Semantiska webben 2010

DFS/Gbg 100112

Olle Olsson
World Wide Web Consortium (W3C)
Swedish Institute of Computer Science (SICS)

With thanks to Ivan for many slides

Trends and forces:

- Technology
- Internet
- Web
- Semantic Web

Trend - Technology

Traditional

- programming, datastructures
- databases
- content, dissemination

Emerging

- stream-processing
- data/information generated continuously
- processed in real-time

Drivers

- sensor networks – Internet-of-Things
- messaging, blogging, micro-blogging

Challenge: how to process data effectively

Trend - Internet

Changed view on Internet

- Content-centric; data-centric; ..

Historical perspective

- telephony: circuit switching; focus on wires
- present Internet: packet switching; focus on hosts
- future Internet: content distribution; focus on data

Reasons include:

- efficiency
- security
- content adressability

Challenge: how to support what the web offers

Trend - Web

Emphasis on

- simplicity
- uniformity
- ubiquity

Strong drivers

- social networking
- growing volumes of data
- mobility
- growing spectrum of devices
- cost-efficiency!

Challenge: how to increase automation

Trend – Semantic Web

Emphasis on

- simplicity

Strong drivers

- web of data
- interoperability frameworks

Challenge: how to support major market needs

Semantics – what?

Semantics web is what?

- about “meaning” and automation

Semantics and meaning

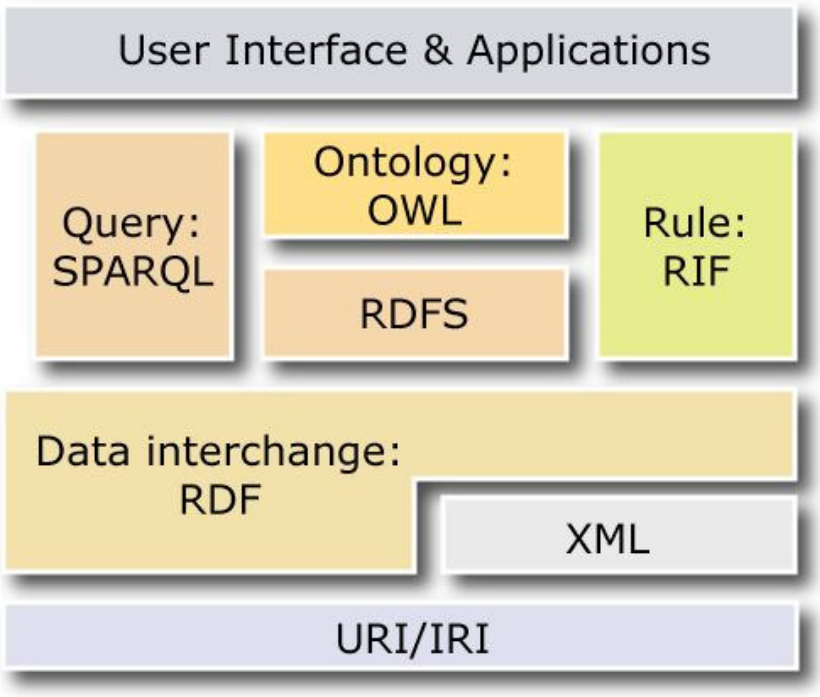
- linguistic/philosophical semantics
- intuitive / everyday / man-on-the-street

Pragmatic approach

- program “knows” what it can do with data
- meaning-as-use

No magic ... instead well-founded engineering

Semantic web building blocks



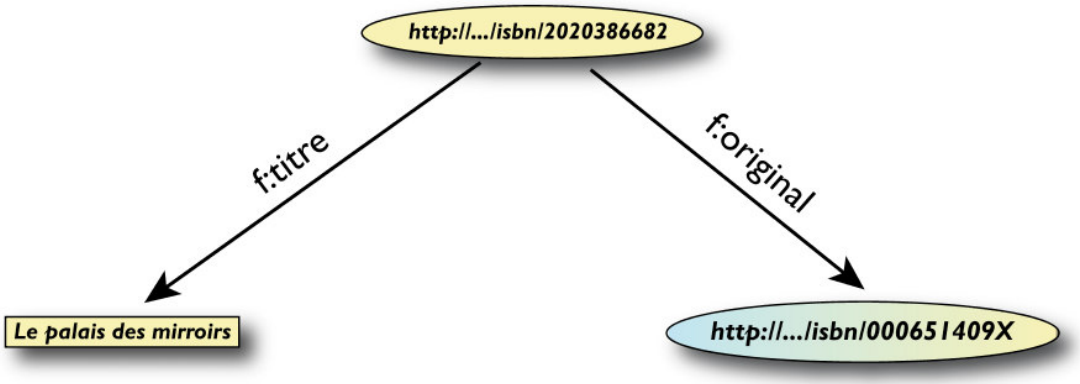
RDF – Resource Description Framework

- Basic data model – a "triple"
  - triple `(s,p,o)` is such that:
    - "`s`", "`p`", and "`o`" stand for "subject", "predicate", and "object", respectively
    - conceptually: "`p`" connects, or relates the "`s`" and "`o`"
- An example triple:

```
(
<http://…isbn…6682>,        # "Le palais des miroirs"
<http://…/original>,        # "is a derivative of the original"
<http://…isbn…409X>         # "The Glass Palace"
)
```

- *RDF* is a general model for such triples
  - machine readable formats like RDF/XML, Turtle, n3, RXR
- *and that's it!*

