An Introduction to the Semantic Web and the Web of Data

Prof. Dr. Philippe Cudré-Mauroux
Director, eXascale Infolab
University of Fribourg, Switzerland

May 12, 2011, VeriSign EMEA, Villars-sur-Glane, Switzerland

eXascale Infolab

UNIVERSITAS FRIBURGENSIS

Overview

- The Vision behind the Semantic Web
- Applications
- Building Blocks
  - RDF/S
  - SPARQL
  - OWL
- A Few Research Topics
  - idMesh
  - GraphDB
- Q&A

My Background in 10 Bullet Points

- HP Europe
- EPFL M.Sc., Eurécom Institute + INRIA M.Sc.
- IBM Watson Research (NY)
- EPFL Ph.D. (best thesis award 2006, EPFL Press Mention 2006)
  - Emergent Semantics, RDF PDMS
- U.C. Berkeley
- Microsoft Research Asia
- M.I.T.
  - Database Systems
- Swiss NSF Prof. at Uni Fribourg

SW and ME

- Interested for years in sharing data on the Web
  - RDF/S is the only reasonable solution!
- International Semantic Web Conference (ISWC)
  - Co-organizer in 2007, 2009, 2010
  - Vice-Chair in 2011 (Bonn)
  - PC-Chair in 2012 (Boston)
    - Hope to see you there!
- Research Topics
  - Decentralized Data Integration
  - RDF Storage (P2P/Cloud)
  - Semantic DNS

The Semantic Web Vision

- Provide interoperable, machine-processable data on the Web
  - Machines cannot meaningfully parse HTML pages
  - Relational DBMSs were designed to be centralized (implicit schema, local data, outside of the Web)
  - XML is document-oriented
- Pushed by Sir Tim Berners-Lee
- Standardized by the W3C (RDF/S, OWL)
- Now promoted by major vendors & actors
  - Google, Yahoo!, Facebook, Oracle,
  - Pharmaceuticals, Governments, News industry, etc.

UNIVERSITAS FRIBURGENSIS

Advantages

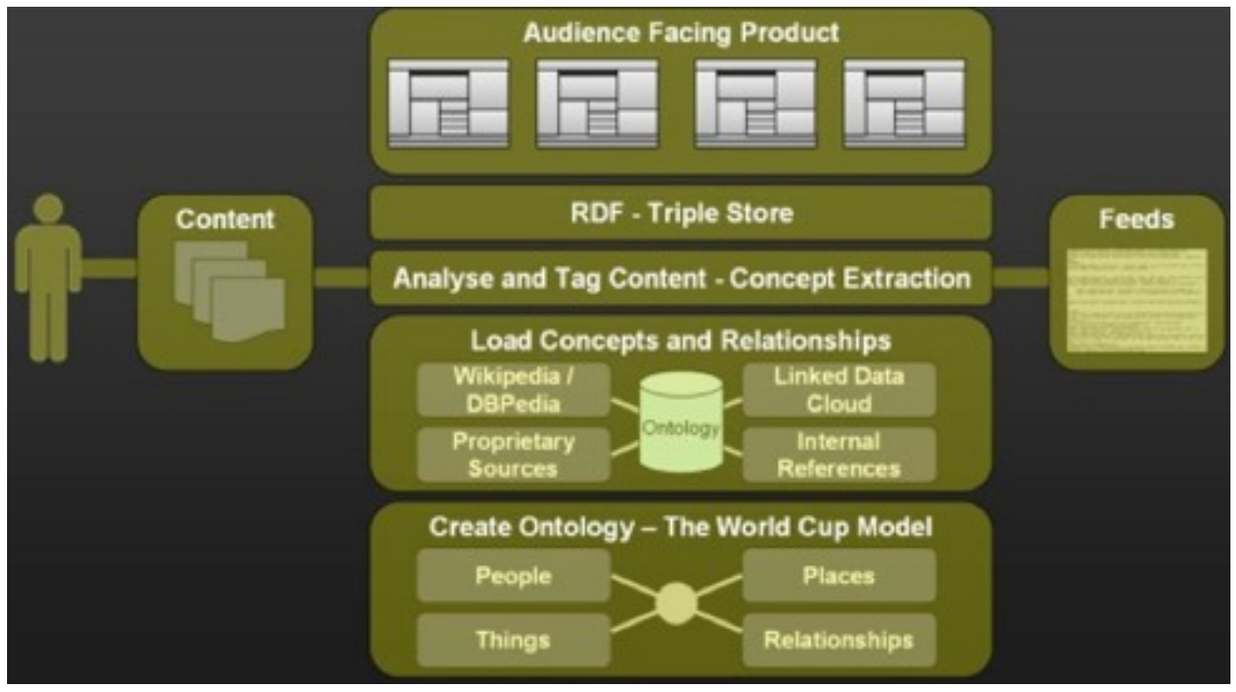
- Standardized!
- Flexible schemas
- Also supports highly-expressive ontologies
  - inference!
- Easy to
  - export (Rel2RDF)
  - publish (SPARQL end-point)
  - discover (SW search-engines)
  - consume (standard libraries)
  - integrate (links bw datasets)

