

Tokyo Linux Users Group
Sun Microsystems, SBS Tower
May 9th, 2009

Semantic Web: Information wants to be useful



Nicolas Raoul – <http://nrw.free.fr>
Working on HTML5 at W3C Tokyo

What Atos Origin says

« What is the most important thing for your organization?
The **information**, or the **system** that manages it? »

« Free your data, don't let I.T. manage it! »

Gautier Poupeau, Data architect

What Oracle says

« But the second big trend then is to **decouple the data from the application** or the application services, so that in that sense what you can do is write your application or **create services independent of the data sources** they have to deal with, which comes full circle back to having a virtual layer between application services and data. »

Robert Shimp, Global Technology Business division vice-president

What the Semantic Web is not

- It will not replace the Web we know. The Semantic Web is for information, the Web has much more than information (style, emotion, advertisements, humor, etc)
- Unlike Web 2.0, it is not a visible revolution from a user's point of view.
- The Semantic Web is not about machines reading English or Japanese text. Some semantic tools do this, but not the Semantic Web.

Goal of the Semantic Web

Allow machines to **easily** use available data

Consequences:

More applications use available data.

Computers become more useful to humans.

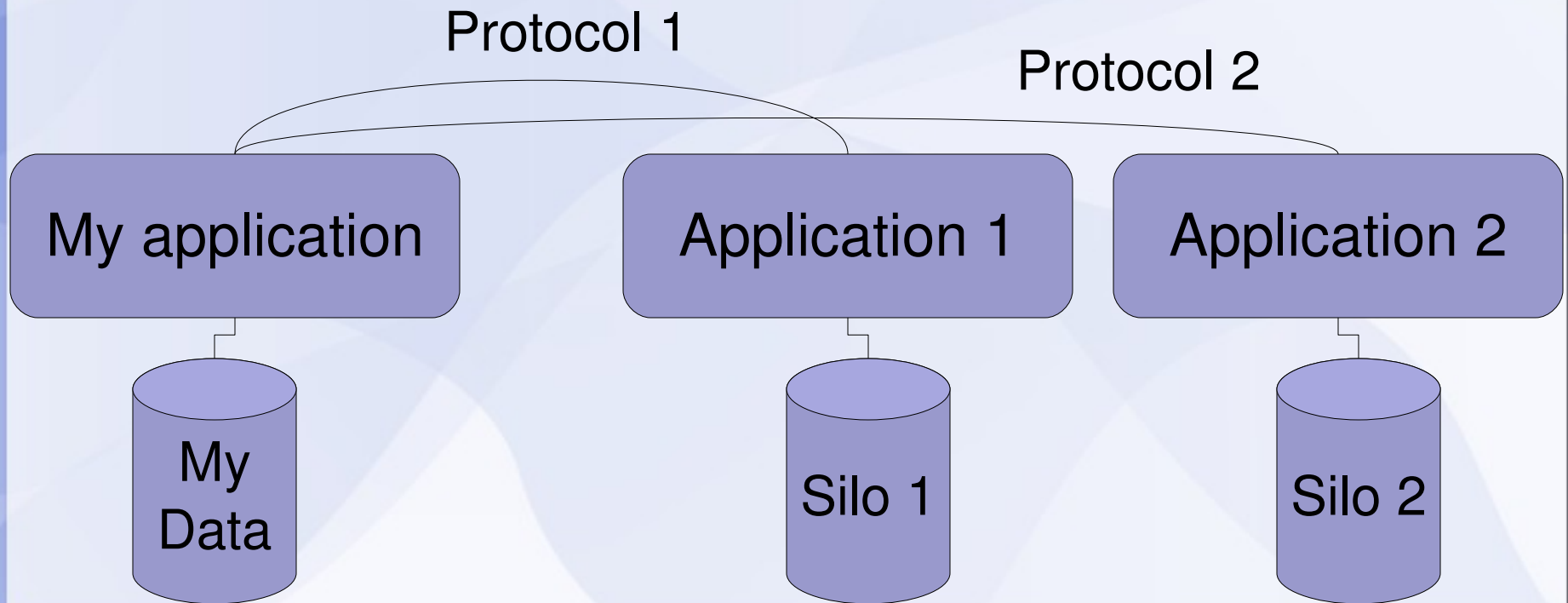
What data is available?

- Molecules
- Maps and places names
- Songs
- Train traffic
- ...

A huge amount of data is already freely available.
Everything that is on the web, basically.

