



## Introduction to the Semantic Web

2<sup>nd</sup> European Semantic Technology Conference  
Vienna, Austria

September 24, 2008

Ivan Herman, W3C

# Towards a Semantic Web

- The current Web represents information using
  - natural language (English, Hungarian, Chinese,...)
  - graphics, multimedia, page layout
- Humans can process this easily
  - can deduce facts from partial information
  - can create mental associations
  - are used to various sensory information
    - (well, sort of... people with disabilities may have serious problems on the Web with rich media!)

# Towards a Semantic Web

- Tasks often require to combine data on the Web:
  - hotel and travel information may come from different sites
  - searches in different digital libraries
  - etc.
- Again, humans combine these information easily
  - even if different terminologies are used!

# However...

- However: machines are ignorant!
  - partial information is unusable
  - difficult to make sense from, e.g., an image
  - drawing analogies automatically is difficult
  - difficult to combine information automatically
    - is <foo:creator> same as <bar:author>?
  - ...

# Example: automatic airline reservation

- Your automatic airline reservation
  - knows about your preferences
  - builds up knowledge base using your past
  - can combine the local knowledge with remote services:
    - airline preferences
    - dietary requirements
    - calendaring
    - etc
- It communicates with remote information
  - (M. Dertouzos: The Unfinished Revolution)

# Example: data(base) integration

- Databases are very different in structure, in content
- Lots of applications require managing several databases
  - after company mergers
  - combination of administrative data for e-Government
  - biochemical, genetic, pharmaceutical research
  - etc.
- Most of these data are accessible from the Web (though not necessarily public yet)

# And the problem is real...

**CoCoDat - Collation of Cortical Data - Mozilla Firefox**

File Edit View History Bookmarks Tools Help

CoCoMac DATABASES ORT EXAMPLES

**CoCoDat: Collation of Cortical [sic] microcircuitry] Data**

CoCoDat is a microcircuitry database that published experimental reports. The data and cellular compartment), as well as the

- Morphology
- Firing properties
- Ionic currents
- Ionic conductances
- Synaptic currents
- Connectivity

The database is available for download under data tables but also a Search Board with p manual or automatic relaxation of the sea

- Brain region
- Layer
- Neuron type

<http://www.cocomac.org/cocodat/catalyzer/index.html>

**Cell Centered Database - Mozilla Firefox**

File Edit View History Bookmarks Tools Help

Cell Centered Database™ National Center for Microscopy and Imaging Research **Gallery**

Data | Search | Gallery | Dictionary | Publications | MyCCDB | Data Download | Contact us | Help

2D image Reconstruction Segmentation Animation

**NeuronDB = Thalamic relay neuron - Overview (A) () - Mozilla Firefox**

File Edit View History Bookmarks Tools Help

Back **Thalamic relay neuron**

Mode: **Overview** Data/Search plus Connectivity plus Classical References/Notes Models

Region: Distal equivalent dendrite Middle equivalent dendrite Proximal equivalent dendrite Soma Axon hillock Axon fiber Axon terminal All Compartments

Properties: Receptors Channels Transmitters **All Properties**

Interoperation: Gene and Chromosome Experimental Data (neurodatabase.org) Microscopy Data (CCDB)

Neuron type: principal Organism: Vertebrates

1. Equivalent dendrite Show other  
2. Distal equivalent dendrite Show other  
3. Middle equivalent dendrite Show other  
4. Proximal equivalent dendrite Show other  
5. Soma Show other

Done

# Example: social networks

- Social sites are everywhere these days (LinkedIn, Facebook, Dopplr, Digg, Plexo, Zyb, ...)
- How many times did you have to add your contacts?
- Applications should be able to get to those data via standard means
  - there are, of course, privacy issues...

# What is needed?

- (Some) data should be available for machines for further processing
- Data should be possibly combined, merged on a Web scale
- Sometimes, data may describe other data...
- ... but sometimes the data is to be exchanged by itself, like my calendar or my travel preferences
- Machines may also need to reason about that data

# In what follows...

- We will use a simplistic example to introduce the main Semantic Web concepts
- We take, as an example area, data integration

# The rough structure of data integration

1. Map the various data onto an abstract data representation
  - make the data independent of its internal representation...
2. Merge the resulting representations
3. Start making queries on the whole!
  - queries not possible on the individual data sets

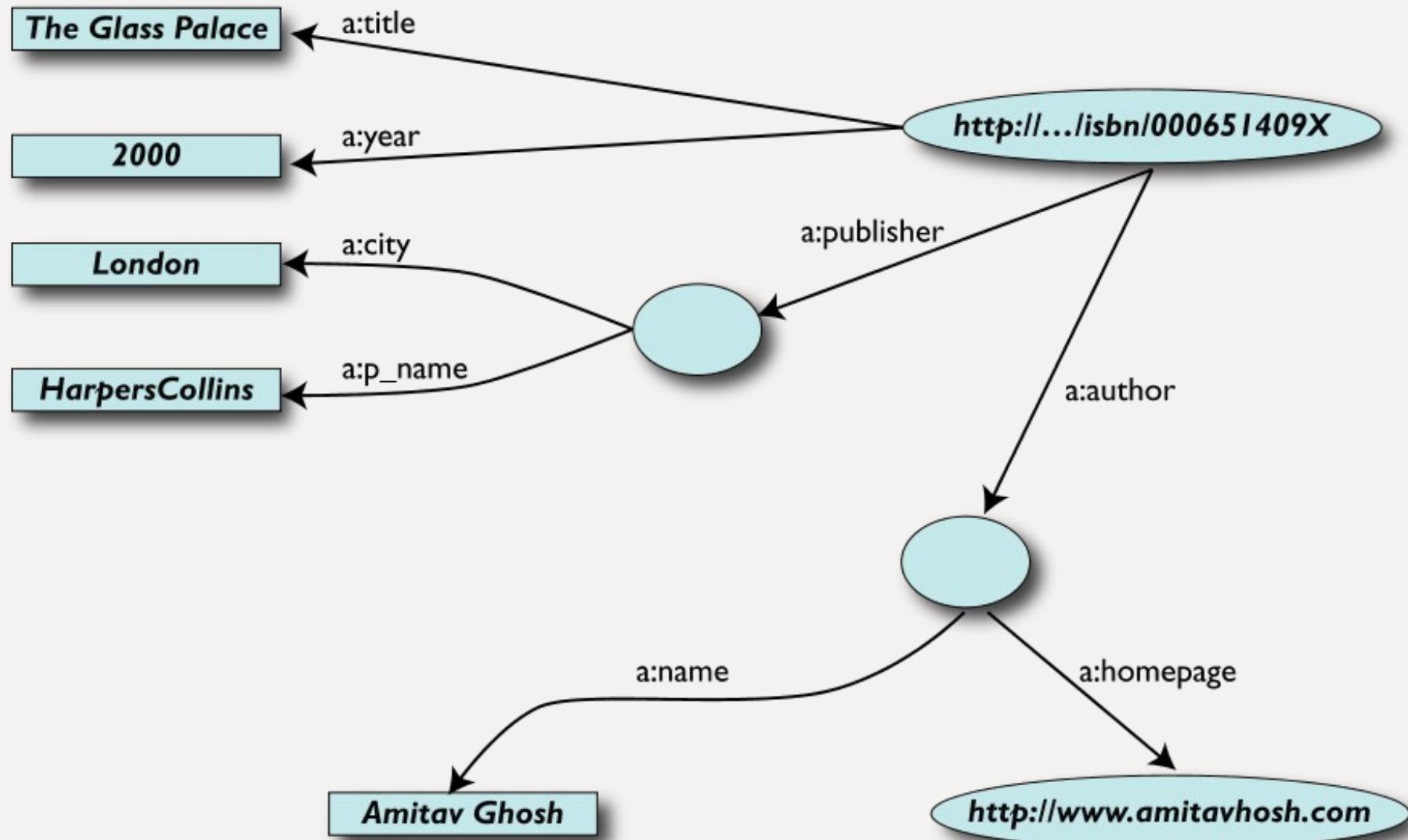
# A simplified bookstore data (dataset “A”)

<b>ID</b>	<b>Author</b>	<b>Title</b>	<b>Publisher</b>	<b>Year</b>
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

<b>ID</b>	<b>Name</b>	<b>Home Page</b>
id_xyz	Ghosh, Amitav	<a href="http://www.amitavghosh.com">http://www.amitavghosh.com</a>

<b>ID</b>	<b>Publ. Name</b>	<b>City</b>
id_qpr	Harpers Collins	London

# 1<sup>st</sup>: export your data as a set of *relations*



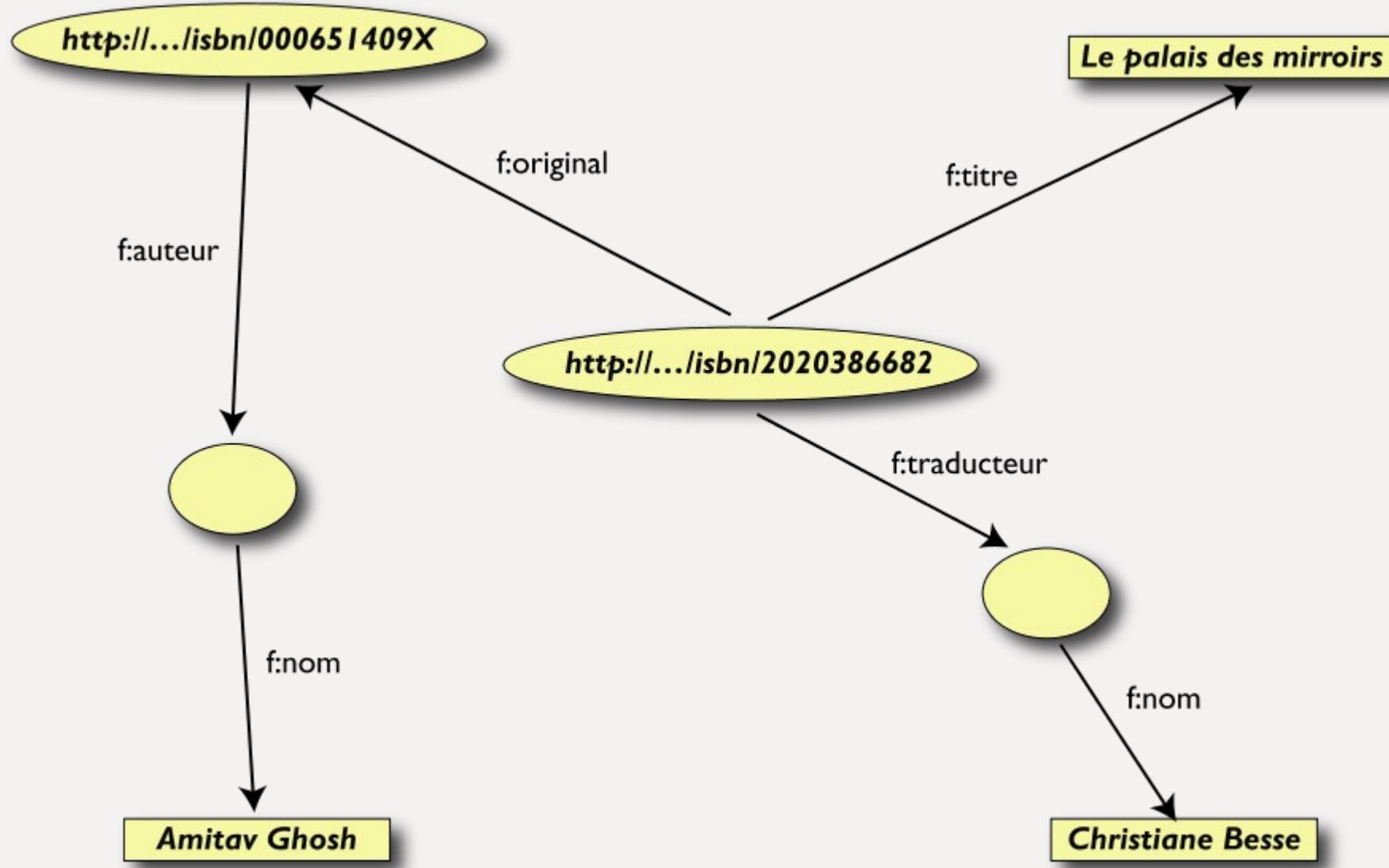
# Some notes on the exporting the data

- Data export does not necessarily mean physical conversion of the data
  - relations can be generated on-the-fly at query time
    - via SQL “bridges”
    - scraping HTML pages
    - extracting data from Excel sheets
    - etc.
- One can export part of the data

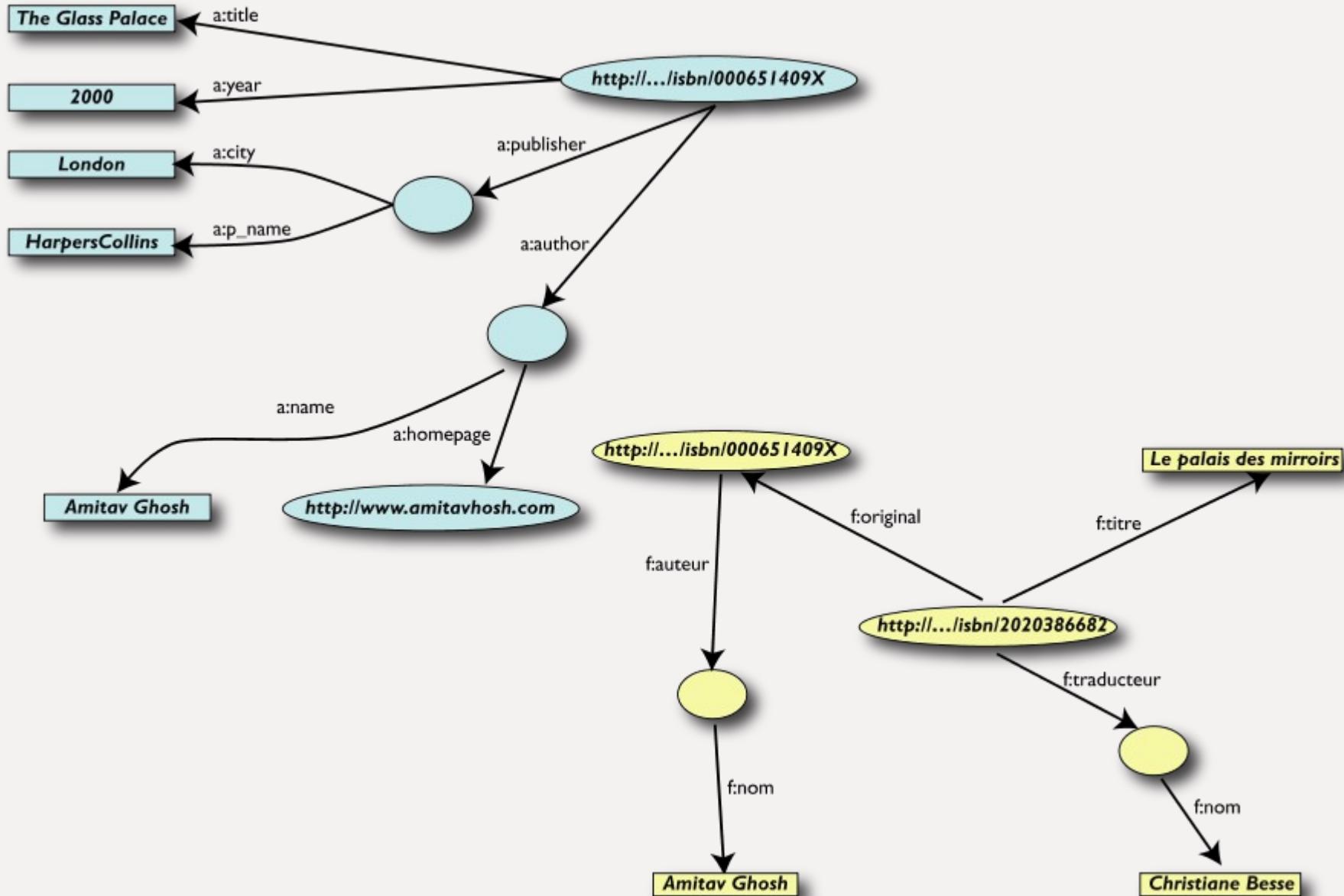
# Another bookstore data (dataset “F”)

	A	B	C	D	E
1	<b>ID</b>	<b>Titre</b>	<b>Auteur</b>	<b>Traducteur</b>	<b>Original</b>
2	ISBN0 2020386682	Le Palais des miroirs	A7	A8	ISBN-0-00-651409-X
3					
4					
5					
6		<b>Nom</b>			
7		Ghosh, Amitav			
8		Besse, Christianne			

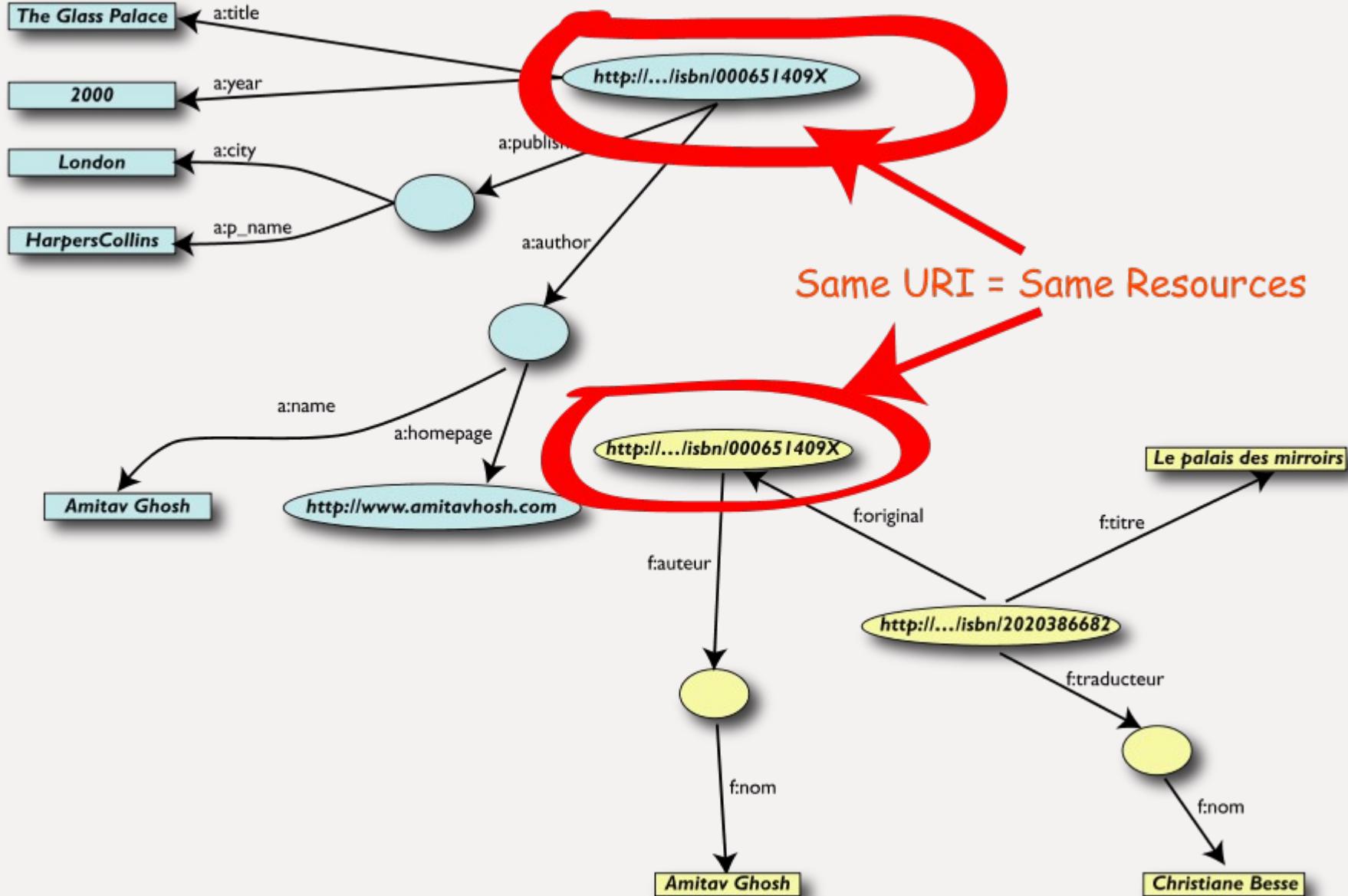
## 2<sup>nd</sup>: export your second set of data



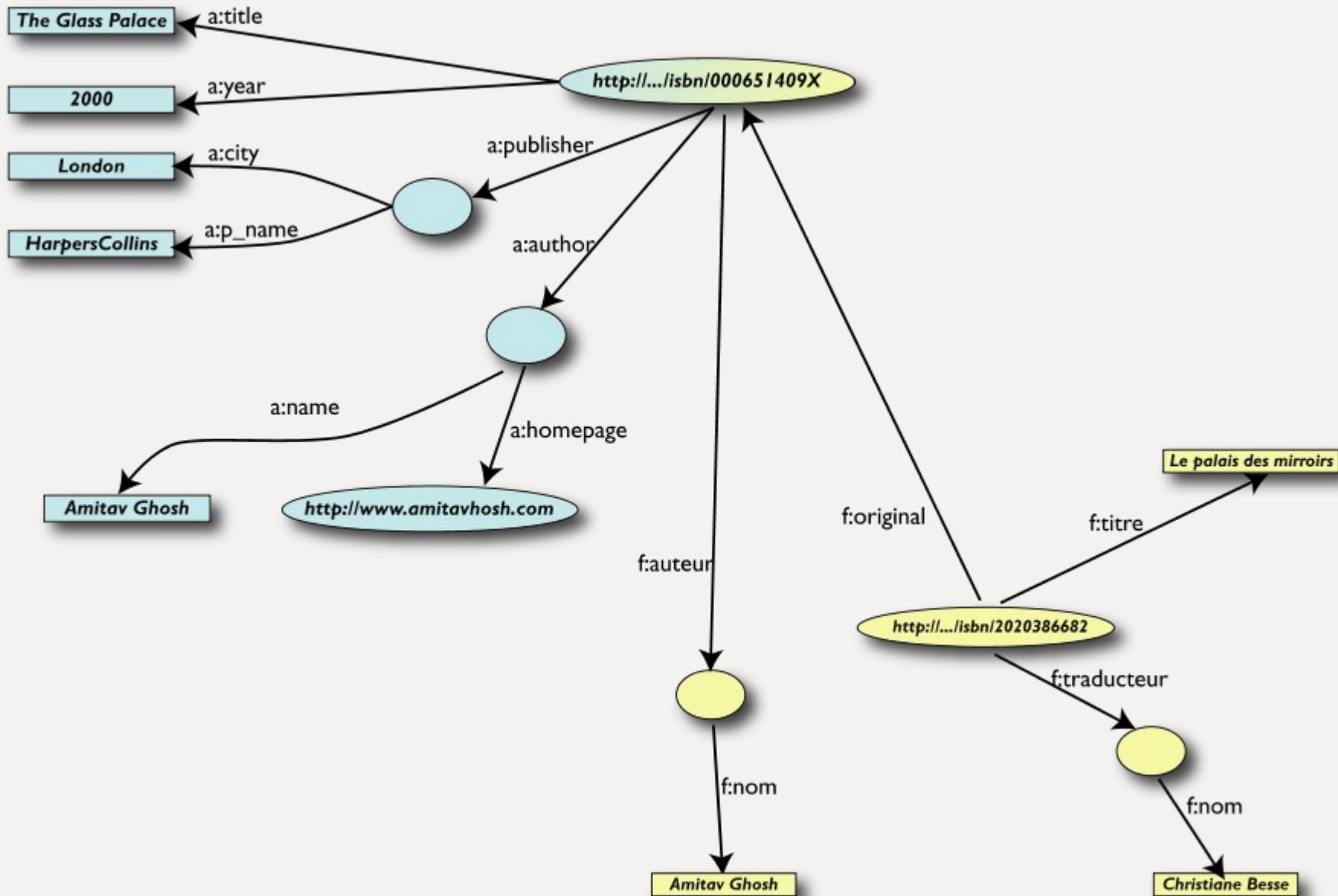
# 3<sup>rd</sup>: start merging your data



