

State of the Semantic Web

Karl Dubost and Ivan Herman, W3C

INTAP Semantic Web Conference, Tokyo, Japan, March 7, 2008

> Significant buzz...



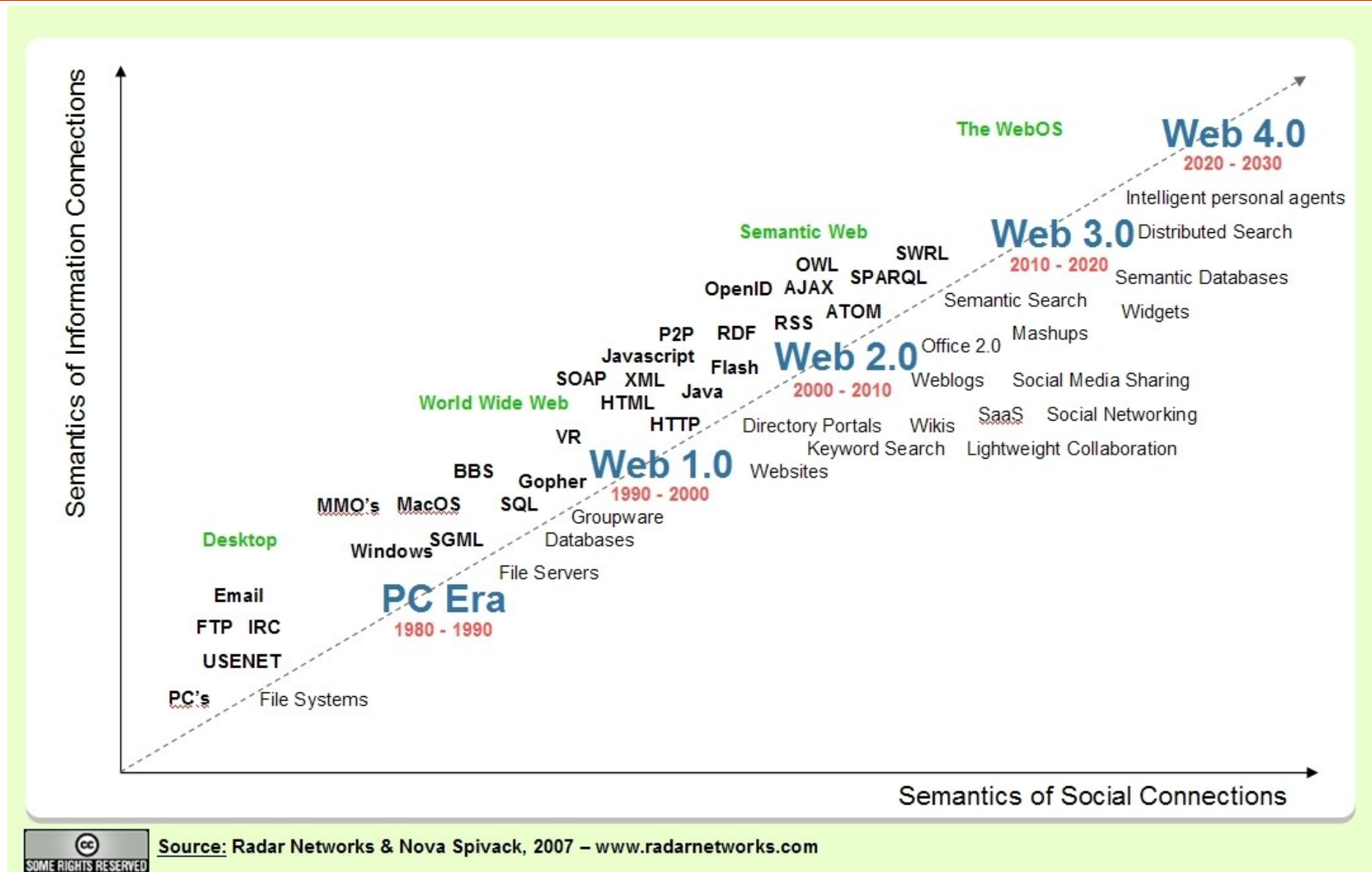
- There is quite a buzz around “Semantics”, “Semantic Technologies”, “Semantic Web”, “Web 3.0”, “Data Web”, etc, these days
- New applications, companies, tools, etc, come to the fore frequently
- It is, of course, not always clear what these terms all mean:
 - “Semantic Web” is a way to specify data and data relationships; it is also a collection of specific technologies (RDF, OWL, GRDDL, SPARQL, ...)
 - “Semantic Technologies”, “Web 3.0” often mean more, including intelligent agents, usage of complex logical procedures, etc

> Significant buzz... (cont.)



- Predicting the exact evolution in terms of Web 3.0, Web 4.0, etc, is a bit as looking into a crystal ball
- But the *Semantic Web technologies are already here, are used and deployed*
- They are at the basis of further evolution

> A vision on the evolution...



(this Web 3.0 is not identical to the “journalistic” Web3.0; merely timing)

> The 2007 Gartner predictions



During the next 10 years, Web-based technologies will improve the ability to embed semantic structures [... it] will occur in multiple evolutionary steps...

By 2017, we expect the vision of the Semantic Web [...] to coalesce [...] and the **majority of Web pages are decorated with some form of semantic hypertext.**

By **2012, 80% of public Web sites will use some level of semantic hypertext to create SW documents [...]** 15% of public Web sites will use more extensive Semantic Web-based ontologies to create semantic databases

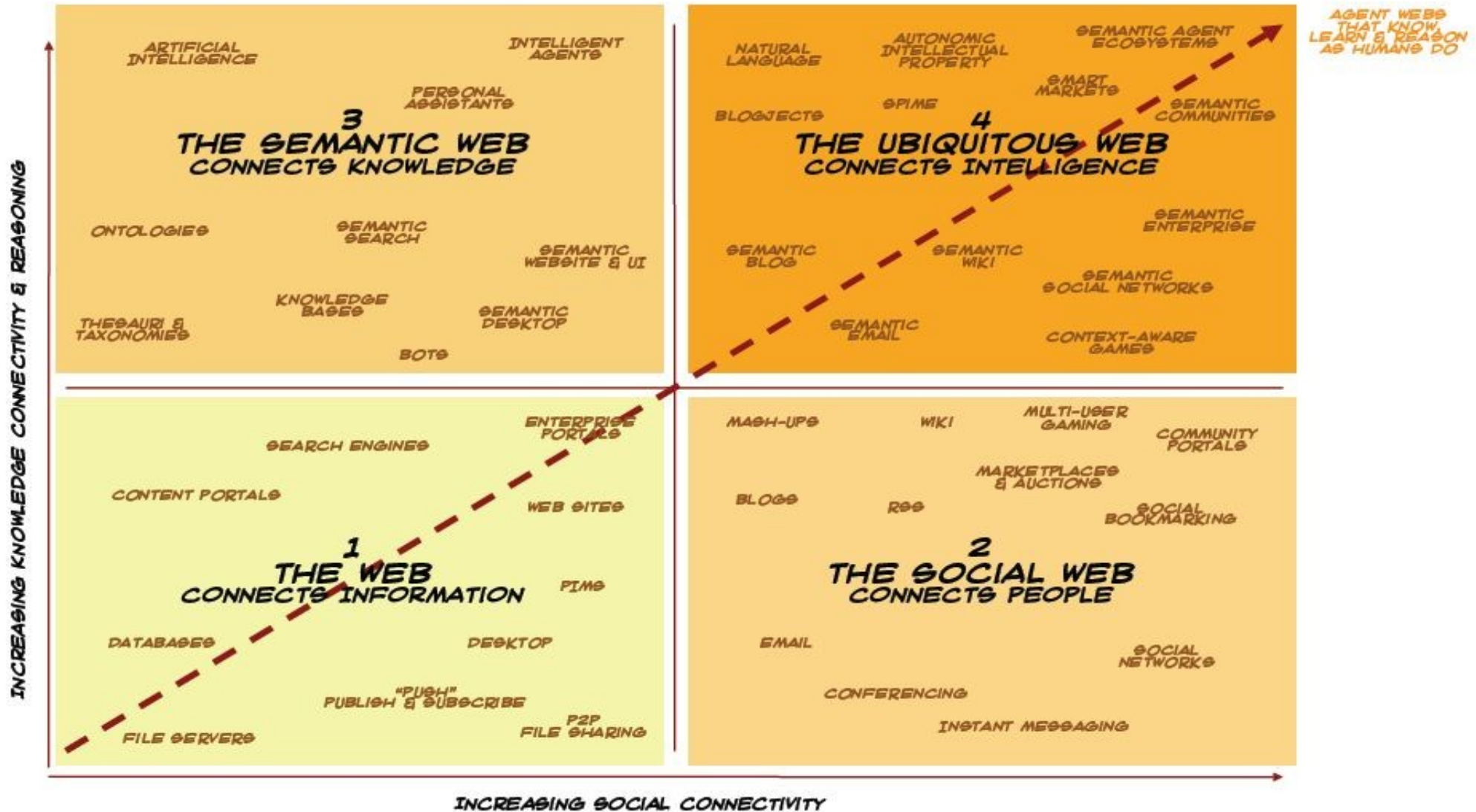
(note: “semantic hypertext” refers to, eg, RDFa, microformats with possible GRDDL, etc.)

Source: “Finding and Exploiting Value in Semantic Web Technologies on the Web”, Gartner Research Report, May 2007

> Another longer term vision...



(from the “Semantic Wave 2008” report, from Project10X)

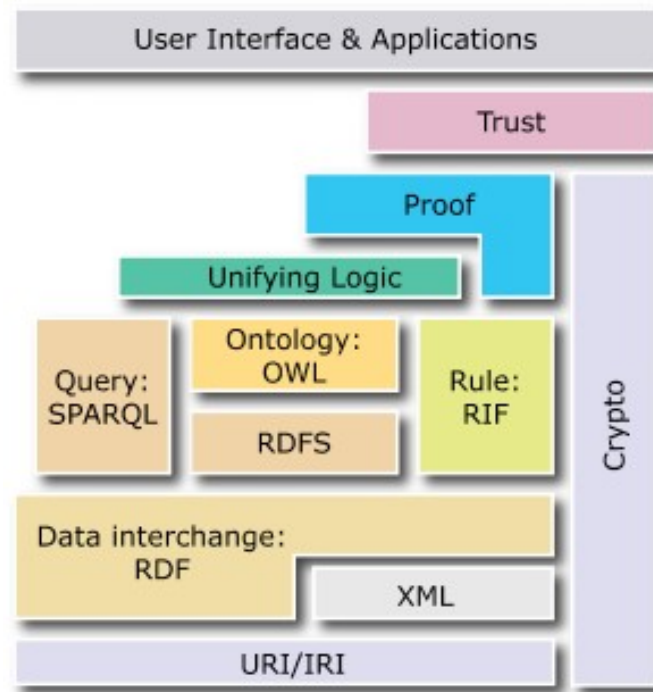


Courtesy of Mills Davis, Project10X; source: Nova Spivack, Radar Networks and John Breslin, DERI

> Let us keep to the Semantic Web for now...



- In what follows we will restrict ourselves to the Semantic Web
 - a way to specify data and data relationships
 - allows data to be shared and reused across application, enterprise, and community boundaries
 - a collection of fundamental technologies (RDF/S, OWL, GRDDL, SPARQL, ...)



> The “corporate” landscape is moving



- Major companies offer (or will offer) Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, GE, Northrop Gruman, Altova, Microsoft, Dow Jones, ...
- Others are using it (or consider using it) as part of their own operations: Novartis, Boeing, Pfizer, Telefónica, ...
- Some of the names of active participants in W3C SW related groups: ILOG, HP, Agfa, SRI International, Fair Isaac Corp., Oracle, Boeing, IBM, Chevron, Siemens, Nokia, Pfizer, Sun, Eli Lilly, ...

> Some SW Tools (not and exhaustive list!)



• Triple Stores

- RDFStore, AllegroGraph, Tucana
- RDF Gateway, Mulgara, SPASQL
- Jena's SDB, D2R Server, SOR
- Virtuoso, Oracle11g
- Sesame, OWLIM, Tallis Platform
- ...

• Reasoners

- Pellet, RacerPro, KAON2, FaCT++
- Ontobroker, Ontotext
- SHER, Oracle 11g, AllegroGraph
- ...

• Converters

- flickurl, TopBraid Composer
- GRDDL, Triplr, jpeg2rdf
- ...

• Search Engines

- Falcon, Sindice, Swoogle
- ...

• Middleware

- IODT, Open Anzo, DartGrid
- Ontology Works, Ontoprise
- Profium Semantic Information Router
- Software AG's EII
- Thetus Publisher, Asio, SDS
- ...

• Semantic Web Browsers

- Disco, Tabulator, Zitgist, OpenLink Viewer
- ...

• Development Tools

- SemanticWorks, Protégé
- Jena, Redland, RDFLib, RAP
- Sesame, SWI-Prolog
- TopBraid Composer
- DOME
- ...

• Semantic Wiki systems

- Semantic Media Wiki, Platypus
- Visual knowledge

Inspired by “Enterprise Semantic Web in Practice”, Jeff Pollock, Oracle. See also [W3C's Wiki Site](#).

> Some SW tools (cont.)



- Significant speed, store capacity, etc, improvements are reported every day
- 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!
- We still need more “middleware” tools to properly *combine* what is already available...
- *Anybody can start developing RDF-based applications today*



Let us look at the technical state of the SW first

> Querying RDF: SPARQL



- Querying RDF graphs is essential (can you imagine Relational Databases without SQL?)
- SPARQL is
 - a query language based on graph patterns
 - a protocol layer to use SPARQL over, eg, HTTP
 - an XML return format for the query results
- Is a W3C Standard (since January 2008)
- Numerous implementations are already available (eg, built in triple stores)

> Some new technologies at W3C



- SPARQL
- GRDDL
- RDFa
- SKOS
- OWL 1.1
- RIF (Rules)

> SPARQL (cont.)



- There are also SPARQL “endpoints” services on the Web:
 - send a query and a reference to data over HTTP GET, receive the result in XML or JSON
 - big datasets often offer “SPARQL endpoints” to query local data
 - applications may not need any direct RDF programming any more, just use a SPARQL processor
- SPARQL can also be used to construct graphs!

> The power of CONSTRUCT



```
CONSTRUCT {  
  <http://dbpedia.org/resource/Amitav_Ghosh> ?p1 ?o1.  
  ?s2 ?p2 <http://dbpedia.org/resource/Amitav_Ghosh>.  
}  
WHERE {  
  <http://dbpedia.org/resource/Amitav_Ghosh> ?p1 ?o1.  
  ?s2 ?p2 <http://dbpedia.org/resource/Amitav_Ghosh>.  
}
```

- SPARQL endpoint
- returns RDF/XML



```
SELECT  
FROM <http://dbpedia.org/sparql/?query=CONSTRUCT+%7B++...>  
WHERE {  
  ?author_of dbpedia:author res:Amitav_Ghosh.  
  res:Amitav_Ghosh dbpedia:reference ?homepage;  
                  rdf:type          ?type;  
                  foaf:name         ?foaf_name.  
  FILTER regex(str(?type), "foaf")  
}
```

- Data reused in a query elsewhere...

> 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...

> Bridge to relational databases



- Most of the data on the Web are stored in relational databases
 - “RDFying” them is an impossible
 - relational databases are here to stay...
- “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 languages/approaches are being used
 - a number of systems can now be used as relational database as well as triple stores (eg, Oracle, OpenLink, ...)
- Work for a survey on mapping techniques benchmarks may start soon at W3C
- SPARQL is becoming the tool of choice to query the data
 - ie, “SPARQL endpoints” are defined to query the databases

> How to get RDF data?



- Of course, one could create RDF data manually...
 - ... but that is unrealistic on a large scale
- Goal is to generate RDF data automatically when possible and “fill in” by hand only when necessary
- We have already seen the work relating to “traditional” databases
- But there are also other types of data out there, too...

> Data may be extracted (a.k.a. “scraped”)



- Different tools, services, etc, come around:
 - get RDF data associated with images, for example:
 - service to [get RDF from flickr images](#)
 - service to [get RDF from XMP](#)
 - scripts to convert spreadsheets to RDF
 - etc
- Many of these tools are still individual “hacks”, but show a general tendency
- Hopefully more tools will emerge



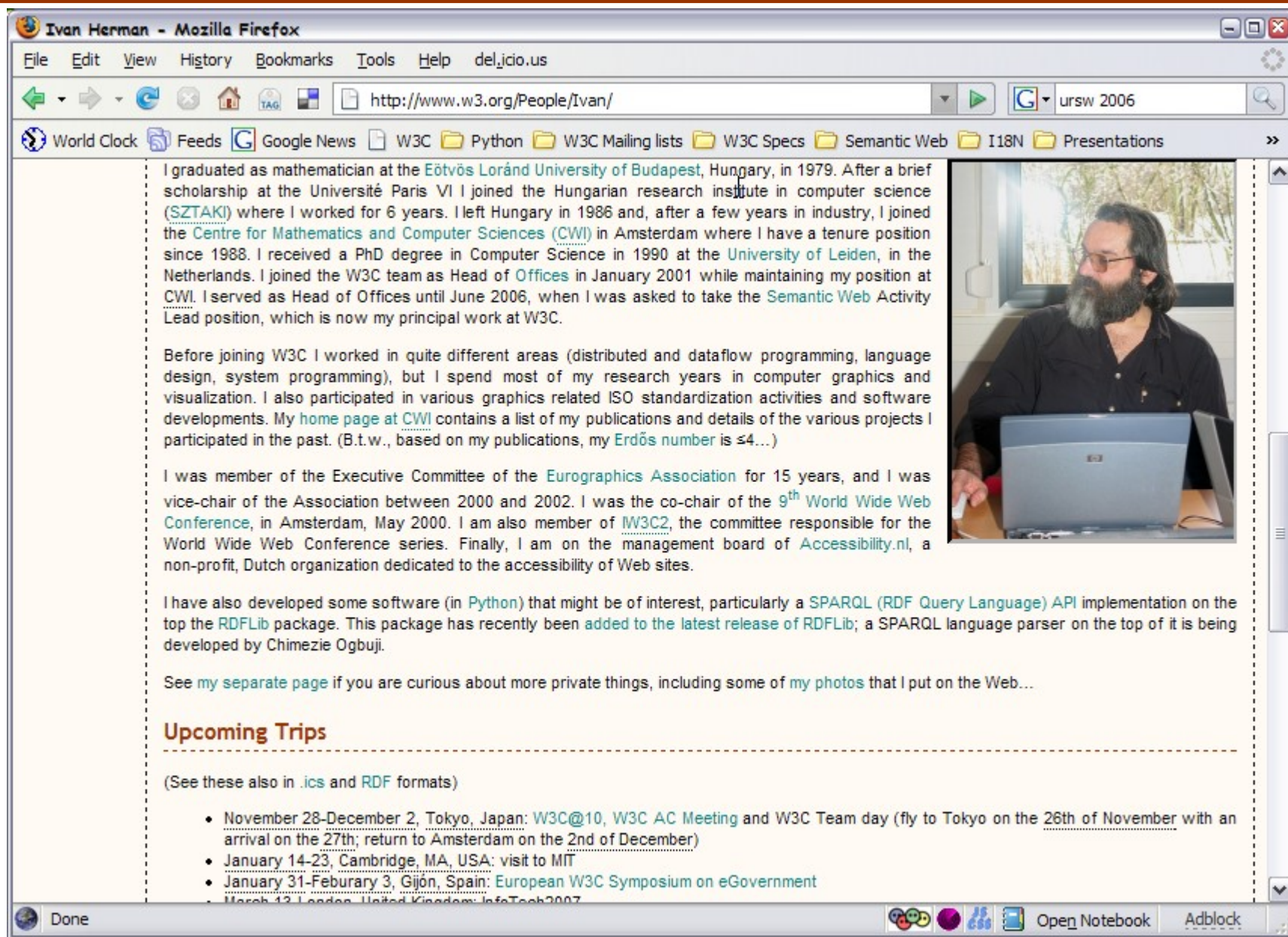
- GRDDL is a way to access structured data in XML/XHTML and turn it into RDF:
 - defines XML attributes to bind a suitable script to transform (part of) the data into RDF
 - script is usually XSLT but not necessarily
 - has a variant for XHTML
 - a “GRDDL Processor” runs the script and produces RDF on–the–fly
- A way to access existing structured data and “bring” it to RDF
 - eg, a possible link to microformats
 - exposing data from large XML use bases, like XBRL

> Getting structured data to RDF: RDFa



- RDFa extends XHTML with a set of attributes to include structured data into XHTML
- Makes it easy to “bring” existing RDF vocabularies into XHTML
- Uses namespaces for an easy mix of terminologies
- It can also be used with GRDDL
 - but: no need to implement a separate transformation per vocabulary

> GRDDL & RDFa: Ivan's home page...



