Semantic Web

Andrea Giovanni Nuzzolese

10/11/2016 - Bologna

What do you know about the Web? and Internet?

- Basic protocols
  - MAC, PPP, DSL, … (link layer, physical networks)
  - TCP, UDP, … (transport layer)
  - IPv4, IPv6, … (internet layer)
  - HTTP (GET, POST, …), SSH, FTP, DNS, POP, IMAP, SMTP, … (application layer)
- Identification scheme
  - URI (Uniform Resource Identifier)
    - URL (Uniform Resource Locator)
      - PURL (Persistent URL)
    - URN (Uniform Resource Name) e.g. ISBN, ISSN, DOI (Digital Object Identifier)
    - IRI (International Resource Identifier)
  - Okkam ENS (Entity Name System, http://api.okkam.org/)
- Languages
  - HTML, CSS
  - XML, XSD
  - PHP
  - Javascript
- Architectures
  - Client-Server, LAMP, WISA
- Request-Response Loop / Peer-to-Peer models
- A lot of hardware (copper, fibres, computers, routers, switches, air conditioning, farms, data centers, …)
- …

Internet protocol suite

Application layer

BGP · DHCP · **DNS** · FTP · HTTP · IMAP · LDAP · MGCP · NNTP · NTP · POP · ONC/RPC · RTP · RTSP · RIP · SIP · SMTP · SNMP · SSH · Telnet · TLS/SSL · XMPP · *more...*

Transport layer

TCP · UDP · DCCP · SCTP · RSVP · *more...*

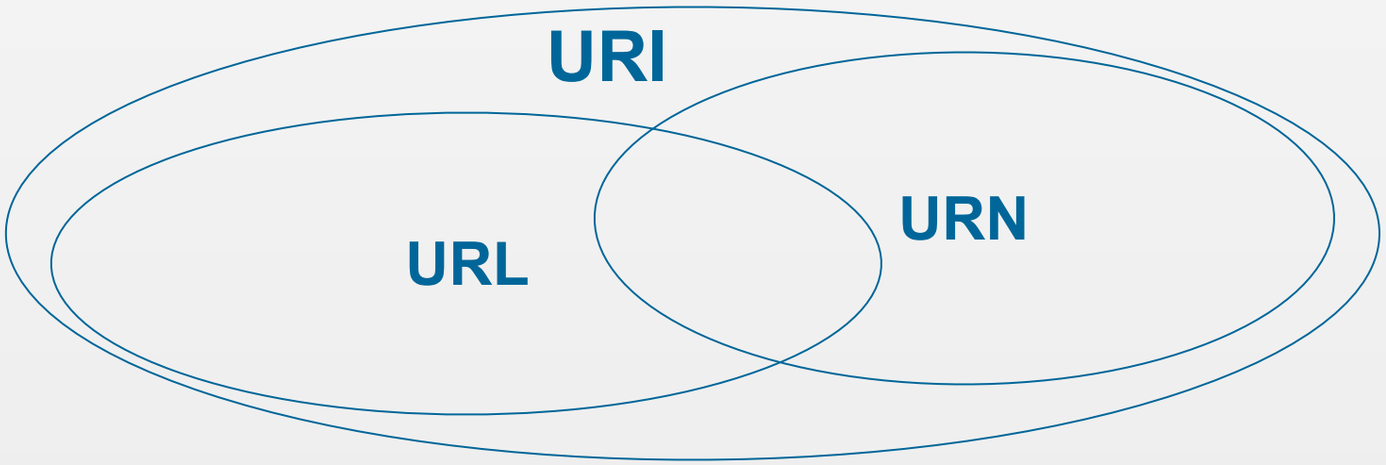
Internet layer

IP (IPv4 · IPv6) · ICMP · ICMPv6 · ECN · IGMP · IPsec · *more...*

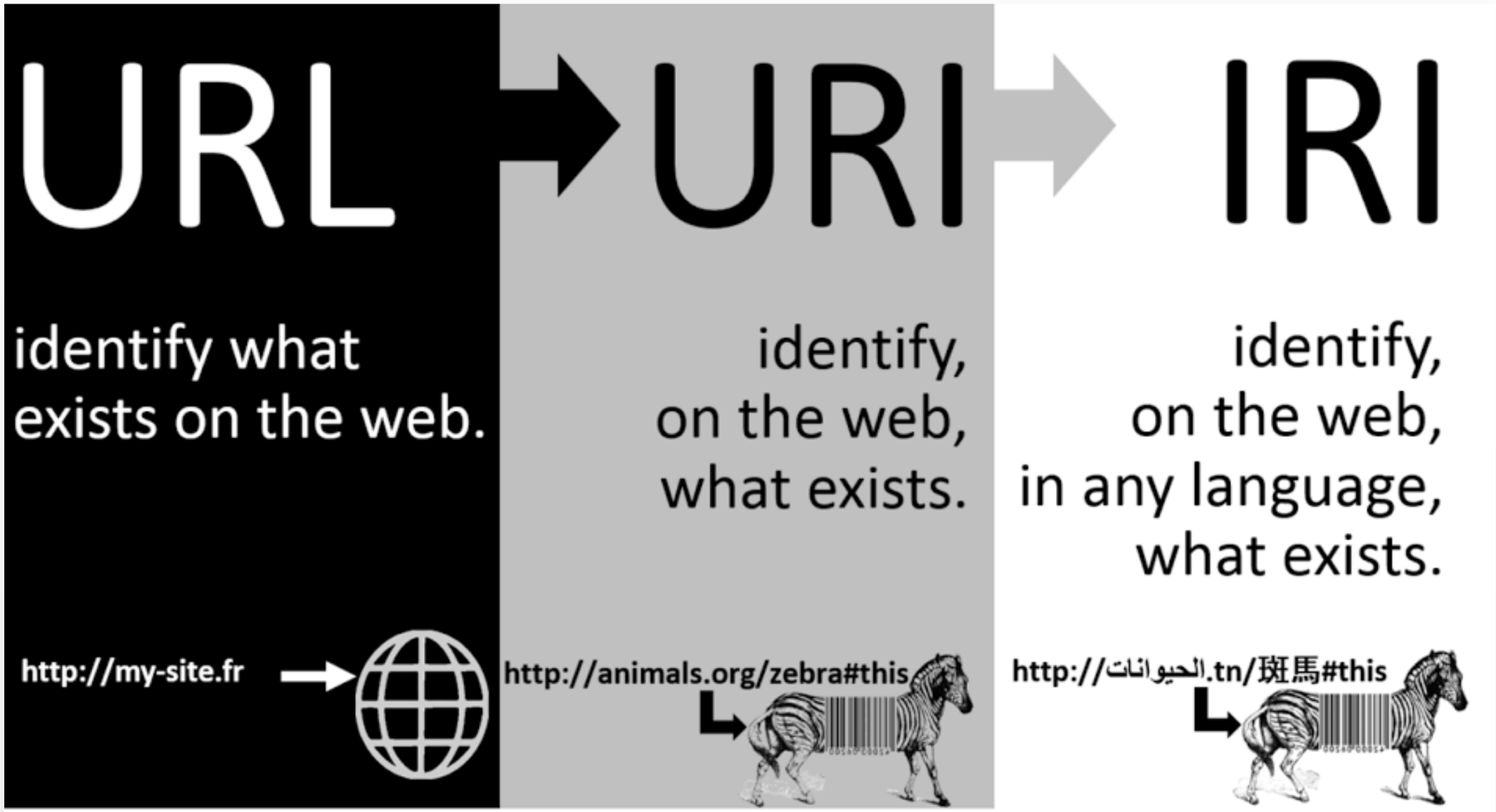
Link layer

ARP · NDP · OSPF · Tunnels (L2TP) · PPP · MAC (Ethernet · DSL · ISDN · FDDI) · *more...*

