

W3C Semantic Web technologies for the OKS



TAMPERE UNIVERSITY OF TECHNOLOGY

Hypermedia Laboratory



WORLD WIDE WEB
consortium

Suomen toimisto

Dr. Ossi Nykänen,
Special Researcher at the
Tampere University of Technology, Hypermedia Laboratory,
Manager of the W3C Finnish Office

**Tutorial at the
OPAALS 2008 Conference
7-8 October, Tampere, Finland**



Introduction

- Background
 - **World Wide Web Consortium (W3C)** is an international body of some 400+ member organisations that develops Web standards, guides, and practices (see <http://www.w3.org/>)
 - “*The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework.*” (see <http://www.w3.org/2001/sw/>)
- Today’s presentation, outline
 - Machine-understandable semantics?
 - News aggregation example using Semantic Web
 - W3C Semantic Web activity
 - Technologies for the OPAALS OKS?
 - Conclusion
- The main objective of the tutorial is twofold: getting familiar with the basic Semantic Web techniques and getting an overview of the W3C Semantic Web Activity



Prelude

Machine-understandable semantics?

- Recall a dream from the past:
 - *If controversies were to arise, "there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, and say to each other: **Let us calculate.**"*

— G.W.Leibnitz, 1666.

⟨⟨Rewinding few hundred years...⟩⟩

- Today, “**Machine-understandable**” **semantics** is about modelling useful aspects of the world, using a data model that can be **processed with software**, e.g., in terms of “**semantic computing**”
- Effectively, this means **externalising** useful descriptions and metadata
- Or even better, transforming **existing** externalised descriptions into a suitable format...



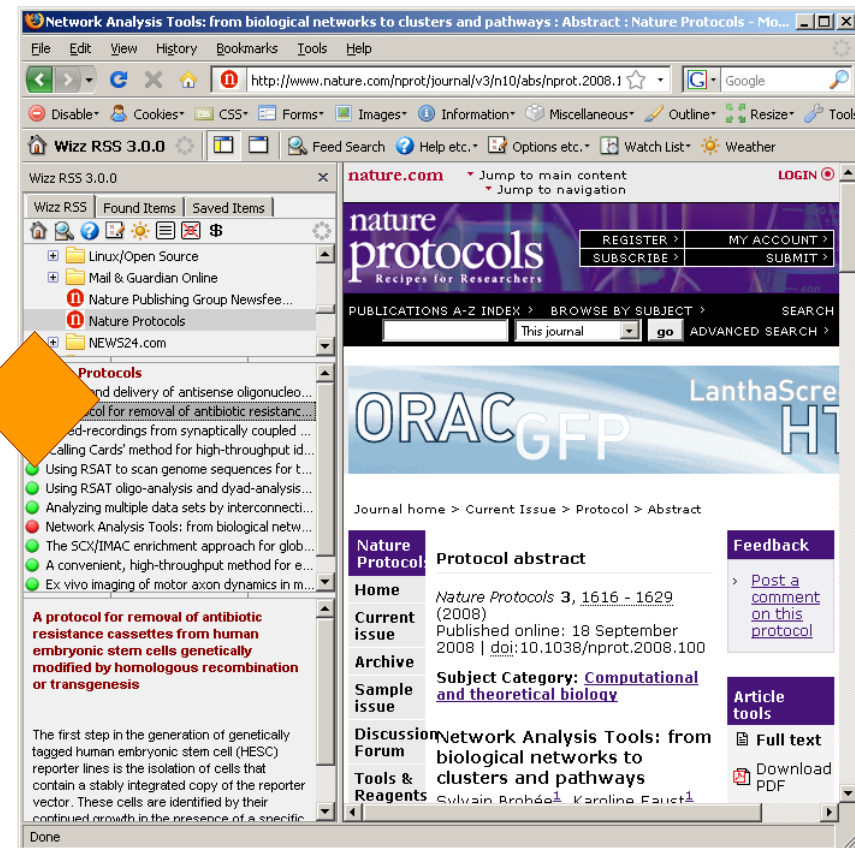
Gottfried Wilhelm Leibnitz.