Where is the data I want to use?

Most of today's data is in “data silos”



In the Semantic Web

The world's public data cloud

My application

Application 1

Application 2

My organization's
private data cloud
for confidential data

Who is a data silo?

- FindRestaurant.co.jp: Website interface only.
- LegiFrance: Data in RTF files.
- Google Weather: Web service with specific method names.
- My data: SQL or similar with specific schema names and value types.

Nice promise, but how?

The key is that any information in the world can be expressed as a set of very simple sentences.

Each sentence is called a “triple” :

Subject **Predicate (kind of verb)** **Object**

Examples of triples

Hamlet **is** **a book**

Mosquitoes **are a kind of** **Insect**

I **was born in** **1981**

Stéphane Corlosquet **knows** **me**

Every sentence has this grammar we have seen:

Subject **Predicate** **Object**

Words are ambiguous

Consider this triple:

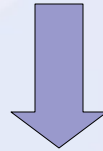
I **interested in** **Java**

- “Java” is ambiguous (island or coffee?).
- “interested in” has a vague meaning.
- “I” will mean something different for everybody.

URIs are not ambiguous

The solution already exists: instead of a word, use a URI (Uniform Resource Identifier)

I interested in Java



<http://nrw.free.fr/data/contact/foaf-nicolas-raoul.rdf#me>

http://xmlns.com/foaf/spec/#term_topic_interest

[http://dbpedia.org/resource/Java_\(programming_language\)](http://dbpedia.org/resource/Java_(programming_language))

Expressing information

A lot of things and verbs have URIs, and we will see how to create new ones, so I can express any information in triples:

Nicolas Raoul interested in **Java**

Nicolas Raoul was born in **1981**

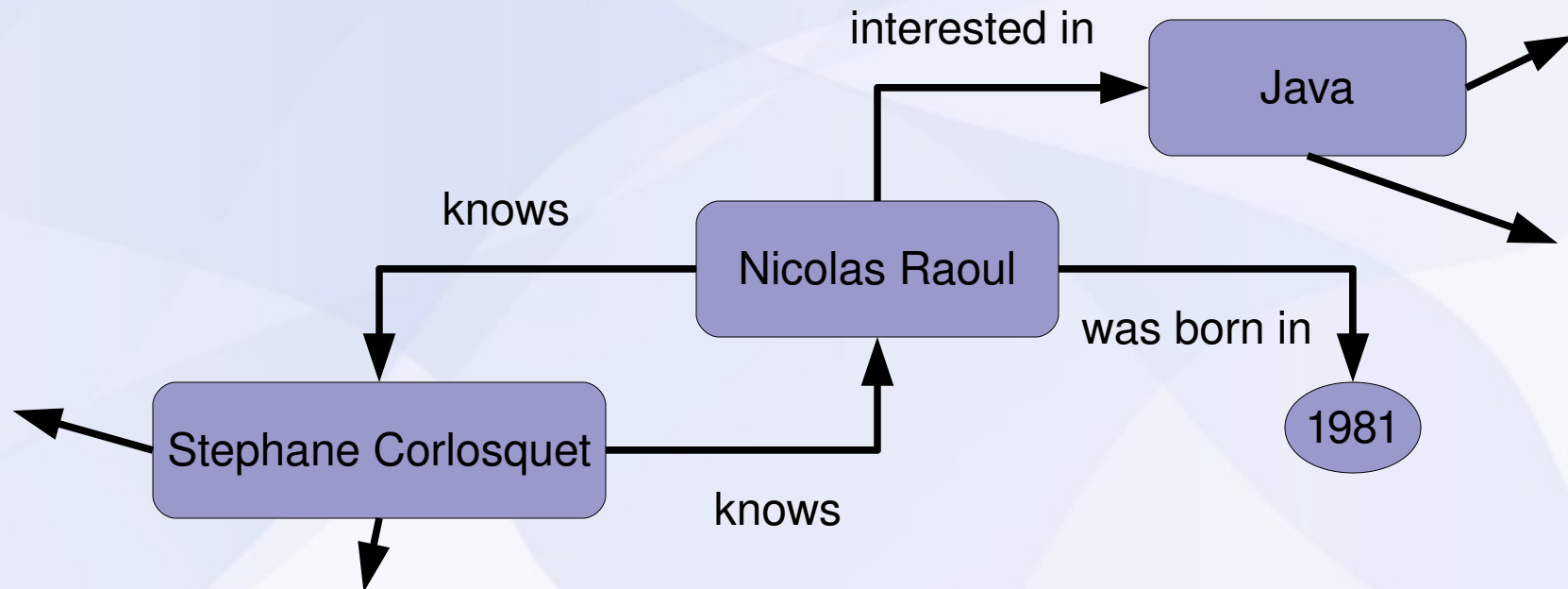
Stéphane Corlosquet knows **Nicolas Raoul**

Stéphane Corlosquet interested in **Java**

This web is semantic

Triples form a web between things.

