

Semantic Web

**Introduction**

# Where are we?



#	Title
1	Introduction
<b>2</b>	<b>Semantic Web Architecture</b>
3	Resource Description Framework (RDF)
4	Web of data
5	Generating Semantic Annotations
6	Storage and Querying
7	Web Ontology Language (OWL)
8	Rule Interchange Format (RIF)
9	Reasoning on the Web
10	Ontologies
11	Social Semantic Web
12	Semantic Web Services
13	Tools
14	Applications

# Agenda

## 1. Motivation

1. Development of the Web
  1. Internet
  2. Web 1.0
  3. Web2.0
2. Limitations of the current Web

## 2. Technical solution

1. Introduction to Semantic Web
2. Semantic Web – architecture and languages
3. Semantic Web - data
4. Semantic Web – processes

## 3. Recent trends

## 4. Summary

## 5. References

# DEVELOPMENT OF THE WEB

# Development of the Web

1. Internet
2. Web 1.0
3. Web 2.0

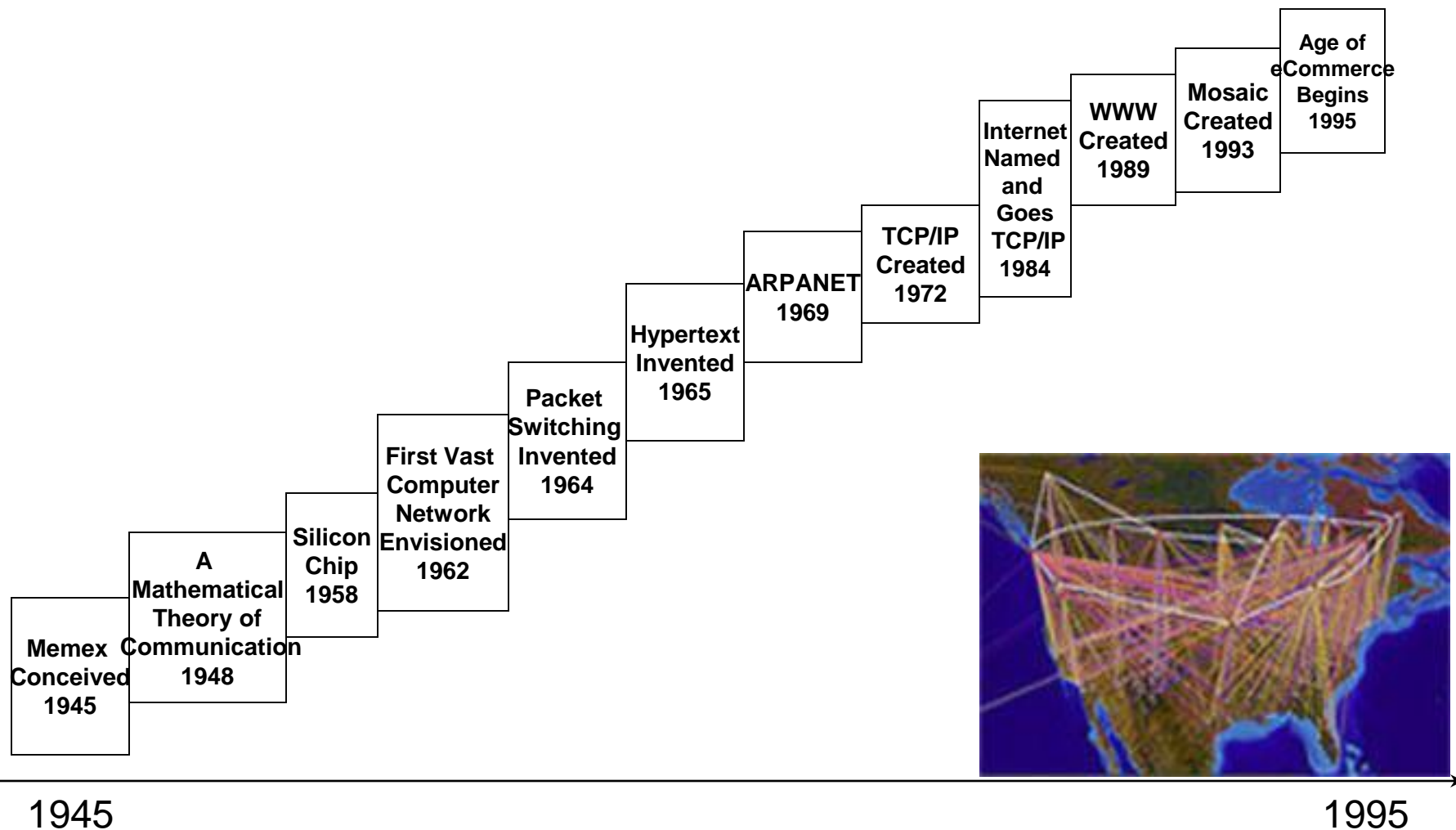
INTERNET

# Internet

- “The **Internet** is a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide. It is a network of networks that consists of millions of private and public, academic, business, and government networks of local to global scope that are linked by a broad array of electronic and optical networking technologies.”

<http://en.wikipedia.org/wiki/Internet>

# A brief summary of Internet evolution



Source:

[http://www.isoc.org/internet/history2002\\_0918\\_Internet\\_History\\_and\\_Growth.ppt](http://www.isoc.org/internet/history2002_0918_Internet_History_and_Growth.ppt)



WEB 1.0

# 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”.

[http://en.wikipedia.org/wiki/World\\_Wide\\_Web](http://en.wikipedia.org/wiki/World_Wide_Web)

# Web 1.0

- **Netscape**

- *Netscape* is associated with the breakthrough of the Web.
- *Netscape* had rapidly a large user community making attractive for others to present their information on the Web.



- **Google**

- *Google* is the incarnation of Web 1.0 mega grows
- *Google* indexed already in 2008 more than 1 trillion pages [\*]
- *Google* and other similar search engines turned out that a piece of information can be faster found again on the Web than in the own bookmark list



[\*] <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>

# Web 1.0 principles

- The success of Web1.0 is based on three simple principles:
  1. A simple and uniform addressing schema to indentify information chunks i.e. ***Uniform Resource Identifiers (URIs)***
  2. A simple and uniform representation formalism to structure information chunks allowing browsers to render them i.e. ***Hyper Text Markup Language (HTML)***
  3. A simple and uniform protocol to access information chunks i.e. ***Hyper Text Transfer Protocol (HTTP)***

# 1. Uniform Resource Identifiers (URIs)

- Uniform Resource Identifiers (URIs) are used to name/identify resources on the Web
- URIs are pointers to resources to which request methods can be applied to generate potentially different responses
- Resource can reside anywhere on the Internet
- Most popular form of a URI is the Uniform Resource Locator (URL)

## 2. Hyper-Text Markup Language (HTML)

- Hyper-Text Markup Language:
  - A subset of Standardized General Markup Language (SGML)
  - Facilitates a hyper-media environment
- Documents use elements to “mark up” or identify sections of text for different purposes or display characteristics
- HTML markup consists of several types of entities, including: elements, attributes, data types and character references
- Markup elements are not seen by the user when page is displayed
- Documents are rendered by browsers

# 3. Hyper-Text Transfer Protocol (HTTP)

- Protocol for client/server communication
  - The heart of the Web
  - Very simple request/response protocol
    - Client sends request message, server replies with response message
  - Provide a way to publish and retrieve HTML pages
  - Stateless
  - Relies on URI naming mechanism

WEB 2.0



# 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”

[http://en.wikipedia.org/wiki/Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0)

# Web 2.0

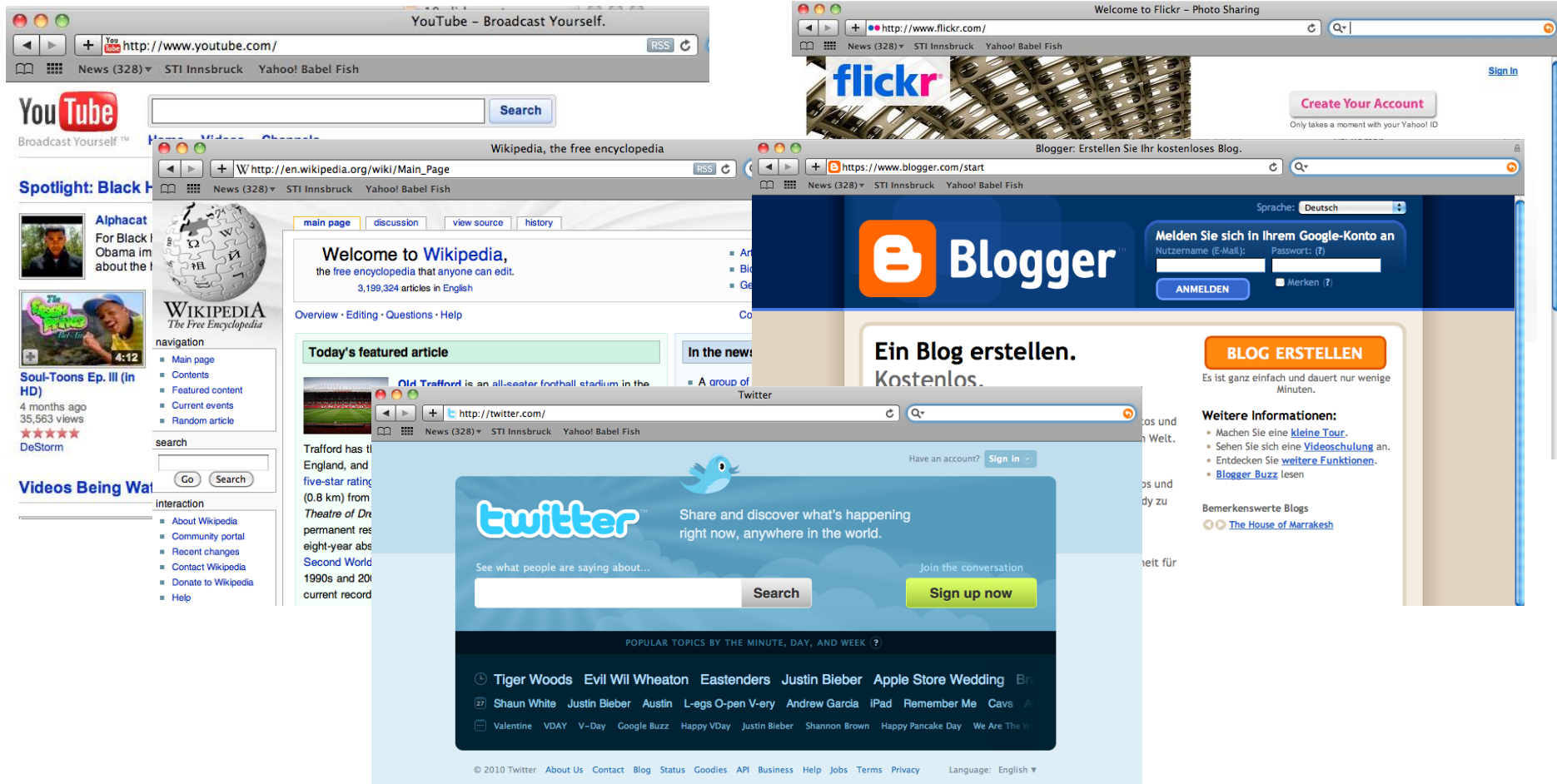
- Web 2.0 is a vaguely defined phrase referring to various topics such as social networking sites, wikis, communication tools, and folksonomies.
- Tim Berners-Lee is right that all these ideas are already underlying his original web ideas, however, there are differences in emphasis that may cause a qualitative change.
- With Web 1.0 technology a significant amount of software skills and investment in software was necessary to publish information.
- Web 2.0 technology changed this dramatically.

# Web 2.0 major breakthroughs

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

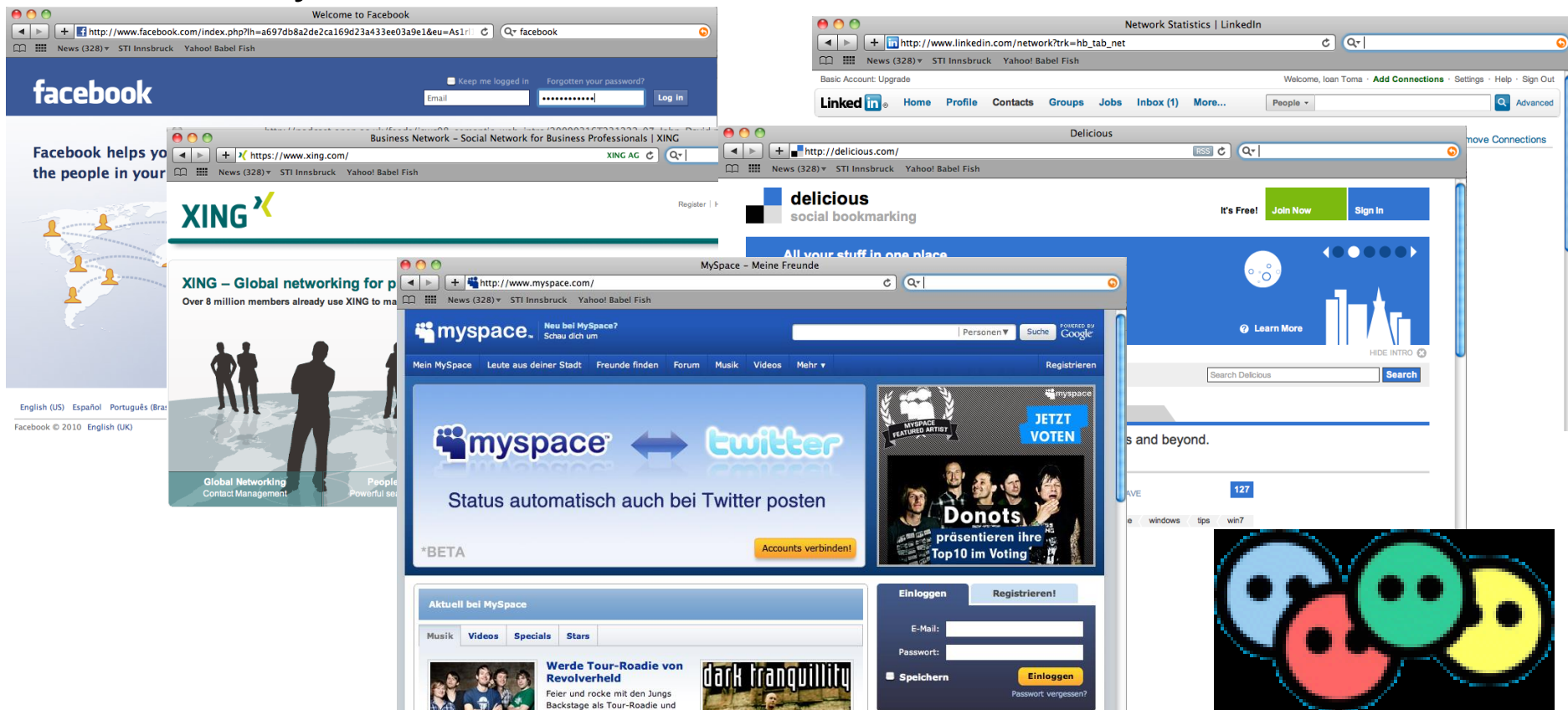
# 1. Blurring the distinction between content consumers and content providers

Wiki, Blogs, and Twiter turned the publication of text in mass phenomena, as flickr and youtube did for multimedia



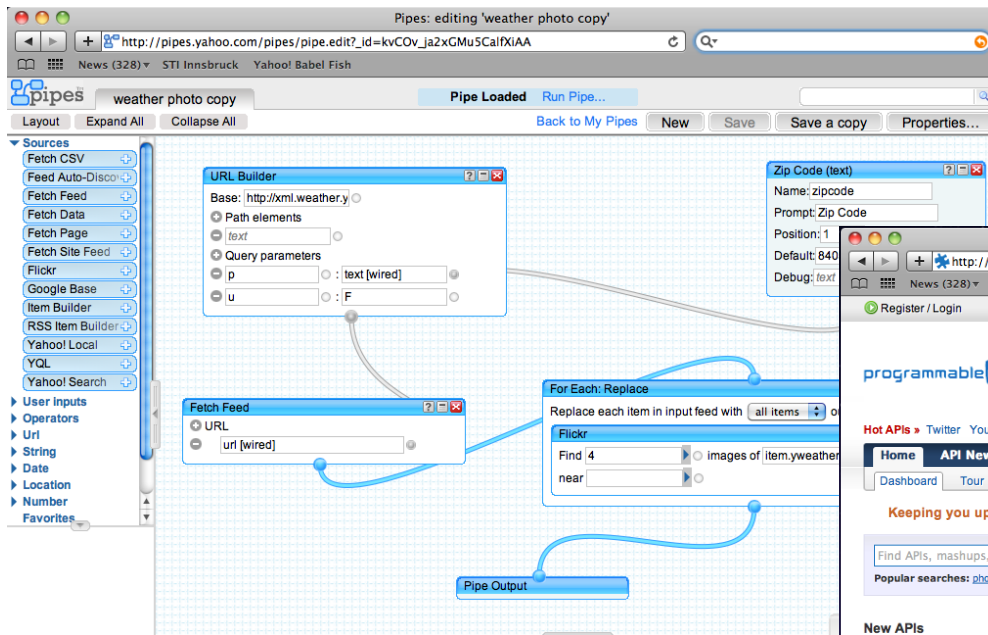
## 2. Moving from a media for individuals towards a media for communities

Social web sites such as del.icio.us, facebook, FOAF, linkedin, myspace and Xing allow communities of users to smoothly interweave their information and activities



# 3. Blurring the distinction between service consumers and service providers

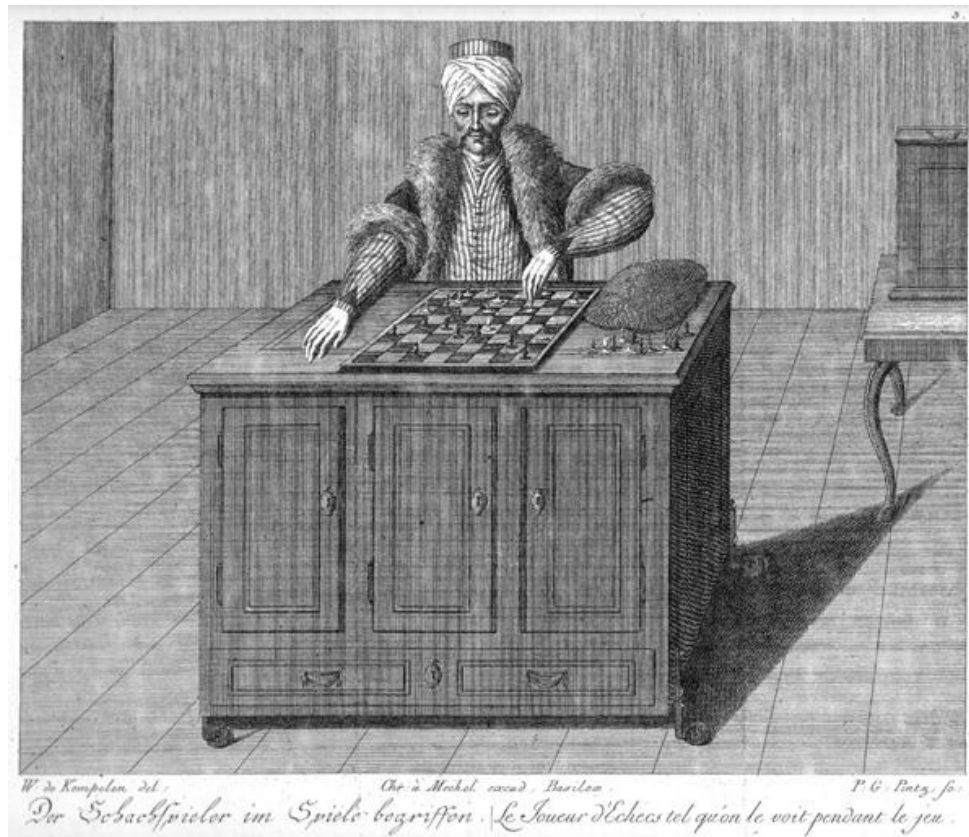
Mashups allow web users to easily integrate services in their web site that were implemented by third parties



The screenshot shows the homepage of ProgrammableWeb, a website dedicated to APIs, mashups, and the web as a platform. The page features a navigation menu with links for Home, API News, API Directory, Mashups, Community, and How-to. A prominent banner for 'apigee' offers 'Free API analytics and protection'. The main content area includes a search bar, statistics showing '1725 APIs' and '4659 Mashups', and sections for 'New APIs', 'Mashup of the Day', and 'New Mashups'. A news article titled 'Microsoft Outlook is Now a Mashup: Uses APIs From LinkedIn, Facebook and MySpace' is featured at the bottom.

