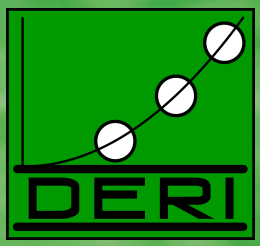
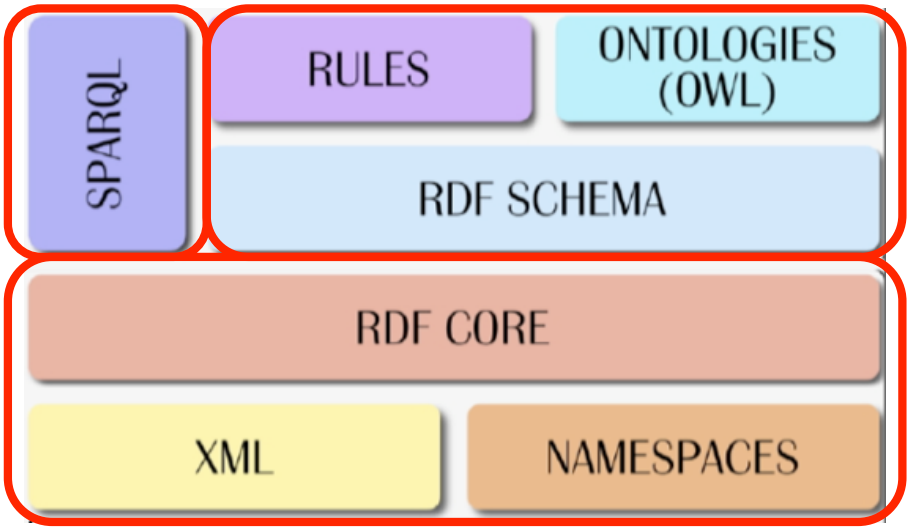


Semantic Web Technologies: From Theory to Standards

Axel Polleres
Digital Enterprise Research Institute, NUI Galway

The Semantic Web in W3C's view:

3. Shall allow us to ask **structured queries** on the Web



2. Shall allow us to describe the structure of information in machine readable form: **RDFS+OWL+RIF**

1. Shall allow us to publish structured information on the Web: **XML+RDF**

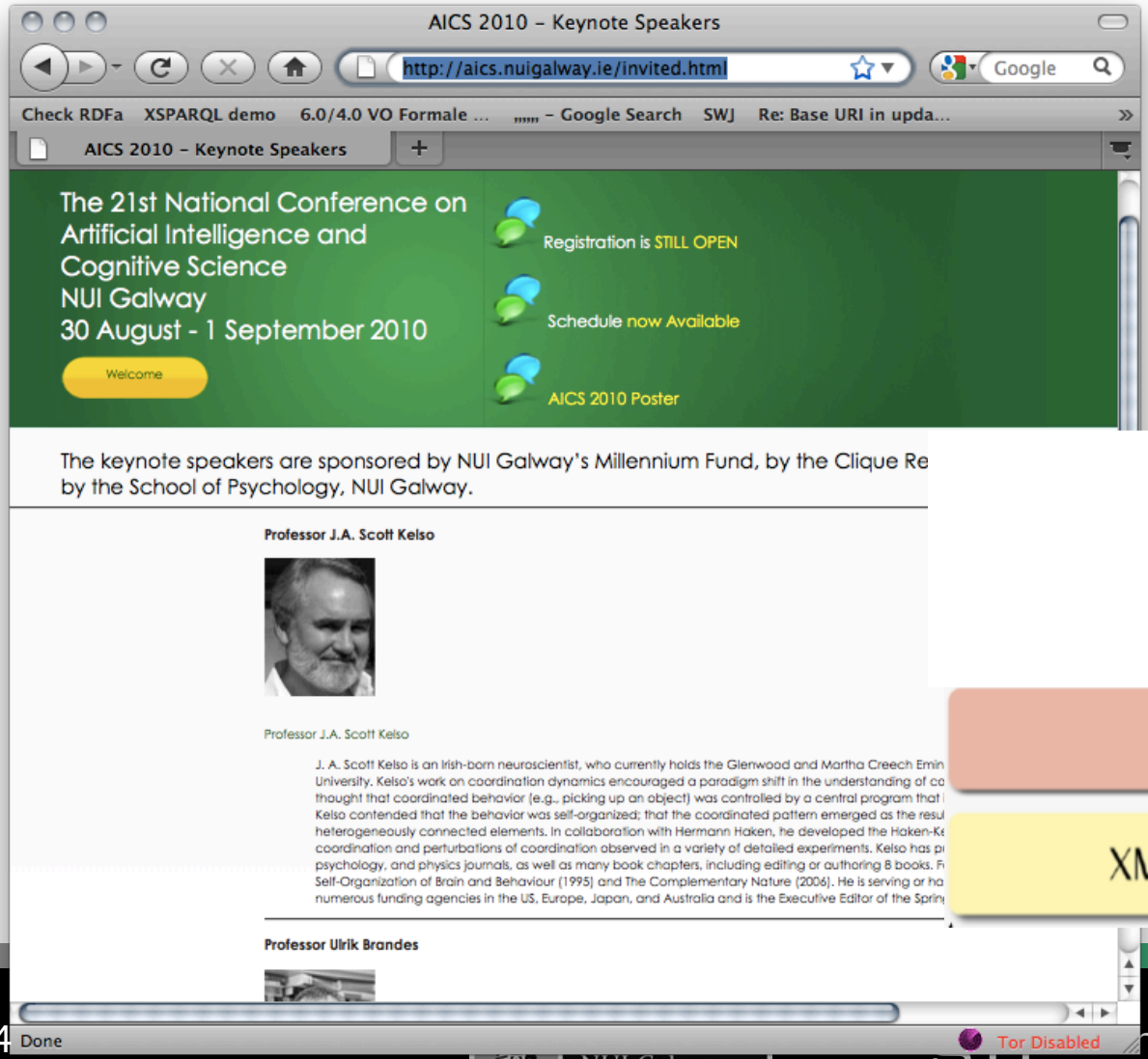
Focus in this talk/paper:

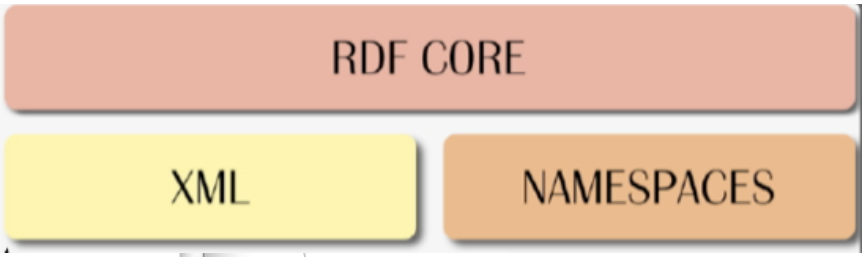
- Which theory do these Sem. Web standards base on?
- What's missing? (= Do these standards work together)

- (Brief overview of own contributions/solutions in this area, details in the references, paper is meant as a literature survey, etnry point)

1. Structured Data on the Web

"Prof. Scott Kelso gives a Keynote at AICS"

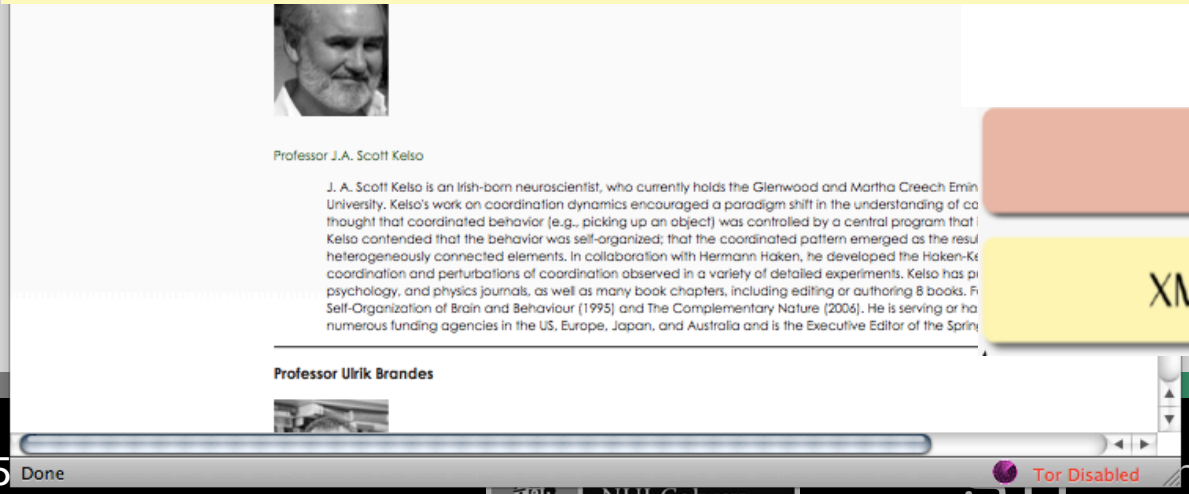


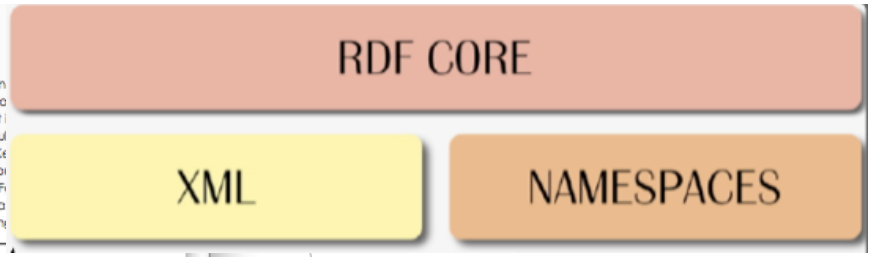


1. Structured Data on the Web

"Prof. Scott Kelso gives a Keynote at AICS"

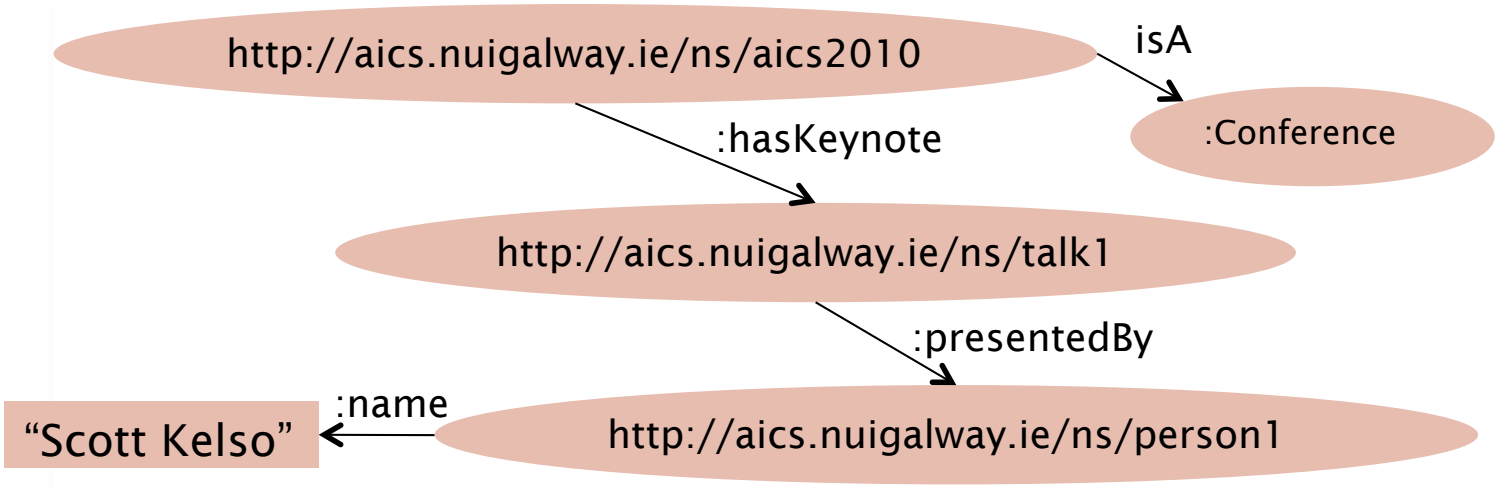
```
<conference xmlns="http://aics.nuigalway.ie/ns/">
  <name>The 21st National Conference on Artificial Intelligence
and Cognitive Science</name>
  <keynote id="talk1" href="http://aics.nuigalway.ie/
invited.html">
    <presentedBy ref="person1">Scott Kelso</presentedBy>
  </keynote>
  <keynote>
  ...
</conference>
```