“Semantic” means URIs remove ambiguity.



N-Triples

The easiest way to express triples is an N-Triple file:

<Subject1> <Predicate1> <Object1>.

<Subject2> <Predicate2> <Object2>.

<Subject3> <Predicate3> value3.

...

RDF/XML

RDF/XML is another way to express triples:

```
<rdf:RDF (namespaces declarations omitted) >
<Person rdf:about="http://nrw.free.fr/data/contact/foaf-nicolas-raoul.rdf">
  <foaf:name>Nicolas Raoul</foaf:name>
  <foaf:dateOfBirth>1981-11-23</foaf:dateOfBirth>
  <foaf:knows>
    <foaf:Person>
      <foaf:name>Stéphane Corlosquet</foaf:name>
      <rdfs:seeAlso rdf:resource="http://www.deri.ie/fileadmin/scripts/foaf.php?id=305"/>
    </foaf:Person>
  </foaf:knows>
</Person>
</rdf:RDF>
```

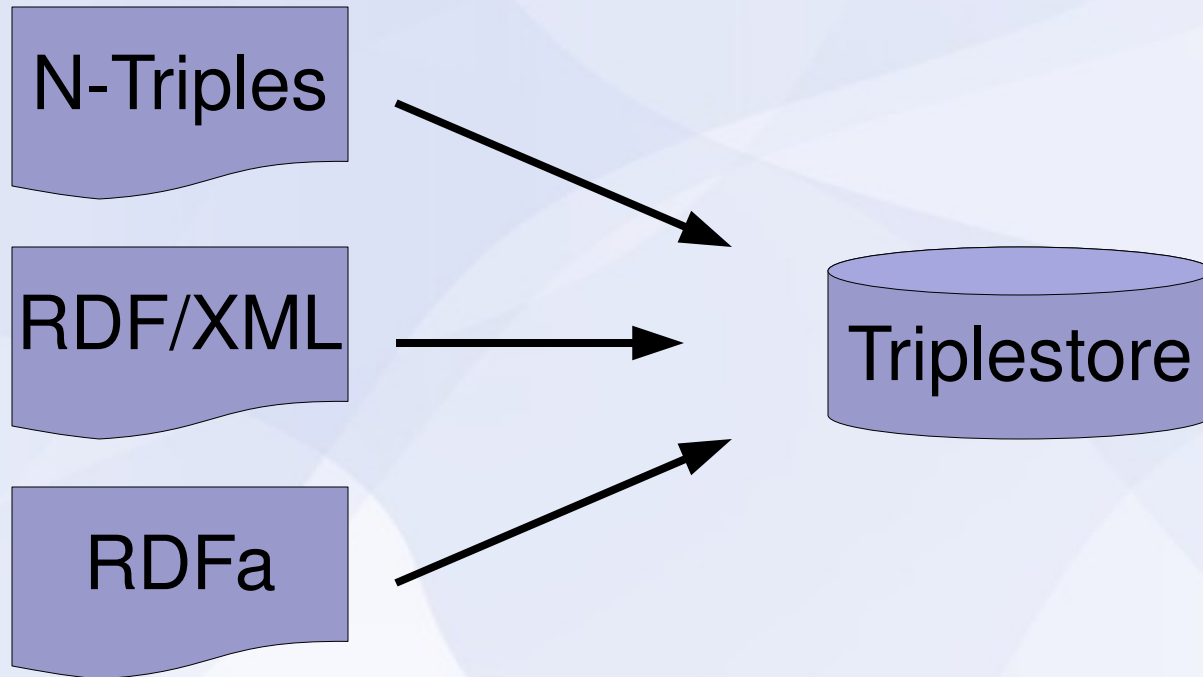
XHTML+RDFa

Express triples within a normal web page:

```
<div xmlns:dc="http://purl.org/dc/terms/"  
  about="urn:isbn:978-1741047882">  
  Famous author  
  <span property="dc:creator">Matthew Firestone</span>  
  had his famous  
  <span property="dc:title">Tokyo City Guide</span>  
  published on  
  <span property="dc:date" content="2008-08-01">August 1, 2008</span>.  
</div>
```

