Knowledge Management using Semantic Web Technologies

Axel Polleres

NUI Galway
OÉ Gaillimh

The goal of this talk is...

...to introduce & illustrate the potential of
Semantic Web Standard Technologies & **Linked Data**
in Knowledge Management...

... show examples of rapidly increasing adoption

... industrial uptake
- Google, Facebook, BestBuy, BBC, NYT, Cisco, Alcatel-Lucent, etc.

... and still a lot of research challenges!

NUI Galway
OÉ Gaillimh

Enabling **networked** knowledge.

Emerging Enterprise 2.0 technologies promise to help me structure my knowledge...

Do they?

Let's ask someone else...

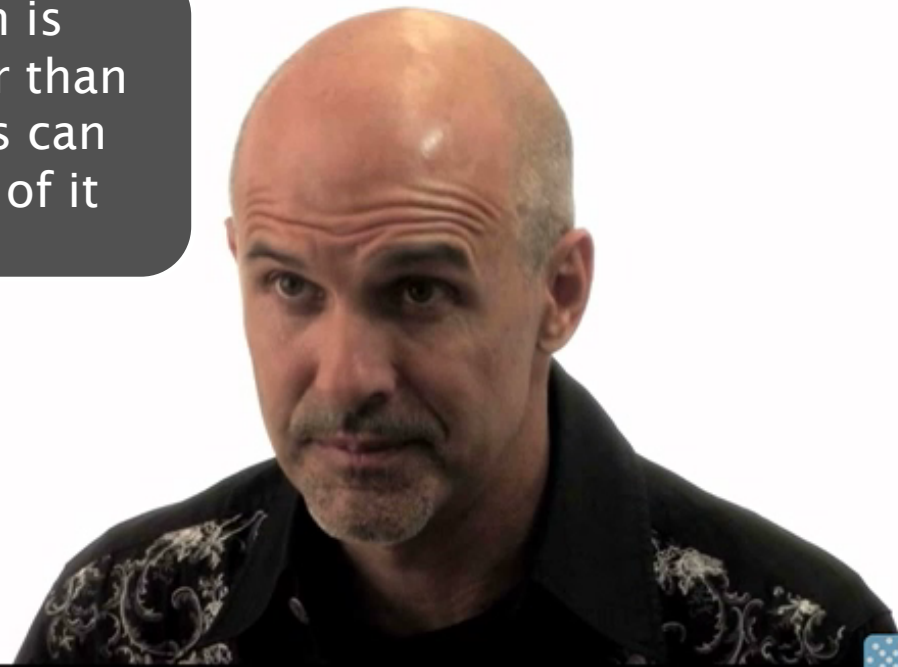Found that on Twitter yesterday, thanks to @kidehen... quite a **bold** statement:

> Information is growing faster than organizations can make sense of it

http://gigaom.com/2010/10/11/jeff-jonas-big-data/

Enabling **networked** knowledge.

- **So, what did we gain with all those tools?**
  - Some limited search functionality +Data silos

Is this Knowledge Management?

Enabling **networked** knowledge.

■ **Some well known models and their current reality...**

> *Sveiby, 1996 - What is Knowledge Management?*
>
> **Knowledge** = objects that can be identified and handled in information systems (*Management of Information*)
>
> **Knowledge** = are processes, a complex set of dynamic skills, know-how, etc, that is constantly changing. (*Management of people*)

Observation: By Web 2.0/Enterprise 2.0 we can observe convergence, ie. most of the processes and also social interactions people involve in, become visible as information items!

Challenge1: We can't access these information **items in a uniform way**: scattered over closed data silos, using different formats (e.g. XML, RDB, specific APIs).
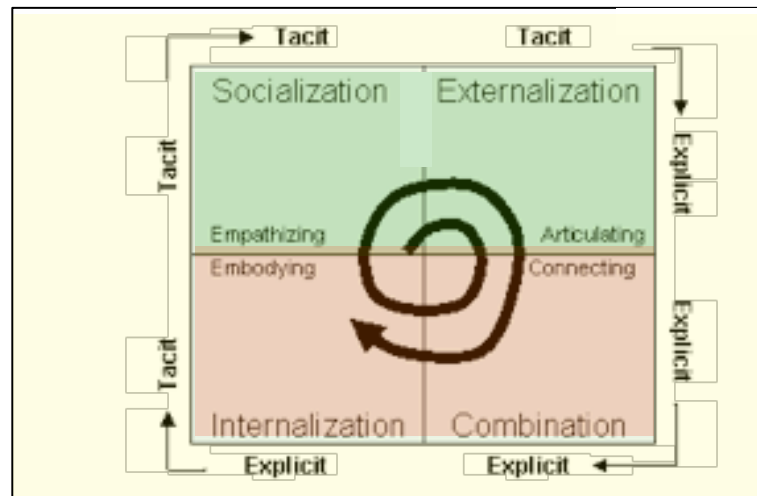
- ## *Nonaka & Takeuchi, 1995 -*

Observation: *By Web 2.0/Enterprise 2.0 we can observe convergence, ie. Socialization and Externalization, that is "spreading" and "publishing" knowledge become one!*

Challenge2: *Assume we have overcome Problem 1, how can we **enrich, use, and share gained knowledge and make it reusable** in possibly unforeseen ways?*
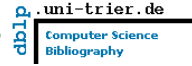
Main challenges

Challenge1: access information items in a uniform way: *scattered over closed data silos, using different formats.*

Challenge2: Enrich, use, and share gained knowledge *and make it reusable in possibly unforeseen ways?*

Enabling **networked** knowledge.

NUI Galway
OÉ Gaillimh

- **Use case 1:** Almost "Traditional" Web Search: "*Find Organizations active in Knowledge Management?*" Google ? bing ? *Not quite…*

- **Use case 2:** Intranet + Web Data: DERI + dblp.uni-trier.de Computer Science Bibliography

  "*My colleagues just told me that another colleague in DERI had an interesting paper about "**blogger analysis**"… but I don't remember his name and. How to find the colleague and his telephone number?*"

Enabling **networked** knowledge.
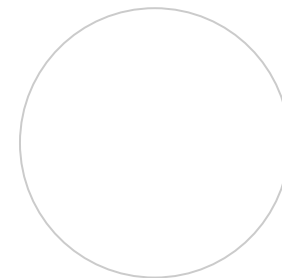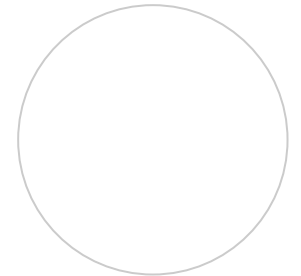
Linked Data to the rescue!

Idea: rely on widely deployed existing infrastructure

☐ **Strong Standards** (standard **Web** technologies)

☐ **Simple Principles**

(+ Advanced Technologies)