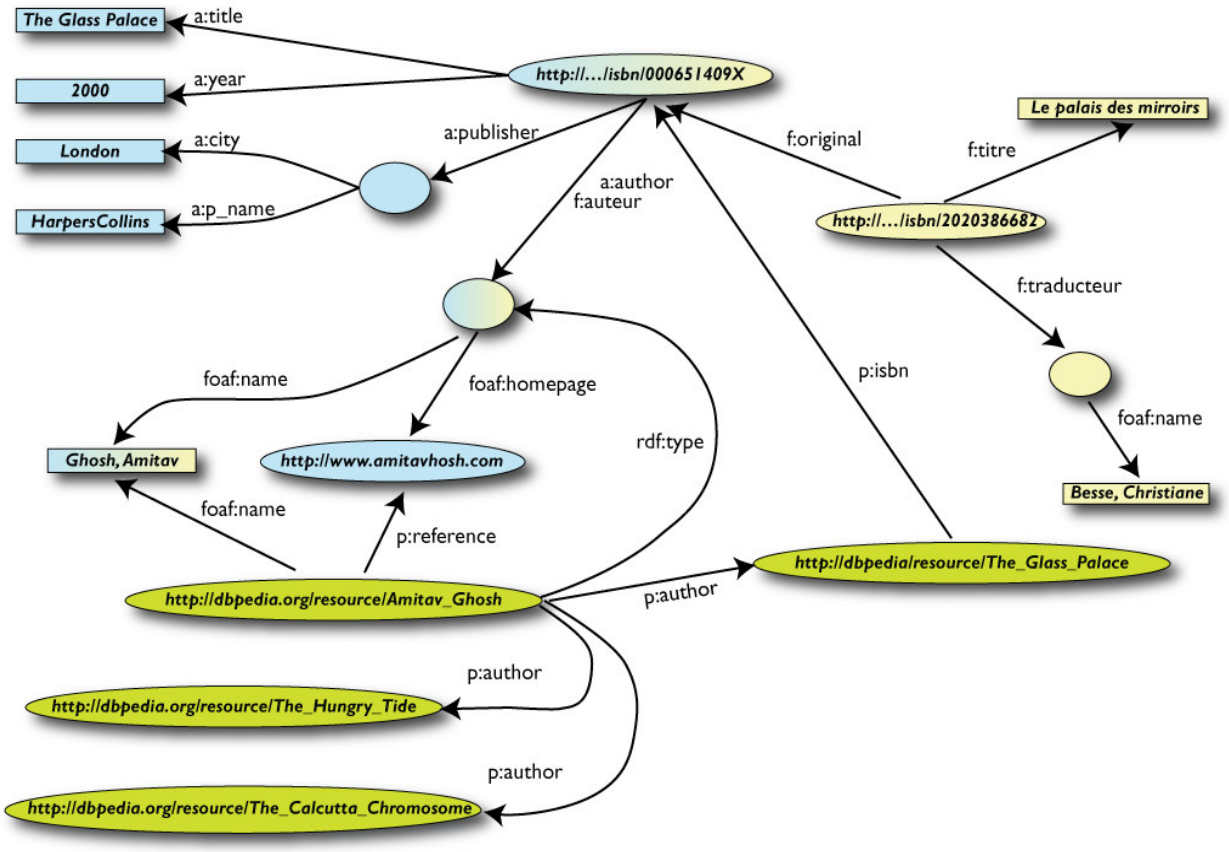
RDF Example




```
<rdf:Description rdf:about="http://…/isbn/2020386682">
    <f:titre xml:lang="fr">Le palais des mirroirs</f:titre>
    <f:original rdf:resource="http://…/isbn/000651409X"/>
</rdf:Description>
```

Set of triples form a graph – the *RDF graph*

RDF Graph



A thread that binds lots of of the issues together…

Linking Open Data Project

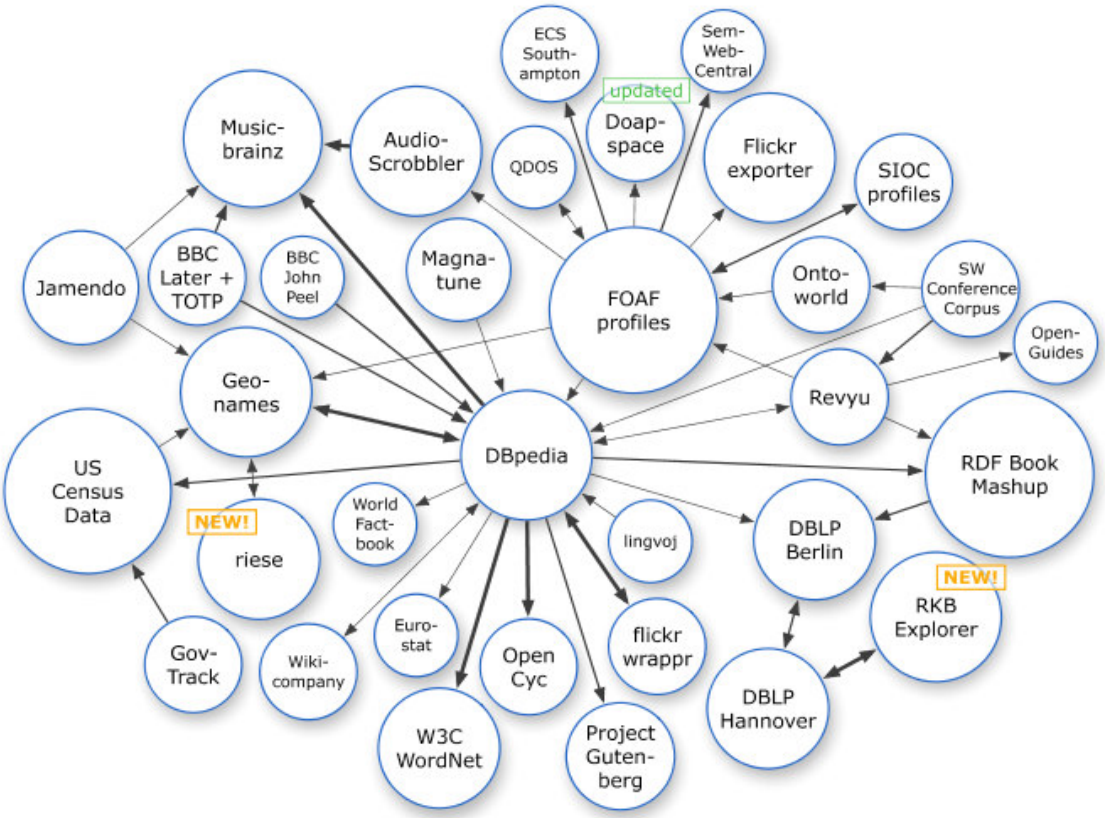
Goal:

- "expose" open datasets via RDF
- *set RDF links among the data items* from different datasets
  - a typical example is to set an owl:sameAs between two items in different datasets that refer to the same "thing"
- set up query endpoints (usually SPARQL)