Part 1/3:
**A simplified, application-centric tour
through the key technologies**

A simple application: Web syndication

- Consider a simple application: **Web syndication** of news (etc.) feeds
 - Several organisations **publish** news items using a common data format
 - “Subscribers” may easily **read and integrate** different news sources
- May be implemented in several formats, e.g., using **RDF Site Summary (RSS 1.0)**

Psst. This looks much easier than it actually is! Several (in principle adaptable) formats exist, and **syntactic and semantic inconsistencies** easily lead into problems



Code example of a news item description

- An example news item (in N3) looks something like this (real data from the *nature protocols*):

```
...
<http://dx.doi.org/10.1038/nprot.2008.100>
  rdf:type <http://purl.org/rss/1.0/item> ;
  <http://purl.org/rss/1.0/title> "Network Analysis Tools: from biological networks
to clusters and pathways" ;
  <http://purl.org/rss/1.0/link> "http://dx.doi.org/10.1038/nprot.2008.100" ;
  <http://purl.org/rss/1.0/description> "Network Analysis Tools (NeAT) is a suite
of computer tools that integrate various algorithms for the analysis of
biological networks: comparison between graphs, between clusters, or between
graphs and clusters; network randomization; analysis of degree distribution;
network-based clustering and path finding. The tools are interconnected " ;
  dc:title "Network Analysis Tools: from biological networks to clusters and
pathways" ;
  dc:creator "Sylvain Broh\u00E9e" ,
    "Karoline Faust" ,
    "Gipsi Lima-Mendez" ,
    "Gilles Vanderstocken" ,
    "Jacques van Helden" ;
  dc:identifier "doi:10.1038/nprot.2008.100" ;
  dc:source "Nature Protocols 3, 1616 (2008)" ;
  dc:date "2008-09-18" ;
  prism:publicationName "Nature Protocols" ;
  prism:publicationDate "2008-09-18" ;
  prism:volume "3" ;
  prism:number "10" ;
  prism:section "Protocol" ;
  prism:startingPage "1616" ;
  prism:endingPage "1629" .
```

[Network Analysis Tools: from biological networks to clusters and pathways](#)

18. syyskuuta 2008, 3:00:00 | Sylvain Brohée →

Network Analysis Tools: from biological networks to clusters and pathways

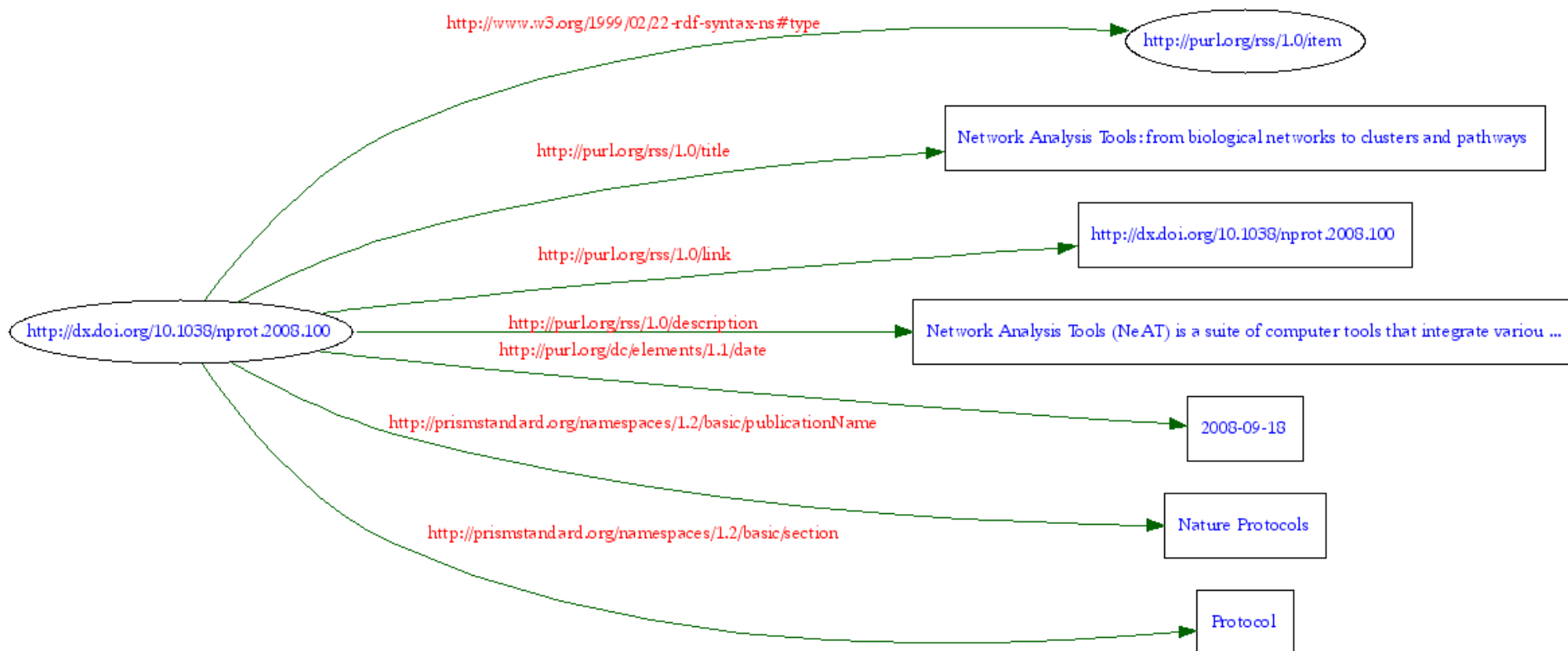
Nature Protocols 3, 1616 (2008). [doi:10.1038/nprot.2008.100](https://doi.org/10.1038/nprot.2008.100)