NUI Galway
OÉ Gaillimh

Enabling **networked** knowledge.
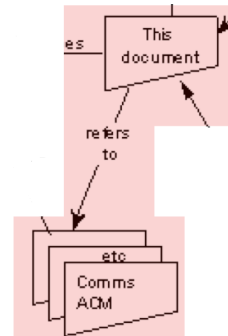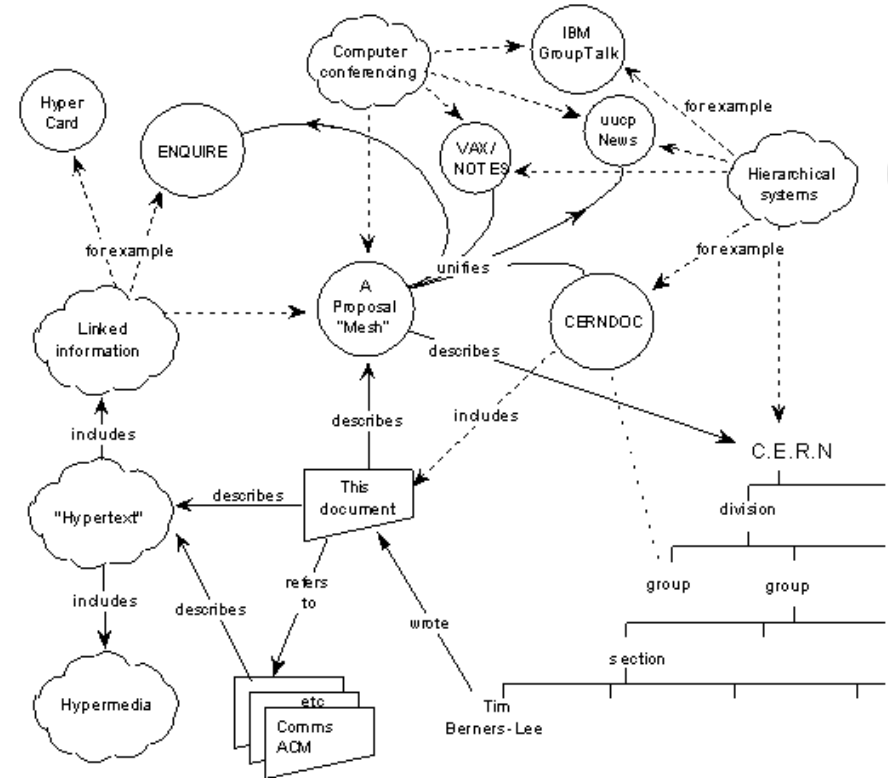
- Globally Unique identifiers **URIs**
- Links between Documents (href)
- A common protocol **HTTP**

The original idea had more….

"This proposal concerns the **management of general information** about accelerators and experiments at CERN […] based on a **distributed hypertext system**. "
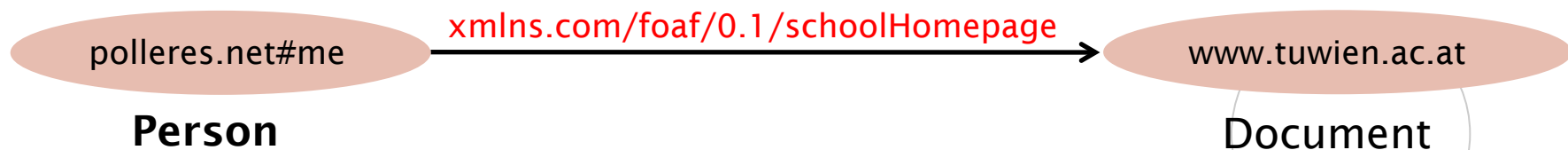
Enabling **networked** knowledge.

NUI Galway
OÉ Gaillimh

The Web **of Data 2011**...

- Globally Unique identifiers   URIs
- **Typed** Links between **Entities**   RDF
- A common protocol   HTTP

Axel Polleres' Personal Web Page

http://www.polleres.net/   rdfa tutorial

Axel Polleres' Personal Web Page   +

`<p about="#me">I studied <a rel="foaf:schoolHomepage" href="http://www.tuwien.ac">here</a></p>`

Done   RDFa   Tor Disabled

polleres.net#me   xmlns.com/foaf/0.1/schoolHomepage   www.tuwien.ac.at

Person   Document

Enabling **networked** knowledge.
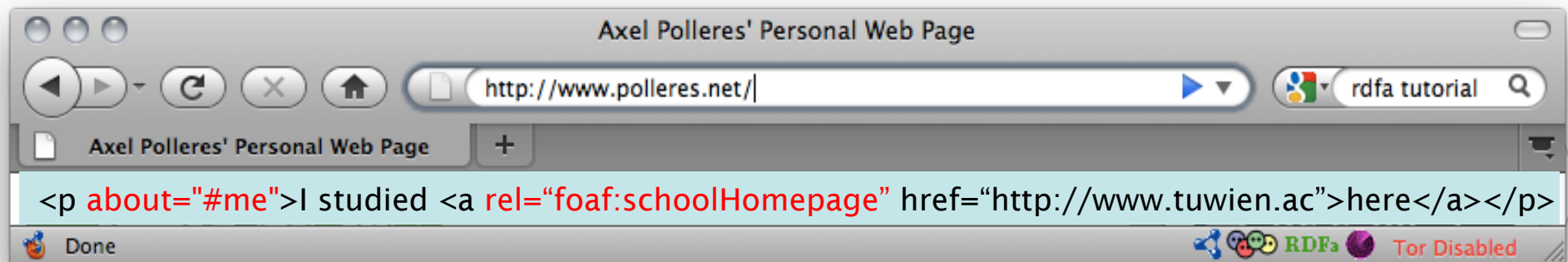
Linked Data Principles

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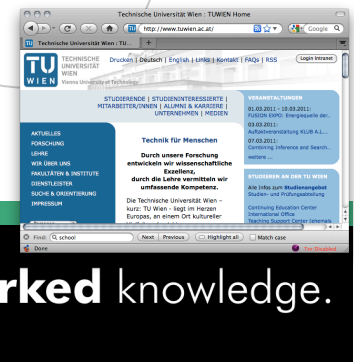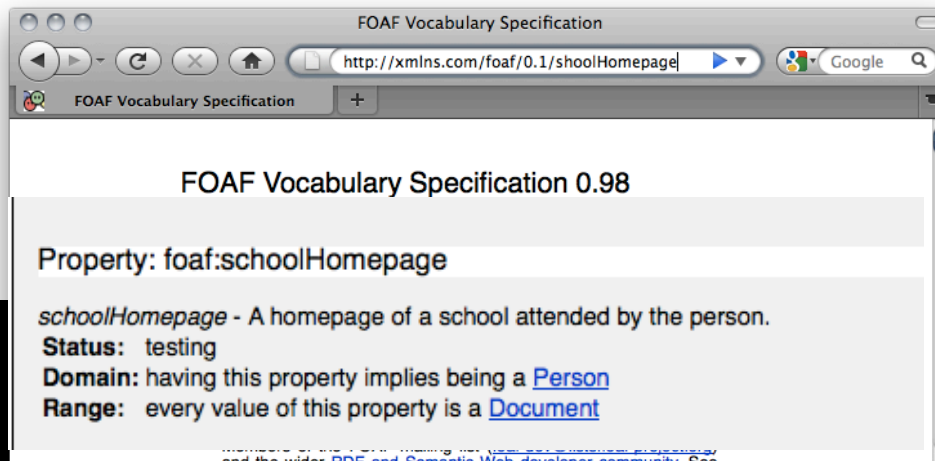2. These URIs are linked via RDF describing relations
3. Relations are URIs again (e.g. :name)
4. **When I dereference the URIs, I should find more information about them, defining them.**

polleres.net#me

xmlns.com/foaf/0.1/schoolHomepage

www.tuwien.ac.at

FOAF Vocabulary Specification

http://xmlns.com/foaf/0.1/shoolHomepage

FOAF Vocabulary Specification 0.98

Property: foaf:schoolHomepage

schoolHomepage - A homepage of a school attended by the person.
Status: testing
Domain: having this property implies being a Person
Range: every value of this property is a Document

networked knowledge.