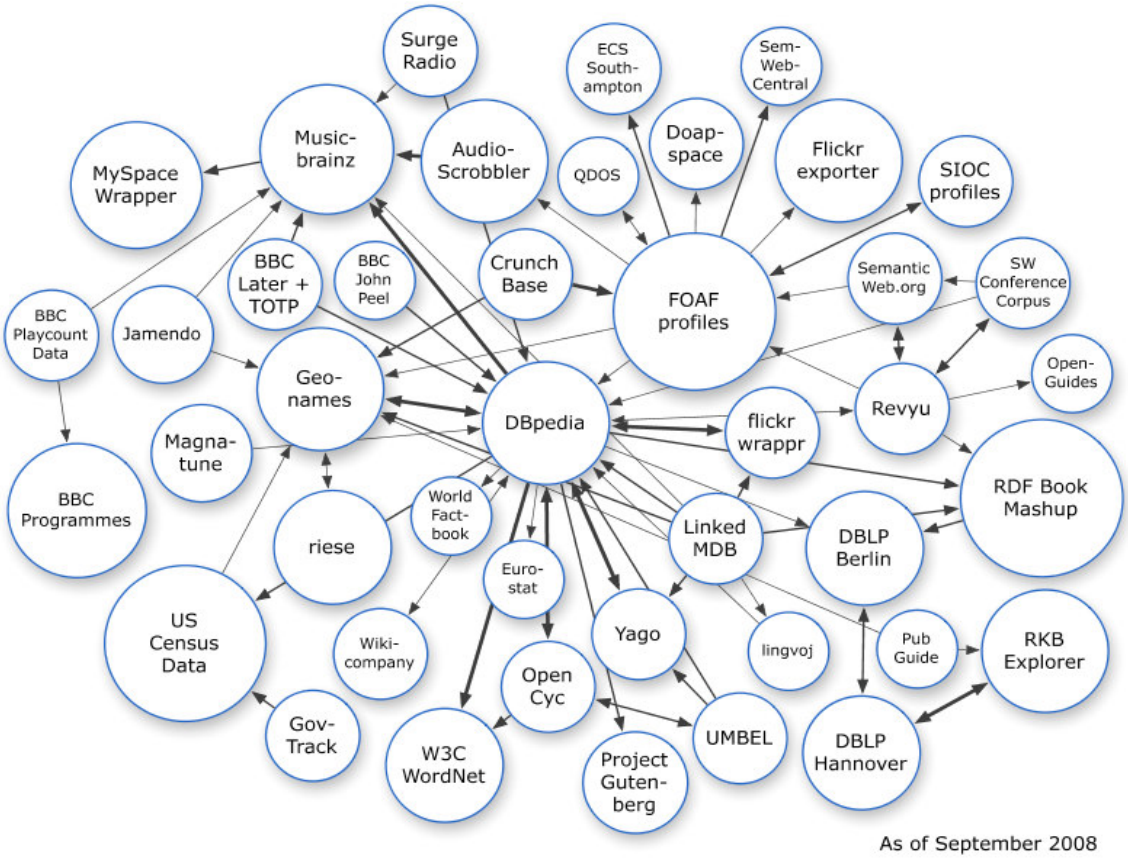
Altogether billions of triples, millions of links…

The "seed" for a general Web of Data

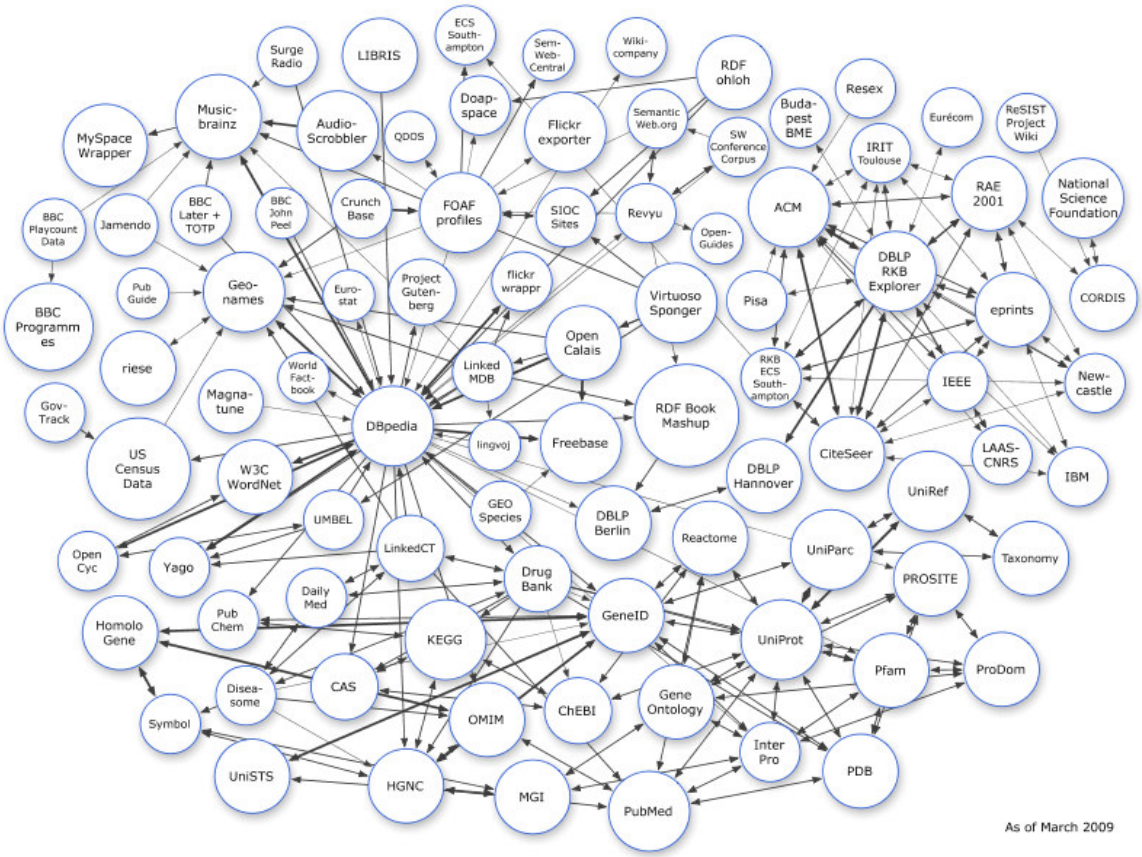
The LOD "cloud", March 2008



The LOD "cloud", September 2008



The LOD "cloud", March 2009



Applications using the cloud begin to emerge

- Bookmarking systems, exploration of social graphs, financial reporting
- LOD nodes (eg, DBPedia) provide a set of referenceable URI-s for many things
- Worth looking at the proceedings of the latest workshop, for example
  - April 2009, at WWW2009
  - http://events.linkeddata.org/ldow2009

Challenge: get the data out there!

How to access a database

Many of the LOD blobs come from relational databases

Issue: how to "map" a relational database content to RDF

- different tools exist (Virtuoso's RDF view, D2RQ, Triplify, R2O, Dartgrid toolkit, Asio, RDBToOnto)
- the W3C RDB2RDF Incubator Group published a survey:
  - http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf

How to access a database (cont.)

