



Uncertainty and the Semantic Web

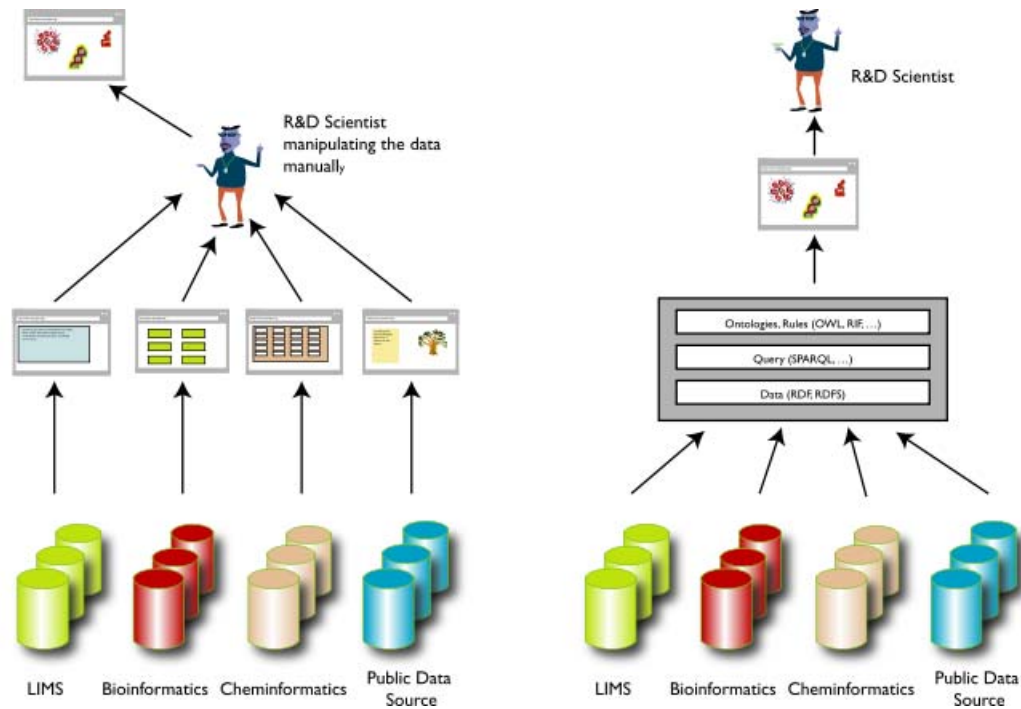
URSW Workshop, Athens, GA, USA, 2006-11-05

Ivan Herman, W3C

Caveat

- You guys know more about the mathematics and modelling issues than I do
- Ie, take everything I say with a pinch of salt 😊

A typical SW usage example: data integration



General approach

1. Map the various data onto RDF
 - *assign URI-s to your data*
 - *“mapping” may mean on-the-fly SPARQL to SQL conversion, “scraping”, etc*
2. Merge the resulting RDF graphs (by identifying URI-s)
3. Start making queries on the whole

General approach (cont.)

- Ontologies and/or rules play a pivotal role
 - *by providing some extra knowledge, further nodes in the graph can be merged, combined, related to one another*
 - *the system can deduce new relationships using some entailment regime*
- These ontologies are not necessarily complex!
 - *even a few lines of RDFS can make wonders*

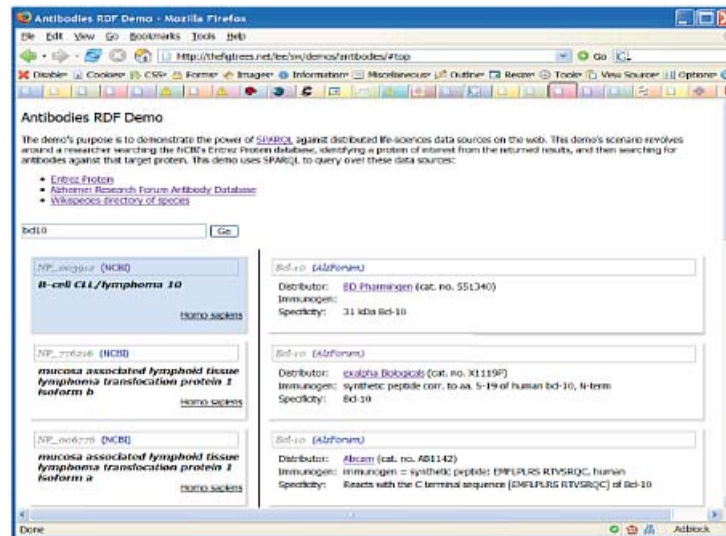
Large application areas with this pattern

- Health care and life science
- eGovernmental initiatives
- Financial services
- Oil exploration
- Legal profession
- ...