Authors: Sylvain Brohée, Karoline Faust, Gipsi Lima-Mendez, Gilles Vanderstocken & Jacques van Helden

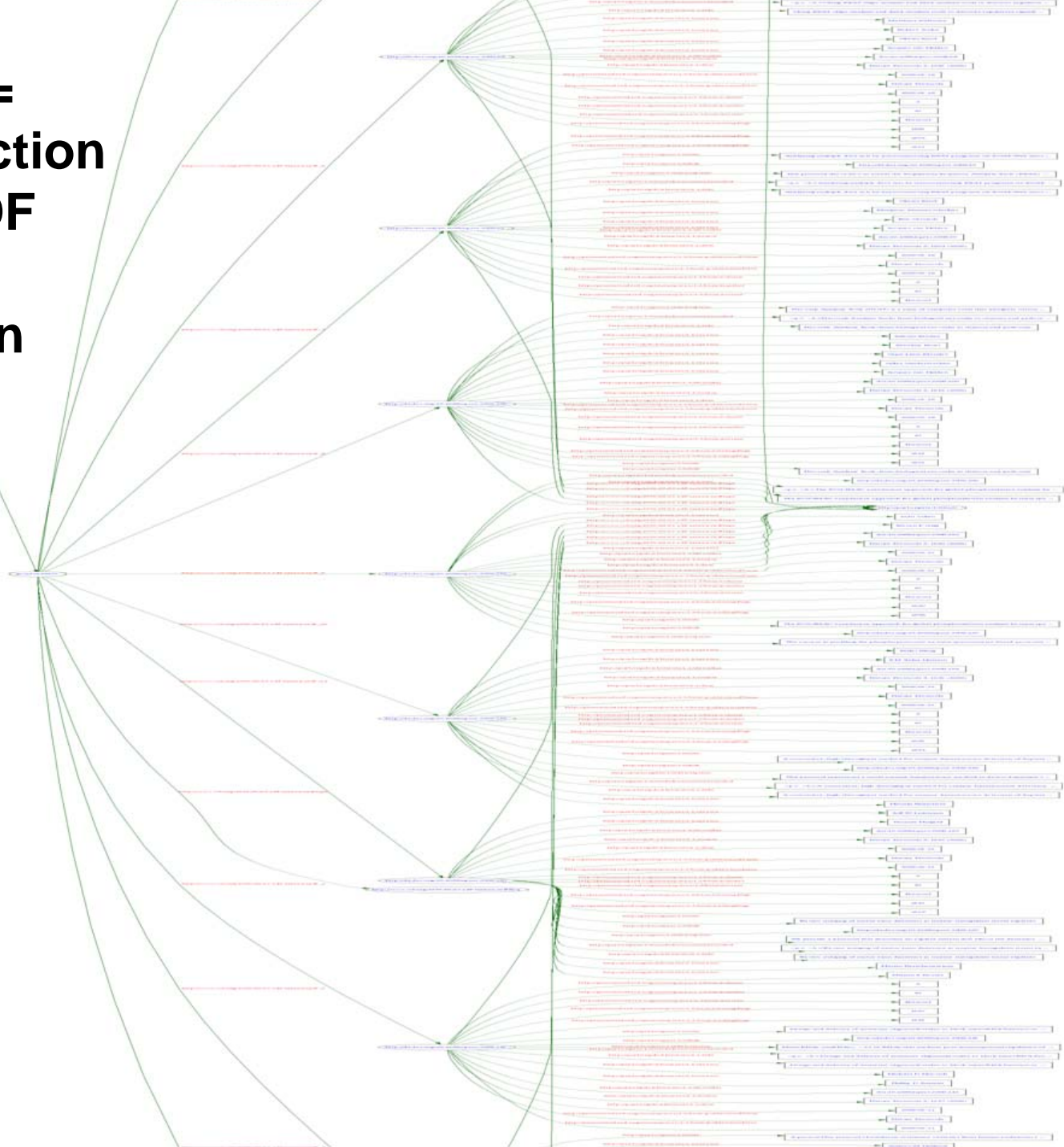
Network Analysis Tools (NeAT) is a suite of computer tools that integrate various algorithms for the analysis of biological networks: comparison between graphs, between clusters, or between graphs and clusters; network randomization; analysis of degree distribution; network-based clustering and path finding. The tools are interconnected