## 4. Integrating human and machine computing in a new way

Amazon Mechanical Turk - allows to access human services through a web service interface blurring the distinction between manually and automatically provided services



# LIMITATIONS OF THE CURRENT WEB



# Limitations of the current Web

- The current Web has its limitations when it comes to:
  1. finding relevant information
  2. extracting relevant information
  3. combining and reusing information

# Limitations of the current Web

## *Finding relevant information*

- 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) even though the user is interested only in one of them



# Limitations of the current Web

*Finding relevant information*

- Keyword search has a limited recall and precision due also to:
  - ***Spelling variants:***
    - e.g. “organize” in American English vs. “organise” in British English
  - ***Spelling mistakes***
  - ***Multiple languages***
    - i.e. information about same topics is published on the Web on different languages (English, German, Italian,...)
- Current search engines provide no means to specify the relation between a resource and a term
  - e.g. sell / buy

# Limitations of the current Web

*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

The screenshot shows the Amazon.com search results page for the query 'web'. The browser window title is 'Amazon.com: All Products Search Results: web - Mozilla Firefox'. The search bar contains 'Amazon.com' and 'for web'. The search results are displayed in a list format. The first result is 'Creating Web Pages for Dummies, Sixth Edition' by Bud E. Smith, et al (Paperback). The price is listed as 'Buy new: \$16.49' and 'Used & new from \$14.45'. The second result is 'By Order of the President' by W. E. B. Griffin (Hardcover). The price is listed as 'Buy new: \$16.17' and 'Used & new from \$12.90'. The third result is 'Search Engine Optimization for Dummies' by Peter Kent (Paperback). The price is listed as 'Buy new: \$16.17' and 'Used & new from \$12.90'. Red circles highlight the titles of the first two books, and blue circles highlight their prices. Red arrows point from the text on the right to these highlighted elements.

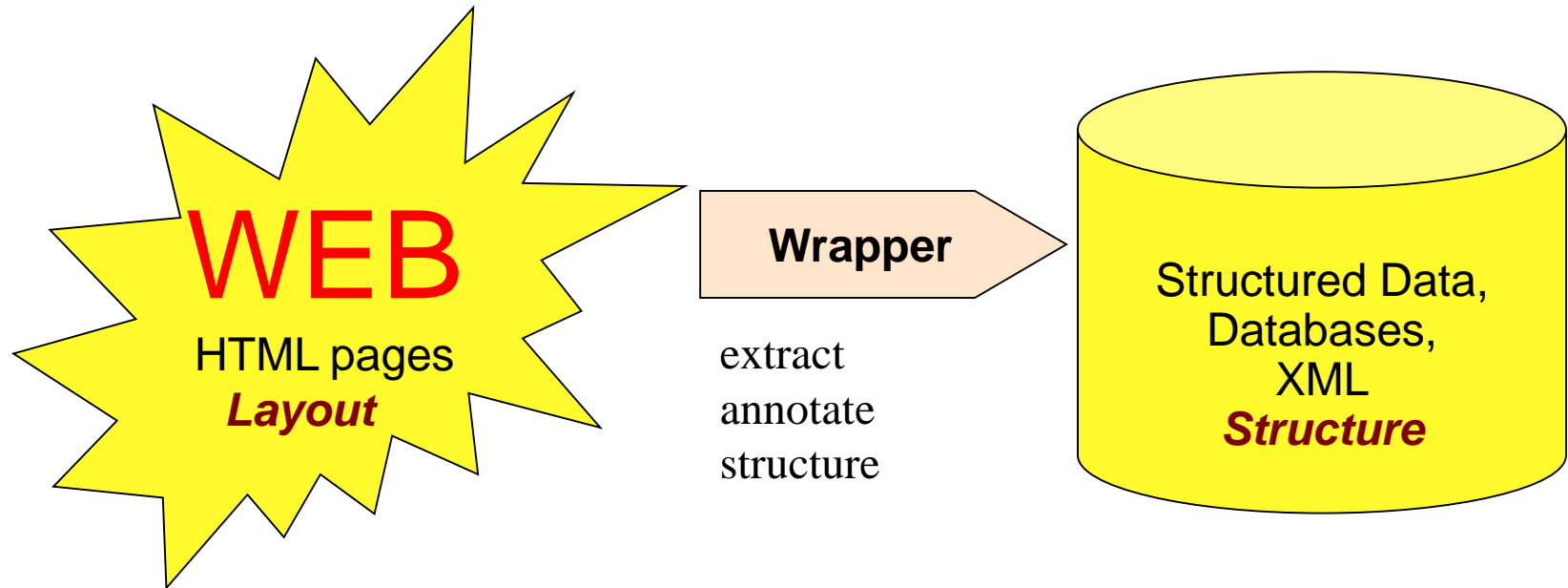
**Which book is about the Web?**

**What is the price of the book?**

# Limitations of the current Web

*Extracting relevant information*

- Extracting information from current web sites can be done using **wrappers**



# Limitations of the current Web

*Extracting relevant information*

- The actual extraction of information from web sites is specified using standards such as XSL Transformation (XSLT) [1]
- Extracted information can be stored as structured data in XML format or databases.
- However, using wrappers do not really scale because the actual extraction of information depends again on the web site format and layout

[1] <http://www.w3.org/TR/xslt>

# Limitations of the current Web

*Combining and reusing information*

- Tasks often require to combine data on the Web
  1. Searching for the same information in different digital libraries
  2. Information may come from different web sites and needs to be combined

# Limitations of the current Web

*Combining and reusing information*

## 1. Searches for the same information in different digital libraries

**Example: I want travel from Innsbruck to Rome.**

The screenshot shows the OBB website interface. At the top, there is a search bar with 'Suche' and a language selector set to 'Deutsch'. Below the search bar, there are navigation links for 'VORTEILScard', 'Online-Ticket', 'CallCenter 05-1747', 'ÖBB-Webshop', and 'Lost&Found'. The main search area has 'Von: innsbruck', 'Nach: rome', 'Datum: 12.02.2010', and 'Zeit: 22:41'. There are buttons for 'Abfahrt' and 'Ankunft', and a 'Verbindung suchen' button. Below the search area, there are several promotional banners, including one for 'SPARSCHIENE EUROPA' for 19€ and another for 'redticket Langstrecke und Europa'. The website also features a 'Country of Residence' dropdown and a 'Suche' button.

The screenshot shows the Lufthansa website interface. At the top, there is a search bar with 'Suche' and a language selector set to 'Deutsch'. Below the search bar, there are navigation links for 'Booking', 'Top Offers', 'Information & Service', and 'Miles & More'. The main search area has 'From: Innsbruck', 'To: Rome', 'Depart on: 13.02.2010', and 'Return on: 21.02.2010'. There are buttons for 'Round-trip' and 'One-way', and a 'Suche' button. Below the search area, there are several promotional banners, including one for 'Europe from 99€' and another for 'Fly to Europe from 99€'.

The screenshot shows the Lufthansa website interface. At the top, there is a search bar with 'Suche' and a language selector set to 'Deutsch'. Below the search bar, there are navigation links for 'Booking', 'Top Offers', 'Information & Service', and 'Miles & More'. The main search area has 'From: Innsbruck', 'To: Rome', 'Depart on: 13.02.2010', and 'Return on: 21.02.2010'. There are buttons for 'Round-trip' and 'One-way', and a 'Suche' button. Below the search area, there are several promotional banners, including one for 'Europe from 99€' and another for 'Fly to Europe from 99€'.



# Limitations of the current Web

*Combining and reusing information*

2. Information may come from different web sites and needs to be combined

**Example: I want to travel from Innsbruck to Rome where I want to stay in a hotel and visit the city**

The screenshot shows the Austrian Airlines website interface. At the top, there is a navigation bar with links for 'Flüge buchen', 'Angebote, etix® & mehr', 'Hotels, Autos & mehr', 'Rund um den Flug', 'Miles & More', 'Austrian Airlines Group', and 'Suche'. Below this is a login section with fields for 'Username' and 'Passwort', and a 'Login' button. A main section titled 'Wie möchten Sie fliegen?' offers options for 'Hin- und Rückflug', 'Nur Hinflug', and 'Mehrere Strecken'. The 'Von:' field is set to 'Innsbruck' and the 'Nach:' field is set to 'Rome'. There are also fields for 'Hinflug:' and 'Rückflug:'. A 'redticket Langstrecke und Europa' banner is visible. At the bottom, the Expedia logo is present, along with a 'NO FEES!' badge and a navigation bar with links for 'Home', 'Vacation Packages', 'Hotels', 'Cars', 'Flights', 'Cruises', 'Activities', 'DEALS & OFFERS', 'Maps', and 'Busi'.

The screenshot shows the Expedia website interface. A prominent banner for 'Lone Star Sale Hotels from \$40' is displayed, with a 'Book now on Expedia' button. Below the banner, there are radio button options for 'Hotel only', 'Hotel + Flight', 'Hotel + Flight + Car', and 'Hotel + Car'. A search field is labeled 'Search near: (e.g. an airport, attraction, address)' and contains the text 'A city'. A 'Destination:' field is set to 'Rome'. The background of the banner shows silhouettes of people riding horses at sunset.

The screenshot shows the AVIS website interface. At the top, there is a navigation bar with links for 'Reservations', 'Locations', 'Deals', 'Cars & Services', and a 'SEARCH' button. Below this is a 'Make a Reservation' section with tabs for 'View / Modify / Cancel', 'Get a Receipt', and 'Car in the Shop?'. The 'DATE & TIME' section shows 'Pick-up' on 02/13/201 at 9:00 AM and 'Return' on 02/14/201 at 9:00 AM. The 'LOCATION' section has a 'Pick-up' field and checkboxes for 'Returning to same location (Uncheck for different location)' and 'Resident of United States (USA)'. An 'OPTIONAL INFO' section includes a 'Wizard Number' field and a note that 'LAST NAME is required ONLY when entering a WIZARD NUMBER'. A red Chevrolet Camaro is featured on the right side of the page.

The screenshot shows the Rome Guide website interface. At the top, there is a navigation bar with links for 'versione in italiano' and 'english version'. Below this is a main section titled 'ROME guide' with a large image of the Roman Forum. A navigation bar at the bottom includes links for 'ART & HISTORY', 'Rome History', 'Monuments >>', 'Museums >>', 'VISITING ROME', 'Hotels in Rome by Map', and 'Rome Hotels | list'. A central image shows the St. Peter's Basilica at night. On the right side, there is a 'FREE TIME' section with links for 'Annual Events', 'Rome Events', 'Music', 'Rome by Night', 'Shopping', 'Roman Cuisine', 'Games', and 'Museums Reservation'.

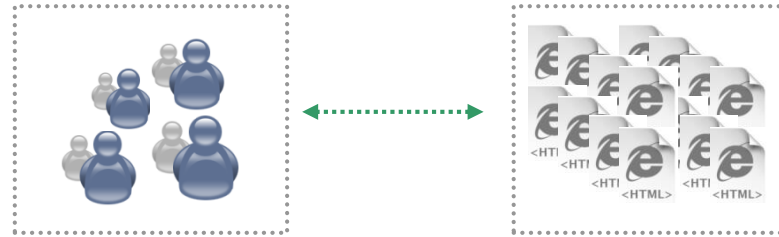
# How to improve current Web?

- 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!

TECHNICAL SOLUTION

# INTRODUCTION TO SEMANTIC WEB

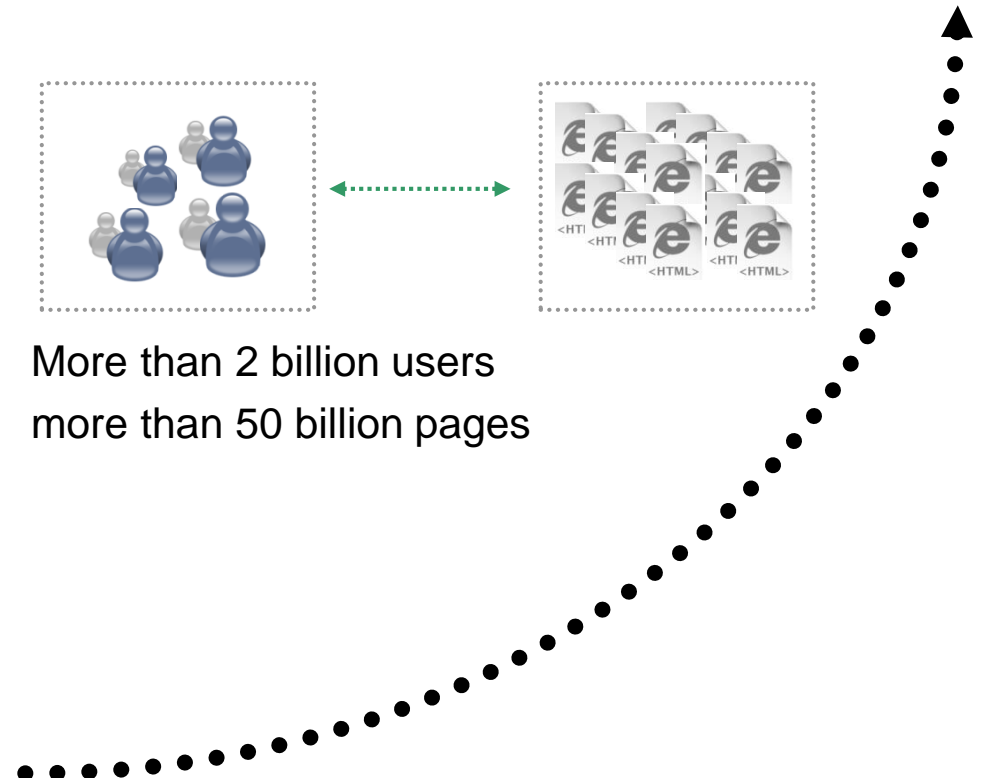
# The Vision



More than 2 billion users  
more than 50 billion pages

Static

**WWW**  
URI, HTML, HTTP



# The Vision (contd.)

## Serious problems in

- information finding,
- information extracting,
- information representing,
- information interpreting and
- and information maintaining.

Static

**WWW**  
URI, HTML, HTTP



**Semantic Web**  
RDF, RDF(S), OWL

# 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.**”*

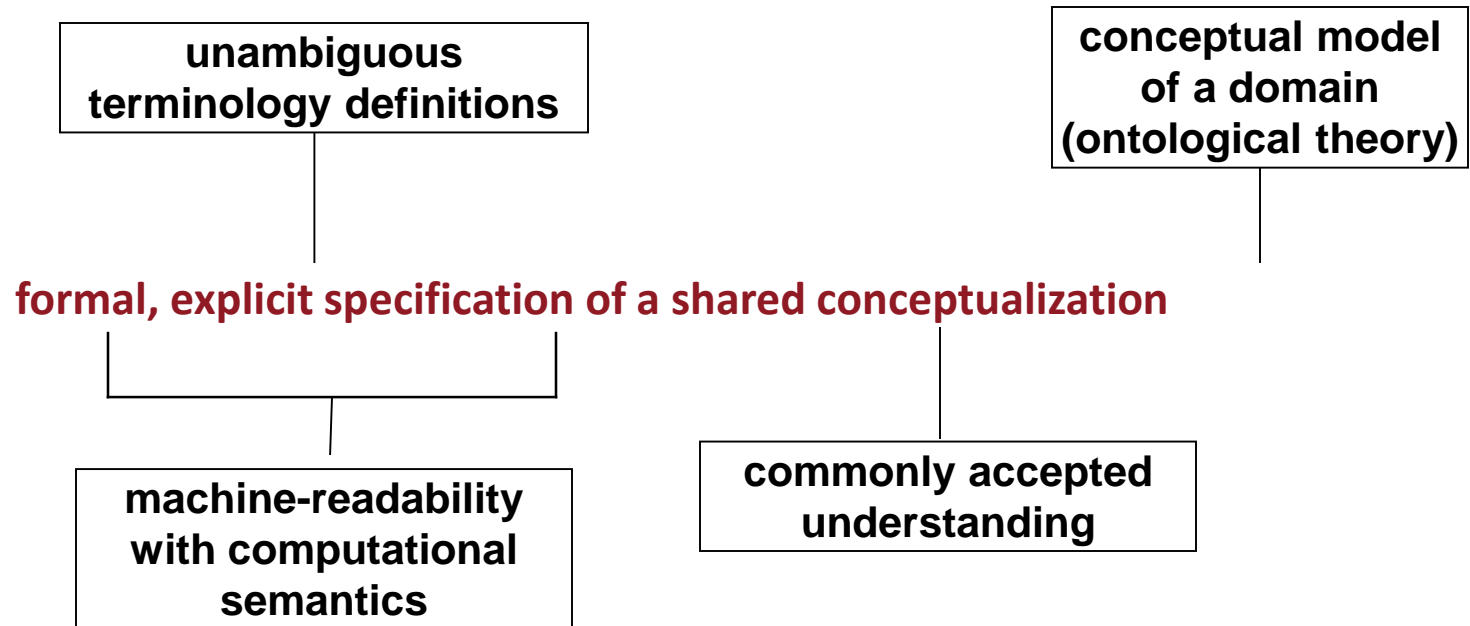
T. Berners-Lee, J. Hendler, O. Lassila, “The Semantic Web”, Scientific American, May 2001

## What is the Semantic Web?

- 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***



# 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

... “*well-defined meaning*” ...