# 3<sup>rd</sup>: start merging your data (cont.)

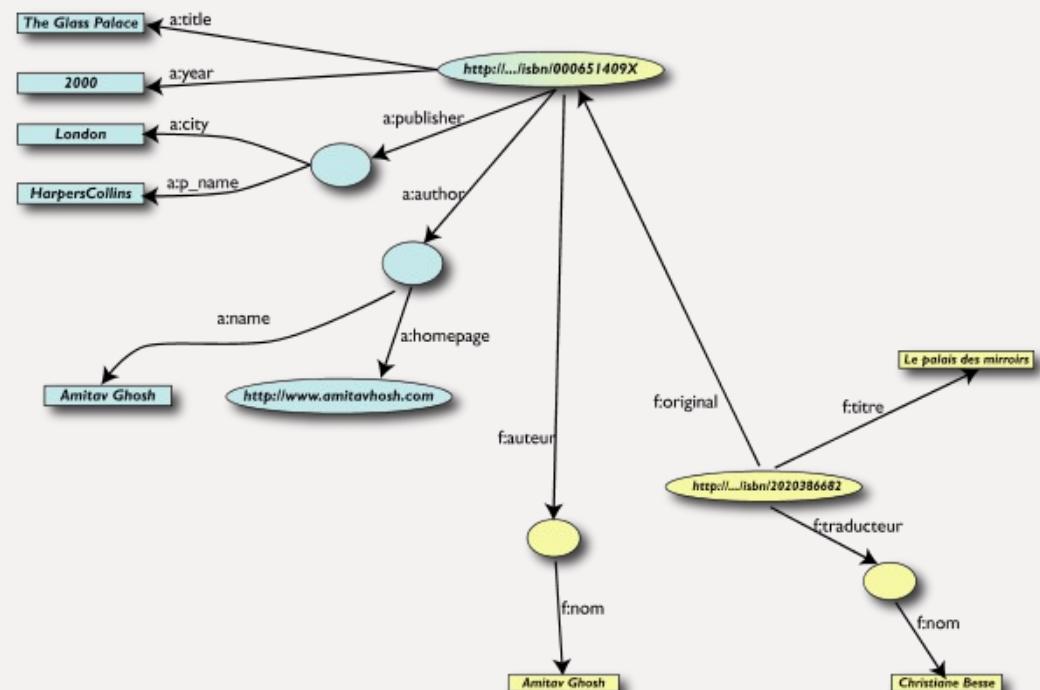


# 3<sup>rd</sup>: merge identical resources



# Start making queries...

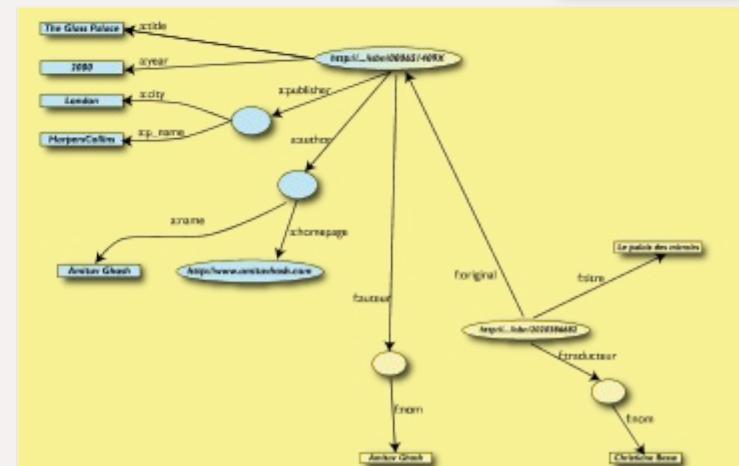
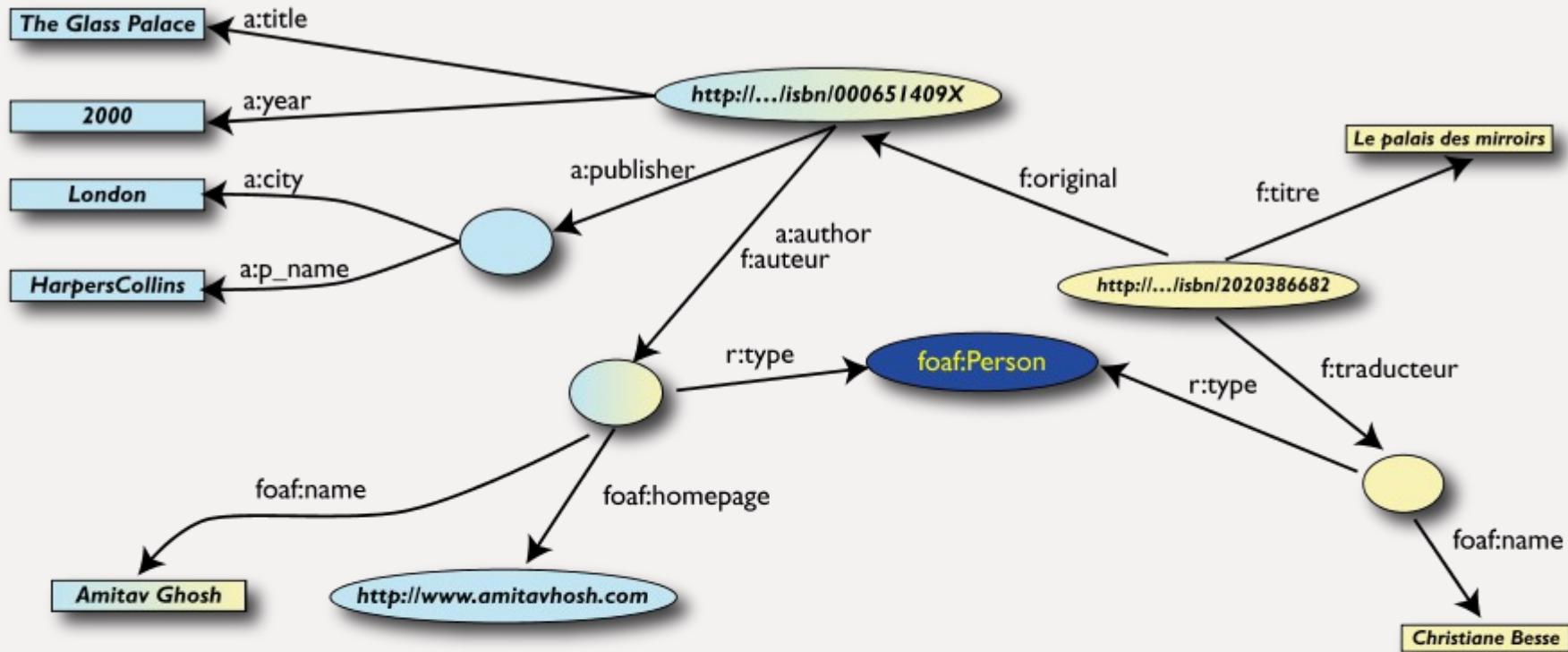
- User of data “F” can now ask queries like:
  - “give me the title of the original”
- This information is not in the dataset “F”...
- ...but can be retrieved by merging with dataset “A”!



# However, more can be achieved...

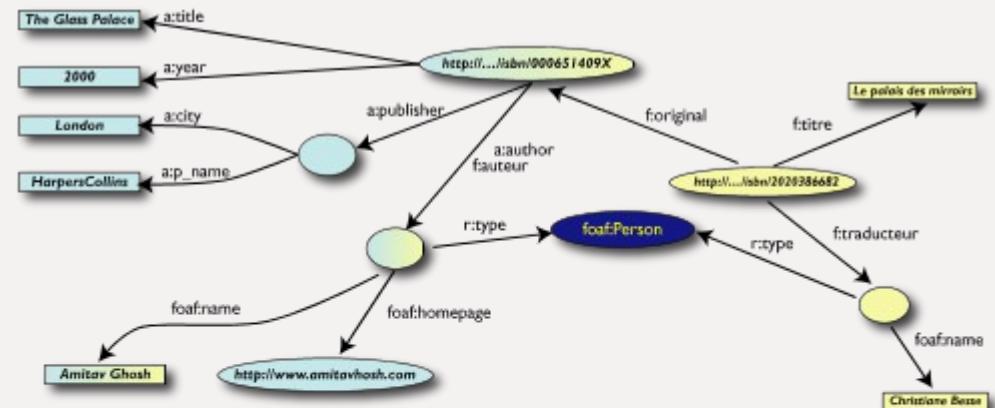
- We “feel” that **a:author** and **f:auteur** should be the same
- But an automatic merge does not know that!
- Let us add some extra information to the merged data:
  - **a:author** same as **f:auteur**
  - both identify a “Person”
  - a term that a community may have already defined:
    - a “Person” is uniquely identified by his/her name and, say, homepage
    - it can be used as a “category” for certain type of resources

# 3<sup>rd</sup> revisited: use the extra knowledge



# Start making richer queries!

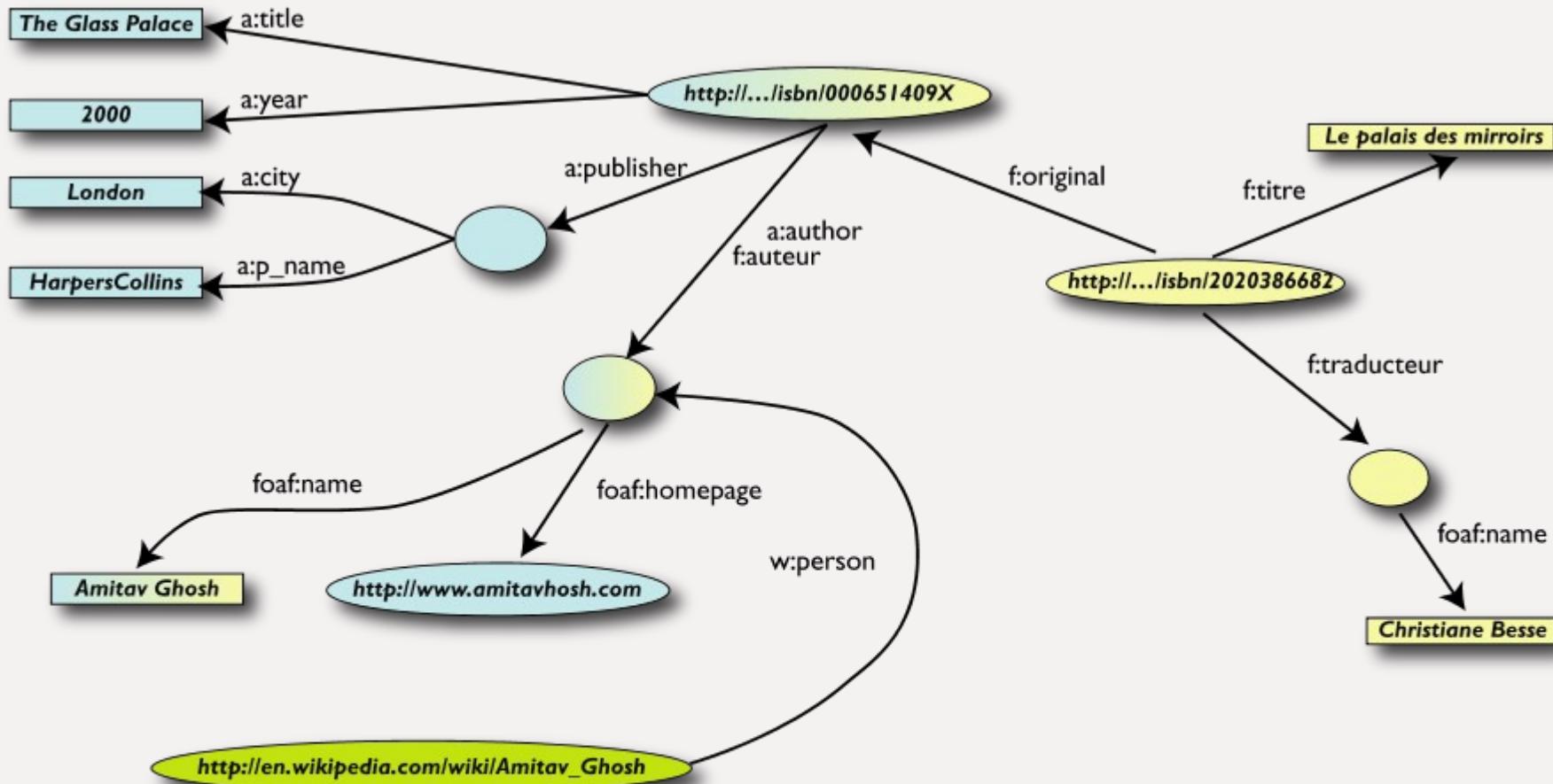
- User of dataset “F” can now query:
  - “give me the home page of the original’s author”
- The information is not in datasets “F” or “A”...
- ...but was made available by:
  - merging datasets “A” and datasets “F”
  - adding three simple extra statements as an extra “glue”



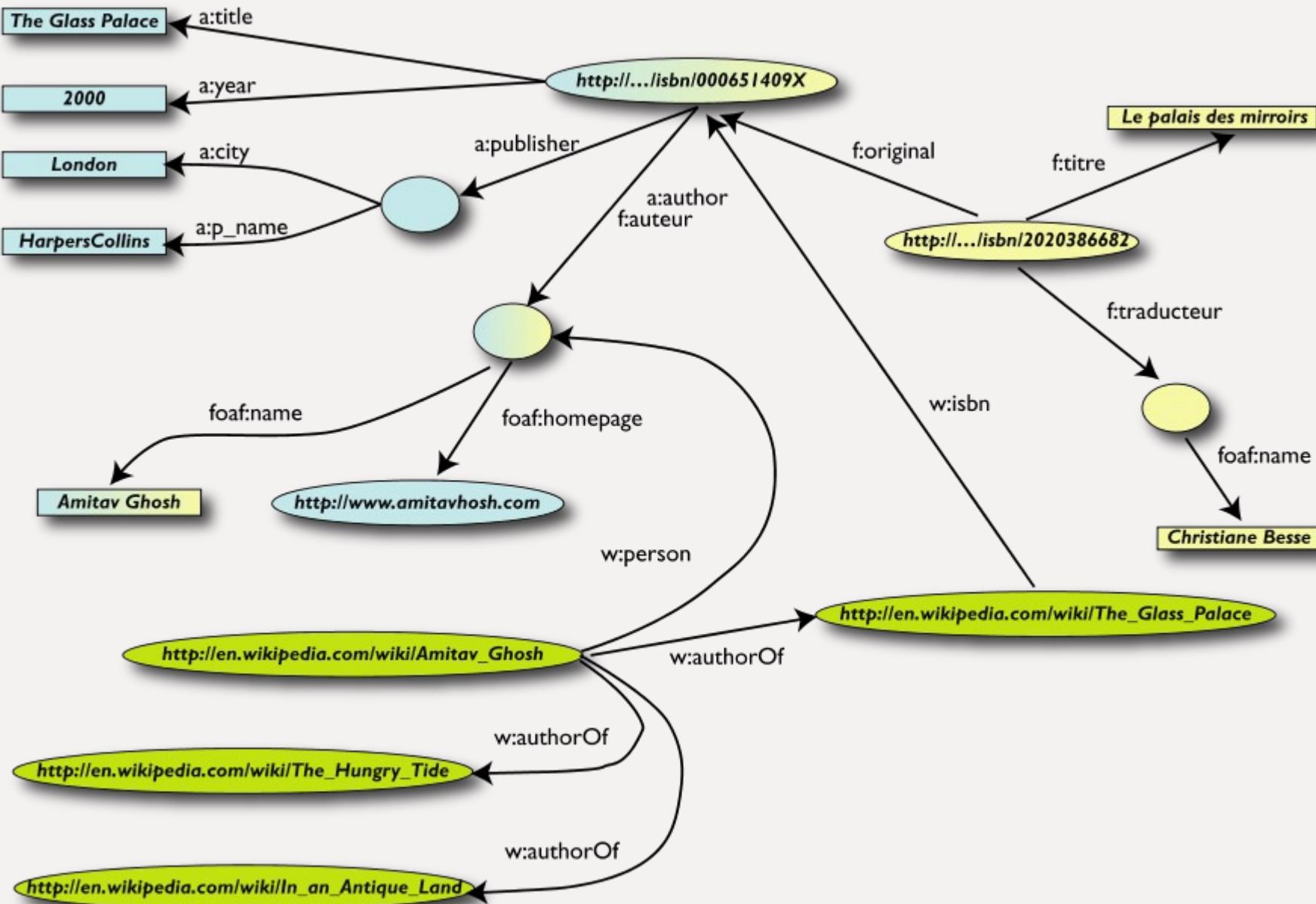
# Combine with different datasets

- Via, e.g., the “Person”, the dataset can be combined with other sources
- For example, data in Wikipedia can be extracted using dedicated tools

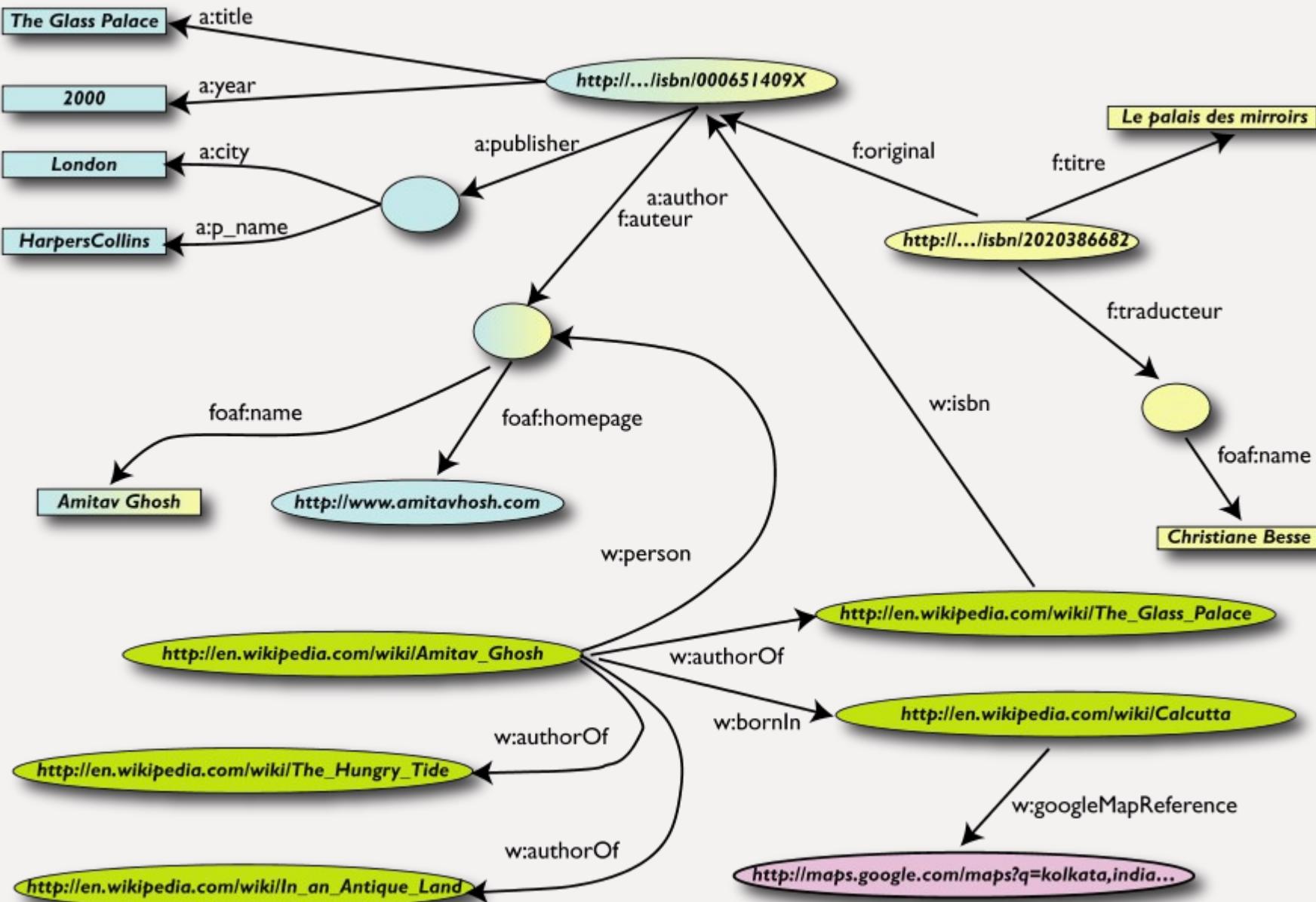
# Merge with Wikipedia data



# Merge with Wikipedia data



# Merge with Wikipedia data



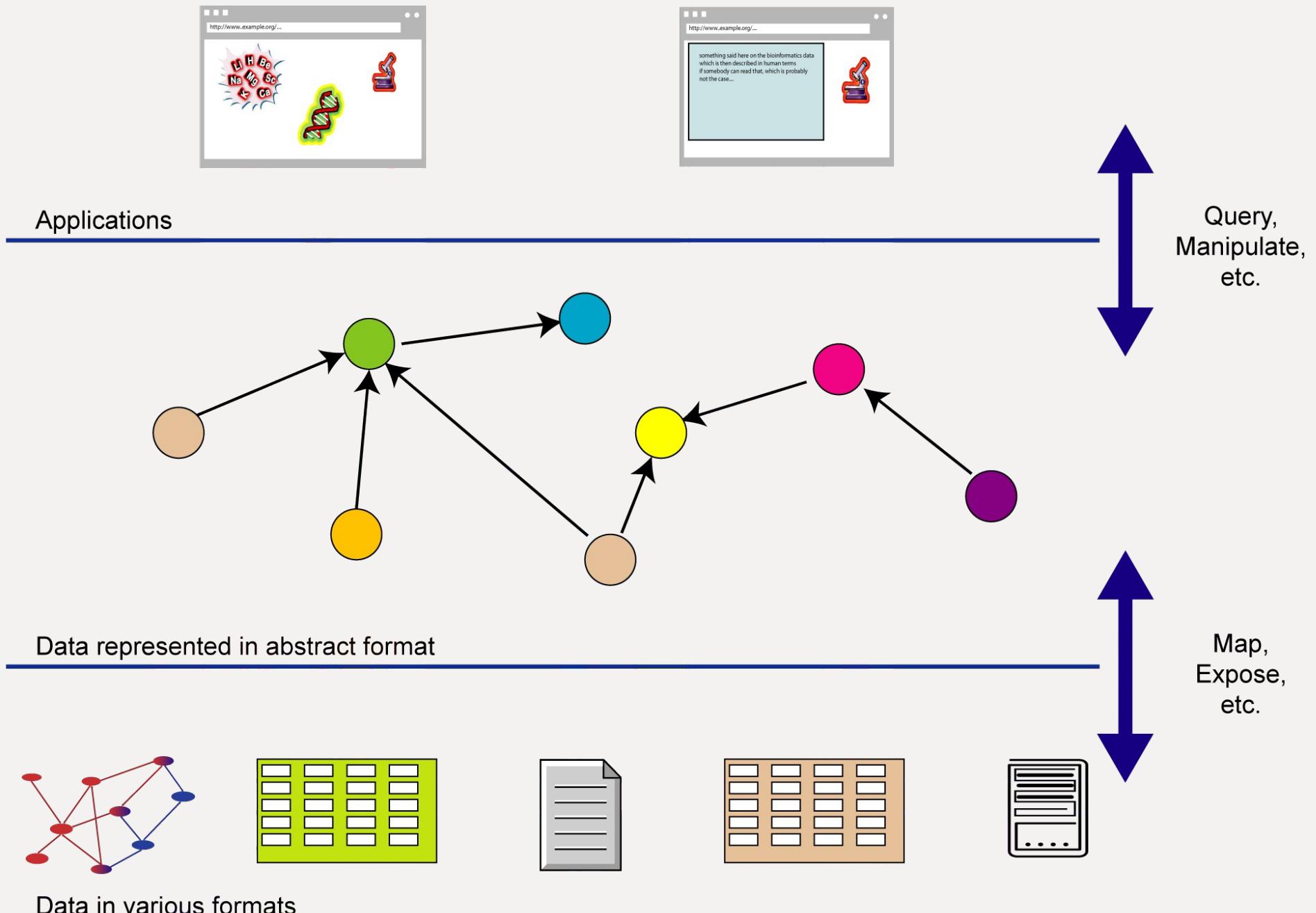
# Is that surprising?