Triplestore

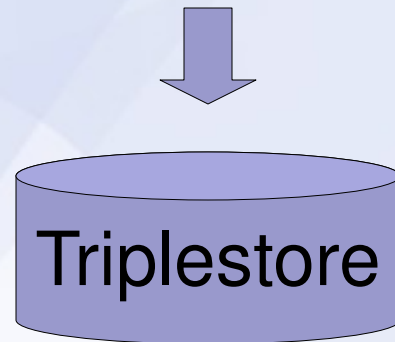
Once expressed, triples are stored in a triplestore.



SPARQL

How to get information from the Semantic Web?
SPARQL: a W3C standard since last year.

Barack Obama was born in <some place>
<some place> has name ?



? = "Honolulu"

SPARQL syntax

```
PREFIX dbpedia:<http://dbpedia.org/resource/>
```

```
PREFIX dbpedia-owl:<http://dbpedia.org/ontology/>
```

```
PREFIX property:<http://dbpedia.org/property/>
```

```
SELECT ?placename
```

```
WHERE {
```

```
  dbpedia:Barack_Obama dbpedia-owl:birthplace ?place .
```

```
  ?place property:name ?placename .
```

```
}
```

People born in Tokyo before 1923

(Prefix declarations omitted)

```
SELECT ?name ?person WHERE {  
  ?person dbpedia2:birthPlace <http://dbpedia.org/resource/Tokyo> .  
  ?person dbo:birthdate ?birth .  
  ?person foaf:name ?name .  
  FILTER (?birth < "1923-01-01"^^xsd:date) .  
}  
ORDER BY ?name
```

What if I cannot find an URI for something?

Search Swoogle and Sindice, ask in your field's semantic community.

If nothing exists yet, you will have to create an URI for the thing you want to signify.

For instance, I was not in Freebase nor anywhere, so I decided to signify myself by this URI:
<http://nrw.free.fr/data/contact/foaf-nicolas-raoul.rdf#me>

What is a cool URI?

Cool URIs are short and don't change.

Don't include words that reflect what is behind, because words' meaning can change, or become politically incorrect.

Don't include technology terms such as “.jsp” or “/cgi/”, because it would change sooner or later.

Example: <http://arxiv.org/abs/0904.0564>

See <http://www.w3.org/TR/2007/WD-cooluris-20071217>

What if I cannot find the appropriate predicate?

Predicates (and what they apply to) are defined by ontologies.

If your company deals in nattō, and no ontology exists yet to describe all subtleties of the nattō industry, then you will have to create a new ontology for this field.

The ontology will be useful if everybody uses it, so you should:

- Seek feedback from your field, including competitors.
- Build a community around this ontology.

Writing an ontology is a bit similar to writing a hierarchy of classes and attributes in Object-Oriented Programming.

What if two organizations defined the same thing?

Sometimes two URIs signify the exact same thing.
For instance:

<http://dbpedia.org/resource/Carbon>

<http://sw.opencyc.org/2008/06/10/concept/Mx4rvVjIQpwpEbGdrcN5Y29ycA>

In that case, write a triple with the “sameAs” predicate:

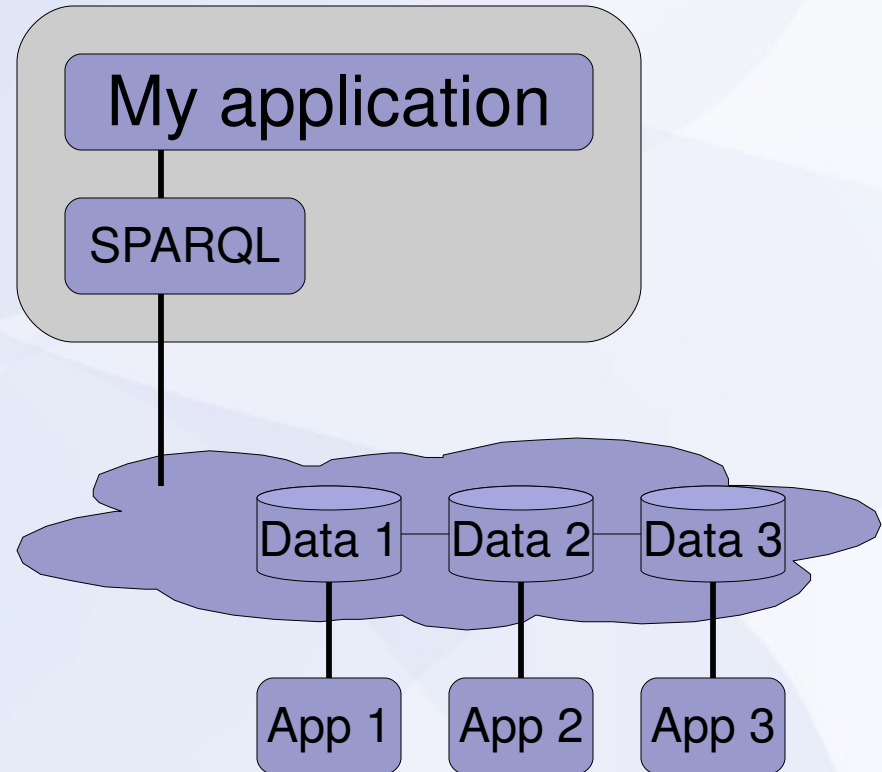
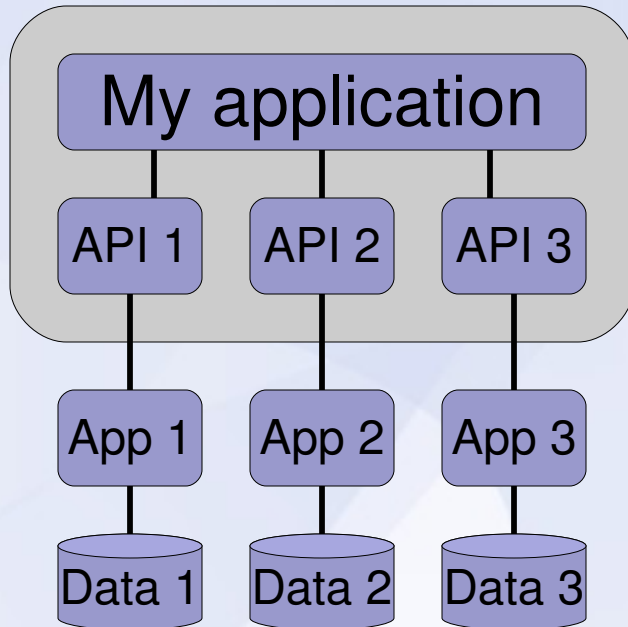
<http://dbpedia.org/resource/Carbon>

<http://www.w3.org/2002/07/owl#sameAs>

<http://sw.opencyc.org/2008/06/10/concept/Mx4rvVjIQpwpEbGdrcN5Y29ycA>

Silos vs. Semantic Web

Specific APIs.
No links between two applications' data.