- “An ontology is an **explicit specification** of a **conceptualization**”  
Gruber, “Toward principles for the design of ontologies used for 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

*... explicit, ... specification, ...  
conceptualization, ...*

An 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 (the use of the natural language is excluded)
- 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.

## Ontology example

### Concept

conceptual entity of the domain

### Property

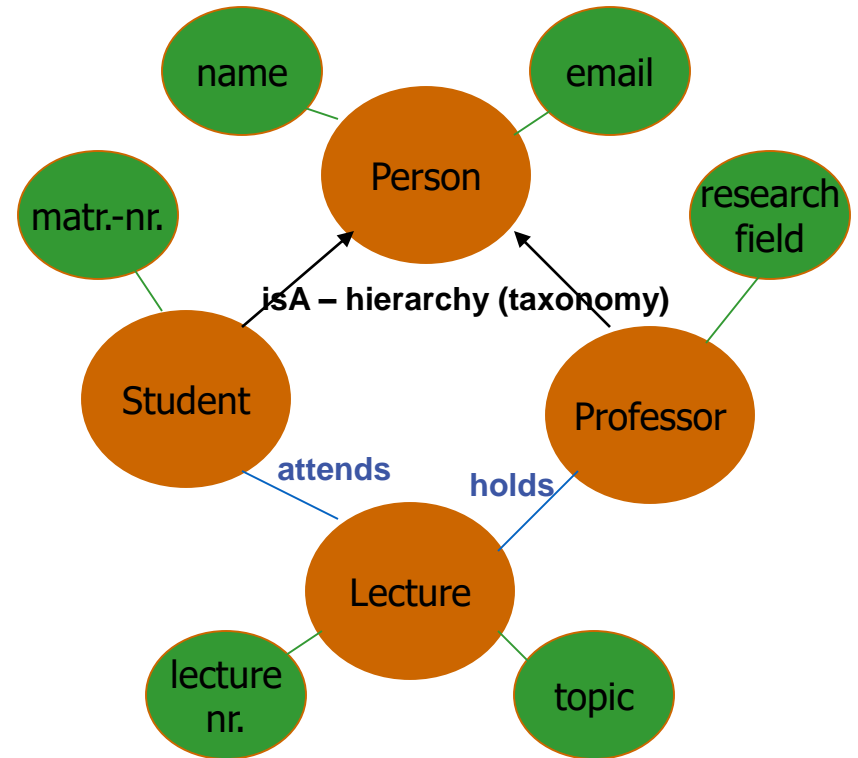
attribute describing a concept

### Relation

relationship between concepts or properties

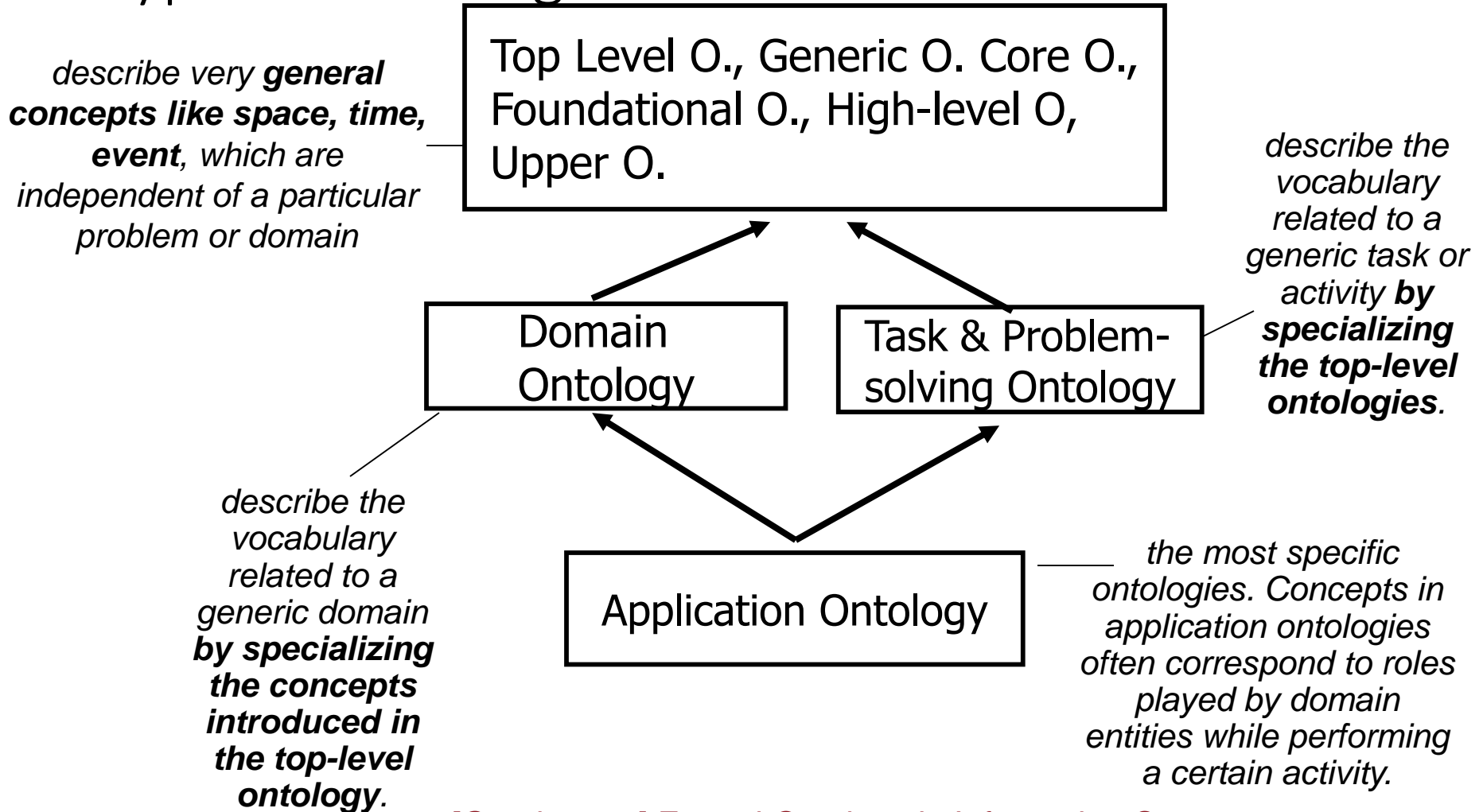
### Axiom

coherency description between Concepts / Properties / Relations via logical expressions



$\text{holds}(\text{Professor}, \text{Lecture}) \Rightarrow$   
 $\text{Lecture.topic} = \text{Professor.researchField}$

# Types of ontologies



[Guarino, 98] Formal Ontology in Information Systems

<http://www.loa-cnr.it/Papers/FOIS98.pdf>

# The 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)

# Web Data Annotating

article discussion **annotate** edit history protect delete move watch refresh

## Annotating Hydrogen

With an atomic mass of 1.000794 g/mole (unit), hydrogen is the lightest element. It is also the most abundant, constituting roughly 75% of the universe's elemental mass. Hydrogen in the Universe, NASA Website. URL accessed on 2 June 2006. Stars in the main sequence are mainly composed of hydrogen in its plasma state. Elemental hydrogen is relatively rare on Earth, and is industrially produced from hydrocarbons, after which most free hydrogen is used "captively" (meaning locally at the production site), with the largest markets about equally divided between fossil fuel upgrading (e.g., hydrocracking) and in ammonia production (mostly for the fertilizer market). However, hydrogen can easily be produced from water using the process of electrolysis.

The most common naturally occurring isotope of hydrogen has a single proton and no neutrons. In ionic compounds it can take on either a positive charge (becoming a cation composed of a bare proton) or a negative charge (becoming an anion known as a hydride). Hydrogen can form compounds with most elements and is present in water and most organic compounds. It plays a particularly important role in acid-base chemistry, in which hydrogen bonds between soluble molecules. As the only neutral atom for which the quantum mechanics of the atom has been studied, it is a subject of theoretical and experimental study.

**Specify this property.**  
Property:   
Page:

### History

#### Discovery of H<sub>2</sub>

Hydrogen gas, H<sub>2</sub>, was first artificially produced and formally described by Paracelsus (also known as T. von Hohenheim, 1493–1541) via the mixing of metals with strong acids. He was unaware that the flammable gas produced by this chemical reaction was a new chemical element. In 1671, Robert Boyle rediscovered and described the reaction between iron filings and dilute acids, which results in the production of hydrogen gas. In 1766, Henry Cavendish was the first to recognize hydrogen gas as a discrete substance, by identifying the gas from a metal-acid reaction as "inflammable air", and further finding that the gas produces water when burned. Cavendish had stumbled on hydrogen when experimenting with acids and mercury. Although he wrongly assumed that hydrogen was a liberated component of the mercury rather than the acid, he was still able to accurately describe several key properties of hydrogen. He is usually given credit for its discovery as an element. In 1783, Antoine Lavoisier gave the element the name of hydrogen when he (with Laplace) reproduced Cavendish's finding that water is produced when hydrogen is burned. Lavoisier's name for the gas won out.

One of the first uses of H<sub>2</sub> was for balloons. The H<sub>2</sub> was obtained by reacting sulphuric acid and metallic iron. Infamously, H<sub>2</sub> was used in the Hindenburg airship that was destroyed in a midair fire.

#### Role in history of quantum theory

Because of its relatively simple atomic structure, consisting only of a proton and an electron, the hydrogen atom,

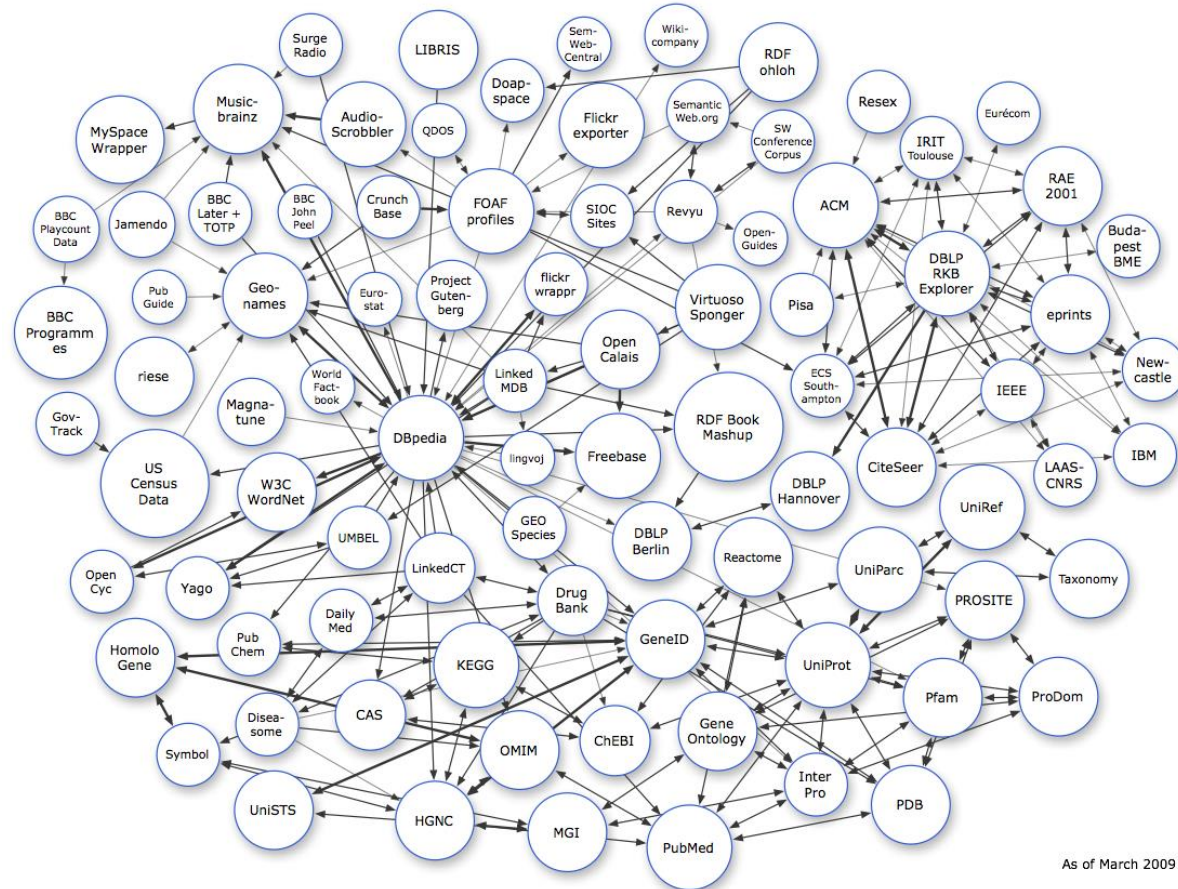
**Annotations & Help**

- Categories**
  - Annotate
  - Chemical elements
  - Nonmetal
- Properties**
  - Create | Has part
  - Can be produced by: hydrocarbon
  - Was discovered by: Paracelsus
  - was first synthesized: Paracelsus
  - Group, F: by
  - Is used as: an ingredient in some rocket fuels
- Help**
  - How can I annotate a part in the AAM?
  - How can I change an annotated category?
  - How can I change an annotated property?
  - What is the context sensitive auto completion?
  - How can I annotate or create a category?
- Annotation hints**
  - No hints for this article.
- Don't forget to save your work!**
  -

Critical point	25.35 K, 12.063 MPa					
Heat of fusion	0.117 kJ·mol <sup>-1</sup>					
Heat of vaporization	0.904 kJ/mol					
Heat capacity	(25 °C) 28.836 J·mol <sup>-1</sup> ·K <sup>-1</sup>					
Vapor pressure						
<i>P</i> (Pa)	1	10	100	1 k	10 k	100 k
at <i>T</i> (K)				15	20	

<http://www.ontoprise.de/>

# LOD Cloud March 2009



As of March 2009

Linked Data, <http://linkeddata.org/> (last accessed on 18.03.2009)



# Data Linking on the Web

- **Linked Open Data statistics:**

- **data sets: 121**
- **total number of triples: 13.112.409.691**
- **total number of links between data sets: 142.605.717**

- **Statistics available at (last accessed on 04.02.2010):**

- <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/Statistics>
- <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/LinkStatistics>

# Data linking on the Web principles

- Use URIs as names for things
  - anything, not just documents
  - you are not your homepage
  - information resources and non-information resources
- Use HTTP URIs
  - globally unique names, distributed ownership
  - allows people to look up those names
- Provide useful information in RDF
  - when someone looks up a URI
- Include RDF links to other URIs
  - to enable discovery of related information

# DBpedia



- DBpedia is a community effort to:
  - Extract structured information from Wikipedia
  - Make the information available on the Web under an open license
  - Interlink the DBpedia dataset with other open datasets on the Web
- DBpedia is one of the central interlinking-hubs of the emerging Web of Data

# The DBpedia Dataset

- 91 languages
- Data about 2.9 million “things”. Includes for example:
  - 282.000 persons
  - 339.000 places
  - 119.00 organizations
  - 130.000 species
  - 88.000 music albums
  - 44.000 films
  - 19.000 books
- Altogether 479 million pieces of information (RDF triples)
  - 807.000 links to images
  - 3.840.000 links to external web pages
  - 4.878.100 data links into external RDF datasets

# LinkedCT

- LinkedCT is the Linked Data version of ClinicalTrials.org containing data about clinical trials.
- Total number of triples:  
6,998,851
- Number of Trials:  
61,920
- RDF links to other data sources:  
177,975
- Links to other datasets:
  - DBpedia and YAGO(from intervention and conditions)
  - GeoNames (from locations)
  - Bio2RDF.org's PubMed (from references)

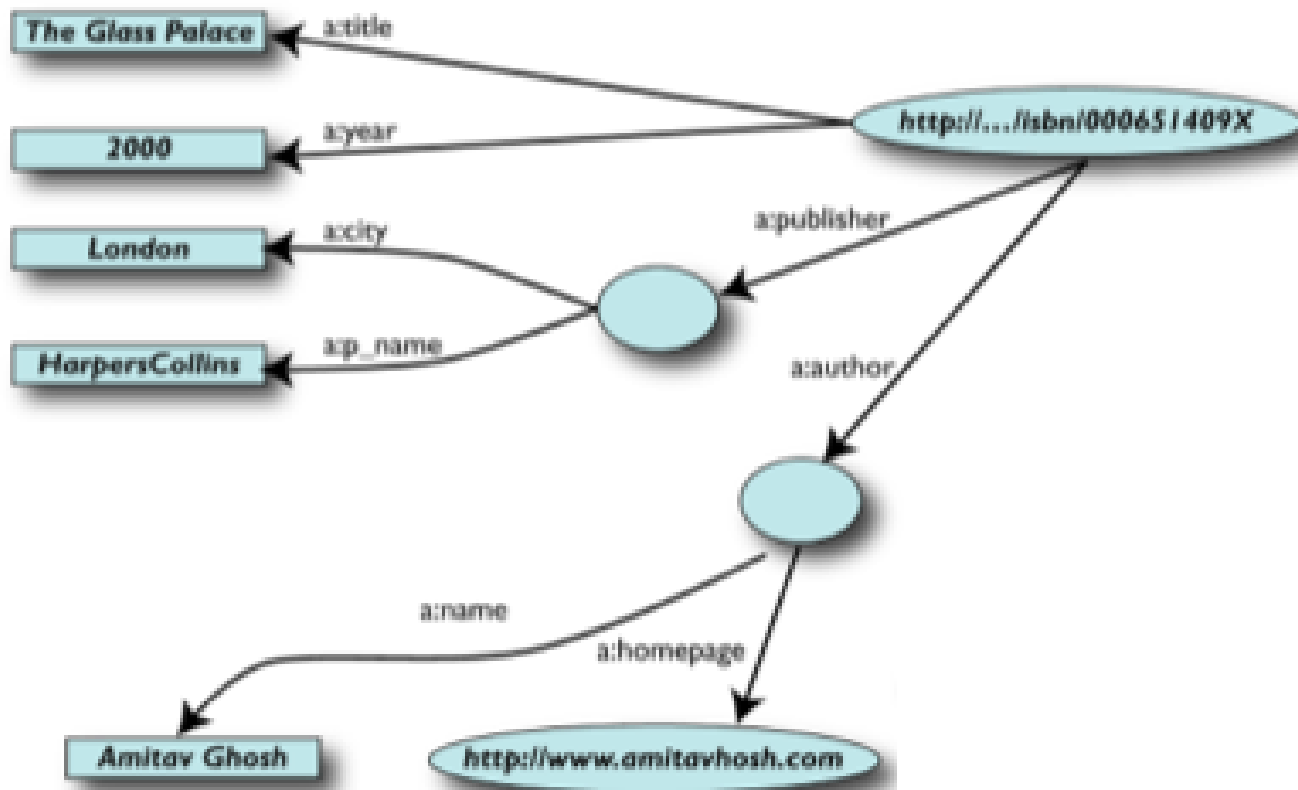
# Data integration over the Web

- Data integration involves combining data residing in different sources and providing user with a unified view of these data
- Data integration over the Web can be implemented as follows:
  - 1. Export the data sets to be integrated as RDF graphs**
  - 2. Merge identical resources (i.e. resources having the same URI) from different data sets**
  - 3. Start making queries on the integrated data, queries that were not possible on the individual data sets.**