- Maybe but, in fact, no...
- What happened via automatic means is done every day by Web users!
- The difference: a bit of extra rigour so that machines could do this, too

# It could become even more powerful

- We could add extra knowledge to the merged datasets
  - e.g., a full classification of various types of library data
  - geographical information
  - etc.
- This is where ontologies, extra rules, etc, come in
  - ontologies/rule sets can be relatively simple and small, or huge, or anything in between...
- Even more powerful queries can be asked as a result

# What did we do? (cont)



# The network effect

- Through URI-s we can link any data to any data
- The “network effect” is extended to the (Web) data
- “Mashup on steroids” become possible

# So where is the Semantic Web?

- The Semantic Web provides technologies to make such integration possible!
- Hopefully you get a full picture at the end of the tutorial...

# Basic RDF

# RDF triples

- Let us begin to formalize what we did!
  - we “connected” the data...
  - but a simple connection is not enough... it should be named somehow
  - hence the RDF Triples: a labelled connection between two resources

# RDF triples (cont.)

- An RDF Triple ( $s, p, o$ ) is such that:
  - “ $s$ ”, “ $p$ ” are URI-s, ie, resources on the Web; “ $o$ ” is a URI or a literal
    - “ $s$ ”, “ $p$ ”, and “ $o$ ” stand for “subject”, “property”, and “object”
  - here is the complete triple:

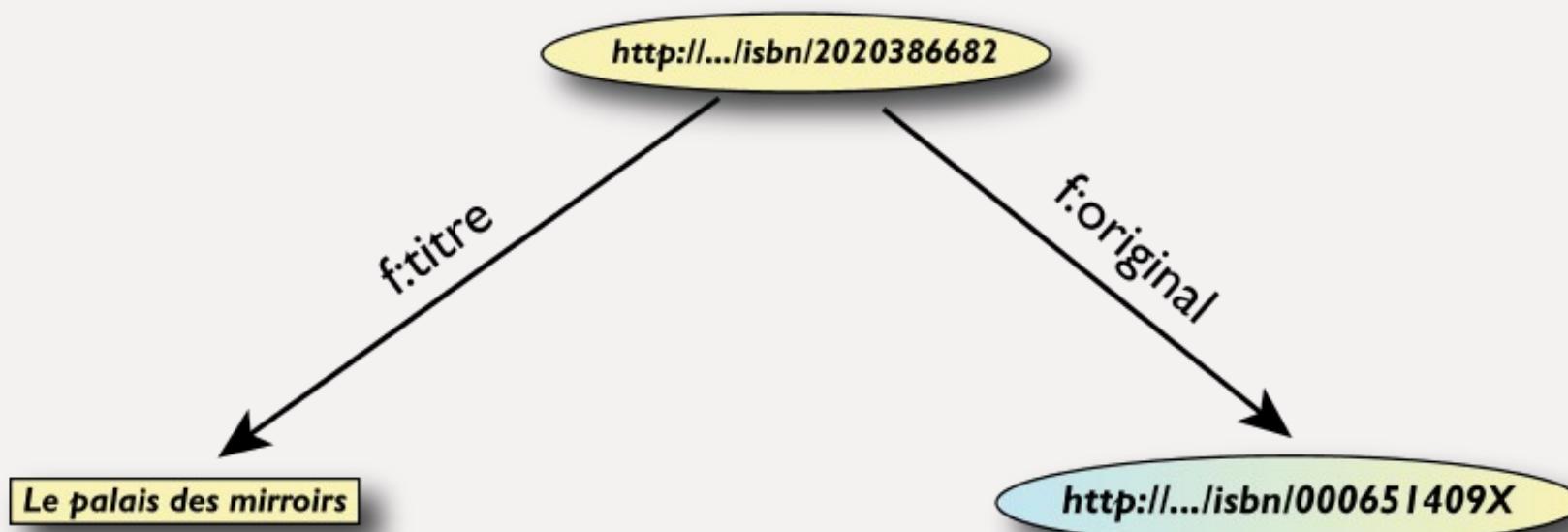
```
(<http://...isbn...6682>, <http://.../original>, <http://...isbn...409X>)
```

- RDF is a general model for such triples (with machine readable formats like RDF/XML, Turtle, N3, RXR, ...)

# RDF triples (cont.)

- Resources can use any URI; it can denote an element within an XML file on the Web, not only a “full” resource, e.g.:
  - `http://www.example.org/file.xml#element(home)`
  - `http://www.example.org/file.html#home`
  - `http://www.example.org/file2.xml#xpath1(//q[@a=b])`
- RDF triples form a directed, labelled graph (best way to think about them!)

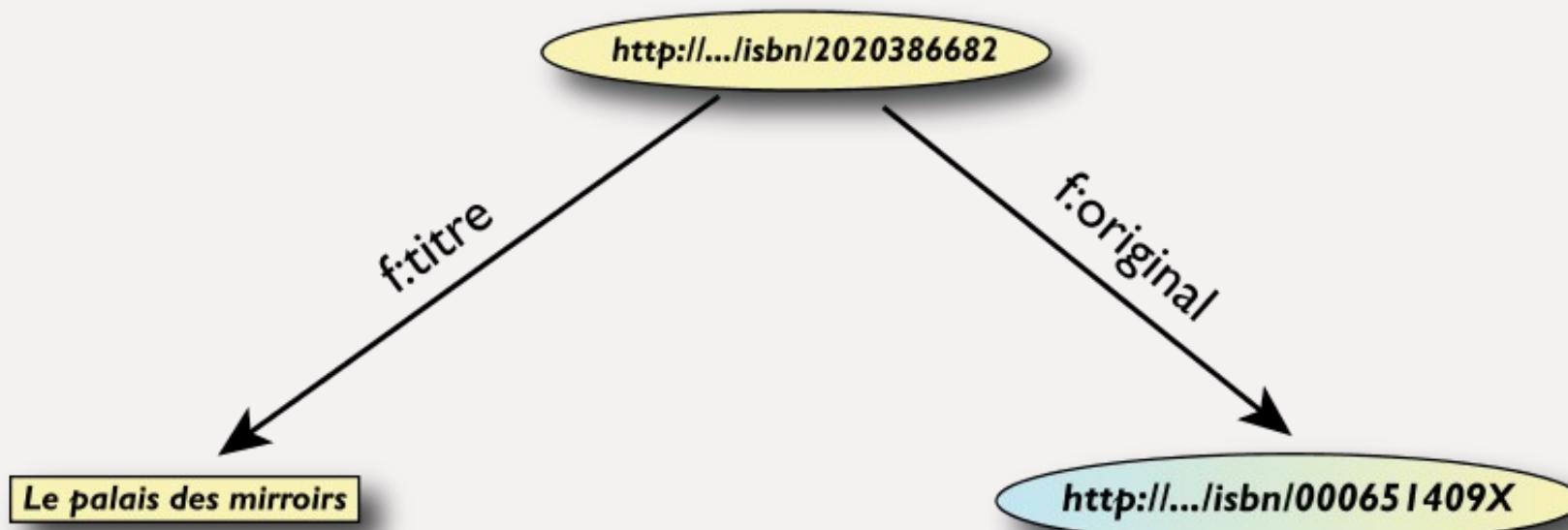
# A simple RDF example (in RDF/XML)



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

(Note: namespaces are used to simplify the URI-s)

# A simple RDF example (in Turtle)



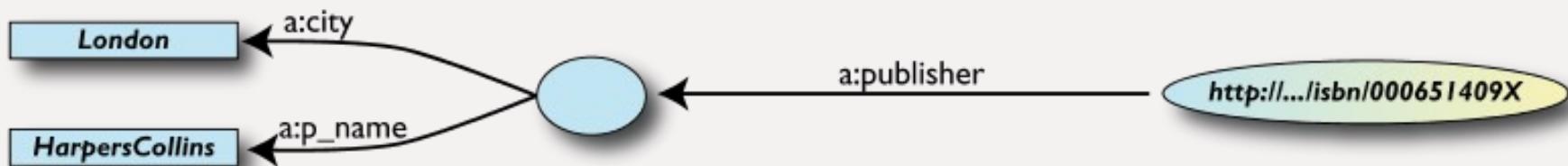
```
<http://.../isbn/2020386682>
  f:titre "Le palais des mirroirs"@fr ;
  f:original <http://.../isbn/000651409X> .
```

# URI-s play a fundamental role

- URI-s made the merge possible
- URI-s ground RDF into the Web
  - information can be retrieved using existing tools
  - this makes the “Semantic Web”, well... “Semantic Web”

# “Internal” nodes

- Consider the following statement:
  - “the publisher is a «thing» that has a name and an address”
- Until now, nodes were identified with a URI. But...
- ...what is the URI of «thing»?



# One solution: create an extra URI

```
<rdf:Description rdf:about="http://.../isbn/000651409X">
  <a:publisher rdf:resource="urn:uuid:f60ffb40-307d-..." />
</rdf:Description>
<rdf:Description rdf:about="urn:uuid:f60ffb40-307d-...">
  <a:p_name>HarpersCollins</a:p_name>
  <a:city>HarpersCollins</a:city>
</rdf:Description>
```

- The resource will be “visible” on the Web
  - care should be taken to define unique URI-s
- Serializations may give syntactic help to define local URI-s

# Internal identifier (“blank nodes”)

```
<rdf:Description rdf:about="http://.../isbn/000651409X">
  <a:publisher rdf:nodeID="A234"/>
</rdf:Description>
<rdf:Description rdf:nodeID="A234">
  <a:p_name>HarpersCollins</a:p_name>
  <a:city>HarpersCollins</a:city>
</rdf:Description>
```

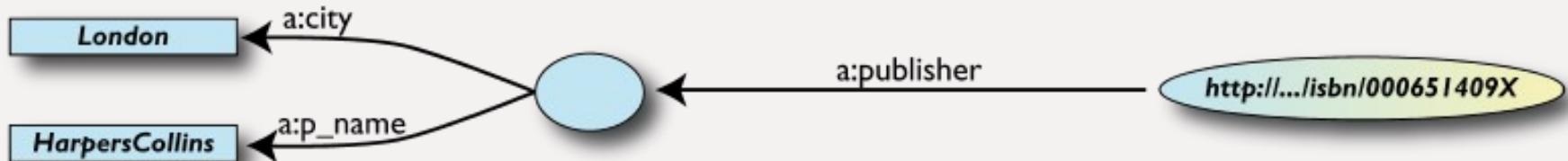
```
<http://.../isbn/2020386682> a:publisher _:A234.
_:A234 a:p_name "HarpersCollins".
```

- Syntax is serialization dependent
- A234 is invisible from outside (it is not a “real” URI!); it is an internal identifier for a resource

# Blank nodes: the system can also do it

- Let the system create a “nodeID” internally (you do not really care about the name...). In Turtle:

```
<http://.../isbn/000651409X> a:publisher [  
    a:p_name "HarpersCollins";  
    ...  
].
```



# Blank nodes: some more remarks

- Blank nodes require attention when merging
  - blanks nodes with identical nodeID-s in *different* graphs are *different*
  - implementations must be careful...
- Many applications prefer not to use blank nodes and define new URI-s “on-the-fly”
  - eg, when triples are in a database
- From a logic point of view, blank nodes represent an “existential” statement
  - “there is a resource such that...”

# RDF in programming practice

- For example, using Java+Jena (HP's Bristol Lab):
  - a “Model” object is created
  - the RDF file is parsed and results stored in the Model
  - the Model offers methods to retrieve:
    - triples
    - (property,object) pairs for a specific subject
    - (subject,property) pairs for specific object
    - etc.
  - the rest is conventional programming...
- Similar tools exist in Python, PHP, etc.

# Jena example

```
// create a model
Model model=new ModelMem();
Resource subject=model.createResource("URI_of_Subject")
// 'in' refers to the input file
model.read(new InputStreamReader(in));
StmtIterator iter=model.listStatements(subject,null,null);
while(iter.hasNext()) {
    st = iter.next();
    p = st.getProperty();
    o = st.getObject();
    do_something(p,o);
}
```

# Merge in practice

- Environments merge graphs automatically
  - e.g., in Jena, the Model can load several files
  - the load merges the new statements automatically

# A relatively simple RDF application

- Goal: reuse of older experimental data
- Keep data in databases or XML, just export key “fact” as RDF
- Use a faceted browser to visualize and interact with the result

**Internal Compound Repurposing Example**

Welcome, Allergy & Respiratory Team Member

This tool allows you to identify opportunities for additional uses of compounds from other teams within your project. It combines internal data, public data and the results of data mining experiments to provide testable hypotheses.

Control Panel & Item Filtering					
Area	5: Approach	3: Term+Reason	1: Max_Stage_Reached	1: Literature Links	
29 Pain	<input checked="" type="checkbox"/> 7: Antibody	<input type="checkbox"/> 31 ACTIVE	<input type="checkbox"/> 51 Candidate	<input checked="" type="checkbox"/> 0 - 50	
16 Metabolic Disease	<input checked="" type="checkbox"/> 1: Recombinant	<input type="checkbox"/> 12 BIOMARKER	<input type="checkbox"/> 10 Discovery	<input type="checkbox"/>	
3: Cancer	<input type="checkbox"/> 18 SM_Agonist	<input checked="" type="checkbox"/> 51 EFFICACY	<input checked="" type="checkbox"/> 41 Exploratory	<input type="checkbox"/>	

**Internal Compound Repurposing Example**

Welcome, Allergy & Respiratory Team Member

This tool allows you to identify opportunities for additional uses of compounds from other teams within your project. It combines internal data, public data and the results of data mining experiments to provide testable hypotheses.

Control Panel & Item Filtering					
Area	5: Approach	3: Term+Reason	1: Max_Stage_Reached	1: Literature Links	
29 Pain	<input checked="" type="checkbox"/> 7: Antibody	<input type="checkbox"/> 31 ACTIVE	<input type="checkbox"/> 51 Candidate	<input checked="" type="checkbox"/> 0 - 50	
16 Metabolic Disease	<input checked="" type="checkbox"/> 1: Recombinant	<input type="checkbox"/> 12 BIOMARKER	<input type="checkbox"/> 10 Discovery	<input type="checkbox"/>	
3: Cancer	<input type="checkbox"/> 18 SM_Agonist	<input checked="" type="checkbox"/> 51 EFFICACY	<input checked="" type="checkbox"/> 41 Exploratory	<input type="checkbox"/>	
3: Sexual Health	<input checked="" type="checkbox"/> 12 SM_Antagonist	<input type="checkbox"/> 11 MARKET	<input type="checkbox"/> 19 HTS	<input type="checkbox"/>	
2: Infectives	<input checked="" type="checkbox"/> 21 SM_Inhibitor	<input type="checkbox"/> 11 REORG	<input type="checkbox"/> 11 Phase I	<input type="checkbox"/>	
1: Urogenitals	<input checked="" type="checkbox"/>	<input type="checkbox"/> 10 TOXIC	<input type="checkbox"/> 13 Phase III	<input type="checkbox"/>	
			<input type="checkbox"/> 41 Screening	<input type="checkbox"/>	

51 items filtered from 710 originally (Reset All Filters)

Area	Original + Indication	Target_Name	Approach	Start	Term+Reason	Max_Stage_Reached	Owner	OMIM Lit_All Lit_2007 Lit_Mech IMA GEO Pathway Compounds
Metabolic Disease	Diabetes	Liver glycogen phosphorylase	SM_Inhibitor	2007-Q2	EFFICACY	Candidate	P. Person	SW-030072
Sexual Health	Erectile Dysfunction	Integrin alpha-3 (Glycoprotein B3) (VLA-3) (CD49c)	SM_Antagonist	2006-Q3	EFFICACY	Candidate	P. Person	SW-029782
Sexual Health	Erectile Dysfunction	Leukotriene C4 synthase	SM_Agonist	2006-Q3	EFFICACY	Candidate	M. Manager	SW-029638
Sexual Health	Erectile Dysfunction	transcription elongation factor A (STF)-like 4	SM_Inhibitor	2005-Q2	EFFICACY	Candidate	P. Person	SW-029926
Infectives	HIV	Putative four-repeat ion channel (ix)	SM_Inhibitor	2006-Q2	EFFICACY	Candidate	L. Leader	SW-029994
Infectives	HIV	Voltage-gated potassium channel protein KV.2 (ix)	SM_Agonist	2007-Q1	EFFICACY	Candidate	A. Scientist	SW-029653
Urogenital Incontinence		Human RNA binding motif (RBM) gene, partial cdk.	SM_Agonist	2007-Q3	EFFICACY	Candidate	L. Leader	SW-029684
Pain	Migraine	Monocarboxylate transporter homologue2294064CD1 (ix)	SM_Inhibitor	2007-Q3	EFFICACY	Candidate	L. Leader	SW-030085

Courtesy of Nigel Wilkinson, Lee Harland, Pfizer Ltd, Melliyal Annamalai, Oracle (SWEO Case Study)

# RDF schemas

# Need for RDF schemas

- First step towards the “extra knowledge”:
  - define the terms we can use
  - what restrictions apply
  - what extra relationships are there?
- Officially: “RDF Vocabulary Description Language”
  - the term “Schema” is retained for historical reasons...

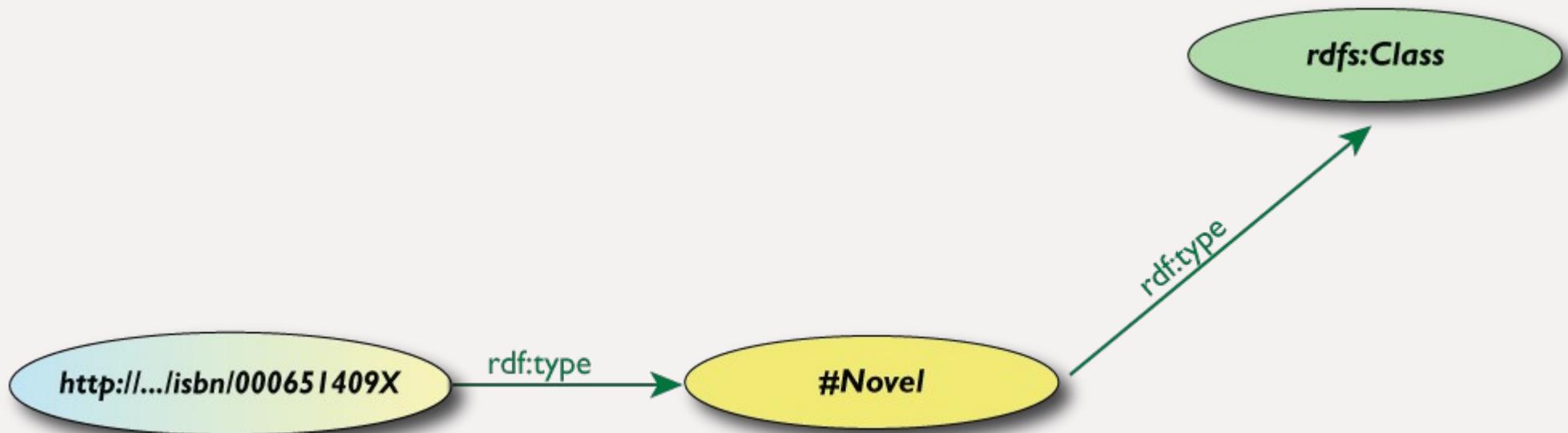
# Classes, resources, ...

- Think of well known traditional ontologies or taxonomies:
  - use the term “novel”
  - “every novel is a fiction”
  - “«The Glass Palace» is a novel”
  - etc.
- RDFS defines resources and classes:
  - everything in RDF is a “resource”
  - “classes” are also resources, but...
  - ...they are also a collection of possible resources (i.e., “individuals”)
    - “fiction”, “novel”, ...

# Classes, resources, ... (cont.)

- Relationships are defined among classes/resources:
  - “typing”: an individual belongs to a specific class:
    - “«The Glass Palace» is a novel”
    - to be more precise: “«<http://.../000651409x>» is a novel”
  - “subclassing”: *all* instances of one are also the instances of the other (“every novel is a fiction”)
- RDFS formalizes these notions in RDF

# Classes, resources in RDF(S)



- RDFS defines the meaning of these terms
  - (these are all special URI-s, we just use the namespace abbreviation)

# Schema example in RDF/XML

- The schema part:

```
<rdf:Description rdf:id="Novel">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
</rdf:Description>
```

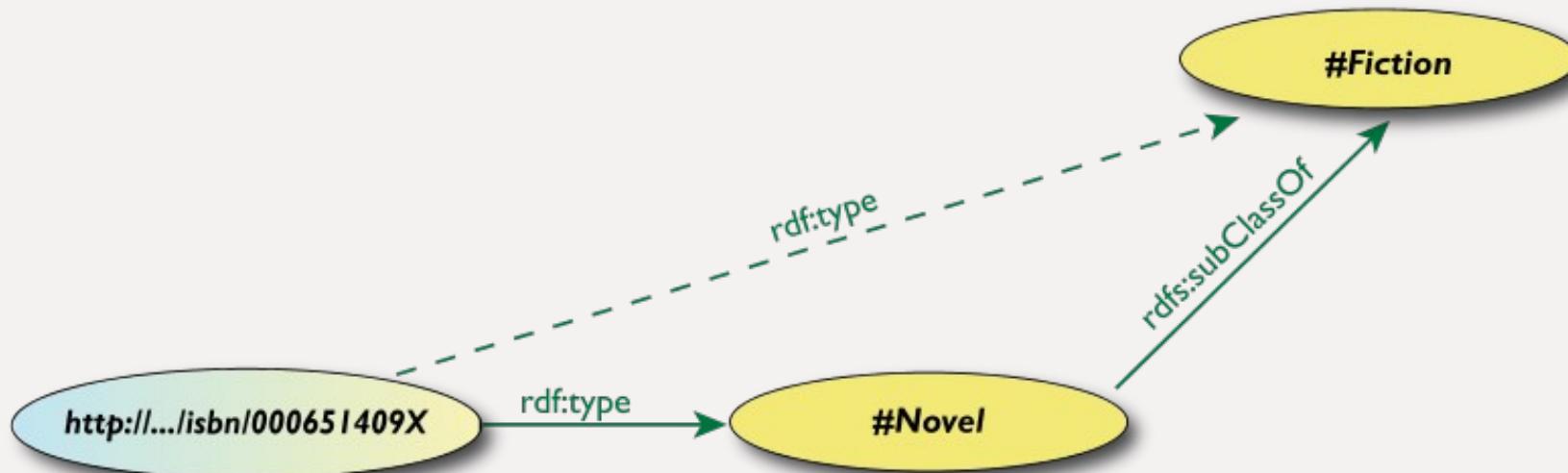
- The RDF data on a specific novel:

```
<rdf:Description rdf:about="http://.../isbn/000651409X">
  <rdf:type rdf:resource="http://.../bookSchema.rdf#Novel"/>
</rdf:Description>
```

# On types

- The type information may be very important for applications
  - e.g., it may be used for a categorization of possible nodes
  - probably the most frequently used RDF property...
  - (remember the “Person” in our example?)