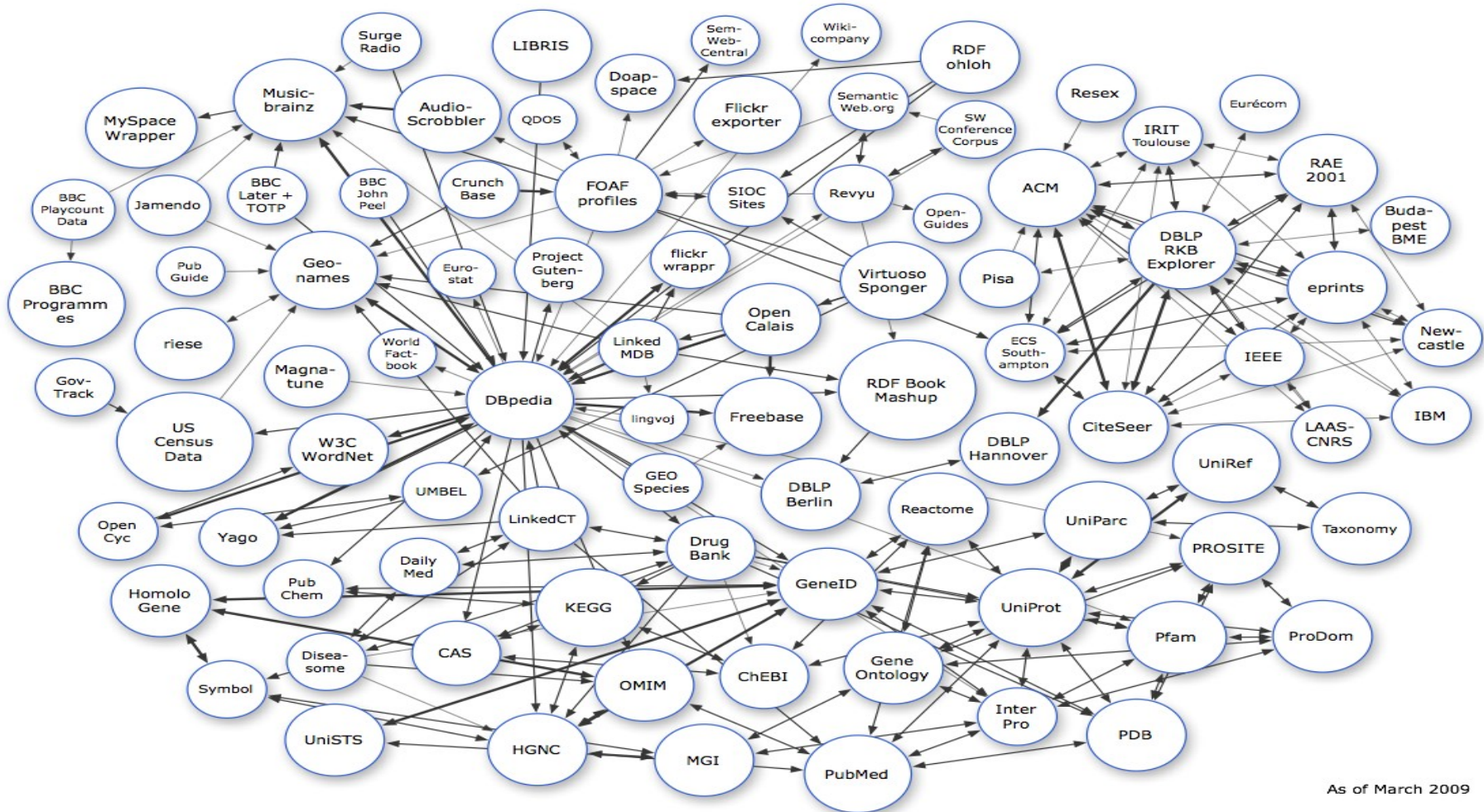
XML vs. Semantic Web

	XML	Semantic Web
Model	Tree	Graph
Links target ...	Other nodes of the same document	Anything (universal)
Best use	Ideal for a document	Ideal for knowledge

Relational DB vs. Semantic Web

	Relational DB	Semantic Web
Beyond the company, your data is ...	meaningless.	Self-describing and universally unambiguous.
Incoming links	cannot point to your data, only to some views of it.	point exactly to your data.
Structure design	Design before using. Any change is a pain.	Easily add new relationships on the fly.
Sharing data	Must create a bridge format and translate.	Nothing to do in theory, sometimes bridge rules.
Merging two companies	Big data mess.	Usage may need adjustments, but data is safe.

Data is already here



As of March 2009

Real-world applications

- DBpedia (extracts data from Wikipedia)
- Faviki (a del.icio.us where tags are meaningful)
- BnF (tons of metadata)
- pubmed.org (metadata)
- BBC (smarter context information)
- Drupal (improve search engines ranking)
- GoodRelations (description of businesses)

How to publish your data

Refrain from creating new ontologies when one already exists.

Like the normal web, the Semantic Web would not be a web without links: link to anything that makes sense, using existing URIs.

Follow the best practices:

<http://www.w3.org/TR/swbp-vocab-pub>

Conclusion

- Have fun exploring things on DBpedia.org to get familiar with data graphs.
- Publish your data, it will become useful!
- Use the data, create mashups and applications, it is really easy!

Links

- <http://en.wikipedia.org/wiki/URI>
- <http://esw.w3.org/topic/SweolG/TaskForces/CommunityProjects/LinkingOpenData>
- <http://www.w3.org/DesignIssues/LinkedData.html>
- <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial>
- <http://www.dfki.uni-kl.de/~sauermann/2006/11/cooluris>
- <http://www.w3.org/2001/sw/RDFCore/ntriples>
- <http://neologism.deri.ie>
- <http://www.semantic-conference.com/ataglance>



Source available at <http://nrw.free.fr/tlugsw.odp>