# Data integration over the Web

## 1. Export first data set as RDF graph

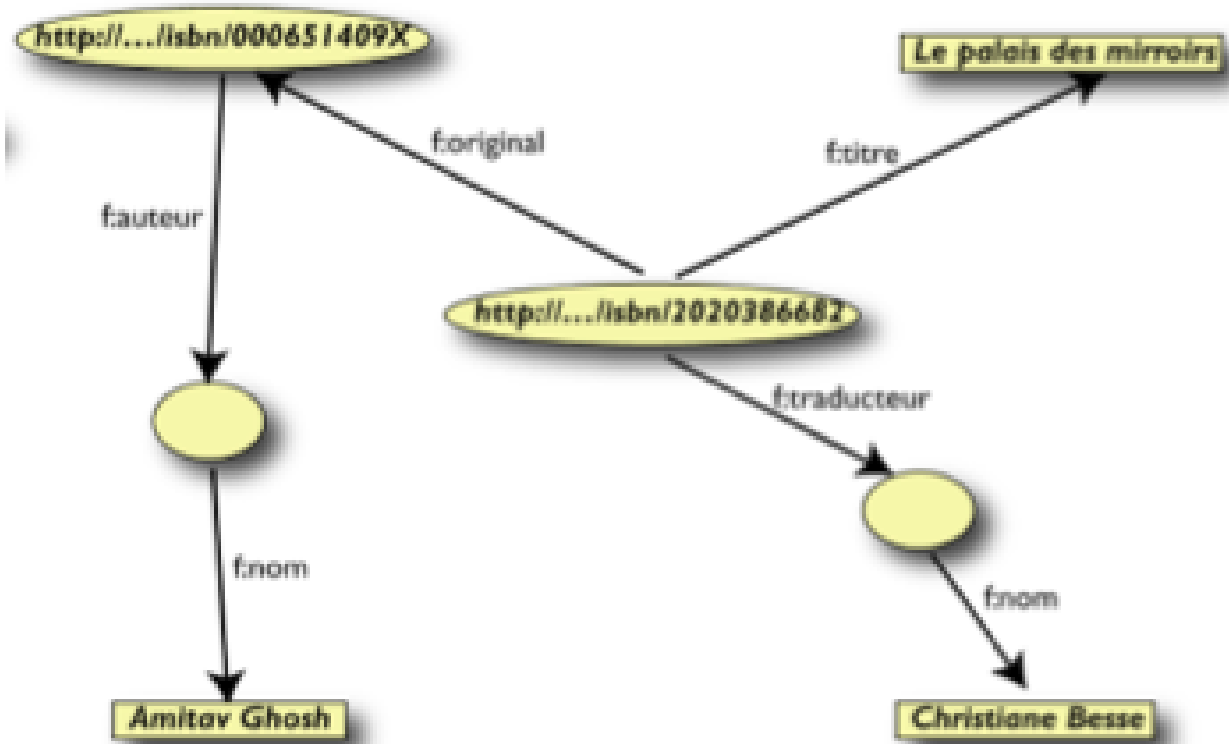
For example the following RDF graph contains information about book “The Glass Palace” by Amitav Ghosh



# Data integration over the Web

## 1. Export second data set as RDF graph

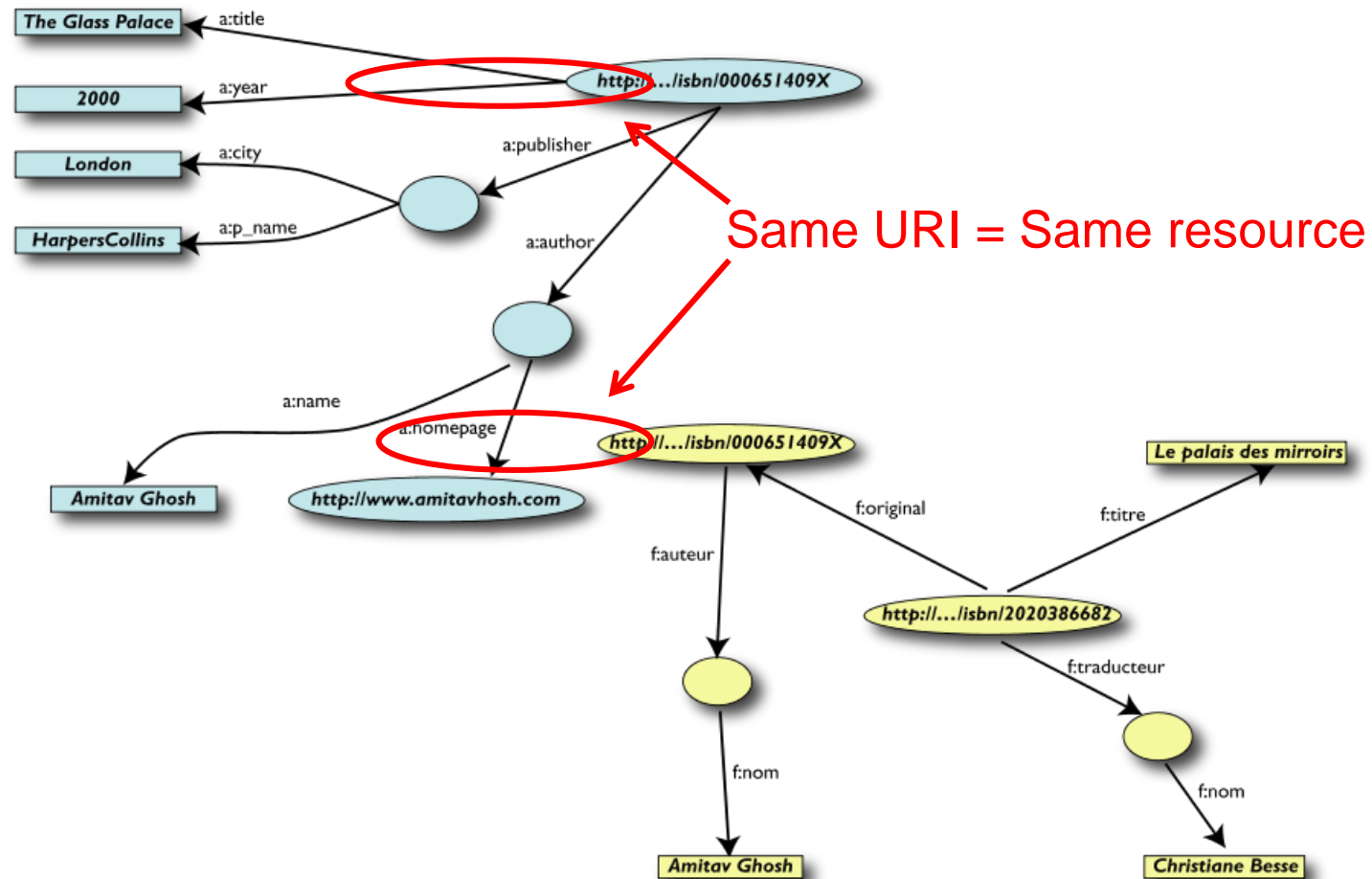
Information about the same book but in French this time is modeled in RDF graph below





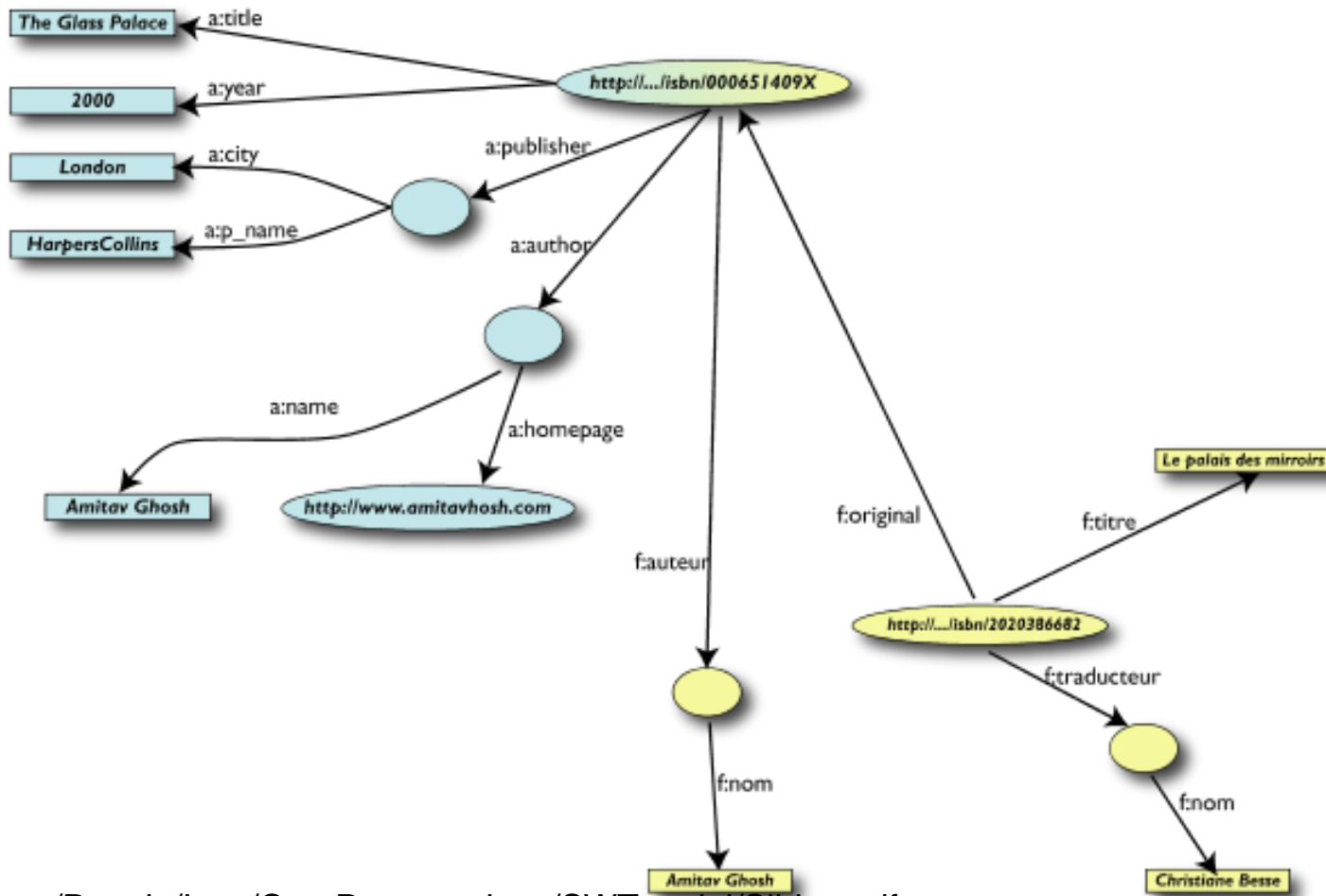
# Data Integration over the Web

## 2. Merge identical resources (i.e. resources having the same URI) from different data sets



# Data integration over the Web

## 2. Merge identical resources (i.e. resources having the same URI) from different data sets



# Data integration over the Web

## 3. Start making queries on the integrated data

- A user of the second dataset may ask queries like: *“give me the title of the original book”*
- This information is not in the second dataset
- This information can be however retrieved from the integrated dataset, in which the second dataset was connected with the the first dataset

# SEMANTIC WEB – ARCHITECTURE AND LANGUAGES

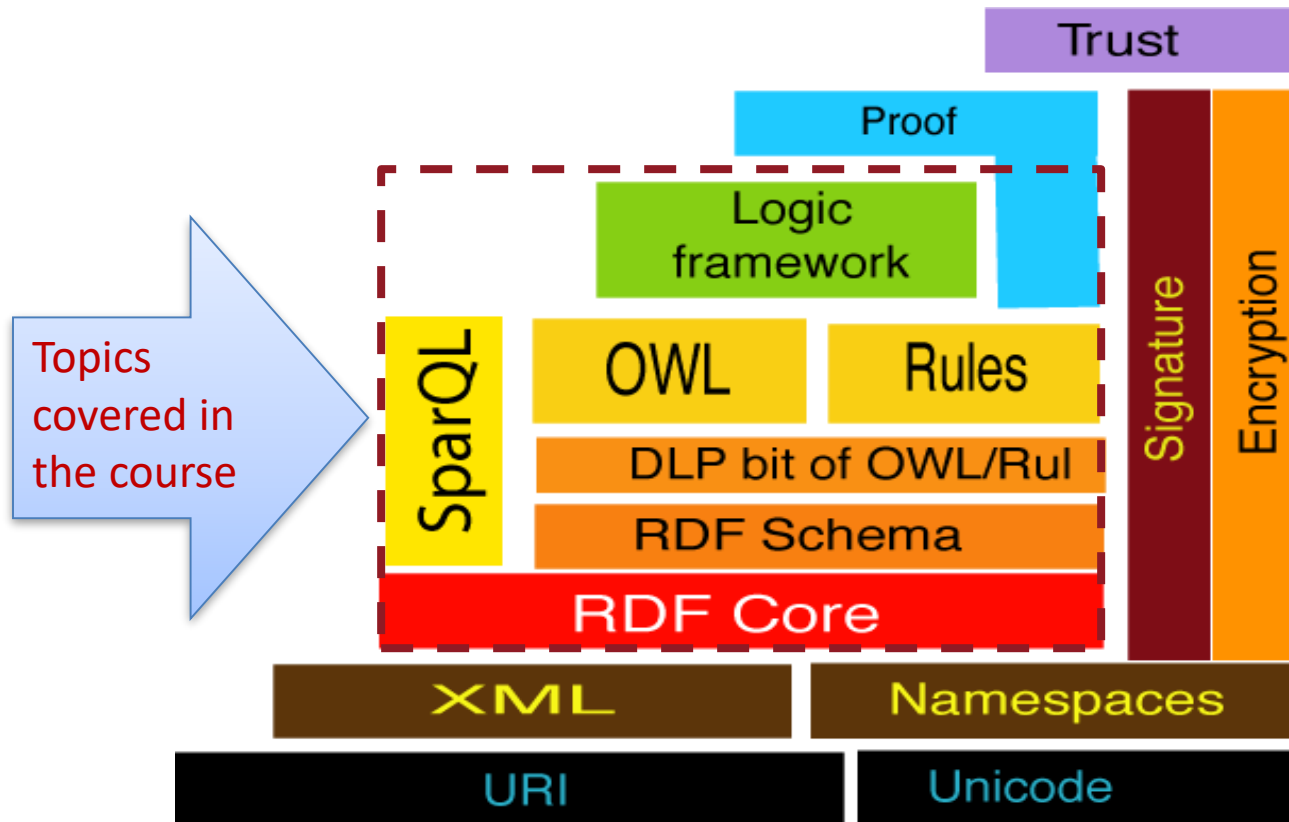
# Web Architecture

- Things are denoted by URIs
- Use them to denote things
- Serve useful information at them
- Dereference them

# Semantic Web Architecture

- Give important concepts URIs
- Each URI identifies one concept
- Share these symbols between many languages
- Support URI lookup

# Semantic Web - Data



# URI and XML

- Uniform Resource Identifier (URI) is the dual of URL on Semantic Web
  - it's purpose is to indentify resources
- eXtensible Markup Language (XML) is a markup language used to structure information
  - fundament of data representation on the Semantic Web
  - tags do not convey semantic information



# RDF and OWL

- **Resource Description Framework (RDF) is the dual of HTML in the Semantic Web**
  - simple way to describe resources on the Web
  - sort of simple ontology language (RDF-S)
  - based on triples (subject; predicate; object)
  - serialization is XML based
- **Ontology Web Language (OWL) a layered language based on DL**
  - more complex ontology language
  - overcome some RDF(S) limitations

# SPARQL and Rule languages

- SPARQL
  - Query language for RDF triples
  - A protocol for querying RDF data over the Web
- Rule languages (e.g. SWRL)
  - Extend basic predicates in ontology languages with proprietary predicates
  - Based on different logics
    - Description Logic
    - Logic Programming

SEMANTIC WEB - DATA

# Semantic Web - Data

- URIs are used to identify **resources**, not just things that exists on the Web, e.g. *Sir Tim Berners-Lee*
- RDF is used to make statements about resources in the form of **triples**  
*<entity, property, value>*
- With RDFS, resources can belong to **classes** (*my Mercedes belongs to the class of cars*) and classes can be **subclasses** or **superclasses** of other classes (*vehicles are a superclass of cars, cabriolets are a subclass of cars*)




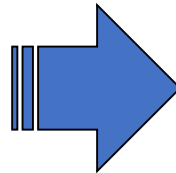
# Semantic Web - Data

Dereferencable  
URI

Disco - Hyperdata Browser (About)  
Tim Berners-Lee

URI:

Property	Value	Sources
type	<a href="http://www.w3.org/2000/10/swap/pim/contact#Male">http://www.w3.org/2000/10/swap/pim/contact#Male</a>	G1 G5
type	Person	G1 G2 G4
value	Tim Berners-Lee	G3
label	Tim Berners-Lee	G1 G5
seeAlso	<a href="http://www.w3.org/People/Berners-Lee/card">http://www.w3.org/People/Berners-Lee/card</a>	G3
seeAlso	<a href="http://www.w3.org/People/Berners-Lee/card.rdf">http://www.w3.org/People/Berners-Lee/card.rdf</a>	G4
assistant	<a href="#">Amy van der Hiel</a>	G1 G5
homePage	<a href="http://www.w3.org/People/Berners-Lee/card.rdf">http://www.w3.org/People/Berners-Lee/card.rdf</a>	G1
work	...	G1
organization	<a href="#">W3C</a>	G1
publicHomePage	<a href="http://www.w3.org/People/Berners-Lee/card.rdf">http://www.w3.org/People/Berners-Lee/card.rdf</a>	G1
sameAs	<a href="#">Tim Berners-Lee</a>	G1
sameAs	<a href="#">Tim Berners-Lee</a>	G1
based near	...	G1
based near	...	G5
family_name	Berners-Lee	G1 G5
Given name	Timothy	G1
homepage	<a href="http://www.w3.org/People/Berners-Lee/">http://www.w3.org/People/Berners-Lee/</a>	G1 G2
image		G1 G5
knows	<a href="#">Coralie Mercier</a>	G1
knows	<a href="#">Edd Dumbill</a>	G5
knows	<a href="#">Henry Story</a>	G1
knows	<a href="http://danbri.org/foaf#danbri">http://danbri.org/foaf#danbri</a>	G1
knows	<a href="#">Ira Fricke</a>	G1



Disco Hyperdata Browser  
navigating the Semantic Web as an  
unbound set of data sources

Person | Disco - Hyperdata Browser - Mozilla Firefox

Person | Disco - Hyperdata ...