A new RDB2RDF Working Group is planned

Goal:

- "standardize a language for mapping relational data and relational database schemas into RDF and OWL"
- how to assign public identifiers to database entries
- group should start in July/August, watch the news and join!

Data in other formats

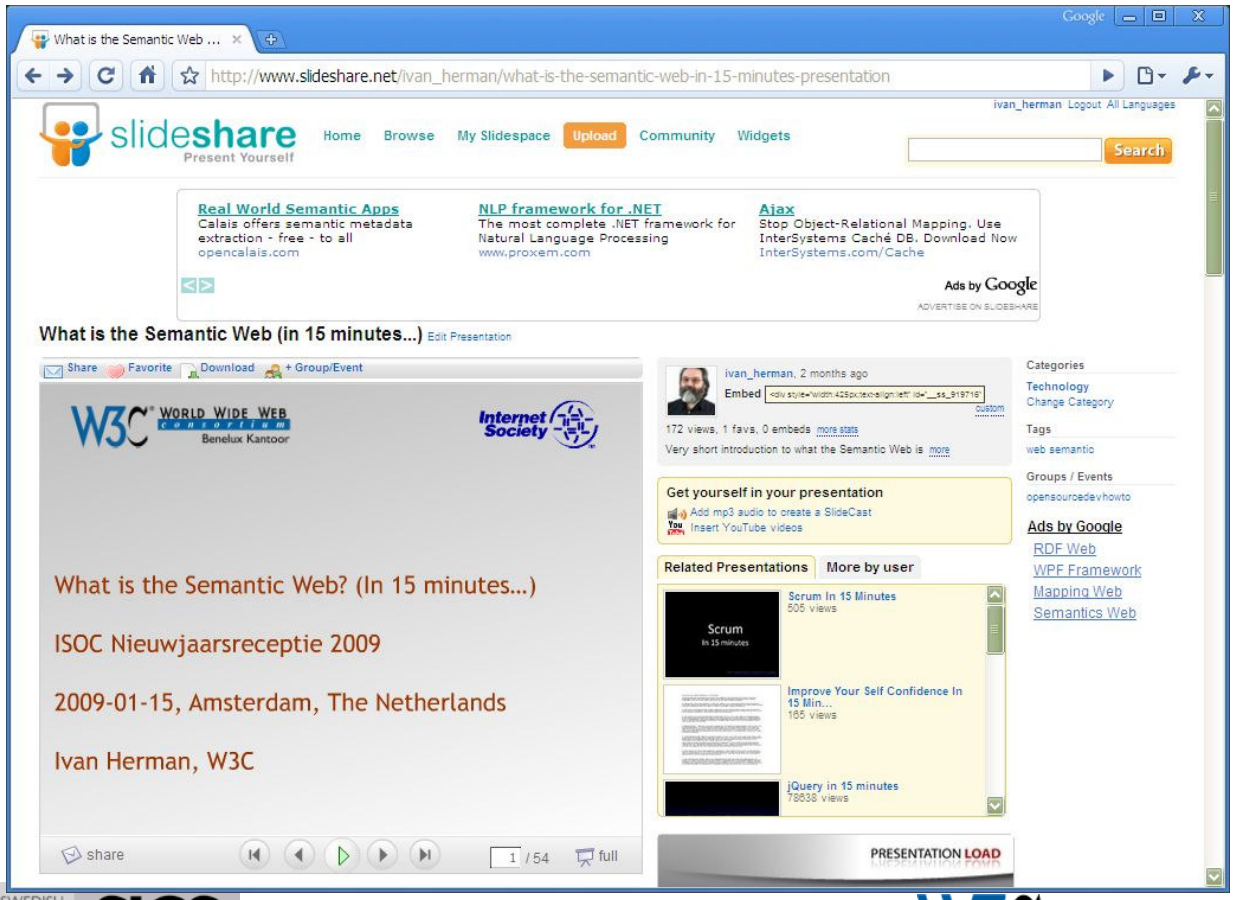
But many data are in XML, HTML and not in databases

Fortunately, GRDDL and RDFa are already around to easily produce (RDF) data

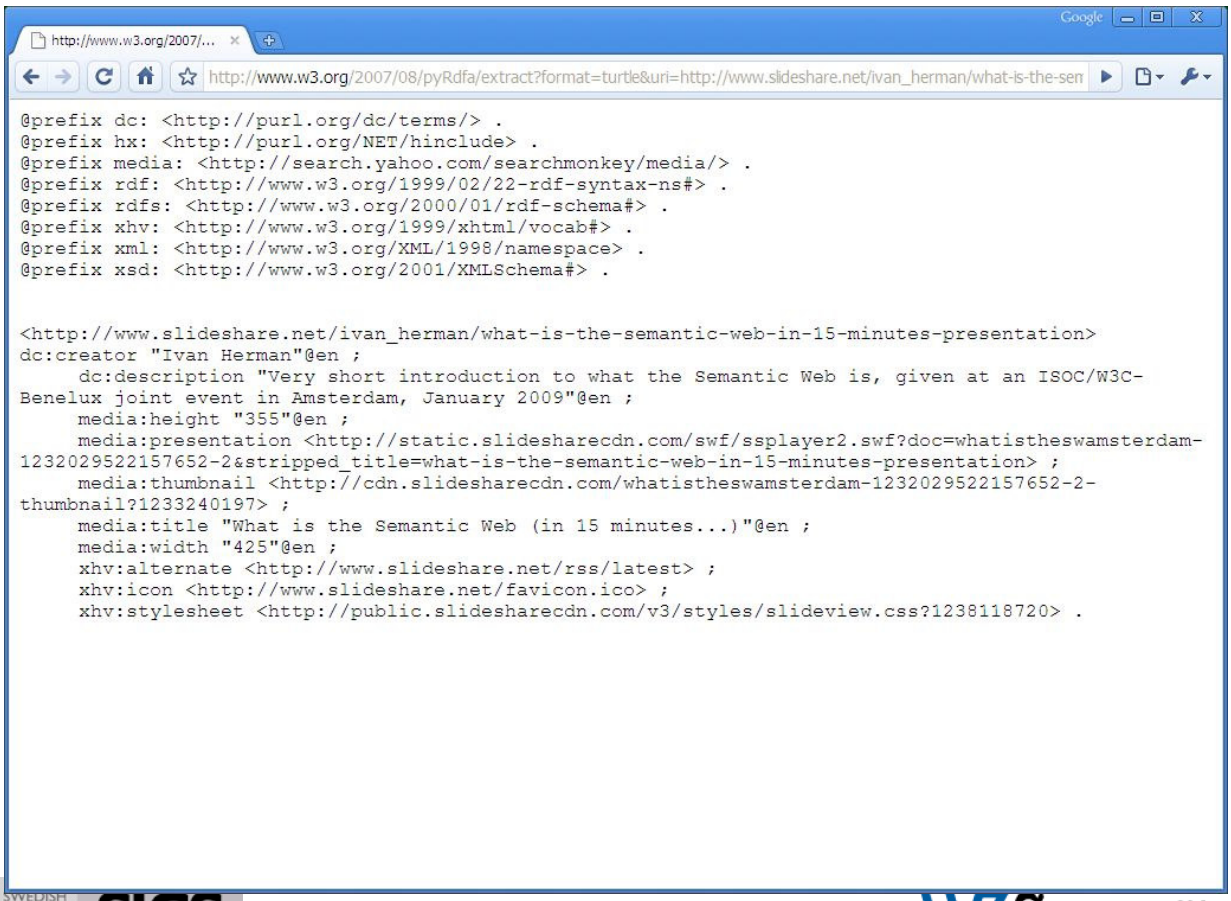
- the usual tools begin to adopt GRDDL and RDFa to retrieve RDF automatically