# Inferred properties



```
(<http://.../isbn/000651409X> rdf:type #Fiction)
```

- is not in the original RDF data...
- ...but can be inferred from the RDFS rules
- RDFS environments return that triple, too

# Inference: let us be formal...

- The [RDF Semantics](#) document has a list of (44) *entailment rules*:
  - “if such and such triples are in the graph, add this and this”
  - do that recursively until the graph does not change
- The relevant rule for our example:

If:

```
uuu rdfs:subClassOf xxx .  
vvv rdf:type uuu .
```

Then add:

```
vvv rdf:type xxx .
```

# Properties

- Property is a special class (**rdf:Property**)
  - properties are also resources identified by URI-s
- Properties' range and domain can be specified
  - i.e., what type of resources can serve as object and subject
- There is also a possibility for a “sub-property”
  - all resources bound by the “sub” are also bound by the other

# Property specification serialized

- In RDF/XML:

```
<rdf:Property rdf:id="title">
  <rdfs:domain rdf:resource="#Fiction"/>
  <rdfs:range rdf:resource="http://...#Literal"/>
</rdf:Property>
```

- In Turtle:

```
:title
  rdf:type    rdf:Property;
  rdfs:domain :Fiction;
  rdfs:range   rdfs:Literal.
```

# What does this mean?

- Again, new relations can be deduced. Indeed, if

```
:title
  rdf:type    rdf:Property;
  rdfs:domain :Fiction;
  rdfs:range   rdfs:Literal.

<http://.../isbn/000651409x> :title "The Glass Palace" .
```

- then the system can infer that:

```
<http://.../isbn/000651409x> rdf:type :Fiction .
```

# A bit of RDFS can take you far...

- Remember the power of merge?
- We could have used, in our example:
  - **f:auteur** is a subproperty of **a:author** and vice versa (although we will see other ways to do that...)
- Of course, in some cases, more complex knowledge is necessary (see later...)

# Vodafone live!

- Integrate various vendors' product descriptions via RDF
  - ring tones, games, wallpapers
  - manage complexity of handsets, binary formats
- A portal is created to offer appropriate content
- Significant increase in content download after the introduction



Courtesy of Kevin Smith, Vodafone Group R&D ([SWEO Case Study](#))

# Get to RDF(S) data

# Simple approach

- Write RDF/XML or Turtle “manually”
- In some cases that is necessary, but it really does not scale...

# RDF can also be extracted/generated

- Use intelligent “scrapers” or “wrappers” to extract a structure (hence RDF) from a Web pages or XML files...
- ... and then generate RDF automatically (e.g., via an XSLT script)

# Formalizing the scraper approach: GRDDL

- GRDDL formalizes scrapers:

```
<html xmlns="http://www.w3.org/1999/">
  <head profile="http://www.w3.org/2003/g/data-view">
    <title>Some Document</title>
    <link rel="transformation" href="http:.../dc-extract.xsl"/>
    <meta name="DC.Subject" content="Some subject"/>
    ...
  </head>
  ...
  <span class="date">2006-01-02</span>
  ...
</html>
```

- yields, via **dc-extract.xsl**:

```
<rdf:Description rdf:about="...">
  <dc:subject>Some subject</dc:subject>
  <dc:date>2006-01-02</dc:date>
</rdf:Description>
```

# GRDDL

- The transformation itself has to be provided for each set of conventions (meta-s, class id-s, etc...)
- A “bridge” to “microformats”
- A method to get data in other formats to RDF (e.g., XBRL)

# Another solution: RDFa

- For example:

```
<div about="http://uri.to.newsitem">
  <span property="dc:date">March 23, 2004</span>
  <span property="dc:title">Rollers hit casino for £1.3m</span>
  By <span property="dc:creator">Steve Bird</span>. See
  <a href="http://www.a.b.c/d.avi" rel="dc:type:MovingImage">
    also video footage</a>...
</div>
```

- yields:

```
<http://uri.to.newsitem>
  dc:date          "March 23, 2004";
  dc:title         "Rollers hit casino for £1.3m;
  dc:creator       "Steve Bird";
  dc:type:MovingImage <http://www.a.b.c/d.avi>.
```

# RDFa (cont.)

- RDFa extends (X)HTML a bit by:
  - defining general attributes to add data to any elements
  - an almost complete “serialization” of RDF in XHTML
- It is a bit like the microformats/GRDDL approach but with more rigour and fully generic

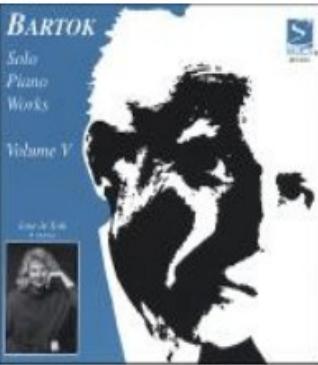
# A typical RDFa usage: the visible page

Bitmunk: Collection: Bartok Solo Piano Works, Volume 5 by June de Toth - Mozilla Firefox

File Edit View History Bookmarks Tools Help

<http://bitmunk.com/media/6071604>

 audio news downloads help



**Bartok Solo Piano Works, Volume 5**

**June de Toth**

June de Toth's interpretation of Bartk's solo piano music is decidedly her own. She expresses a lyrical and romantic vision of his early poetic works, which were very much influenced by the French impressionist composer Claude Debussy.

**Purchase**

WebBuy directly from Bitmunk

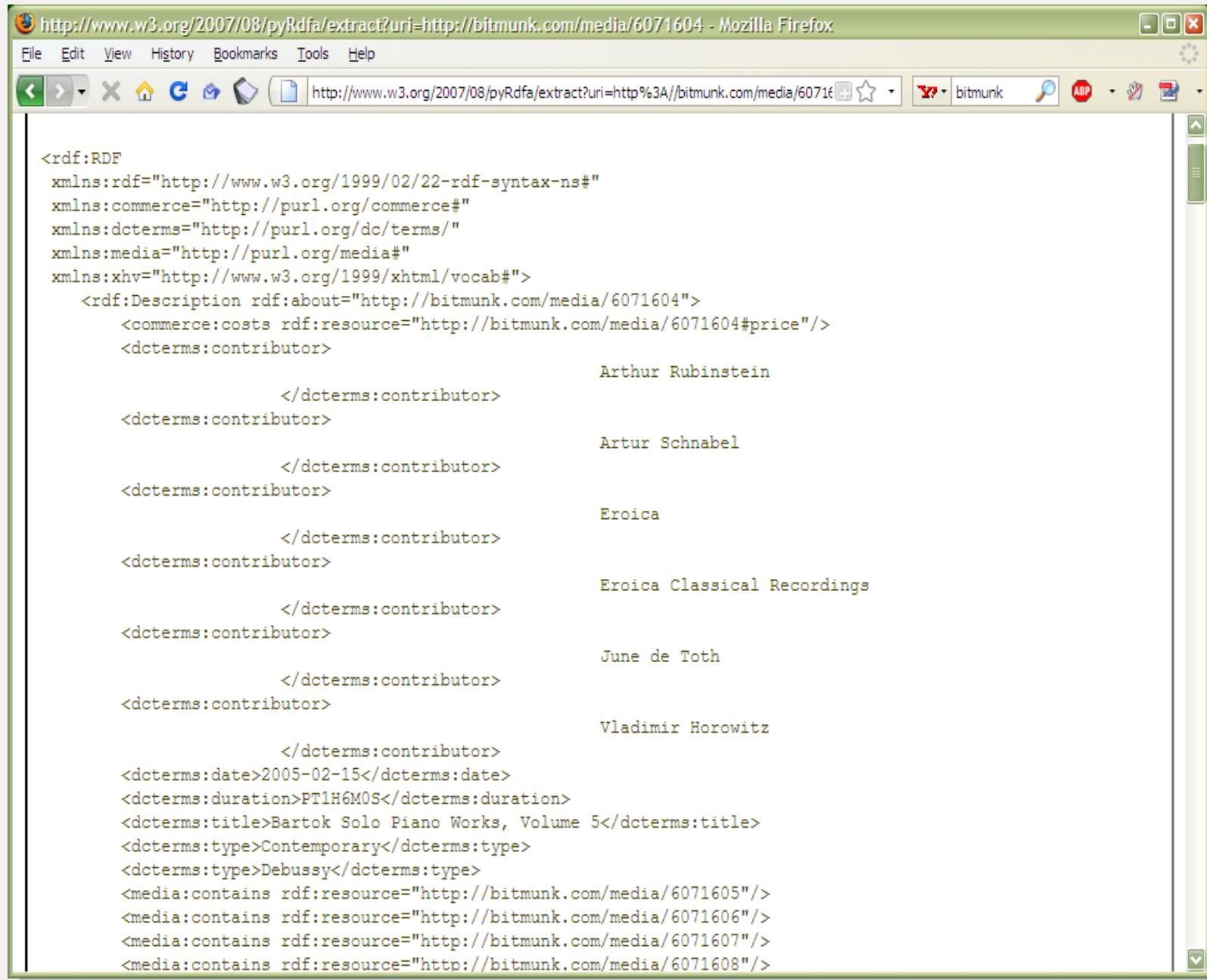
**WebBuy ➔** for USD \$8.15

**Sample**

 Play all samples in this collection.

<b>Type</b>	<input checked="" type="radio"/> Collection (audio)
<b>Copyright Owner</b>	Eroica
<b>Music Label</b>	Eroica
<b>Performer</b>	June de Toth
<b>Publisher</b>	Eroica Classical Recordings
<b>Similar Artist</b>	Vladimir Horowitz Arthur Rubinstein Artur Schnabel
<b>Released</b>	2005-02-15
<b>Genre</b>	Audio > Debussy Audio > Contemporary
<b>Items</b>	27 (1:06:00)
<b>Owner</b>	cdbaby

# ... and what is inside as metadata



The screenshot shows a Mozilla Firefox window displaying the results of an RDFa extraction. The address bar shows the URL: <http://www.w3.org/2007/08/pyRdfa/extract?uri=http://bitmunk.com/media/6071604>. The page content is an RDFa dump with various namespaces defined:

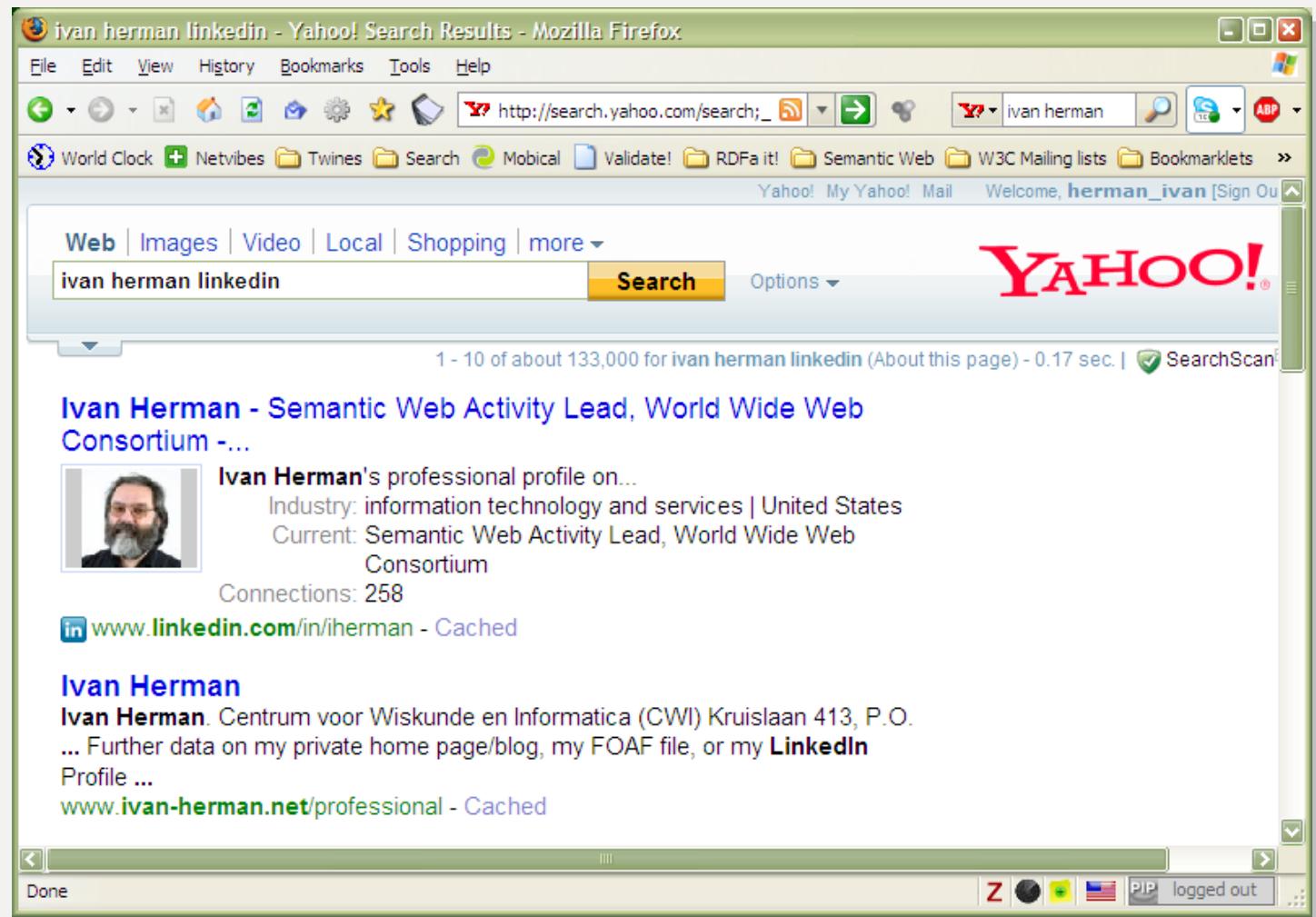
```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:commerce="http://purl.org/commerce#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:media="http://purl.org/media#"
  xmlns:xhv="http://www.w3.org/1999/xhtml/vocab#">
  <rdf:Description rdf:about="http://bitmunk.com/media/6071604">
    <commerce:costs rdf:resource="http://bitmunk.com/media/6071604#price"/>
    <dcterms:contributor>
      Arthur Rubinstein
    </dcterms:contributor>
    <dcterms:contributor>
      Artur Schnabel
    </dcterms:contributor>
    <dcterms:contributor>
      Eroica
    </dcterms:contributor>
    <dcterms:contributor>
      Eroica Classical Recordings
    </dcterms:contributor>
    <dcterms:contributor>
      June de Toth
    </dcterms:contributor>
    <dcterms:contributor>
      Vladimir Horowitz
    </dcterms:contributor>
    <dcterms:date>2005-02-15</dcterms:date>
    <dcterms:duration>PT1H6M0S</dcterms:duration>
    <dcterms:title>Bartok Solo Piano Works, Volume 5</dcterms:title>
    <dcterms:type>Contemporary</dcterms:type>
    <dcterms:type>Debussy</dcterms:type>
    <media:contains rdf:resource="http://bitmunk.com/media/6071605"/>
    <media:contains rdf:resource="http://bitmunk.com/media/6071606"/>
    <media:contains rdf:resource="http://bitmunk.com/media/6071607"/>
    <media:contains rdf:resource="http://bitmunk.com/media/6071608"/>
  </rdf:Description>
</rdf:RDF>

```

# Yahoo's SearchMonkey

- Search results may be customized via small applications using content metadata in, eg, RDFa
- Users can customize their search pages

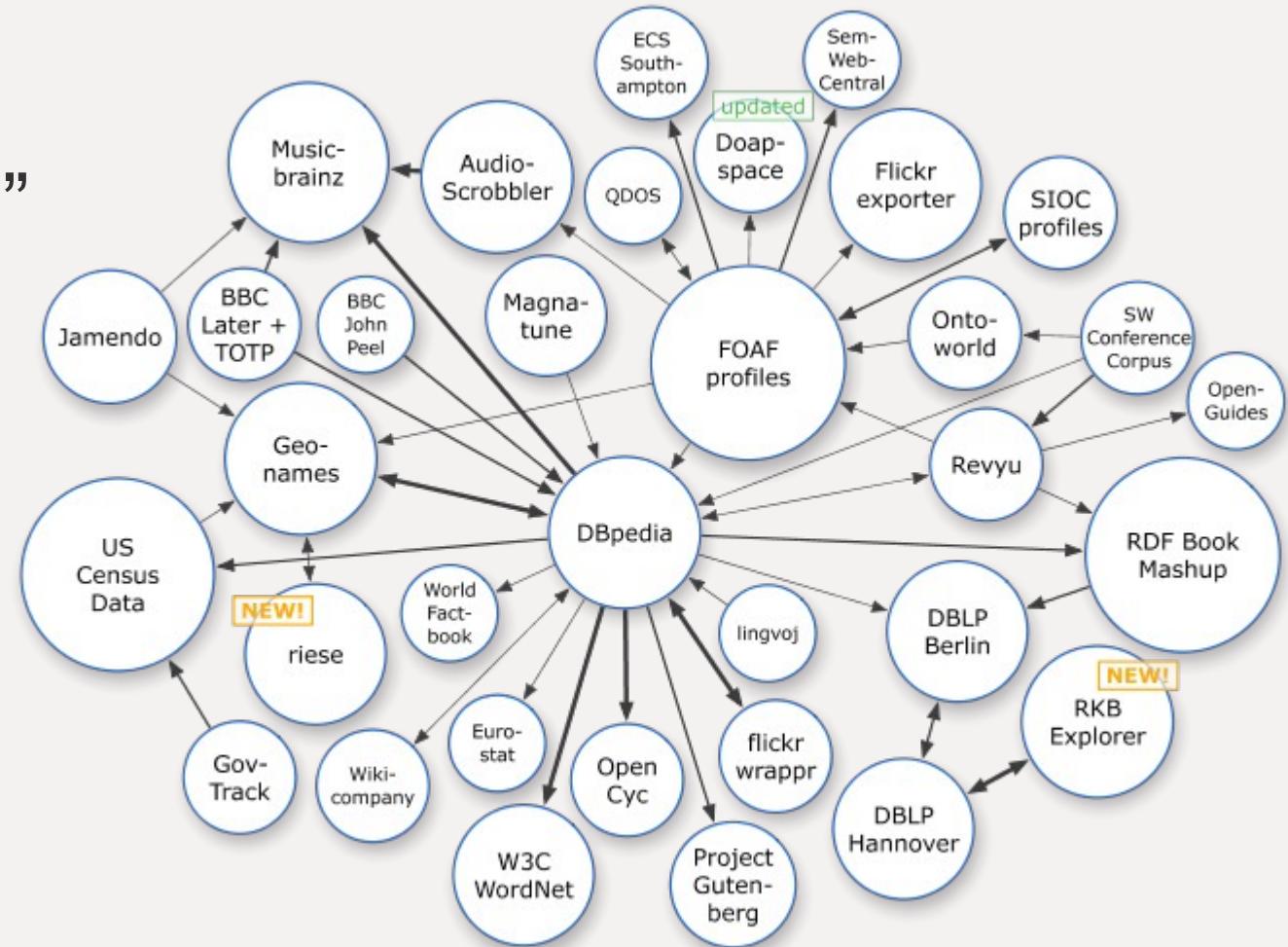


# Bridges to relational databases

- Data on the Web are mostly stored in databases
- “Bridges” are being defined:
  - a layer between RDF and the relational data
    - RDB tables are “mapped” to RDF graphs, possibly on the fly
    - different mapping approaches are being used
  - a number RDB systems offer this facility already (eg, Oracle, OpenLink, ...)
- Work for a survey on mapping techniques at W3C

# Linking Open Data Project

- Goal: “expose” open datasets in RDF
- Set *RDF links among the data items* from different datasets
- Billions triples,  
millions of “links”



# DBpedia: Extracting structured data from Wikipedia

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

```
<http://dbpedia.org/resource/Kolkata>
  dbpedia:native_name "Kolkata (Calcutta)"@en;
  dbpedia:altitude "9";
  dbpedia:populationTotal "4580544";
  dbpedia:population_metro "14681589";
  geo:lat "22.56970024108887"^^xsd:float;
  ...
  
```

Kolkata (Chutiyon ka Shahar) ?

West Bengal • India



Victoria Memorial

Coordinates:  22.5697, 88.3697

Time zone	IST (UTC+5:30)
Area	1,480 km <sup>2</sup> (571 sq mi)
• Elevation	• 9 m (30 ft)
District(s)	Calcutta †
Population	14,681,589 (2001)
• Density	• 9,920/km <sup>2</sup> (25,693/sq mi)
• Metro	• 4,580,544
Mayor	Bikash Ranjan Bhattacharya
Codes	
• Pincode	• 700 xxxx
• Telephone	• +91 (33)
Website:	<a href="http://www.kolkatamyacity.com">www.kolkatamyacity.com</a> 

† The Kolkata urban agglomeration also includes portions of North 24 Parganas and South 24 Parganas districts.

# Automatic links among open datasets

```
<http://dbpedia.org/resource/Kolkata>
owl:sameAs <http://sws.geonames.org/1275004/>;
...
...
```

DBpedia

Geonames

```
<http://sws.geonames.org/1275004/>
owl:sameAs <http://DBpedia.org/resource/Kolkata>
wgs84_pos:lat "22.5697222";
wgs84_pos:long "88.3697222";
sws:population "4631392"
...
...
```



Processors can switch automatically from one to the other...

# Example usage: “Review Anything”

Licence to Kill - Things - Revyu.com - Mozilla Firefox

File Edit View History Bookmarks Tools Help

World Clock Netvibes Twines Search Mobical Validate! RDFa it! Semantic Web W3C Mailing lists Bookmarklets Python

[Home](#) | [Browse Things](#) | [Search Things](#) | [Browse People](#)

[Login/Register](#) | [New Review](#)

**Licence to Kill**

**Links**  
See Also: [http://en.wikipedia.org/wiki/Licence\\_to\\_kill](http://en.wikipedia.org/wiki/Licence_to_kill)

**Tags**  
[action](#) [film](#) [james-bond](#) [movie](#)

**Reviews (1)**

 by tom on 31 Dec 2006  
Utterly forgettable Bond film. Over the top action sequences, unconvincing romances, and a disjointed storyline. There aren't even any good Bond one-liners. Passes the time but not much else.