URI:

Property	Value
type	<a href="#">Class</a>
type	<a href="#">Class</a>
comment	A person.
isDefinedBy	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>
label	Person
seeAlso	<a href="http://www4.wiwiss.fu-berlin.de/dblp/spaq?query=DESCRIBE+%3Chttp%3A%2F%2Fxmlns.com%2Ffoaf%2F0.1%2Fxmlns.com/foaf/0.1/workInfoHomepage">http://www4.wiwiss.fu-berlin.de/dblp/spaq?query=DESCRIBE+%3Chttp%3A%2F%2Fxmlns.com%2Ffoaf%2F0.1%2Fxmlns.com/foaf/0.1/workInfoHomepage</a>
subClassOf	<a href="http://www.w3.org/2000/10/swap/pim/contact#Person">http://www.w3.org/2000/10/swap/pim/contact#Person</a>
subClassOf	<a href="#">SpatialThing</a>
subClassOf	<a href="#">Agent</a>
subClassOf	<a href="http://xmlns.com/wordnet/1.6/Agent">http://xmlns.com/wordnet/1.6/Agent</a>
subClassOf	<a href="http://xmlns.com/wordnet/1.6/Person">http://xmlns.com/wordnet/1.6/Person</a>
disjointWith	<a href="#">Document</a>
disjointWith	<a href="#">Organization</a>
disjointWith	<a href="#">Project</a>
term status	stable

**Sources**

Displayed information originates from the following RDF graphs:

- G1. <http://xmlns.com/foaf/0.1/workInfoHomepage>
- G2. <http://www4.wiwiss.fu-berlin.de/dblp/resource/person/110566>
- G3. <http://xmlns.com/foaf/0.1/schoolHomepage>

## KIM platform

The KIM platform provides a novel infrastructure and services for:

- automatic semantic annotation,
- indexing,
- retrieval of unstructured and semi-structured content.

# KIM Constituents

The KIM Platform includes:

- **Ontologies** (PROTON + KIMSO + KIMLO) and KIM World KB
- **KIM Server** – with a set of APIs for remote access and integration
- **Front-ends**: Web-UI and plug-in for Internet Explorer.

# KIM Ontology (KIMO)

- light-weight upper-level ontology
- 250 NE classes
- 100 relations and attributes:
- covers mostly NE classes, and ignores general concepts
- includes classes representing lexical resources





## KIM KB

- KIM KB consists of above 80,000 entities (50,000 locations, 8,400 organization instances, etc.)
- Each location has geographic coordinates and several aliases (usually including English, French, Spanish, and sometimes the local transcription of the location name) as well as co-positioning relations (e.g. **subRegionOf**.)
- The organizations have **locatedIn** relations to the corresponding Country instances. The additionally imported information about the companies consists of short description, URL, reference to an industry sector, reported sales, net income, and number of employees.

# KIM is Based On...

KIM is based on the following open-source platforms:

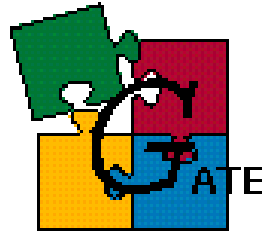
- **GATE** – the most popular NLP and IE platform in the world, developed at the University of Sheffield. Ontotext is its biggest co-developer.

[www.gate.ac.uk](http://www.gate.ac.uk) and [www.ontotext.com/gate](http://www.ontotext.com/gate)

- **OWLIM** – OWL repository, compliant with **Sesame** RDF database from Aduna B.V.

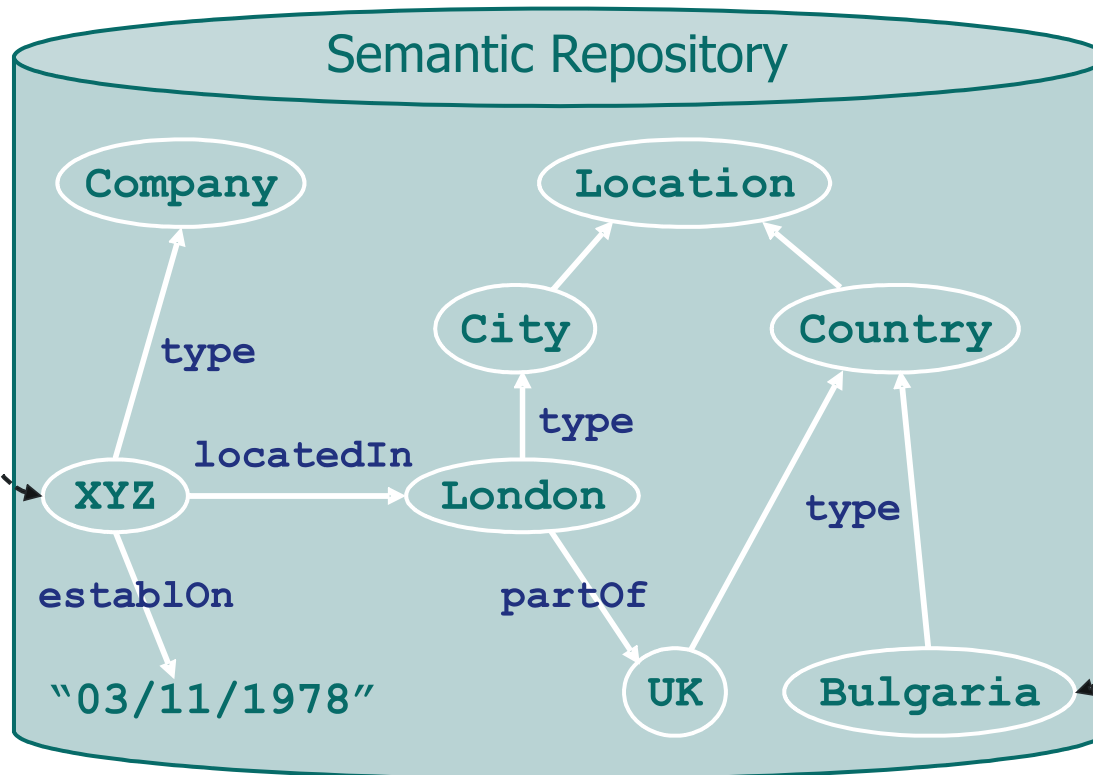
[www.ontotext.com/owlim](http://www.ontotext.com/owlim)

- **Lucene** – an open-source **IR engine** by Apache. [jakarta.apache.org/lucene/](http://jakarta.apache.org/lucene/)



# KIM Platform – Semantic Annotation

XYZ announced profits in Q3, planning to build a \$120M plant in Bulgaria, and more and more and more and more text...

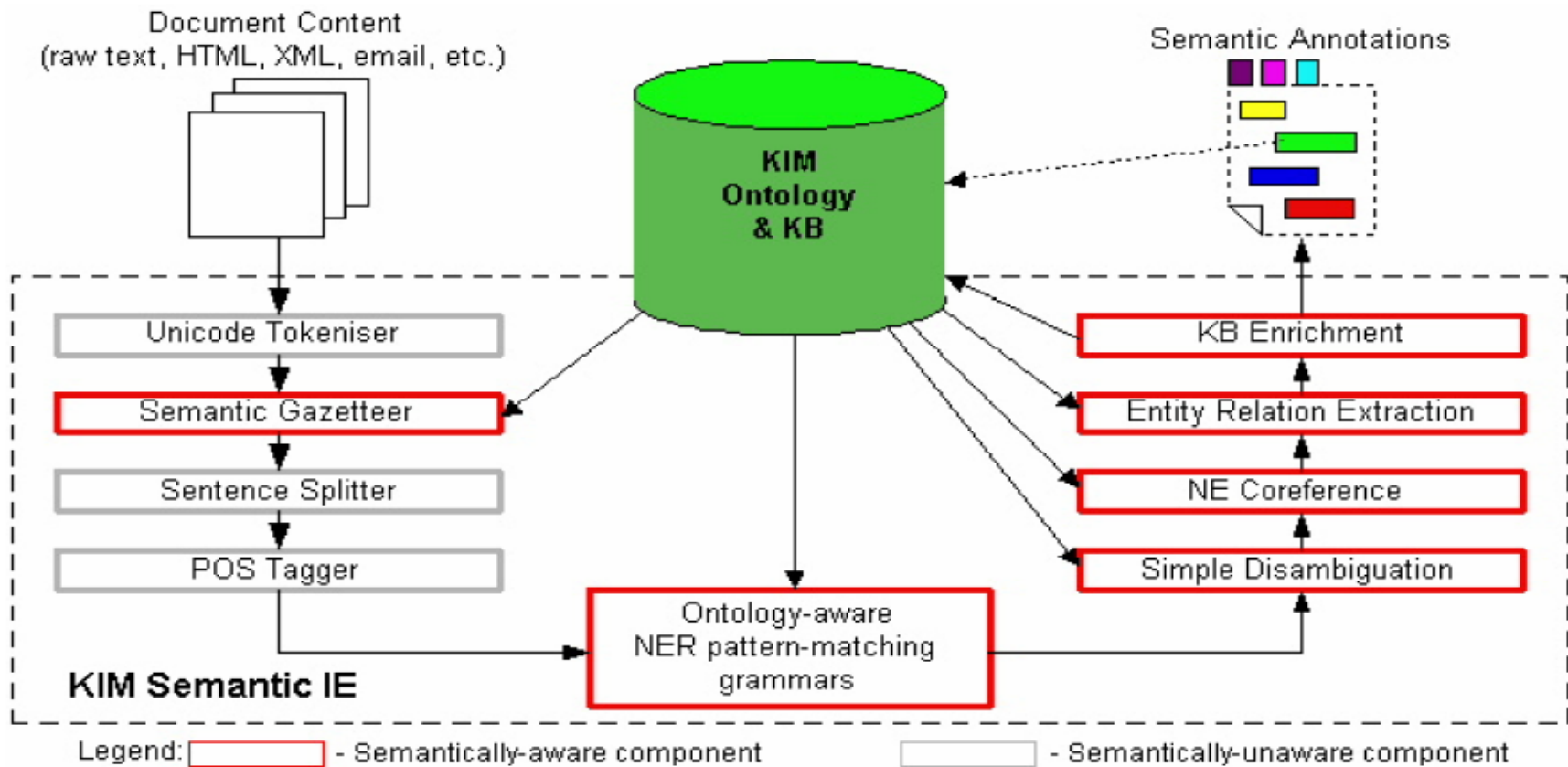


# KIM platform – Semantic Annotation

- The automatic semantic annotation is seen as a named-entity recognition (NER) and annotation process.
- The traditional flat NE type sets consist of several general types (such as Organization, Person, Date, Location, Percent, Money). In KIM the NE type is specified by reference to an ontology.
- The semantic descriptions of entities and relations between them are kept in a knowledge base (KB) encoded in the KIM ontology and residing in the same semantic repository. Thus KIM provides for each entity reference in the text (i) a link (URI) to the most specific class in the ontology and (ii) a link to the specific instance in the KB. Each extracted NE is linked to its specific type information (thus Arabian Sea would be identified as **Sea**, instead of the traditional – **Location**).

# KIM platform – Information Extraction

- KIM performs IE based on an ontology and a massive knowledge base.



# KIM platform - Browser Plug-in

- KIM Browser Plugin
  - Web content is annotated using ontologies
  - Content can be searched and browsed intelligently

Exploring: **Netherlands**, a Country, Trusted<sup>tip!</sup>

Property	value
hasMainAlias	Netherlands
hasAlias	Nederland
hasAlias	Holland
subRegionOf	Western Europe
subRegionOf	Europe
hasCapital	Amsterdam

*Select one or more concepts from the ontology...  
Currently loaded  
Annotation*

Exploring: **Amsterdam**, a CountryCapital, Trusted<sup>tip!</sup>

Property	value
hasMainAlias	Amsterdam
hasAlias	Mokum
Latitude	52.373056
Longitude	4.892222
subRegionOf	North Holland
subRegionOf	Netherlands
subRegionOf	Western Europe
subRegionOf	Europe

**Trusted Content**

# SEMANTIC WEB - PROCESSES

# Processes

- The Web is moving from static data to dynamic functionality
  - Web services: a piece of software available over the Internet, using standardized XML messaging systems over the SOAP protocol
  - Mashups: The compounding of two or more pieces of web functionality to create powerful web applications



# Semantic Web - Processes

The image consists of two side-by-side screenshots from a Mozilla Firefox browser window.

**Left Screenshot: Pipes - editing 'Aggregated News Alerts'**

- Address Bar:** `http://pipes.yahoo.com/pipes/pipe.edit?id=fELaGmGz2xGBTC3qe5kA&opendesc=true`
- Page Title:** Aggregated News Alerts
- Left Panel:** A sidebar menu with categories like Sources, User inputs, Operators, and Depreciated. Under Sources, options include Fetch CSV, Feed Auto-Discover, Fetch Feed, Fetch Data, Fetch Site Feed, Flickr, Google Base, Item Builder, Yahoo! Local, and Yahoo! Search.
- Main Canvas:** A workflow diagram with two 'URL Builder' components. The first component has a Base of `rss.findory.com/rss` and a Query parameter `q` with a value of `text [wired]`. The second component has a Base of `api.technorati.com/` and a Query parameter `format` with a value of `rss`. Arrows indicate data flow between these components and other parts of the pipeline.

**Right Screenshot: The New York Times - RSS**

- Address Bar:** `http://www.nytimes.com/services/xml/rss/`
- Page Title:** The New York Times RSS
- Header:** NYTimes.com RSS feeds. Includes a search bar and a 'Go' button.
- Content Area:** A grid of news categories, each with an RSS icon. Categories include BUSINESS, ARTS, EDUCATION, HEALTH, MEDIA & ADVERTISING, WORLD BUSINESS, SMALL BUSINESS, YOUR MONEY, DEATHBOOK, DESIGN, MUSIC, TELEVISION NEWS, AUTOMOBILES, BOOKS, and DINING & WINE.
- Text Content:**
  - NYTimes.com RSS feeds:** Really Simple Syndication (RSS) is an XML-based format for content distribution. NYTimes.com offers several RSS feeds for use in news readers and Web logs (blogs). These feeds include headlines, summaries and links back to NYTimes.com for the full article. RSS feeds are free and NYTimes.com currently uses RSS 2.0.
  - How to sign up for NYTimes.com RSS feeds:** Click on the below content area(s) you're interested in subscribing to, and follow the instructions to add to your news reader or your personal My Yahoo! or My AOL page.
  - NYTimes.com Podcasts:** Subscribe to NYTimes.com podcasts to get audio automatically delivered to you. **Free podcasts** are available.
  - How to access RSS feeds:** There are a number of ways to access RSS feeds. You can install a news reader that displays RSS feeds from the Web sites you select, enabling you to view hundreds of headlines at once. After installing the news reader, you can add each feed manually from the Web site by clicking on the "Subscribe" or the "XML" orange button next to the feed.
  - Alternative:** An alternative to downloading a dedicated news reader is to use a Web-based news reader. For example, My Yahoo! or My AOL users can now add RSS feeds directly to their personal page.
  - Terms for NYTimes.com RSS feeds:** We encourage the use of NYTimes.com RSS feeds for personal use in a news reader or as part of a non-commercial Web site or blog. We require proper format and attribution whenever New York Times content is posted on your Web site, and we reserve the right to require that you cease distributing NYTimes.com content.