Sep. 2010

As of September 2010 (cc)

Linked Data Ontologies =
RDF Vocabularies (OWL, RDFS)
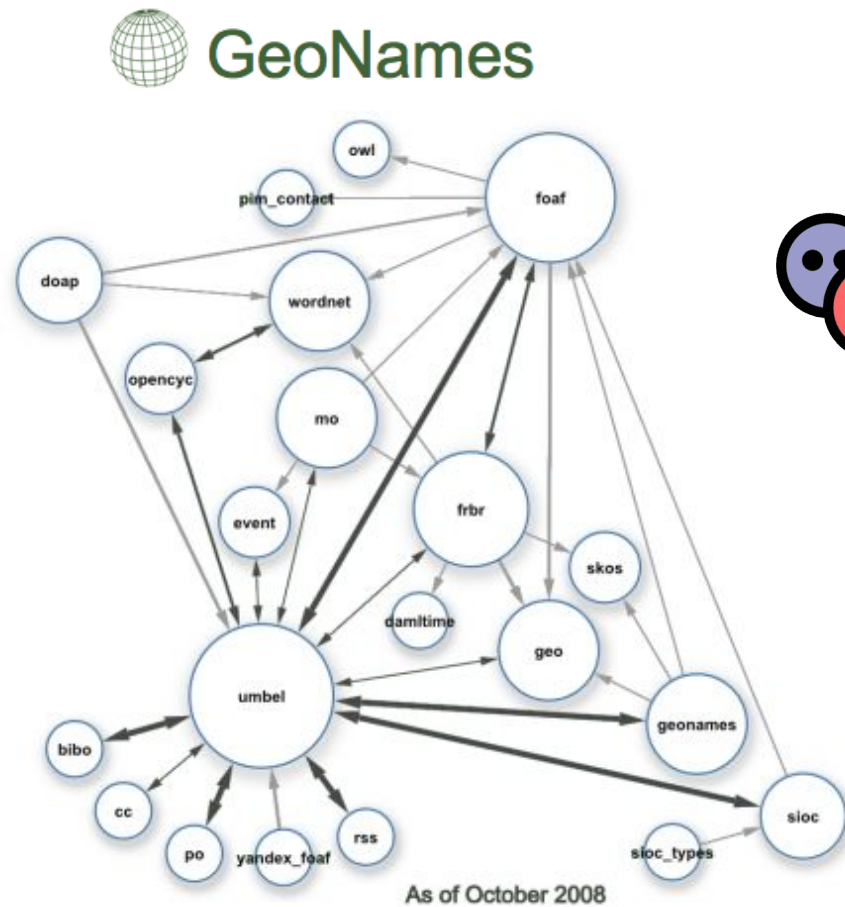
As of October 2008
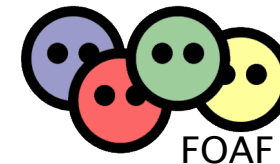
Image from http://blog.dbtune.org/public/.081005_lod_constellation_m.jpg:; Giasson, Bergman

Example: *Find Organizations active in Knowledge Management?*

RDF *addresses Challenge1: access information items in a uniform way*

Subject	Predicate	Object
www.deri.ie	type	o1:SciOrg
www.deri.ie	ont:name	"Digital Ent…"
www.deri.ie	ont:topic	o1:KMngmt
www.know.at	type	o1:SciOrg
www.know.at	ont:label	"KNOW Cen…"
www.know.at	ont:topic	o1:WissMngmt
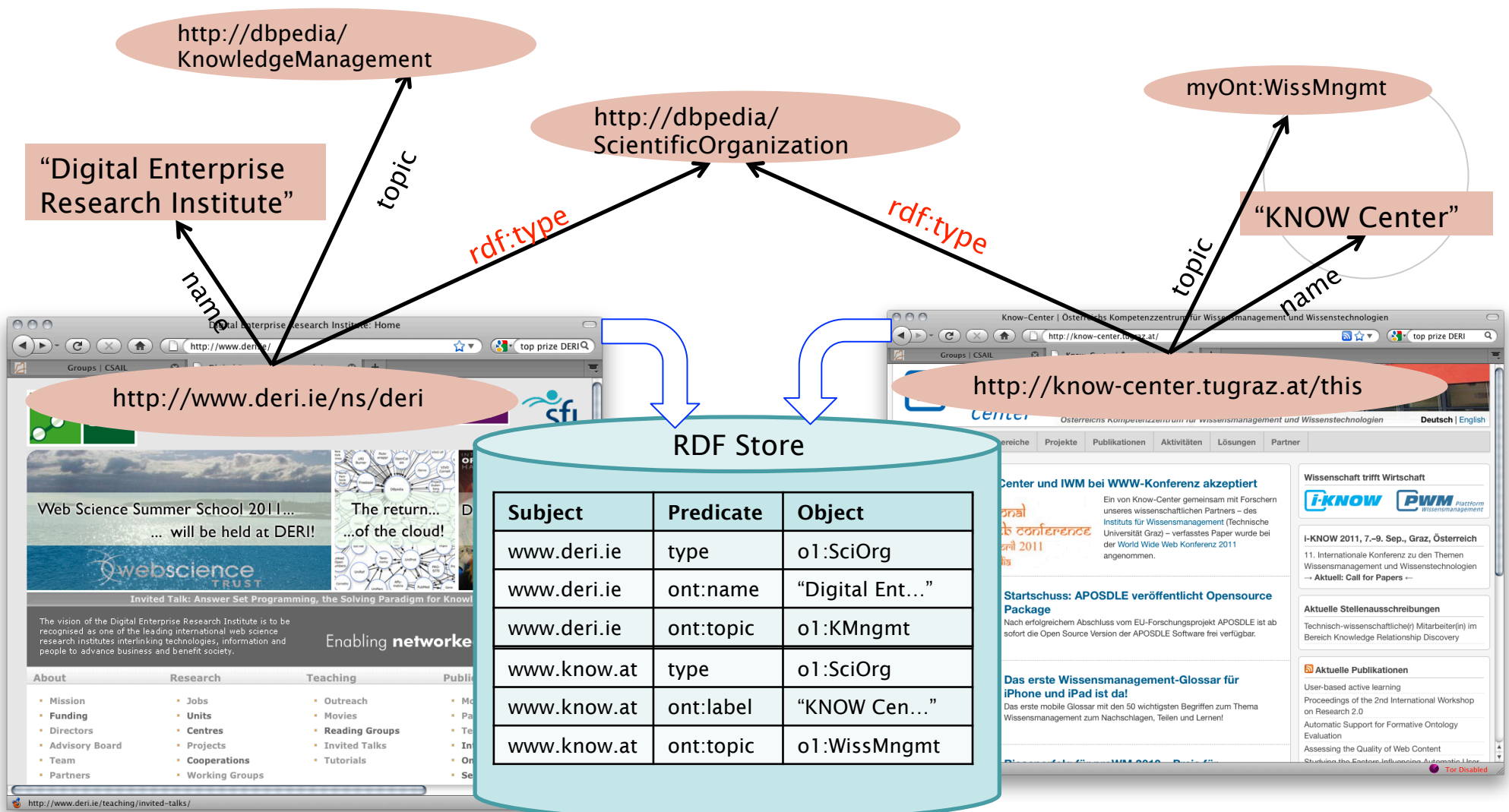
Linked Data "organizes itself"!

FOAF-Ontology

Organization

subclassOf

MyOntology

sameAs

myOnt:WissMngmt

KnowledgeManagement

"Digital Enterprise Research Institute"