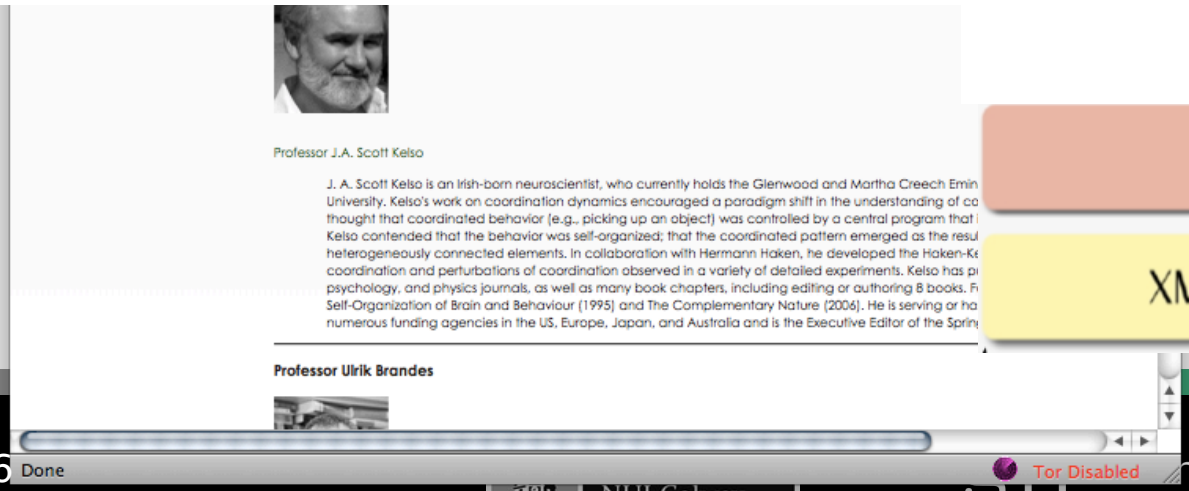


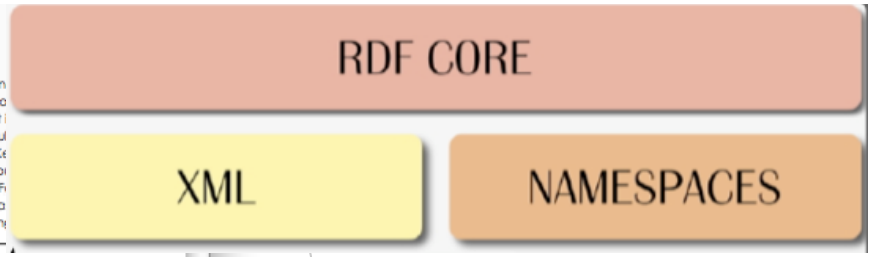


1. Structured Data on the Web

"Prof. Scott Kelso gives a Keynote at AICS"







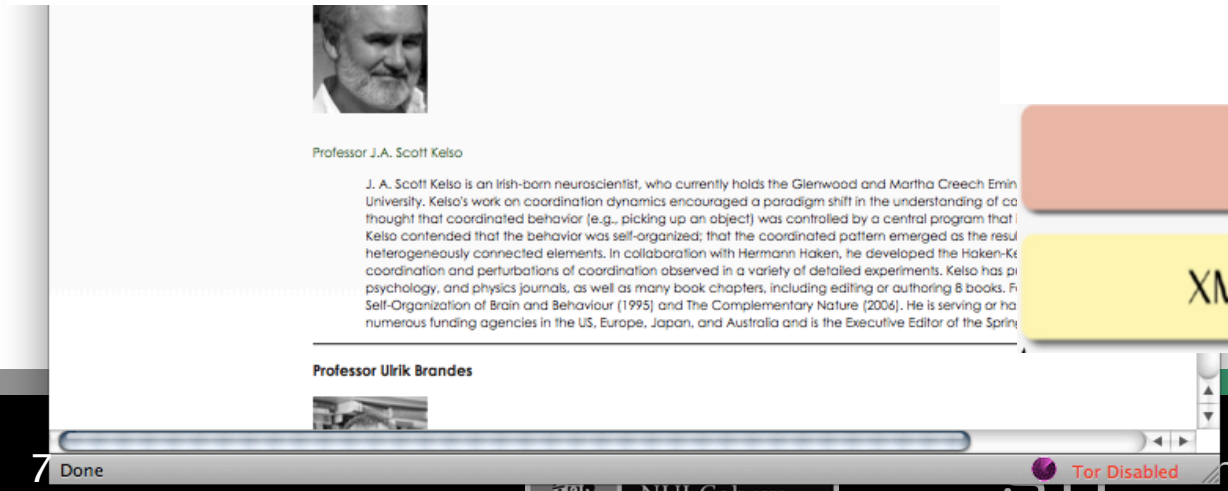
1. Structured Data on the Web

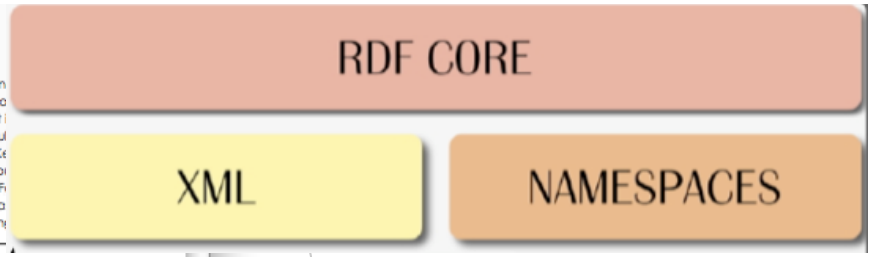
"Prof. Scott Kelso gives a Keynote at AICS"

```
<http://aics.nuigalway.ie/ns/person1>
        :name   "Scott Kelso" .

<http://aics.nuigalway.ie/ns/aics2010>
        rdf:type :Conference ;
        :hasKeynote <http://aics.nuigalway.ie/ns/talk1> .

<http://aics.nuigalway.ie/ns/talk1>
        :presentedBy
        <http://aics.nuigalway.ie/ns/person1> .
```





1. Structured Data on the Web

"Prof. Scott Kelso gives a Keynote at AICS"