> ...marked up with GRDDL headers...



```
Source of: http://www.w3.org/People/Ivan/ - Mozilla Firefox
File Edit View Help

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-
<html xmlns="http://www.w3.org/1999/xhtml" lang="en"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:con="http://www.w3.org/2000/10/swap/pim/contact#"
  xmlns:wot="http://xmlns.com/wot/0.1/"
  xmlns:vccard="http://www.w3.org/2001/vcard-rdf/3.0#"
  xmlns:cal="http://www.w3.org/2002/12/cal/ical#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:bio="http://purl.org/vocab/bio/0.1/">
  <head profile="http://www.w3.org/2003/g/data-view">
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
    <title>Ivan Herman</title>
    <link rel="stylesheet" type="text/css" href="StyleSheets/Private.css" />
    <link rel="meta" type="application/rdf+xml" title="FOAF" href="http://www.ivan-herman.net/foaf.
    <script src="/2004/08/TalkFiles/popup.js" type="text/javascript"><!-- The popup script --></scri
    <link rel="transformation" href="http://www.w3.org/2002/12/cal/glean-hcal.xsl" />
    <link rel="transformation" href="http://www-sop.inria.fr/acacia/soft/RDFa2RDFXML.xsl"/>
  </head>
  <body xml:lang="en" lang="en">
    <div class="left">
      <a href="/"></a><br>
    </div>

    <div class="right" id="me" role="foaf:Person">
      <link rel="owl:sameAs" href="http://www4.wiwiw.fu-berlin.de/dblp/resource/person/103481"/>
      <link rel="owl:sameAs" href="http://ivan-herman.net/Ivan_Herman"/>
      <link rel="owl:sameAs" href="http://ivan-herman.net/foaf.rdf#me"/>
      <h1 property="foaf:name">Ivan Herman</h1>
      Accessibility.nl</a>,
a non-profit, Dutch organization dedicated to the accessibility of Web sites. </p>

<p>I have also developed some software (in <a href="http://www.python.org">Python</a>) that might
a <a href="http://dev.w3.org/cvsweb/%7Echeckouts/2004/PythonLib-IH/Doc/sparqlDesc.html">SPARQL
on the top of the <a href="http://rdflib.net/">RDFLib</a> package.
This package has recently been <a href="http://www.ivan-herman.net/WebLog/WorkRelated/SemanticWe
added to the latest release of RDFLib</a>; a SPARQL language parser on the top of it is being de
Chimezie Ogbuji.
</p>

See <a rel="foaf:homepage" href="http://www.ivan-herman.net">my separate page</a> if you are curious about
private things, including some of <a href="http://www.ivan-herman.net/Photos/">my photos</a> that I put

<h2><a id="trips" name="trips">Upcoming Trips</a></h2>

<p>(See these also in <a href="http://feeds.technorati.com/events/http%3A//www.w3.org/People/Ivan
and <a href="http://www.w3.org/2000/06/webdata/xslt?xslfile=http%3A%2F%2Fwww.w3.org%2F200
<ul>

<li><span class="vevent" id="ac06"><abbr class="dtstart" title="2006-11-28">November 28<
<span class="location vcard"><abbr class="geo" title="35.670685;139.770813">Tokyo, Japan
<span class="summary"><a href="http://www.w3.org/Member/Meeting/2006ac/November/" class=
W3C AC Meeting</a> and W3C Team day</span></span> (<span class="vevent" id="flyTok06"><s
<abbr class="dtstart" title="2006-11-26">26th of November</abbr> with an arrival on the<
27th</abbr></span>; <span class="vevent" id="flyTokAms06"><span class="summary">return
<abbr class="dtstart" title="2006-12-02">2nd of December</abbr></span></li>

<li class="vevent" id="mitjan"><abbr class="dtstart" title="2007-01-14">January 14</abbr>
<span class="location vcard"><abbr class="geo" title="42.365;-71.105">Cambridge, MA, USA
<span class="summary">visit to MIT</span></li>

<li class="vevent" id="logovian"><abbr class="dtstart" title="2007-01-31">January 31</abbr>
```

> ...yielding; ...



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xml:base="http://www.w3.org/People/Ivan/">
  <c:Vcalendar xmlns:r="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
              xmlns:ical=... >
    <c:component>
      <c:Vevent r:about="#ac06">
        <ical:summary>W3C@10, W3C AC Meeting and W3C Team day</ical:summary>
        <ical:dtstart>2006-11-28</ical:dtstart>
        <ical:dtend>2006-12-03</ical:dtend>
        <ical:url
          r:resource="http://www.w3.org/Member/Meeting/2006ac/November/" />
        <loc:location xml:lang="en">Tokyo, Japan</location>
        <geo:geo r:parseType="Resource">
          <r:first>35.670685</r:first>
          <r:rest r:parseType="Resource">
            ...
          </r:rest>
        </geo:geo>
      </c:Vevent>
    </c:component>
  </c:Vcalendar>
</rdf:RDF>
```

> ...marked up with RDFa tags...



```
Source of: http://localhost:8001/People/Ivan/OverviewNewRDFa.php - Mozilla Firefox
File Edit View Help

<dt>Postal address:</dt>
<dd>C/o <a href="http://www.w3c.nl">W3C Benelux Office</a> at
      <a href="http://www.cwi.nl">Centre for Mathematics and Computer Sciences
(CWI)</a> <br />
      <a
href="http://www.google.com/maps?f=q&hl=en&q=kruislaan+413,+amsterdam,+the+netherlands&ie=UTF8&z=13&
413, P.O. Box 94079,
      1090 GB Amsterdam, The Netherlands</a>.</dd>
<dt>Phone numbers:</dt>
<dd>
      phone: <a rel="foaf:phone" href="tel:+31-20-5924163">+31-20-5924163</a><br />
      mobile phone: <a rel="foaf:phone" href="tel:+31-641044153">+31-641044153</a> <br />
      fax: +31-20-5924312
</dd>
<dt>PGP/GPG:</dt>
<dd>My <a rel="wot:pubkeyAddress" href="http://www.ivan-herman.net/pgpkey.html">GnuPGP
key and signature</a> is
      available on-line.</dd>
<dt>FOAF:</dt>
<dd>You can either extract a short FOAF information from this page using an
      <a href="http://triplr.org/rdfa-rdf/http://www.w3.org/People/Ivan/">RDFa service</a>,
      of consult <a href="http://www.ivan-herman.net/foaf.rdf" rel="rdfs:seeAlso">my more
complete, public FOAF file</a>.</dd>
<dt>Misc:</dt>
<dd>I am often on the freenode #swig irc channel, usually using the nickname
      <span instanceof="foaf:OnlineChatAccount" rel="foaf:holdsAccount">
        <span rel="foaf:accountServiceHomepage"
resource="http://www.freenode.net/irc_servers.shtml"/>
        <span property="foaf:accountName">IvanHerman</span> or 'ivan'
      </span>
      <br/>I am also on <a href="http://www.facebook.com/p/Ivan_Herman/555188827"
rel="rdfs:seeAlso">Facebook</a>,
      <a href="http://www.dopplr.com/traveller/IvanHerman" rel="rdfs:seeAlso">Dopplr</a>,
      <a href="http://www.linkedin.com/profile?viewProfile=&key=2352277"
rel="rdfs:seeAlso">LinkedIn</a>, and
      <a href="http://www.flickr.com/photos/ivan_herman/" rel="rdfs:seeAlso">Flickr</a>
      <br/>
      My URI (as a real person): <a
href="http://www.ivan-herman.net/Ivan_Herman"><code>http://www.ivan-herman.net/Ivan_Herman</code></a>
      </dd>
</dl>
```

> ... yielding



```
@prefix foaf: <http://xmlns.com/foaf/0.1/>
@prefix wot: <http://xmlns.com/wot/0.1/>
...
@base <http://www.w3.org/People/Ivan/>

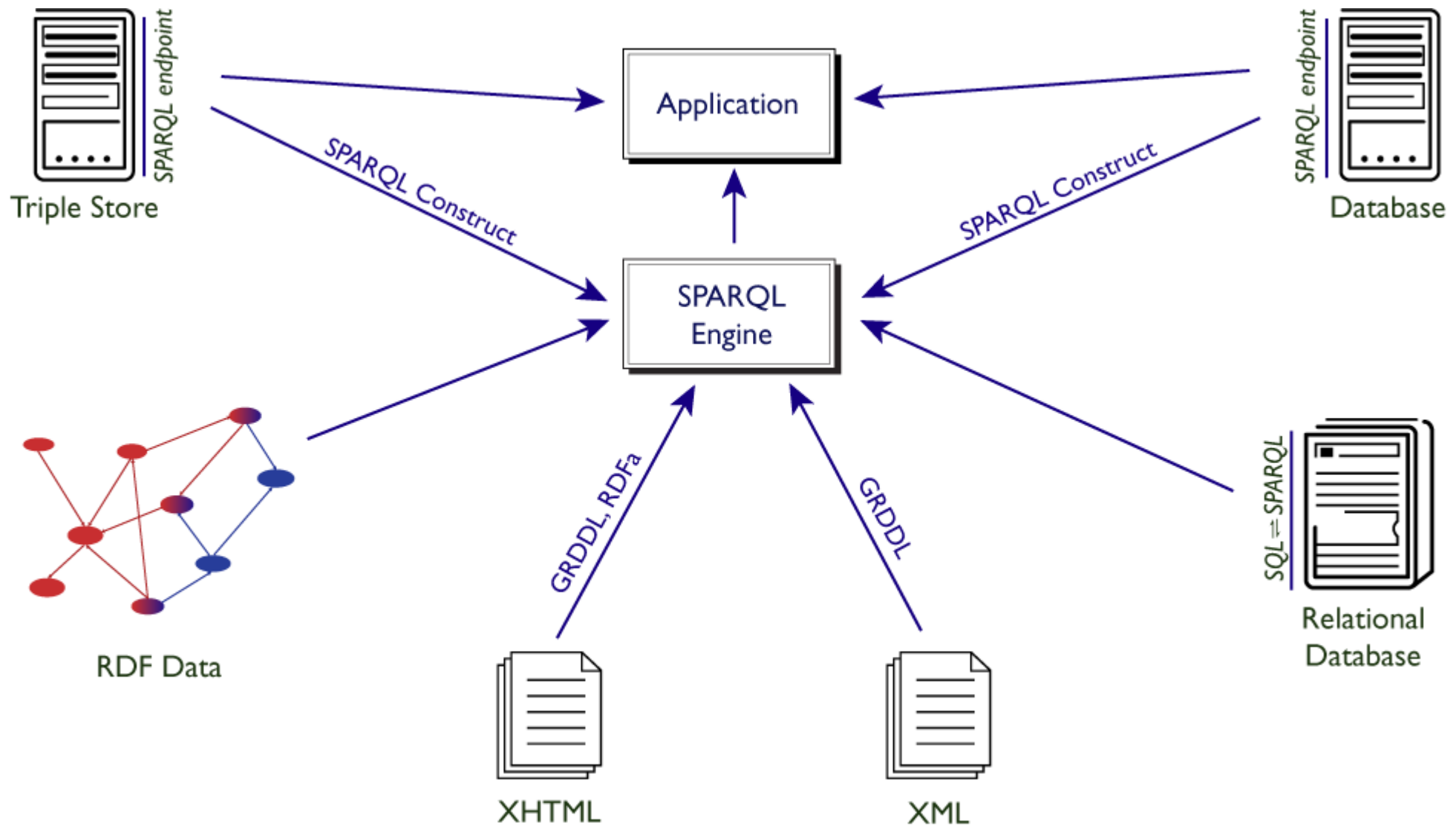
<#me>
  foaf:phone <tel:+31-20-5924163>;
  foaf:phone <tel:+31-641044153>;
  wot:pubkeyAddress <http://www.ivan-herman.net/pgpkey.html>;
  rdfs:seeAlso <http://www.ivan-herman.net/foaf.rdf>;
  foaf:holdsAccount
    [ a foaf:OnlineChatAccount;
      foaf:accountServiceHomepage
        <http://www.freenode.net/irc_servers.html>;
      foaf:accountName "IvanHerman";
    ];
  rdfs:seeAlso <http://www.facebook.com/p/Ivan_Herman/555188824>;
  ...
```


> Such data can be SPARQL-ed



```
SELECT DISTINCT ?name ?home ?orgRole ?orgName ?orgHome
# Get RDFa from my home page:
FROM <http://www.w3.org/People/Ivan/>
# GRDDL-ing http://www.w3.org/Member/Mail:
FROM <http://www.w3.org/Member/Mail/>
WHERE {
  ?foafPerson    foaf:mbox ?mail;
                  foaf:homepage ?home.
  ?individual    contact:mailbox ?mail;
                  contact:fullName ?name.
  ?orgUnit ?orgRole ?individual;
                  org:name ?orgName;
                  contact:homePage ?orgHome.
}
```

> SPARQL as a unifying point!



> Simple Knowledge Organization System



- Goal: representing and sharing classifications, glossaries, thesauri, etc, as developed in the “Print World”. For example:
 - [Dewey Decimal Classification](#), Art and Architecture Thesaurus, ACM classification of keywords and terms...
 - DMOZ categories (a.k.a. [Open Directory Project](#))
- The system must be simple to allow for a quick port of traditional data (done by non-experts in, say, Semantic Web)
- This is where SKOS comes in: define classes, properties, where those structures can be added

> Example: thesaurus



Term

Economic cooperation

Used For

Economic co-operation

Broader terms

Economic policy

Narrower terms

Economic integration, European economic cooperation, ...

Related terms

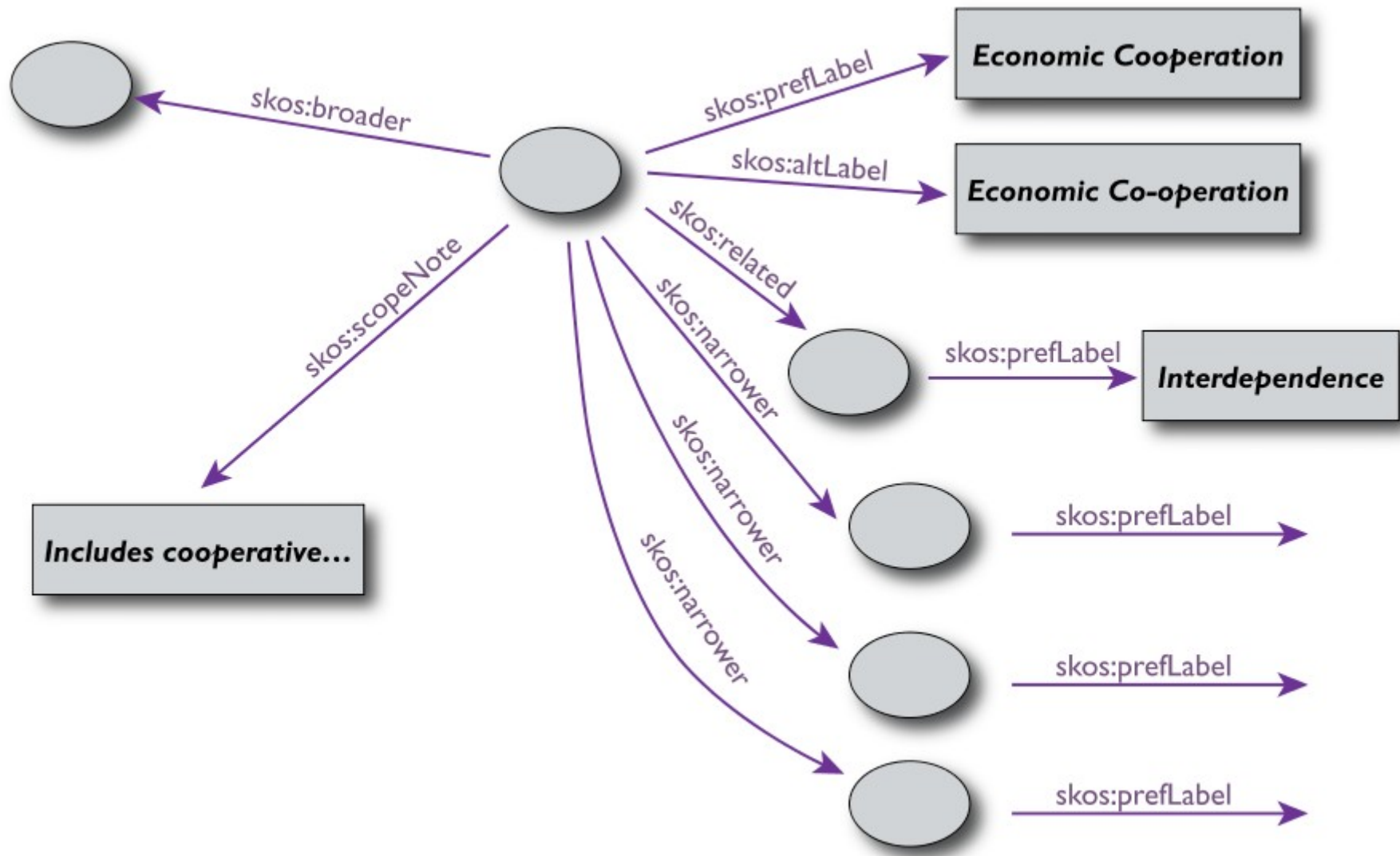
Interdependence

Scope Note

Includes cooperative measures in banking, trade, ...

(from the UK Archival Thesaurus)

> Example: thesaurus in SKOS



> SKOS and digital libraries



- SKOS plays an important role in “bridging” to digital libraries
- A huge community out there with its own traditions, style...
 - ... but huge amount of data to be “linked” to the Semantic Web!
- Major library metadata standards are being re-defined in terms of RDF (and SKOS),
 - eg, “Resource Description and Access” (RDA)
 - a major cataloging rule set for librarians
 - potentially, all major library catalogs around the globe could be translated into RDF and, eg, linked as an Open Linked Data...



- Large ontologies are being developed (converted from other formats or defined in OWL). For example:
 - [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; or [UniProt](#) for protein sequence and annotation terminology and data
 - [BioPAX](#): for biological pathway data
 - [ISO 15926](#): “Integration of life-cycle data for process plants including oil and gas production facilities”

> OWL in applications



- An increasing number of applications rely on OWL (Pfizer, Nasa, Eli Lilly, Elsevier, FAO, ...)
 - see some more example at the end of the talk
- Not all use complex reasoning; in many cases a small fraction of OWL is used

> New OWL Working Group



- A new Working Group just started on the revision of OWL
- The goal of the group:
 1. add a few extensions to current OWL that are useful, and is known to be implementable
 - many things happened in research since 2004
 - features should (if possible) be valid both in the DL and OWL Full world
 2. define fragments, ie, “profiles” of OWL that are:
 - smaller, easier to implement and deploy
 - cover important application areas and are easily understandable to non-expert users

> “OWL 1.1”: new proposed features



- “Qualified cardinality restrictions” (eg, “class instance must have two black cats”)
- Disjoint, reflexive, irreflexive properties; disjoint union of classes
- Property chains (eg, the uncle example: “if y is father x of y and y is brother of z, then z is uncle of x”)
- Own datatype constructs instead of complex XML Schema datatypes
 - eg, to express restrictions like number intervals easily

> “OWL 1.1”: new proposed features (cont)



- Metamodeling (a.k.a. “punning”): the same symbol may be used both as, e.g., a Class and an Instance, or for a datatype and an object property
 - this is not a problem in OWL Full, but is a significant restriction in OWL DL
 - in the DL there would still be some restrictions on how that can be used (eg, not all “natural” inferences can be drawn)

> “OWL 1.1”: small fragments



- For a number of applications RDFS is not enough, but even OWL Lite is too much (and too complex to implement)
- There is a need for (very) “light” versions of OWL: just a few extra possibilities added to RDFS
- Some can be as simple as having only (on top of RDFS):

`equivalentClass`
`equivalentProperty`
`sameAs`

`inverseOf`
`TransitiveProperty`
`SymmetricProperty`
`FunctionalProperty`
`InverseFunctionalProperty`

> “OWL 1.1”: small fragments (cont.)



- There are a number of proposals, papers, prototypes (and implementations!). Eg:
 - EL++, DLP: all DL dialects (e.g., EL++ is already in use by the health care community for medical ontologies)
 - pD*, OWLPrime: OWL Full dialects, that can be implemented with rule engines on top of, say, database engines
- It may be possible to create a (or more) dialect that may have both a DL and an OWL Full semantics (eg, OWLPrime~DLP)
- The Working Group will have to settle on the final list and structure



- There is a long history of rule languages and rule-based systems
 - eg: logic programming (Prolog), production rules
- Lots of small and large rule systems (from mail filters to expert systems)
- Hundreds of niche markets

> Why rules on the Semantic Web?



- There are conditions that ontologies (ie, OWL) cannot express (or only with difficulties)
 - a well known examples is Horn rules: $(P1 \wedge P2 \wedge \dots) \rightarrow C$
- There are conditions that are complicated in rules and ontologies are better (eg, complex classification of terms)
- Simple rule engines might be easier to implement (eg, on top of database engines)
- A different way of thinking — people may feel more familiar in one or the other

> Things you may want to express



- An example:
 - “if two Persons have the same name and the same email, or the same name and the same home page, then they are identical”
- Something like (with an ad-hoc syntax):