Application (1/3): Knowledge Bases

- US National Center for Biotechnology Information, "Oncology Metathesaurus"
  - 50,000+ classes, ~8 people supporting full time
  - OWL DL rigorously followed
  - Provably consistent

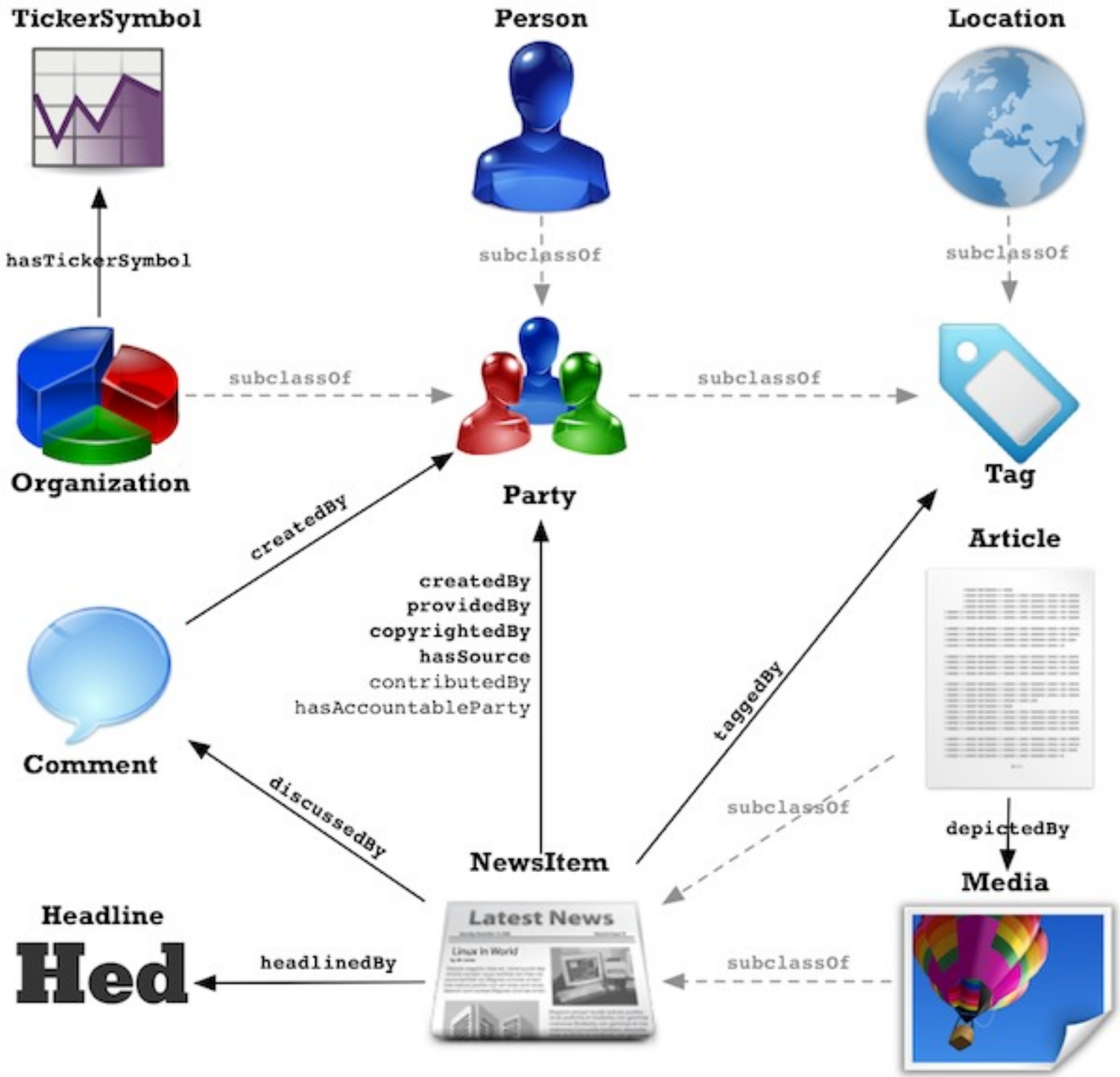
NCBI
National Center for Biotechnology Information