ScientificOrganization

topic

rdf:type

rdf:type

rdf:type

rdf:type

topic

"KNOW Center"

name

topic

name

http://www.deri.ie/ns/deri

http://know-center.tugraz.at/this
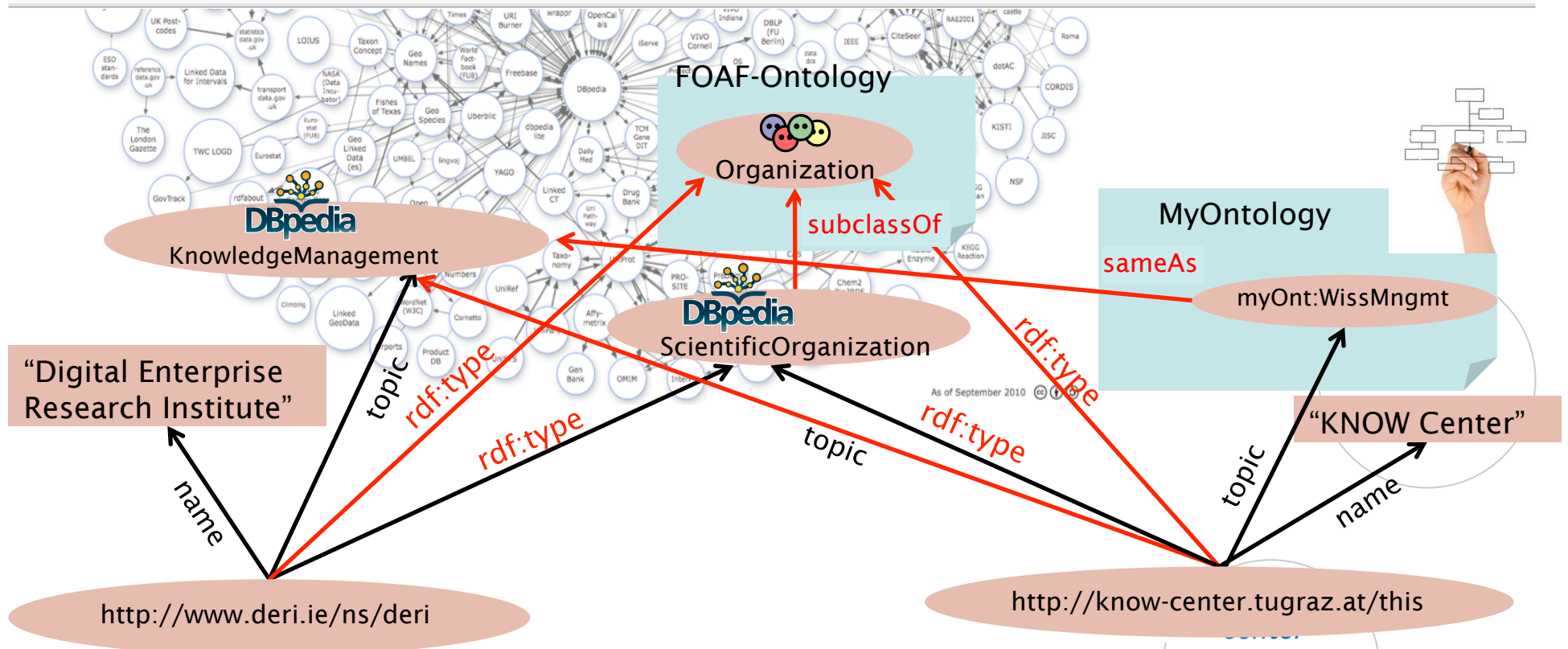
- **Socially:** by **Reuse** of links and URLs
- **Automatically:** by **inference** using ontologies (OWL) semantics!

Enabling **networked** knowledge.

Example: *Find Organizations active in Knowledge Management?*

Linked Data + **OWL** *+* **SPARQL** *address Challenge2: Enrich, use, and share gained knowledge and make it reusable in possibly unforeseen ways*
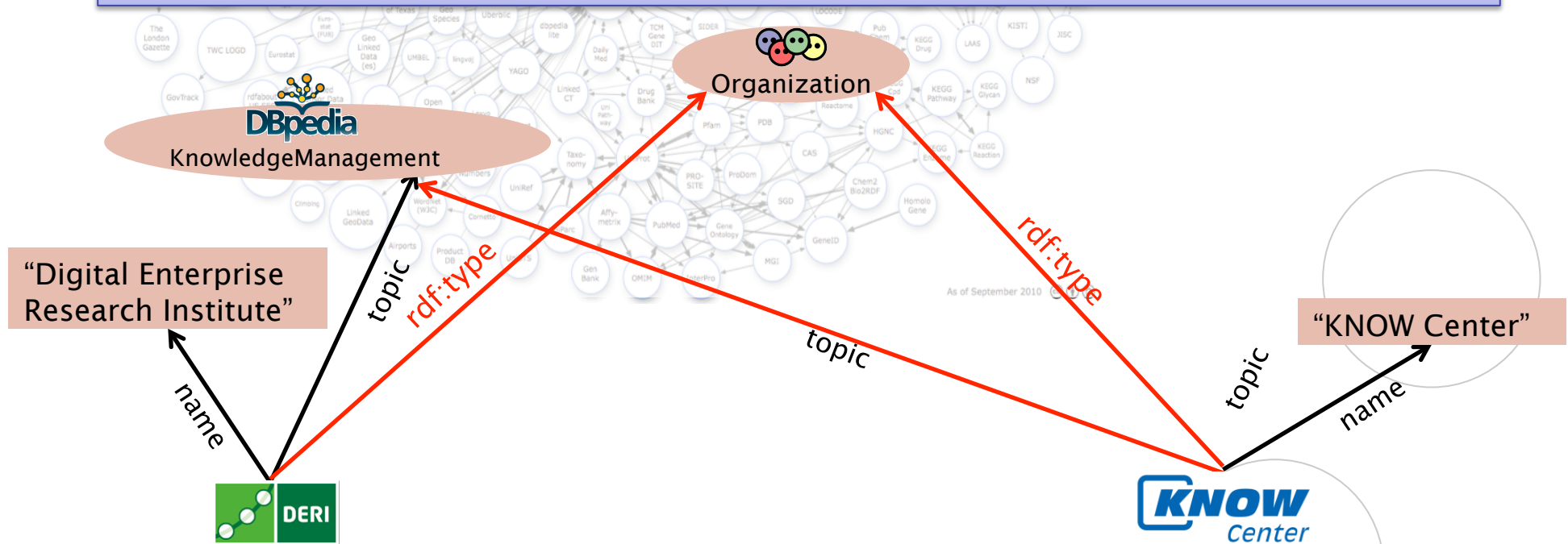
SPARQL `SELECT ?X WHERE { ?X rdf:type foaf:Organization .`
`?X foaf:topic dbpedia:KnowledgeManagement . }`

RDF Store

Enabling **networked** knowledge.

NUI Galway
OÉ Gaillimh

Semantic Web Standards provide the technical basis we need

DERI

Digital Enterprise Research Institute

RIF — **Rules**

{ X lengthInFt **LF** } AND
LM = LF * 0.3048
→
{ X lengthInM **LF** }

OWL — **Ontologies**

Classes   **Attributes**

subclassOf
= subPropertyOf
sameAs

Relations

W3C

RDF

enrich

describe

Publish/
retrieve

Web — URIs/HTTP

HTTP Server guidelines
META-Links
Embedding RDF in e.g. HTML
SPARQL Endpoints

As of September 2010

Applicability

Web Data, Cities, Enterprises, Personal Desktop, **strong industry interest!**

SPARQL1.1 — **Query**

SELECT ?X
WHERE
{?X a foaf:Organization.
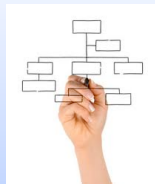?X foaf:topic dbpedia:KM.}

query

+ off-the-shelf components:

- Standard Web Servers
- Crawlers
- RDF Stores
- Query Engines
- Inference Engines

Semantic Web Standards provide the technical basis we need

RIF

{ X length
LM = LF
→
{ X leng

Applicability

...terprises,
...rong

OWL

Classes
subclass
su
sam
⊂ =

Relat

1 **Query**

...anization.
...bpedia:KM.}

...mponents:
...o Servers
...es
...gines

...d knowledge.

INDUSTRY NEWS, NEWS, PUBLISHING

Cisco Signs a €500,000 Deal with Ireland's DERI

By Angela Guess on February 25, 2011 11:45 AM

Cisco has signed a deal with the Digital Enterprise Research Institute at Irish university NUI Galway "to create the next wave of enterprise social networking tools for the workplace of tomorrow." The deal is estimated at €500,000: €400,000 for the further development of Cisco Quad, an "enterprise social networking and collaboration platform," and €100,000 for a strategic research project entitled "Advances in Real-Time Data Integration."

Murali Sitaram, vice president and general manager of Enterprise Collaboration at Cisco, commented, "DERI is a recognized global leader in the semantic web... And this partnership with DERI enables us to accelerate our research and development in this area, enabling Quad as a social semantic platform for the enterprise."