The RDF Data Model (of triples)

- RSS (1.0) is built on top of the **Resource Description Framework (RDF)** in **RDF/XML**
- Diagrammatic representation of the data (below) illustrates the basic idea:
 - The data model establishes an **RDF Graph** which is a directed, graph associated with string markings
 - Semantic descriptions, **describing resources** (e.g. services, web pages, documents, people, concepts), are asserted as **statements**, each of which constitutes an **subject** node, a **predicate** arc, and an **object** node
 - Predicates (arcs) are always identified with an **URI string**, Subjects (nodes) may in addition be **blank**. Objects may be identified with URI strings, few types of **literal strings**, or they can be blank
 - There is no “closed-world assumption”, names are not unique unless asserted so, “**anyone can say anything about anything**”, etc., which leads into description logics kind of thinking
 - In addition, there are **common modelling primitives** (e.g. `rdf:type`), **new names** are introduced simply by introducing new URI strings, things may be organised hierarchical **classes**, etc.



**Semantic Web =
Union of a selection
of particular RDF
Graphs using
shared names in
descriptions**



A simple semantic query use case: Foreword

- Of course, there is no point in simply publishing data
- While using a common data structure obviously helps **reading semantic data** for purposes of domain-specific applications (and APIs...), perhaps the most easily understandable application is making general-purpose **semantic queries**

A simple semantic query use case

- A query example in SPARQL syntax : *“Find news items published since September 2008 that talk about biology or mathematics, using the two explicitly mentioned RSS sources”*