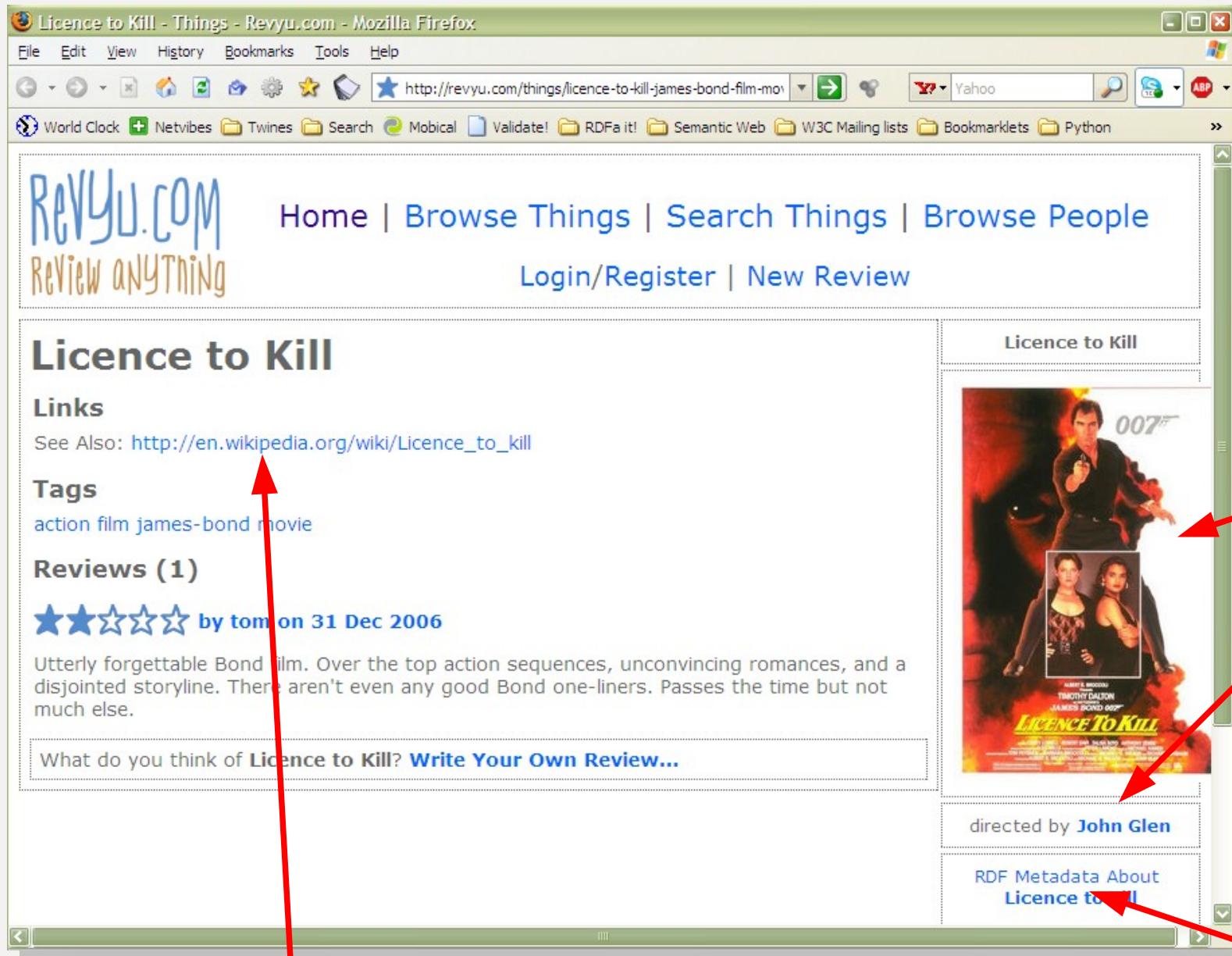
What do you think of Licence to Kill? [Write Your Own Review...](#)

**Licence to Kill**



directed by [John Glen](#)

RDF Metadata About [Licence to Kill](#)



links to, eg, (DB/Wiki)Pedia

enhance output with linked data

data in RDF

# Faviki: social bookmarking with Wiki tagging

- Tag bookmarks via Wikipedia terms/DBpedia URIs
- Helps disambiguating tag usage

The screenshot shows a web browser window for the Faviki website. The title bar reads "RDFa | Faviki. Tags that make sense. - Opera". The address bar shows the URL "http://www.faviki.com/?subject=&object=RDFa;&search=or". The Faviki logo is at the top left, and the user's name "ivan" is at the top right along with "friends". A sidebar on the right lists "Topics: Semantic web, XML-based standards, Metadata, W3C standards, Knowledge representation, Semantic HTML" and a link "W Read more on Wikipedia". The main content area displays a search result for "RDFa" with a list of bookmarks. The results are grouped by month: "This Month" and "August". Each entry includes the user, the title, and the date.

Month	User	Title	Date
This Month	ivan	RDFa Primer	4:41 pm
	ivan	LCSH Search	4:31 pm
	jupy	RDFLib: Home	Sep 07
August	ptraca	The Semantic Web Gang	Aug 25
	ptraca	W3C Semantic Web Activity	Aug 25
	masaka	SearchMonkey Tutorials (Yahoo! Developer Network Blo	Aug 20
	masaka	Yahoo! Search Blog: The Yahoo! Search Open Ecosyste	Aug 20

# RDF data access, a.k.a. query (SPARQL)

# RDF data access

- How do I query the RDF data?
  - e.g., how do I get to the DBpedia data?

# Querying RDF graphs

- Remember the Jena idiom:

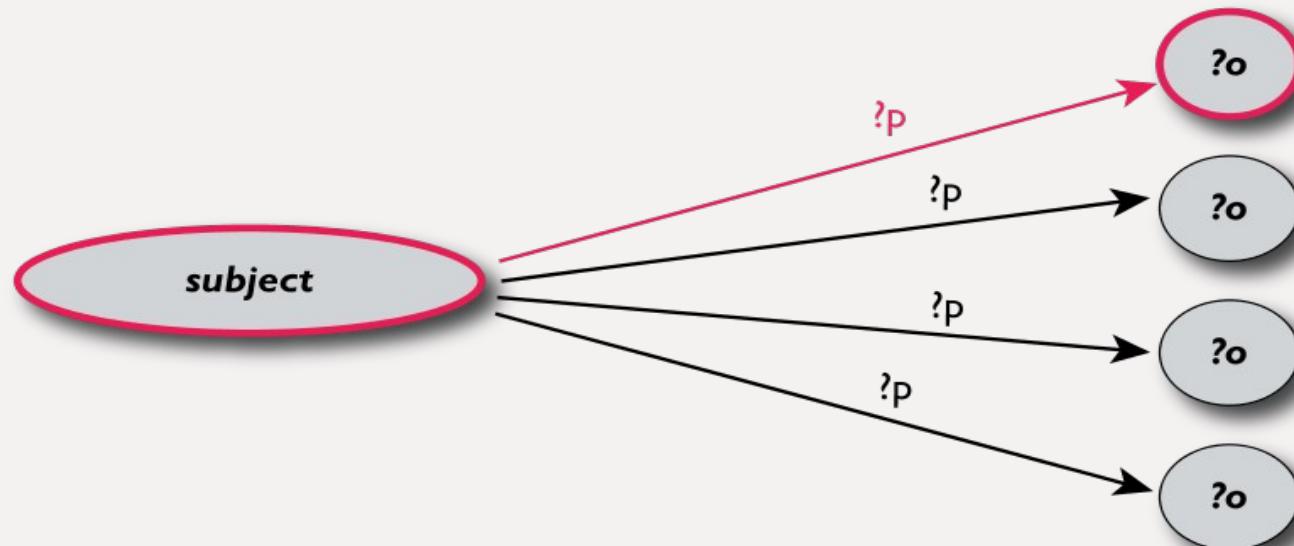
```
StmtIterator iter=model.listStatements(subject,null,null);  
while(iter.hasNext()) {  
    st = iter.next();  
    p = st.getProperty(); o = st.getObject();  
    do_something(p,o);
```

- In practice, more complex queries into the RDF data are necessary
  - something like: “give me the (a,b) pair of resources, for which there is an x such that (x parent a) and (b brother x) holds” (ie, return the uncles)
  - these rules may become quite complex
- The goal of **SPARQL** (Query Language for RDF)

# Analyze the Jena example

```
StmtIterator iter=model.listStatements(subject,null,null);  
while(iter.hasNext()) {  
    st = iter.next();  
    p = st.getProperty(); o = st.getObject();  
    do_something(p,o);
```

- The  $(\text{subject}, ?p, ?o)$  is a *pattern* for what we are looking for (with  $?p$  and  $?o$  as “unknowns”)



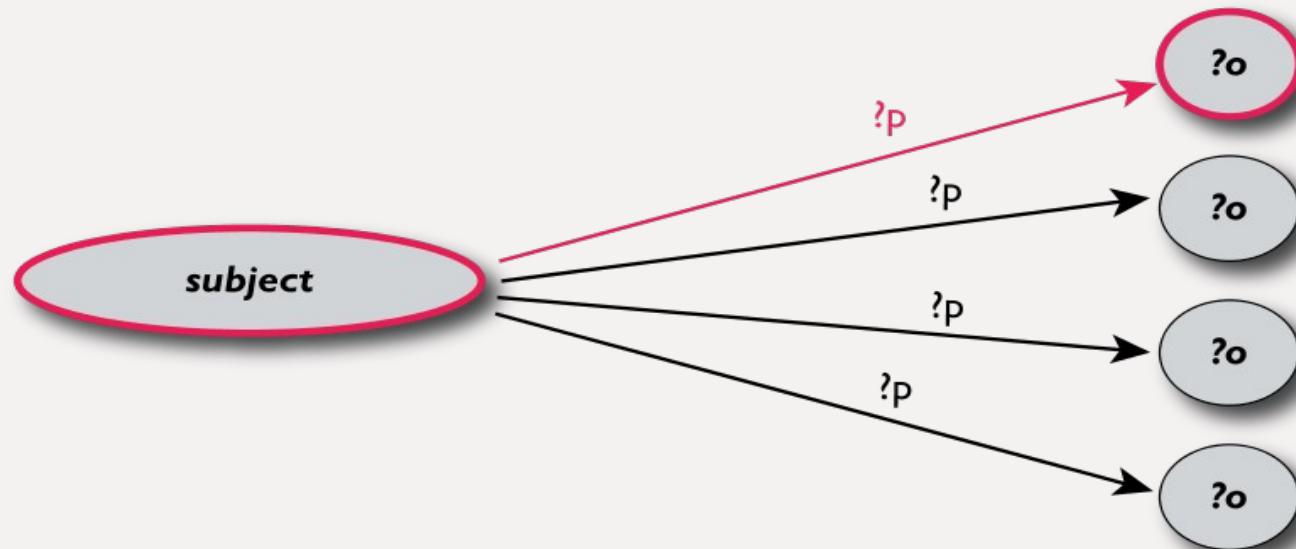
# General: graph patterns

- The fundamental idea: use graph patterns
  - the pattern contains unbound symbols
  - by binding the symbols, subgraphs of the RDF graph are selected
  - if there is such a selection, the query returns bound resources

# Our Jena example in SPARQL

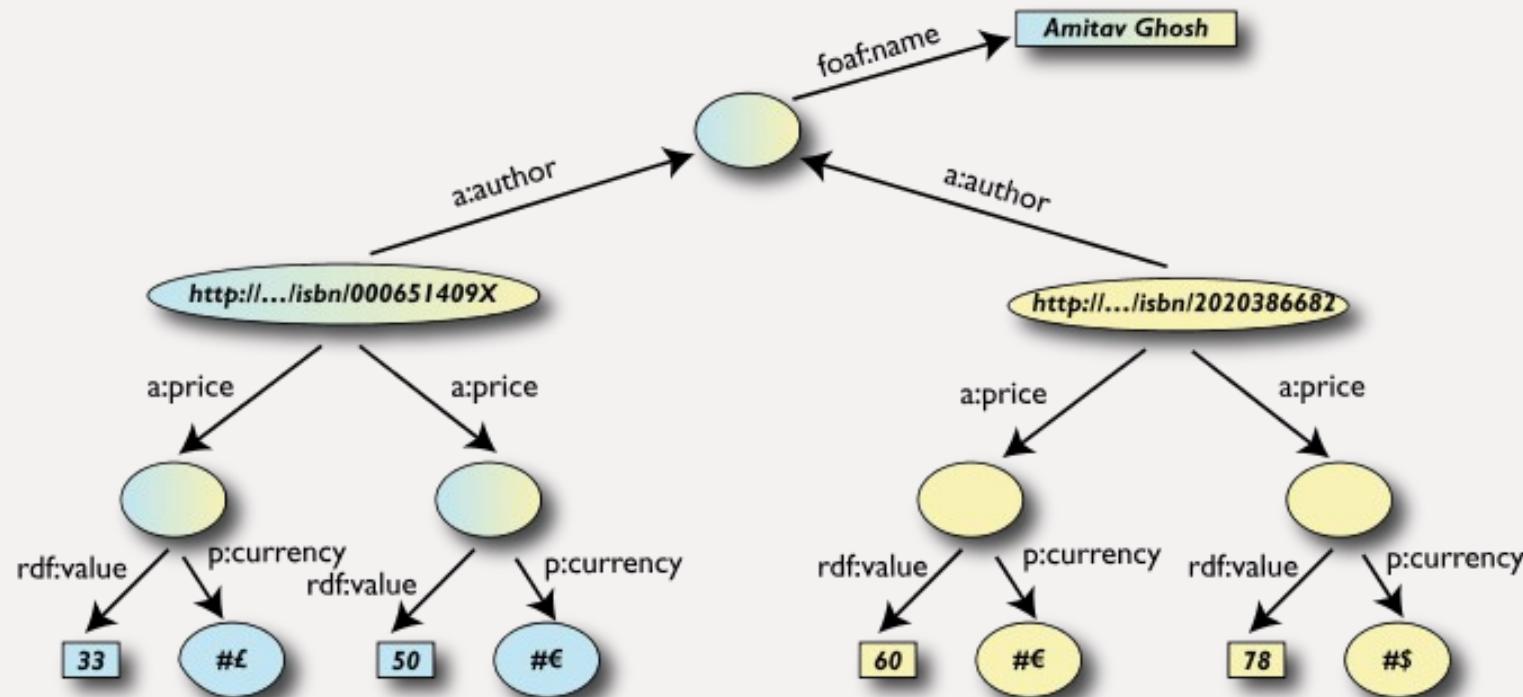
```
SELECT ?p ?o  
WHERE {subject ?p ?o}
```

- The triples in WHERE define the graph pattern, with ?p and ?o “unbound” symbols
- The query returns all p,o pairs



# Simple SPARQL example

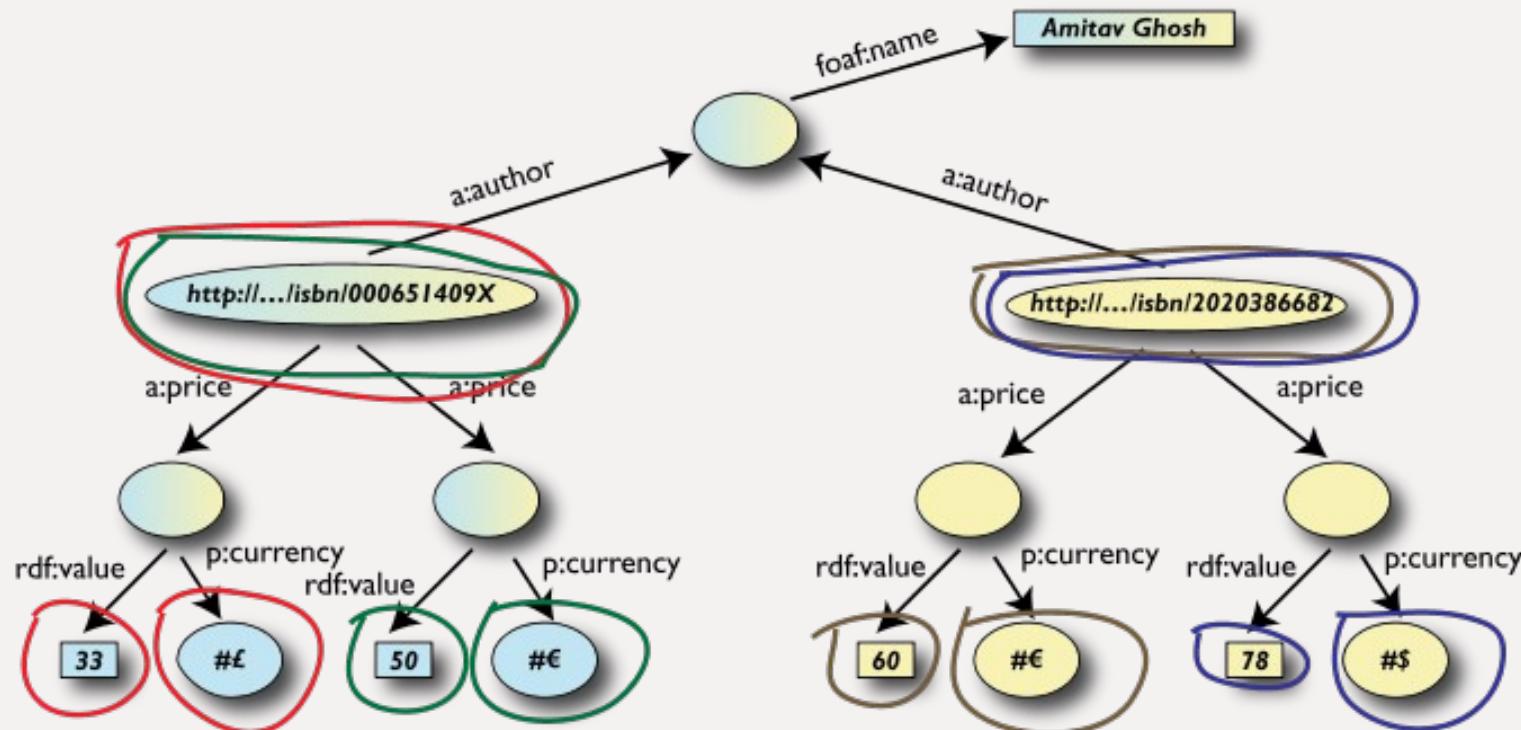
```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```



# Simple SPARQL example

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```

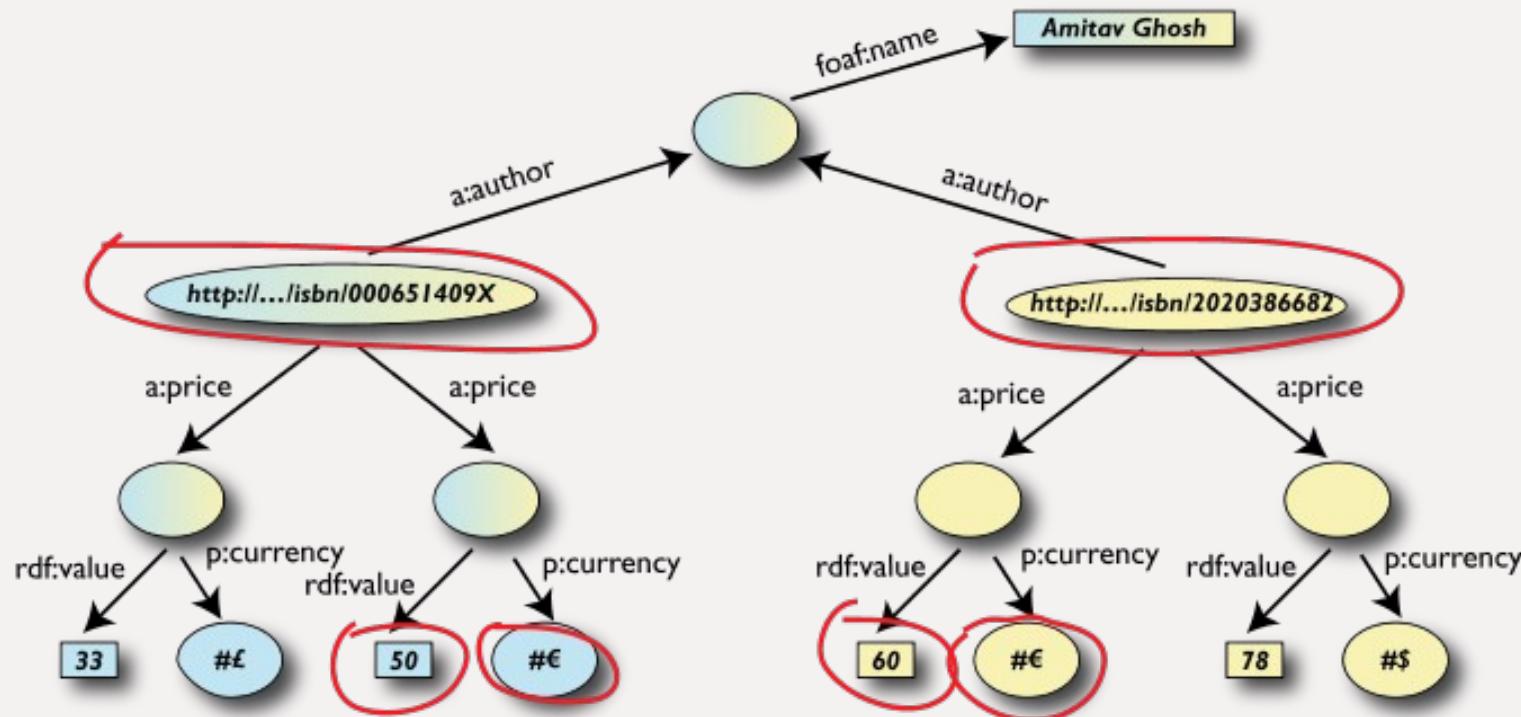
- Returns:  
[[<..49X>,33,£], [<..49X>,50,€], [<..6682>,60,€], [<..6682>,78,\$]]



# Pattern constraints

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE { ?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency .
        FILTER(?currency == € ) }
```

- Returns: [[<..409X>,50,€], [<..6682>,60,€]]



# Some other SPARQL features

- Some of the patterns may be optional
- Limit the number of returned results; remove duplicates, sort them, ...
- Specify several data sources (via URI-s) within the query (essentially, a merge!)
- Construct a graph combining a separate pattern and the query results

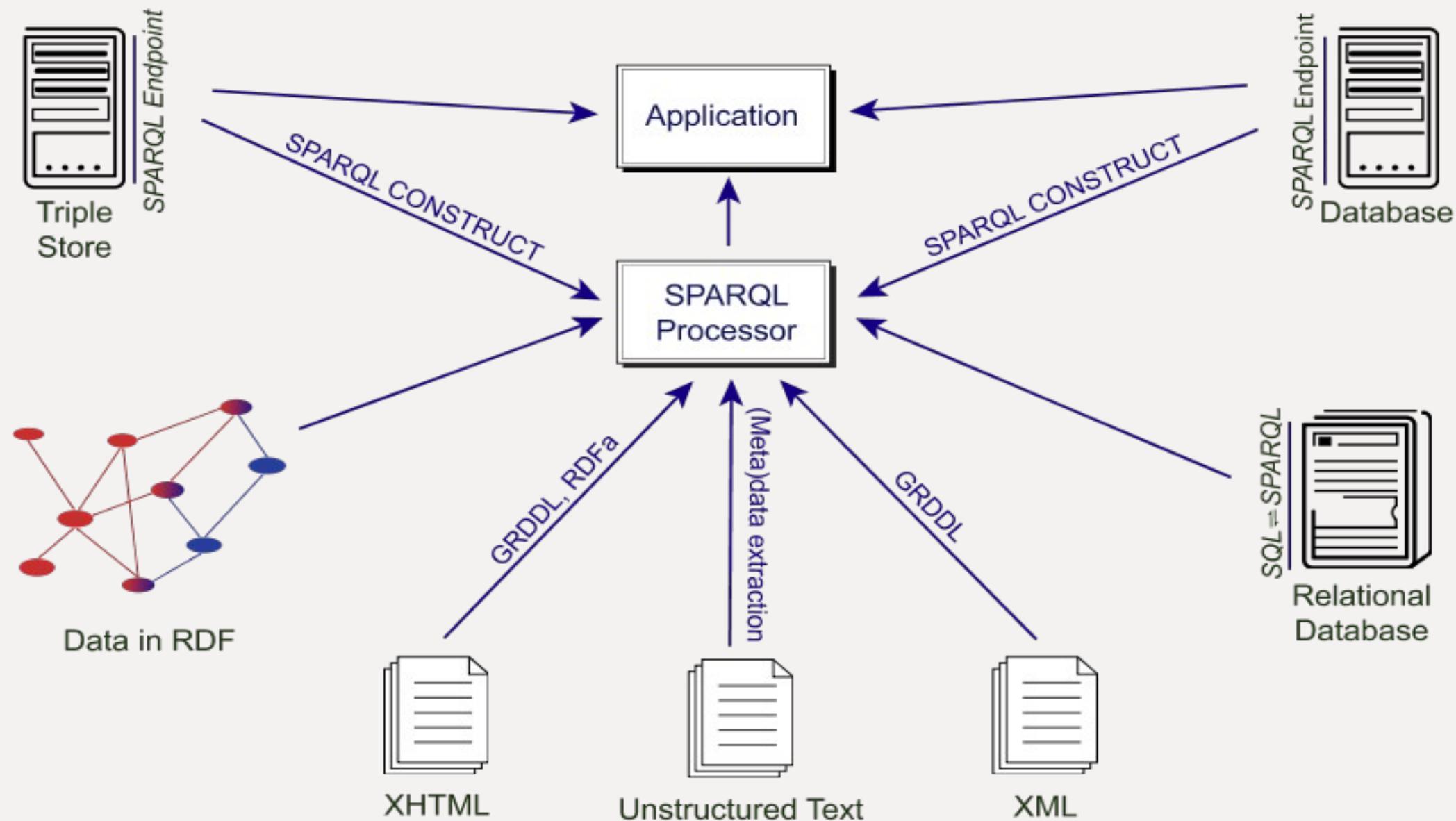
# SPARQL usage in practice

- SPARQL is usually used over the network
  - separate documents define the protocol and the result format
    - SPARQL Protocol for RDF with HTTP and SOAP bindings
    - SPARQL results in XML or JSON formats
- Big datasets usually offer “SPARQL endpoints” using this protocol