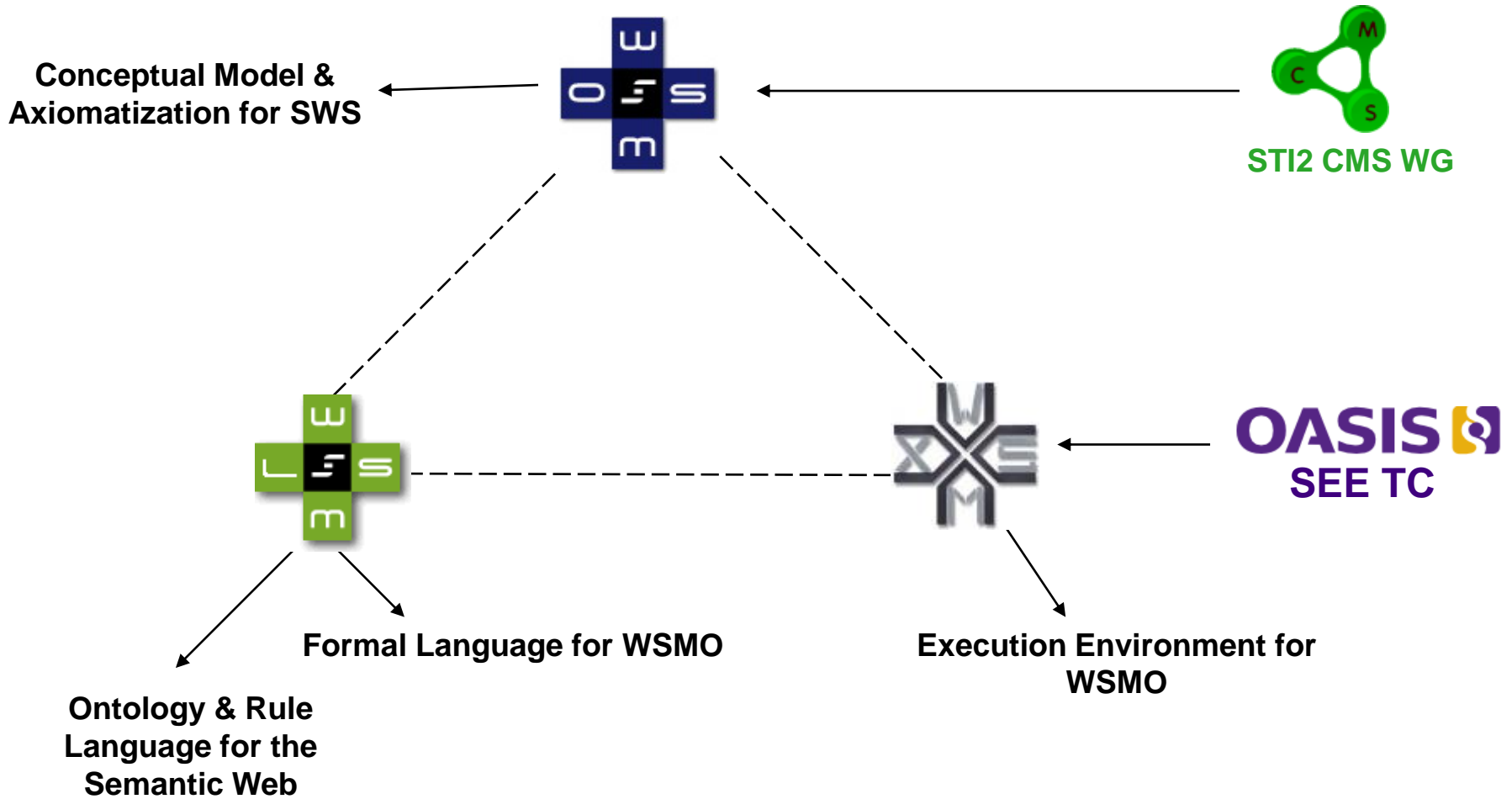
# Semantic Web - Processes

- Web services and mashups are limited by their syntactic nature
- As the amount of services on the Web increases it will be harder to find Web services in order to use them in mashups
- The current amount of human effort required to build applications is not sustainable at a Web scale

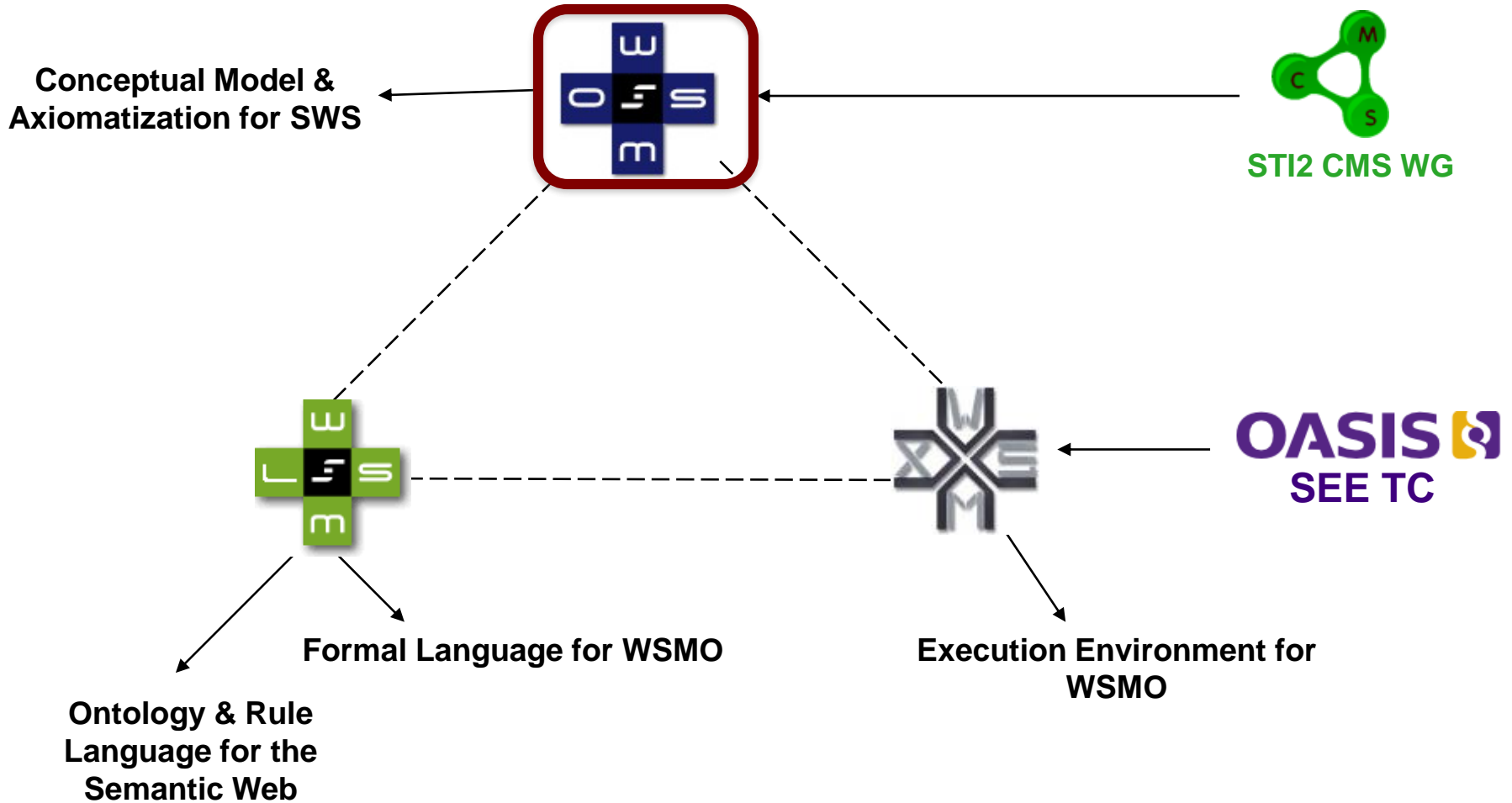
# Semantic Web - Processes

- The addition of semantics to form Semantic Web Services and Semantically Enabled Service-oriented Architectures can enable the automation of many of these currently human intensive tasks
  - Service Discovery, Adaptation, Ranking, Mediation, Invocation
- Frameworks:
  - **OWL-S**: WS Description Ontology (Profile, Service Model, Grounding)
  - **WSMO**: Ontologies, Goals, Web Services, Mediators
  - **SWSF**: Process-based Description Model & Language for WS
  - **SAWSDL (WSDL-S)**: Semantic annotation of WSDL descriptions

# The WSMO Approach

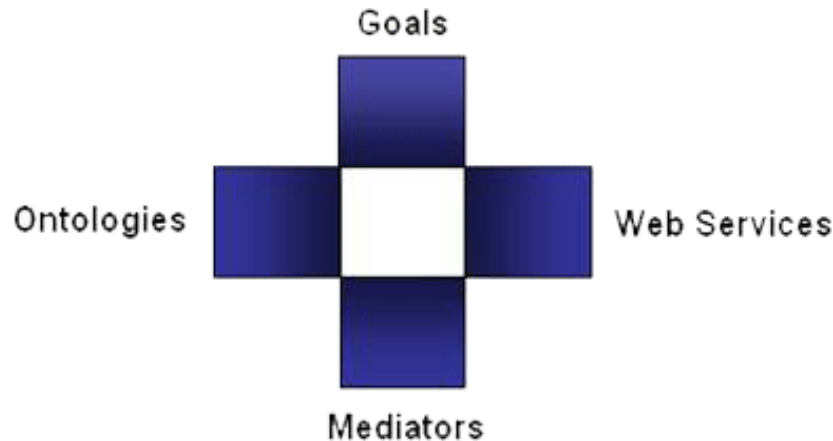


# Web Service Modeling Ontology (WSMO)



# WSMO

Objectives that a client wants to achieve by using Web Services



Provide the formally specified terminology of the information used by all other components

Semantic description of Web Services:

- **Capability** (*functional*)
- **Interfaces** (*usage*)

Connectors between components with mediation facilities for handling heterogeneities

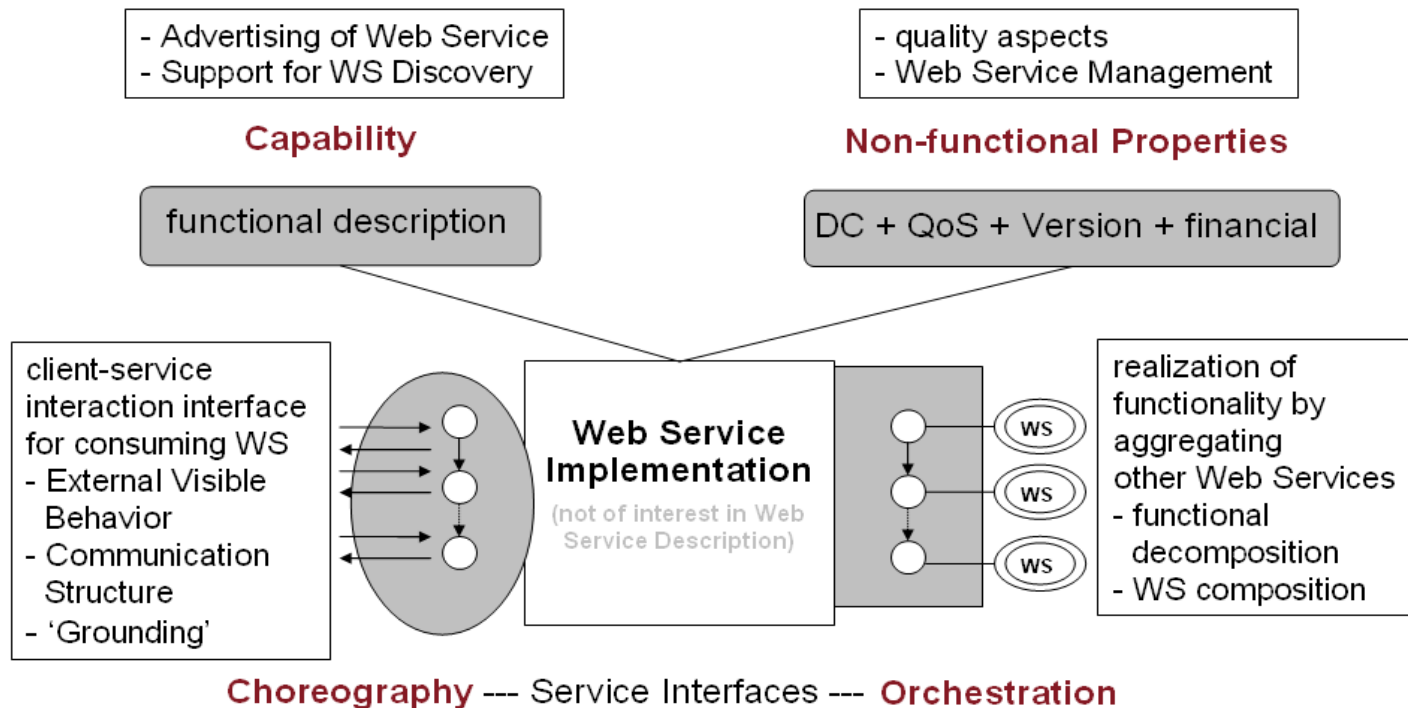
# WSMO Top Elements

- **Ontologies:**

- In WSMO, Ontologies are the key to linking conceptual real-world semantics defined and agreed upon by communities of users

- **Web Services:**

- In WSMO, Web service descriptions consist of non-functional, functional, and the behavioral aspects of a Web service



# WSMO Top Elements (1)

- **Goals:**

- Goals are representations of an objective for which fulfillment is sought through the execution of a Web service. Goals can be descriptions of Web services that would potentially satisfy the user desires

**Class goal sub-Class** wsmoElement

importsOntology **type** ontology

usesMediator **type** {ooMediator, ggMediator}

hasNonFunctionalProperties **type** nonFunctionalProperty

requestsCapability **type** capability *multiplicity = single-valued*

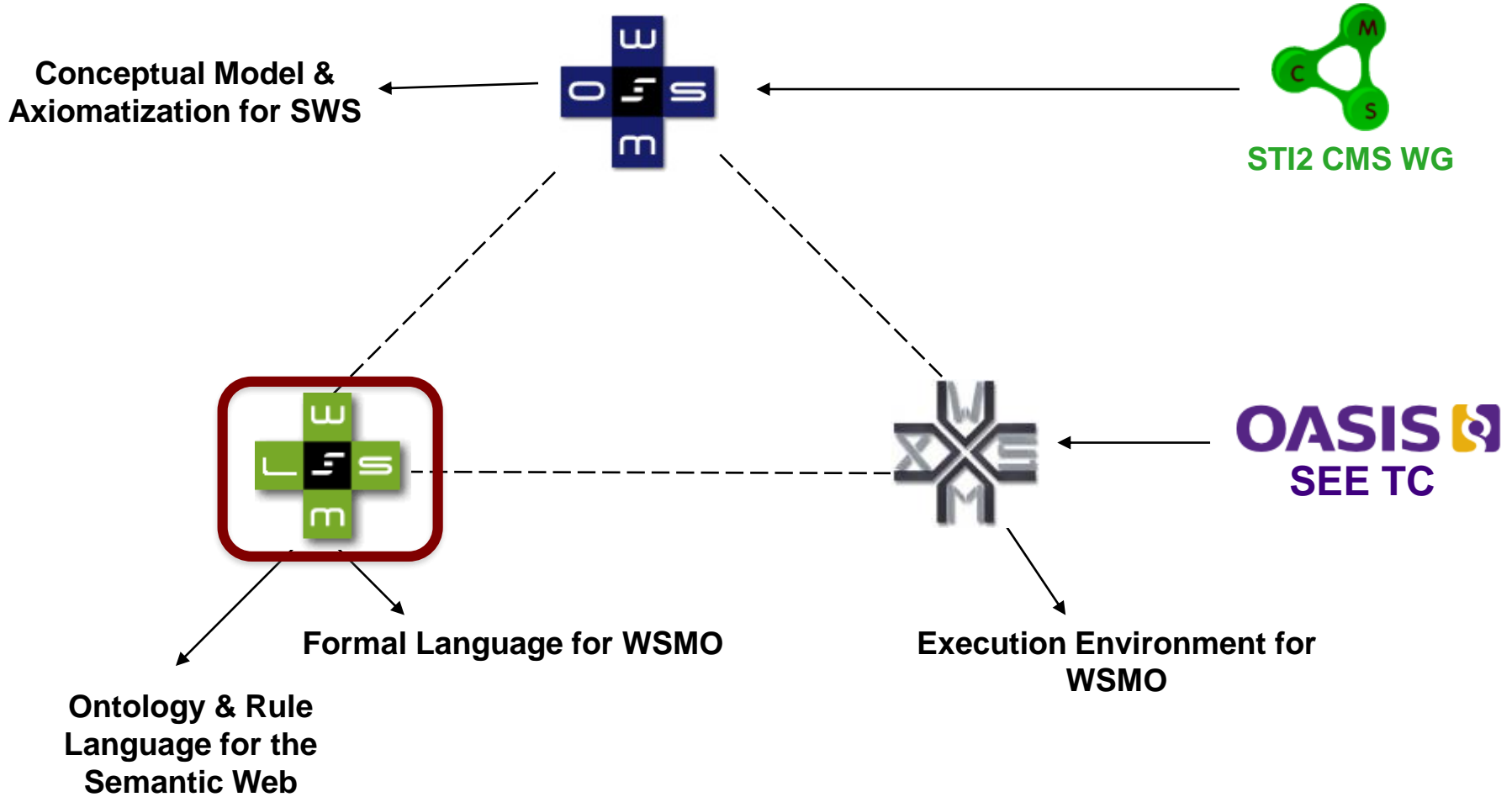
requestsInterface **type** interface

- **Mediators:**

- In WSMO, heterogeneity problems are solved by mediators at various levels:
  - **Data Level** - mediate heterogeneous Data Sources
  - **Protocol Level** - mediate heterogeneous Communication Patterns
  - **Process Level** - mediate heterogeneous Business Processes