```
If { ?x rdf:type foaf:Person.  
    ?y rdf:type foaf:Person.  
    ?x foaf:name ?n.  
    ?x foaf:homepage ?h.  
    ?y foaf:name ?n.  
    ?y foaf:homepage ?h. }  
then { ?x = ?y }
```

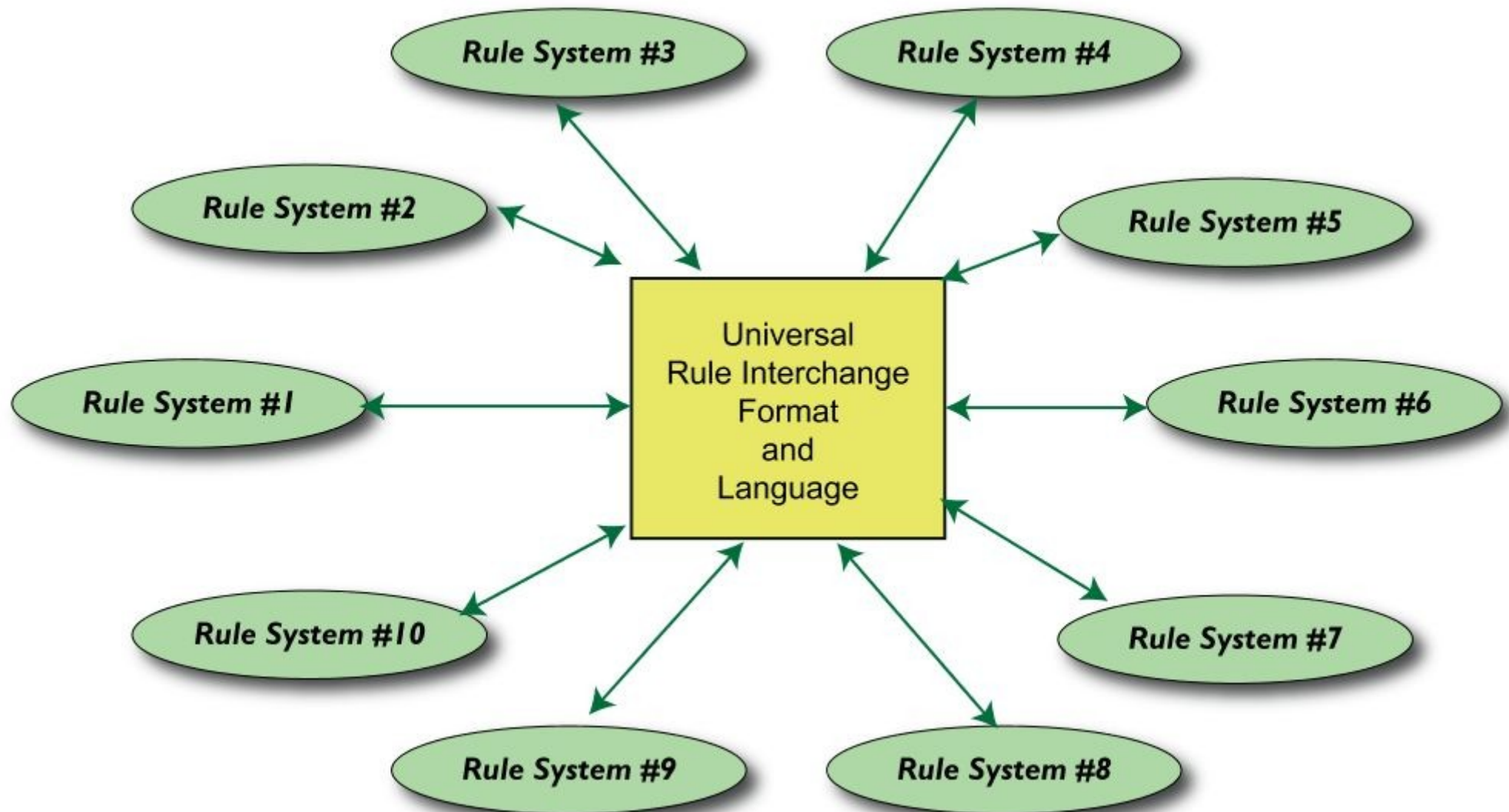
```
If { ?x rdf:type foaf:Person.  
    ?y rdf:type foaf:Person.  
    ?x foaf:name ?n.  
    ?x foaf:mailbox ?h.  
    ?y foaf:name ?n.  
    ?y foaf:mailbox ?m. }  
then { ?x = ?y }
```

> A new requirement: exchange of rules



- Applications may want to *exchange* their rules:
 - negotiate eBusiness contracts across platforms: supply vendor-neutral representation of your business rules so that others may find you
 - describe privacy requirements and policies, and let clients “merge” those (e.g., when paying with a credit card)
- Hence the name of the working group: Rule Interchange Format
 - a language that
 - expresses the rules a bit like a rule language with, eg, RDF
 - can be used to exchange rules among engines

> In an ideal World

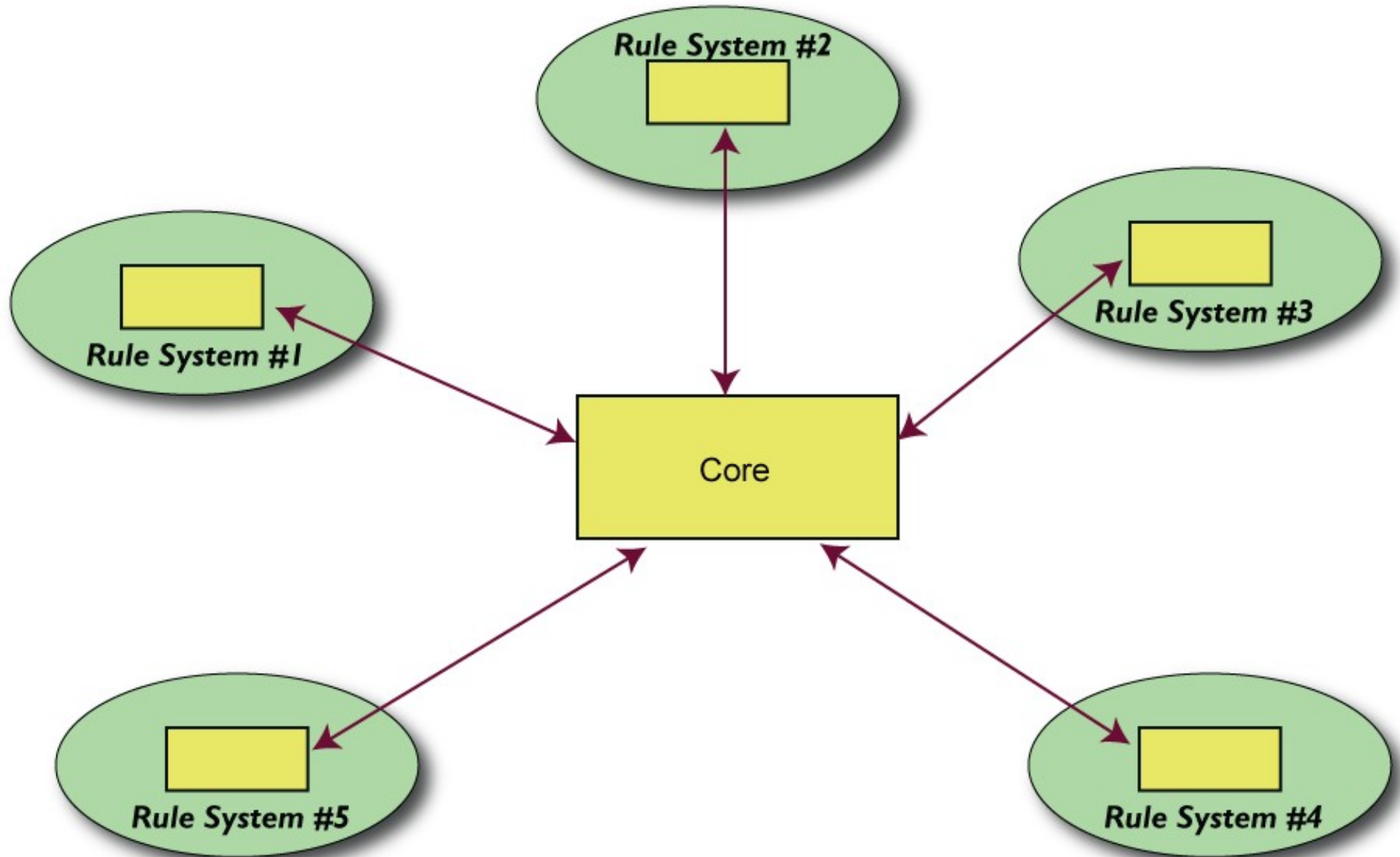


> In the real World...

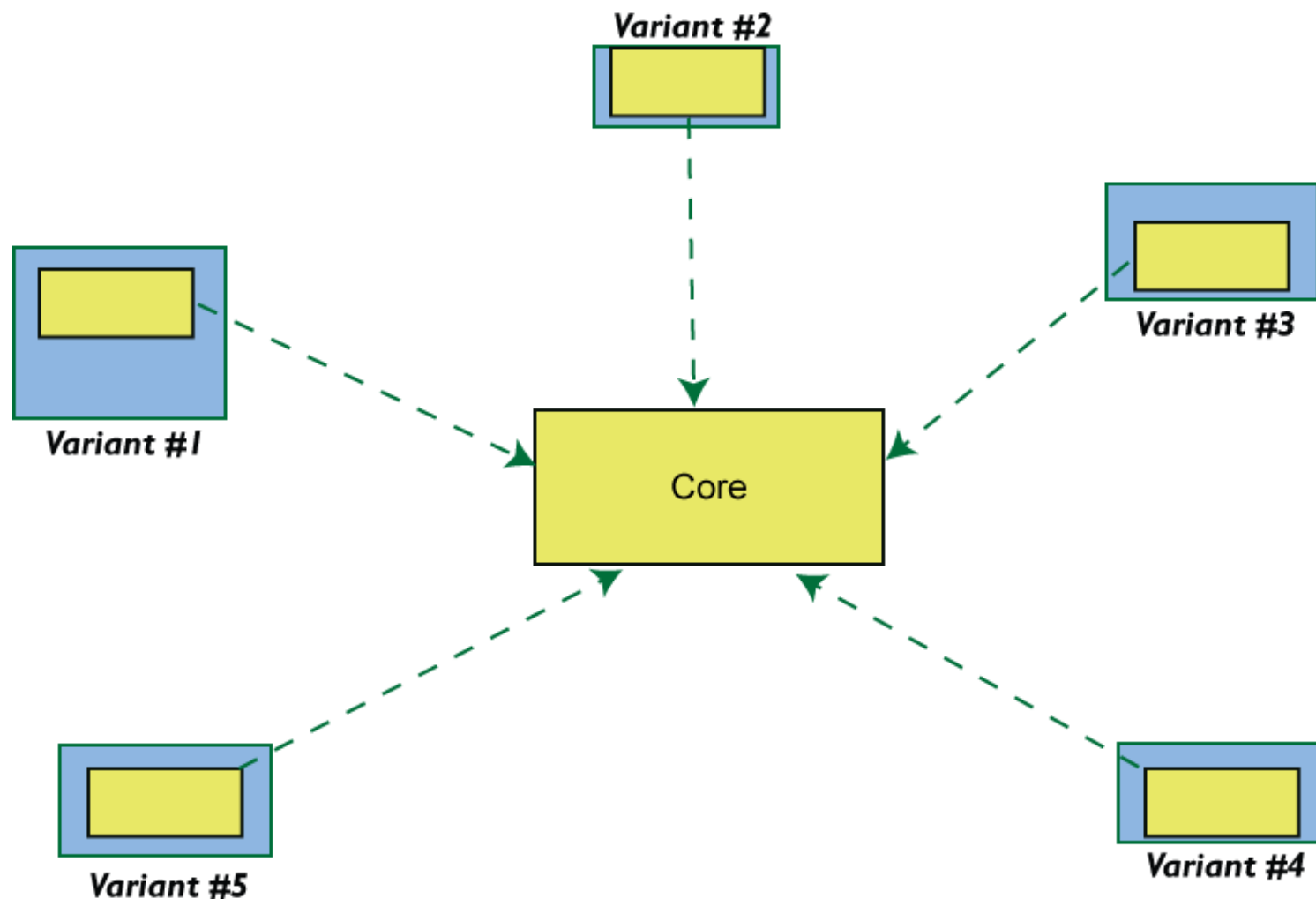


- Rule based systems can be very different
 - different rule semantics (based on various type of model theories, on proof systems, etc)
 - production rule systems, with procedural references, state transitions, etc
- Such universal exchange format is not feasible
- The idea is to define “cores” for a family of languages with “variants”

> RIF “core”: only partial interchange

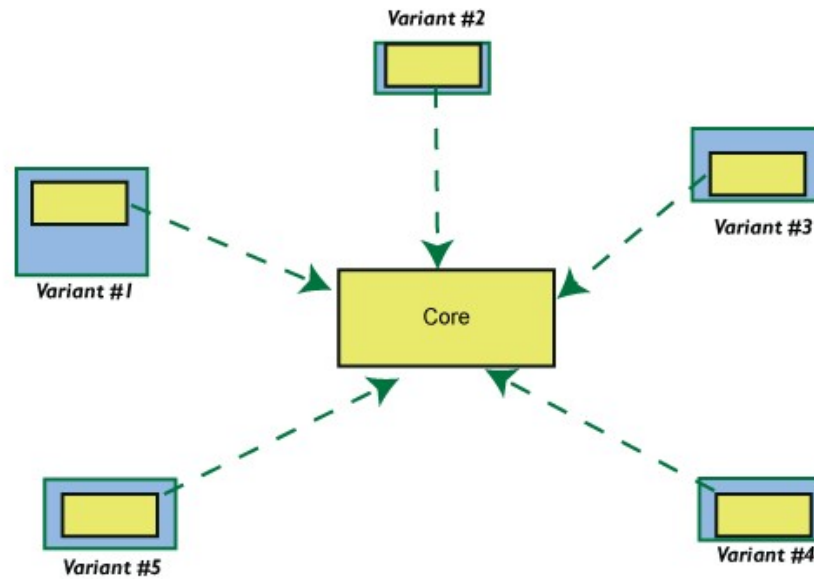


> RIF “variants”

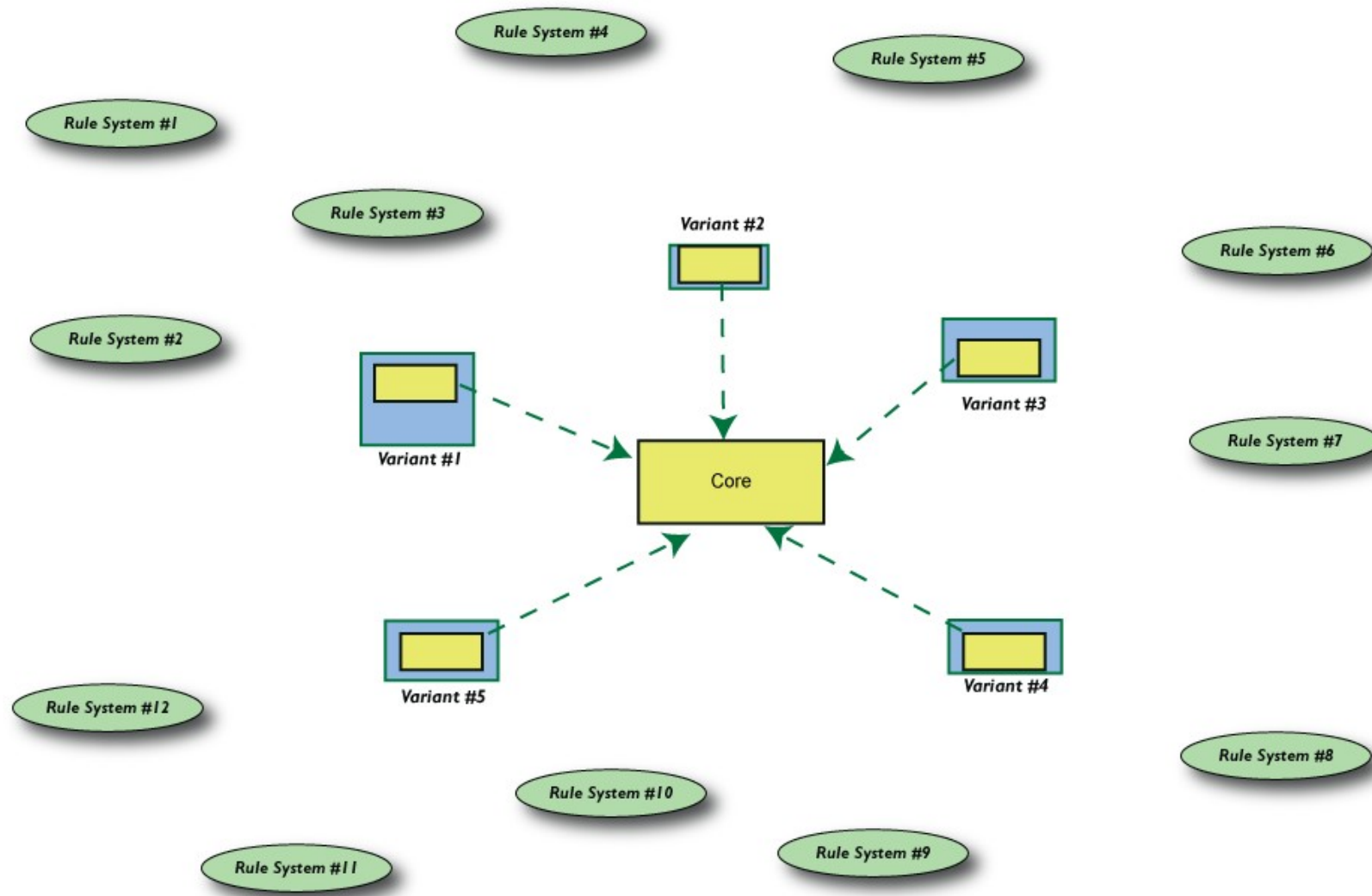


- Possible variants: F-logic, production rules, fuzzy logic systems, ...; none of these have been finalized yet

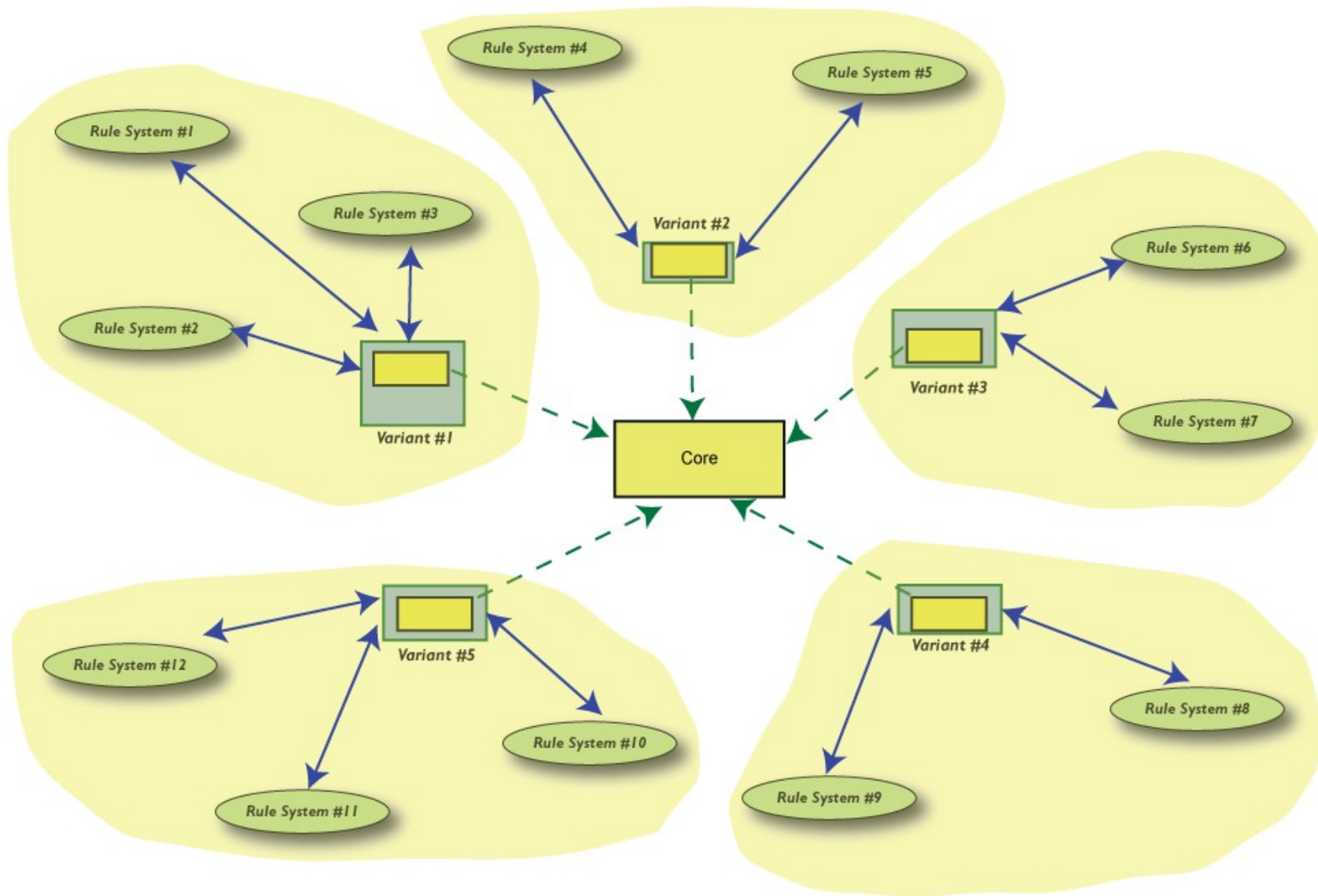
> Role of variants



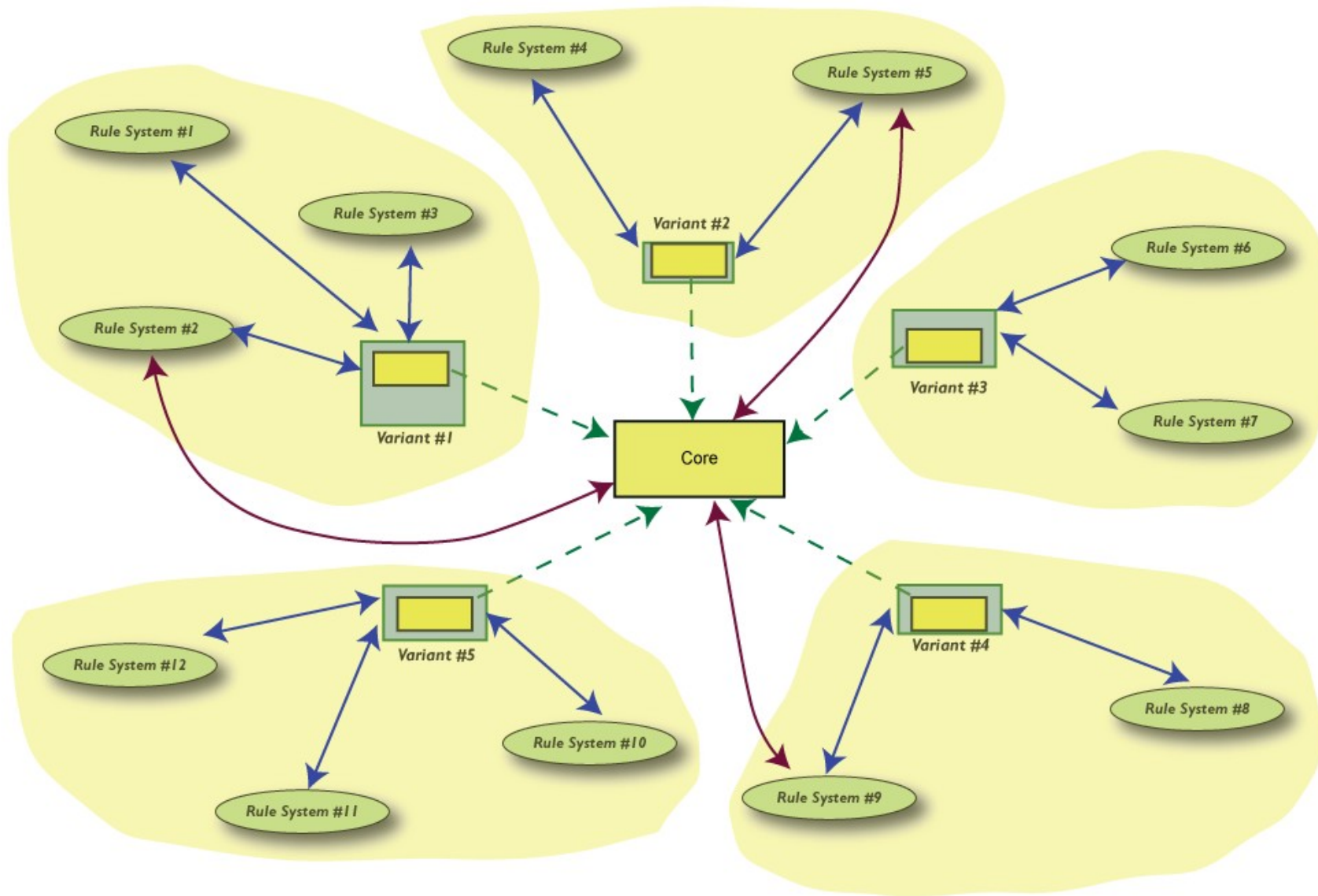
> Role of variants



> Role of variants



> Role of variants

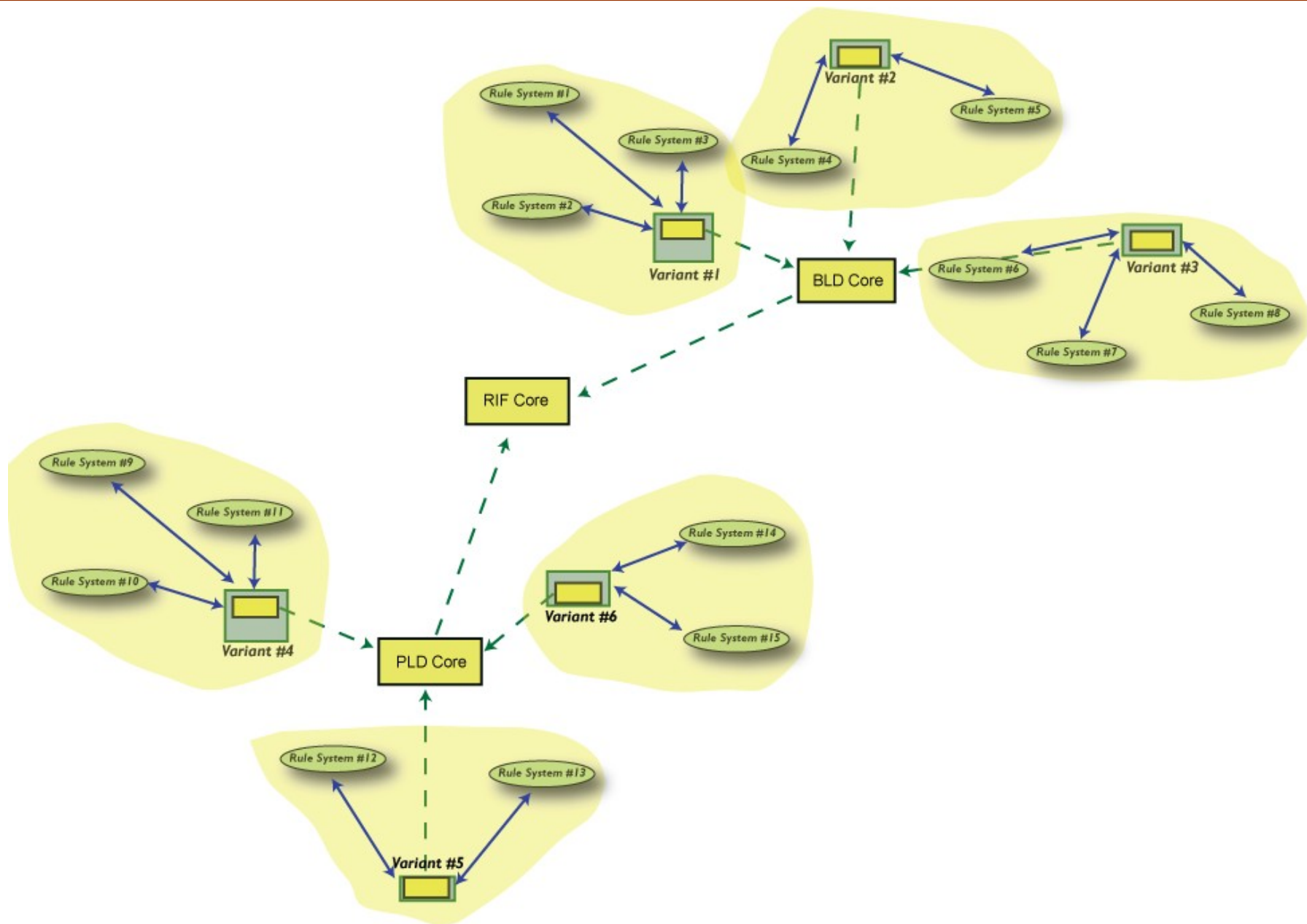


> However...



- Even this model does not completely work 🤖
- The gap between production rules and “traditional” logic systems seems to be large
- A hierarchy of cores may be necessary:
 - a *Basic Logic Dialect* and *Production Rule Dialect* as “cores” for families of languages
 - a common *RIF Core* binding these two

> Hierarchy of cores



> Current status



- There is a draft for the BLD
 - it defines a “positive Horn” language
 - it is a logic based general rule language
 - the language can be used
 - with or without RDF data and/or OWL
 - as a rule language or a rule interchange format
- The plan is to have BLD as a recommendation in 2008
- The work on the PLD Core has also begun



How do applications look like?

> 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)

> Applications can be very simple



- 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	5/7	Approach	3/7	Term+Reason	1/7	Max_Stage_Reached	1/7	Literature Links
29 Pain	<input checked="" type="checkbox"/>	7 Antibody	<input type="checkbox"/>	37 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"/>	
8 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 checked="" type="checkbox"/>	11 MARKET	<input checked="" type="checkbox"/>	19 HTS	<input type="checkbox"/>	
2 Infectives	<input checked="" type="checkbox"/>	21 SM_Inhibitor	<input checked="" type="checkbox"/>	11 REORG	<input type="checkbox"/>	11 Phase I	<input type="checkbox"/>	
1 Urogenitals	<input checked="" type="checkbox"/>			10 TOXIC	<input type="checkbox"/>	13 Phase III	<input type="checkbox"/>	
						41 Screening	<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	5/7	Approach	3/7	Term+Reason	1/7	Max_Stage_Reached	1/7	Literature Links
29 Pain	<input checked="" type="checkbox"/>	7 Antibody	<input type="checkbox"/>	37 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"/>	
8 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 checked="" type="checkbox"/>	11 MARKET	<input checked="" type="checkbox"/>	19 HTS	<input type="checkbox"/>	
2 Infectives	<input checked="" type="checkbox"/>	21 SM_Inhibitor	<input checked="" type="checkbox"/>	11 REORG	<input type="checkbox"/>	11 Phase I	<input type="checkbox"/>	
1 Urogenitals	<input checked="" type="checkbox"/>			10 TOXIC	<input type="checkbox"/>	13 Phase III	<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	UT_Alt	UT_2007	UT_Neoch	IMA	GED	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 (Galactose protein 83)(VLA-3) (CD49C)	SM_Antagonist	2006-Q3	EFFICACY	Candidate	P. Person						1		SW-029782
Sexual Health	Erectile Dysfunction	Leukotriene C4 synthase	SM_Agonist	2006-Q3	EFFICACY	Candidate	M. Manager				1	1			SW-029638
Sexual Health	Erectile Dysfunction	transcription elongation factor A (SII)-like 4	SM_Inhibitor	2005-Q2	EFFICACY	Candidate	P. Person								SW-029926
Infectives	HIV	Putative four repeat ion channel (Jx)	SM_Inhibitor	2006-Q2	EFFICACY	Candidate	L. Leader								SW-029994
Infectives	HIV	Voltage-gated potassium channel protein Kv1.2 (Kv)	SM_Agonist	2007-Q1	EFFICACY	Candidate	A. Scientist						1		SW-029653
Urogenitals	Incontinence	Human Rfx binding motif (RBM) gene, partial cds.	SM_Agonist	2007-Q3	EFFICACY	Candidate	L. Leader						1		SW-029684
Pain	Migraine	Monocarboxylate transporter homologue 2 (SLC16A4) (SLC16A4)	SM_Inhibitor	2007-Q3	EFFICACY	Candidate	L. Leader		18						SW-030085

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

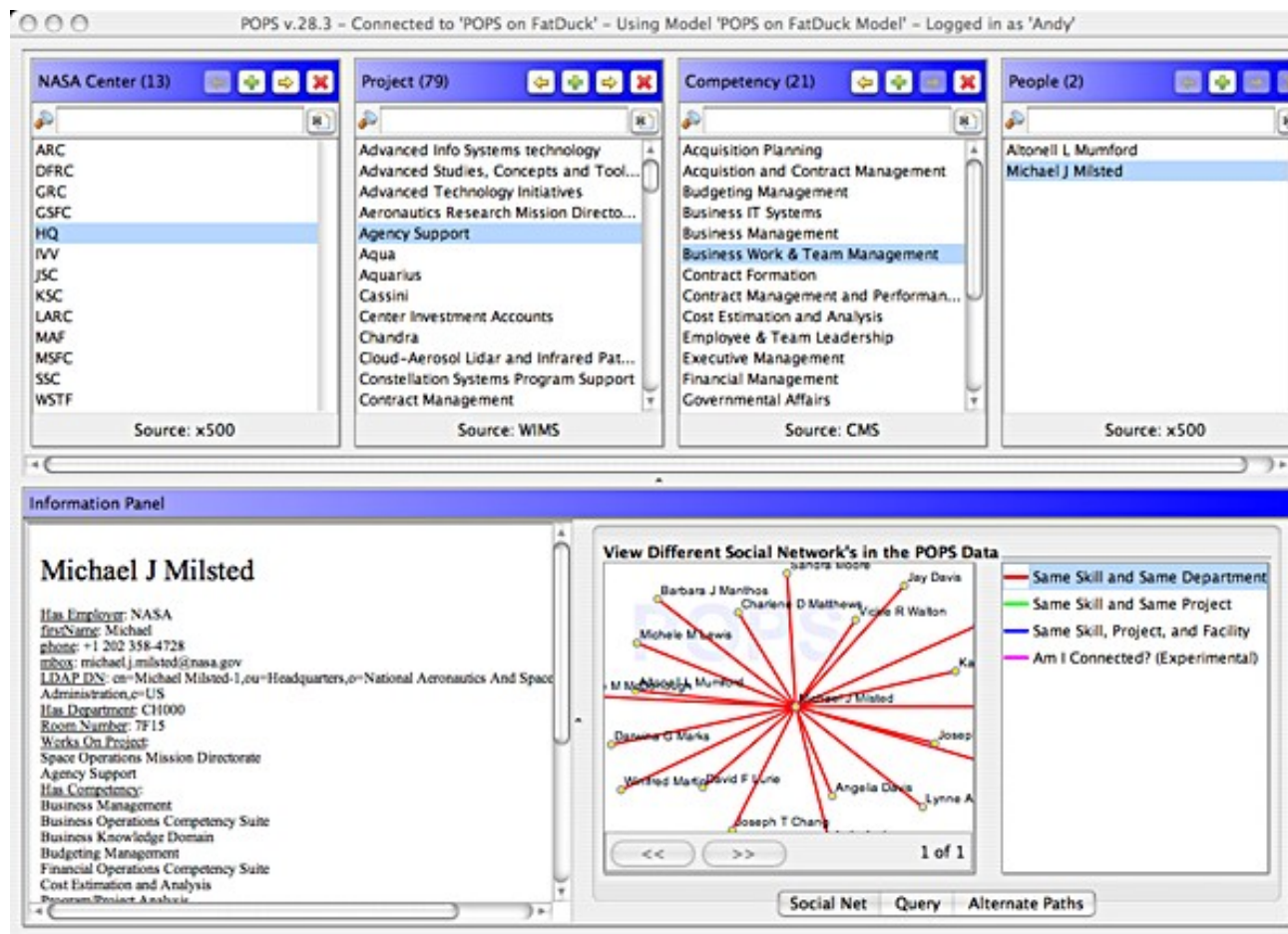
- Form based query interface for end users

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

> Find the right experts at NASA



- Expertise locator for nearly 20,000 NASA civil servants using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services...

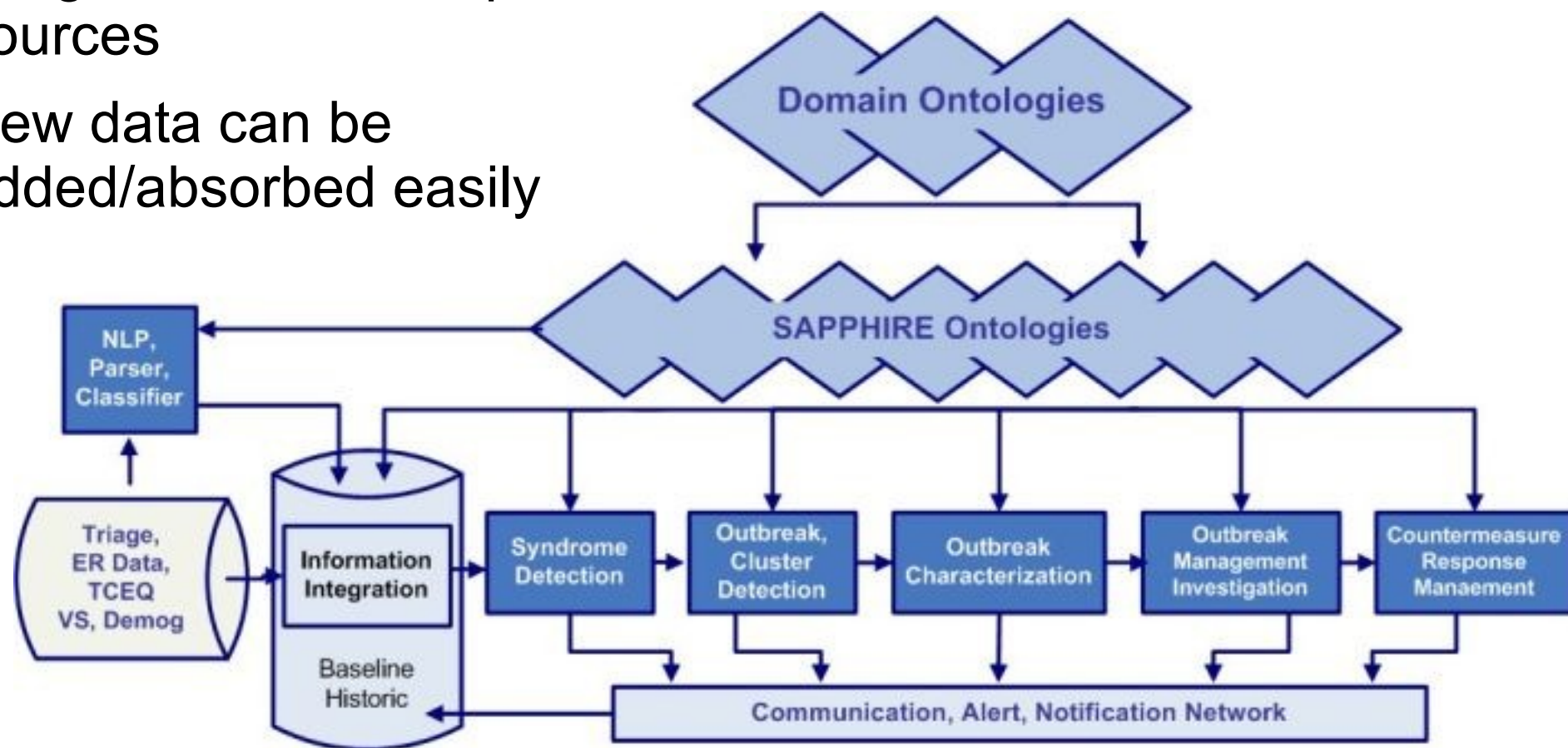


Courtesy of Kendall Clark, Clark & Parsia, LLC

> Public health surveillance (Sapphire)



- Integrated biosurveillance system (biohazards, bioterrorism, disease control, etc)
- Integrates from multiple data sources
- New data can be added/absorbed easily



Courtesy of Parsa Mirhaji, School of Health Information Sciences, University of Texas (SWEO Case Study)

> 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 displays the AKSIO search interface. At the top, there's a search bar with the text "leak in barrier elements" and a search button. Below the search bar, the interface is divided into two main sections: "Search filters" and "Results 1 - 7 of 7".

Search filters:

- discipline:** (dropdown menu)
- operation:** (dropdown menu)
- equipment:** (dropdown menu)
- state:** (checkbox list)
 - ☒ 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:** (dropdown menu)
- wellbore_id_ref:** (dropdown menu)
- field_id:** (checkbox list)
 - ☒ EXPLORATION (1)
 - ☒ GULLFAKS (1)
 - ☒ GULLFAKS SØR (1)
 - ☒ HEIDRUN (1)
 - ☒ HULDRA (1)
 - ☒ MIDGARD (2)
 - ☒ RIMFAKS (1)
 - ☒ SNORRE (1)
 - ☒ VISUND (2)

Results 1 - 7 of 7:

- 1. Top plug/20" EZSV**
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 Forerunner Deep set tubing plug Deep set tubing plug Liner top packer Liner top packer Mechanical tubular plugs Well Integrity Problem Well Integrity Problem
55% [annotate](#) [comment](#)
- 2. 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 M 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 Well test string Well test string components
Erosion Erosion Lack Of Maintenance Lack Of Maintenance Leak in barrier elements Leak in barrier elements Well Integrity Problem Well Integrity Problem
50% [annotate](#) [comment](#)
- 3. 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 fo...
HEIDRUN NO 6507/7-A-20 WIREL Høli Trond OTHER
2007-06-12T07:05:58Z
Cementing Network Sementeringsnettverk Technical Sidetrack Tekniske Sidesteg Bronnintegritet Well Integrity
Snubbing safety head Snubbing safety head USD none return valve USD none return valve
Corrosion Corrosion Erosion Erosion 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 Well Integrity Problem
47% [annotate](#) [comment](#)
- 4. 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 36/2-A-6 8 1/2" Rodvelt Khut 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 Well Integrity Problem
39% [annotate](#) [comment](#)

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

- 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](#))

> Help in choosing the right drug regimen



- Help in finding the best drug regimen for a specific case
 - find the best trade-off for a patient
- Integrate data from various sources (patients, physicians, Pharma, researchers, ontologies, etc)
- Data (eg, regulation, drugs) change often, but the tool is much more resistant against change

Severity	Adverse Drug Effect	v2	Current Regimen	Safety Optimized Profiles
ADE	Moderate Muscle Weakness (Myasthenia)	✓	1	2
ADE	Minor Excessive Sweating (Diaphoresis)	✓	1	2
ADE	Moderate Heart Throbbing or Pounding (Palpitations)	✓	1	2
ADE	Moderate Hives (Urticaria)	✓	1	2
ADE	Major Bladder Inflammation (Cystitis)	✓	1	2
ADE	Major Urinary Tract Infection	✓	1	2

row(s) 1 - 6 of 6

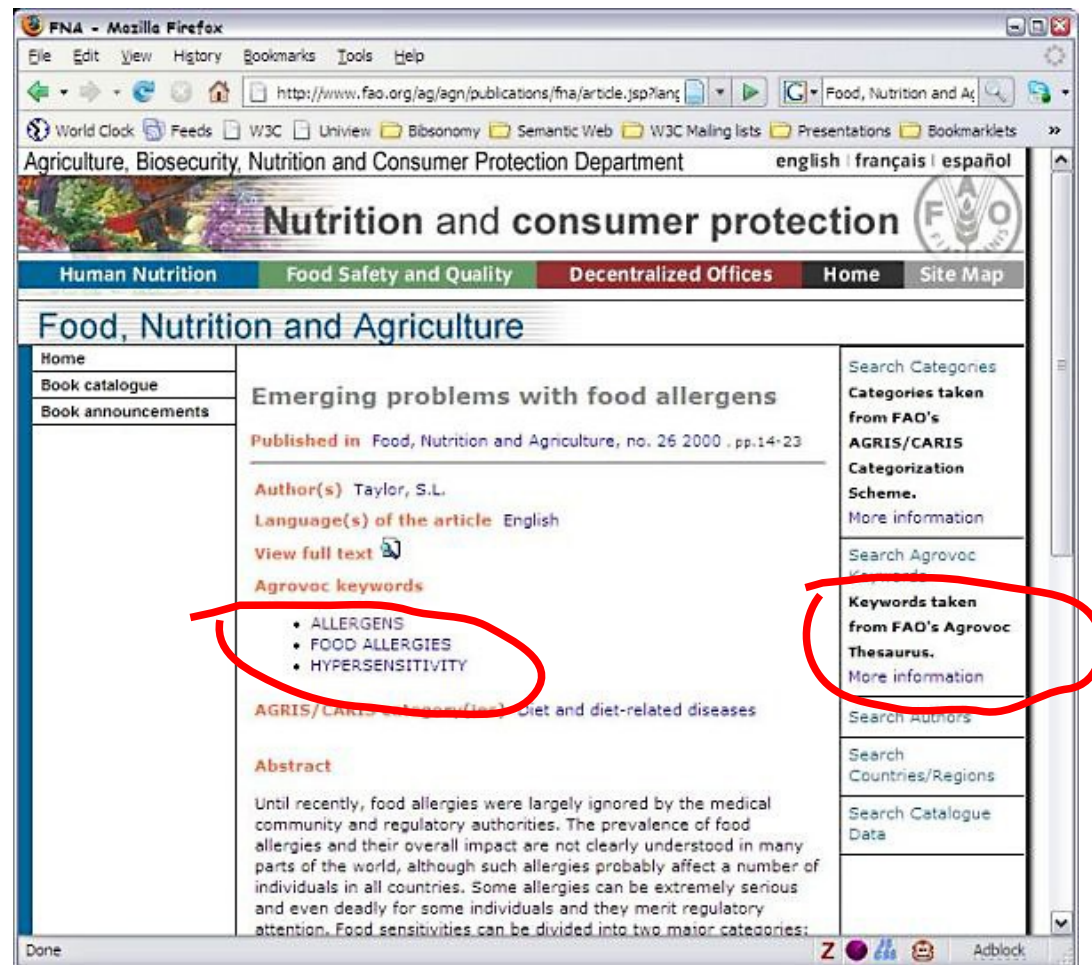
[Export to Excel](#)

Courtesy of Erick Von Schweber, PharmaSURVEYOR Inc., (SWEO Use Case)

> FAO Journal portal



- Improved search on journal content based on an agricultural ontology and thesaurus (AGROVOC)

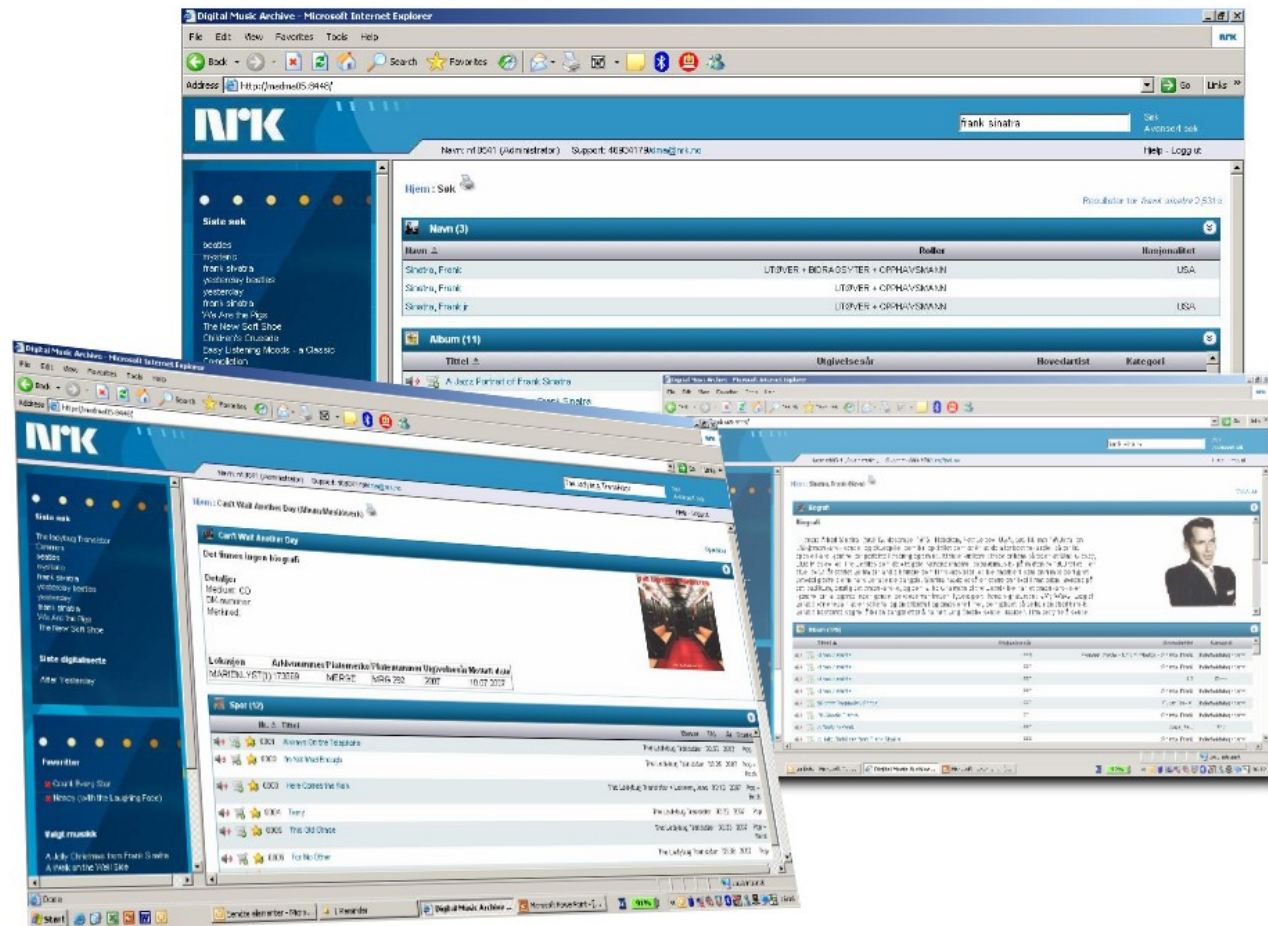


Courtesy of Gauri Salokhe, Margherita Sini, and Johannes Keizer, FAO, (SWEQ Case Study)

> Digital music asset portal at NRK



- Used by program production to find the right music in the archive for a specific show

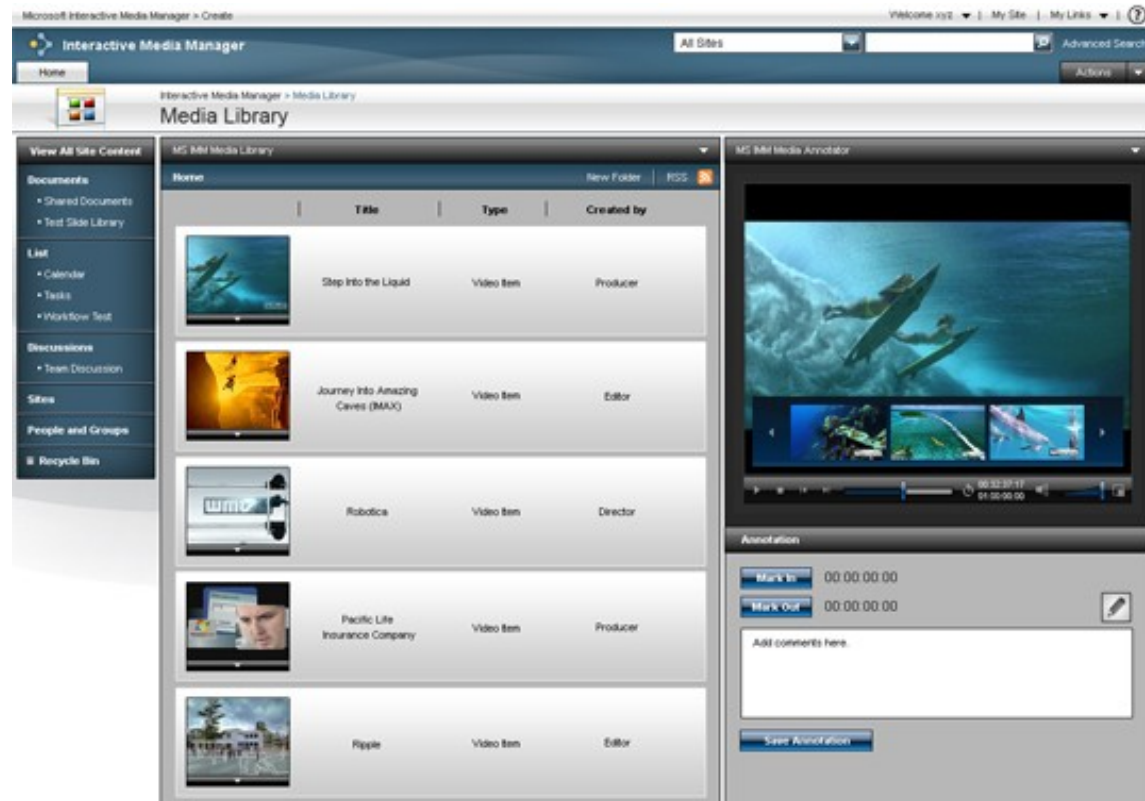


Courtesy of Robert Engels, ESIS, and Jon Roar Tønnesen, NRK (SWEQ Case Study)

> Microsoft Vista's Interactive Media Manager



- Uses an RDF/SPARQL/OWL based metadata framework
 - eg, for a better control over relationships among media assets and categories
- Custom OWL ontologies can be created and imported



-