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rss: <http://purl.org/rss/1.0/>
SELECT ?link ?title
FROM <http://rss.slashdot.org/Slashdot/slashdotScience>
FROM <http://www.nature.com/nprot/current_issue/rss/index.html>
WHERE {
    ?i rss:link ?link .
    ?i dc:date ?date . FILTER (?date > "2008-08-31")
    ?i rss:description ?desc . FILTER regex(?desc,"biolog|mathematic","i")
    ?i rss:title ?title }
```

- Which returns (well, it did return, at the time of writing!):

link	title
"http://dx.doi.org/10.1038/nprot.2008.100"	"Network Analysis Tools: from biological networks to clusters and pathways"
"http://rss.slashdot.org/~r/Slashdot/slashdotScience/~3/VjuQziCKmwc/article.pl"	"The 23 Toughest Math Questions"
"http://rss.slashdot.org/~r/Slashdot/slashdotScience/~3/wL0o8vD8ECQ/article.pl"	"Towards a Wiki For Formally Verified Mathematics"
"http://rss.slashdot.org/~r/Slashdot/slashdotScience/~3/TwErT2npX_g/article.pl"	"Safe Stem Cells Produced From Adult Cells"
"http://dx.doi.org/10.1038/nprot.2008.145"	"Design and delivery of antisense oligonucleotides to block microRNA function in cultured Drosophila and human cells"

- (Of course, while this is useful, end-user application is still needed...)

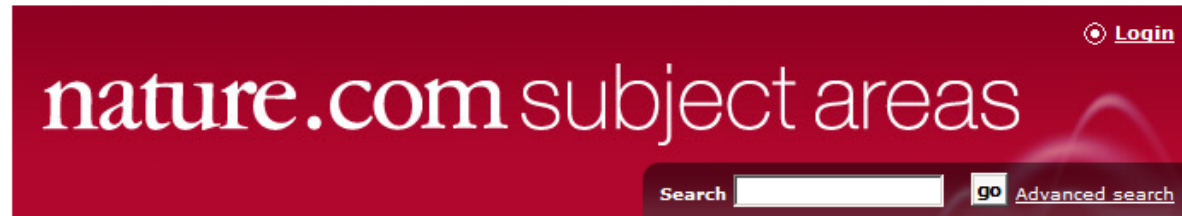
A simple semantic query use case: Notes

- Notes
 - besides a relational table (XML file), a different kind of (SPARQL) query might return a new RDF graph
 - a query finds only things it can reach
 - the example includes also a **keyword (sub)search** (using a regular expression), etc.
- Using RDF, almost anything can be described in applications, not just news. The same application pattern still applies:
 - learn about the available semantic vocabularies and applications, export and publish useful descriptions of data (perhaps owned by you) using appropriate semantic vocabulary, query the data, and exploit it in applications

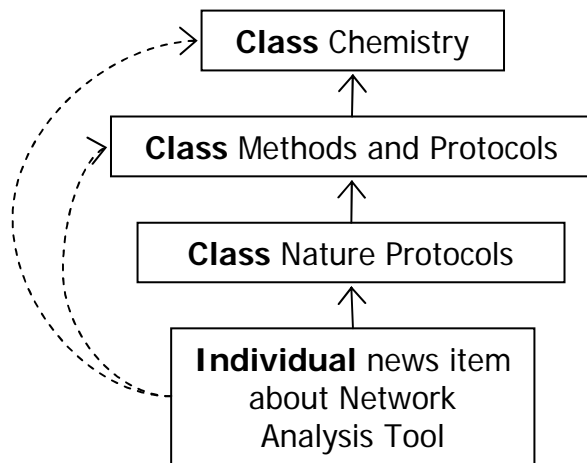
Psst. Of course, all applications might not follow this line of thinking explicitly. Further, graceful error management is typically required, etc.

Reasoning?

- In general, **reasoning** may be conceptualised as a process of **deriving (new, interesting) statements** from a **knowledge base**, using some **valid reasoning procedure**
- Mechanical reasoning is not just for “Artificial Intelligence geeks” – the simple query example already provides a practical use case: **When news items are classified by subject, more expressive semantic search tools can be implemented by adding reasoning functionality**
- Assuming subjects establish a hierarchical structure, simple **transitive reasoner** turns out to be useful in queries since news items may be searched by the inferred (super) categories



Chemistry	Clinical Practice & Research	Earth & Environment	Life Sciences	Physical Sciences
<ul style="list-style-type: none">▶ Chemistry▶ Drug discovery▶ Biotechnology▶ Materials▶ Methods and Protocols	<ul style="list-style-type: none">▶ Cancer▶ Cardiovascular medicine▶ Dentistry▶ Endocrinology▶ Gastroenterology and Hepatology▶ Methods and Protocols▶ Pathology and Pathobiology▶ Urology	<ul style="list-style-type: none">▶ Earth sciences▶ Evolution & Ecology	<ul style="list-style-type: none">▶ Biotechnology▶ Cancer▶ Development▶ Drug discovery▶ Evolution & Ecology▶ Genetics▶ Immunology▶ Medical research▶ Methods and Protocols▶ Microbiology▶ Molecular cell biology▶ Neuroscience▶ Pharmacology▶ Systems biology	<ul style="list-style-type: none">▶ Physics▶ Materials