(these are just some of those we have met at W3C...)

Example: antibodies demo

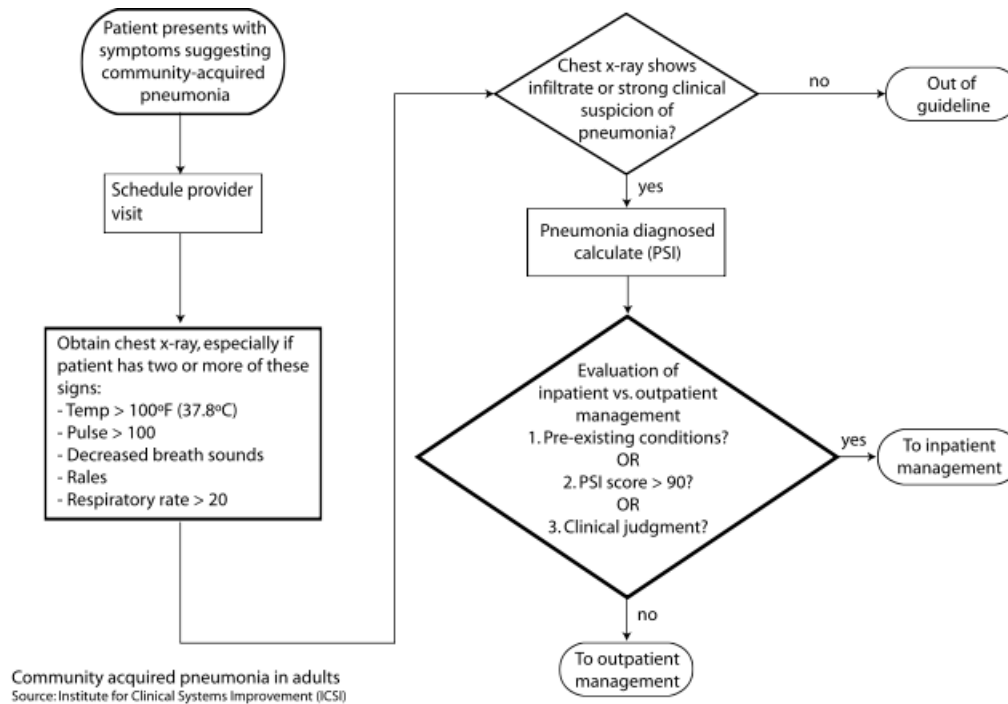
- Scenario: find the known antibodies for a protein in a specific species
- Combine (“scrape”...) three different data sources
- Use SPARQL as an integration tool (see also [demo online](#))



However...

- Things do not always fit into this model nicely
 - *the initial database may include “weights” for certain relations*
 - *ontologies may not cleanly separate or bind terms, there are “shades”*
 - *etc*

A specific case



A specific case (cont.)

- A Semantic Web application would:
 - *encode such flowchart (“clinical pathway”) in ontologies and/or rules*
 - *combine it with data coming from other sources (drug information, hospital administration, etc...)*
 - *provide a comprehensive tool to help doctors*
- (Go to the HCLS Workshop tomorrow if you are interested...)

What are the proper conclusions?

Obtain chest X-ray, especially if patient has two or more of these signs:

- 1. Temp > 37.8°C*
- 2. Pulse > 100*
- 3. Decreased breath sounds*
- 4. Respiratory rate > 20*

■ But, surely:

- *Temp > 37.5°C is in the “danger zone” already*
- *symptoms carry different weights in decision making*
- *etc*

Similar scenarii arise in...

- Biochemical, biological research
- Oil exploration
- Exchanging/modelling spam control rules
- Legal profession
- Media content rating rules
- Multilingual setting
- Even the ground data may bear some weights/uncertainties:
 - *E.g., geotagging via Google, if you are in a country not “covered” by precise cartography data*

An aside...