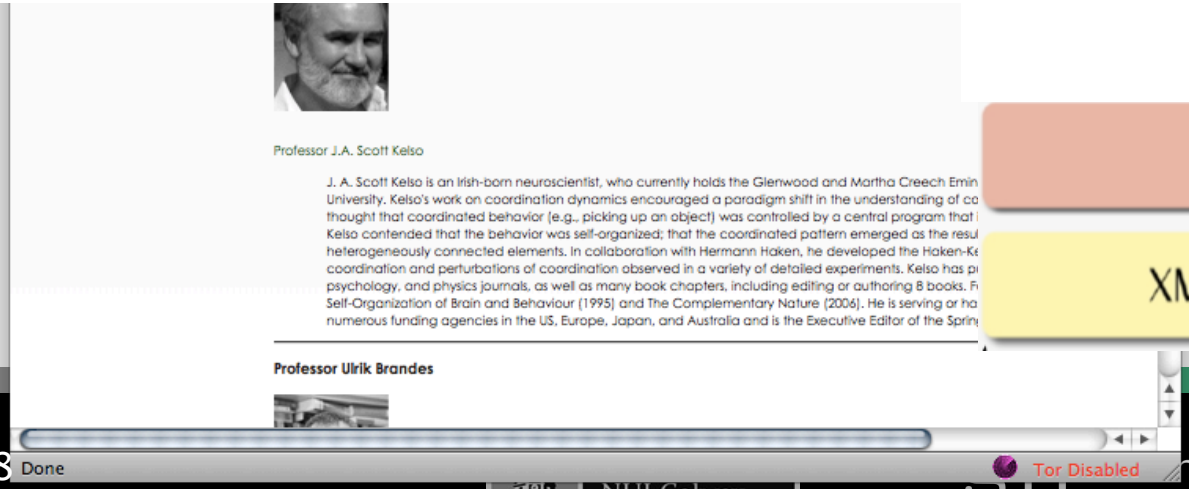
$$name(person1, "Scott\ Kelso") \wedge$$

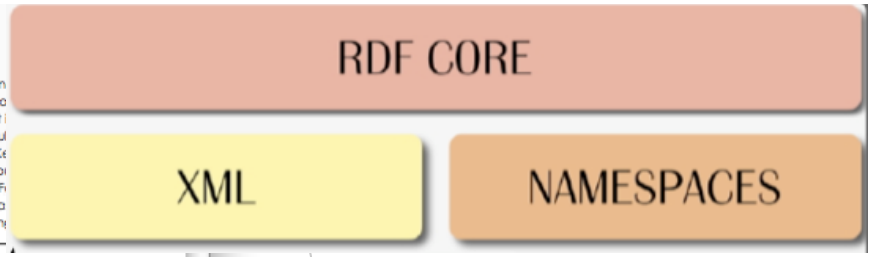
$$Conference(aics2010) \wedge$$

$$hasKeynote(aics2010, talk1) \wedge$$

$$presentedBy(talk1, person1)$$

RDF+RDF Schema can be embedded in ***FOL*** [deBruijn et al. 2005] …
…or ***Datalog*** [deBruijn et al. 2007] [Ianni et al. 2009]





1. Structured Data on the Web

"Prof. Scott Kelso gives a Keynote at AICS"

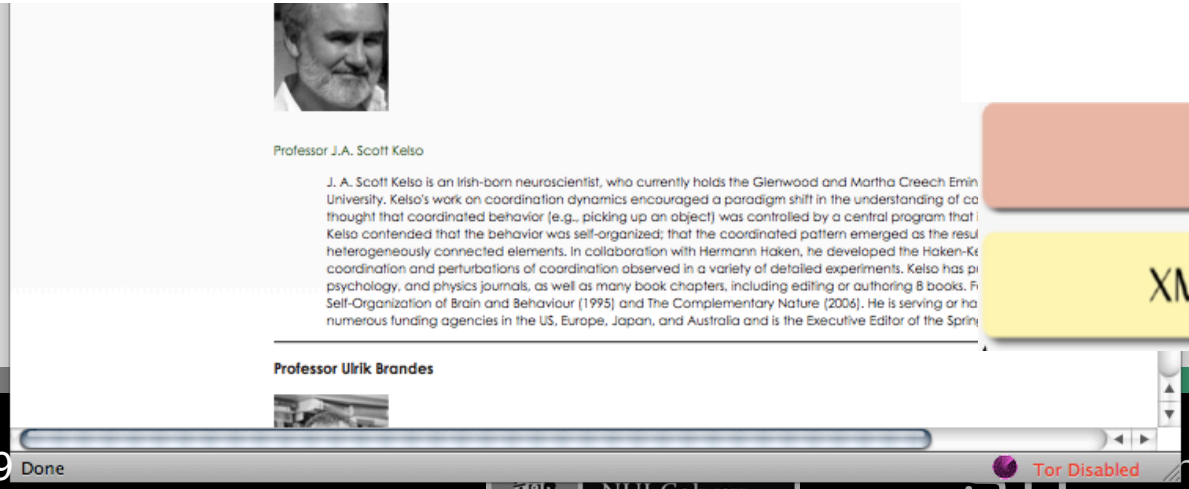
name(person1, "Scott Kelso").

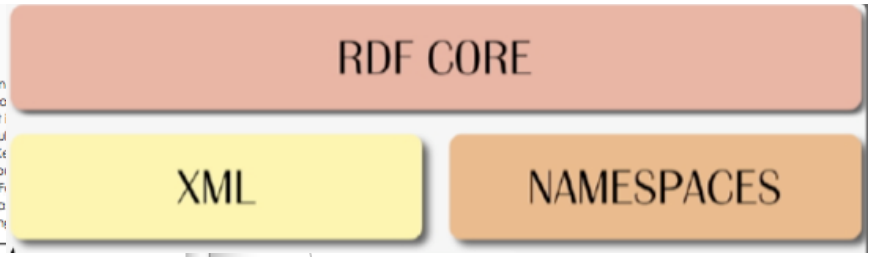
Conference(aics2010).

hasKeynote(aics2010, talk1).

presentedBy(talk1,person1).