The article continues, "DERI software engineers, based at Cisco's research and development facility in Oranmore, Galway, will use semantic search and integration technology to link information in more intelligent and useful ways to improve communication and collaboration within companies. The two-year contract builds on existing research agreements between the two organizations."

Representatives from Cisco will be speaking on Using Semantic Technology to Revolutionize Traditional Business Models at SemTech 2011 in June.

Image: Courtesy Cisco

Semantic Web Technologies vs. Knowledge Management Challenges

Challenge1: access information items in a uniform way: *scattered over closed data silos, using **different formats**.*

RDF   +   URIs (+ HTTP)   =

addresses Challenge 1 … partially:

- ☐ *Open Question… How to get structured non-RDF data into RDF?*

Challenge2: Enrich, use, and share gained knowledge *and make it reusable in possibly unforeseen ways?*

SPARQL1.1
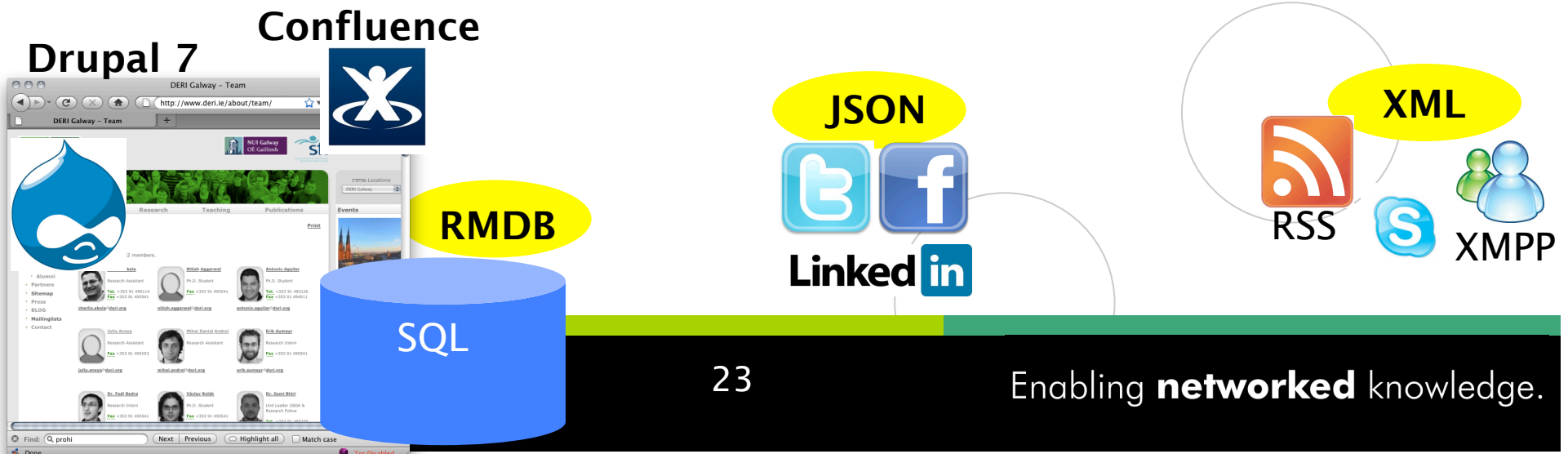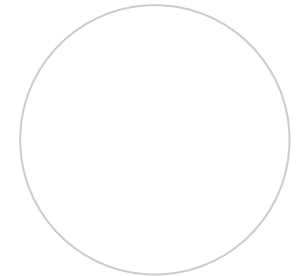RIF   OWL   ■   address Challenge 2 … partially:

- ☐ *Open Question… How to build scalable infrastructures and applications that leverage these technologies?*

Enabling **networked** knowledge.

NUI Galway
OÉ Gaillimh

How to get structured non-RDF data into RDF?

NUI Galway
OÉ Gaillimh

Enabling **networked** knowledge.

- **RDF** processors available off-the shelf, scale reasonably well, but existing systems don't **provide** nor **consume** RDF readily

- Change those systems? No! … Prohibitively expensive

  → Fetch users/developers **where they are**:

  1. Cater for their **formats**
  2. Cater for their **tools**

Drupal 7

Confluence

RMDB

SQL

JSON

Linked in

XML

RSS

XMPP

23   Enabling **networked** knowledge.

- Our approach: A unified Query/Transformation Layer…

  □ Language & Engine to transform between any of these formats

  □ **[ESWC2008]**

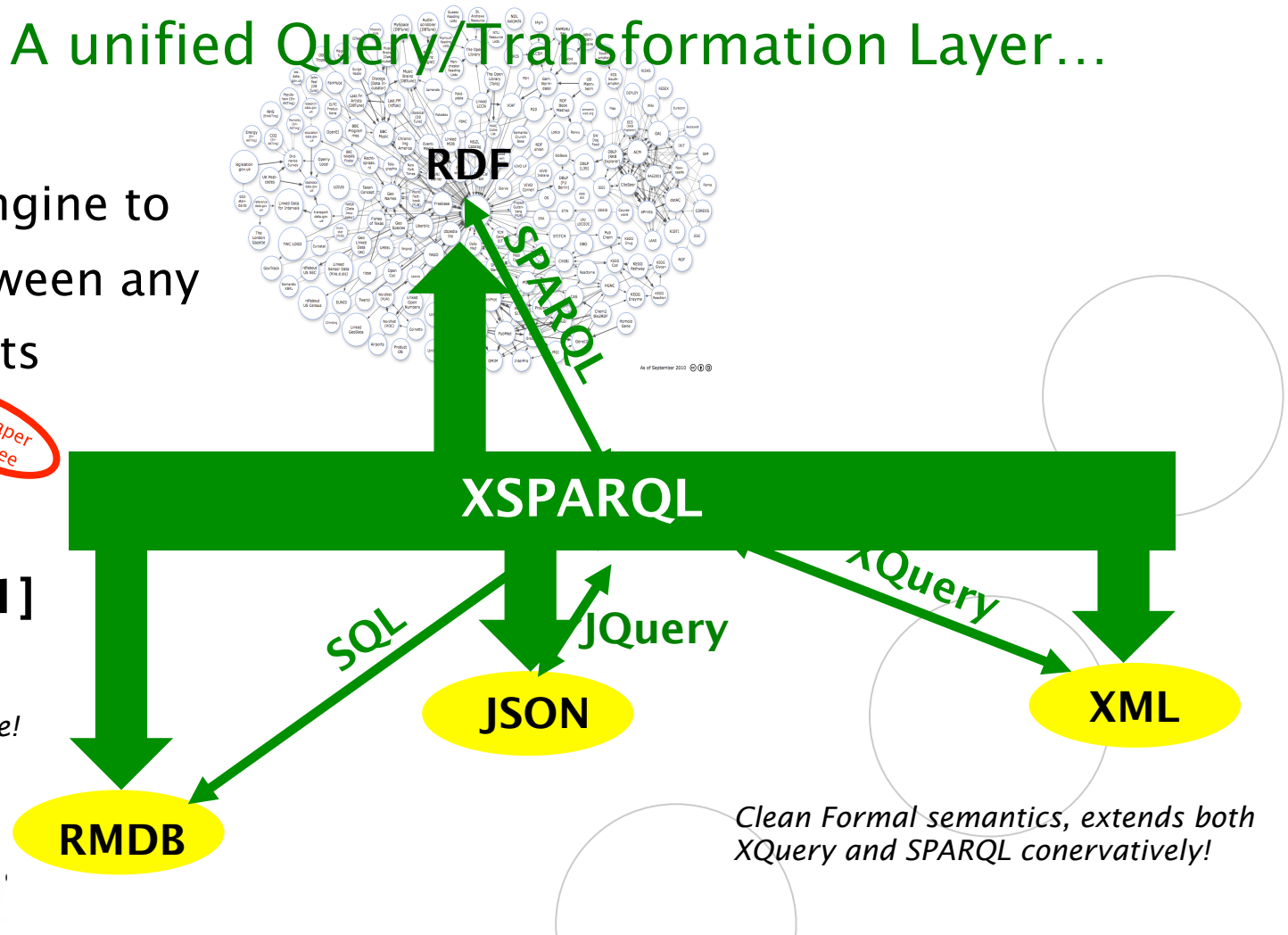   *Best Paper Nominee*

    W3C **member submission**

    [SemTech2011]

    New version soon available on sourceforge!
    http://xsparql.deri.org

    Industry partner:
    Alcatel·Lucent
    Bell Labs

RDF

SPARQL

XSPARQL

SQL   **JQuery**   XQuery

JSON   **XML**

RMDB

Clean Formal semantics, extends both XQuery and SPARQL conervatively!