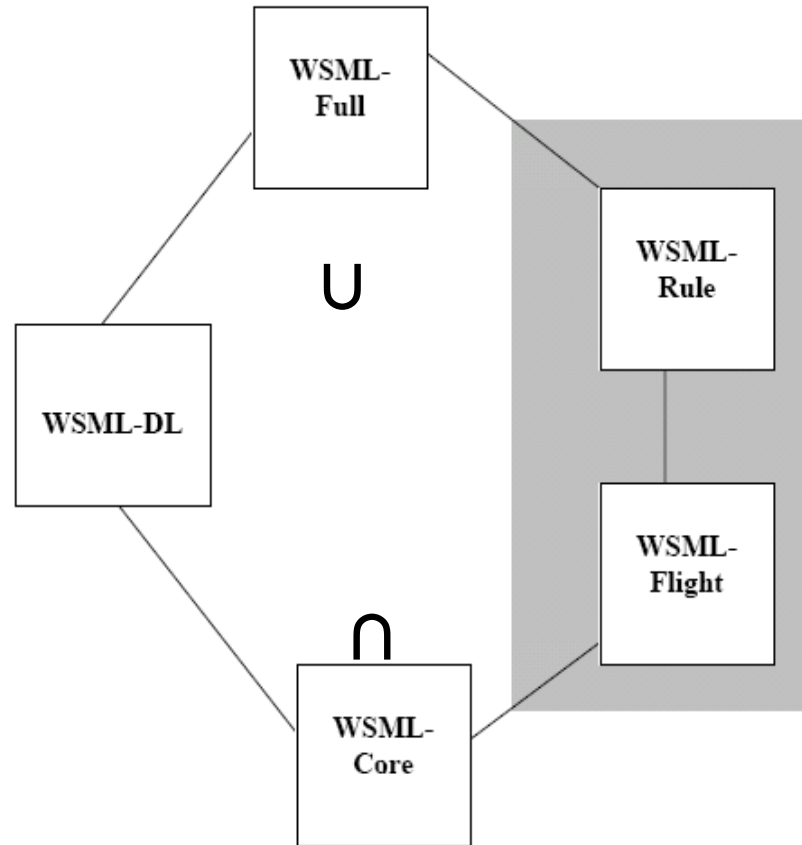


# Web Service Modeling Language (WSML)

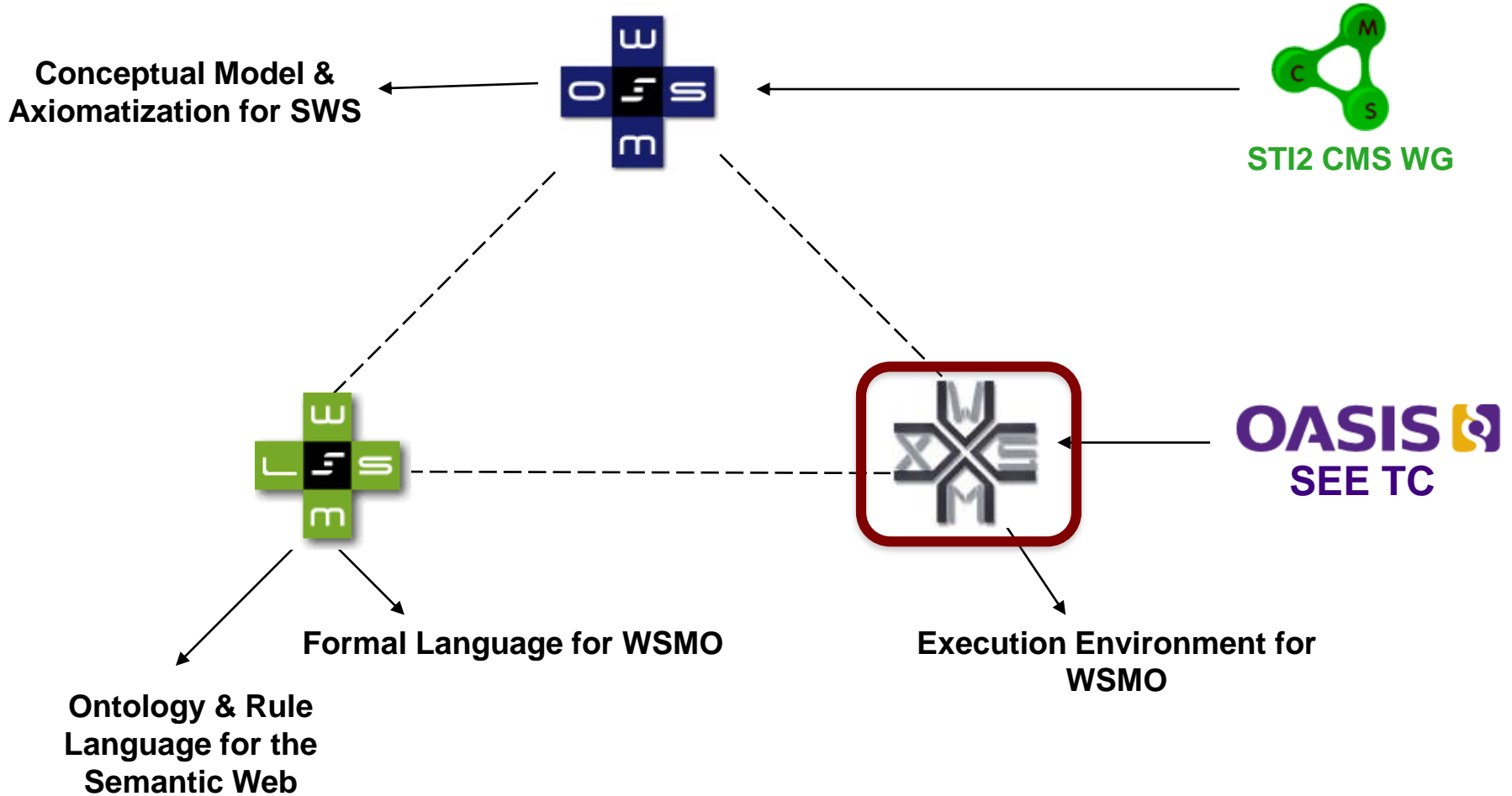


# WSML Variants

- WSML Variants - allow users to make the trade-off between the provided **expressivity** and the implied **complexity** on a per-application basis



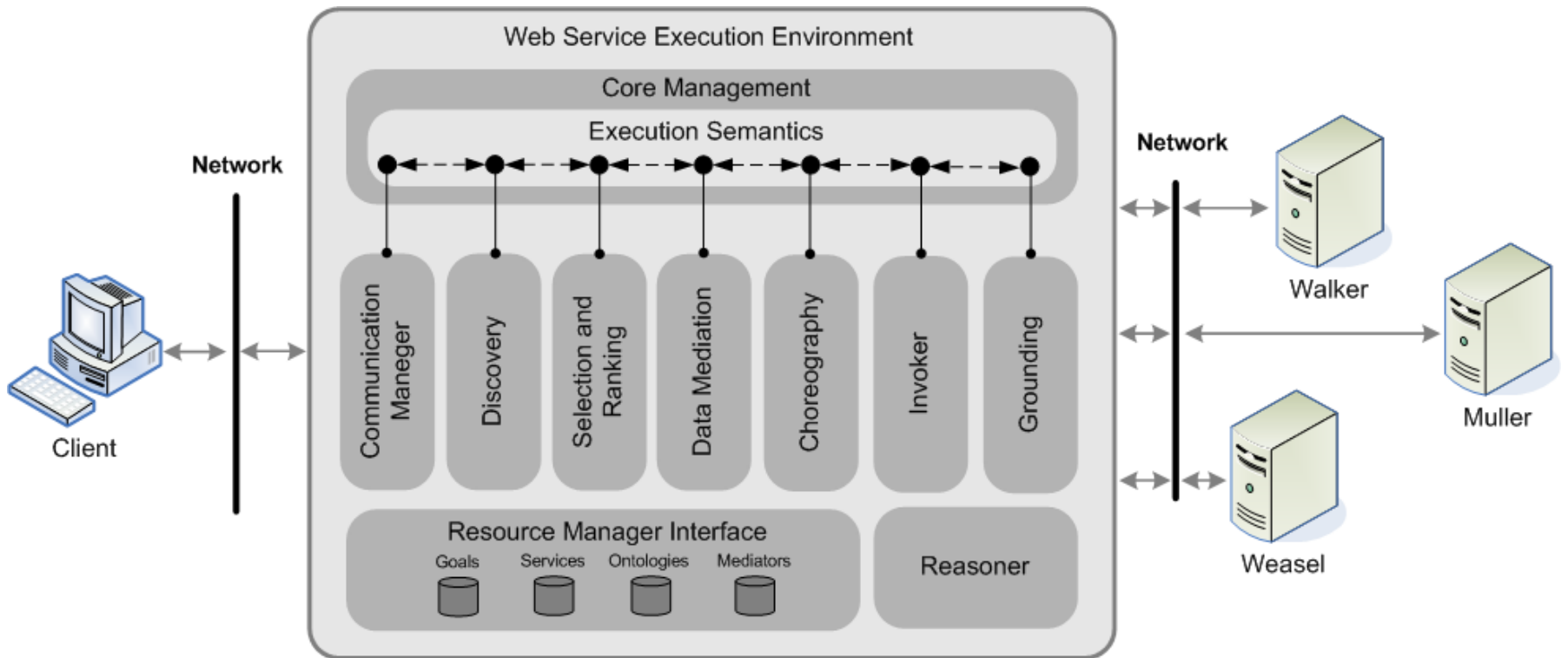
# Web Service Execution Environment (WSMX)



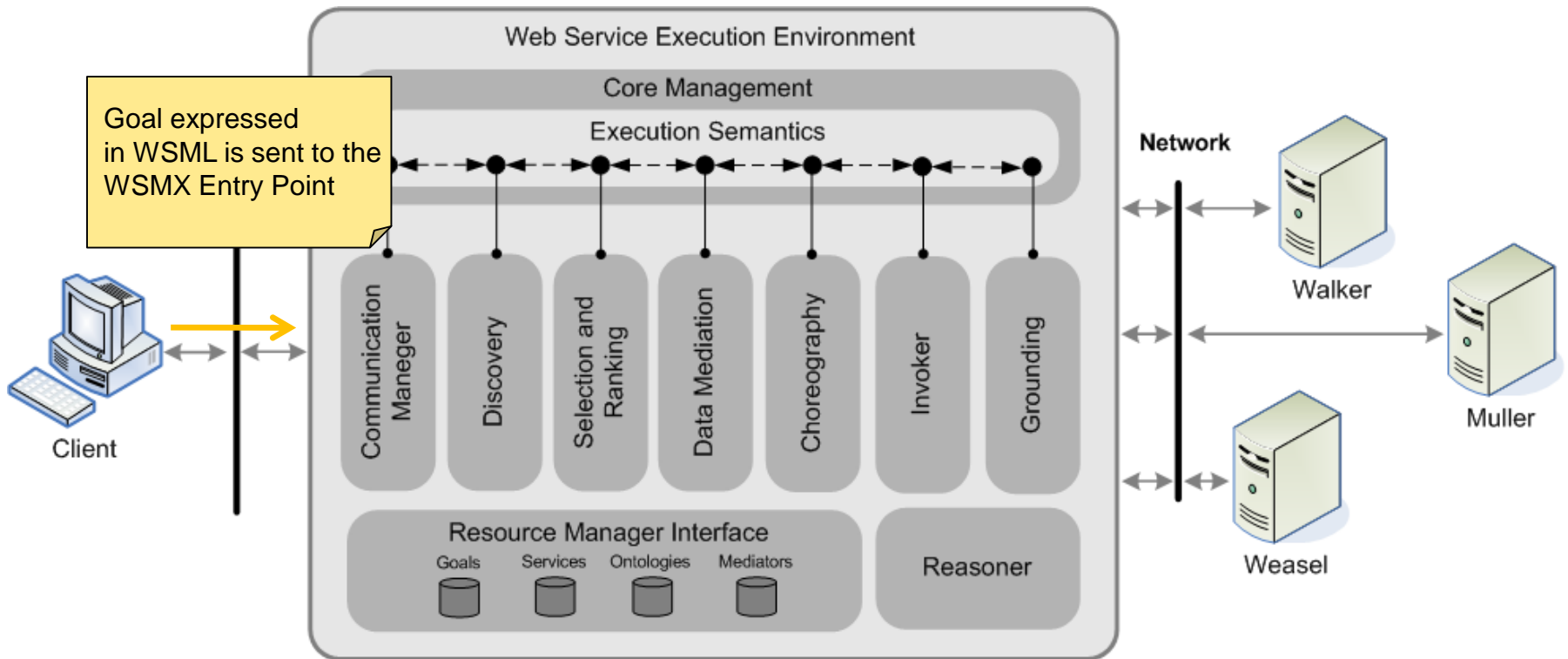
# Web Service Execution Environment (WSMX)

- ... is comprehensive software framework for runtime binding of service requesters and service providers,
- ... interprets service requester's goal to
  - discover matching services,
  - select (if desired) the service that best fits,
  - provide data/process mediation (if required), and
  - make the service invocation,
- ... is reference implementation for WSMO,
- ... has a formal execution semantics, and
- ... is service oriented, event-based and has pluggable architecture
  - Open source implementation available through Source Forge,
  - based on microkernel design using technologies such as JMX.