# A word of warning on SPARQL...

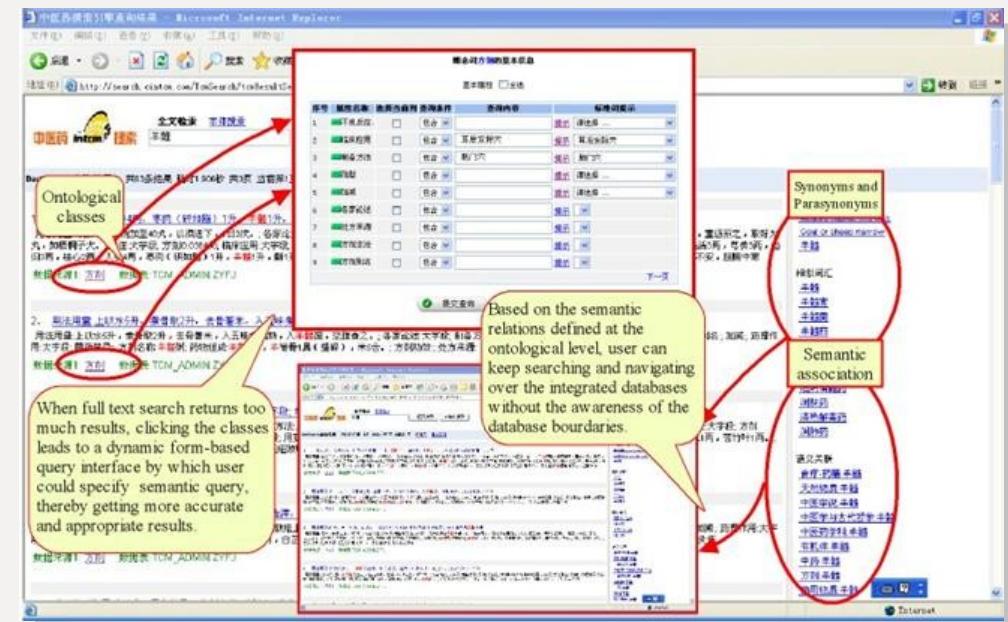
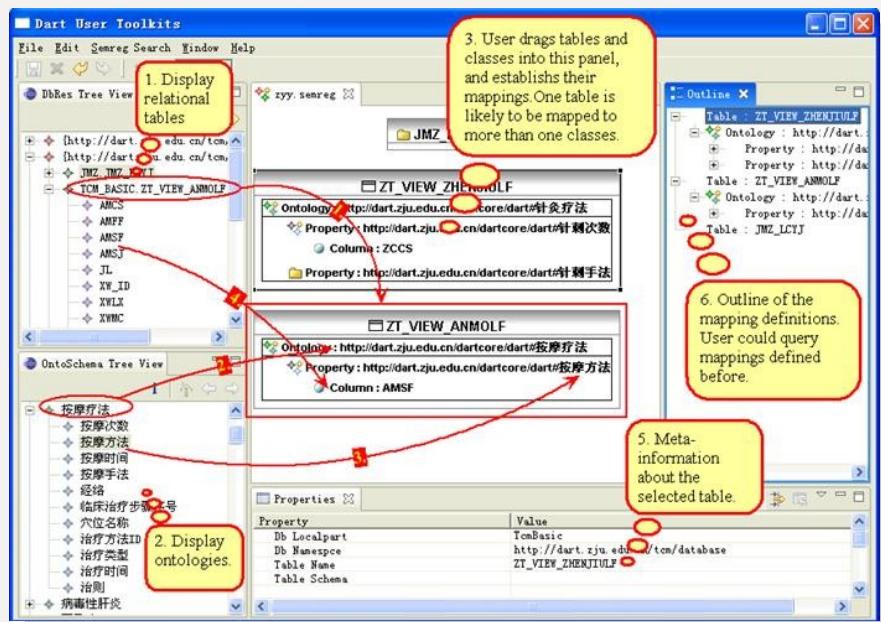
- Some features are missing
  - control and/or description on the entailment regimes of the triple store (RDFS? OWL-DL? OWL-Lite? ...)
  - modify the triple store
  - querying collections or containers may be complicated
  - no functions for sum, average, min, max, ...
  - ways of aggregating queries
  - ...
- Delayed for a next version...

# SPARQL as a unifying point



# Integrate knowledge for Chinese Medicine

- Integration of a large number of TCM databases
  - around 80 databases, around 200,000 records each
- A visual tool to map databases to the semantic layer using a specialized ontology
- Form based query interface for end users



Courtesy of Huajun Chen, Zhejiang University, (SWEO Case Study)

# Ontologies (OWL)

# Ontologies

- RDFS is useful, but does not solve all possible requirements
- Complex applications may want more possibilities:
  - characterization of properties
  - identification of objects with different URI-s
  - disjointness or equivalence of classes
  - construct classes, not only name them
  - more complex classification schemes
  - can a program reason about some terms? E.g.:
    - “if «Person» resources «A» and «B» have the same «`foaf:email`» property, then «A» and «B» are identical”
  - etc.

# Ontologies (cont.)

- The term ontologies is used in this respect:

“defines the concepts and relationships used to describe and represent an area of knowledge”

- Ie, there is a need for Web Ontology Language(s)
  - RDFS can be considered as a simple ontology language
- Languages should be a compromise between
  - rich semantics for meaningful applications
  - feasibility, implementability

# Web Ontology Language = OWL

- OWL is an extra layer, a bit like RDF Schemas
  - own namespace, own terms
  - it relies on RDF Schemas
- It is a separate recommendation
- There is an active W3C Working Group working on extensions of the current standards
  - the new version will be called “OWL 2”
  - in what follows, some features will be referred to as “may come in future”, i.e., under consideration by that group

# OWL is complex...

- OWL is a large set of additional terms
- We will not cover the whole thing here...

# First some simple features

# Term equivalence

- For classes:
  - **owl:equivalentClass**: two classes have the same individuals
  - **owl:disjointWith**: no individuals in common
- For properties:
  - **owl:equivalentProperty**
    - remember the **a:author** vs. **f:auteur**?
- For individuals:
  - **owl:sameAs**: two URIs refer to the same concept (“individual”)
  - **owl:differentFrom**: negation of **owl:sameAs**

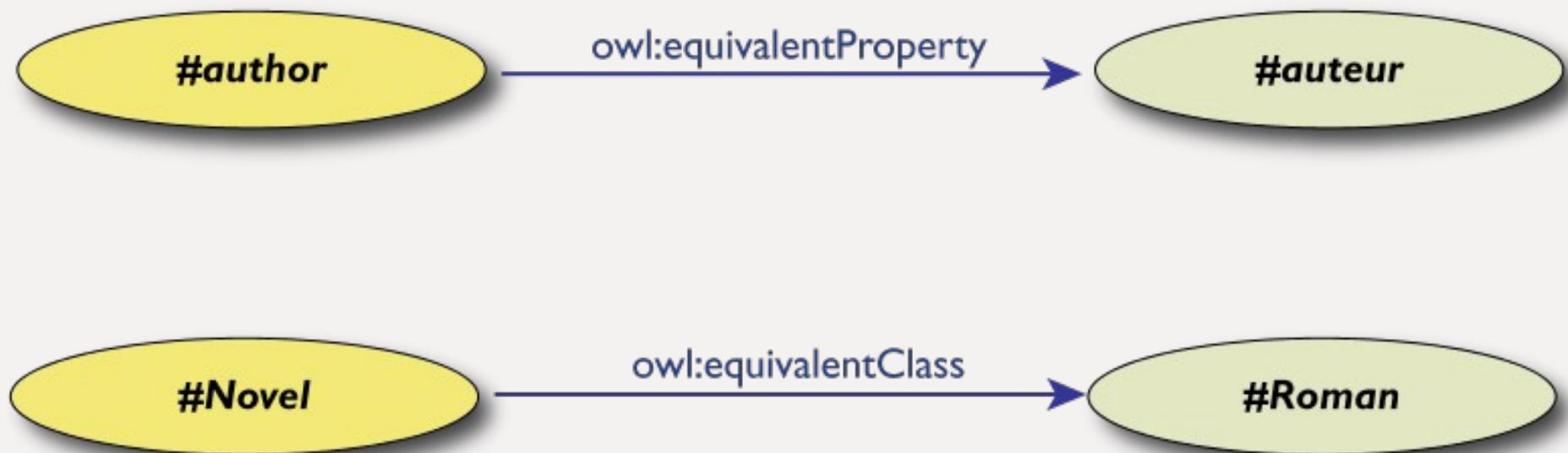
# Typical usage of owl:sameAs

- Linking our example of Kolkata from one data set (DBpedia) to the other (Geonames):

```
<http://dbpedia.org/resource/Kolkata>
owl:sameAs <http://sws.geonames.org/1275004/>;
```

- This is the main mechanism of “Linking” in the Linking Open Data project

# Other example: connecting to French

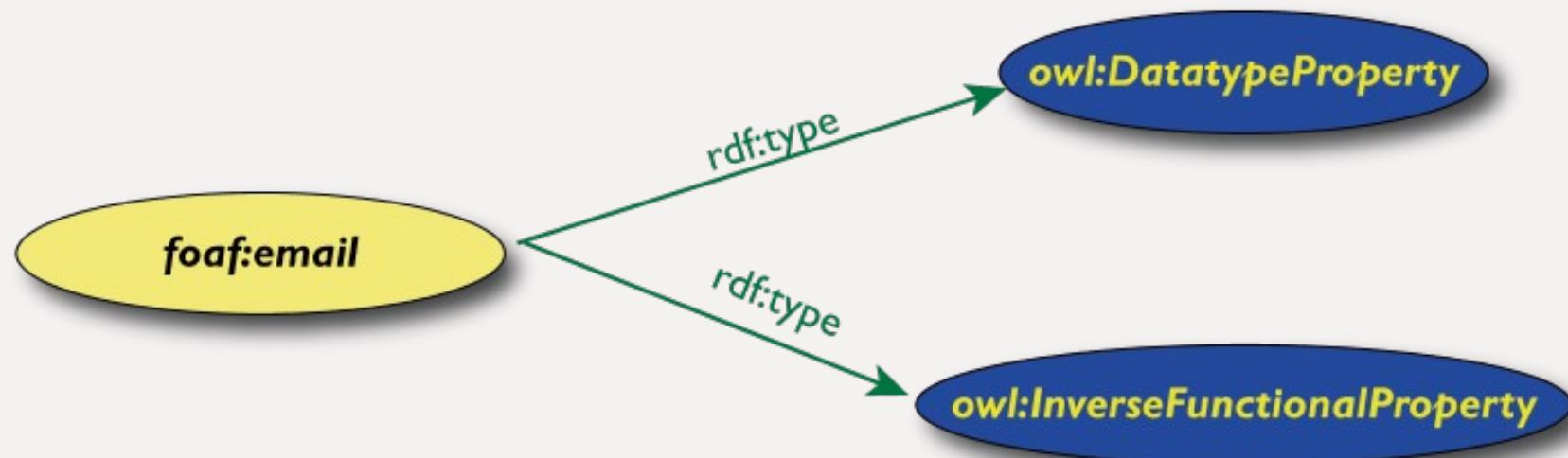


# Property characterization

- In OWL, one can characterize the behaviour of properties (symmetric, transitive, functional, inverse functional...)
- OWL also separates *data* and *object* properties
  - “datatype property” means that its range are typed literals

# Characterization example

- “**foaf:email**” is inverse functional (i.e., two different subjects cannot have identical objects)



# What this means is...

- If the following holds in our triples:

```
:email rdf:type owl:InverseFunctionalProperty.  
<A> :email "mailto:a@b.c".  
<B> :email "mailto:a@b.c".
```

- then the following holds, too:

```
<A> owl:sameAs <B>.
```

- I.e., *new relationships* were discovered again  
(beyond what RDFS could do)

# Other property characterizations

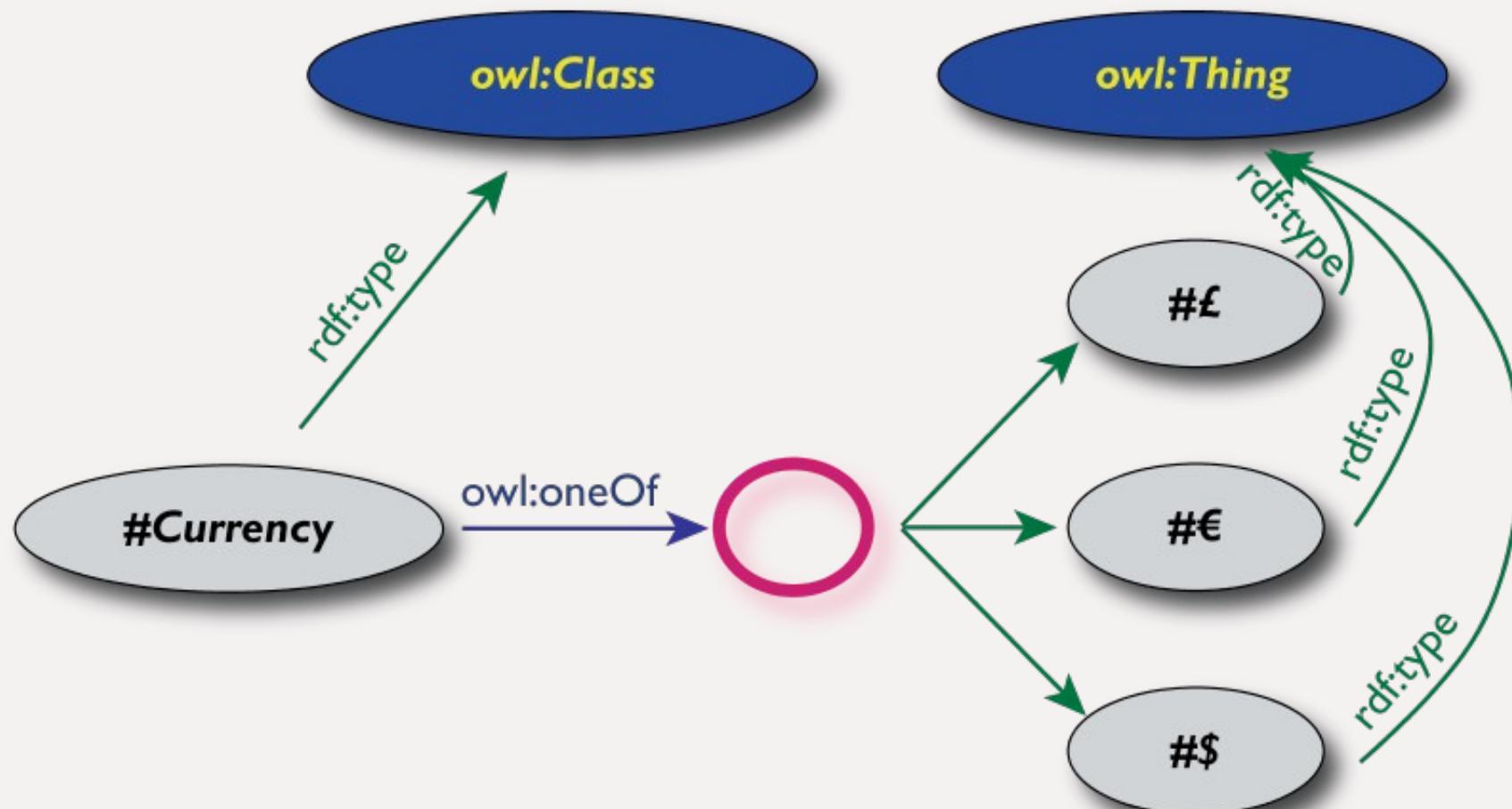
- Functional property (“owl:FunctionalProperty”)
- Transitive property (“owl:TransitiveProperty”)
- Symmetric property (“owl:SymmetricProperty”)
- Inverse of another property (“owl:inverseOf”)
- May come in future:
  - reflexive and irreflexive object properties
  - specify that properties are “disjoint”

# Classes in OWL

- In RDFS, you can subclass existing classes... that's all
- In OWL, you can construct classes from existing ones:
  - enumerate its content
  - through intersection, union, complement
  - etc
- OWL makes a stronger distinction between classes and individuals
  - referring to its own **Class** and to “Thing”, respectively

# OWL classes can be “enumerated”

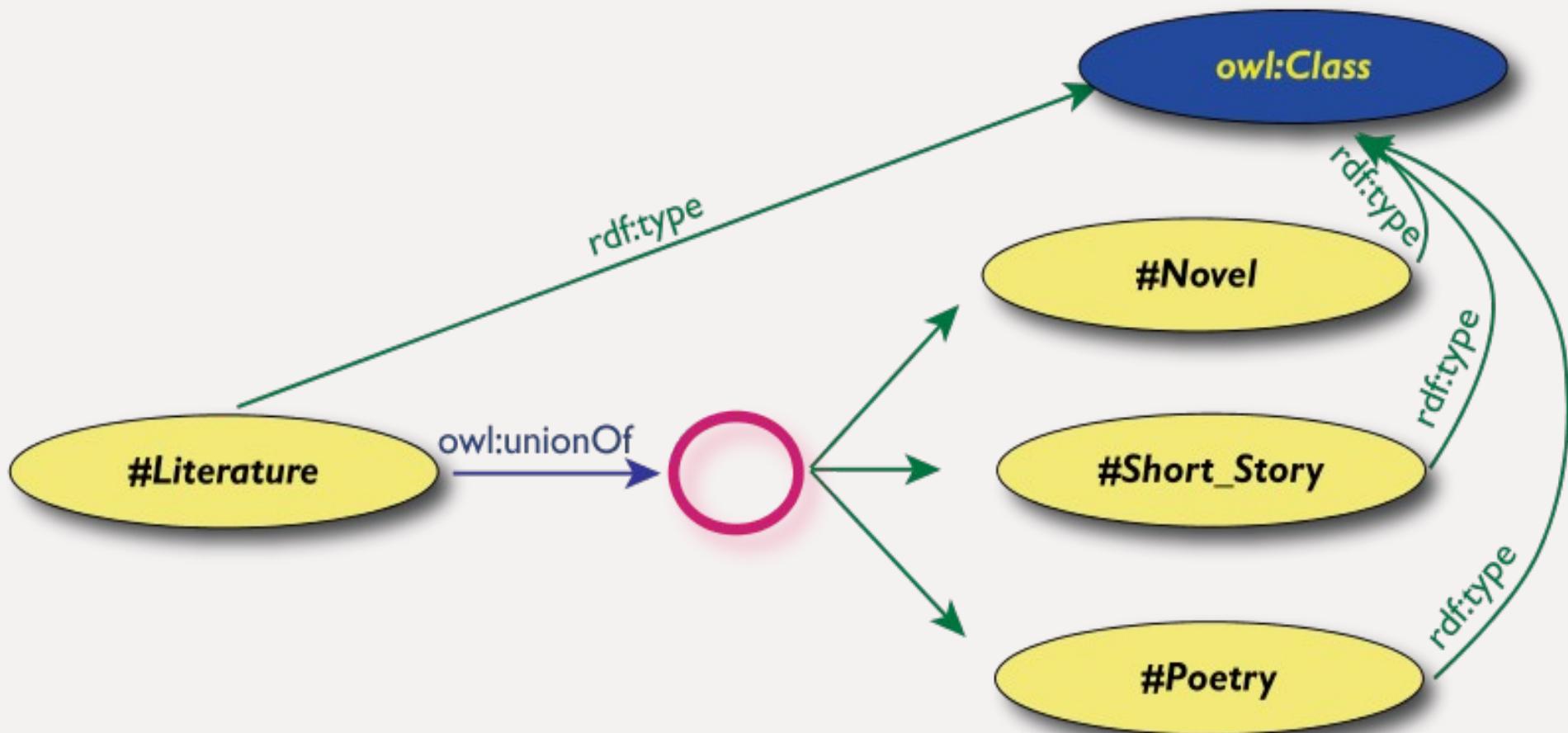
- The OWL solution, where possible content is explicitly listed:



(don't worry about the syntax mapping...)

# Union of classes

- Essentially, like a set-theoretical union:



- Other possibilities: intersection, complement

# What we have so far...

- The OWL features listed so far are already fairly powerful
- E.g., various databases can be linked via **owl:sameAs**, functional or inverse functional properties, etc.
- It is still possible to find inferred relationship using a traditional rule engine
  - (more or less... there are some restrictions on details)

# Oracle's Technology Network portal

Aggregates many source of content

This screenshot shows a semantic web interface for Oracle's Technology Network. The top navigation bar includes links for Home, Headlines, Downloads, Forum, Articles, Podcasts, Blogs, and More. A search bar is located at the top right. The main content area is organized into several columns:

- NAVIGATE BY DATABASE:** Includes links for Oracle Database (288), Oracle Database 10g (146), Oracle Options (122), Extended Database (9), Java Core at Oracle (23), and Secure Memory Database (12).
- NAVIGATE BY MIDDLEWARE:** Includes links for Oracle Collaboration Suite (5), Oracle Database Tools (65), Oracle Fusion Middleware (122), SOA Suite (11), and Oracle Business Intelligence (11).
- NAVIGATE BY APPLICATIONS:** Includes links for Oracle Database (17), Oracle Business Intelligence (10), and Oracle SOA Suite (1).
- NAVIGATE BY TECHNOLOGIES:** Includes links for Business Integration (222), Business Intelligence (73), Database (632), Enterprise Management (146), Grid Computing (110), Java (276), Linux (27), Open Source (122), Security (76), and Service Oriented Architecture (37).
- NAVIGATE BY DATA:** Includes links for 2007 (1116), 2006 (980), 2005 (227), and 2004 (41).
- ARTICLES:** A column listing various articles such as "Getting into SOA (Part 1)" and "The 'Bible' on Blogs with Oracle® Paul".
- BLOGS & NEWS:** A column listing various blog posts and news items.
- FORUMS & PODCASTS:** A column listing forums and podcasts.
- DISCUSSION FORUMS:** A column listing discussion forums.

This screenshot shows a blog post from the Oracle Blogs section. The top navigation bar includes links for Home, Headlines, Downloads, Forum, Articles, Podcasts, Blogs, and More. The main content area displays a blog post titled "Ready to say 'Hello' to E-Bus P12?" by Mark C. Miller. The post discusses the release of Oracle Application Express 12c and includes a chart showing the number of releases per month from 2007 to 2008. Below the chart, there are comments and a sidebar with links to other blog posts and categories like Data, Grid, and SOA.

Re-group, categorize, etc content (using a taxonomy)

Courtesy of Mike DiLascio, Siderean Software, and Justin Kestelyn, Oracle Corporation ([SWEO Case Study](#))

# However... that may not be enough

- Very large vocabularies might require even more complex features
  - typical example: definition of all concepts in a health care environment
- One major issue is the way classes (i.e., “concepts”) are defined
- OWL includes those extra features but... the inference engines become (much) more complex 😞

# Property value restrictions

- Classes are created by restricting the property values on a superclass
- For example: how would I characterize a “listed price”?
  - it is a price (which may be a general term), but one that is given in one of the “allowed” currencies (say, €, £, or \$)
  - more formally:
    - the value of “**p:currency**”, when applied to a resource on listed price, must take one of those values...
    - ...thereby defining the class of “listed price”

# Restrictions expressed in OWL/RDF

```
:Listed_Price rdf:type owl:Class;
  rdfs:subClassOf [
    rdf:type          owl:Restriction;
    owl:onProperty   <http://...#currency>;
    owl:allValuesFrom :Currency .
  ] .
```

- “**allValuesFrom**” could be replaced by “**someValuesFrom**” to express another type of restriction
  - e.g., I could have said: there should be a price given in at least one of those currencies
- or “**hasValue**”, when restricted to one specific value

# Similar concept: cardinality restriction

- In a property restriction, the issue was to restrict the possible values of a property
- In a cardinality restriction, the number of relations with that property is restricted
- Eg: “a book being on offer” could be characterized as having at least one price property (i.e., the price of the book has been established)

# Cardinality restriction

```
:Book_on_sale rdf:type owl:Class;
  rdfs:subClassOf [
    rdf:type          owl:Restriction;
    owl:onProperty   <http://...#price>;
    owl:minCardinality "1"^^xsd:integer.
  ] .
```

- could also be “owl:cardinality” or “owl:maxCardinality”

# Find the right experts at NASA

- Expertise locator for nearly 70,000 NASA civil servants, using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services...