RDF+RDF Schema can be embedded in ***FOL*** [deBruijn et al. 2005] …
…or ***Datalog*** [deBruijn et al. 2007] [Ianni et al. 2009]





RDF is the basis for Linked Data:

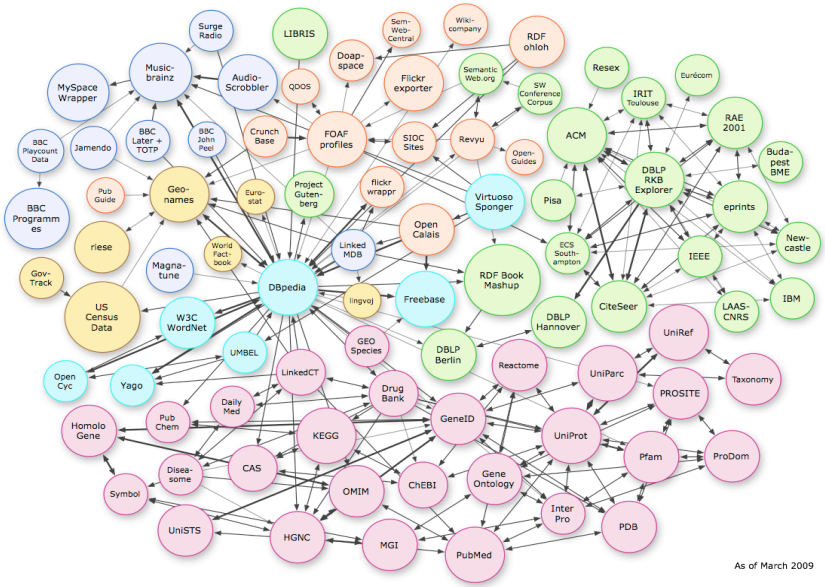
1. Everything gets a URI (conferences, people, talks, …)
2. These URIs are linked via RDF describing relations
3. Relations are URIs again (e.g. :name)
4. When I dereference the URIs, I should find more information about them

GOV-data

dblp .uni-trier.de Computer Science Bibliography

…

The New York Times



Freebase

WIKIPEDIA

facebook

OÉ Gaillimh NUI Galway

sfi science foundation ireland fondúireacht eolaíochta éireann

Enabling **networked** knowledge.

2. RDF can be described in terms of Ontologies and Rules ➔ *allows Reasoning!*

name(person1, "Scott Kelso")

Conference(aics2010)
hasKeynote(aics2010, talk1)
presentedBy(talk1,person1).

Attendee(person1).
Attendee(person2).

$\exists hasKeynote^{-}.\top \sqsubseteq Talk$
$Talk \sqcap \exists givenAt\{aics2010\} \sqsubseteq \exists hasTopic\{AI\}$

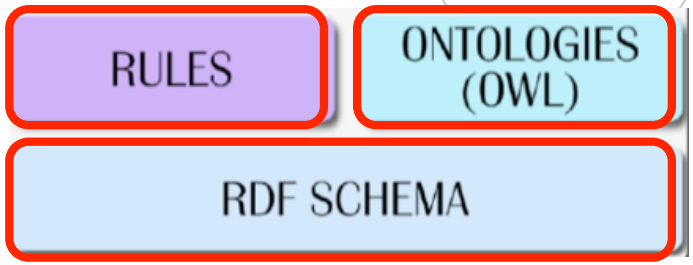
givenAt(E,T) :- hasKeynote(E,T).
*attendedBy(T,P) :- Attendee(P), **not** presentedBy(T,P).*

:talk1 :hasTopic dbpedia:AI .
:talk1 :attentedBy :person2 . ?

➔ RDF Schema (RDFS)

➔ Web Ont. Lang. (OWL)

➔ Rule Interchange Format (RIF)

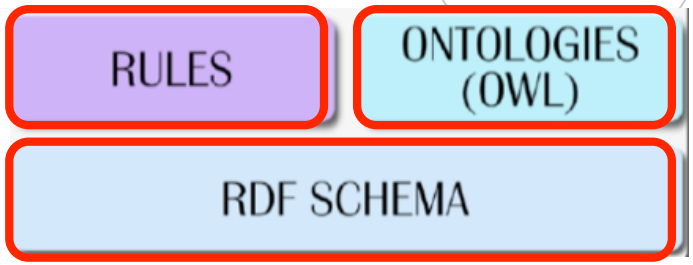


2. RDF can be described in terms of Ontologies and Rules ➔ *allows Reasoning!*

*"Every **keynote** at an event is a **talk**"* ➔ RDF Schema (RDFS)

*"Every **talk given at AICS2010** is **about AI**"* ➔ Web Ont. Lang. (OWL)

*"If an event **has a keynote**, it is a speech **given at** the event"*
*"Every AICS attendee **not** presenting a talk is attending the talk."* ➔ Rule Interchange Format (RIF)




```
:hasKeynote rdfs:range :Talk .
```

$$\exists hasKeynote^{-}.\top \sqsubseteq Talk$$
$$Talk \sqcap \exists givenAt\{aics2010\} \sqsubseteq \exists hasTopic\{AI\}$$

givenAt(E,T) :- hasKeynote(E,T).
*attendedBy(T,P) :- Attendee(P), **not** presentedBy(T,P).*

2. RDF can be described in terms of Ontologies and Rules ➔ *allows Reasoning!*

OWL's theoretical foundation: Description Logics,

SHOIN [Horrocks and Patel-Schneider, 2004]

SROIQ [Horrocks et al. 2006]

RIF's theoretical foundation: Logic programming, F-Logic,

but also Datalog/Answer Set Programming, Deductive Databases

(some RIF dialects allow negation as failure)

RDF Schema: in essence in the intersection

(but strictly speaking more liberal than Description Logics)

2. Structured queries over Web data

- SPARQL = "SQL look-and-feel query language for the Web"
- allows us to ask structured queries such as:

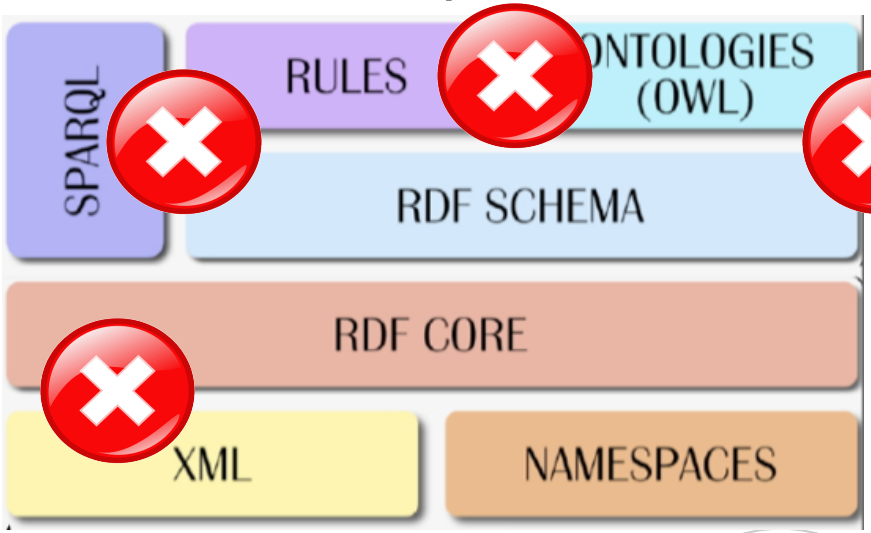
"Give me names of people presenting AI or SemanticWeb talks"

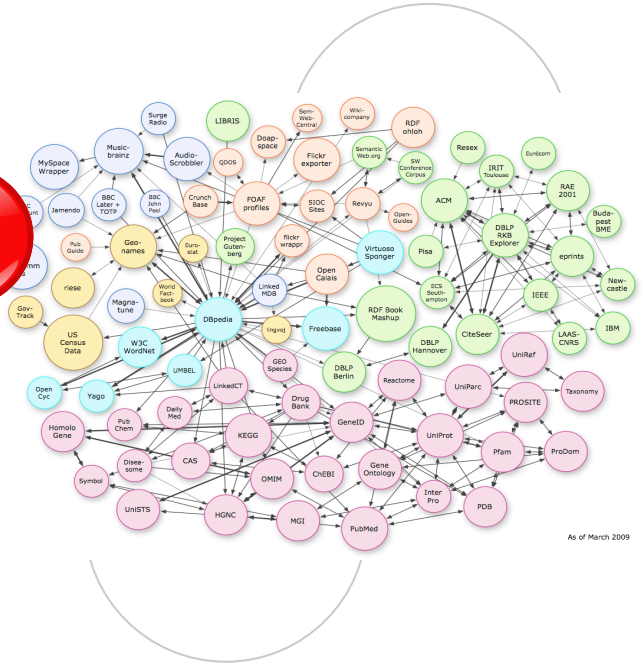
```
SELECT ?Talk ?N
 { ?Talk :presentedBy ?P . ?P :name ?N
   { { ?Talk :hasTopic dbpedia:AI . }
     UNION
     { ?Talk :hasTopic dbpedia:Semantic_Web . }
   } }
```

Unions of conjunctive queries, but also advanced features such as outer joins (NOT EXISTS), value filtering, etc.

How do the standards interplay?

- Challenges:
  - Ontologies & Rules: OWL2 & RIF
  - Querying Ontologies & Rules: SPARQL/OWL+RIF
  - Data on the Web is NOT clean/consistent!
  - Querying XML & RDF: XQuery & SPARQL



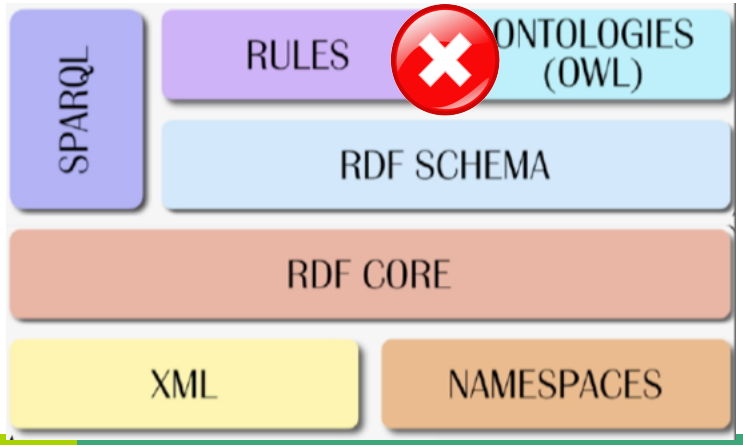


- Some of these challenges in Detail & current solutions to follow…

Ontologies and Rules:

- Decidability:
  - OWL is decidable, Datalog with negation is decidable, but their union isn't.
- Nonmonotonicity:
  - OWL/Description Logics are subsets of classical FO-Logic
  - Rule Languages with Negation as failure (Answer Set Programming, Well-founded semantics) rely on non-classical logics

➔ ***Can't arbitrarily mix RIF with OWL without trouble!***



Approaches:

OWA vs (L)CWA:

Has person2 presented talk1?

$$\exists hasKeynote^{-}.\top \sqsubseteq Talk$$
$$Talk \sqcap \exists givenAt\{aics2010\} \sqsubseteq \exists hasTopic\{AI\}$$

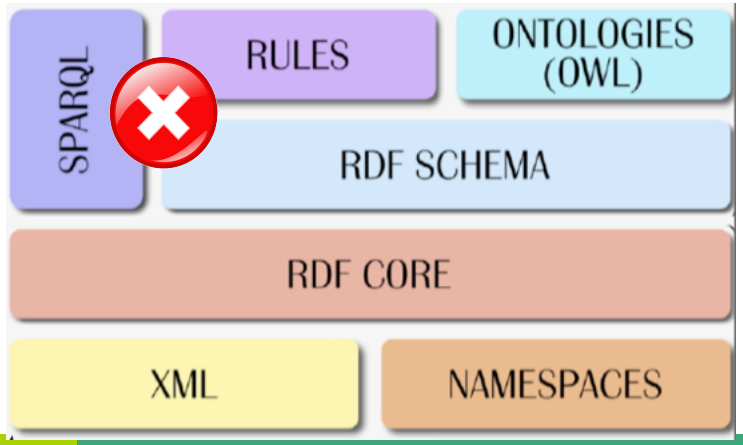
givenAt(E,T) :- hasKeynote(E,T).
*attendedBy(T,P) :- Attendee(P), **not** presentedBy(T,P).*

- Combinations of LP and DL still a vivid field of research…
  - Embedding LP and DL into common non-classical Logics: e.g.
    - first-order autoepistemic Logics [deBruijn, Eiter, Polleres, Tompits et al. 2007,2010]
    - Quantified Equilibrium Logics [deBruijn, Pearce,Polleres, Valverde, 2007, 2010]
  - Defining decidable language fragments to combine: e.g. Horn-SHIQ, OWL2RL, DL-safe rules, cf. Bibliography in the paper)
- … which also means not yet mature for standardisation.