These data can be added to the cloud easily

Publication of data: SlideShare



Publication of data: SlideShare



How to "assign" RDF data to a collection of resources?

Instead of spelling out information for each resource, is it possible to "generate" those?

Some examples:

copyright information for all of your photographs
is a Web page collection usable on a mobile phone and how?
bibliographical data for a series of publications
provenance data for a collection of resources
annotation of the data resulting from a scientific experiment
etc

How to "assign" RDF data to a collection of resources?

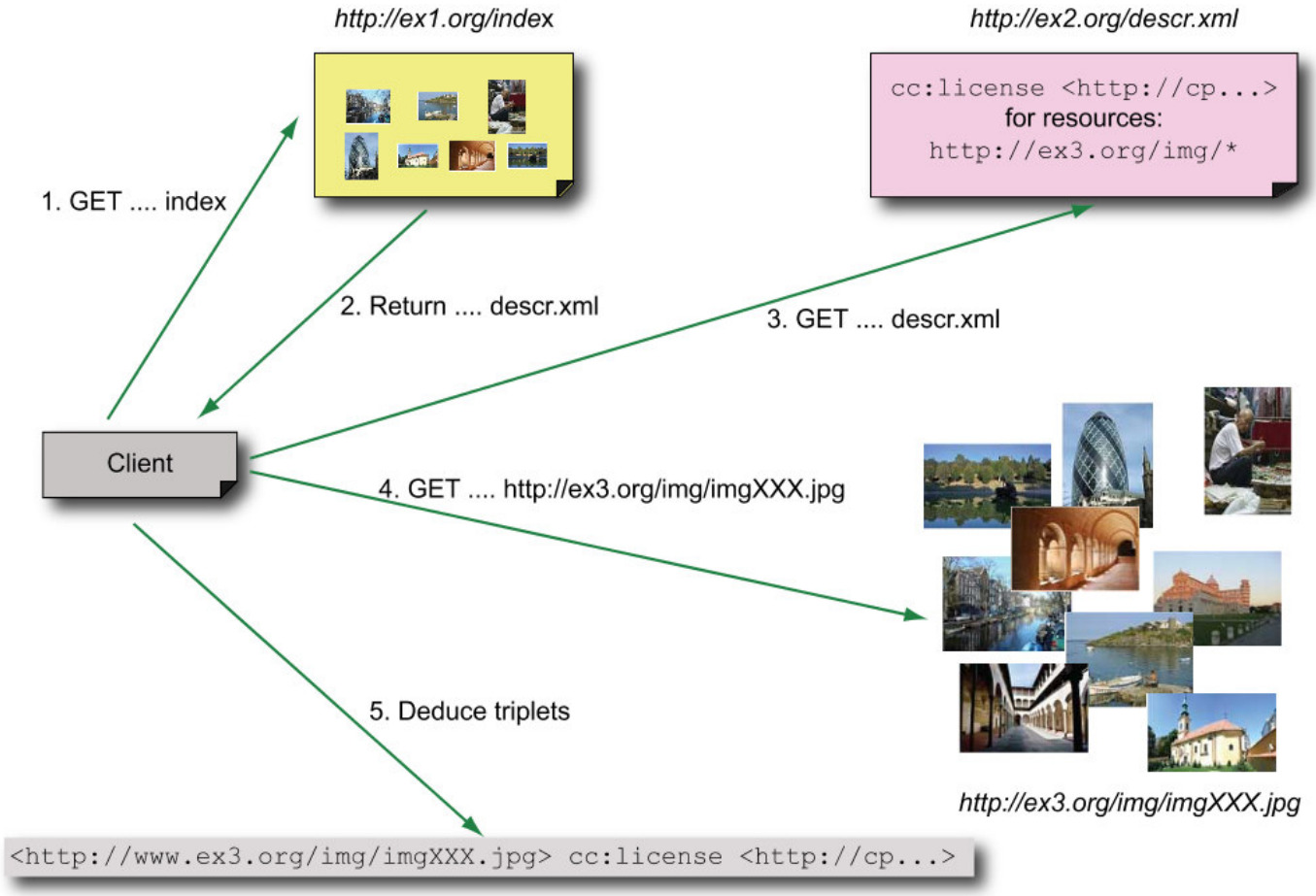
The issue:

given the URI of the resource (photograph, publication, etc), how do I find the relevant RDF data?