- The screenshot displays the Lilly Science Grid application. On the left, a search bar contains 'TNF', and the 'Search Options' section is checked for 'Match case' and 'Exact match'. Below, the 'Ontology Tree' shows a hierarchical view of 'TNF' associated with various diseases and pathways. The main area features a 'Star Tree' visualization, a radial network diagram where 'Tumour necrosis factor (TNF)' is the central node, connected to numerous peripheral nodes representing specific drugs (e.g., 'an', 'Can', 'Sr', 'Gr', 'Uc', 'Boehr', 'Fulcrum', 'ynr', 'B', 'tr', 'Hadasi', 'Encysa', 'Pharm', 'UCB', 'Johr', 'Ta', 'Anar', 'Uc', 'Leuk', 'Biok', 'Cy', 'Unach', 'Yeda', 'Asthma', 'Acne', 'Arthritis, rheum', 'Crohn's disease', 'Ischaemia, gener', 'Inflammation, ocular', 'Sepsis', 'Infection', 'Inflammation', 'Cancer, gener', 'Arthritis, gener', 'Inflammatory bow', 'Cancer, pancreat', 'Arthritis, psori', 'Psoriasis') and diseases. The bottom of the interface includes a toolbar with icons for navigation and a search bar, and the text 'Star Tree™ created with InSight VizServer™'.

Karl Dubost and Ivan Herman, The state of the Semantic Web

> Improved Search via Ontology: GoPubMed



- Improved search on top of pubmed.org
 - search results are ranked using ontologies
 - related terms are highlighted, usable for further search

> Radar Network's Twine

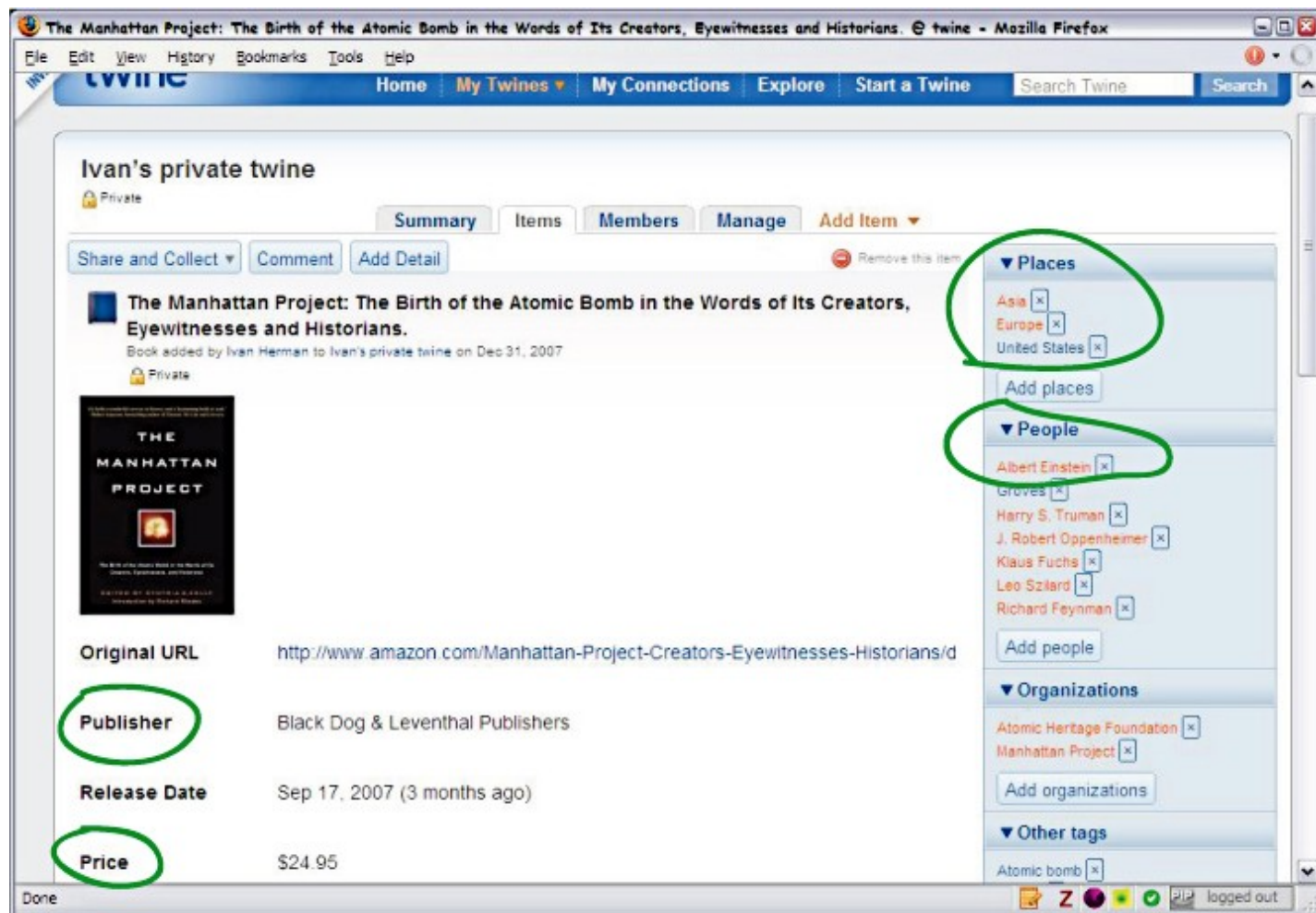


- “Social bookmarking on steroids”

- Item relationships are based on ontologies

- evolving over time
- possibly enriched by users

- Internals in RDF, will be available via APIs and SPARQL



> Other application areas come to the fore



- Content management
- Business intelligence
- Collaborative user interfaces
- Sensor-based services
- Linking virtual communities
- Grid infrastructure
- Multimedia data management
- Etc

> Thank you for your attention!



- These slides are publicly available on:

<http://www.w3.org/2008/Talks/0307-Tokyo-IH/>

- There is also a collection of use cases at:

<http://www.w3.org/2001/sw/sweo/public/UseCases/>

State of the Semantic Web

Karl Dubost and Ivan Herman, W3C

INTAP Semantic Web Conference, Tokyo, Japan, March 7, 2008

This is just a generic slide set. Should be adapted, reviewed, possibly with slides removed, for a specific event. Rule of thumb: on the average, a slide is a minute...

> Significant buzz...



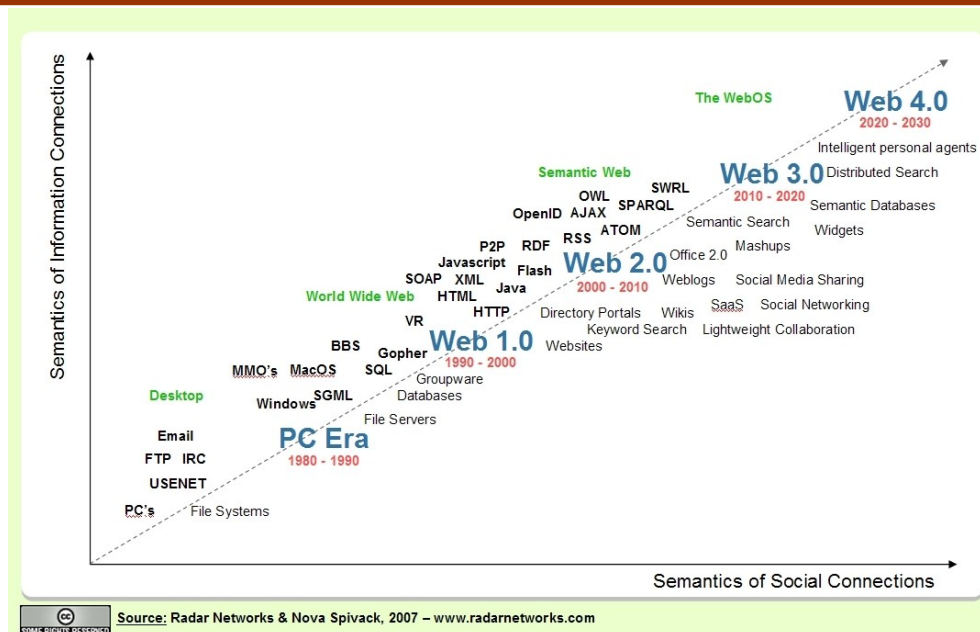
- There is quite a buzz around “Semantics”, “Semantic Technologies”, “Semantic Web”, “Web 3.0”, “Data Web”, etc, these days
- New applications, companies, tools, etc, come to the fore frequently
- It is, of course, not always clear what these terms all mean:
 - “Semantic Web” is a way to specify data and data relationships; it is also a collection of specific technologies (RDF, OWL, GRDDL, SPARQL, ...)
 - “Semantic Technologies”, “Web 3.0” often mean more, including intelligent agents, usage of complex logical procedures, etc

> Significant buzz... (cont.)



- Predicting the exact evolution in terms of Web 3.0, Web 4.0, etc, is a bit as looking into a crystal ball
- But the *Semantic Web technologies are already here, are used and deployed*
- They are at the basis of further evolution

> A vision on the evolution...



(this Web 3.0 is not identical to the "journalistic" Web3.0; merely timing)

This Web 3.0 is not the 'usual' Web 3.0. It is simply an evolutionary, well, versioning step, whereas, often, W3b 3.0 has an emphasis on the role of Artificial intelligence...

> The 2007 Gartner predictions



During the next 10 years, Web-based technologies will improve the ability to embed semantic structures [...] it will occur in multiple evolutionary steps...

By 2017, we expect the vision of the Semantic Web [...] to coalesce [...] and the **majority of Web pages are decorated with some form of semantic hypertext.**

By **2012, 80% of public Web sites will use some level of semantic hypertext to create SW documents [...]** 15% of public Web sites will use more extensive Semantic Web-based ontologies to create semantic databases

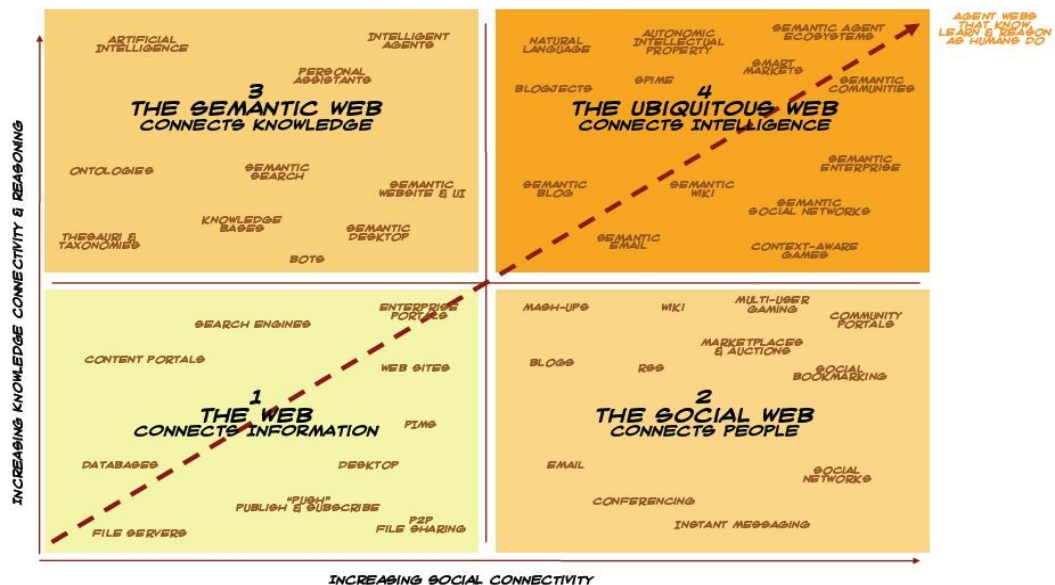
(note: "semantic hypertext" refers to, eg, RDFa, microformats with possible GRDDL, etc.)

Source: "Finding and Exploiting Value in Semantic Web Technologies on the Web", Gartner Research Report, May 2007

> Another longer term vision...



(from the "Semantic Wave 2008" report, from Project10X)



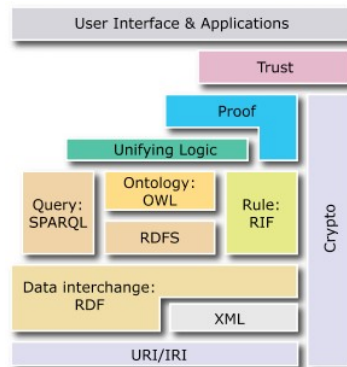
Courtesy of Mills Davis, Project10X; source: Nova Spivack, Radar Networks and John Breslin, DERI

The W3C's terminology is more to say that the SW 'connects data' rather than the (much more vague) term of connecting knowledge, but that is a minor issue. The upper right hand corner is certainly one grand vision for these analysts.

> Let us keep to the Semantic Web for now...



- In what follows we will restrict ourselves to the Semantic Web
 - a way to specify data and data relationships
 - allows data to be shared and reused across application, enterprise, and community boundaries
 - a collection of fundamental technologies (RDF/S, OWL, GRDDL, SPARQL, ...)



> The “corporate” landscape is moving



- Major companies offer (or will offer) Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, GE, Northrop Gruman, Altova, Microsoft, Dow Jones, ...
- Others are using it (or consider using it) as part of their own operations: Novartis, Boeing, Pfizer, Telefónica, ...
- Some of the names of active participants in W3C SW related groups: ILOG, HP, Agfa, SRI International, Fair Isaac Corp., Oracle, Boeing, IBM, Chevron, Siemens, Nokia, Pfizer, Sun, Eli Lilly, ...

> Some SW Tools (not and exhaustive list!)



• Triple Stores

- RDFStore, AllegroGraph, Tucana
- RDF Gateway, Mulgara, SPASQL
- Jena's SDB, D2R Server, SOR
- Virtuoso, Oracle11g
- Sesame, OWLIM, Tallis Platform
- ...

• Reasoners

- Pellet, RacerPro, KAON2, FaCT++
- Ontobroker, Ontotext
- SHER, Oracle 11g, AllegroGraph
- ...

• Converters

- flickurl, TopBraid Composer
- GRDDL, Triplr, jpeg2rdf
- ...

• Search Engines

- Falcon, Sindice, Swoogle
- ...

• Middleware

- IODT, Open Anzo, DartGrid
- Ontology Works, Ontoprise
- Profium Semantic Information Router
- Software AG's EII
- Thetus Publisher, Asio, SDS
- ...

• Semantic Web Browsers

- Disco, Tabulator, Zitgist, OpenLink Viewer
- ...

• Development Tools

- SemanticWorks, Protégé
- Jena, Redland, RDFLib, RAP
- Sesame, SWI-Prolog
- TopBraid Composer
- DOME
- ...

• Semantic Wiki systems

- Semantic Media Wiki, Platypus
- Visual knowledge

Inspired by "Enterprise Semantic Web in Practice", Jeff Pollock, Oracle. See also [W3C's Wiki Site](#).

Not an exhaustive list of tools. Some of the tools are open source (eg, Jena), some of them are products (Ontotext). Some of them are from big, established companies (Oracle), some of them are from smaller, specialized companies (AllegroGraph from Franc Inc), etc. *It is the usual picture of the Web industry*; in this sense, nothing special any more...

> Some SW tools (cont.)



- Significant speed, store capacity, etc, improvements are reported every day
- 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!
- We still need more “middleware” tools to properly *combine* what is already available...
- *Anybody can start developing RDF-based applications today*

The last point is important. Some years ago the problem was that application developers had to start from scratch because (almost) only the specifications were around plus some initial, mostly not-well-tested open source project results (or academic work output). Since about 2 years (rough estimate) this is not true any more.



Let us look at the technical state of the SW first

> Querying RDF: SPARQL



- Querying RDF graphs is essential (can you imagine Relational Databases without SQL?)
- SPARQL is
 - a query language based on graph patterns
 - a protocol layer to use SPARQL over, eg, HTTP
 - an XML return format for the query results
- Is a W3C Standard (since January 2008)
- Numerous implementations are already available (eg, built in triple stores)

The fact that SPARQL is not only a query language, but a full protocol over the Web is important to emphasize. *This* makes it deployable on the Web.

> Some new technologies at W3C



- SPARQL
- GRDDL
- RDFa
- SKOS
- OWL 1.1
- RIF (Rules)

> SPARQL (cont.)



- There are also SPARQL “endpoints” services on the Web:
 - send a query and a reference to data over HTTP GET, receive the result in XML or JSON
 - big datasets often offer “SPARQL endpoints” to query local data
 - *applications may not need any direct RDF programming any more, just use a SPARQL processor*
- SPARQL can also be used to construct graphs!

“service” means that these are running SPARQL processors that people can simply use by sending RDF reference data URI-s and the query, and they do the query for you. For some of these public services the RDF data can be anywhere on the web, not necessarily on the same site. I.e., these services make it possible to query RDF data anywhere in the world. Of course, these services usually have limitations in size, so one cannot do very serious applications, but it is good for simpler ones. Also: it is very easy to install some of these services locally on one's own machine. Typical example: Jena's sparql service, or Virtuoso's free version.

The last bulleted item is important: for many applications, one can rely on the query language only and it is not necessary to know about the details of how RDF environment store and manage triples, what programming language they use, etc. SPARQL makes it much easier to develop applications that mash up RDF data.

The last point is showed more in details in the next few slides. It is an essential, but not very well known feature of SPARQL, good to show for an already RDF aware audience

> The power of CONSTRUCT