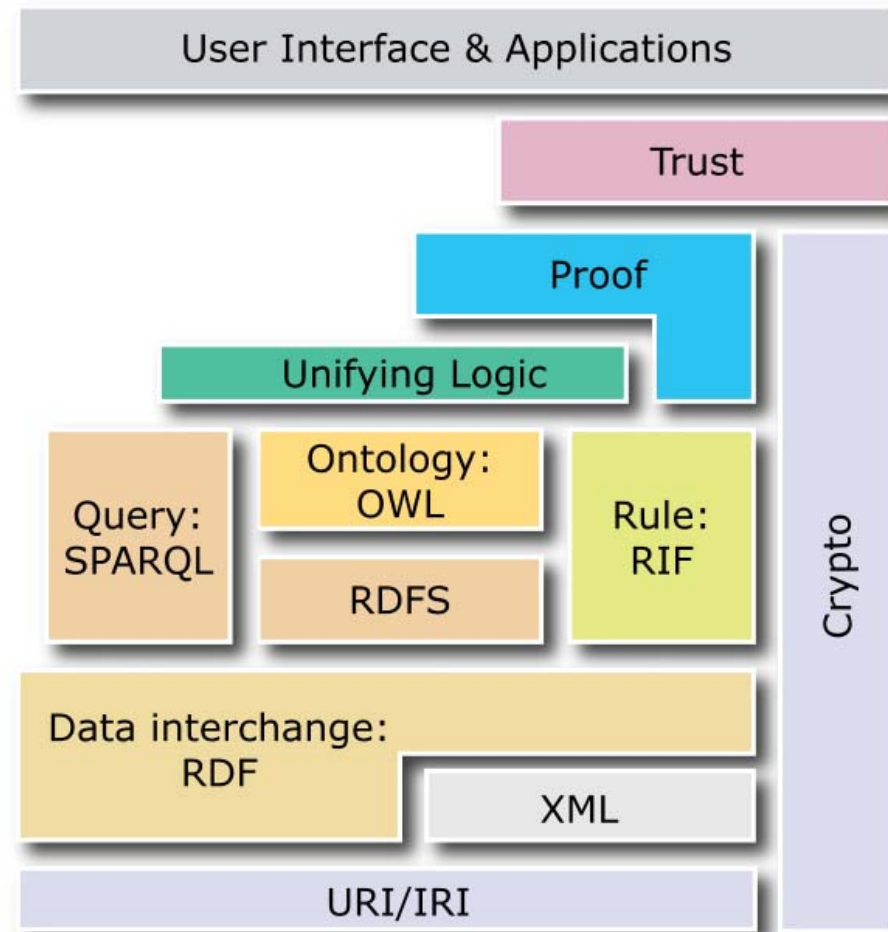


Psst. As suspected, common modelling primitives are required for asserting the appropriate descriptions (e.g. **RDF Schema**). Further, the SPARQL language is still useful, since a SPARQL query engine may provide options for using different reasoning systems

Web Ontology Language (OWL)? Rules? The other standard technologies?

- Sometimes more expressive concept systems are needed: The Semantic Web standards include the **knowledge representation ontology language OWL** for expressing, e.g., **domain ontologies**
- (An ontology identifies/formalises the concepts and relationships of a particular application.)
- With OWL, it is possible to assert things like...
 - what classes and predicates are supposed to be used in particular descriptions
 - which (individual, class or property) names are equivalent
 - whether a user-defined property *p* is e.g. functional or transitive and what values are supposed to be used, etc.
- Ontologies may also provide a terminology for asserting (e.g. Horn-like “if-then”) **rules** (standard underway)
- The other needed logical components include, e.g., establishing **trust**, e.g., via digital signatures and known trustees

Psst. Using OWL per se for making semiformal concept systems might be an overkill in restricted applications. (See, e.g., **SKOS Simple Knowledge Organization System**)



**Part 2/3:
Semantic Web specifications and
working groups at the W3C**

Core W3C Semantic Web specifications (Oct 2008)

- **Resource Description Framework (RDF):**
 - [RDF/XML Syntax Specification \(Revised\)](#), W3C Recommendation, February 10, 2004, Dave Beckett, ed.
 - [RDF Vocabulary Description Language 1.0: RDF Schema](#), W3C Recommendation, February 10, 2004, Ramanathan V. Guha, Dan Brickley, eds.
 - [Resource Description Framework \(RDF\): Concepts and Abstract Syntax](#), W3C Recommendation, February 10, 2004, Graham Klyne, Jeremy J. Carroll, eds.
 - [RDF Semantics](#), W3C Recommendation, February 10, 2004, Patrick Hayes, ed.
 - [RDF Test Cases](#), W3C Recommendation, February 10, 2004, Jan Grant, Dave Beckett, eds.
- **Gleaning Resource Descriptions from Dialects of Languages (GRDDL):**
 - [Gleaning Resource Descriptions from Dialects of Languages \(GRDDL\)](#), W3C Recommendation, September 11, 2007, Dan Connolly, ed.
 - [GRDDL Test Cases](#), W3C Recommendation, September 11, 2007, Chimezie Ogbuji, ed.
- **SPARQL Query Language for RDF:**
 - [SPARQL Query Language for RDF](#), W3C Recommendation, January 15, 2008, Andy Seaborne, Eric Prud'hommeaux, eds.
 - [SPARQL Protocol for RDF](#), W3C Recommendation, January 15, 2008, Elias Torres, Lee Feigenbaum, Kendall Grant Clark, eds.
 - [SPARQL Query Results XML Format](#), W3C Recommendation, January 15, 2008, Dave Beckett, Jeen Broekstra, eds.
- **Web Ontology Language (OWL):**
 - [OWL Web Ontology Language Reference](#), W3C Recommendation, February 10, 2004, Guus Schreiber, Mike Dean, eds. Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, Lynn Andrea Stein, authors.
 - [OWL Web Ontology Language Use Cases and Requirements](#), W3C Recommendation, February 10, 2004, Jeff Heflin, ed.
 - [OWL Web Ontology Language Semantics and Abstract Syntax](#), W3C Recommendation, February 10, 2004, Peter F. Patel-Schneider, Patrick Hayes, Ian Horrocks, eds.
 - [OWL Web Ontology Language Overview](#), W3C Recommendation, February 10, 2004, Deborah L. McGuinness, Frank van Harmelen, eds.