POWDER (Protocol for Web Description Resources)

Lets you define predicates that can be automatically assigned to a *set* of resources

POWDER scenario: copyright for photos



The technical details…

The “description resource” is an XML file:

```
<powder xmlns="http://www.w3.org/2007/05/powder#"
        xmlns:cc="http://creativecommons.org/ns#">
  <attribution>
    <issuedby src="http://www.ivan-herman.net/me"/>
  </attribution>
  <dr>
    <iriset>
      <includehosts>www.ex2.org</includehost>
      <includepathstartswith>/img/</includepathstartswith>
    </iriset>
    <descriptorset>
      <cc:license rdf:resource="http://cp:..."/>
    </descriptorset>
  </dr>
```

The technical details…

Powder processors may then return

- direct RDF triples, eg:

```
<http://www.ex2.org/img/imgXXX.jpg> cc:license <http://cp:...>.
```

- or can transform this XML file into an OWL for more generic processors
  - a canonical processing of the XML file is defined by the POWDER specification

POWDER Service

Online POWDER service can also be set up:

- a Web service with
  - submit a URI and a resource description file
  - return the RDF statements for that URI
- such service should be set up, eg, at W3C

Challenge: get the data organized

Just getting the data out there is not enough

Some sort of organization, categorization of data is necessary

Ie, the LOD needs various sorts of vocabularies to rely on

SKOS
(Simple Knowledge Organization System)

Represent and share classifications, glossaries, thesauri, etc

- for example:
  - Dewey Decimal Classification, Art and Architecture Thesaurus, ACM classification of keywords and terms…
  - classification of Web 2.0 type tags

Define classes and properties to add those structures to an RDF universe

- allow for a quick port of this traditional data, combine it with other data

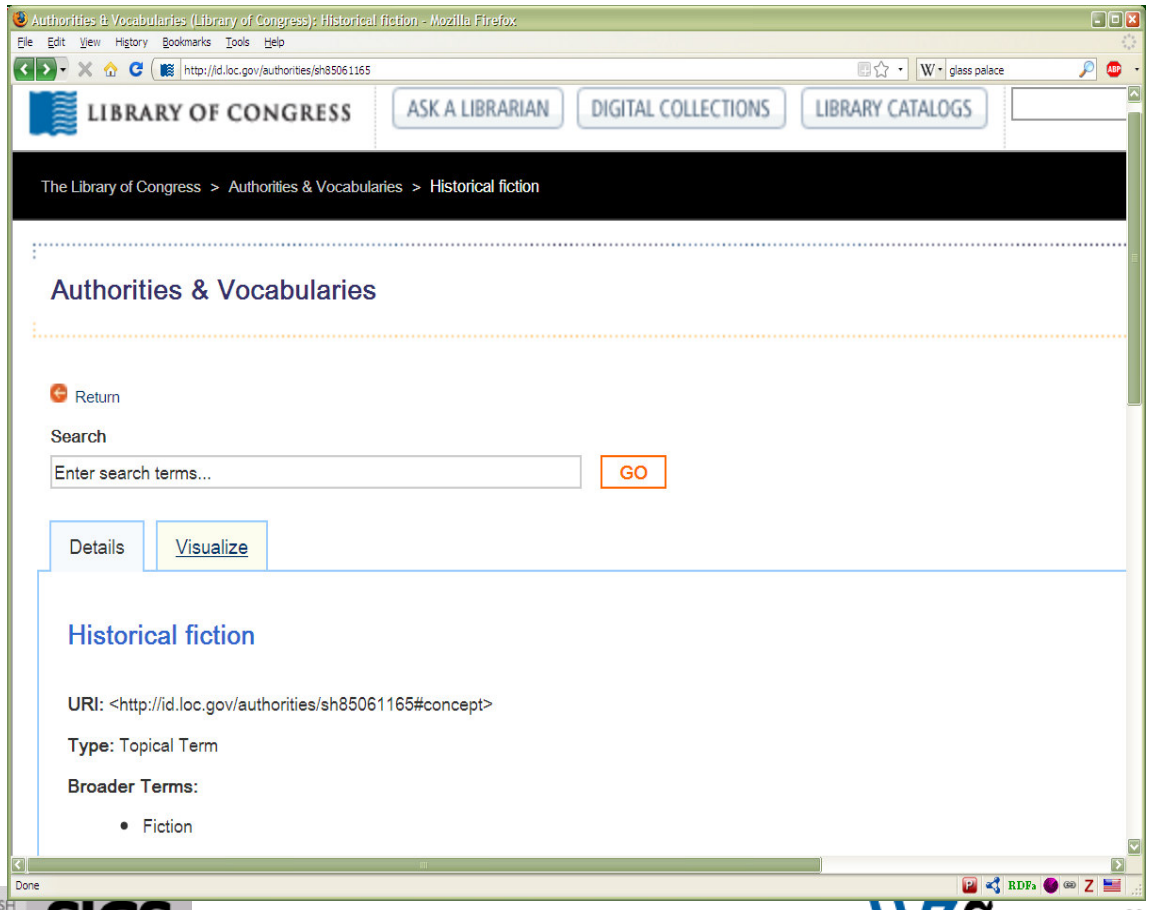
SKOS

SKOS is based on a simple structure

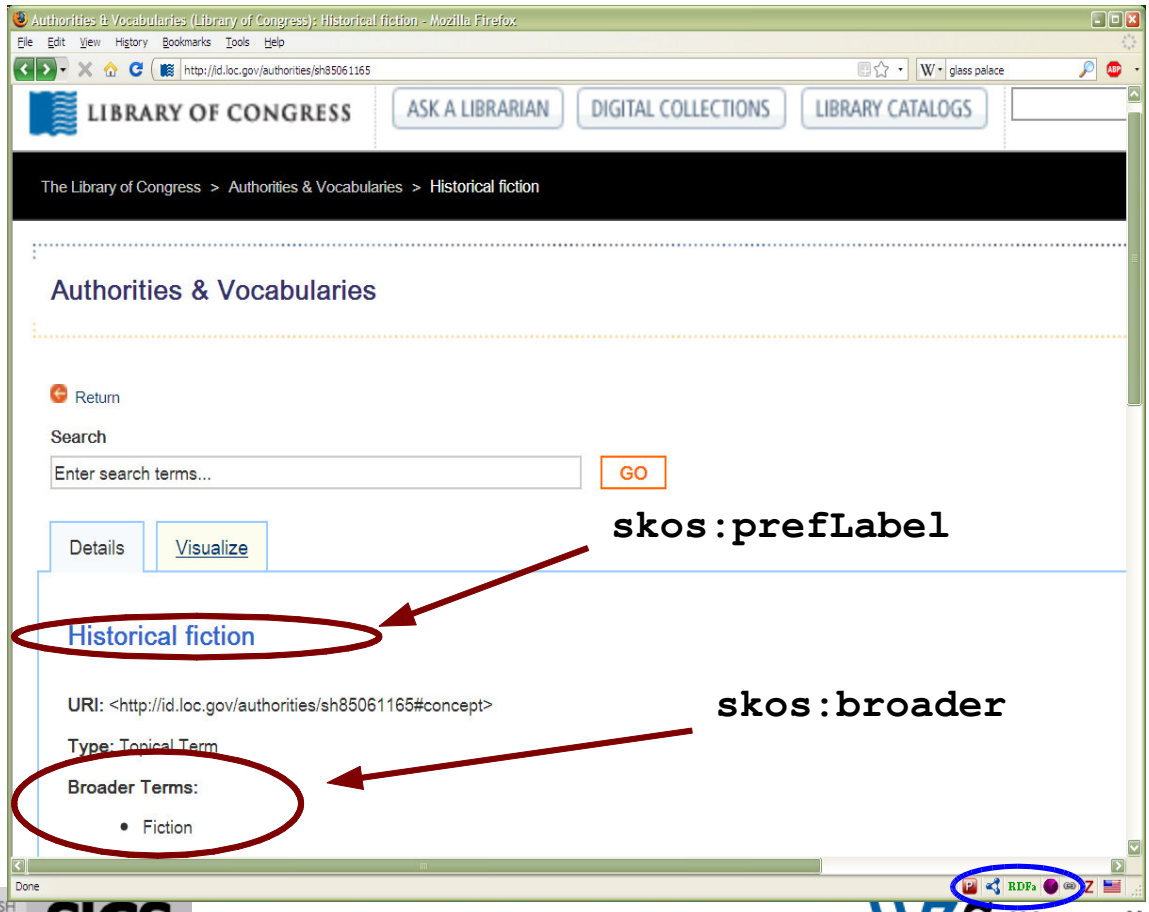
- the central concept is, well, a “SKOS concept”
- concepts can have preferred and alternate labels
- a concept may be narrower or broader then another one
- concepts may be related to one another
- concepts can be collected in “concept schemes”
- and that is it (well, almost...)