XSPARQL Example: SIOC-2-RSS

■ XSPARQL+SIOC enables **customised RSS export**:
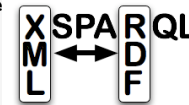
```
<channel>
<title>
        {for $name
         from <http://www.johnbreslin.com/blog/index.php?sioc_type=site>
         where { [a sioc:Forum] sioc:name $name }
         return $name}
</title>
        {for $seeAlso
         from <http://www.johnbreslin.com/blog/index.php?sioc_type=site>
         where { [a sioc:Forum] sioc:container_of [rdfs:seeAlso $seeAlso] }
return <item>
                {for $title $descr $date
                 from $seeAlso
                 where { [a sioc:Post] dc:title $title ;
                                       sioc:content $descr;
                                       dcterms:created $date
                 return <title>$title</title>
                        <description>$descr</description>
                        <pubDate>$date</pubDate>}
```

RSS2.0

"Great stuff,... I have not seen any SIOC to RSS xslt examples or vice versa" (John Breslin, creator of SIOC)

Enabling **networked** knowledge.

Cater for existing tools: Semantic Drupal

- **Drupal**: One of the most Popular Content Management Systems on Internet and Intranet sites… (estimated +7M Websites)

RDFCCK "One-click-install module" to **export** Linked Data RDF

SPARQL_EP "One-click-install module" to create a **SPARQL Endpoint** for your Drupal Website

SPARQLViews: dynamically query external SPARQL endpoints to enrich your site.

EVOC: Module to **link to existing ontologies**.

Deployments of some of these modules. e.g. **ISWC2010, FIS2010** websites

Best In-Use Paper Award

[ISWC2009]

Most notable achievement:

Drupal 7 **(out now!) supports RDF natively!**

26

How to build scalable infrastructures and applications on Linked Data?

Enabling **networked** knowledge.
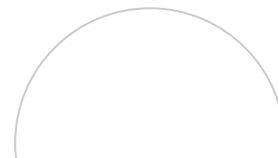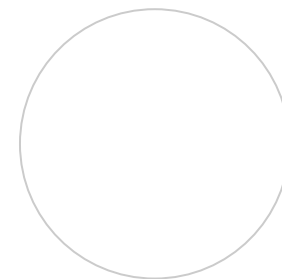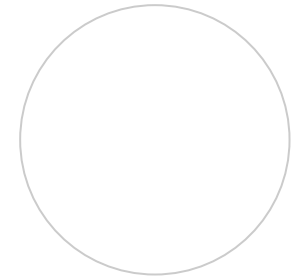
Web scalable inference?
SAOR: *Scalable Authoritative OWL Reasoner*

explicit data

implicit data

SAOR: *Make this implicit knowledge explicit & ready for Search*
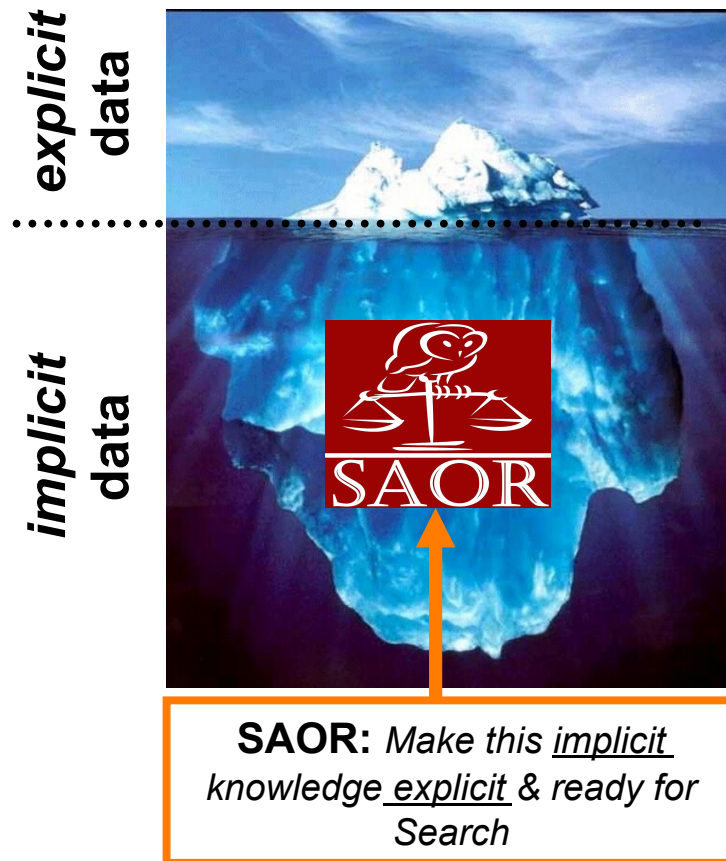
Observation: *Sound & complete OWL Reasoning does NOT scale to the Web!*

Our Goal:

- **Efficient** **rule-based OWL Reasoner,** implements **lightweight/scalable** subset of **OWL2RL** standard

- **Distributed**

- **Robust:** determines trustworthiness of Web documents using Linked Data/Web Architecture principles

- *Field tested on real Linked Data*

- *Evaluated reasoning over ~1 billion RDF facts from 4 million Web documents*

[IJSWIS2009,ISWC2010,Hogan2011]

Enabling **networked** knowledge.

1. Extract **Tbox** (i.e. Ontologies) from Data in memory

2. Eliminate *unauthoritative* Tbox

3. Encode OWL in **Datalog** rules that don't need ABox-joins

$C \sqsubseteq D$	?C rdfs:subClassOf ?D . ?s a ?C . $\Rightarrow$?s a ?D .
$C \equiv D$	?C :equivalentClass ?D . ?s a ?C . $\Rightarrow$?s a ?D .
	?C :equivalentClass ?D . ?s a ?D . $\Rightarrow$?s a ?C .
$P \sqsubseteq Q$	?P rdfs:subPropertyOf ?Q . ?s ?P ?o . $\Rightarrow$?s ?Q ?o .
$P \equiv Q$	?P :equivalentProperty ?Q . ?s ?P ?o . $\Rightarrow$?s ?Q ?o .
	?P :equivalentProperty ?Q . ?s ?Q ?o . $\Rightarrow$?s ?P ?o .
$P \equiv P_0^-$	?P :inverseOf ?Q . ?s ?P ?o . $\Rightarrow$?o ?Q ?s .
	?P :inverseOf ?Q . ?s ?Q ?o . $\Rightarrow$?o ?P ?s .
$\top \sqsubseteq \forall P^-.C$	?P rdfs:domain ?C . ?s ?P ?o . $\Rightarrow$?s a ?C .
$\top \sqsubseteq \forall P.C$	?P rdfs:range ?C . ?s ?P ?o . $\Rightarrow$?o a ?C .
$P \equiv P^-$	?P a :SymmetricProperty . ?s ?P ?o . $\Rightarrow$?o ?P ?s .
$\exists P.x$	?C :hasValue ?x; :onProperty ?P . ?y ?P ?x . $\Rightarrow$?y a ?C .
	?C :hasValue ?x; :onProperty ?P . ?y a ?C . $\Rightarrow$?y ?P ?x .
$C_1 \sqcup ... \sqcup C_n$	?C :unionOf (?C$_1$...?C$_i$...?C$_n$) . ?x a ?C$_i$. $\Rightarrow$?x a ?C .
$(\geq 1P)$	?C :minCardinality 1; :onProperty ?P . ?x ?P ?y . $\Rightarrow$?x a ?C .
$C_1 \sqcap ... \sqcap C_n$	?C :intersectionOf (?C$_1$ ... ?C$_n$) . ?y a ?C . $\Rightarrow$?y a ?C$_1$, ..., ?C$_n$.
$C_1 \sqcap ... \sqcap C_n$	?C :intersectionOf (?C$_1$) . ?y a ?C$_1$. $\Rightarrow$?y a ?C .