V · T · E



URL/URI/IRI

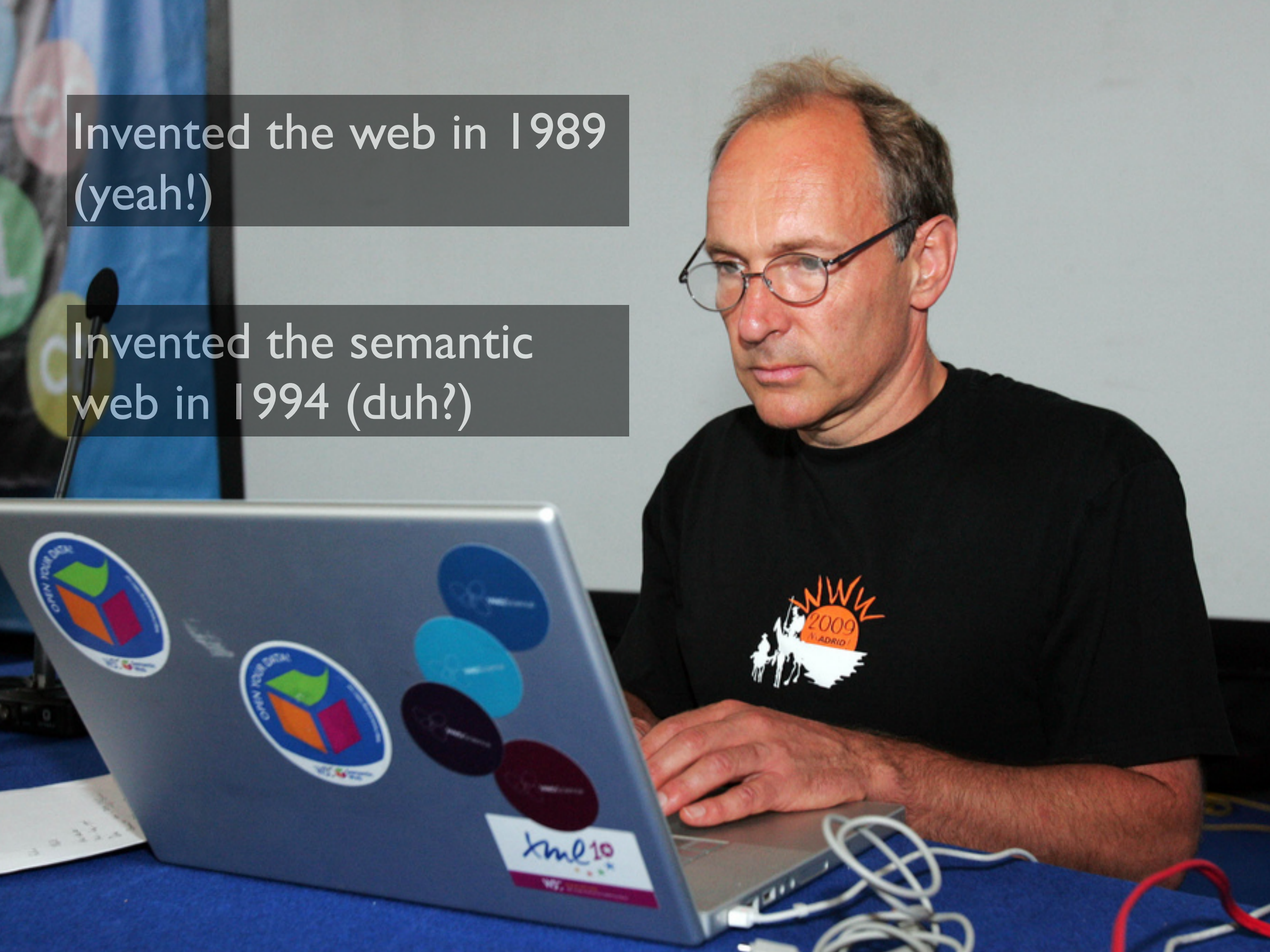

URL
identify what exists on the web.
http://my-site.fr
URI
identify, on the web, what exists.
http://animals.org/zebra#this
IRI
identify, on the web, in any language, what exists.
http://الحيوانات.tn/斑馬#this

WWW
2009
MADRID
xml 10

Invented the web in 1989
(yeah!)

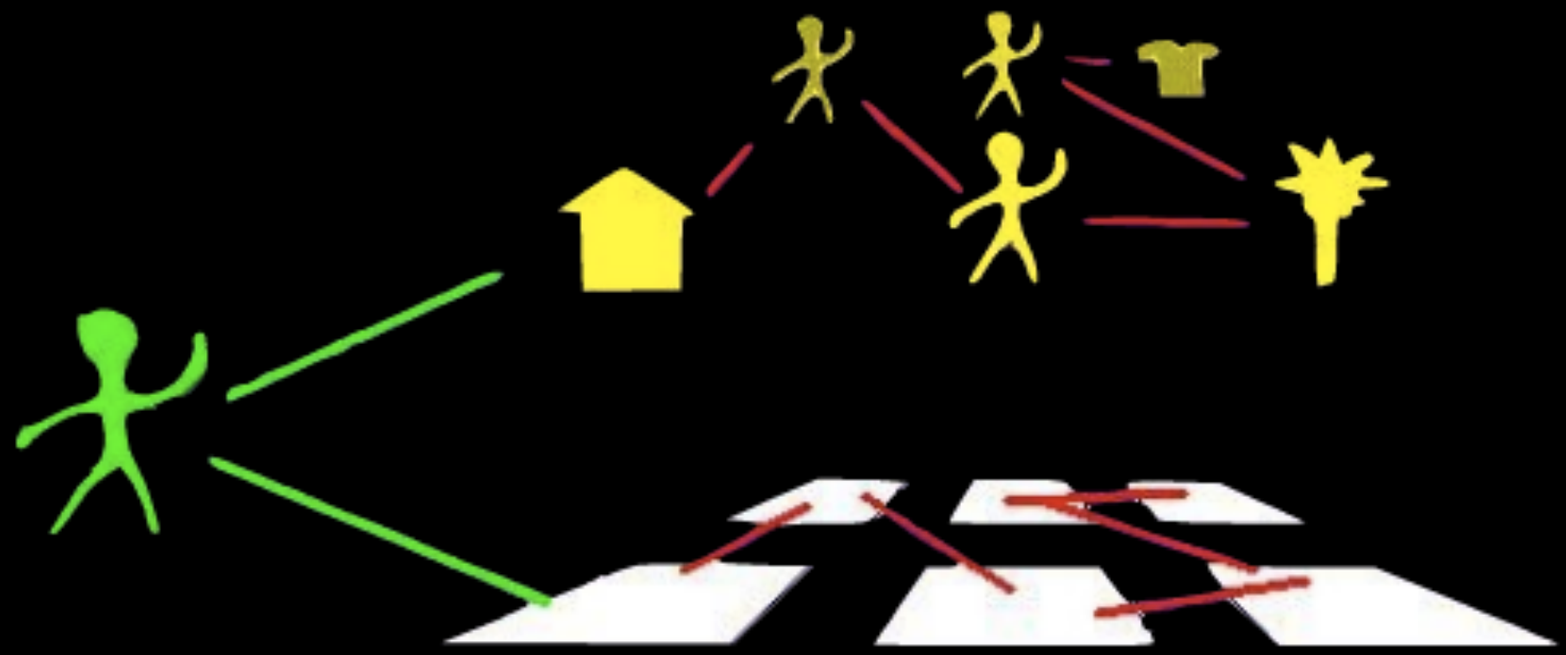
Invented the web in 1989 (yeah!)

Invented the semantic web in 1994 (duh?)



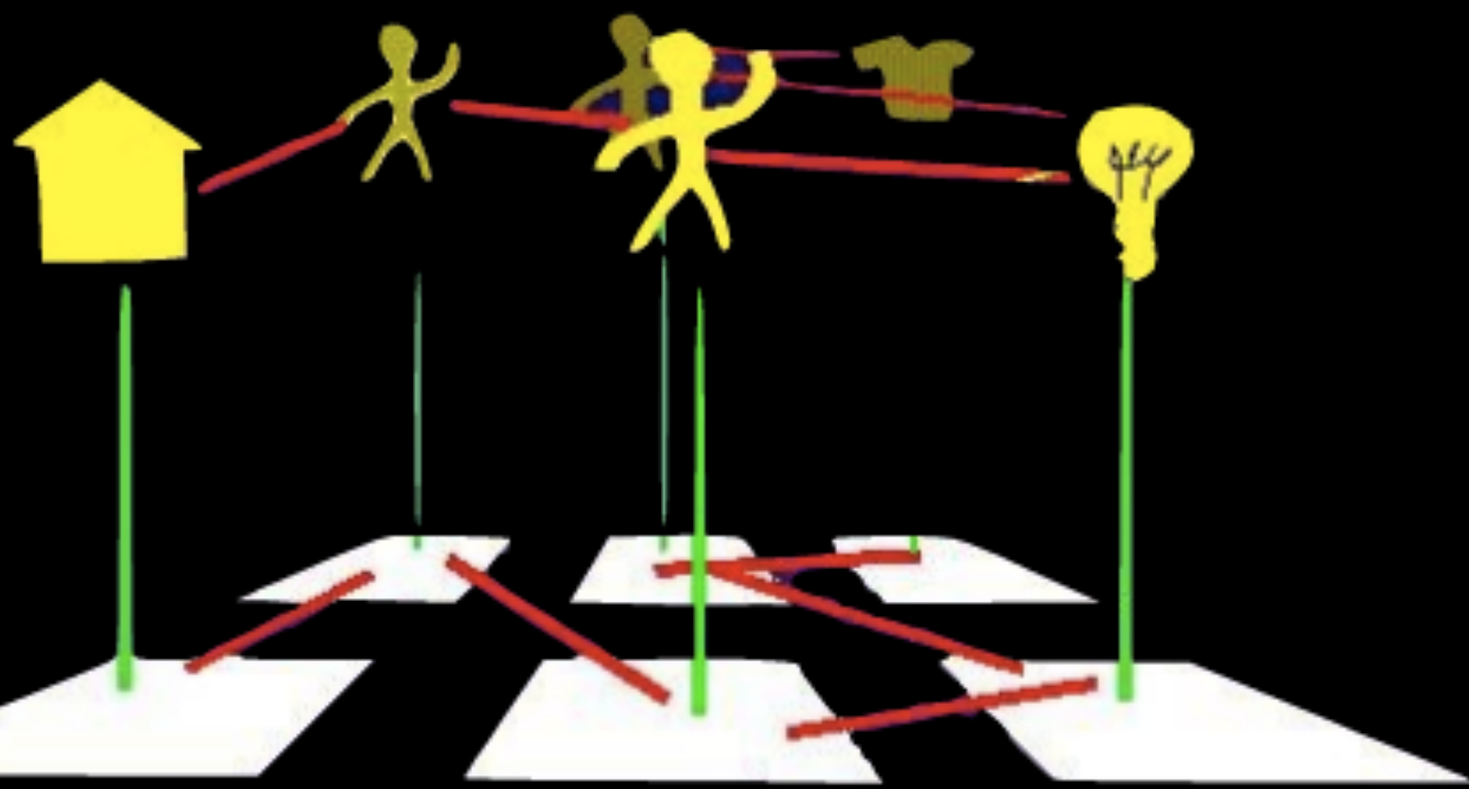
"To a computer, then, the web is a **flat**, **boring** world devoid of **meaning**"