The screenshot shows the POPS v.28.3 application window with the following panels:

- NASA Center (15):** Lists facilities: ARC, DFRC, GRC, GSFC (selected), HQ, IIV, JPL, JSC, KSC, LARC, MAF, MSFC. Source: x500.
- Project (176):** Lists projects: Mars Global Surveyor, Mars Odyssey 2001, Mars R&A, Mars Reconnaissance Orbiter 2005 (...), Messenger, Minor Revital, Mission Operations, Mission Science Guest Investigator, Mission Success – Center Specific, Multi-Mission Operations, NMP Program Management and Future..., NPOESS Preparatory Project (NPP). Source: WIMS.
- Competency (21):** Lists competencies: Astrobiology, Astronomy and Astrophysics, Climate Change and Variability, Earth Atmosphere, Earth Science Applications Research, Earth System Modeling, Fluid Physics, Fundamental Physics, Geophysical/Geologic Science, Geospatial Science and Technologies (selected), Icing Physics, Laser Technology. Source: CMS.
- People (1):** Lists people: Jeanne M (selected).

**Information Panel:**

**View Different Social Network's Present in the Data:**

Diagram showing connections between people based on shared skills, projects, and facilities:

- Jeanne M (Skill: Earth Sciences Competency Suite, Project: Center Investment Accounts) is connected to Michael H Grove (Facility: HQ) via a pink line labeled "Jeffrey T".
- Michael H Grove (Name: Michael Grove, Email: [redacted]@nasa.gov, Phone: 301. [redacted], Employer: Clark and Parsia) is connected to Michael H Grove (Facility: HQ) via a pink line labeled "Jeffrey T".

**Legend:**

- Red line: Same Skill and Same Department
- Green line: Same Skill and Same Project
- Blue line: Same Skill, Project, and Facility
- Pink line: Am I Connected?

**Social Net:**

Michael Grove, Clark & Parsia, LLC, and Andrew Schain, NASA, ([SWEO Case Study](#))

# But: OWL is hard!

- The combination of class constructions with various restrictions is extremely powerful
- What we have so far is following the same logic as before
  - extend the basic RDF and RDFS possibilities with new features
  - expect to infer new relationships based on those
- However... a full inference procedure is hard 😞
  - not implementable with simple rule engines, for example
  - in some cases, it may even be impossible

# OWL profiles

- The term OWL “profiles” comes to the fore:
  - restricting which terms can be used and under what circumstances (restrictions)
  - if one abides to those restrictions, then simpler inference engines can be used

# OWL profiles (cont.)

- In the *current* OWL standard, three such “profiles” are defined:
  - OWL Full: no restrictions whatsoever
  - OWL DL (and its “sub profile” OWL Lite): major restrictions to ensure implementability
- The OWL 2 work will add new profiles
  - profiles that are simple enough to be implementable with simple rule engines (like the first few examples we had)
  - profiles that are optimized to a small number of class and property definition but a large amount of data
  - etc.

# OWL Full

- No constraints on the various constructs
  - this means that:
    - Class can also be an individual, a URI can denote a property as well as a Class
      - e.g., it is possible to talk about class of classes, etc.
    - one can make statements on RDFS constructs (e.g., declare `rdf:type` to be functional...)
    - etc.
- But: *an OWL Full ontology may be undecidable!*

# OWL Description Logic (DL)

- A number of restrictions are defined
  - classes, individuals, properties strictly separated: a class *cannot* be an individual of another class
  - strict separation of the user's and the reserved (RDFS, OWL) terms
    - no statements on RDFS and OWL resources, for example
  - the values of user's object properties must be individuals
    - i.e., they are used to create relationships *between individuals*
  - no characterization of *datatype* properties
  - ...
- But: well known inference algorithms exist!

# Note on OWL profiles

- OWL profiles are defined to reflect compromises:
  - expressibility vs. implementability
- Some application just need to express and interchange terms (with possible scruffiness): OWL Full is fine
  - they may build application-specific reasoning instead of using a general one
- Some applications need rigour, but only a simple set of statements: a rule engine based profile might be o.k.
- Some applications need rigour and complex term classification: OWL DL might be the good choice

# Ontology development

- The hard work is to create the ontologies
  - requires a good knowledge of the area to be described
  - some communities have good expertise already (e.g., librarians)
  - OWL is just a tool to formalize ontologies
- Large scale ontologies are often developed in a community process
- Ontologies should be shared and reused
  - can be via the simple namespace mechanisms...
  - ...or via explicit inclusions
- Applications can also be developed with very small ontologies, though

# Ontologies examples

- International Country List
  - example for an OWL Lite ontology
- Large ontologies are being developed
  - [eClassOwl](#): eBusiness ontology for products and services, 75,000 classes and 5,500 properties
  - [National Cancer Institute's ontology](#): about 58,000 classes
  - [Open Biomedical Ontologies Foundry](#): a collection of ontologies, including the [Gene Ontology](#) to describe gene and gene product attributes in any organism or protein sequence and annotation terminology and data ([UniProt](#))
  - [BioPAX](#): for biological pathway data

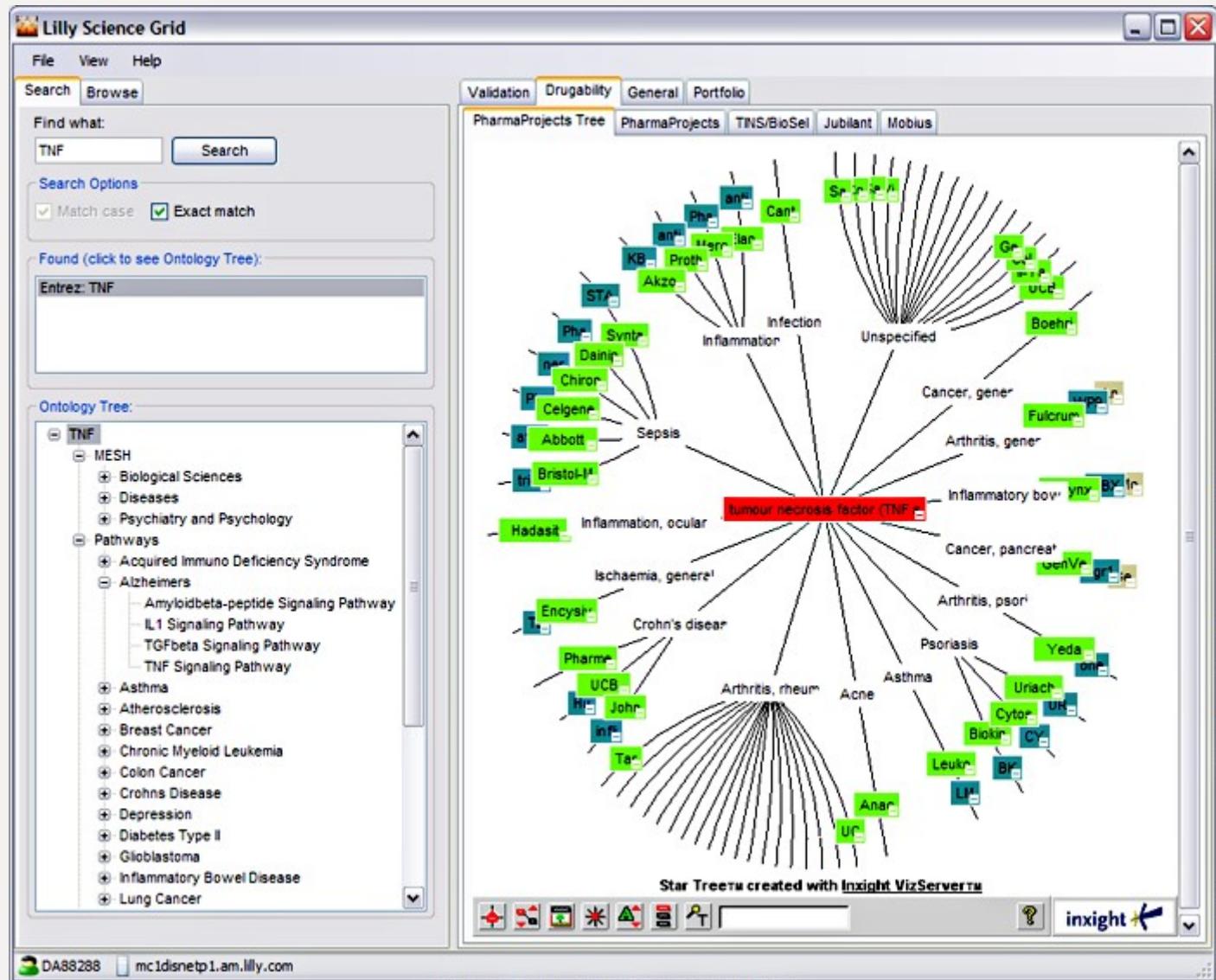
# Portal to Principality of Asturias' documents

- Search through governmental documents
- A “bridge” is created between the users and the juridical jargon using SW ontologies and tools

Courtesy of Diego Berrueta and Luis Polo, CTIC, U. of Oviedo, and the Principality of Asturias, ([SWEO Case Study](#))

# Eli Lilly's Target Assessment Tool

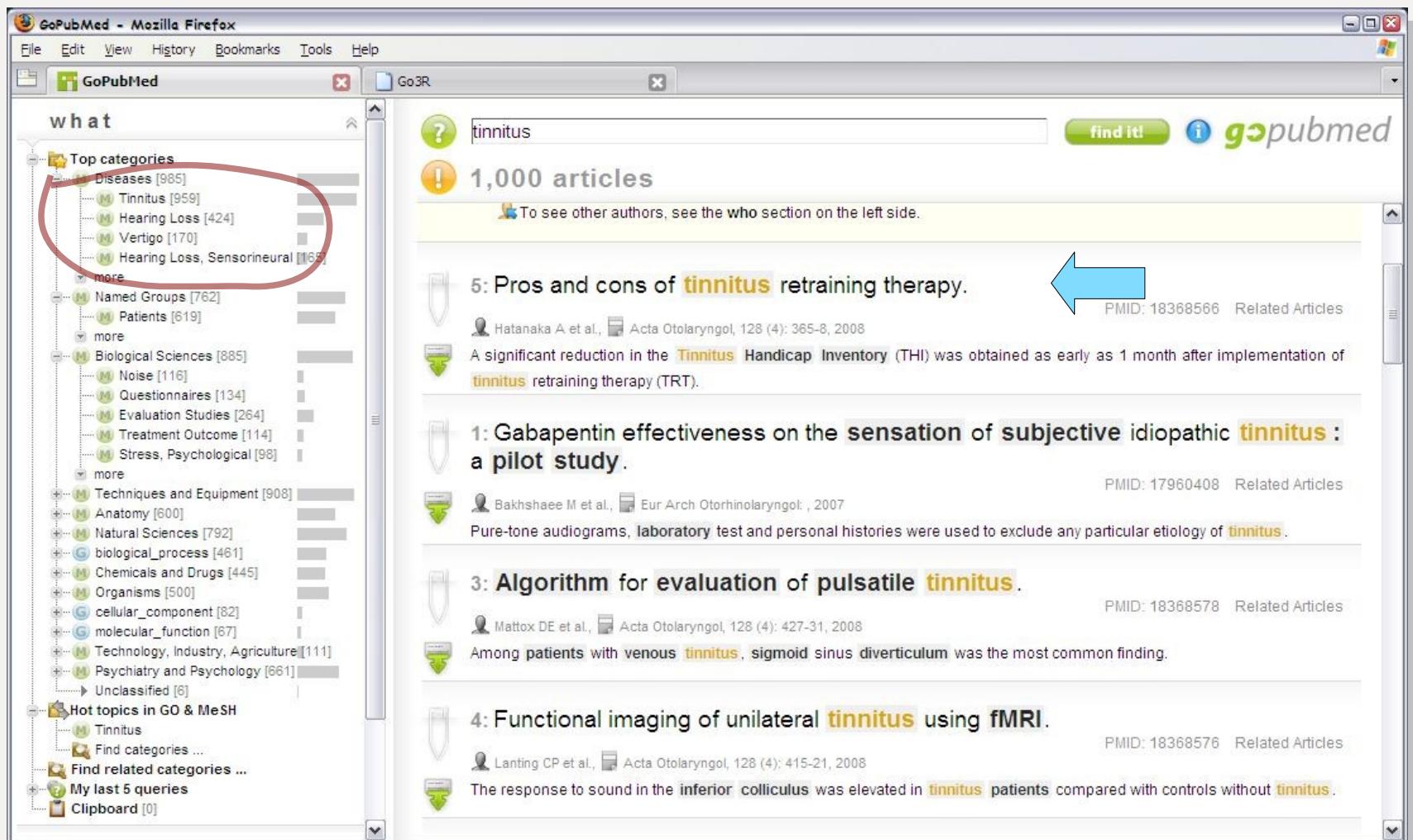
- Prioritization of drug target, integrating data from different sources and formats
- Integration, search via ontologies (proprietary and public)



Courtesy of Susie Stephens, Eli Lilly (SWEO Case Study)

# Improved Search via Ontology (GoPubMed)

- Search results are re-ranked using ontologies
- Related terms are highlighted



# Improved Search via Ontology (Go3R)

- Same dataset, different ontology
  - (ontology is on non-animal experimentation)

The screenshot shows a Mozilla Firefox window with two tabs: "GoPubMed" and "Go3R". The "Go3R" tab is active, displaying search results for the term "tinnitus".

**Search Results:**

- 1: Microvascular decompression of cochleovestibular nerve.**  
Yap L et al., Eur Arch Otorhinolaryngol, 2008  
This report provides a review of all the published studies on MVD of the eighth (8th) nerve in alleviating cochleovestibular symptoms and presents three additional patients who underwent MVD of the eighth nerve for **tinnitus** or vertigo.
- 2: Algorithm for evaluation of pulsatile tinnitus.**  
Mattox DE et al., Acta Otolaryngol, 128 (4): 427-31, 2008  
Among patients with arterial **tinnitus**, carotid atherosclerotic disease was the most common.
- 3: Functional imaging of unilateral tinnitus using fMRI.**  
Lanting CP et al., Acta Otolaryngol, 128 (4): 415-21, 2008  
This article shows that the inferior colliculus plays a key role in unilateral subjective **tinnitus**.
- 4: Pros and cons of tinnitus retraining therapy.**  
Hatanaka A et al., Acta Otolaryngol, 128 (4): 365-8, 2008  
A significant reduction in the **Tinnitus Handicap Inventory (THI)** was obtained as early as 1 month after implementation of **tinnitus** retraining therapy (TRT).
- 5: Mass casualty incident management triage, injury distribution of casualties and**

**Left Panel (3R Relevance Filters):**

- Diseases & Symptoms [601]
  - Tinnitus [547]
  - Hearing Loss [248]
  - Vertigo [98]
  - Disease [118]
  - Hearing Loss, Sensorineural [95]
- Methodology [408]
- Life Sciences [503]
- Body Systems & Structures [401]
  - Bioethics [102]
  - Reduction [90]
- Statistics [125]
- Substances, Preparations & Products [277]
- Biological Material & Organisms for Animal U
- Method Specification [36]
- Animal Species [40]
- Product Properties & Effects [62]
- Product Testing & Assessment [20]
- 3Rs Methods in the Life Sciences [6]
- Animal Experiment [6]
- 3Rs Relevant [5]
  - In Vitro Experimental Design [20]
  - In Vivo Experimental Design [5]
  - Animal Condition, Physiological or Psycholog
  - Animal Care & Handling [3]
  - Toxic Actions of Substances [7]
- Unclassified [390]
- Find related categories ...
- My last 5 queries
- Clipboard [0]

# Same problem, different solution...

How do I use Anatomy Lens? [BASIC](#) [ADVANCED](#)

Best viewed in Firefox at 1280 x 1024

## ANATOMY LENS: Semantic Search Over PubMed

**PUBMED SEMANTIC QUERY**

Search Stop Clear All

**Specify Medical Subject Heading(s) (MeSH):**

Lung

**Specify Anatomical Part(s) (FMA):**

Lung

**Specify Genetic Process(es) (GO):**

respiratory tube development

Published between 2005 and 2008

Result Limit 100

**Query Results**

23 result(s) found between 2005 and 2008 in 1.612 seconds Continue semantic search on additional related concepts?

Compare results with standard PubMed search on [Lung respiratory tube development](#).

2005 [Lysyl oxidase](#) is essential for normal development and function of the respiratory system and for the integrity of elastic and collagen fibers in various tissues. [A]

2005 [Nmyc](#) plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation. [A]

2005 [Pathophysiological consequences following inhibition of a CFTR-dependent developmental cascade in the lung](#). [A]

2005 [The transcription factor gene Nfib](#) is essential for both lung maturation and brain development. [A]

2005 [Vascular endothelial growth factor](#) co-ordinates proper development of lung epithelium and vasculature. [A]

6 articles found related to [Lung] and [alveolus development] Why? Rate:

**Date PubMed Article**

2007 [Fgf10 dosage](#) is critical for the amplification of epithelial cell development. [A]

2007 [Foxp2 and Foxp1](#) cooperatively regulate lung and esophagus. [A]

2007 [Respiratory distress and neonatal lethality](#) in mice lacking Gαo. [A]

2006 [Alterations in gene expression in T1 alpha null lung](#): a model. [A]

2005 [Inactivation of tensin3](#) in mice results in growth retardation and death. [A]

2005 [Vascular endothelial growth factor gene therapy increases survival in hyperoxia-induced lung injury](#): evidence that angiogenesis plays a role. [A]

**Annotations on Article**

**MeSH Terms:** [abnormalities, Epithelial Cells, Male, embryology, Mice, Transgenic, Animals, Newborn, Gene Dosage, metabolism, Heterozygote, Myocytes, Smooth Muscle, Gene Expression Regulation, Developmental, Embryonic Stem Cells, Vascular Endothelial Growth Factor A, Mice, Knockout, Fibroblast Growth Factor 10, growth & development, Female, Lac Operon, Wnt Proteins, Mesoderm, cytology, Phenotype, Pregnancy, Animals, genetics, Lung, Platelet-Derived Growth Factor, Mice]

**FMA Concepts:** [Embryonic stem cell, Mesoderm, Lung]

**Result Explanation**

These articles talk about **alveolus development** which is related to the queried term **respiratory tube development**

This is because...

[alveolus development](#) is a [part\\_of](#) [lung development](#)  
 | [lung development](#) is a [part\\_of](#) [respiratory tube development](#)  
 | [part\\_of](#) is Transitive

Courtesy of Kavitha Srinivas, IBM J Watson Research Center

Copyright © 2008, W3C

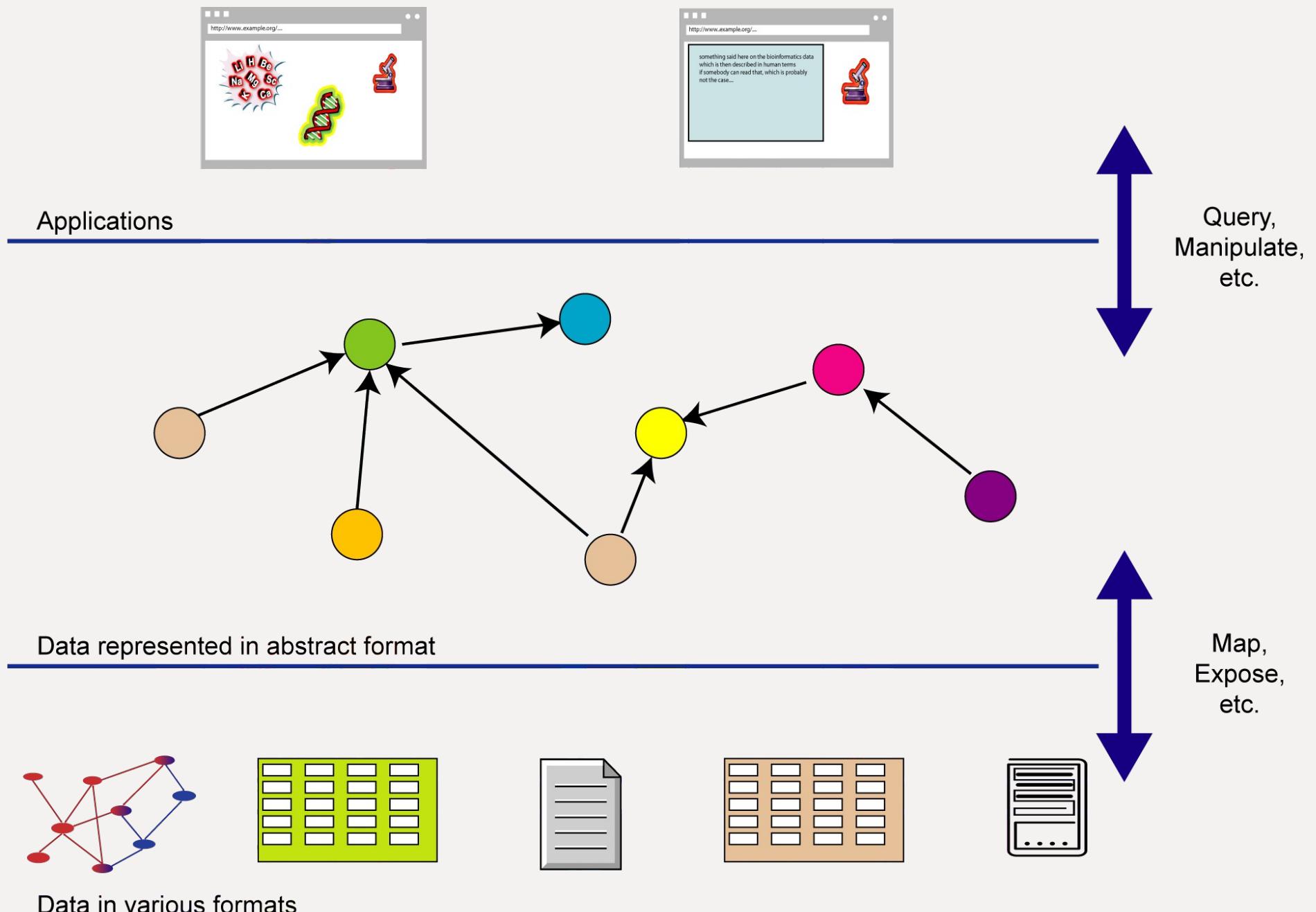
(129)

# What have we achieved?

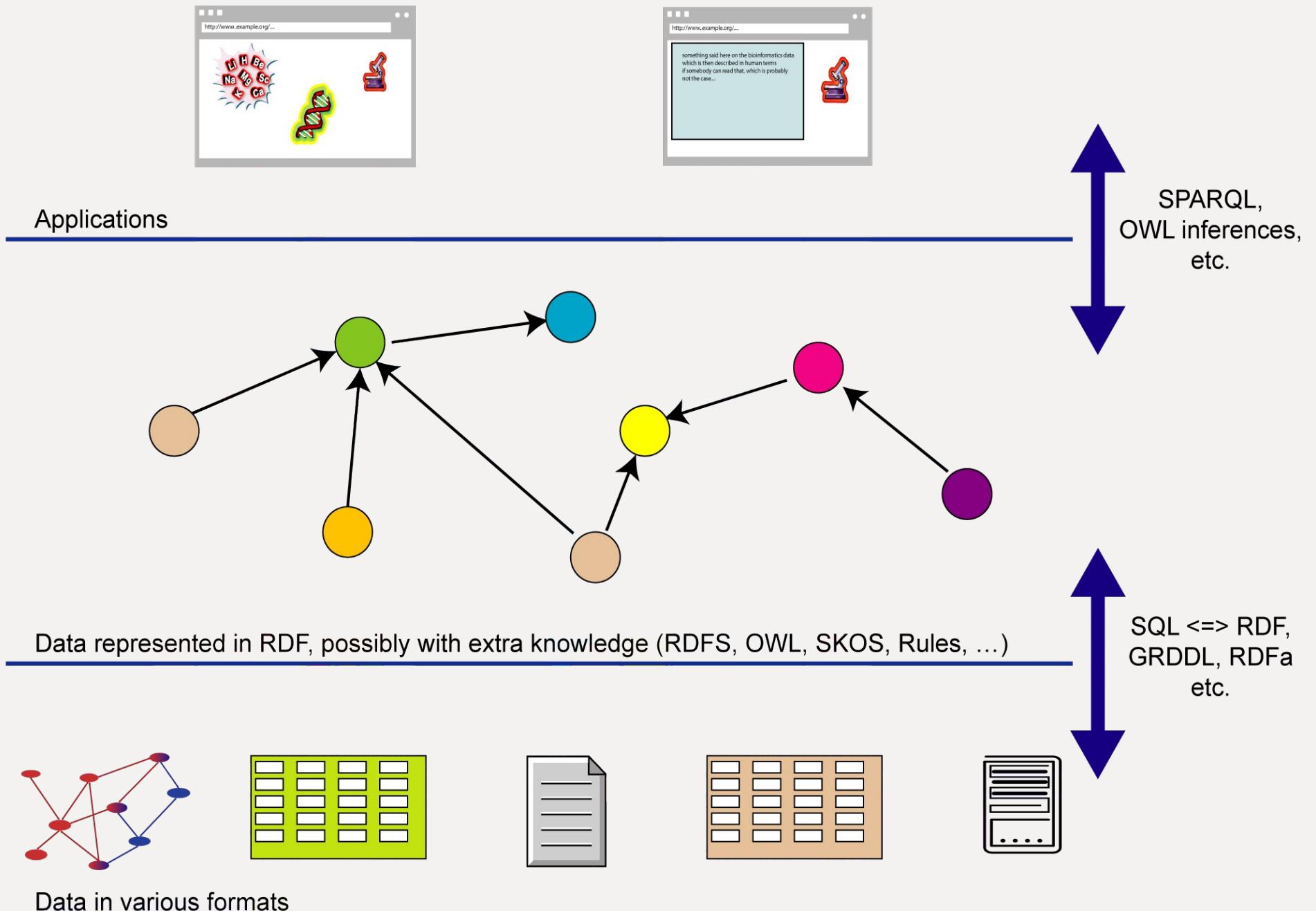
# We have not covered everything...