$\leftarrow$ Distributable
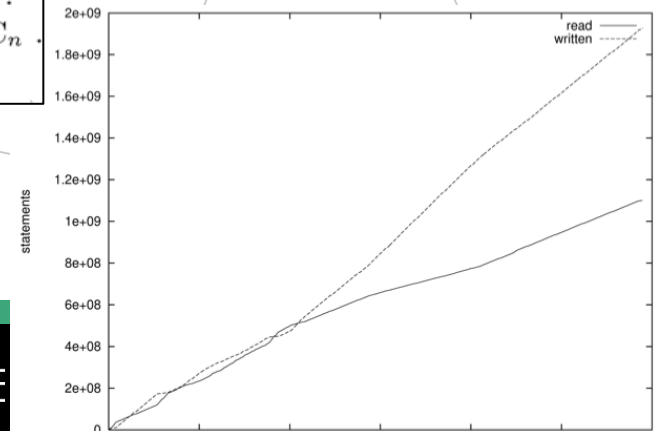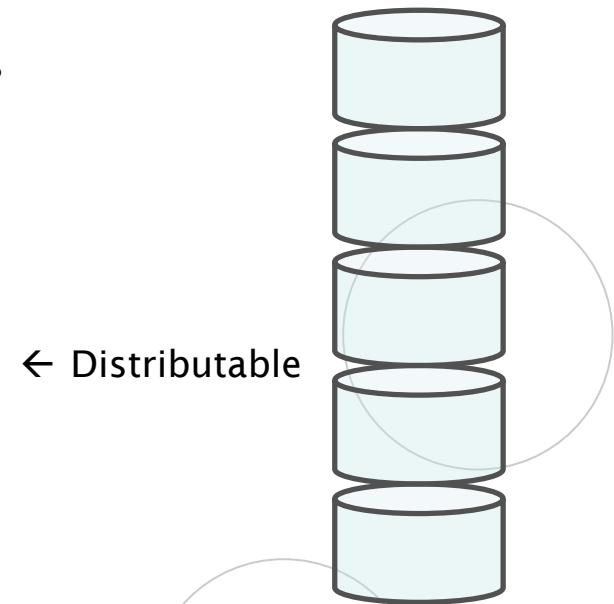
4. Run rules **linearly** over non-ontology RDF Data (Abox)

[IJSWIS2009,ISWC2010,Hogan2011]

29

Consolidation: *Find out when different Web sources are speaking about the **same *thing* (**document, book, person, event...) and **"join the dots"***

- *Reasoning:* use OWL to find new matches – this is not enough!

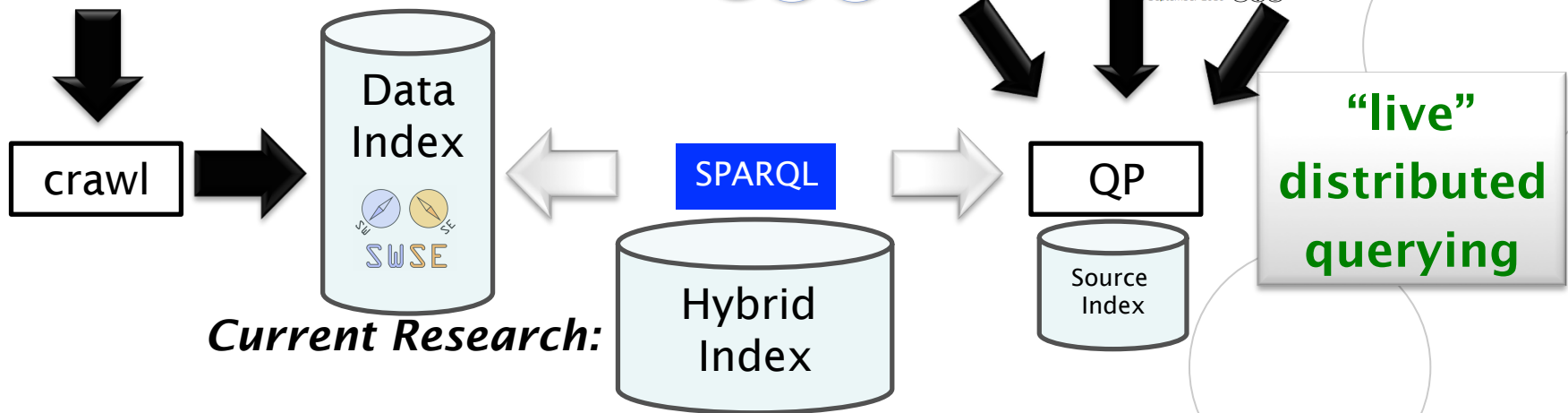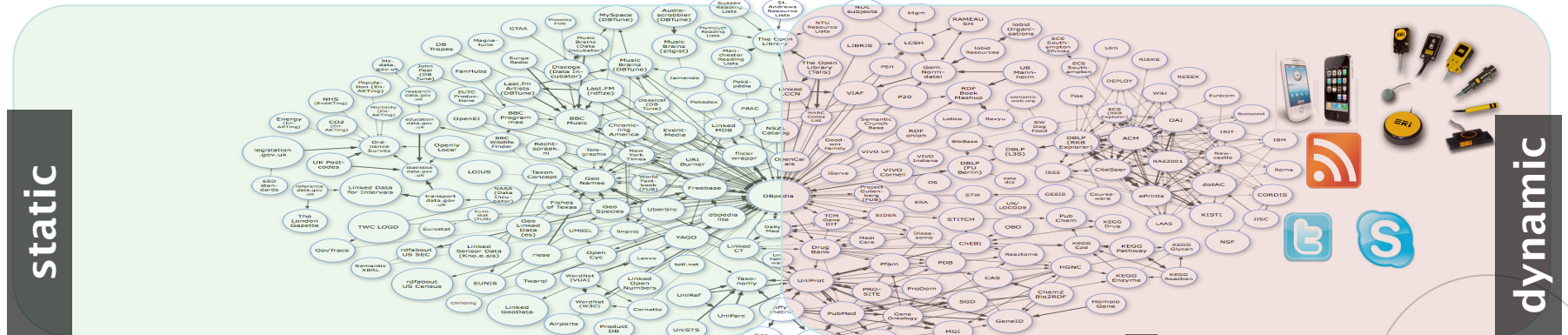- *Needs Statistical methods:* use **patterns** in the data, Idea: exploit shared **discriminating** properties

Combined Approach - *Preliminary results:* **[NeFoRS2010]**

<u>Scalable:</u> **distributed** sorts and scans

<u>~Robust:</u> *takes some measures to ensure correctness of results*

- *Accuracy?... good for reasoning... (statistical? **open question**)*

- *Field tested on real Linked Data*
- *Evaluated reasoning/consolidation over ~1 billion facts from 4 million Web documents*

RDF Store

NUI Galway
OÉ Gaillimh

Enabling **networked** knowledge.