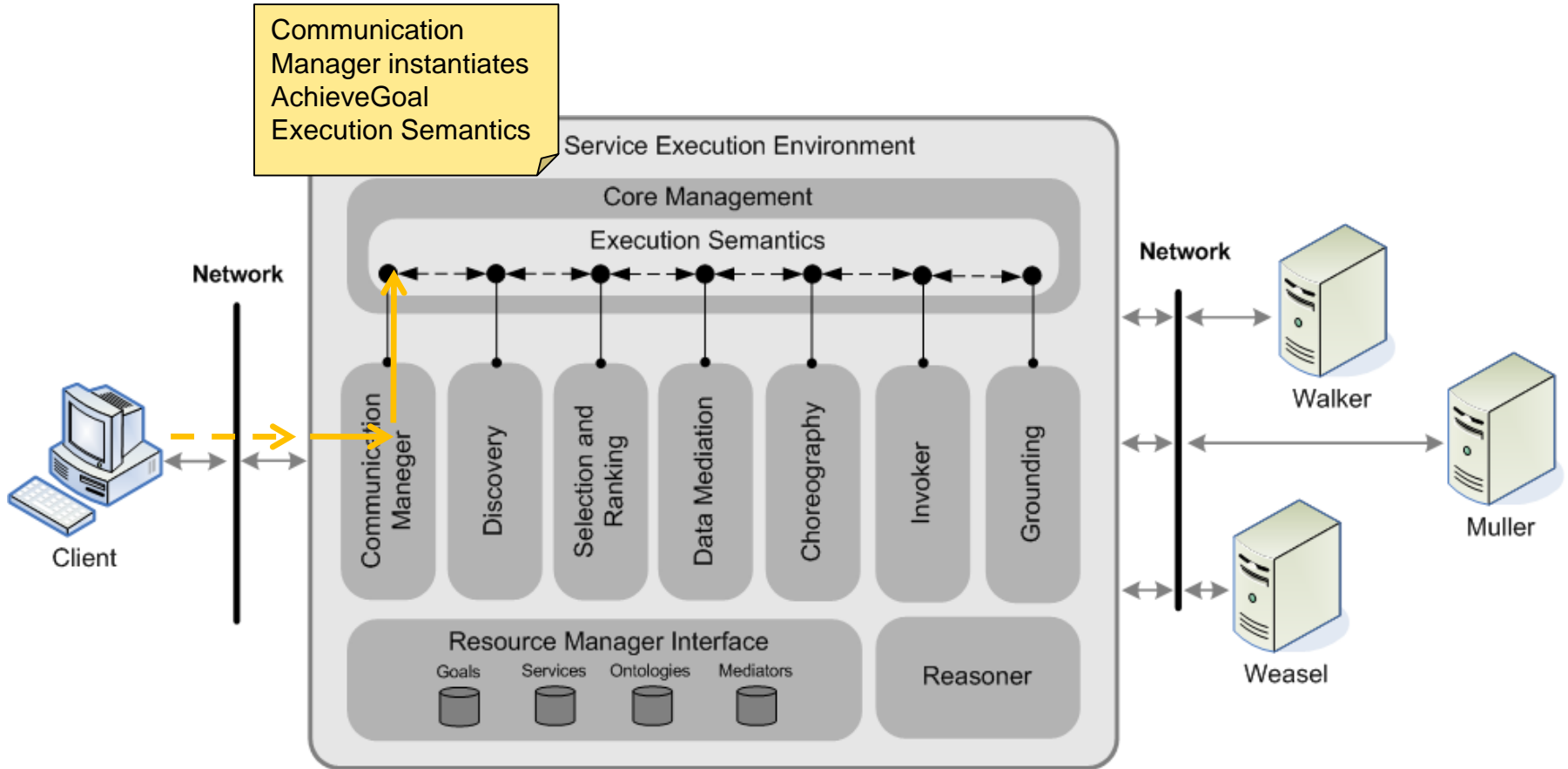
# WSMX Illustration



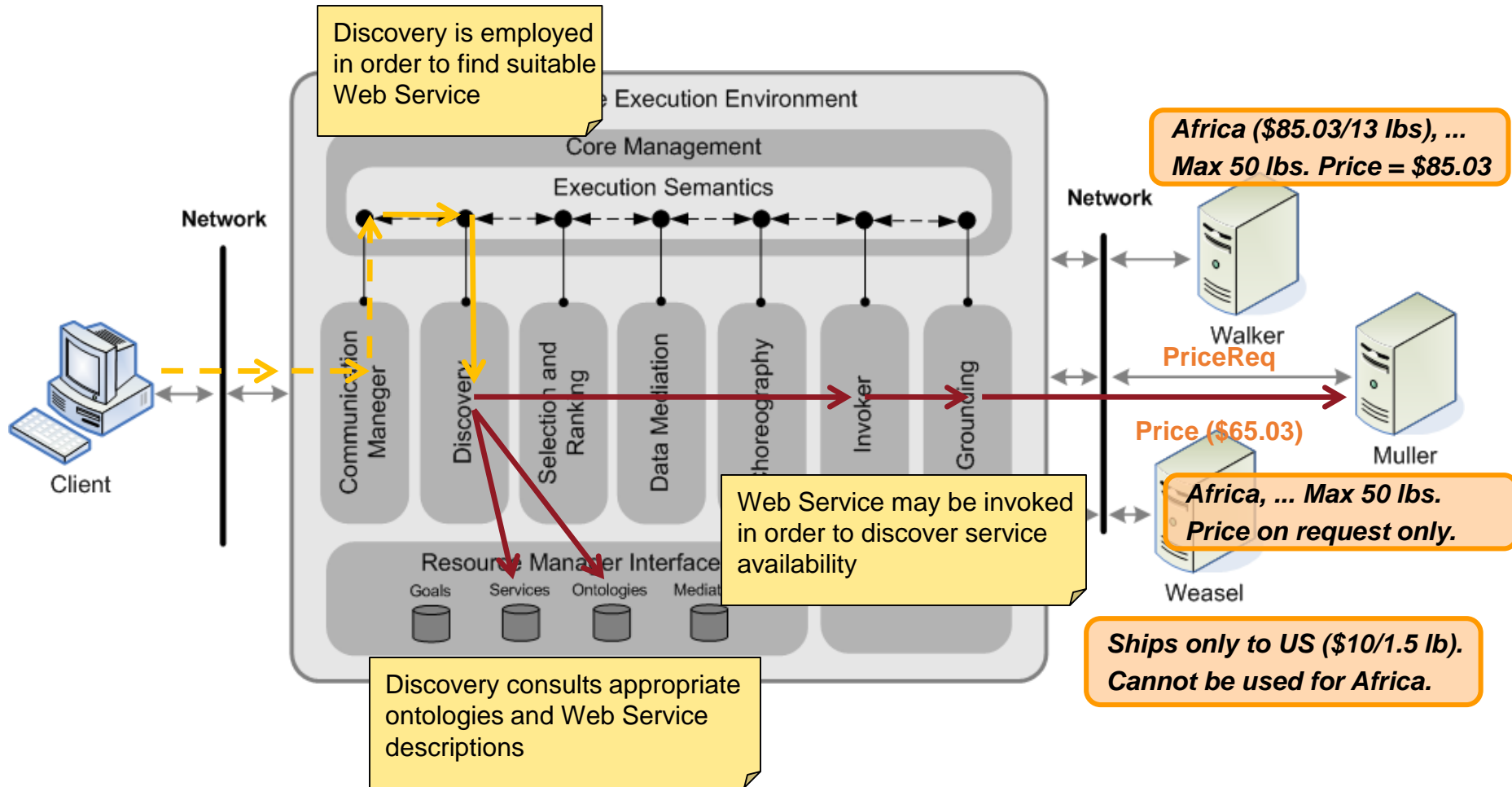
# WSMX Illustration



# WSMX Illustration

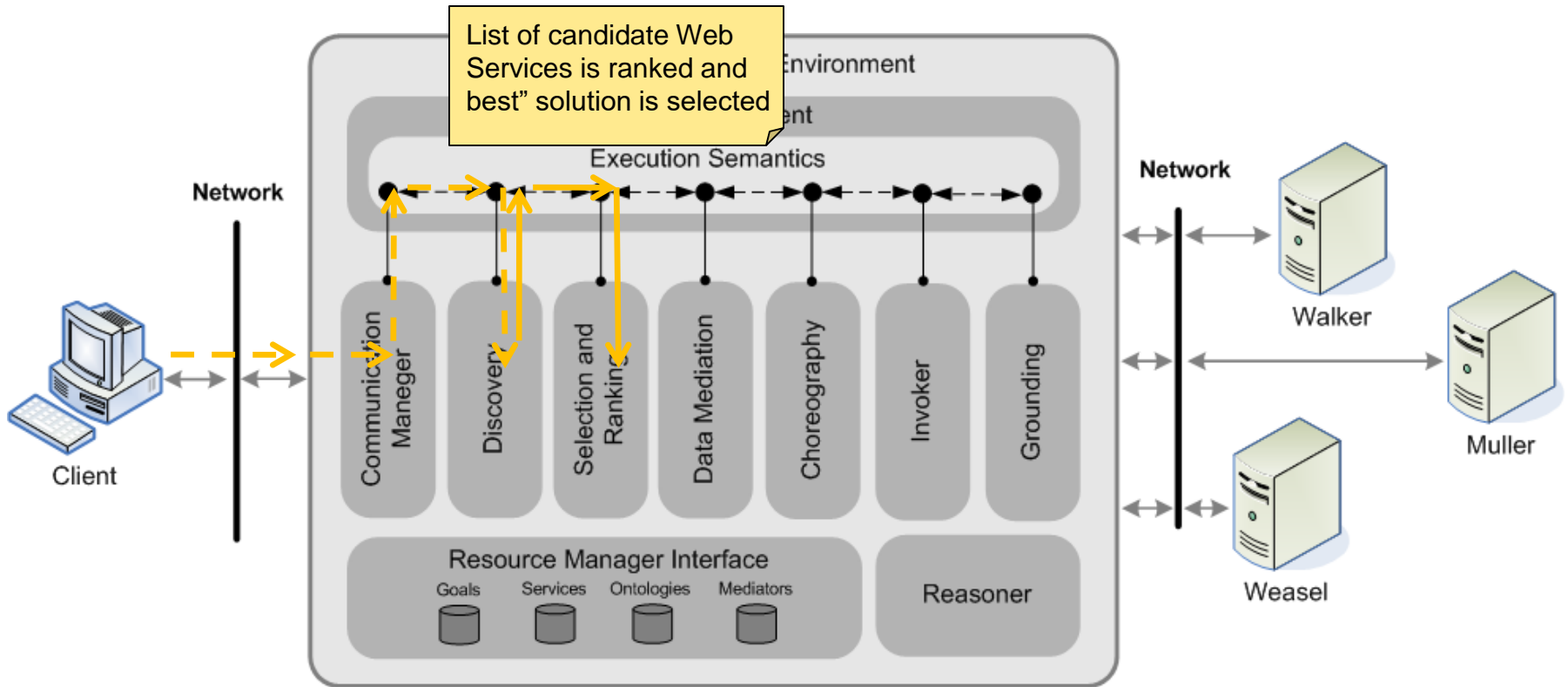


# WSMX Illustration

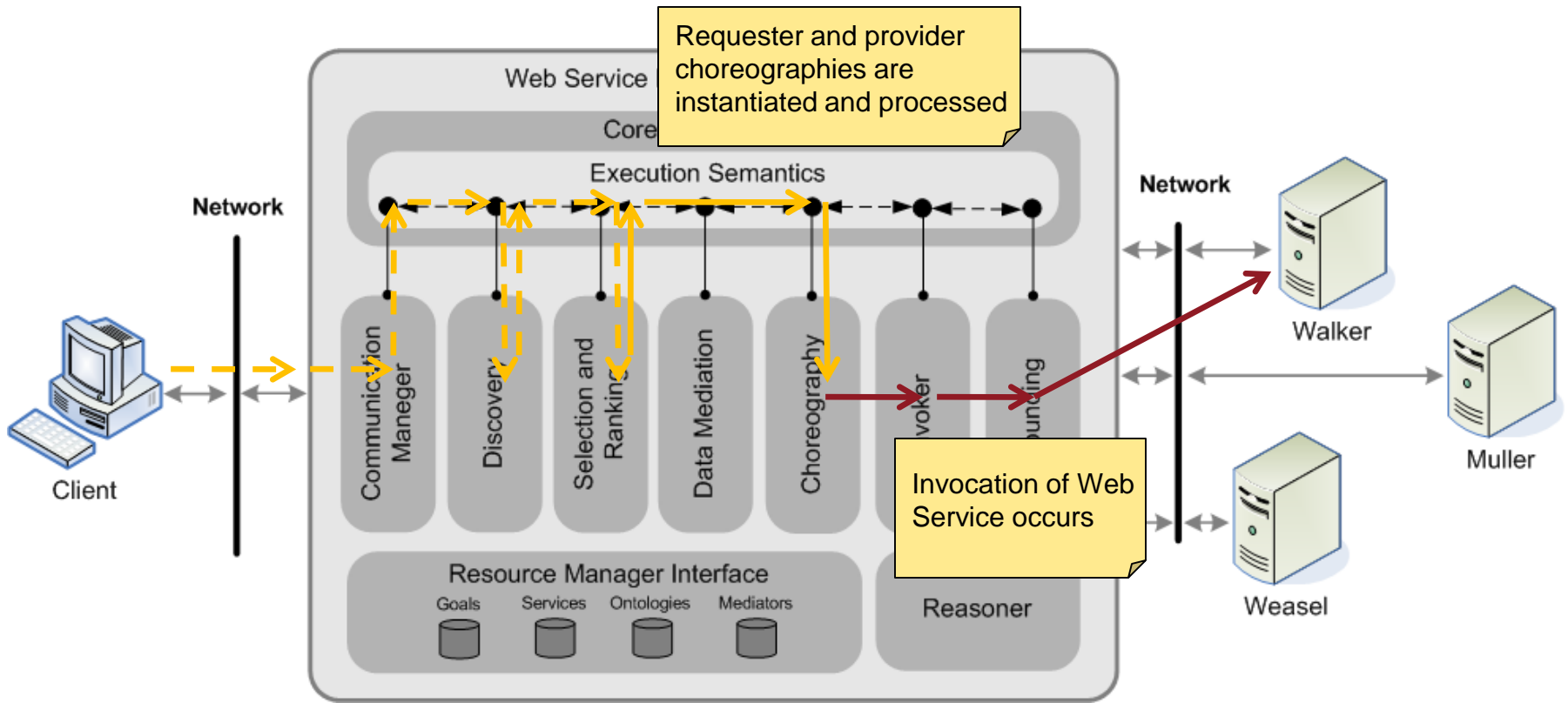




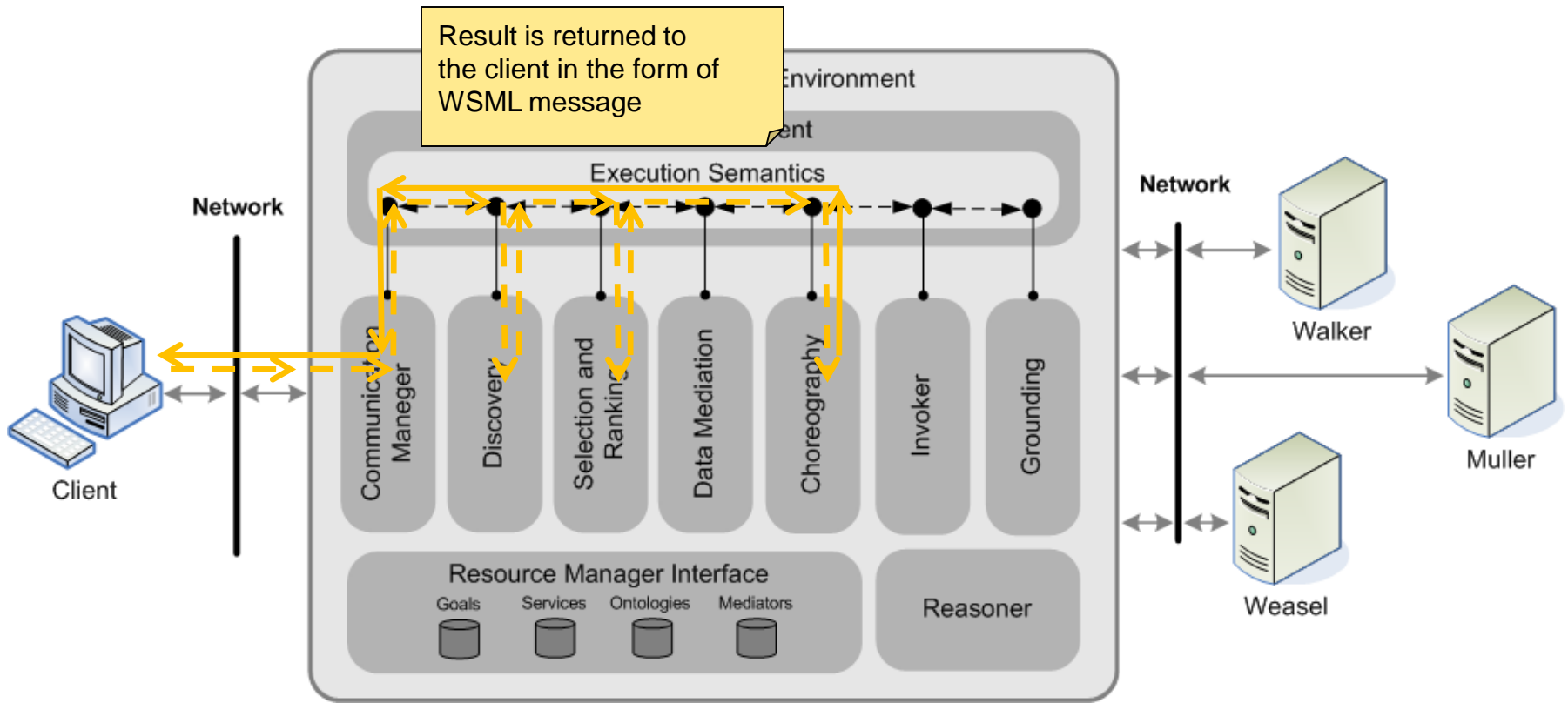
# WSMX Illustration



# WSMX Illustration



# WSMX Illustration



# RECENT TRENDS

# Open government UK

The image shows the homepage of data.gov.uk. At the top left is the HM Government logo. The top right corner displays the URL 'data.gov.uk'. A dark navigation bar contains links for Home, Blog, Data, SPARQL, Apps, Ideas, Forum, Wikl, Resources, and About. A search bar is located on the right side of this bar. Below the navigation bar is a large banner with the text 'Unlocking innovation' and 'Working with UK Public Sector information and data', accompanied by a blue molecular structure graphic. To the right of the banner is a 'Subscribe by RSS' button and social media icons for G+ and Facebook. Below the banner is a 'User login' section with input fields for 'Username: \*' and 'Password: \*', a 'Log in' button, and a link for 'Request new password'. At the bottom, there are two main sections: 'Search Data' with a search input field and a 'Search' button, and 'Browse for Data' with links for 'List all datasets' and 'Common tags'. A footer section on the right contains the text 'What is the Semantic Web?' and 'Combining different data'.

HM Government

data.gov.uk

Home Blog Data SPARQL Apps Ideas Forum Wikl Resources About

Unlocking innovation

Working with UK Public Sector information and data

Subscribe by RSS

G+ Facebook

User login

Username: \*

Password: \*

Log in

Request new password

Search Data

Enter keyword(s) Search

e.g. education, NHS, crime, transport, environment

Browse for Data

List all datasets

Common tags

What is the Semantic Web?

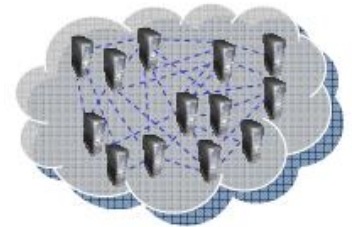
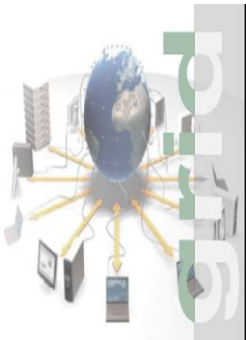
Combining different data

# Open government UK

- British government is opening up government data to the public through the website **data.gov.uk**.
- **data.gov.uk** has been developed by Sir Tim Berners-Lee, founder of the Web and Prof. Nigel Shadbolt at the University of Southampton.
- **data.gov.uk** was launched in January 2010
- **data.gov.uk** will publish governmental non-personal data using the *Resource Description Framework (RDF)* data model
- Query of data is possible using *SPARQL*

# Cloud computing

- Grid Computing
  - solving large problems with parallel computing
- Utility Computing
  - Offering computing resources as a metered service
- Software as a service
  - Network-based subscription to applications
- Cloud Computing
  - Next generation internet computing
  - Next generation data centers



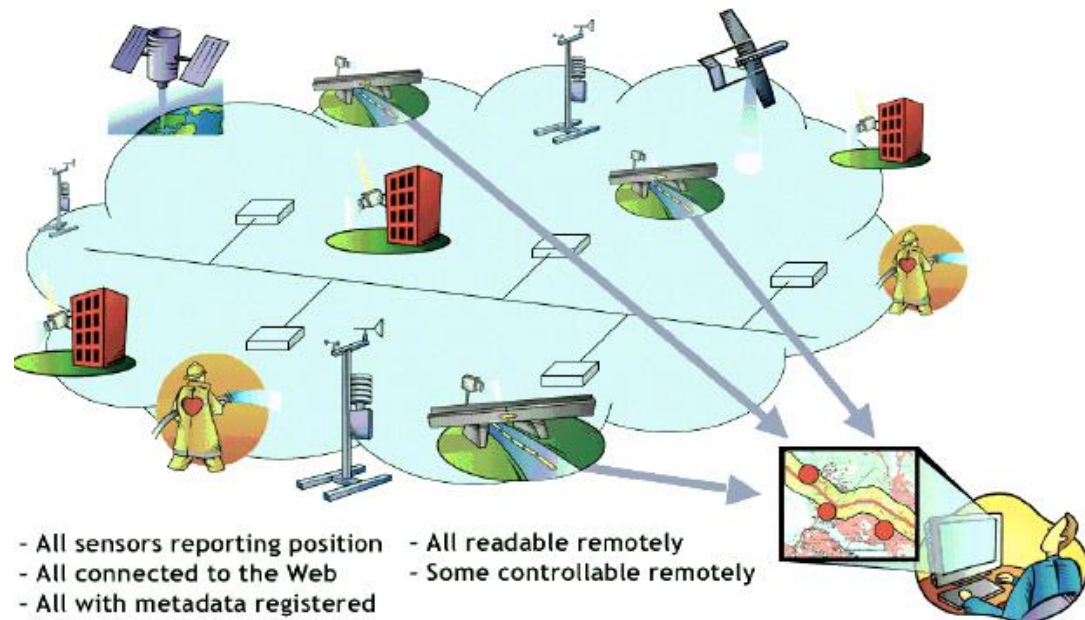
# Cloud computing

- Including semantic technologies in Cloud Computing will enable:
  - Flexible, dynamically scalable and virtualized data layer as part of the cloud
  - Accurate search and acquire various data from the Internet,



# Mobiles and Sensors

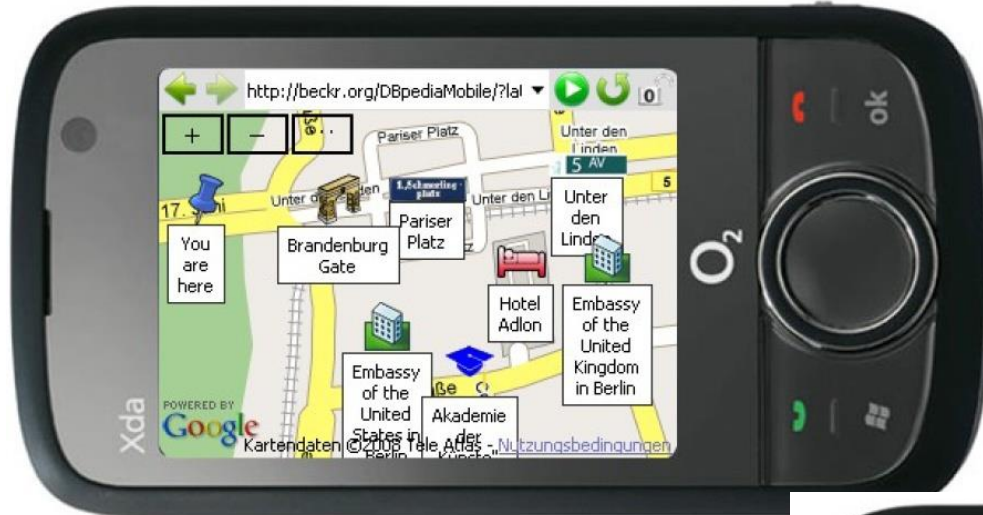
- Extending the mobile and sensors networks with Semantic technologies, Semantic Web will enable:
  - Interoperability at the level of sensors data and protocols
  - More precise search for mobile capabilities and sensors with desired capability



# Linked Open Data and Mobiles

- Combination of Linked Open Data and Mobiles has triggered the emergence of new applications
- One example is **DBpedia Mobile** that based on the current GPS position of a mobile device renders a map containing information about nearby locations from the DBpedia dataset.
- It exploits information coming from DBpedia, Revyu and Flickr data.
- It provides a way to explore maps of cities and gives pointers to more information which can be explored

# Linked Open Data and Mobiles



Pictures from DBpedia Mobile



Try yourself: <http://wiki.dbpedia.org/DBpediaMobile>

# SUMMARY

# Summary

- Semantic Web is not a *replacement* of the current Web, it's an *evolution* of it
- Semantic Web is about:
  - annotation of data on the Web
  - data linking on the Web
  - data Integration over the Web
- Semantic Web aims at *automating* tasks currently carried out by humans
- Semantic Web is becoming *real* (maybe not as we originally envisioned it, but it is)

# REFERENCES

# References

- **Mandatory reading:**

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

- **Further reading:**

- D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd Edition, Springer 2003.
- G. Antoniou and F. van Harmelen. *A Semantic Web Primer*, (2nd edition), The MIT Press 2008.
- H. Stuckenschmidt and F. van Harmelen. *Information Sharing on the Semantic Web*, Springer 2004.
- T. Berners-Lee. *Weaving the Web*, HarperCollins 2000
- T.R. Gruber, *Toward principles for the design of ontologies used or knowledge sharing?*, Int. J. Hum.-Comput. Stud., vol. 43, no. 5-6,1995

# References

- Wikipedia and other links:
  - [http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)
  - [http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework)
  - [http://en.wikipedia.org/wiki/Linked\\_Data](http://en.wikipedia.org/wiki/Linked_Data)
  - <http://www.w3.org/TR/rdf-primer/>
  - <http://www.w3.org/TR/rdf-mt/>
  - <http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial>
  - <http://linkeddata.org/>
  - <http://www.opengeospatial.org/projects/groups/sensorweb>
  - <http://www.data.gov.uk/>



Questions?