- BBC World Cup 2010 Web Site
  - Dynamic publishing



UNIVERSITAS FRIBURGENSIS

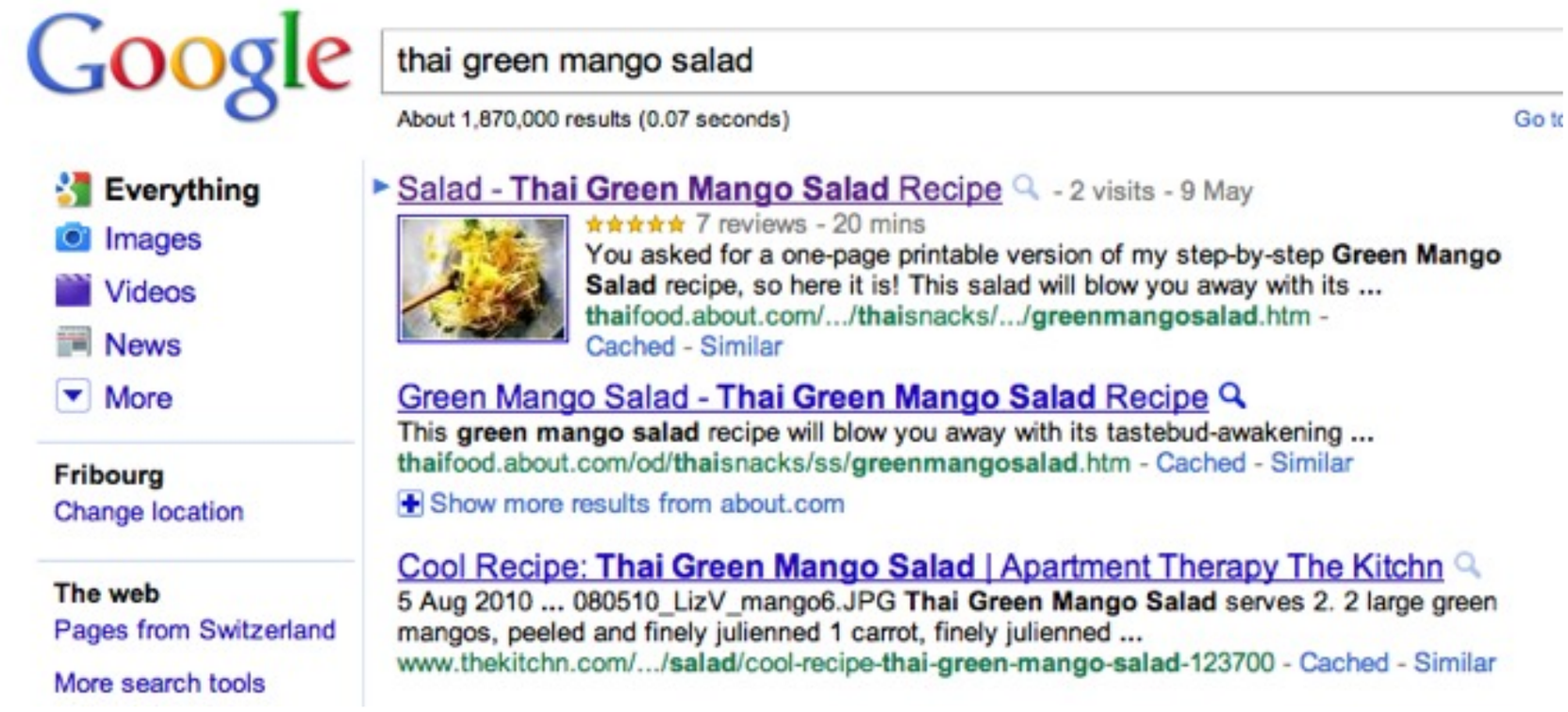
Applications (2/3): Web Metadata



- *RNews*
  - New IPTC standard (News)
  - Better news search
  - Better news alerts
  - Better ad placement
  - Better analytics