```
CONSTRUCT {  
  <http://dbpedia.org/resource/Amitav_Ghosh> ?p1 ?o1.  
  ?s2 ?p2 <http://dbpedia.org/resource/Amitav_Ghosh>.  
}  
WHERE {  
  <http://dbpedia.org/resource/Amitav_Ghosh> ?p1 ?o1.  
  ?s2 ?p2 <http://dbpedia.org/resource/Amitav_Ghosh>.  
}
```

- SPARQL endpoint
- returns RDF/XML



```
SELECT  
FROM <http://dbpedia.org/sparql/?query=CONSTRUCT+%7B++>  
WHERE {  
  ?author of dbpedia:author res:Amitav_Ghosh.  
  res:Amitav_Ghosh dbpedia:reference ?homepage;  
                  rdf:type      ?type;  
                  foaf:name     ?foaf_name.  
  FILTER regex(str(?type), "foaf")  
}
```

- Data reused in a query elsewhere...

This means: one can have a URI that refers to a specific graph as returned by a SPARQL query somewhere on the WEB. This URI can then be incorporated into the query of *another* SPARQL processor. Another way of putting it is that SPARQL queries can be, sort of, “chained” together.

> 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...

Note: W3C is in the process of setting up an appropriate mechanism to gather feedbacks and will, probably, start work for a “SPARQL2” (*provisional* name) within 1-2 years. Undecided, though.

> Bridge to relational databases



- Most of the data on the Web are stored in relational databases
 - “RDFying” them is an impossible
 - relational databases are here to stay...
- “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 languages/approaches are being used
 - a number of systems can now be used as relational database as well as triple stores (eg, Oracle, OpenLink, ...)
- Work for a survey on mapping techniques benchmarks may start soon at W3C
- SPARQL is becoming the tool of choice to query the data
 - ie, “SPARQL endpoints” are defined to query the databases

On the work coming up: we are in discussion for two XG-s on those issues. It is not yet 100% sure they will happen, there is currently a bigger probability for the mapping one to come and the other is still unclear. Of course, members interested in this work would be welcome!

> How to get RDF data?



- Of course, one could create RDF data manually...
- ... but that is unrealistic on a large scale
- Goal is to generate RDF data automatically when possible and “fill in” by hand only when necessary
- We have already seen the work relating to “traditional” databases
- But there are also other types of data out there, too...

> Data may be extracted (a.k.a. “scraped”)



- Different tools, services, etc, come around:
 - get RDF data associated with images, for example:
 - service to [get RDF from flickr images](#)
 - service to [get RDF from XMP](#)
 - scripts to convert spreadsheets to RDF
 - etc
- Many of these tools are still individual “hacks”, but show a general tendency
- Hopefully more tools will emerge

> Getting structured data to RDF: GRDDL



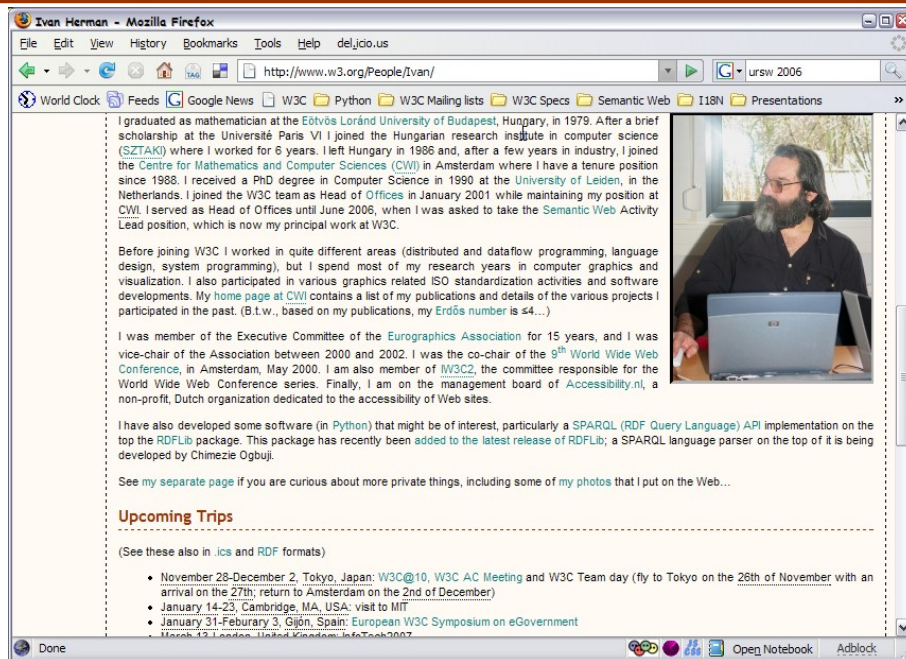
- GRDDL is a way to access structured data in XML/XHTML and turn it into RDF:
 - defines XML attributes to bind a suitable script to transform (part of) the data into RDF
 - script is usually XSLT but not necessarily
 - has a variant for XHTML
 - a “GRDDL Processor” runs the script and produces RDF on-the-fly
- A way to access existing structured data and “bring” it to RDF
 - eg, a possible link to microformats
 - exposing data from large XML use bases, like XBRL

> Getting structured data to RDF: RDFa



- RDFa extends XHTML with a set of attributes to include structured data into XHTML
- Makes it easy to “bring” existing RDF vocabularies into XHTML
- Uses namespaces for an easy mix of terminologies
- It can also be used with GRDDL
 - but: no need to implement a separate transformation per vocabulary

> GRDDL & RDFa: Ivan' home page...



> ...marked up with GRDDL headers...