- Studies have shown that traditions in East Asian cultures (China, Korea, Japan,...) are very different
 - *not based on the rational Greek heritage*
 - *the “shades” and the overall picture is taken more seriously (think of traditional Chinese medicine)*
 - *yes-or-no logic is less natural than for others*
- Ie, modelling, eg, Chinese medicinal approaches on the Semantic Web might be more difficult...



Things can be hacked, of course...

- One can imagine complex modelling, e.g., (in the pneumonia example)
 1. *separate classes with various temperature intervals and with an extra property for a weight*
 2. *class axioms using the various combinations of these*
 3. *etc*
- Combine a traditional reasoning on some part of the knowledge base, and a statistical reasoning on other parts
- SPARQL queries may contain FILTER-s evaluating numerical values
- But all these lead easily to the equivalent of a spaghetti code... 🤔

Instead...

- A clean (reasoning) model may make great sense
 - *and that is where workshops like this are important*
- However: it should be part of the Semantic Web landscape!

By the way...

- Are we talking about probabilistic reasoning, or...
- ... reasoning based on fuzzy logic, or...
- ...both?
- It seems that these two approaches (or the communities?) are fairly independent of one another...

Where are we now (reminder)?

- Solid specification suite since 2004: well defined (formal) semantics for RDF, RDFS, OWL, clear RDF/XML syntax
- *Lots* of tools are available. Are listed [on W3C's wiki](#):
 - *RDF programming environment for 14+ languages, including C, C++, Python, Java, Javascript, Ruby, PHP,...*
(no Cobol or Ada yet sad smiley!)
 - *13+ Triple Stores, ie, database systems to store (sometimes huge!) datasets*
 - *etc*
- Some of the tools are Open Source, some are not; some are very mature, some are not 😊:
it is the usual picture of software tools, nothing special any more!
- *Anybody can start developing RDF-based applications today*

Where are we now? (Cont.)

- Separate layers have been defined for OWL, balancing expressibility vs. implementability (OWL-Lite, OWL-DL, OWL-Full)
- SPARQL is coming to the fore, with lots of implementations already
 - *it will play an important role in integrating/federating RDF data*

Lots of developments are going on

- “Scraping” RDF data from various sources: images (XMP), XHTML/XML sources (e.g., GRDDL, RDFa)
- Building SQL \Leftrightarrow RDF “bridges” to export data to RDF
- Developing rule interchange formats (RIF)
- etc

Revisions of RDF and OWL?

- Such specifications have their own life
- Missing features come up, errors show up
- There will probably be a next version at some point

Revision of the RDF model?

- Some restrictions in RDF may be unnecessary (bNodes as predicates, literals as subject, ...)
- Issue of “named graph”: possibility to give a URI to a set of triplets and make statements on those
- Alternative serializations (XML or otherwise)?
- Add a time tag to statements?
- Internationalization issues with literals (how do I set “bidi” writing?)

These are just ideas floating around...

“Light” ontologies

- For some applications RDFS is not enough, but even OWL Lite is too much
- There may be a need for a “light” version of OWL, just a few extra possibilities v.a.v. RDFS
- There are a number of proposals, papers, prototypes around: RDFS++, OWL Feather, pD*, ...
 - *pD**, for example, has property characterization (symmetric, transitive, inverse), class and property equivalence, and property restrictions with some or all values
- This might consolidate in the coming years

Consequences for uncertainty reasoning work

- The Semantic Web is more than just RDF and OWL
- *Any development should be part of this overall picture!*
- this means:
 - *think of querying, not only reasoning (à la SPARQL)*
 - *relying only on OWL-DL or higher may not be a fully satisfactory*
 - *think whether the core RDF semantics should be changed (and how and when...)*
 - *think of how to generate core data with embedded uncertainty*
 - *looking at SQL \rightleftharpoons RDF issues may become necessary*
 - *consider the RIF (Phase II) work, it may be a good forum to look at some of the issues*



Thank you for your attention!

These slides are publicly available on:

<http://www.w3.org/2006/Talks/1105-Athens-IH/>

in XHTML and PDF formats; the XHTML version has active links that you can follow