Tim Berners Lee, http://www.w3.org/Talks/WWW94Tim/



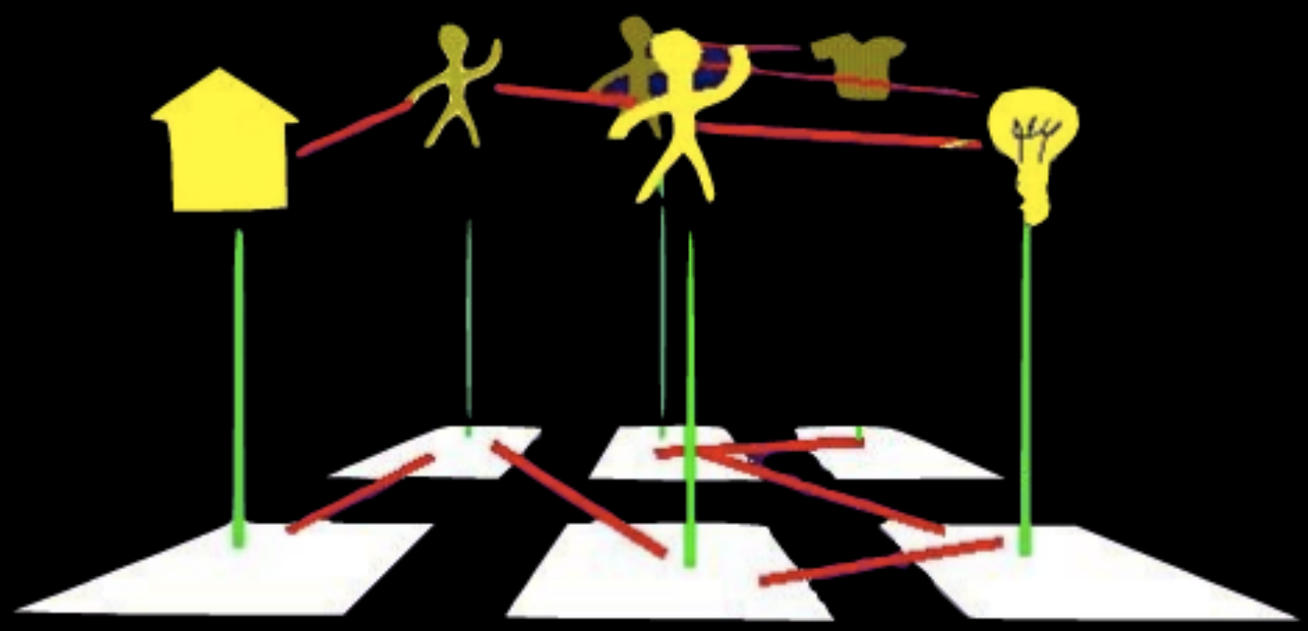
"This is a pity, as in fact **documents** on the web describe **real objects** and imaginary **concepts**, and give particular **relationships** between them"

Tim Berners Lee, http://www.w3.org/Talks/WWW94Tim/



"Adding semantics to the web involves two things: allowing **documents** which have information in **machine-readable** forms, and allowing **links** to be created with **relationship values**."

Tim Berners Lee, http://www.w3.org/Talks/WWW94Tim/



"The Semantic Web is not a separate Web but an **extension** of the current one, in which information is given well-defined **meaning**, better enabling **computers and people** to work in cooperation."

Tim Berners Lee, http://www.w3.org/Talks/WWW94Tim/

The Web

- A principle: hypertext
- A protocol: HTTP
- An identification scheme: URNs/URIs
- A language: HTML

The Semantic Web

- A principle: hypertext
- A protocol: HTTP
- An identification scheme: URNs/URIs
- Languages: HTML, RDF, OWL, SPARQL

The Semantic Web

- Is an extension of the Web through standards by the World Wide Web Consortium (W3C)

- The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries

Who is using the Semantic Web?

STLAB
SEMANTIC TECHNOLOGY LABORATORY

DATA.GOV
EMPOWERING PEOPLE

IBM

The New York Times

DBpedia

NASA

YAHOO!

ELSEVIER

data.gov.uk BETA
Opening up government

facebook

Google

O'REILLY
Spreading the knowledge of innovators.

NeOn
Foundation

protégé

Consiglio Nazionale delle Ricerche

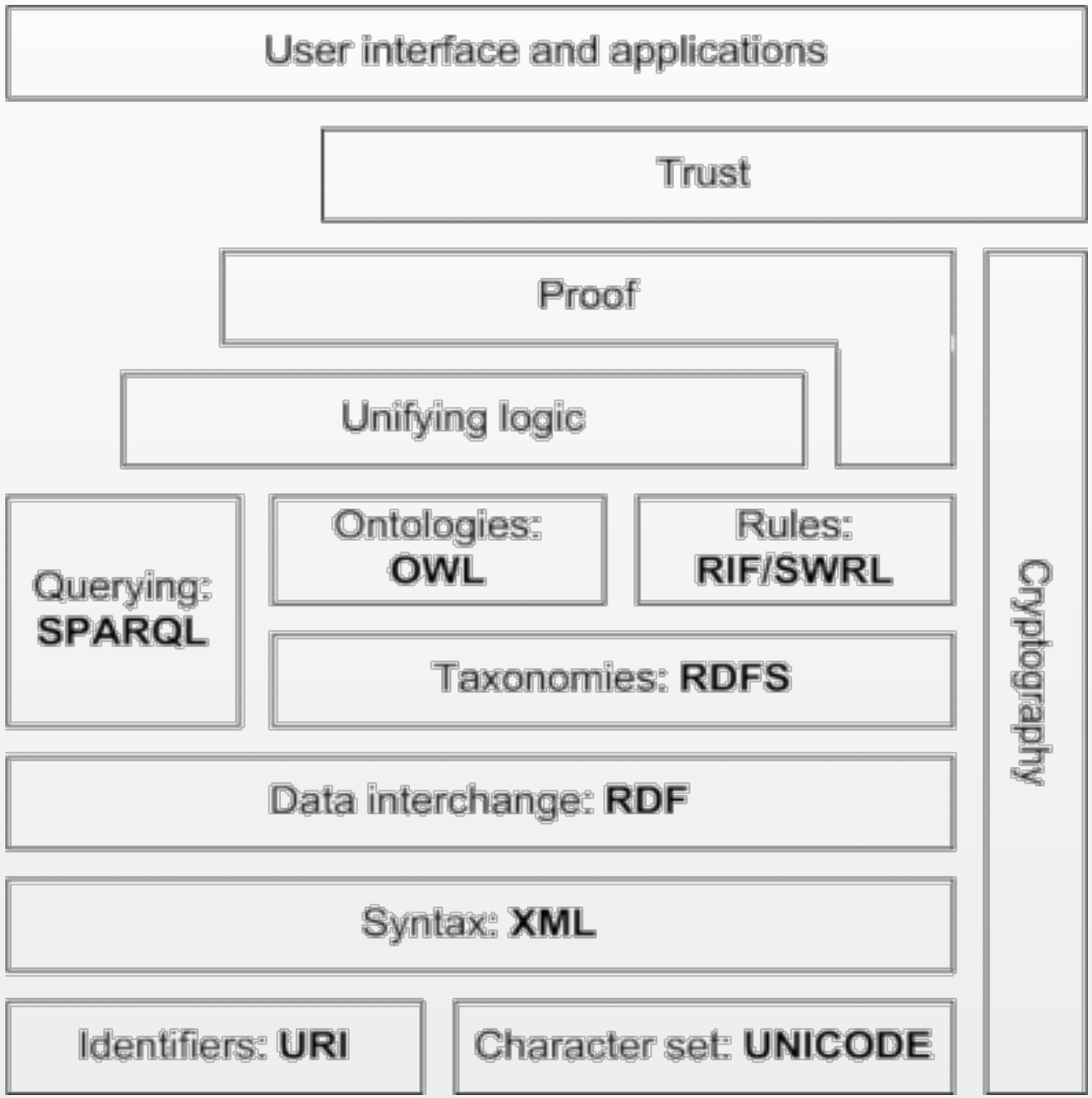
BBC

Smithsonian

the Gene Ontology

CNR
ISTC
ISTITUTO DI SCIENZE E TECNOLOGIE DELLA COGNIZIONE

The W3C "Layer Cake"



URIs and the Web of Things

- URIs (Unique Resource Identifiers) are used to identify things (also called entities) in the real world

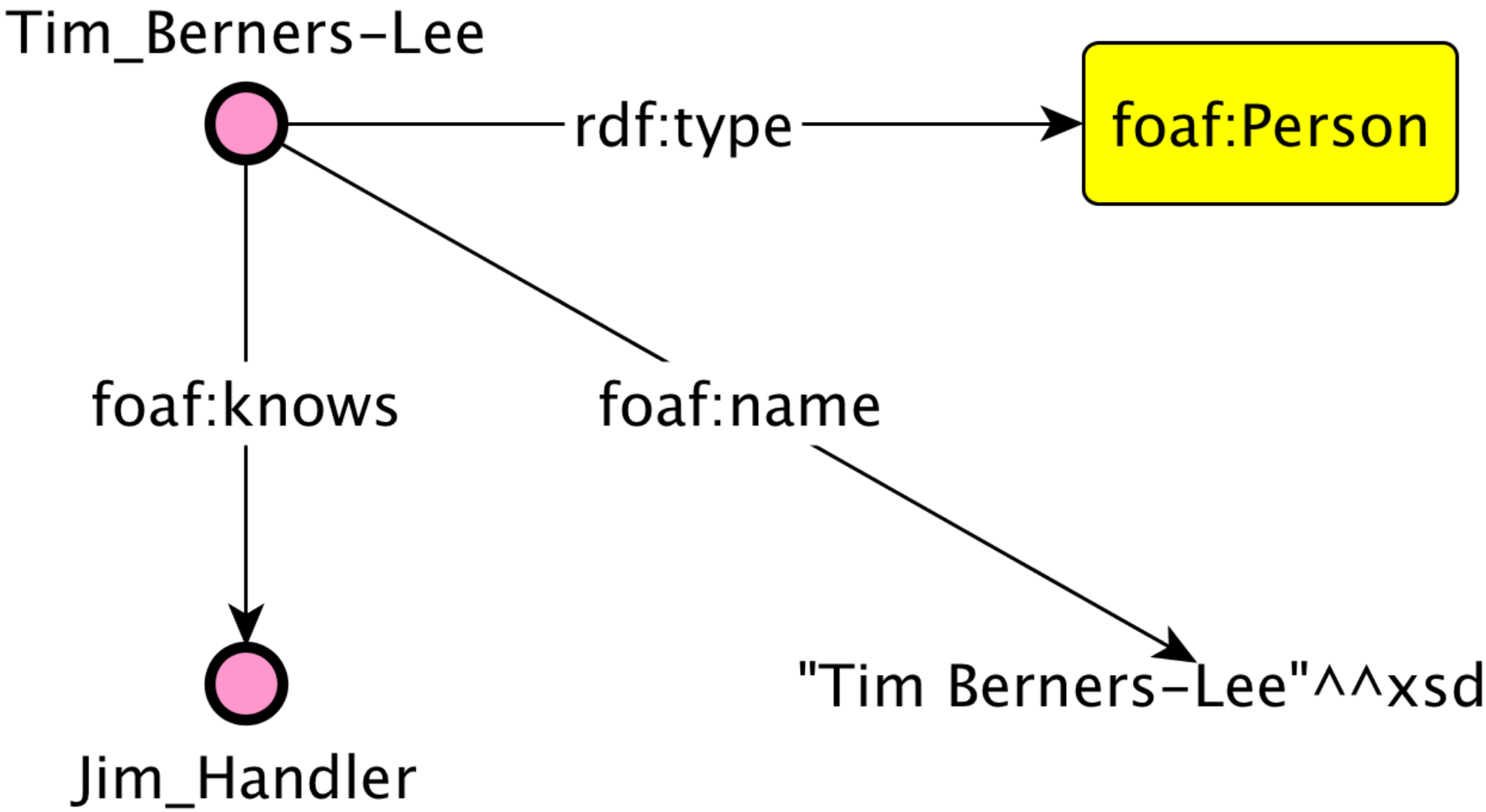
- For instance: people, places, events, companies, products, movies, etc.