Other resources can then refer to these concepts as, eg, their subject

Typical example: LC Subject Headings



Typical example: LC Subject Headings



Typical example: LC Subject Headings

Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///C:/DOCUME~1/IVANHE~1/LOCALS~1/Temp/extract-4

glass palace

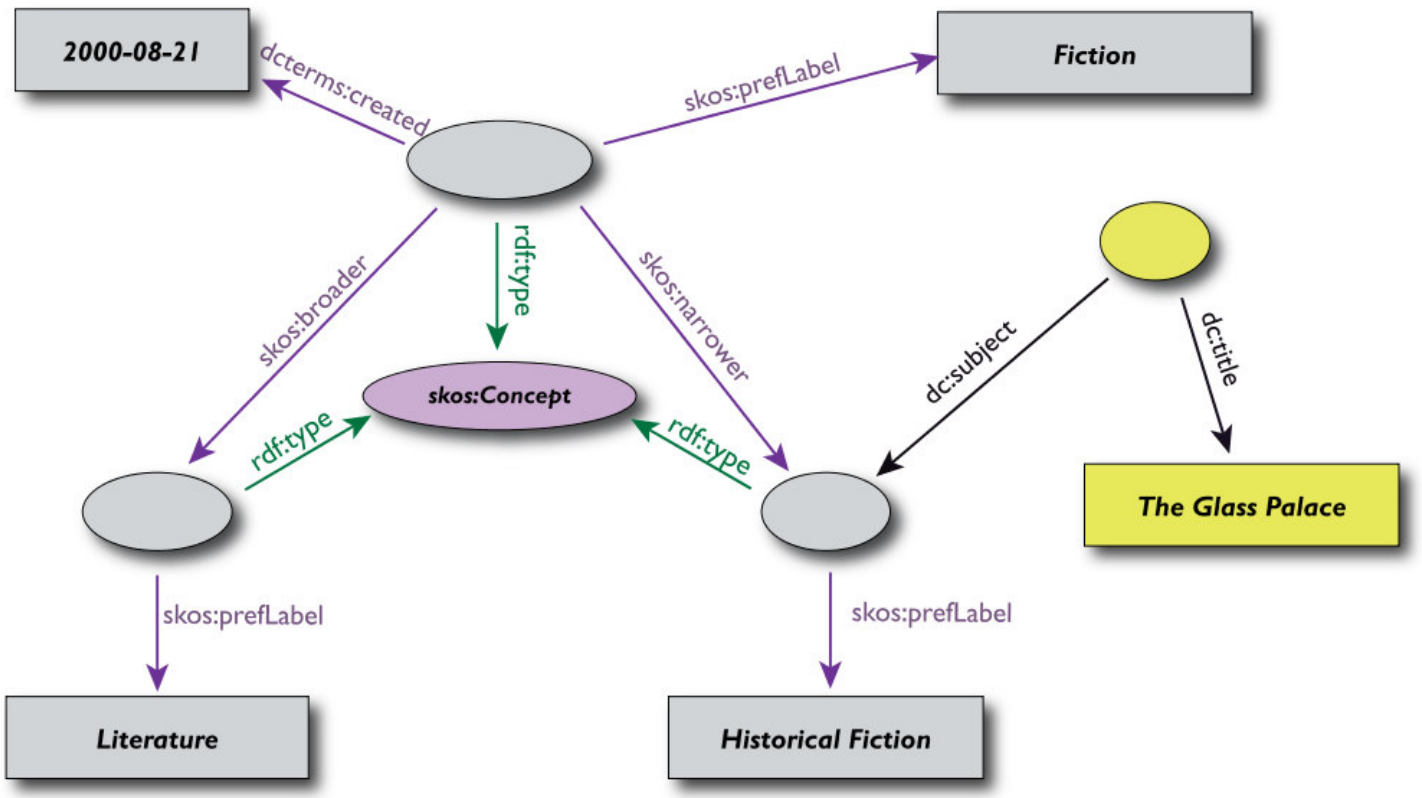
```
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xhv: <http://www.w3.org/1999/xhtml/vocab#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xs: <http://www.w3.org/2001/XMLSchema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .


<http://id.loc.gov/authorities/sh85061165> xhv:alternate
        <http://id.loc.gov/authorities/feed/>,
        <http://id.loc.gov/authorities/sh85061165.json>,
        <http://id.loc.gov/authorities/sh85061165.nt>,
        <http://id.loc.gov/authorities/sh85061165.rdf> ;
    xhv:icon <http://www.loc.gov/favicon.ico> ;
    xhv:stylesheet <http://id.loc.gov/static/css/subject_headings_print.css>, <http://id.l

<http://id.loc.gov/authorities/sh85061165#concept> a skos:Concept ;
    dcterms:created "2000-08-21T00:00:00-04:00"^^xsd:dateTime ;
    dcterms:modified "2000-10-04T10:47:15-04:00"^^xsd:dateTime ;
    dcterms:source "GSAFD, 2000 (Historical fiction. UF Fiction, Historical; Historical no
    skos:broader <http://id.loc.gov/authorities/sh85048050#concept> ;
    skos:closeMatch <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb119808101> ;
    skos:inScheme <http://id.loc.gov/authorities#conceptScheme>, <http://id.loc.gov/author
    skos:prefLabel "Historical fiction"@en .
```