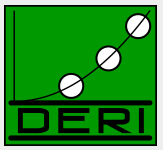
SPARQL & Ontologies:

Similar problems:

- Decidability:
  - Conjunctive queries with non-distinguished variables for expressive DLs is an avtive field of research… OWL2? Not yet known. [Glimm, Rudolph, KR2010]
- Nonmonotonicity:
  - SPARQL has NOT EXISTS/OPTIONAL ~ similar negation as failure.



Approaches:



$$Talk \sqsubseteq \exists hasChair$$

- *"Give me all talks that have a chair?"*

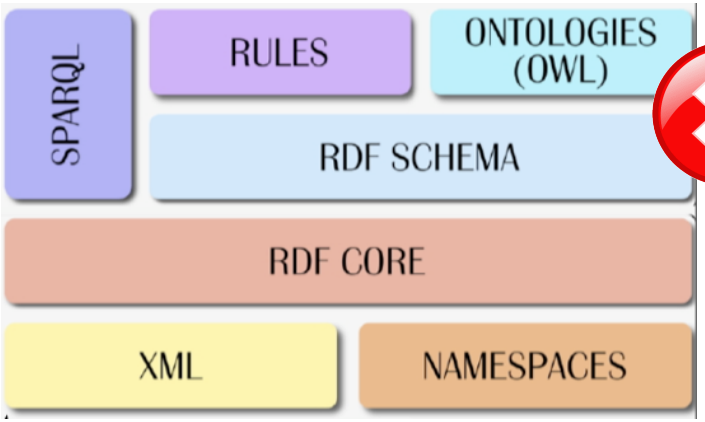
```
SELECT ?T { ?T :hasChair ?C }
```

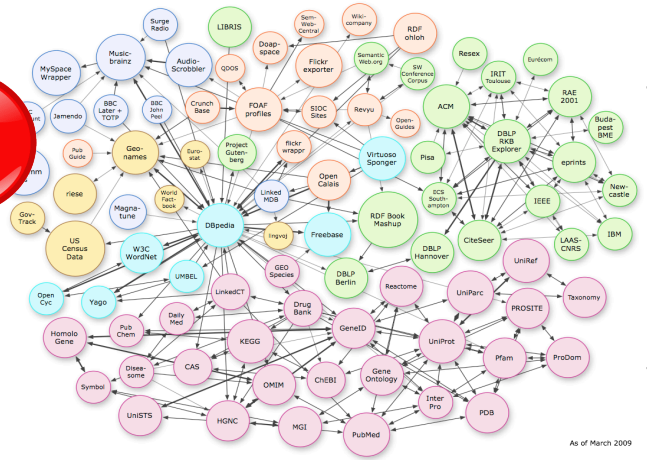
Do I need to know the actual chairs to answer this question?
Two possible views on this query:

- **Yes:** Treat all query variables as distinguished (=output variables):
  - Non-monotonic constructs on top not a problem for this approach
  - SPARQL1.1 is currently exploring this route.
- **No:** in certain subsets of OWL this can be answered:
  - Subset of OWL translatable to SQL: OWL2QL
  - Subset of OWL translatable to extended versions of Datalog:
    Datalog± [Cali et al. 2009]
  - ➔ BTW, query answering not only decidable but also tractable

Is OWL suitable for Linked Data

- OWL DL Reasoning on data crawled from the Web almost certainly yields inconsistencies
- Assuming that the Semantic Web would be less messy than the HTML Web is very optimistic
- Example:
  - Source A says: $Document\ (\ <http://www.nuigalway.ie>\ )$
  - Source B says: $Organisation\ (\ <http://www.nuigalway.ie>\ )$
  - Ontology C says: $Document \sqsubseteq \neg Organisation$





Approaches

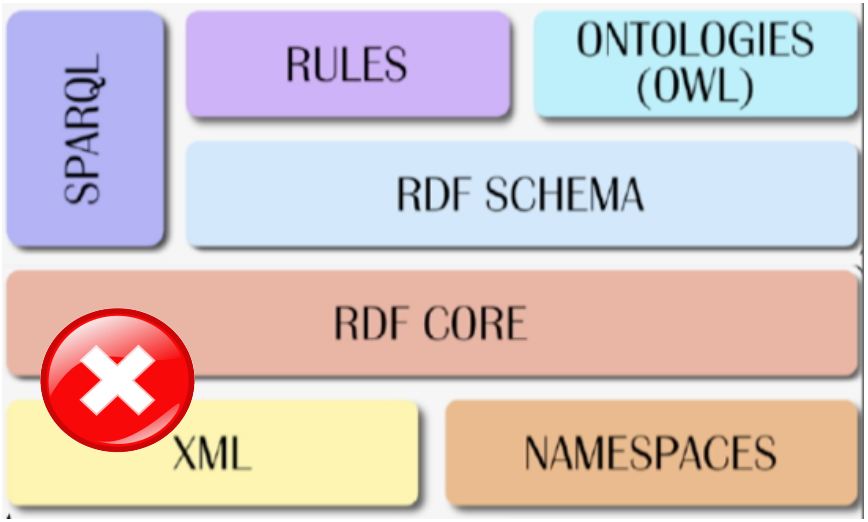
- OWL Reasoning on Web data needs to be scalable & noise tolerant

- Our approach
  - Sound but incomplete reasoning
  - Use a robust/scalable fragment of OWL (OWL2RL)
  - Exploit authority of Web documents
  - Used in Sindice [Delbru et al. 2008], SWSE [Hogan et al. 2009]

- Alternatives?
  - Para-consistent reasoning?
  - RankingSources & Probabilistic Fuzzy Reasoning?