The RDF model

Abstract
data representation

RDF is used to describe relationships between objects, identified by their URIs

RDF Example



:	http://www.foo.org/person/
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
foaf:	http://xmlns.com/foaf/0.1/

RDF serialization

Formal language
for data representation

As XML:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

Others, ex: N3:

```
<http://www.example.com/alice/posts/trouble_with_bob>
    <http://purl.org/dc/elements/1.1/title> "The Trouble with Bob";
    <http://purl.org/dc/elements/1.1/creator> "Alice" .
```

SPARQL Sample

Query language to RDF

```
prefix db:      <http://dbpedia.org/resource/>
prefix dbo:     <http://dbpedia.org/ontology/>
prefix foaf:    <http://xmlns.com/foaf/0.1/>

select distinct ?name, ?website, ?abstract where {
  ?software a                dbo:Software .
  ?software dbo:license      db:LGPL .
  ?software dbo:genre        db:Enterprise_Content_Management .
  ?software foaf:name        ?name .
  ?software foaf:homepage ?website .
  ?software dbo:abstract ?abstract .
  FILTER ( lang(?abstract) = "en" ) .
}
```

OWL example

Formal language
for automated reasoning

If:

```
:Novel              rdf:type owl:Class.
:Short_Story        rdf:type owl:Class.
:Poetry             rdf:type owl:Class.
:Literature rdf:type owl:Class;
   owl:unionOf (:Novel :Short_Story :Poetry).

<myWork> rdf:type :Novel .
```

then the following holds, too:

```
<myWork> rdf:type :Literature .
```

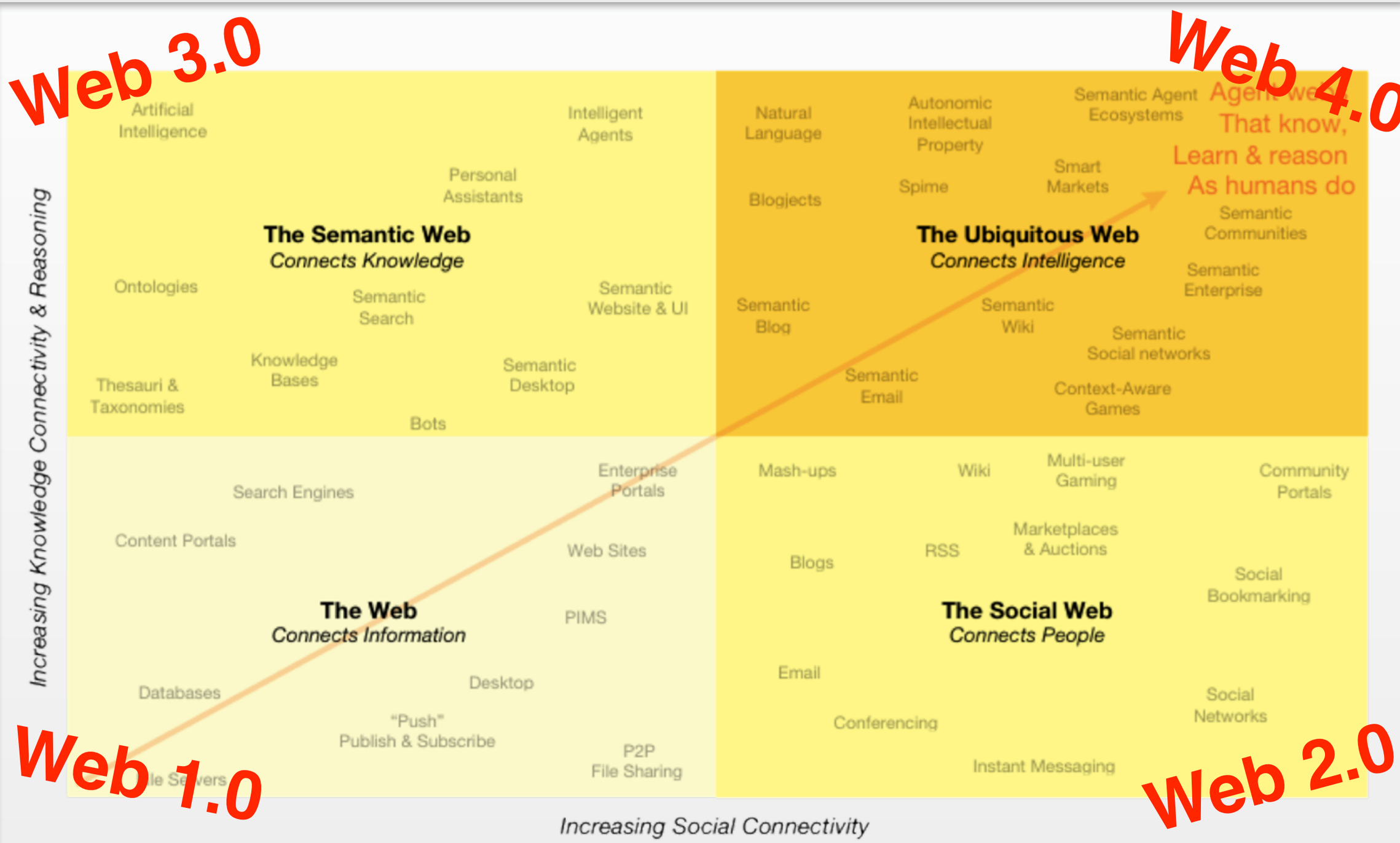
Historical perspective

- From web 1.0: web of sites and pages, aka the World Wide Web

- To web 2.0: web of people and of participation, aka the Social Web (Blogs, RSS, tags, Facebook, Wikipedia, etc.)

- To web 3.0: web of data, of meaning and connected knowledge, aka the Semantic Web

Web Evolution

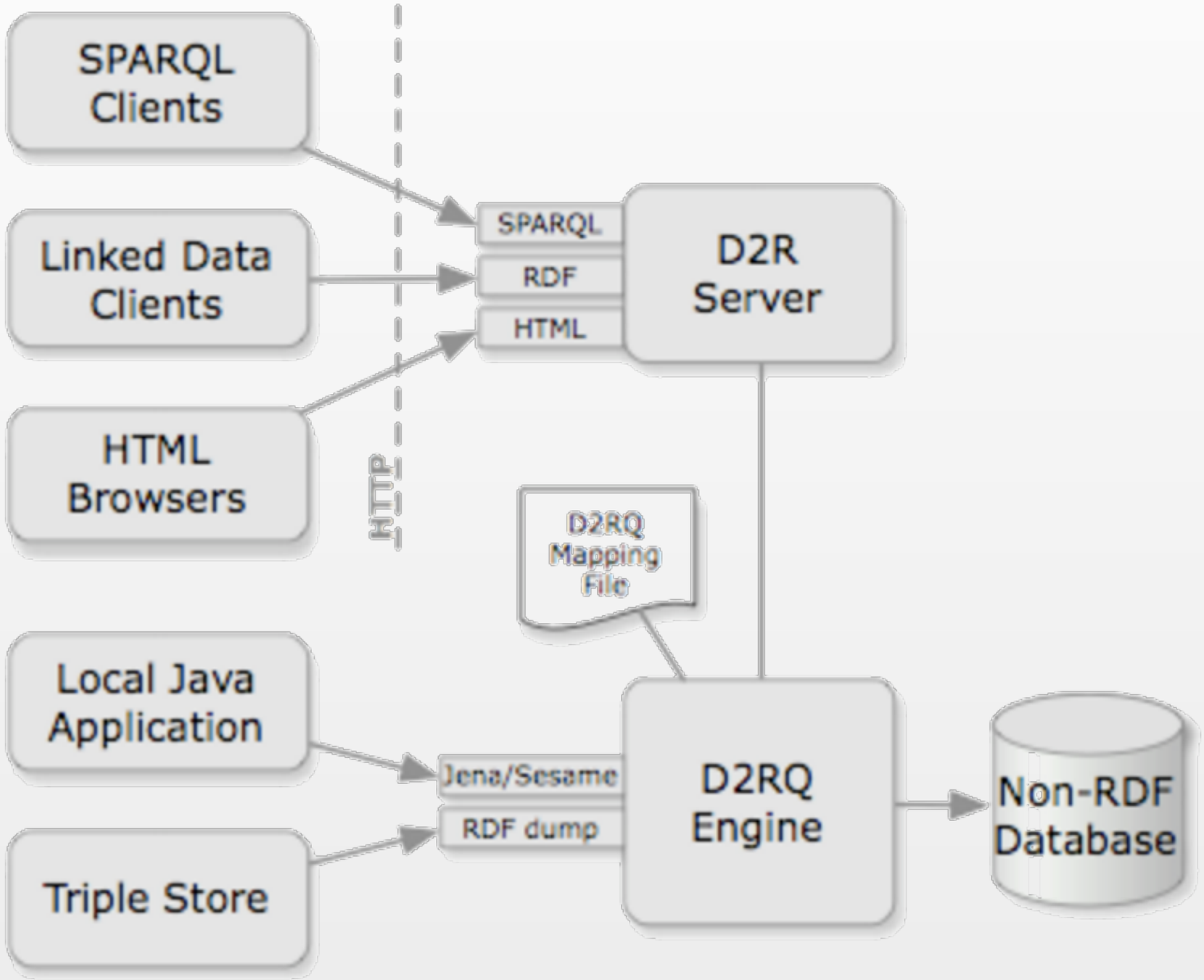


Where and how to find, access and use semantic web data and schemas?