 - [OWL Web Ontology Language Test Cases](#), W3C Recommendation, February 10, 2004, Jeremy J. Carroll, Jos De Roo, eds.
 - [OWL Web Ontology Language Guide](#), W3C Recommendation, February 10, 2004, Michael K. Smith, Deborah L. McGuinness, Chris Welty, eds.
- You may want to look at the collection of [SW Case Studies and Use Cases](#) to see how organizations are using these technologies today.

Currently active working groups at the W3C

- **Semantic Web Coordination Group**
 - The [Semantic Web Coordination Group](#) is tasked to provide a forum for managing the interrelationships and interdependencies among groups focusing on standards and technologies that relate to this goals of the Semantic Web Activity. This group is designed to coordinate, facilitate and (where possible) help shape the efforts of other related groups to avoid duplication of effort and fragmentation of the Semantic Web by way of incompatible standards and technologies.
- **Rules Interchange Format Working Group**
 - This Working Group is chartered to produce a core rule language plus extensions which together allow rules to be translated between rule languages and thus transferred between rule systems. The Working Group will have to balance the needs of a community diverse including Business Rules and Semantic users Web specifying extensions for which it can articulate a consensus design and which are sufficiently motivated by use cases. (**OPAALS note: RIF may includes user perspective/use cases for SBVR.**)
- **OWL Working Group**
 - The mission of the OWL Working Group, is to produce a W3C Recommendation that refines and extends the 2004 version of [OWL](#). The proposed extensions are a small set that: have been identified by users as widely needed, and have been identified by tool implementers as reasonable and feasible extensions to current tools.
- **Semantic Web Deployment Working Group**
 - The mission of this Working Group is to provide guidance in the form of W3C Technical Reports on issues of practical RDF development and deployment practices in the areas of publishing vocabularies, OWL usage, and integrating RDF with HTML documents.
 - This group is also responsible for the development of the [RDFa](#) and [SKOS](#) specifications.
- **Semantic Web Interest Group**
 - The Semantic Web Interest Group is a forum for W3C Members and non-Members to discuss innovative applications of the Semantic Web. The Interest Group also initiates discussion on potential future work items related to enabling technologies that support the Semantic Web, and the relationship of that work to other activities of W3C and to the broader social and legal context in which the Web is situated.
- **Semantic Web Health Care and Life Sciences Interest Group**
 - The Semantic Web Health Care and Life Sciences Interest Group is designed to improve collaboration, research and development, and innovation adoption in the health care and life science industries. Aiding decision-making in clinical research, Semantic Web technologies will bridge many forms of biological and medical information across institutions.

**Part 3/3:
What about the Open Knowledge Space (OKS)?**



- Accepting that it is useful to consider that a certain aspect the OKS can technically modelled as a repository, Semantic Web technologies have things to offer
 - A **common data model** for asserting semantic descriptions