- Some other aspects of SW are being developed
  - some at W3C, others are still research
- For example:
  - RIF: using general rule engines with SW data; also *interchange* rule descriptions (just like data are interchanged)
  - SKOS: general framework to express term structures like vocabularies, taxonomies, glossaries
    - eg, to interface bibliographic records

# Remember the integration example?



# Same with what we learned



# eTourism: provide personalized itinerary

The screenshot shows a personalized itinerary for June 17th, 2008. The itinerary is divided into Morning and Afternoon sessions.

- Morning:**
  - Tourist Sites:
    - (10:10) Basílica de la Virgen del Pilar
    - (10:55) Ibercaja Camón Aznar Museum
    - (11:40) Cathedral of San Salvador o La Seo
    - (12:25) The Caesaraugusta Forum Museum
    - (12:55) The Caesaraugusta River Port Museum
    - (13:25) The Caesaraugusta public baths museum
    - (13:50) Iglesia Parroquial de San Gil Abad
    - (14:15) Molins house

No itinerary at these times
- Afternoon:**
  - Tourist Sites:
    - (16:15) Church of la Mantería
    - (16:35) Church of San Ildefonso o de Santiago el Mayor
    - (17:00) Church of Santo Tomás de Aquino (Escuelas Pías)
    - (17:20) Church of San Pablo
    - (17:50) Casa Armas
    - (18:10) Central market
    - (18:30) Church of Santa Isabel de Portugal o San Cayetano
    - (18:45) Samaritana Fountain
    - (19:05) Church of San Felipe y Santiago el Menor
    - (19:25) Church of San Juan de los Panetes
    - (19:50) Church of Santa Cruz
    - (20:20) Church of Santa María Magdalena

No itinerary at these times

**Zaragoza street plan:**

**Tourist site:** BASILICA OF THE PILAR

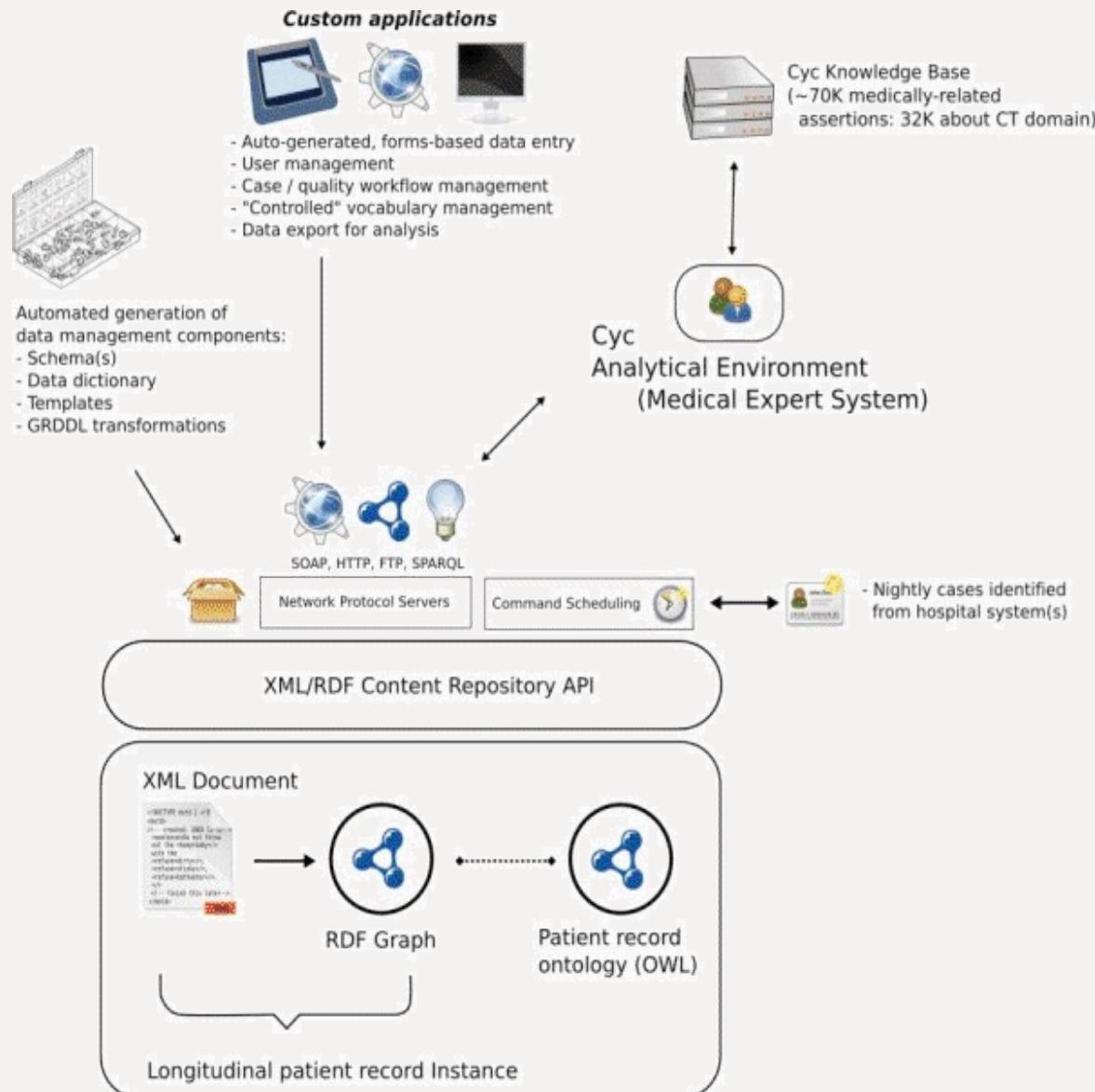
Full accessibility  
The construction of the current Basilica of the Pillar is closely linked to the increase in devotion to the Pillar throughout the 17th century. The previous Gothic-Mudejar building was not big...  
[Know more...](#)

- Integration of relevant data in Zaragoza (using RDF and ontologies)
- Use rules on the RDF data to provide a proper itinerary

Courtesy of Jesús Fernández, Municipality of Zaragoza, and Antonio Campos, CTIC (SWEO Use Case)

# Semantic DB at the Cleveland Clinic

- Problem: extreme compartmentalization of medical knowledge
- Unified repository collects and stores various data
- Usage of OWL and rules allow high level operations on the data



Courtesy of Chimezie Ogbuji, ClevelandClinic, (SWEO Case Study)

# National Archives of Korea

- Ontology based metadata infrastructure for NAK (over 12 million metadata statements)
- Usage of rules to retrieve information



Courtesy of Tony Lee, Jin Woo Kim, and Bok Ju Lee, Saltlux, Kyu Hyup Kim, Yoon Jung Kang, NAK ([SWEO Case Study](#))

# Available documents, tools

# Available specifications: Primers, Guides

- The “[RDF Primer](#)” and the “[OWL Guide](#)” give a formal introduction to RDF(S) and OWL
- [GRDDL Primer](#) and [RDFa Primer](#) have been published
- The [W3C Semantic Web Activity Homepage](#) has links to all the specifications

# “Core” vocabularies

- There are also a number “core vocabularies” (not necessarily OWL based)
  - [Dublin Core](#): about information resources, digital libraries, with extensions for rights, permissions, digital right management
  - [FOAF](#): about people and their organizations
  - [DOAP](#): on the descriptions of software projects
  - [SIOC](#): Semantically-Interlinked Online Communities
  - [vCard in RDF](#)
  - ...
- One should never forget: ontologies/vocabularies must be shared and reused!

# Some books

- P. Mika: Social Networks and the Semantic Web, 2007
- L. Yu: Semantic Web and Semantic Web Services, 2007
- G. Antoniu and F. van Harmelen: Semantic Web Primer, 2<sup>nd</sup> edition in 2008
- D. Allemang and J. Hendler: Semantic Web for the Working Ontologist, 2008
- ...

See the separate [Wiki page](#) collecting book references

# Further information

- **Dave Beckett's Resources** at Bristol University
  - huge list of documents, publications, tools, ...
- **Planet RDF** aggregates a number of SW blogs
- **Semantic Web Interest Group**
  - a forum developers with archived (and public) mailing list, and a constant IRC presence on freenode.net#swig
  - anybody can sign up on the list

# Lots of Tools (not an exhaustive list!)

- Categories:

- Triple Stores
- Inference engines
- Converters
- Search engines
- Middleware
- CMS
- Semantic Web browsers
- Development environments
- Semantic Wikis
- ...

- Some names:

- Jena, AllegroGraph, Mulgara, Sesame, flickurl, ...
- TopBraid Suite, Virtuoso environment, Falcon, Drupal 7, Redland, Pellet, ...
- Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, Tallis Platform, ...
- RDF Gateway, RDFLib, Open Anzo, DartGrid, Zitgist, Ontotext, Protégé, ...
- Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore...
- ...

# Tools

- Worth noting: major companies offer (or will offer) Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, webMethods, Northrop Grumman, Altova, Dow Jones, BBN, ...
- See also the [W3C Wiki page on tools](#)

# Application patterns

- It is fairly difficult to “categorize” applications (there are always overlaps)
- With this caveat, some of the application patterns:
  - data integration (ie, integrating data from major databases)
  - intelligent (specialized) portals (with improved local search based on vocabularies and ontologies)
  - content and knowledge organization
  - knowledge representation, decision support
  - X2X integration (often combined with Web Services)
  - data registries, repositories
  - collaboration tools (eg, social network applications)

# Help for deep sea drilling operations

- Integration of experience and data in the planning and operation of deep sea drilling processes
- Discover relevant experiences that could affect current or planned drilling operations
  - uses an ontology backed search engine



The screenshot shows a search interface for 'AKSIO-search'. At the top, there's a search bar with the query 'leak in barrier elements'. Below it is a sidebar with 'Search filters' and a main area titled 'Results 1 - 7 of 7'.

**Search filters:**

- discipline
- operation
- equipment
- state
  - select all unselect all
  - Corrosion (1)
  - Erosion (2)
  - Lack Of Maintenance (2)
  - Leak in barrier elements (5)
  - Scale Deposition (4)
  - Too High Mud Density (1)
  - Well Integrity Problem (7)
- keywords\_ref
- wellbore\_id\_ref
- field\_id
  - select all unselect all
  - EXPLORATION (1)
  - GULLFAKS (1)
  - GULLFAKS SØR (1)
  - HEIDRUN (1)
  - HULDRA (1)
  - MIDGARD (2)
  - RIMFAKS (1)
  - SNORRE (1)
  - VISUND (2)

**Results 1 - 7 of 7**

- Top plug / 20" E2SV**  
2002-06-26T10:00:00Z  
**Description:** Experience: In a "standard" OPR design, the upper cement plug would cover the 13 3/8" cut as well as...  
EXPLORATION NO 6406/1-1 PA PLUGBACK/KICK-OFF  
2007-06-12T07:05:58Z  
Cementing Network Sementeringsnettverk Directional Drilling Network Directional Drilling Network Bronnintegritet Well Integrity Casing Foringar Deep set tubing plug Deep set tubing plug Liner top packer Liner top packer Mechanical tubular plugs Well Integrity Problem Well Integrity Problem
- RIH with drill stem teststring.**  
2002-06-13T10:00:00Z  
**Description:** RIH with drill stem teststring. Took weight when entering 7" liner with test string. Worked same pas...  
RIMFAKS NO 34/10-3-4 H DST DRILL STEM TEST  
2007-06-12T07:05:58Z  
Bronnintegritet Well Integrity Snubbing Snubbing Completion string component Completion string component Downhole tester valve Downhole tester valve Borestreng Drillingstring Subsea production tree Subsea test tree Subsea test tree Surface test tree Surface test tree Well test packer Well test string Wall test string Wall test string components Erosion Erosjon Lack Of Maintenance Lack Of Maintenance Leak in barrier elements Leak in barrier elements Well Integrity Problem Well Integrity Problem
- Flowing well**  
2003-02-20T11:00:00Z  
**Description:** The well was temporary handed back to production during changeover from slick line to 5/16" cable to...  
HEIDRUN NO 6507/7-A-20 WIREL Holi Trond OTHER  
2007-06-12T07:05:58Z  
Cementing Network Sementeringsnettverk Technical Sidetrack Tekniske Sidesteg Bronnintegritet Well Integrity Snubbing safety head Snubbing safety head UBD none return valve UBD none return valve Corrosion Corrasjon Erosion Erosjon Lack Of Maintenance Lack Of Maintenance Leak in barrier elements Leak in barrier elements Scale Deposition Scale Deposition Too High Mud Density Too high mud density Well Integrity Problem
- Fill drop sub assy prior to making up packer for barrier assy to avoid possible trapped pressure.**  
2002-06-24T10:00:00Z  
**Description:** Fill drop sub assy prior to making up packer for barrier assy to avoid possible trapped pressure...  
HULDRA NO 30/2/A-6 8 1/2" Rodvet Knut T/A PLUGS & MECH. PLUGS  
2007-06-12T07:05:58Z  
Cementing Network Sementeringsnettverk Bronnintegritet Well Integrity Deep set tubing plug Deep set tubing plug Leak in barrier elements Leak in barrier elements Scale Deposition Scale Deposition Well Integrity Problem

Courtesy of David Norheim and Roar Fjellheim, Computas AS ([SWEO Use Case](#))

# Radar Network's Twine

- “Social bookmarking on steroids”

- Relationships are based on ontologies

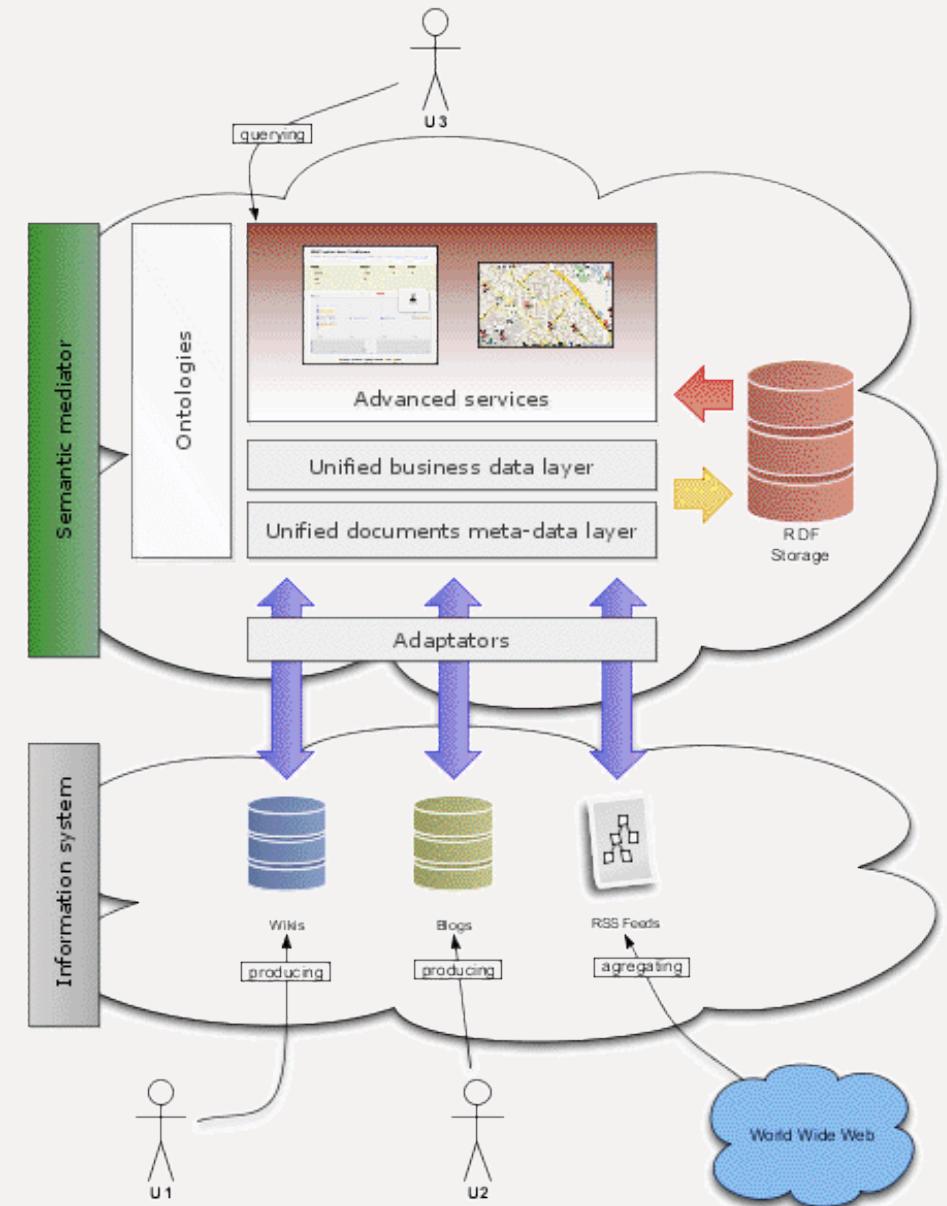
- evolving over time, possibly enriched by users

- Internals in RDF

The screenshot shows a Mozilla Firefox browser window displaying a Twine interface. The title bar reads "The Manhattan Project: The Birth of the Atomic Bomb in the Words of Its Creators, Eyewitnesses and Historians. © twine - Mozilla Firefox". The main content area is titled "Ivan's private twine" and shows a summary of the item "The Manhattan Project: The Birth of the Atomic Bomb in the Words of Its Creators, Eyewitnesses and Historians." Below the summary, there is a thumbnail image of the book cover, the "Original URL" (http://www.amazon.com/Manhattan-Project-Creators-Eyewitnesses-Historians/d...), the "Publisher" (Black Dog & Leventhal Publishers), the "Release Date" (Sep 17, 2007 (3 months ago)), and the "Price" (\$24.95). To the right of the summary, there is a sidebar with semantic annotations categorized under "Places", "People", "Organizations", and "Other tags". The "Places" section includes "Asia", "Europe", and "United States". The "People" section includes "Albert Einstein", "Groves", "Harry S. Truman", "J. Robert Oppenheimer", "Klaus Fuchs", "Leo Szilard", and "Richard Feynman". The "Organizations" section includes "Atomic Heritage Foundation" and "Manhattan Project". The "Other tags" section includes "Atomic bomb". Three specific annotations are circled in green: "Publisher", "Release Date", and "Price".

# Integration of “social” software data

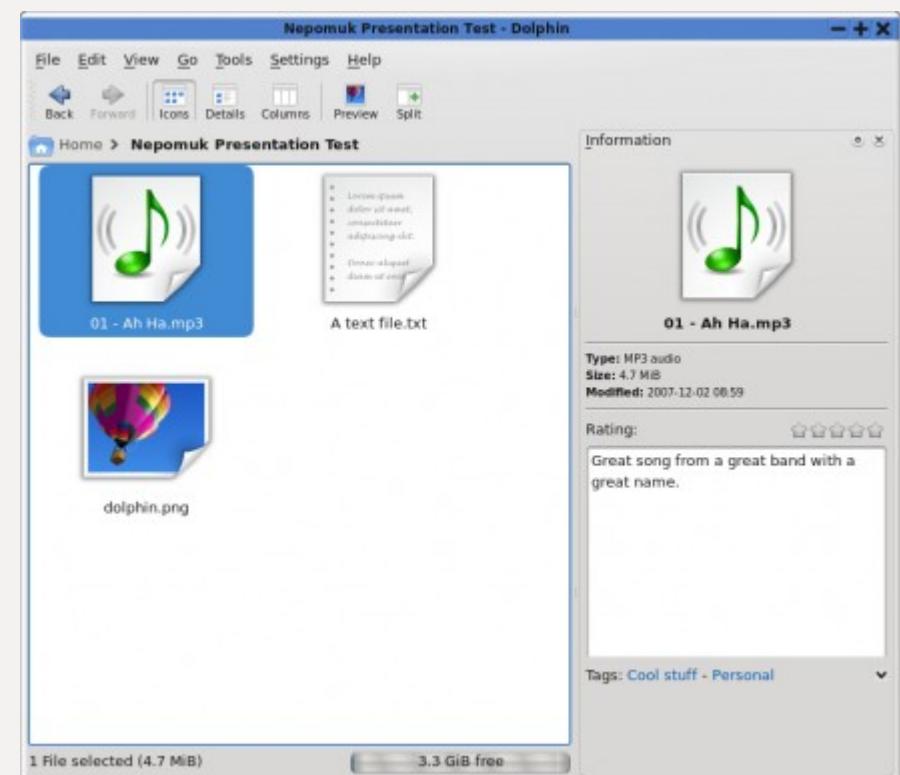
- Internal usage of wikis, blogs, RSS, etc, at EDF
  - uses:
    - public ontologies (SIOC, FOAF, DC, Geonames)
    - public datasets for tagging
    - SPARQL as integration tool for queries
- Details are hidden from end users (via plugins, extra layers, etc)



Courtesy of Alexandre Passant, EDF R&D and LaLIC, Université Paris-Sorbonne, ([SWEO Case Study](#))

# Semantic tagging and search in KDE

- Metadata backend fully based on RDF
- Each file can be tagged, rated, commented, and automatically indexed
- Queries are made via the combination of all those



Courtesy of Leo Sauermann, DFKI, and Sebastian Trüg, Mandriva Linux (SWEO Case Study)

# Conclusions

- The Semantic Web is there to integrate data on the Web
- The goal is the creation of a *Web of Data*

# Thank you for your attention!

- These slides are publicly available on:

<http://www.w3.org/2008/Talks/0924-Vienna-IH/>