RDF views over non-RDF sources

- Accessing non-RDF sources (e.g., relational databases) with SPARQL and as Linked Data
- Recipes based on **mapping configurations** to import data from relational databases
  - the mapping specify ho to interpret non-RDF data as RDF
- Example: D2R Server
  - a tool for publishing the content of relational databases on the Semantic Web, a global information space consisting of Linked Data.

D2R Server Architecture



Knowledge Extraction

- HTML scraping and natural language processing (NLP) techniques to extract semantic information from existing content / sites

- Generic solutions: OpenCalais, Zemanta, Apache Stanbol, NERD, FOX, FRED

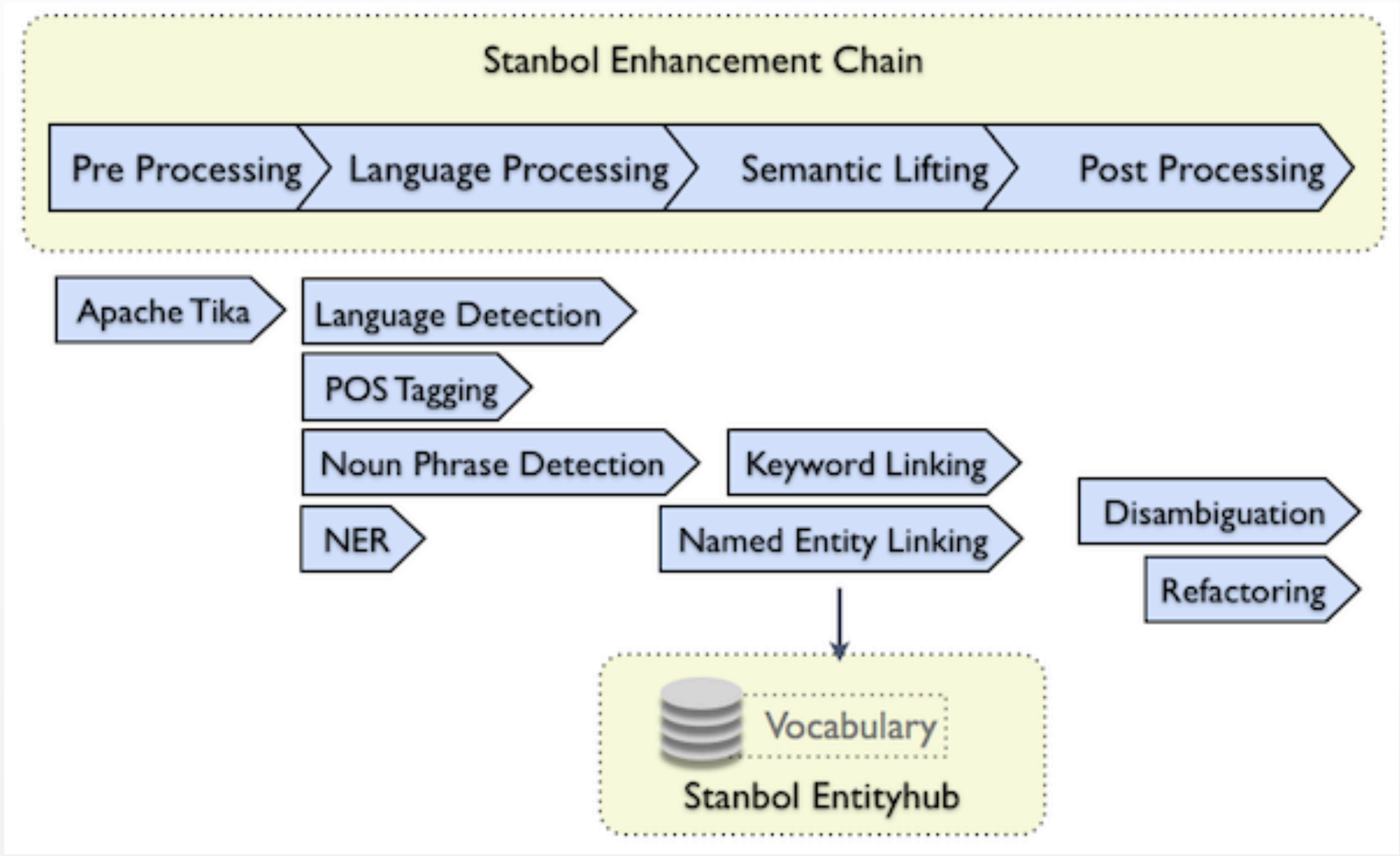
- Pro: no need to change existing content
- Con: error prone, needs human checks

Apache Stanbol

- Apache Stanbol provides a set of reusable components for semantic content management.

- Apache Stanbol's intended use is to extend traditional content management systems with semantic services.
  - e.g., semantic enhancement of natural language text, named entity extraction and linking

Apache Stanbol enhancement chain



Miles Davis was an american jazz musician.

dbpedia:Miles_Davis

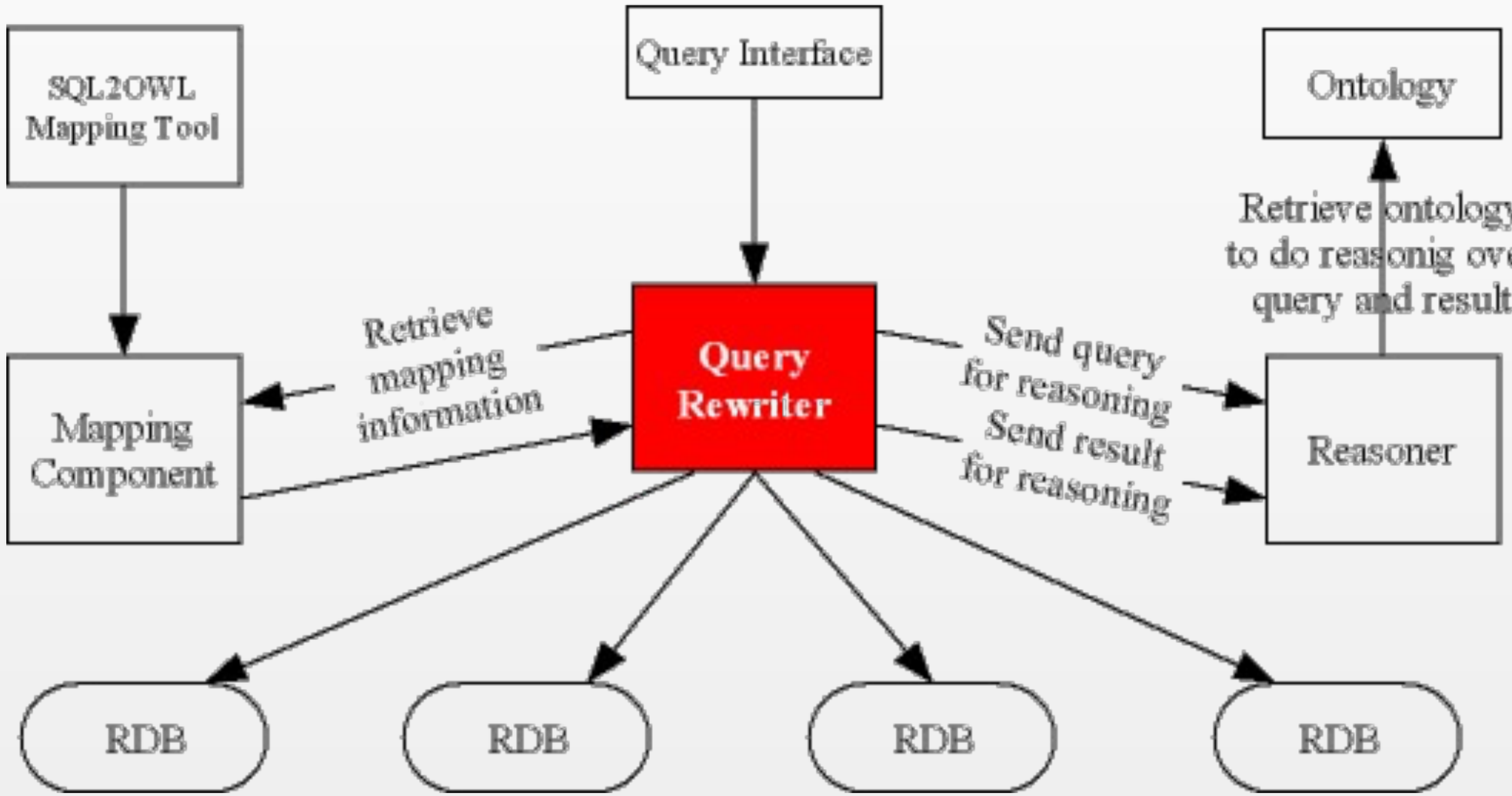
dbpedia:American_jazz

dbpedia:Musician

Reasoning

- Infer new knowledge from asserted one
- Based on logic axioms defined in a ontology modelled as OWL
- Reasoning can be used for query expansion in RDF knowledge bases