 - A common set of **modelling primitives and core vocabularies**
 - **Established applications**, e.g., for expressing concept systems
- This line of thinking, **the use case of a semantic search**, has essentially driven the specification of the (first versions of the) [OKS Data Model](#)
 - Further, when considering semantic integration (and the “Unifying Logic”), [Interpretation Logics](#) may provide one potential solution
- What the Semantic Web specifications do *not* (as such) provide, include:
 - The actual distributed **repository tools & services** for capturing, e.g., documents (representations), providing the query interface
 - **Community process** for collaboratively constructing knowledge

Endgame

Conclusion

- The W3C Semantic Web activities provide both technologies and a vision for making semantic computing applications
- While the standardisation process provides stable & usable specifications, Semantic Web is not “done yet” since new applications and standards are constantly needed
- The OPAALS OKS clearly benefits from adopting the basic logical semantic modelling stance for descriptions
 - ...while the OKS process description and a physical implementation, given the extra requirements, must be dealt in a more application-specific manner
- Thank you! (Questions?)



Useful resources

- Visit <http://www.w3.org> to learn more about the **World Wide Web Consortium (W3C)**
- Visit <http://www.w3.org/2001/sw/> to learn more about the **W3C Semantic Web**
- If interested in this area, why not contribute in the current and the future Web standards by **joining the W3C?** (See <http://www.w3.org/Consortium/join>)

- The examples of this tutorial were compiled using the following sources:
 - RSS feeds were picked from **Slashdot** (see <http://slashdot.org/>) and **nature protocols** (see <http://www.nature.com/nprot/journal/v3/n10/>)
 - The example RSS reader was **Mozilla Firefox with the Wizz RSS News Reader** (see <http://www.wizzrss.com/index.php>)
 - SPARQL processing was done using **Jena & ARQ** (see <http://jena.sourceforge.net/ARQ/>)
- When interested in other applications, see, e.g.,
 - **Dave Beckett's Resource Description Framework (RDF) Resource Guide** (see <http://planetrdf.com/guide/>)
- The **few OPAALS-specific references** mentioned in the presentation include
 - Nykänen, O., Huhtamäki, J., Salonen, J., Pohjolainen, S., & Silius, K. (Eds.) 2008. Proceedings of the 2nd International OPAALS Conference on Digital Ecosystems: OPAALS 2008. 7-8 October 2008, Tampere, Finland. 108 pages. A local copy is available at <http://matriisi.ee.tut.fi/hypermedia/julkaisut/opaals2008-proceedings.pdf> (PDF 3.3 MB)
 - Nykänen, O., Salonen, J., Huhtamäki, J., & Haapaniemi, M. 2008. *OKS Data Model, version 1.01*. A milestone specification for the OPAALS PROJECT (Contract number IST-034824), WP10: Sustainable Research Community Building in the Open Knowledge Space. Contribution to the Milestone M10.10: OKS Data Model (M24), 1 July 2008 (29 pages). Available at http://wiki.opaals.org/M10.10_OKSDataModel . A local copy is available at <http://matriisi.ee.tut.fi/hypermedia/julkaisut/20080701-oks-dm-v1-01.pdf> (PDF 0.5 MB)
 - Nykänen, O. 2007. Interpretation Logics. Proceedings of the 1st OPAALS conference, 26-27 November 2007, Rome, Italy. A local copy is available at <http://matriisi.ee.tut.fi/hypermedia/julkaisut/2007-nykanen-ilogics.pdf> (PDF 191 Kb) (Note that the proceedings was published afterwards, 15 July 2008.)