```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-
<html xmlns="http://www.w3.org/1999/xhtml" lang="en"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:con="http://www.w3.org/2000/10/swap/pim/contact#"
  xmlns:wot="http://xmlns.com/wot/0.1/"
  xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#"
  xmlns:cal="http://www.w3.org/2002/12/cal/ical#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:bio="http://purl.org/vocab/bio/0.1/"
  <head profile="http://www.w3.org/2003/g/data-view">
  <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
  <title>Ivan Herman</title>
  <link rel="stylesheet" type="text/css" href="StyleSheets/Private.css" />
  <link rel="meta" type="application/rdf+xml" title="FOAF" href="http://www.ivan-herman.net/foaf.
  <script src="/2004/08/TalkFiles/popup.js" type="text/javascript"><!-- The popup script --></scri
  <link rel="transformation" href="http://www.w3.org/2002/12/cal/glean-hcal.xsl" />
  <link rel="transformation" href="http://www.sop.inria.fr/acacia/soft/RDfa2RDFXML.xsl"/>
  </head>
  <body xml:lang="en" lang="en">
  <div class="left">
    <a href="/"></a><h1>
  </div>

  <div class="right" id="me" role="foaf:Person">
    <link rel="owl:sameAs" href="http://www4.wiwiiss.fu-berlin.de/dblp/resource/person/103481"/>
    <link rel="owl:sameAs" href="http://ivan-herman.net/Ivan_Herman"/>
    <link rel="owl:sameAs" href="http://ivan-herman.net/foaf.rdf#me"/>
    <h1 property="foaf:name">Ivan Herman</h1>
    Accessibility.nl</a>,
a non-profit, Dutch organization dedicated to the accessibility of Web sites. </p>

<p>I have also developed some software (in <a href="http://www.python.org">Python</a>) that might
a <a href="http://dev.w3.org/cvsweb/%7Echeckout%7E/2004/PythonLib-IH/Doc/sparqlDesc.html">SPARQL
on the top the <a href="http://rdflib.net/">RDFLib</a> package.
This package has recently been <a href="http://www.ivan-herman.net/WebLog/WorkRelated/SemanticWeb">
added to the latest release of RDFLib</a>; a SPARQL language parser on the top of it is being de
Chimezie Ogbuji.
</p>

See <a rel="foaf:homepage" href="http://www.ivan-herman.net">my separate page</a> if you are curious about
private things, including some of <a href="http://www.ivan-herman.net/Photos/">my photos</a> that I put

<h2><a id="trips" name="trips">Upcoming Trips</a></h2>

<p>(See these also in <a href="http://feeds.technorati.com/events/http%3A/www.w3.org/People/Ivan">
and <a href="http://www.w3.org/2000/06/webdata/xslt?xslfile=http%3A%2F%2Fwww.w3.org%2F20

<ul>

<li><span class="vevent" id="ac06"><abbr class="dtstart" title="2006-11-28">November 28<
<span class="location vcard"><abbr class="geo" title="35.670685;139.770813">Tokyo, Japan<
<span class="summary"><a href="http://www.w3.org/Member/Meeting/2006ac/November/" class="
W3C AC Meeting</a> and W3C Team day</span></span> (<span class="vevent" id="flyTok06"><s
<abbr class="dtstart" title="2006-11-26">26th of November</abbr> with an arrival on the<
27th</abbr></span>; <span class="vevent" id="flyTokAms06"><span class="summary">return
<abbr class="dtstart" title="2006-12-02">2nd of December</abbr></span></li>

<li class="vevent" id="mitjan"><abbr class="dtstart" title="2007-01-14">January 14</abbr>
<span class="location vcard"><abbr class="geo" title="42.365;-71.105">Cambridge, MA, USA
<span class="summary">visit to MIT</span></li>

<li class="summary" id="summary"><abbr class="dtstart" title="2007-01-31">January 31</abbr>
```

The microformat is not defined by W3C...

> ...yielding; ...



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xml:base="http://www.w3.org/People/Ivan/">
  <c:Vcalendar xmlns:r="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:ical=... >
    <c:component>
      <c:Vevent r:about="#ac06">
        <ical:summary>W3C@10, W3C AC Meeting and W3C Team day</ical:summary>
        <ical:dtstart>2006-11-28</ical:dtstart>
        <ical:dtend>2006-12-03</ical:dtend>
        <ical:url
          r:resource="http://www.w3.org/Member/Meeting/2006ac/November/">
        <loc:location xml:lang="en">Tokyo, Japan</location>
        <geo:geo r:parseType="Resource">
          <r:first>35.670685</r:first>
          <r:rest r:parseType="Resource">
            ...
          </r:rest>
        </geo:geo>
      </c:Vevent>
    </c:component>
  </c:Vcalendar>
</rdf:RDF>
```

> ...marked up with RDFa tags...



```
Source of: http://localhost:8001/People/Ivan/OverviewNewRDFa.php - Mozilla Firefox
File Edit View Help

<dt>Postal address:</dt>
<dd>C/o <a href="http://www.w3c.nl">W3C Benelux Office</a> at
    <a href="http://www.cwi.nl">Centre for Mathematics and Computer Sciences
    </a>
    <a href="http://www.google.com/maps?f=q&hl=en&q=kruislaan+413,+amsterdam,+the-netherlands&ie=UTF8&z=13&
href="http://www.google.com/maps?f=q&hl=en&q=kruislaan+413,+amsterdam,+the-netherlands&ie=UTF8&z=13&
413, P.O. Box 94079,
    1090 GB Amsterdam, The Netherlands</a></dd>
<dt>Phone numbers:</dt>
<dd>
    phone: <a rel="foaf:phone" href="tel:+31-20-5924163">+31-20-5924163</a><br />
    mobile phone: <a rel="foaf:phone" href="tel:+31-641044153">+31-641044153</a><br />
    fax: +31-20-5924312
</dd>
<dt>PGP/GPG:</dt>
<dd>My <a rel="wot:pubkeyAddress" href="http://www.ivan-herman.net/pgpkey.html">GnuPGP
key and signature</a> is
    available on-line.</dd>
<dt>FOAF:</dt>
<dd>You can either extract a short FOAF information from this page using an
    <a href="http://trippl.org/rdfa-rdf/http://www.w3.org/People/Ivan/">RDFa service</a>,
    of consult <a href="http://www.ivan-herman.net/foaf.rdf" rel="rdfs:seeAlso">my more
complete, public FOAF file</a>.</dd>
<dt>Misc:</dt>
<dd>I am often on the freenode #swig irc channel, usually using the nickname
    <span instanceof="foaf:OnlineChatAccount" rel="foaf:holdsAccount">
        <span rel="foaf:accountServiceHomepage"
resource="http://www.freenode.net/irc_servers.shtml"/>
        <span property="foaf:accountName">IvanHerman</span> or 'ivan'
    </span>
    <br/>I am also on <a href="http://www.facebook.com/p/Ivan_Herman/555168627"
rel="rdfs:seeAlso">Facebook</a>,
    <a href="http://www.dopplr.com/traveller/IvanHerman" rel="rdfs:seeAlso">Dopplr</a>,
    <a href="http://www.linkedin.com/profile?viewProfile=s&key=2552277"
rel="rdfs:seeAlso">LinkedIn</a>, and
    <a href="http://www.flickr.com/photos/ivan_herman/" rel="rdfs:seeAlso">Flickr</a>
    <br/>
    My URI (as a real person): <a
href="http://www.ivan-herman.net/Ivan_Herman"><code>http://www.ivan-herman.net/Ivan_Herman</code></a>
    </dd>
</dl>
```


> ... yielding



```
@prefix foaf: <http://xmlns.com/foaf/0.1/>
@prefix wot: <http://xmlns.com/wot/0.1/>
...
@base <http://www.w3.org/People/Ivan/>

<#me>
  foaf:phone <tel:+31-20-5924163>;
  foaf:phone <tel:+31-641044153>;
  wot:pubkeyAddress <http://www.ivan-herman.net/pgpkey.html>;
  rdfs:seeAlso <http://www.ivan-herman.net/foaf.rdf>;
  foaf:holdsAccount
    [ a foaf:OnlineChatAccount;
      foaf:accountServiceHomepage
        <http://www.freenode.net/irc_servers.html>;
      foaf:accountName "IvanHerman";
    ];
  rdfs:seeAlso <http://www.facebook.com/p/Ivan_Herman/555188824>;
  ...
```

> Such data can be SPARQL-ed

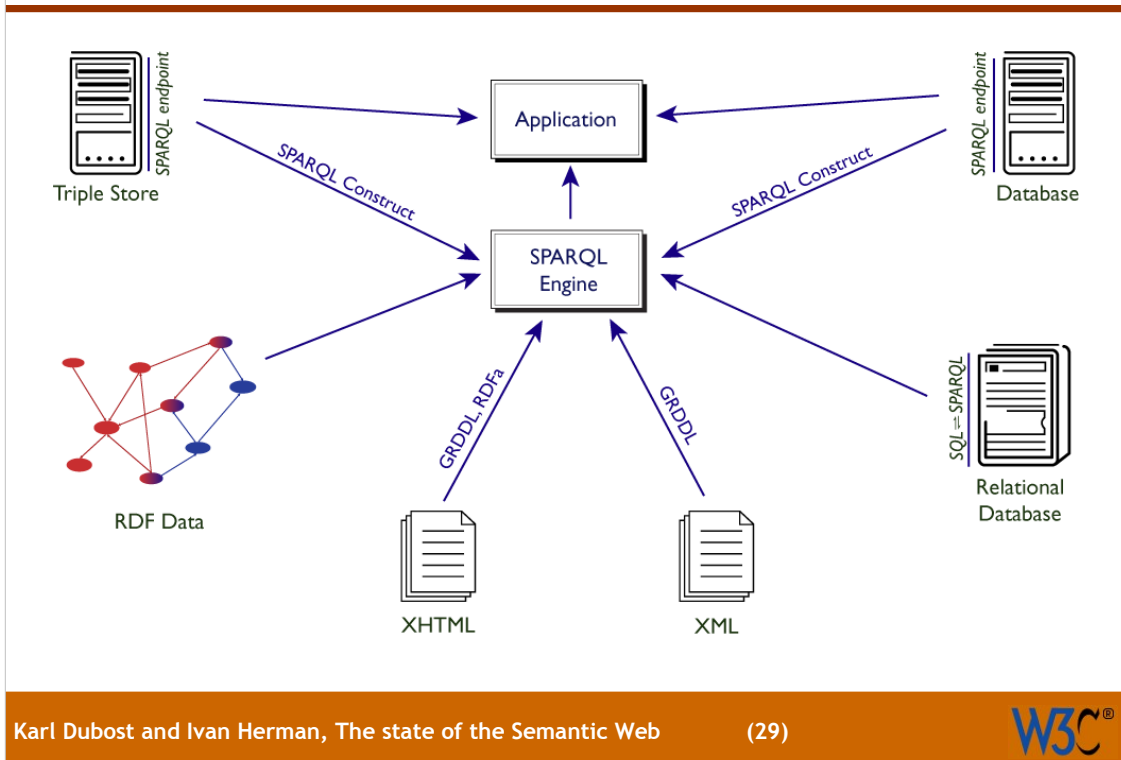


```
SELECT DISTINCT ?name ?home ?orgRole ?orgName ?orgHome
# Get RDFa from my home page:
FROM <http://www.w3.org/People/Ivan/>
# GRDDL-ing http://www.w3.org/Member/Mail:
FROM <http://www.w3.org/Member/Mail/>
WHERE {
  ?foafPerson    foaf:mbox ?mail;
                  foaf:homepage ?home.
  ?individual    contact:mailbox ?mail;
                  contact:fullName ?name.
  ?orgUnit ?orgRole ?individual;
              org:name ?orgName;
              contact:homePage ?orgHome.
}
```

Note that the SPARQL query:

- uses the same URI for the page and the RDF data (some processors, like Virtuoso or Tabulator) are capable of running the converters (well, Tabulator does not do it for RDFa yet)
- the query shows the data coming from different sources, (colour coded) with the ?mail term, sort of, 'binding' the data coming from different places. .Ie, the SPARQL query does the 'mash up' on the query level, regardless of the exact format the data is stored in...

> SPARQL as a unifying point!



This binds back to an earlier remark on SPARQL. For many applications, SPARQL is the only interface to the Semantic Web data, everything else is done under the hood via GRDDL/RDFa, other SPARQL endpoints to data, etc.

> Simple Knowledge Organization System



- Goal: representing and sharing classifications, glossaries, thesauri, etc, as developed in the “Print World”. For example:
 - [Dewey Decimal Classification](#), Art and Architecture Thesaurus, ACM classification of keywords and terms...
 - DMOZ categories (a.k.a. [Open Directory Project](#))
- The system must be simple to allow for a quick port of traditional data (done by non-experts in, say, Semantic Web)
- This is where SKOS comes in: define classes, properties, where those structures can be added

This is a very important spec in accessing to, eg, the (digital) library world, to various thesauri and taxonomies around the globe!

> Example: thesaurus

**Term**

Economic cooperation

Used For

Economic co-operation

Broader terms

Economic policy

Narrower terms

Economic integration, European economic cooperation, ...