Reasoner based architecture

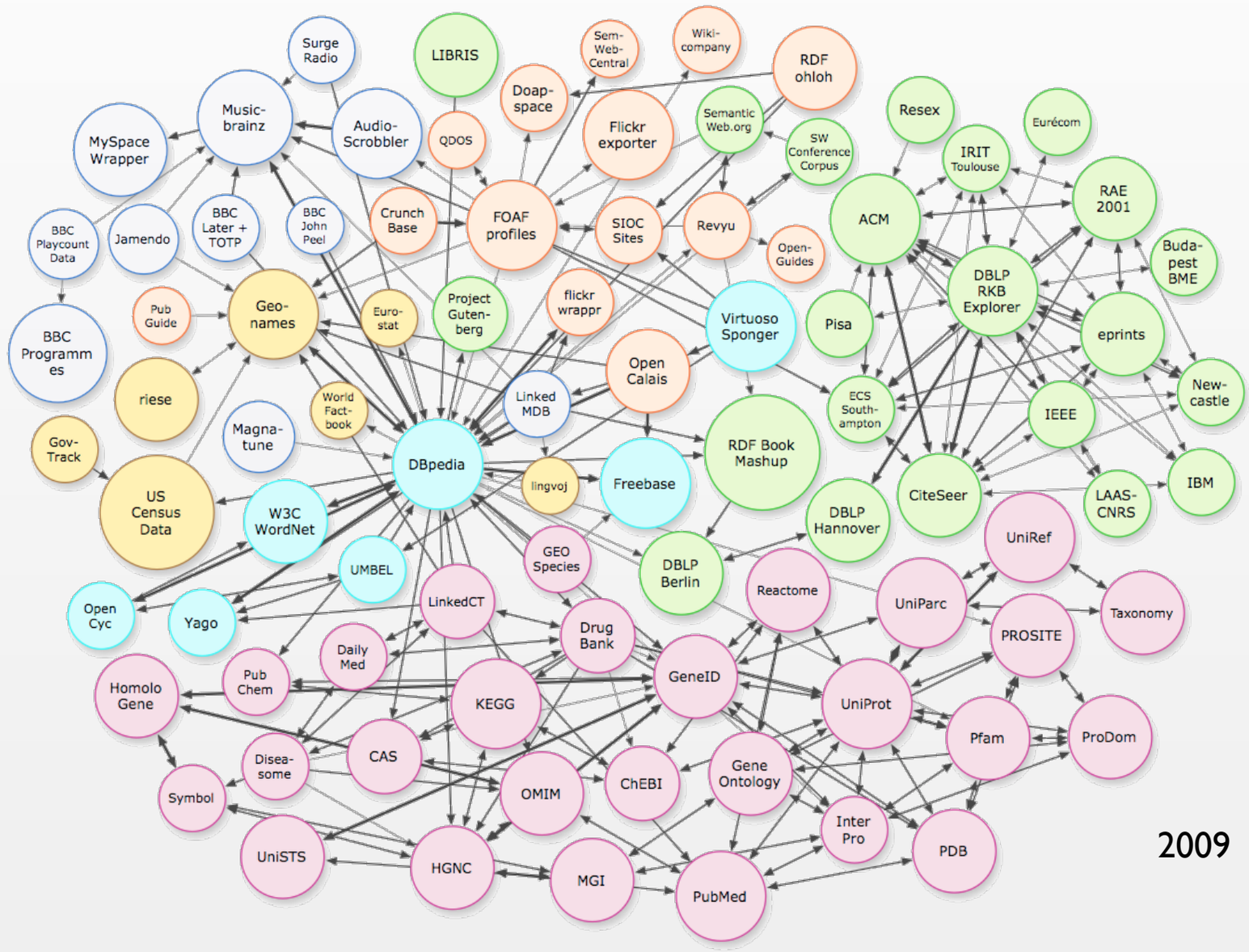


KB reuse

- Linked Open Data: (usually large) data repositories available on the web (for free or not), expressed using the RDF model

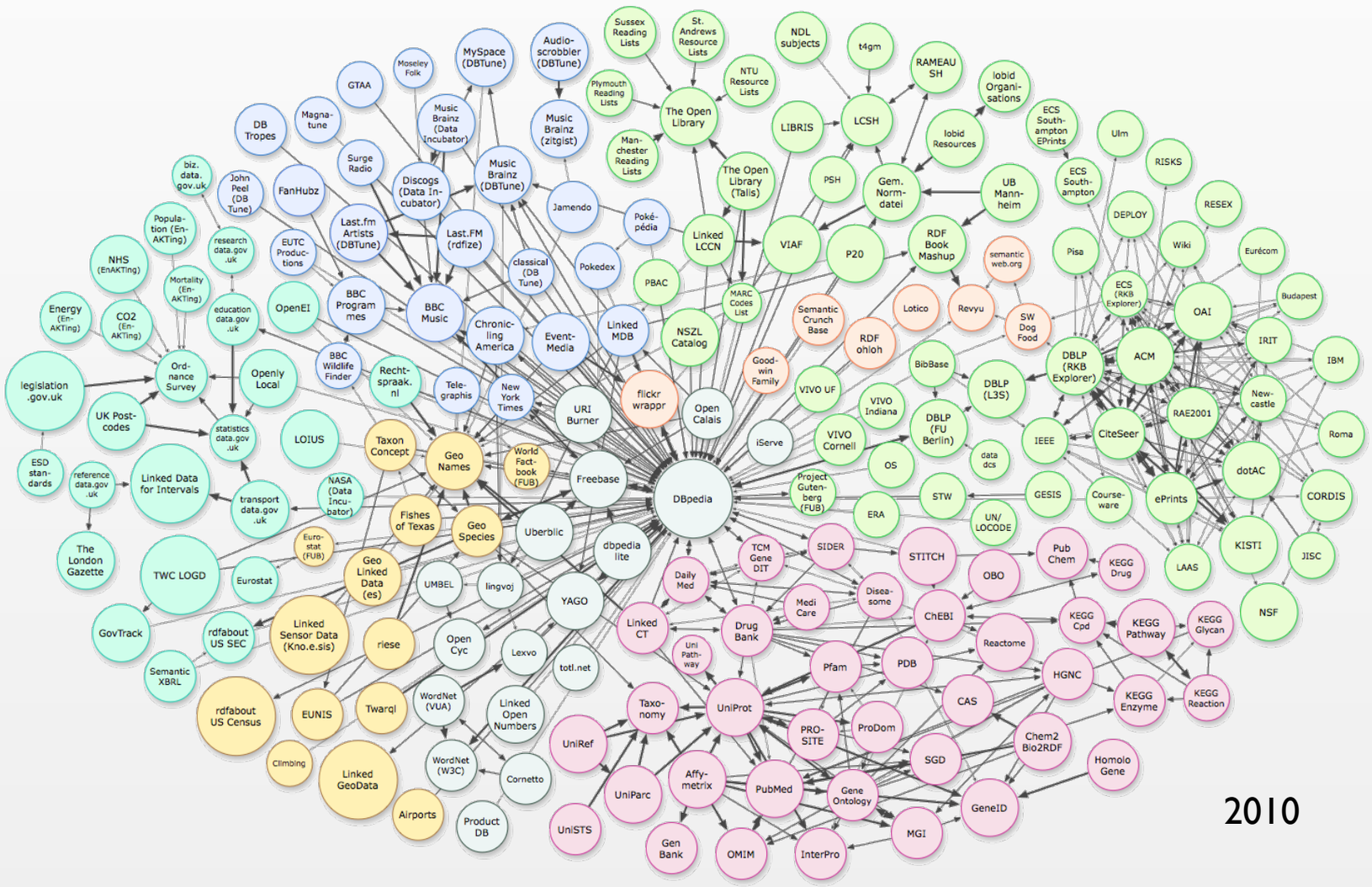
- Interoperability between these repositories (how to align their ontologies and entity names?) is usually partial