static

dynamic

crawl → Data Index / SWSE

SPARQL

QP / Source Index

"live" distributed querying

Current Research: Hybrid Index
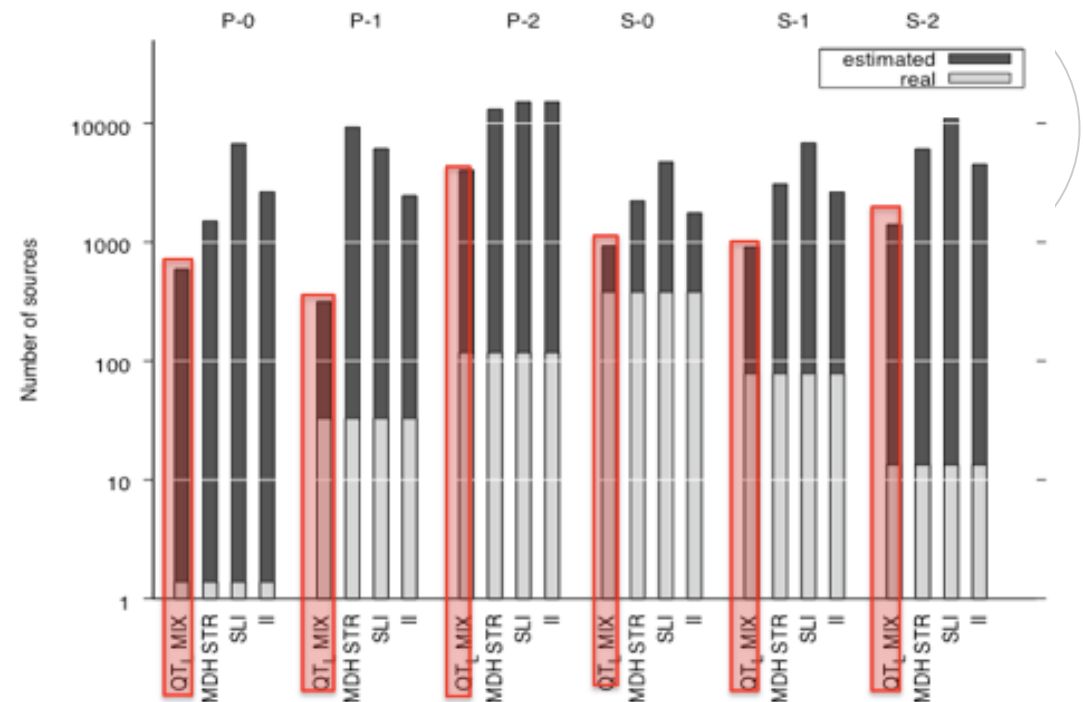
[WWW2010, WWWJ2011]

Hybrid
Index?
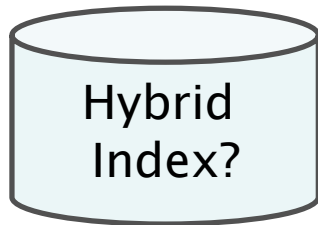
Evaluation of different possible index structures for source selection:
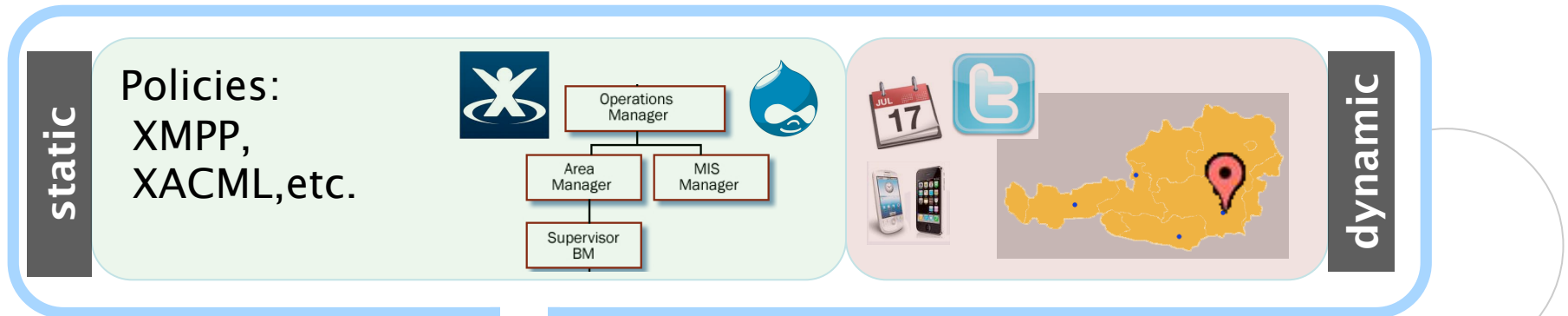
[WWWJ2011]:

- **Inverted URI indices**
- **Schema Level index**
- **Multi-dimensional histograms**
- **QTree**

Next steps:

☐ Data Context (Location, Time, Trust, Provenance) [**AAAI2010,ISWC2010b, FP6-InContext**], Policies….

☐ … and its applications, e.g. Rich Presence [**CollaborateCom2010**]

static

Policies:
XMPP,
XACML,etc.

Operations Manager

Area Manager

MIS Manager

Supervisor BM

dynamic

JUL 17

enrich

What I see:
☐ ✅ Online
☐ 🕐 Away
☑ ❌ Not available
☐ ⛔ Do Not Disturb
☐ ✅ Invisible
☐ ⚪ Offline

What others shall see:
🕐 Out of town (for customers)
⛔ In a meeting (for colleagues)
📞 Available for urgent calls only (for Boss)

NUI Galway
OÉ Gaillimh

33

Enabling **networked** knowledge.

Take home messages

- **As Web 2.0 and the Enterprise are converging**... **information&knowledge become even more unmanageable** ☹

  ... **but:**

- **Semantic Web Standards** & **Linked Data** provide a promising technical basis for Enterprise Knowledge Management:
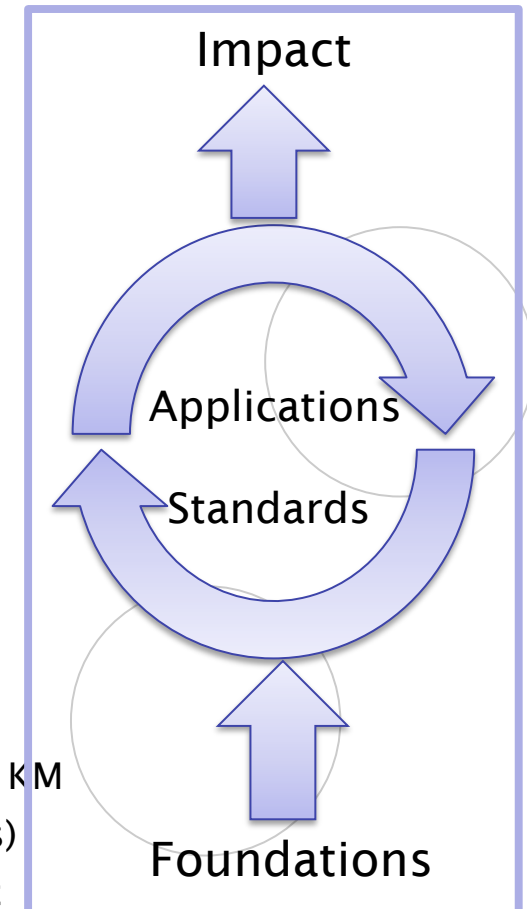 - ☐ **Strong Standards**
  - ☐ Affordable, scalable tools
  - ☐ Allow for Bottom-up deployment

- Promising prospects **for Applied Research**:
  - ☐ Growing Industry Interest (big players)
  - ☐ Potentially new business models (SMEs)

- Urgent needs for **Foundational Research**:
  - ☐ **Web Science** "in its own right" as an emerging discipline rooted in KM
  - ☐ **Social component** of Linked Data (Knowl.Worker+Knowl.engineers)
  - ☐ New ways of thinking needed for **Reasoning & Data Management**

Thank you! Questions?