellbeing Institute

# Linked Data - Community in 2011

31 billion RDF triples
504 million RDF links

# Linked Data - Community (part)

**The distribution of triples by domain**

**The distribution of links by domain**
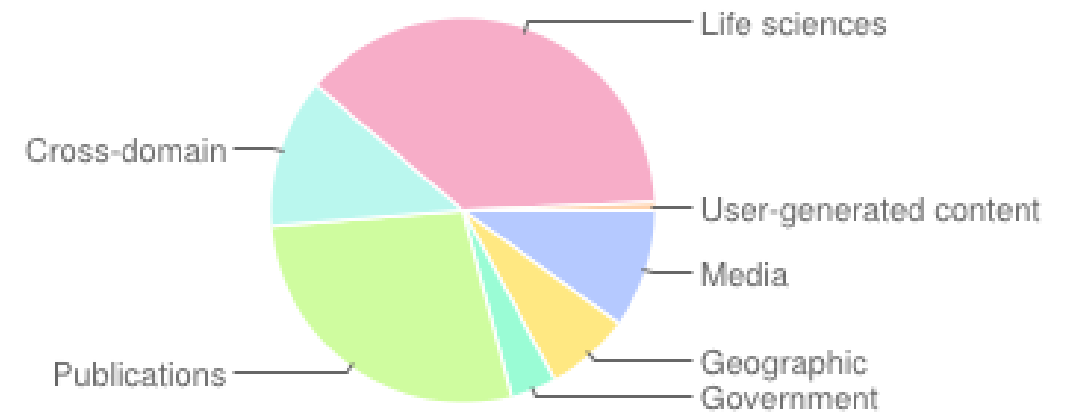


Source: http://lod-cloud.net/state/ (2011)

# Open Data - Open Trends

- The goals of the Open Data movement are similar to those of other "Open" movements

- Open content: is concerned with making resources aimed at a human audience freely available

- Open source: is concerned with the licenses under which computer programs can be distributed and is not normally concerned primarily with data

- Open research: means an approach to open and interconnect scientific assets like data, methods and to enable transparent, reproducible and trans-disciplinary research

# Open Data - Introduction

- Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.

- The concept of open data is not new (but the specifications, yes)

HumanTech
Technology for
Human Wellbeing Institute

# Use Case

MAS – Web Sémantique

# Swiss Parliament

# Conclusion (1)

- The semantic web is based on **machine-processable** semantics of data.

- Its backbone technology are **Ontologies**.

- It is based on new web languages such as **XML, RDF, and OWL** and tools that make use of these languages.

- Ontologies are key enabling technology for the semantic web.

- They interweave human understanding of symbols with their machine processability.

# Conclusion (2)

- Semantic Web step by step:
  - Ontology development in a collaborative and contextual way (RDF, RDFS, OWL)
- Ontology evolution and integration
- Annotation generation
  - Manual, semi-automatic, automatic
- Knowledge retrieval and visualization

HumanTech
Technology for
Human Wellbeing Institute

# Discussion and Questions

- Thank you for your attention

HumanTech
Technology for
Human Wellbeing Institute