Related terms

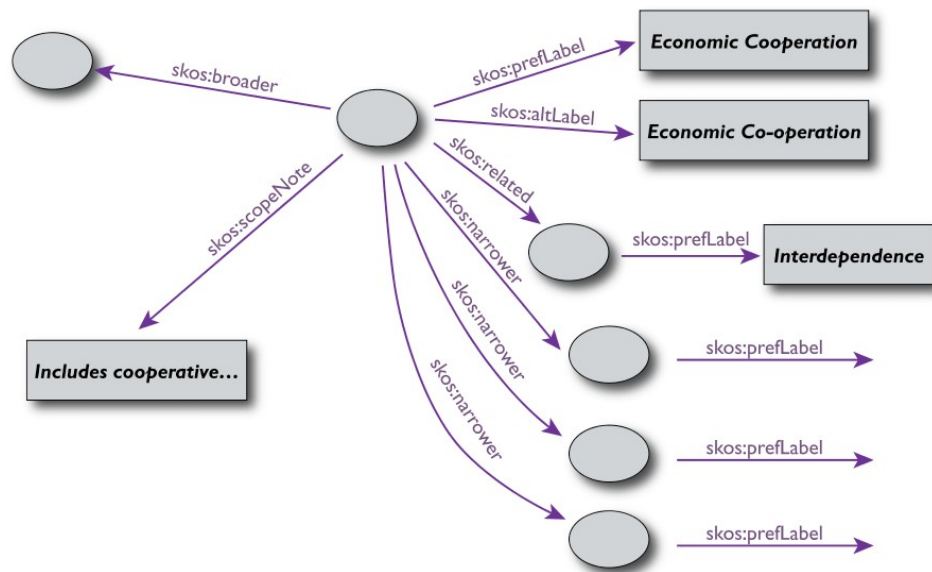
Interdependence

Scope Note

Includes cooperative measures in banking, trade, ...

(from the UK Archival Thesaurus)

> Example: thesaurus in SKOS



> SKOS and digital libraries



- SKOS plays an important role in “bridging” to digital libraries
- A huge community out there with its own traditions, style...
 - ... but huge amount of data to be “linked” to the Semantic Web!
- Major library metadata standards are being re-defined in terms of RDF (and SKOS),
 - eg, “Resource Description and Access” (RDA)
 - a major cataloging rule set for librarians
 - potentially, all major library catalogs around the globe could be translated into RDF and, eg, linked as an Open Linked Data...

> Ontologies



- Large ontologies are being developed (converted from other formats or defined in OWL). For example:
 - [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; or [UniProt](#) for protein sequence and annotation terminology and data
 - [BioPAX](#): for biological pathway data
 - [ISO 15926](#): “Integration of life-cycle data for process plants including oil and gas production facilities”

> OWL in applications



- An increasing number of applications rely on OWL (Pfizer, Nasa, Eli Lilly, Elsevier, FAO, ...)
 - see some more example at the end of the talk
- Not all use complex reasoning; in many cases a small fraction of OWL is used

> New OWL Working Group



- A new Working Group just started on the revision of OWL
- The goal of the group:
 1. add a few extensions to current OWL that are useful, and is known to be implementable
 - many things happened in research since 2004
 - features should (if possible) be valid both in the DL and OWL Full world
 2. define fragments, ie, “profiles” of OWL that are:
 - smaller, easier to implement and deploy
 - cover important application areas and are easily understandable to non-expert users

The work is based on the input of an “ad-hoc” group that looked at the issue in the past 1.5-2 years

> “OWL 1.1”: new proposed features



- “Qualified cardinality restrictions” (eg, “class instance must have two black cats”)
- Disjoint, reflexive, irreflexive properties; disjoint union of classes
- Property chains (eg, the uncle example: “if y is father x of y and y is brother of z, then z is uncle of x”)
- Own datatype constructs instead of complex XML Schema datatypes
 - eg, to express restrictions like number intervals easily

> “OWL 1.1”: new proposed features (cont)



- Metamodeling (a.k.a. “punning”): the same symbol may be used both as, e.g., a Class and an Instance, or for a datatype and an object property
 - this is not a problem in OWL Full, but is a significant restriction in OWL DL
 - in the DL there would still be some restrictions on how that can be used (eg, not all “natural” inferences can be drawn)

> “OWL 1.1”: small fragments



- For a number of applications RDFS is not enough, but even OWL Lite is too much (and too complex to implement)
- There is a need for (very) “light” versions of OWL: just a few extra possibilities added to RDFS
- Some can be as simple as having only (on top of RDFS):

equivalentClass
equivalentProperty
sameAs

inverseOf
TransitiveProperty
SymmetricProperty
FunctionalProperty
InverseFunctionalProperty

Worth noting: the small example is very close to OWLPrime, that Oracle implemented in their newest version (11g) that came out a few months ago

> “OWL 1.1”: small fragments (cont.)



- There are a number of proposals, papers, prototypes (and implementations!). Eg:
 - EL++, DLP: all DL dialects (e.g., EL++ is already in use by the health care community for medical ontologies)
 - pD*, OWLPrime: OWL Full dialects, that can be implemented with rule engines on top of, say, database engines
- It may be possible to create a (or more) dialect that may have both a DL and an OWL Full semantics (eg, OWLPrime~DLP)
- The Working Group will have to settle on the final list and structure

> Rules



- There is a long history of rule languages and rule-based systems
 - eg: logic programming (Prolog), production rules
- Lots of small and large rule systems (from mail filters to expert systems)
- Hundreds of niche markets

> Why rules on the Semantic Web?



- There are conditions that ontologies (ie, OWL) cannot express (or only with difficulties)
 - a well known examples is Horn rules: $(P1 \wedge P2 \wedge \dots) \rightarrow C$
- There are conditions that are complicated in rules and ontologies are better (eg, complex classification of terms)
- Simple rule engines might be easier to implement (eg, on top of database engines)
- A different way of thinking — people may feel more familiar in one or the other

> Things you may want to express



- An example:
 - “if two Persons have the same name and the same email, or the same name and the same home page, then they are identical”
- Something like (with an ad-hoc syntax):

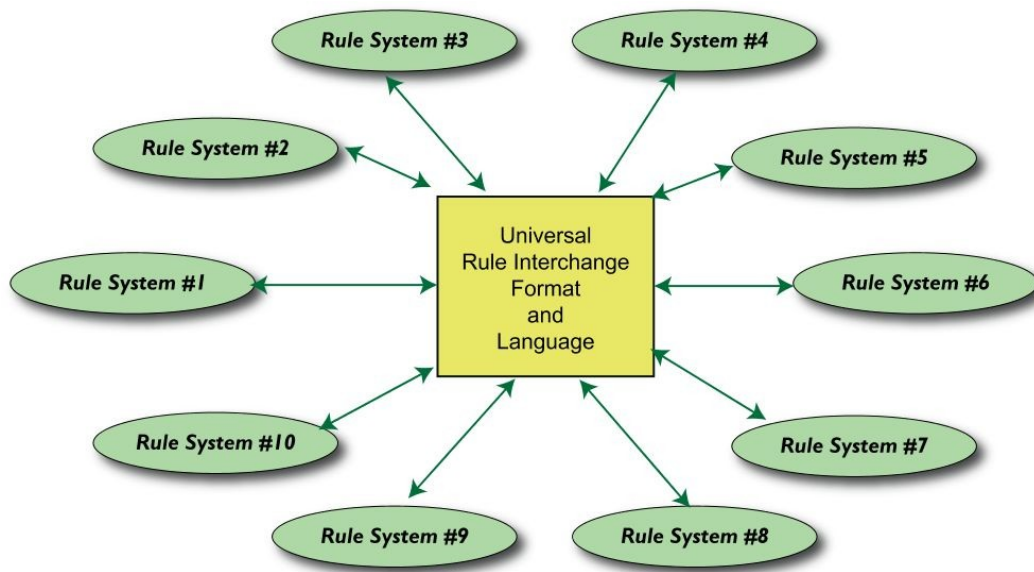
```
If { ?x rdf:type foaf:Person.  
    ?y rdf:type foaf:Person.  
    ?x foaf:name ?n.  
    ?x foaf:homepage ?h.  
    ?y foaf:name ?n.  
    ?y foaf:homepage ?h. }  
then { ?x = ?y }  
  
If { ?x rdf:type foaf:Person.  
    ?y rdf:type foaf:Person.  
    ?x foaf:name ?n.  
    ?x foaf:mailbox ?h.  
    ?y foaf:name ?n.  
    ?y foaf:mailbox ?m. }  
then { ?x = ?y }
```

> A new requirement: exchange of rules



- Applications may want to *exchange* their rules:
 - negotiate eBusiness contracts across platforms: supply vendor-neutral representation of your business rules so that others may find you
 - describe privacy requirements and policies, and let clients “merge” those (e.g., when paying with a credit card)
- Hence the name of the working group: Rule Interchange Format
 - a language that
 - expresses the rules a bit like a rule language with, eg, RDF
 - can be used to exchange rules among engines

> In an ideal World

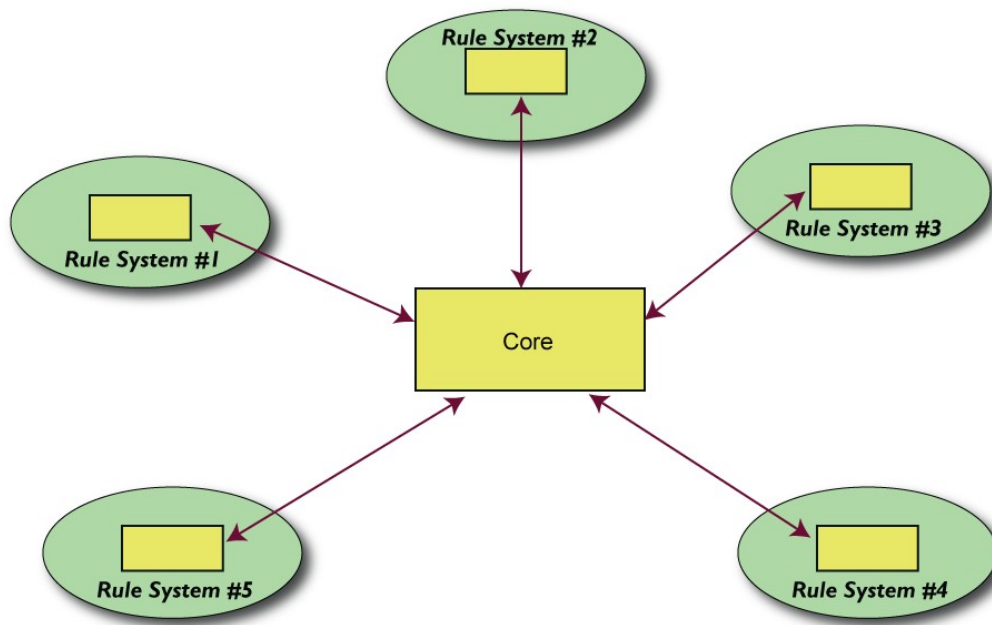


> In the real World...



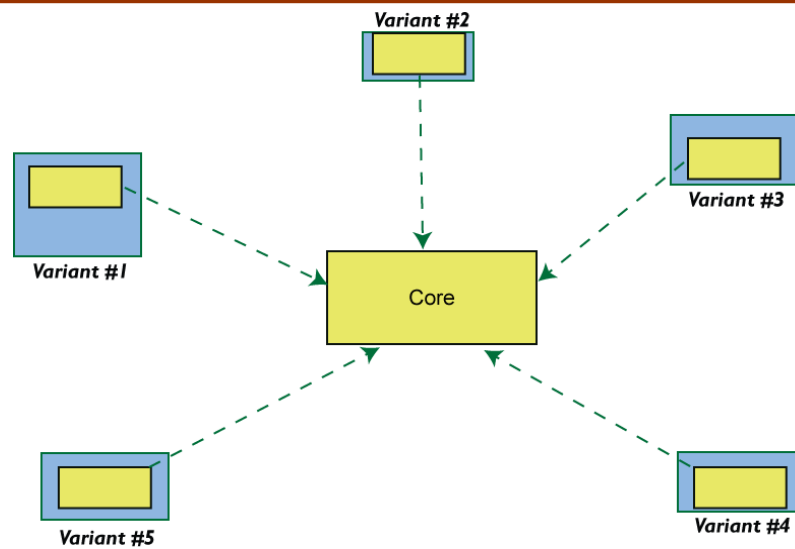
- Rule based systems can be very different
 - different rule semantics (based on various type of model theories, on proof systems, etc)
 - production rule systems, with procedural references, state transitions, etc
- Such universal exchange format is not feasible
- The idea is to define “cores” for a family of languages with “variants”

> RIF “core”: only partial interchange



Ie, only those aspects of, say, Rule System #1 can be exchanged with Rule system #4 that are in the core

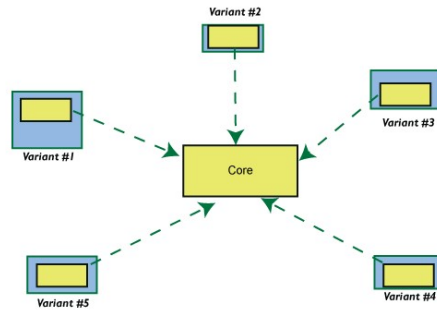
> RIF “variants”



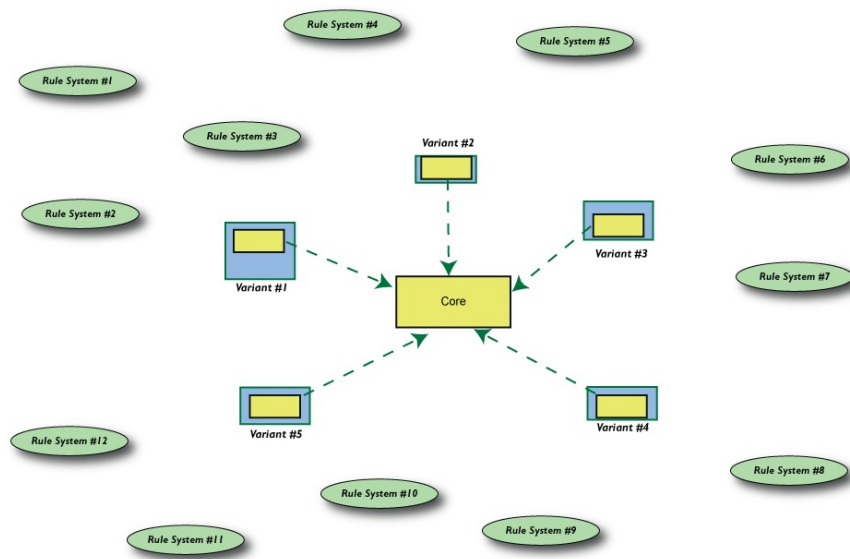
- Possible variants: F-logic, production rules, fuzzy logic systems, ...; none of these have been finalized yet

Variants are, in fact, an extension mechanism to the core...

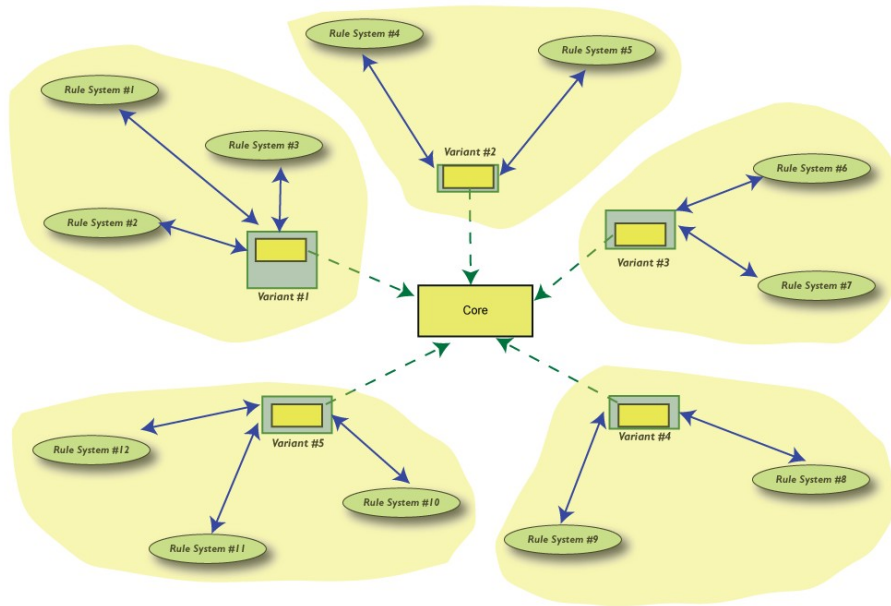
> Role of variants



> Role of variants

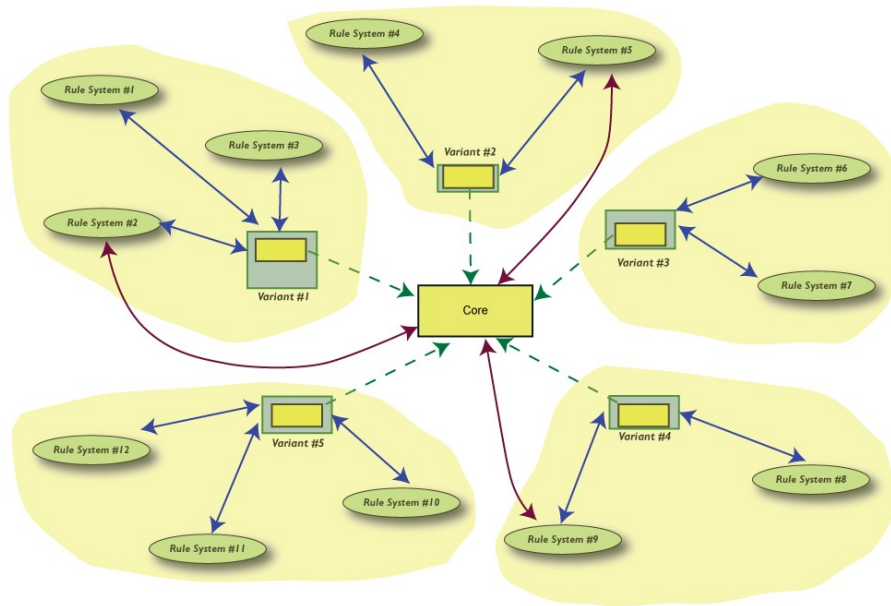


> Role of variants



I.e: variants can play the role of an exchange 'core' within a family of rule systems, but for exchange among families, only the basic core can be applied.

> Role of variants



I.e: variants can play the role of an exchange 'core' within a family of rule systems, but for exchange among families, only the basic core can be applied.

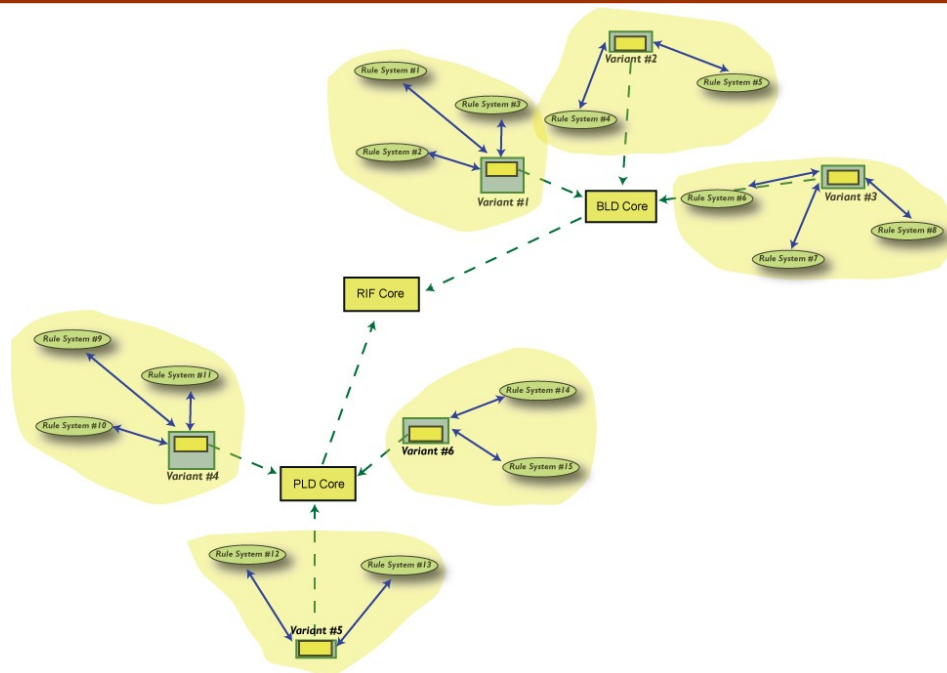
> However...



- Even this model does not completely work 😞
- The gap between production rules and “traditional” logic systems seems to be large
- A hierarchy of cores may be necessary:
 - a *Basic Logic Dialect* and *Production Rule Dialect* as “cores” for families of languages
 - a common *RIF Core* binding these two

The caveat: the model on previous pages was dominating the discussion in the group until around early autumn 2007, but it did not prove to be 100% feasible:-)

> Hierarchy of cores



It is the same model as before but with one more level in the exchange hierarchy. Whether the central core will become feasible is still an open issue at this moment.

> Current status



- There is a draft for the BLD
 - it defines a “positive Horn” language
 - it is a logic based general rule language
 - the language can be used
 - with or without RDF data and/or OWL
 - as a rule *language* or a rule *interchange format*
- The plan is to have BLD as a recommendation in 2008
- The work on the PLD Core has also begun

The publication of the more complete BLD draft is imminent (February 2008)



How do applications look like?

> 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)

X2X means here all the different buzzwords: B2B, B2C, etc...