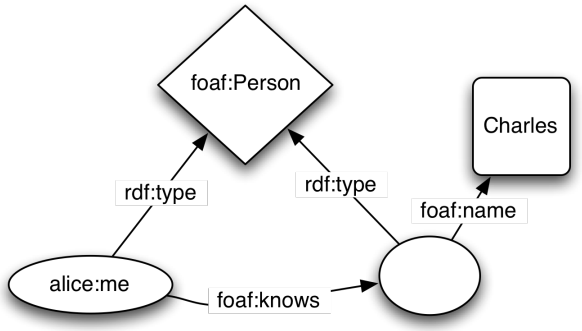
Bringing XML and RDF closer…

- What if I want to translate RDF and OWL data back to XML/HTML ?
  - What to use? Custom Script? XSLT? SPARQL?



Why are XSLT, XQuery not enough?

- Because RDF ≠ RDF/XML !!!



1) many different RDF/XML representations…

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
   <foaf:Person rdf:about="alice/me">
      <foaf:knows>
         <foaf:Person foaf:name="Charles"/>
      </foaf:knows>
   </foaf:Person>
</rdf:RDF>
```

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
   <rdf:Description rdf:nodeID="x">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
      <foaf:name>Charles</foaf:name>
   </rdf:Description>
   <rdf:Description rdf:about="alice/me">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
      <foaf:knows rdf:nodeID="x"/>
   </rdf:Description>
</rdf:RDF>
```

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
   <rdf:Description rdf:about="alice/me">
      <foaf:knows rdf:nodeID="x"/>
   </rdf:Description>
   <rdf:Description rdf:about="alice/me">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="x">
      <foaf:name>Charles</foaf:name>
   </rdf:Description>
   <rdf:Description rdf:nodeID="x">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
   </rdf:Description>
</rdf:RDF>
```

2) … and actually a lot of RDF data residing in RDF stores, accessible via SPARQL endpoints already, rather than in RDF/XML

Our approach: XSPARQL (W3C submission, but not yet a standard)

- New query language… but don't reinvent!

  XQuery + SPARQL = **XSPARQL** [Akhtar et al. 2008]

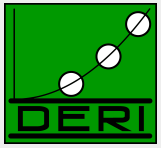
```
<relations>
{ for $Person $Name
  from <relations.rdf>
  where { $Person foaf:name $Name }
  order by $Name
  return
       <person name="{$Name}">
       {for $FName
        from <relations.rdf>
        where {
              $Person foaf:knows $Friend .
              $Person foaf:name $Name .
              $Friend foaf:name $Fname }
        return <knows>{$FName}</knows>
    } </person>
}</relations>
```

```
<relations>
  <person name="Alice">
    <knows>Bob</knows>
    <knows>Charles</knows>
  </person>
  <person name="Bob">
    <knows>Charles</knows>
  </person>
  <person name="Charles"/>
</relations>
```

Conclusions & Outlook (Where's the AI here?):

- Standards (RDF, OWL, SPARQL) are needed to enable structured querying about Web data. Wide adoption already:
  - **RDF** is becoming a ubiquitous standard
  - Lightweight **OWL2** ontologies (FOAF,SIOC, GoodRelations, etc.) emerging
  - Lots of interesting datasets out there! (incl. Twitter, product descriptions/reviews)
  - **SPARQL** becoming quite popular as well, **RIF** to be seen
  - All these standards have clean formal foundations
- BUT:
  - Still not enough data out there
  - Still open KR problems on the border between standards (DL vs. LP vs. Query Languages)
  - Data is not clean (needs AI methods! e.g.: para-consistent reasoning? Ontology matching, NLP, IM/IR,etc.)
  - Query Optimisation in open federated environment is still barely understood, particularly combined with ontological inference.
  - Still a lot to be done ☺

More challenges, interesting pointers:



(not in the paper)

New Journal "***Semantic Web – Interoperability, Usability, Applicability***", *IOS Press* http://www.semantic-web-journal.net/

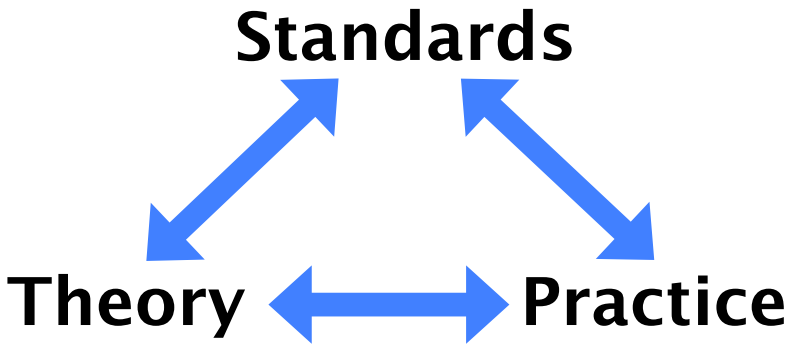
will have some very interesting position papers in its first issue, e.g.:

S. Auer and J. Lehmann. Making the Web a Data Washing Machine - Creating Knowledge out of Interlinked Data. SWJ, accepted for publication, 2010.
http://www.semantic-web-journal.net/content/new-submission-towards-creating-knowledge-out-interlinked-data

P. Hitzler, F. van Harmelen A Reasonable Semantic Web. SWJ, accepted for publication, 2010
http://www.semantic-web-journal.net/content/new-submission-reasonable-semantic-web

A. Polleres, A. Hogan, A. Harth, S. Decker. Can we ever catch up with the Web? SWJ, accepted for publication, 2010.
http://www.semantic-web-journal.net/content/new-submission-can-we-ever-catch-web

Is that the right title?



- Standards: **XML, RDF, OWL, SPARQL, RIF**
- Theory: **Description Logics, Non-monotonic Reasoning, Database Theory**
- Practice/Practically Useful: **Linked Data,** Information Mining?, NLP?

Our approach: XSPARQL (W3C submission)

- New query language… but don't reinvent!
  XQuery + SPARQL = **XSPARQL** [Akhtar et al. 2008]

Prolog:	P	`declare namespace` *prefix*`="`*namespace-URI*`"` or `prefix` *prefix*: <*namespace-URI*>	
Body:	F	`for` *var* `in` *XPath-expression*	
	L	`let` *var* := *XPath-expression*	
	W	`where` *XPath-expression*	
	O	`order by` *expression*	**or**
	F'	`for` *varlist*	
	D	`from` / `from named` <*dataset-URI*>	
	W	`where` {*pattern* }	
	M	`order by` *expression*	
		`limit` *integer* > 0	
		`offset` *integer* > 0	
Head:	C	`construct` { *template (with nested XSPARQL)* }	**or**
	R	`return` *XML + nested XSPARQL*	