LOD Cloud



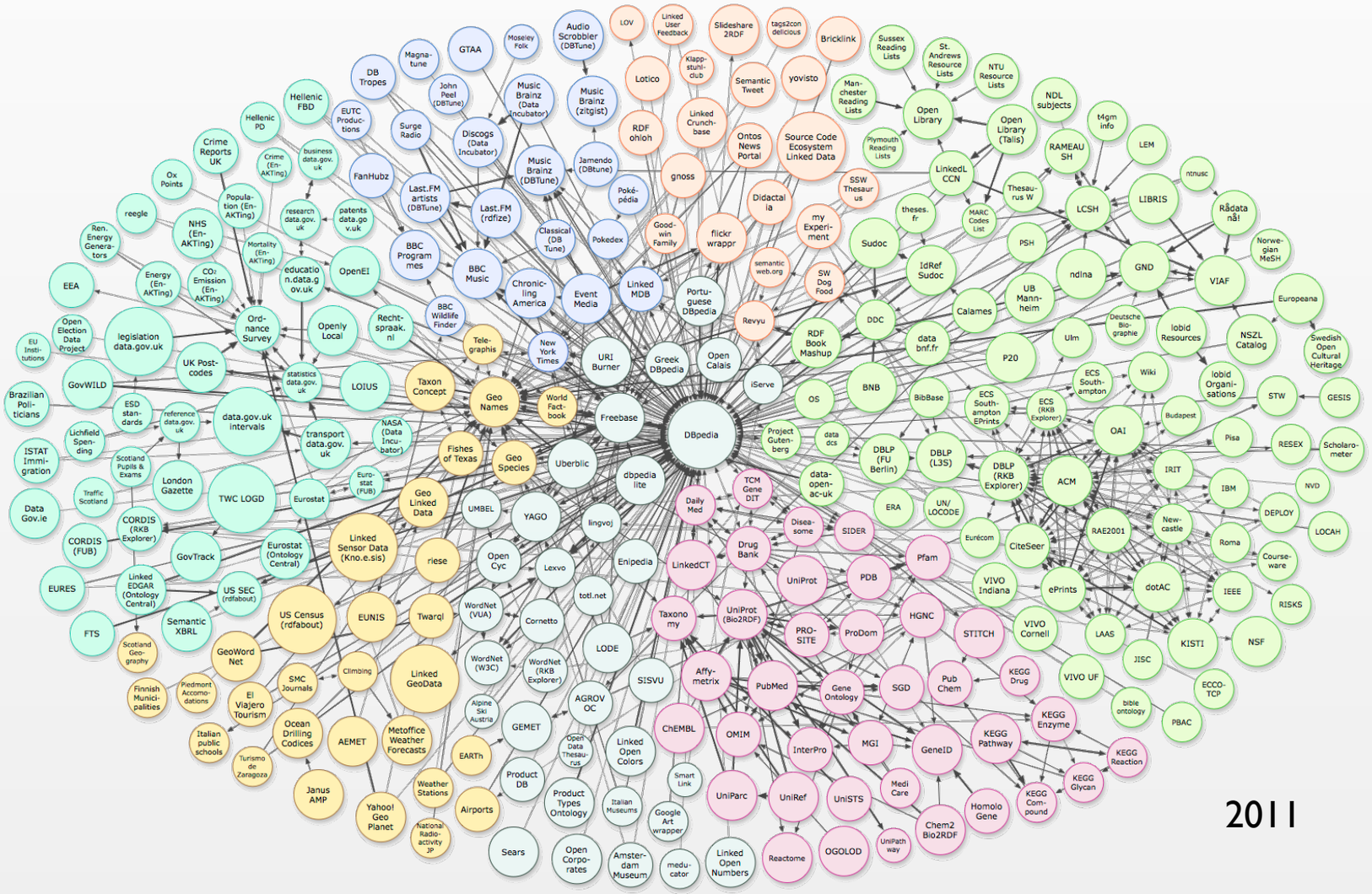
2009

LOD Cloud



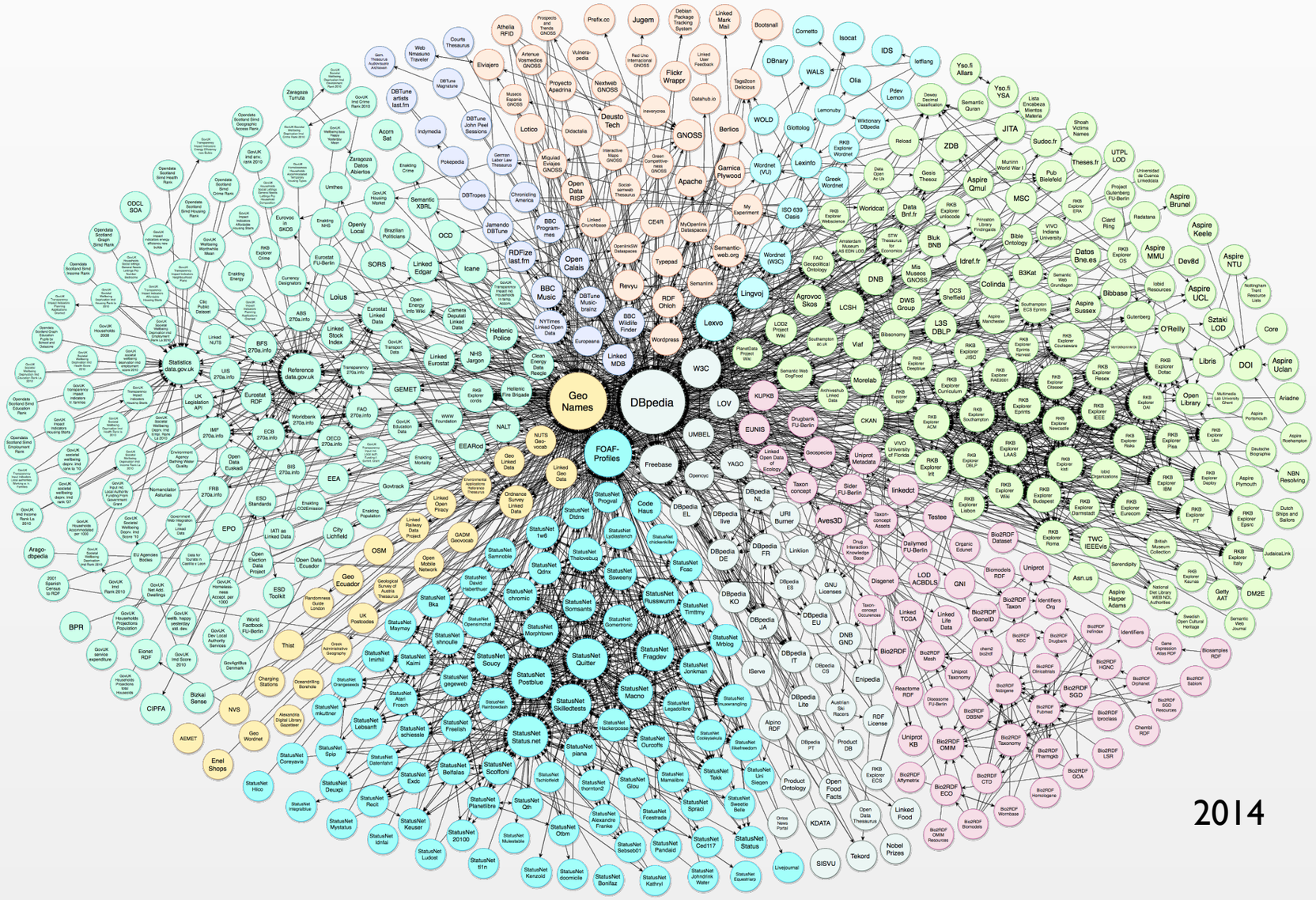
2010

LOD Cloud



2011

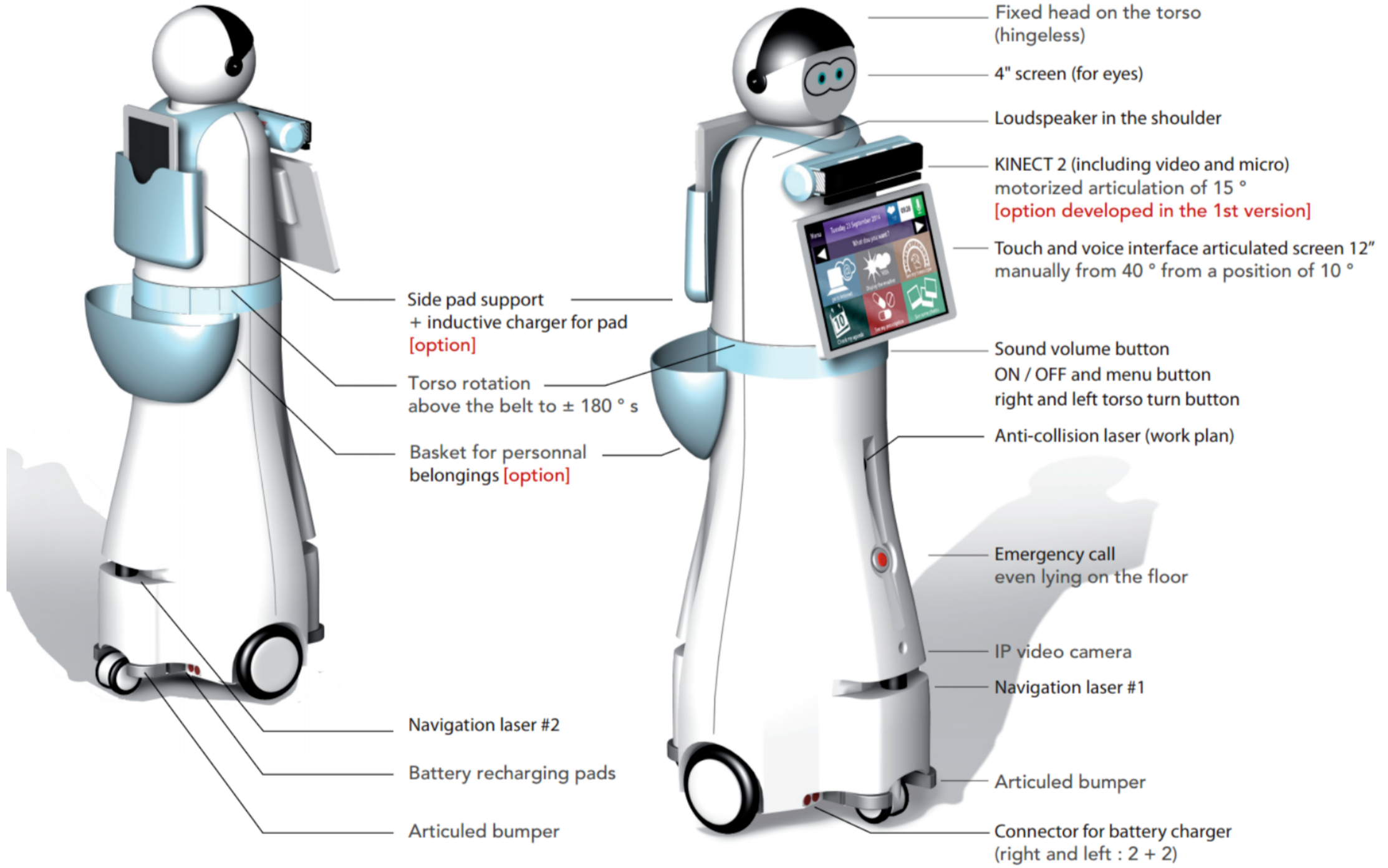
LOD Cloud


Geo Names
DBpedia
FOAF-Profiles
Linked MDB
W3C
Lexvo
Lingvoj
Open Calais
BBC Music
RDFize last.fm
Revyu
RDF Ohloh
Typepad
Wordpress
Semanlink
Semantic-web.org
Open Data RISP
Apache
Garnica Plywood
Berlios
GNOSS
Deusto Tech
Nextweb GNOSS
Proyecto Apadrina
Lotico
Didactalia
Miguiad Evilajes GNOSS
Elviajero
Flickr Wrappr
Linked Mark Mail
Bootsnall
Cometo
Isocat
IDS
Olia
WALS
DBnary
Lexinto
Wordnet (VU)
WOLD
Glottolog
Lemonuby
Pdev Lemon
Yso.fi Allars
Yso.fi YSA
JITA
Sudoc.fr
Theses.fr
UTPL LOD
Aspire Qmul
MSC
Pub Bielefeld
Aspire Brunel
Aspire Keele
Aspire NTU
Dev8d
Aspire MMU
Aspire UCL
Sztaki LOD
Core
DOI
Aspire Uclan
Ariadne
Libris
Open Library
O'Reilly
Gutenberg
Bibbase
Aspire Sussex
Datos Bne.es
B3Kat
Colinda
Idref.fr
Bluk BNB
Data Bnf.fr
Worldcat
Gesis Thesoz
ZDB
Reload
Lexvo
Lingvoj
DNB
LCSH
DWS Group
L3S DBLP
DCS Sheffield
Mis Museos GNOSS
Agrovoc Skos
FAO Geopolitical Ontology
Bibsonomy
Viaf
Morelab
Semantic Web DogFood
LOD2 Project Wiki
Southampton ac.uk
Wordnet (W3C)
Semantic-web.org
Typepad
Revyu
RDF Ohloh
Semanlink
Open Calais
BBC Music
DBTune Music-brainz
BBC Wildlife Finder
NYTimes Linked Open Data
Europeana
Linked MDB
W3C
Lexvo
Wordpress
Hellenic Police
Linked Edgar
Icane
SORS
OCD
Semantic XBRL
Jamendo DBTune
BBC Programmes
Linked Crunchbase
CE4R
MyOpenlink Dataspaces
OpenlinkSW Dataspaces
Open Data RISP
Apache
Garnica Plywood
Berlios
GNOSS
Deusto Tech
Nextweb GNOSS
Proyecto Apadrina
Lotico
Didactalia
Miguiad Evilajes GNOSS
Elviajero
Flickr Wrappr
Linked User Feedback
Datahub.io
Tags2con Delicious
WOLD
Glottolog
Lexinto
Wordnet (VU)
Greek Wordnet
Lemonuby
Wiktionary DBpedia
Pdev Lemon
Dewey Decimal Classification
Semantic Quran
Yso.fi YSA
Yso.fi Allars
Lista Encabeza Mientos Materia
JITA
Sudoc.fr
Theses.fr
Shoah Victims Names
UTPL LOD
Universidad de Cuenca Linkeddata
Project Gutenberg FU-Berlin
Ciard Ring
Radatana
Aspire Brunel
Aspire Keele
Aspire NTU
Dev8d
Aspire MMU
Aspire UCL
Sztaki LOD
Core
DOI
Aspire Uclan
Ariadne
Libris
Open Library
O'Reilly
Gutenberg
Bibbase
Aspire Sussex
Datos Bne.es
B3Kat
Colinda
Idref.fr
Bluk BNB
Data Bnf.fr
Worldcat
Gesis Thesoz
ZDB
Reload
Aspire Qmul
MSC
Pub Bielefeld
Muninn World War I
Bible Ontology
Princeton Library Findingaids
Southampton ECS Eprints
Aspire Manchester
Semantic Web Grundlagen
VIVO Indiana University
RKB Explorer ERA
Data Open Ac Uk
Amsterdam Museum AB EDN LOD
STW Thesaurus for Economics
RKB Explorer Webscience
RKB Explorer unlocode
Data Bnf.fr
Worldcat
Greek Wordnet
Lexinto
Wordnet (VU)
Garnica Plywood
Berlios
GNOSS
Deusto Tech
Nextweb GNOSS
Proyecto Apadrina
Lotico
Didactalia
Miguiad Evilajes GNOSS
Elviajero
Flickr Wrappr
Linked User Feedback
Datahub.io
Tags2con Delicious
WOLD
Glottolog