> Applications can be very simple



- 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's Item Filtering

Area	5/2	Approach	3/2	Term-Reason	1/2	Max_Stage Reached	1/2	Literature Links
2/1 Pain	<input checked="" type="checkbox"/>	7 Antibody	<input checked="" type="checkbox"/>	37 ACTIVE	<input checked="" type="checkbox"/>	51 Candidate	<input checked="" type="checkbox"/>	0 - 50
1/6 Metabolic Disease	<input checked="" type="checkbox"/>	1 Recombinant	<input checked="" type="checkbox"/>	12 BICOMARER	<input checked="" type="checkbox"/>	31 Discovery	<input checked="" type="checkbox"/>	
1 Cancer	<input checked="" type="checkbox"/>	10 SM_Antagonist	<input checked="" type="checkbox"/>	51 EFFICACY	<input checked="" type="checkbox"/>	41 Exploratory	<input checked="" type="checkbox"/>	
1 Cancer	<input checked="" type="checkbox"/>	10 SM_Antagonist	<input checked="" type="checkbox"/>	51 EFFICACY	<input checked="" type="checkbox"/>	41 Exploratory	<input checked="" 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's Item Filtering

Area	5/2	Approach	3/2	Term-Reason	1/2	Max_Stage Reached	1/2	Literature Links
2/1 Pain	<input checked="" type="checkbox"/>	7 Antibody	<input checked="" type="checkbox"/>	37 ACTIVE	<input checked="" type="checkbox"/>	51 Candidate	<input checked="" type="checkbox"/>	0 - 50
1/6 Metabolic Disease	<input checked="" type="checkbox"/>	1 Recombinant	<input checked="" type="checkbox"/>	12 BICOMARER	<input checked="" type="checkbox"/>	31 Discovery	<input checked="" type="checkbox"/>	
1 Cancer	<input checked="" type="checkbox"/>	10 SM_Antagonist	<input checked="" type="checkbox"/>	51 EFFICACY	<input checked="" type="checkbox"/>	41 Exploratory	<input checked="" type="checkbox"/>	
1 Cancer	<input checked="" type="checkbox"/>	10 SM_Antagonist	<input checked="" type="checkbox"/>	51 EFFICACY	<input checked="" type="checkbox"/>	41 Exploratory	<input checked="" type="checkbox"/>	

51 Items (Selected from 110 originally listed 20/07/09)

Area	Original-Indication	Target Name	Approach	Start	Term-Reason	Max_Stage Reached	Owner	QIDM URL	AB URL	2007 UR	Neck	NSA	QED Pathway	Compounds
Metabolic Disease	Diabetes	Insulin	SM_Inhibitor	2007-Q2 EFFICACY	Candidate		P. Person							SA-000072
Sexual Health	Erectile Dysfunction	Phosphodiesterase 5 Inhibitor	SM_Antagonist	2006-Q3 EFFICACY	Candidate		P. Person							SA-000072
Sexual Health	Erectile Dysfunction	Leukotriene C4 synthase	SM_Antagonist	2006-Q3 EFFICACY	Candidate		M. Manager							SA-000072
Sexual Health	Erectile Dysfunction	Transcription elongation factor A (TFIIA)	SM_Inhibitor	2005-Q3 EFFICACY	Candidate		P. Person							SA-000072
Infectious	HIV	Protease Inhibitor	SM_Inhibitor	2006-Q2 EFFICACY	Candidate		L. Leader							SA-000072
Infectious	HIV	Voltage-gated potassium channel protein	SM_Antagonist	2007-Q2 EFFICACY	Candidate		A. Scientist							SA-000072
Urogenital Incontinence	Urinary Incontinence	Alpha-1A adrenergic receptor (ADRA1A)	SM_Antagonist	2007-Q2 EFFICACY	Candidate		L. Leader							SA-000072
Pain	Migraine	Monoclonal antibody	SM_Inhibitor	2007-Q2 EFFICACY	Candidate		L. Leader							SA-000072

Courtesy of Nigel Wilkinson, Lee Harland, Pfizer Ltd, Melliyl Annamalai, Oracle (SWEQ Case Study)

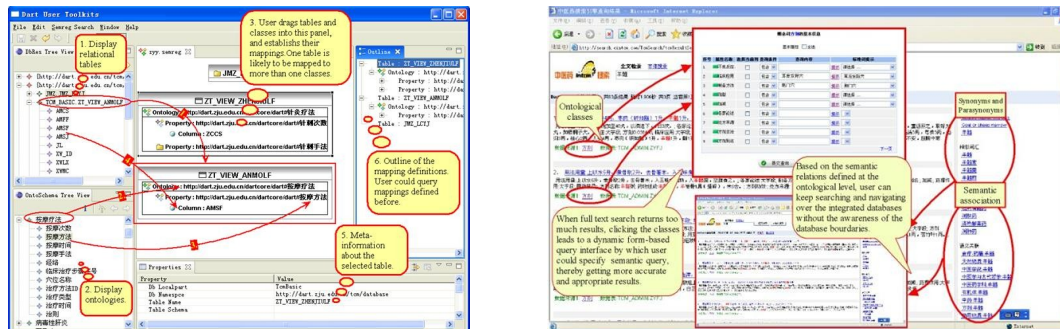
Various types of databases are accessed having an RDF transformation of (part of the data) on the fly. Some of the data may be simple tables, some are the result of continuous background processing analysing the literature (not directly related to the Semantic Web per se).

The integration of the data is done on the RDF level, and is viewed via an off-the-shelf (though experimental) faceted browser (Exhibit). I.e., the Semantic Web portion is very simple but allows for a very quick integration of the data on the screen.

> Integrate knowledge for Chinese Medicine



- Integration of a large number of relational databases (on traditional Chinese medicine) using a Semantic Layer
 - 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)

Karl Dubost and Ivan Herman, The state of the Semantic Web

(59)



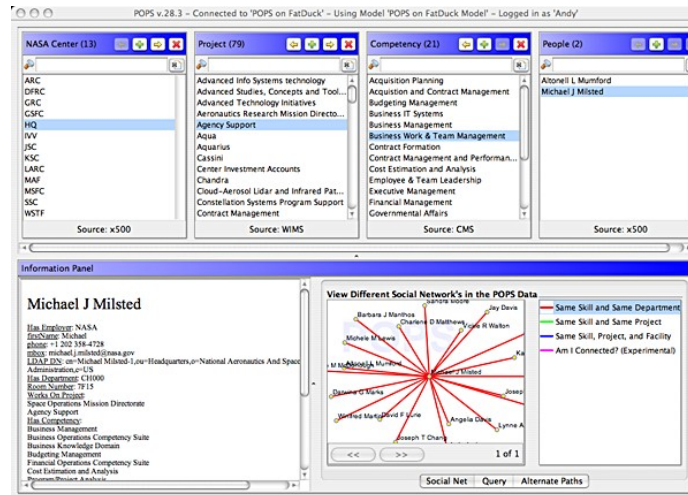
The various databases around the country are handled by independent bodies. A visual query generator creates a SPARQL query, this is then decomposed to access the individual databases, on-the-fly transformed into SQL queries, the result are in RDF and recombined for the answer. The system is used in the Chinese Academy of Sciences' Research institute on traditional Chinese medicine.

The university is also working on a nation-wide ontology on traditional Chinese Medicine that can be combined with the search to improve it. Still in development. That ontology might be bound to western medical ontologies, too (eventually).

> Find the right experts at NASA



- Expertise locator for nearly 20,000 NASA civil servants using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services...



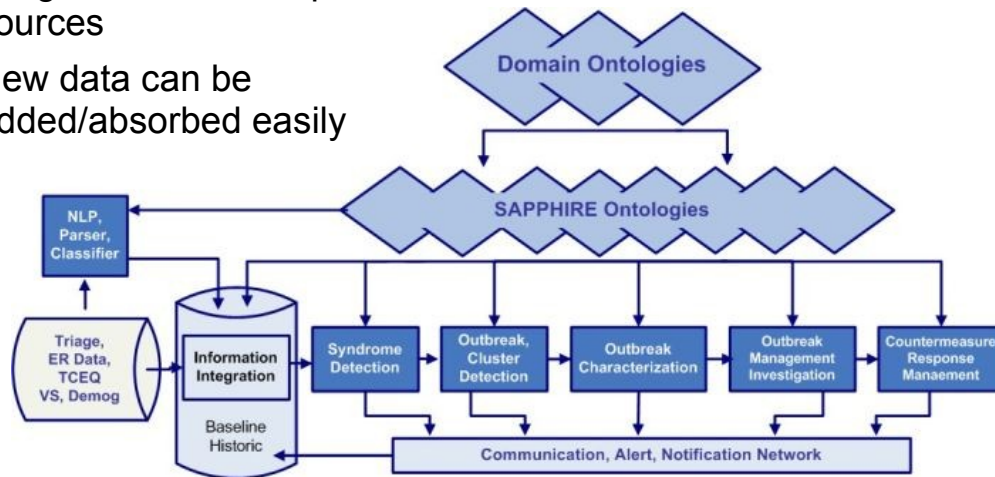
Courtesy of Kendall Clark, Clark & Parsia, LLC

The use internal ontologies/vocabularies to describe the knowledge areas, and a combination of the RDF data and that ontology to search through the (integrated) databases for a specific knowledge expertise. The dump is from a faceted browser developed by the company to view result data.

> Public health surveillance (Sapphire)



- Integrated biosurveillance system (biohazards, bioterrorism, disease control, etc)
- Integrates from multiple data sources
- New data can be added/absorbed easily



Courtesy of Parsa Mirhaji, School of Health Information Sciences, University of Texas (SWEQ Case Study)

The University of Texas Health Science Center (UTHSC) has used Semantic Web technologies to build a prototype system for context-aware interpretation and integration of clinical data, environmental readings, and patient interviews. The system integrates a wide range of health and epidemiological data from local healthcare providers, hospitals and pharmacies. SAPHIRE constructs a collaborative and distributed system to analyze, detect, and respond to public health matters.

Every ten minutes, SAPHIRE receives electronic health records, triage data, patients' complaints, and clinician's notes from eight hospitals spanning four counties and covering more than 30% of all Houston-area emergency-room visits. Using unstructured text analysis and Semantic Web technologies, this information is mined and integrated into a single view of current health conditions across the city.

The flexibility of Semantic Web technologies allows SAPHIRE to operate equally effectively in other contexts. At Hurricane Katrina. Within eight hours of the opening of the shelters, UTHSC configured SAPHIRE to respond to the needs of the disaster.

> 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 displays the AKSIO search interface. On the left, there are search filters categorized by 'Description', 'Operation', 'Equipment', 'State', 'Maintenance', and 'Risk'. The 'State' filter is expanded, showing options like 'Completion', 'Emission', 'Leak of Maintenance', 'Leak in barrier elements', 'Scale Deposition', 'Too high Mud Density', and 'Well Integrity Problem'. The main area shows search results for 'leak in barrier elements'. The results list includes details such as the well name (e.g., 'Top alene/28" K25X'), the description of the experience, the date, and the accuracy of the information (e.g., 55%, 30%, 47%, 39%).

Courtesy of David Norheim and Roar Fjellheim, Computas AS (SWEQ Use Case)

Karl Dubost and Ivan Herman, The state of the Semantic Web

(62)



The system has been developed and tested for Statoil, which is the largest oil company on the Norwegian Continental Shelf, and covers experiences with over 2,500 drilling operations since the early 90s.

The objective of the reuse improvements is to discover relevant experiences that could affect current or planned drilling operations. The shared domain ontology is used for semantic annotation, and for retrieval of information. It is developed collaboratively by the discipline advisors, and covers operations, equipment, events and failure states in drilling operations. It also includes relations between these concepts, for example, to indicate that a particular event may result in a failure state.

> 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)

> Help in choosing the right drug regimen



- Help in finding the best drug regimen for a specific case
 - find the best trade-off for a patient
- Integrate data from various sources (patients, physicians, Pharma, researchers, ontologies, etc)
- Data (eg, regulation, drugs) change often, but the tool is much more resistant against change

The screenshot shows the PharmaSURVEY for Abby interface. It has tabs for 'Optimized Regimens', 'Current Regimen', 'Survey Summary', and 'Survey Set-up'. The 'Current Regimen' tab is active, showing 'Survey 1 Version 2'. Below the tabs are buttons for 'Optimized', 'Differences', 'Ranges', 'Common', 'Severe', 'All', and 'Custom'. The main table displays 'Adverse Drug Effect' with columns for 'Severity', 'ADE', and 'v2'. The table lists six ADEs: 'Moderate Muscle Weakness (Myasthenia)', 'Minor Excessive Sweating (Diaphoresis)', 'Moderate Heart Throbbing or Pounding (Palpitations)', 'Moderate Hives (Urticaria)', 'Major Bladder Inflammation (Cystitis)', and 'Major Urinary Tract Infection'. Each ADE has a checkmark in the 'v2' column. To the right of the table are 'Safety Optimized Profiles' for 'Current Regimen' and two profiles labeled '1' and '2'. At the bottom right, it says 'row(s) 1 - 6 of 6' and 'Export to Excel'.

Severity	Adverse Drug Effect	v2
ADE	Moderate Muscle Weakness (Myasthenia)	✓
ADE	Minor Excessive Sweating (Diaphoresis)	✓
ADE	Moderate Heart Throbbing or Pounding (Palpitations)	✓
ADE	Moderate Hives (Urticaria)	✓
ADE	Major Bladder Inflammation (Cystitis)	✓
ADE	Major Urinary Tract Infection	✓

Courtesy of Erick Von Schweber, PharmaSURVEYOR Inc., ([SWEU Use Case](#))

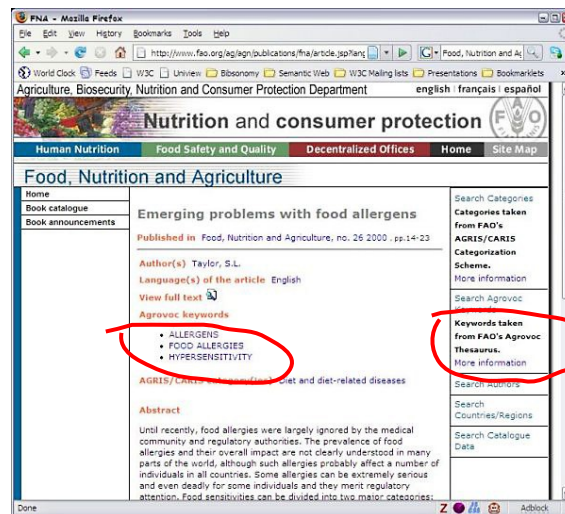
Tool to find the best drug usage adapted to an individual patient. The navigator tool combines various databases, ontologies to provide a better tool.

The flexibility of the interface is important: the structure of the underlying data (eg, databases, regulation) change often, but by localizing the change on the database-to-RDF mapping, the rest of the system is protected against change; one of the main reasons why this approach was chosen

> FAO Journal portal



- Improved search on journal content based on an agricultural ontology and thesaurus (AGROVOC)



Courtesy of Gauri Salokhe, Margherita Sini, and Johannes Keizer, FAO, (SWEQ Case Study)

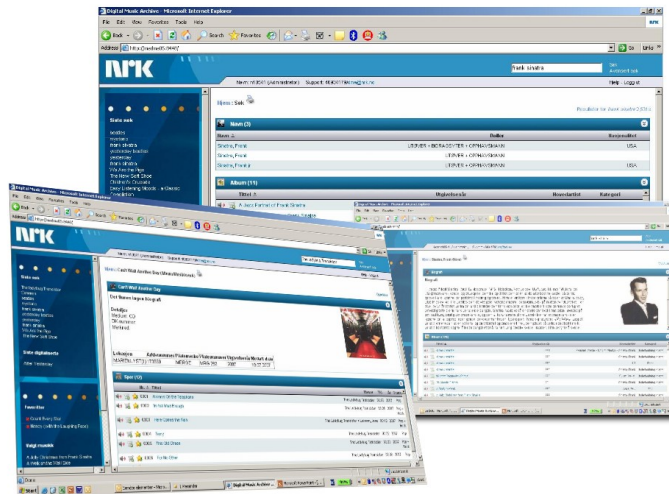
The articles in the Food, Nutrition and Agriculture (FNA) Journal cover topics such as community nutrition, food quality and safety, nutrition assessment, nutrient requirements, food security and rural development. The full-text articles may be in English, French or Spanish.

Metadata about each article in the FNA Journal was available in different formats. Work has been undertaken to combine the metadata and to convert it to a single RDF(S) format, using some ontologies developed internally. A search application was created on top of the ontology and the instance data. A user is guided through the navigation of data by following the links that connect the different metadata elements, such as articles within a specific issue, authors, languages, or keyword.

> Digital music asset portal at NRK



- Used by program production to find the right music in the archive for a specific show



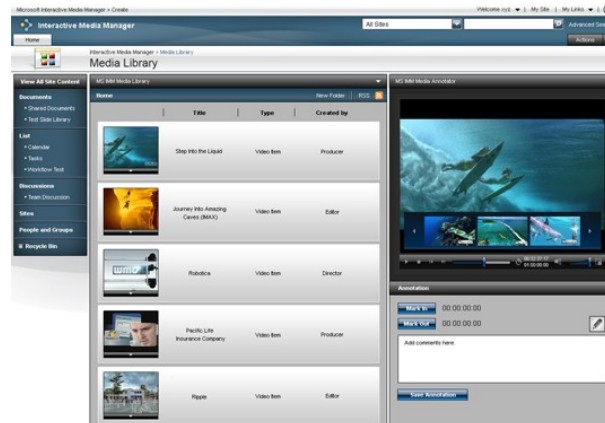
Courtesy of Robert Engels, ESIS, and Jon Roar Tønnesen, NRK (SWEQ Case Study)

NRK is the Norwegian National TV; currently 1.2 million tracks are digitized and only 45,000 are used in practice when, for example, a new program is planned and a music track is to be found to accompany it. Via the metadata, vocabulary, and associated search a much better environment is provided to find appropriate music track. “Hidden assets” could be found much more easily that way. The user interface also provides a much easier and quicker access to data.

> Microsoft Vista's Interactive Media Manager



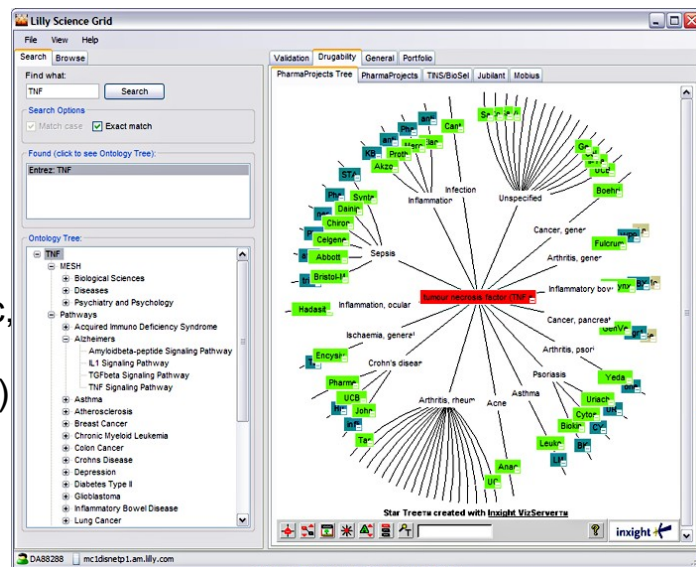
- Uses an RDF/SPARQL/OWL based metadata framework
 - eg, for a better control over relationships among media assets and categories
- Custom OWL ontologies can be created and imported



> Eli Lilly's Target Assessment Tool



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



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

The important point is that the (subsequent) search is not simply done on a (key)word level, like for a traditional search engine, but through the tree of all related terms, where those relations are determined via internal and public ontologies and vocabularies.

The screen snapshot illustrates the user interface of the Target Assessment Tool within Lilly Science Grid. The panels to the left of the screen snapshot show that it is possible to directly search for a term or to navigate the ontology to identify a term of interest. The panel to the right of the screen snapshot demonstrates a graph view of data within the data as it relates to the search term.

> Improved Search via Ontology: GoPubMed



- Improved search on top of pubmed.org
 - search results are ranked using ontologies
 - related terms are highlighted, usable for further search



Karl Dubost and Ivan Herman, The state of the Semantic Web

(69)



Pubmed.org is the 'google' of the medical profession. The result of a search is re-ranked, a better interface is provided, and related terms are also shown based on public medical ontologies. The left hand side refers to the Gene Ontology and Mesh ontologies; actually (red highlight on the screen) and the user is able to follow up related terms, too.

Produced by a German company (transinight)

> Radar Network's Twine

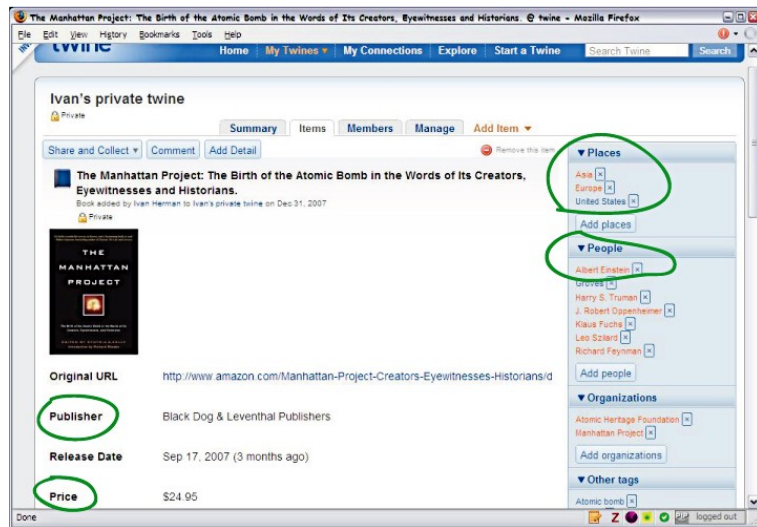


- “Social bookmarking on steroids”

- Item relationships are based on ontologies

- evolving over time
- possibly enriched by users

- Internals in RDF, will be available via APIs and SPARQL



> Other application areas come to the fore



- Content management
- Business intelligence
- Collaborative user interfaces
- Sensor-based services
- Linking virtual communities
- Grid infrastructure
- Multimedia data management
- Etc

> Thank you for your attention!



- These slides are publicly available on:

<http://www.w3.org/2008/Talks/0307-Tokyo-IH/>

- There is also a collection of use cases at:

<http://www.w3.org/2001/sw/sweo/public/UseCases/>