Done

Using the LCSH terms…



Using the LCSH terms…

```
<http://…/isbn/000651409X>
    dc:title "The Glass Palace"@en;
    dc:subject <http://id.loc.gov/authorities/sh85061165#concept>;
    ...

<http://id.loc.gov/authorities/sh85061165#concept>
    a    skos:Concept;
    skos:prefLabel "Historical Fiction"@en;
    dc:created "2000-08-21T00:00:00-04:00"^^xsd:dateTime;
    skos:broader <http://id.loc.gov/authorities/sh85048050#concept>;
    ...

<http://id.loc.gov/authorities/sh85048050#concept>
    a skos:Concept;
    skos:prefLabel "Literature"@en;
    skos:narrower <http://id.loc.gov/authorities/sh85061165#concept>;
    ...
```


SKOS and OWL

SKOS is geared towards specific (though large) use cases, like

taxonomies, glossaries, …

annotations of complex structures (including ontologies)

SKOS is a based on a very simple usage of OWL

using some simple OWL Full constructions

the emphasis is on *organization* and not on logical inferences

"OWL is a Harley-Davison, SKOS is a mountain bike"
— (Tom Baker, co-chair of the relevant WG)

But, of course, there is also OWL

- The LOD does create new challenges due to the amount of data
  - running a full and complete OWL DL inference might be a challenge, to say the least...
- OWL 2 introduces “profiles” that might be better fit for many applications
  - restrictions on which OWL term can be used in under what circumstances

OWL 2 profiles

- Three profiles have been defined
  - Classification and instance queries in polynomial time: *OWL-EL*
  - Implementable on top of conventional relational database engines: *OWL-QL*
  - Implementable on top of traditional rule engines: *OWL-RL*
- Come to the OWL 2 panel if you want to hear more…

Rules

OWL 2 RL shows the importance of rule engines

W3C's Rule work (RIF) is getting to completion

b.t.w., OWL 2 RL can be expressed in RIF

I have no time to go into details here...

Query the data

Querying the data: SPARQL

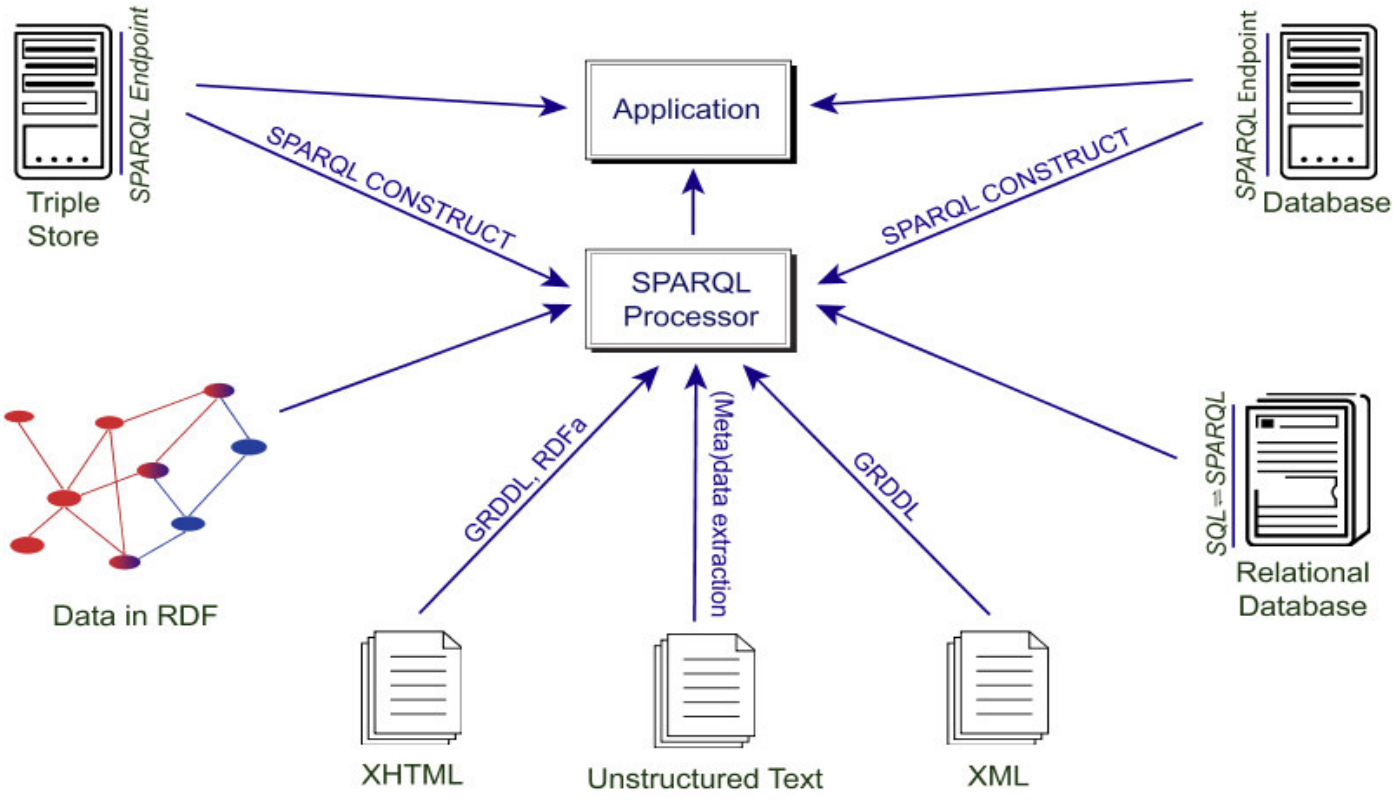
Is a W3C Standard since January 2008

it has already become one of the absolutely essential technologies on the SW

all LOD blobs offer a “SPARQL endpoint”

there is even a SPARQL endpoint for the whole LOD

SPARQL as a unifying point!



New SPARQL WG: Goals

To define a small set of extensions to SPARQL

No complex change, backward compatibility

Listen to user and implementation experiences of the past few years

Group started in February 2009

Planned features

Update, ie, ability to change the RDF store

Service description framework

- what type of extensions, inference possibilities, etc, are available at the endpoint

Addition to the query language

- aggregate functions
- subqueries
- negation
- project expressions

Conclusions

Many things are happening at W3C to evolve the Semantic Web

Many more issues are still to be done…

So join the club! After all, this is really a community effort…

Thank you for your attention!