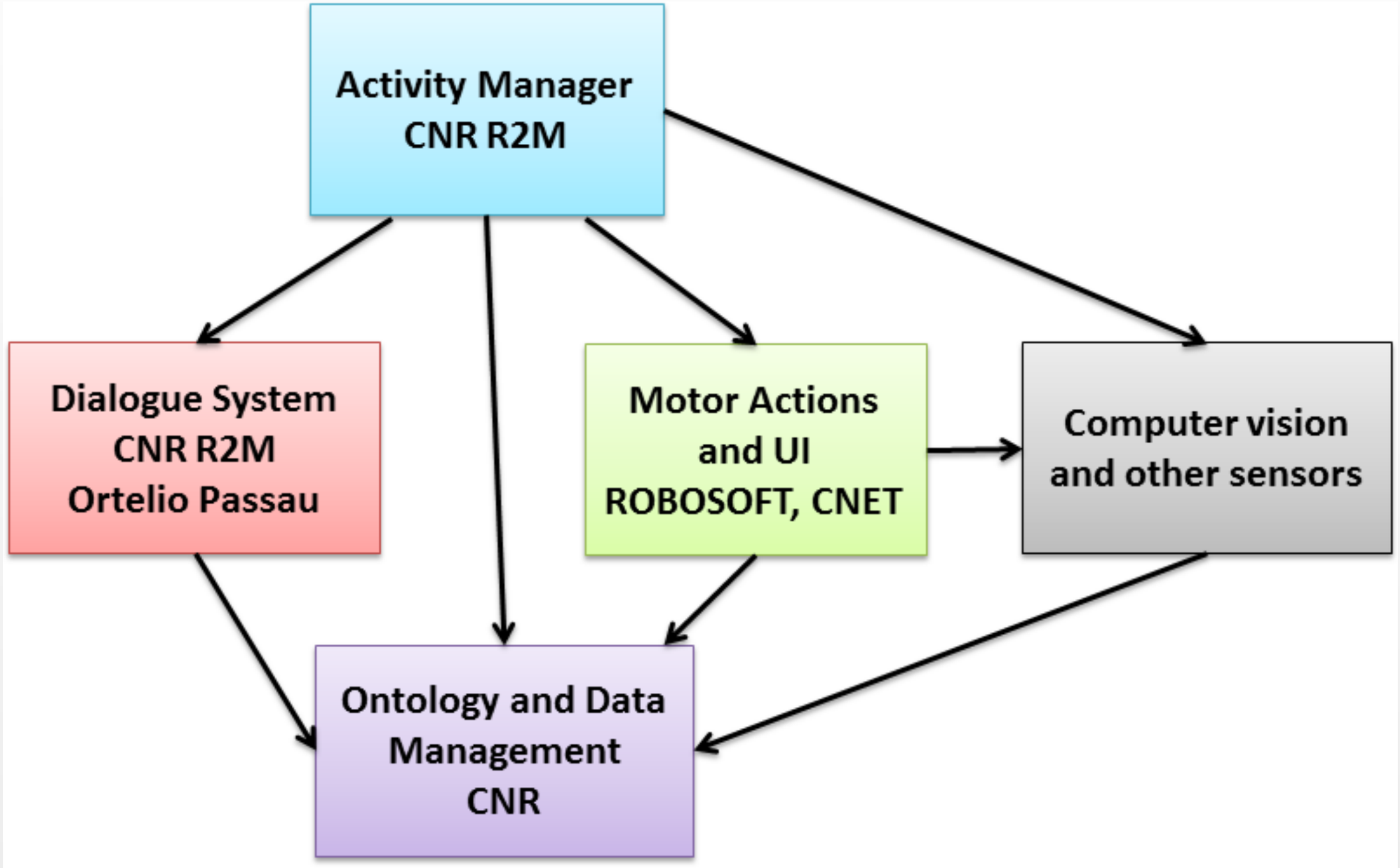
MARIO Project

- Managing Active and healthy aging with use of caRing servIce robots.
- MARIO is a **robot** for supporting older people affected by dementia
- It addresses the difficult challenges of loneliness, isolation and dementia in older persons through innovative and multi-faceted inventions delivered by service robots
- Uses Semantic Web and ontologies extensively

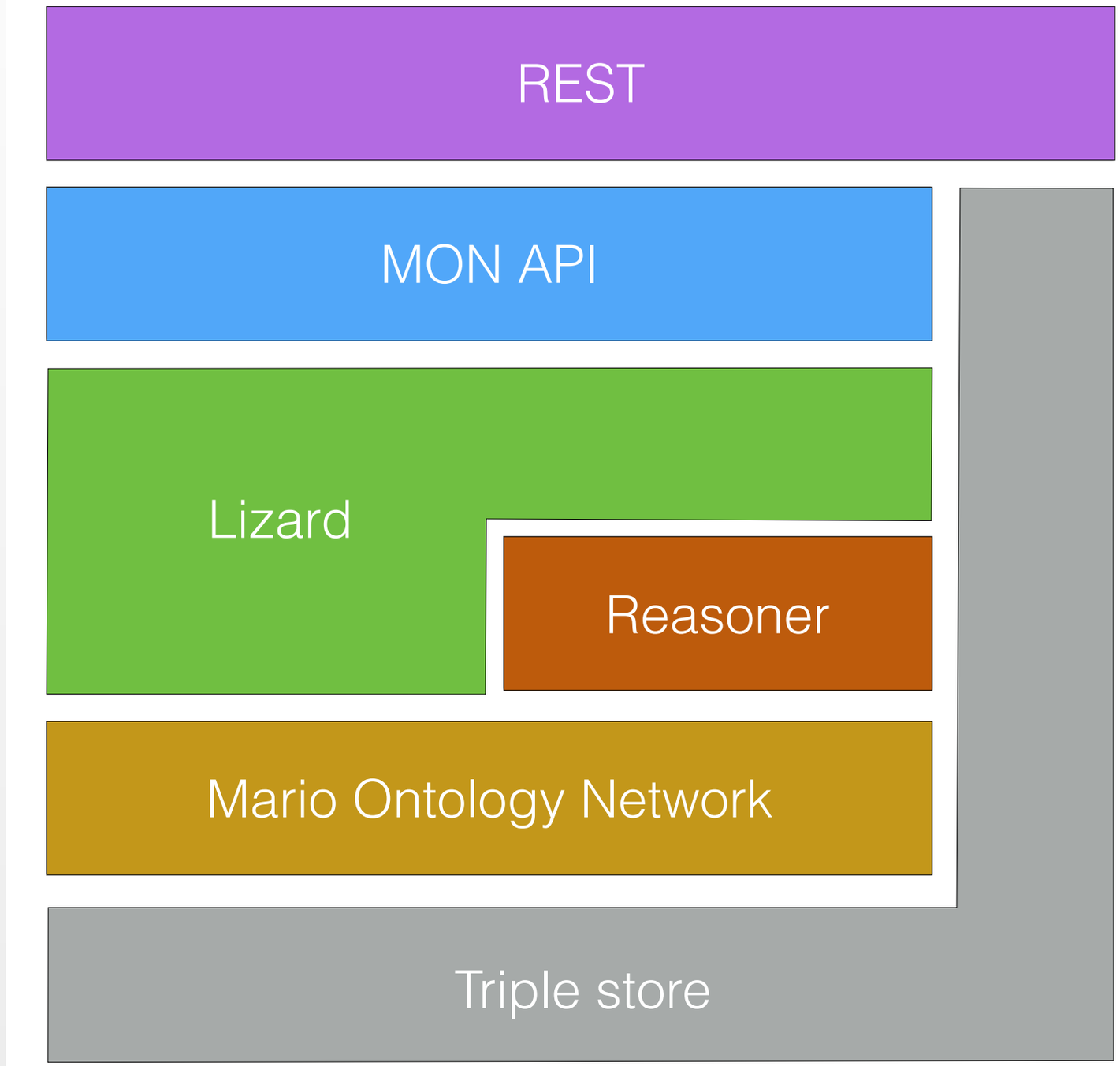
The MARIO robot



MARIO informal architecture



Ontology and Data Management System



Lizard: Ontology API generator

- Lizard is designed and implemented for automatically producing Java and REST API from MON
- Generated API
  - reflect the semantics of MON ontologies
  - enable access via a Knowledge Base modelled with the semantics of MON
- Key features
  - dynamic loading of ontologies to the MARIO Ontology Network (MON)
  - programmatic access to the Knowledge Base of MARIO via Java and HTTP REST
  - transparent access (with respect to client application) to the MON