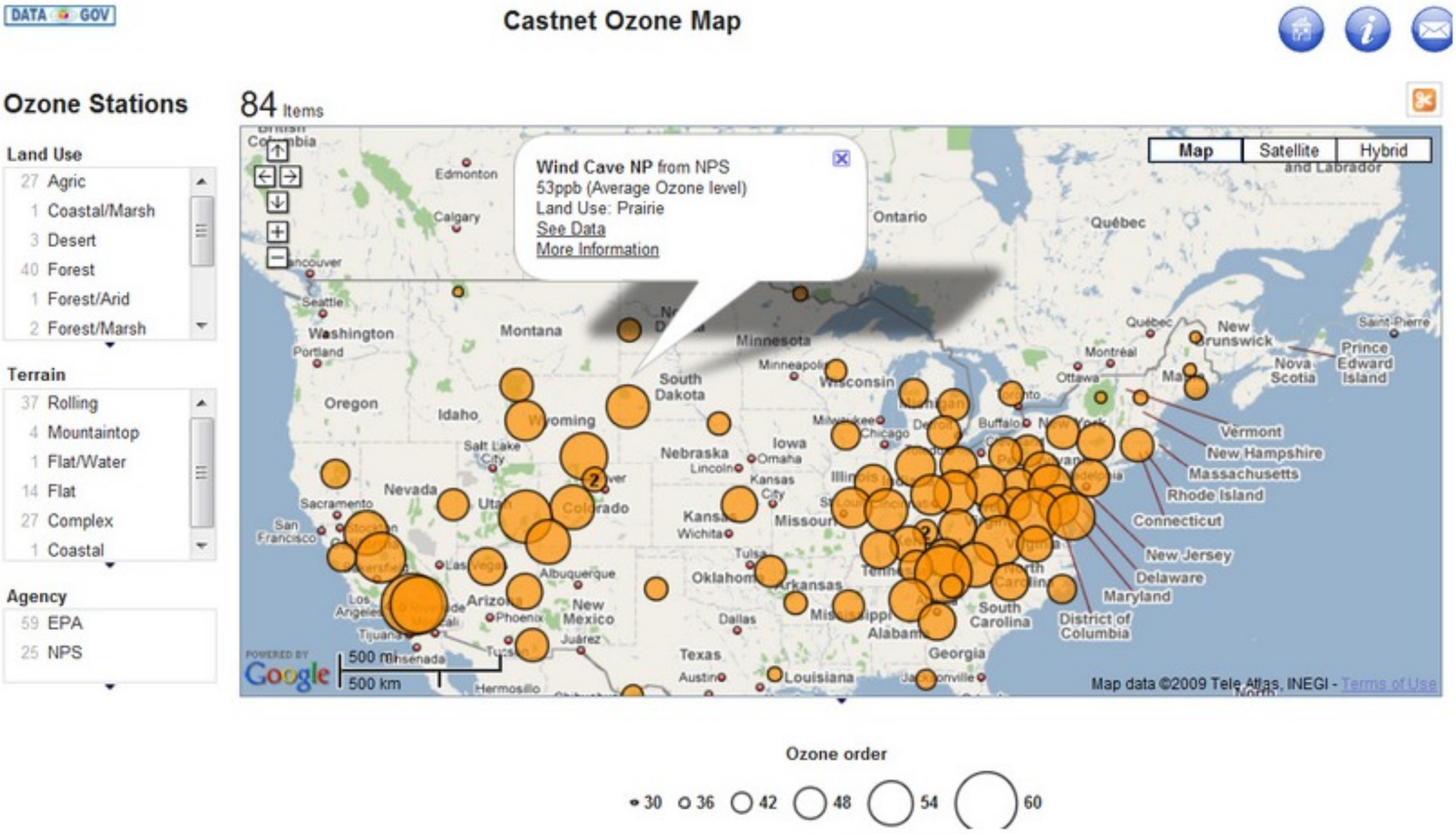
Applications (2/3): Web Metadata

- Metadata parsing supported by major search engines



- Also: Facebook Open Graph API

Application (3/3): Decentralized DBMS

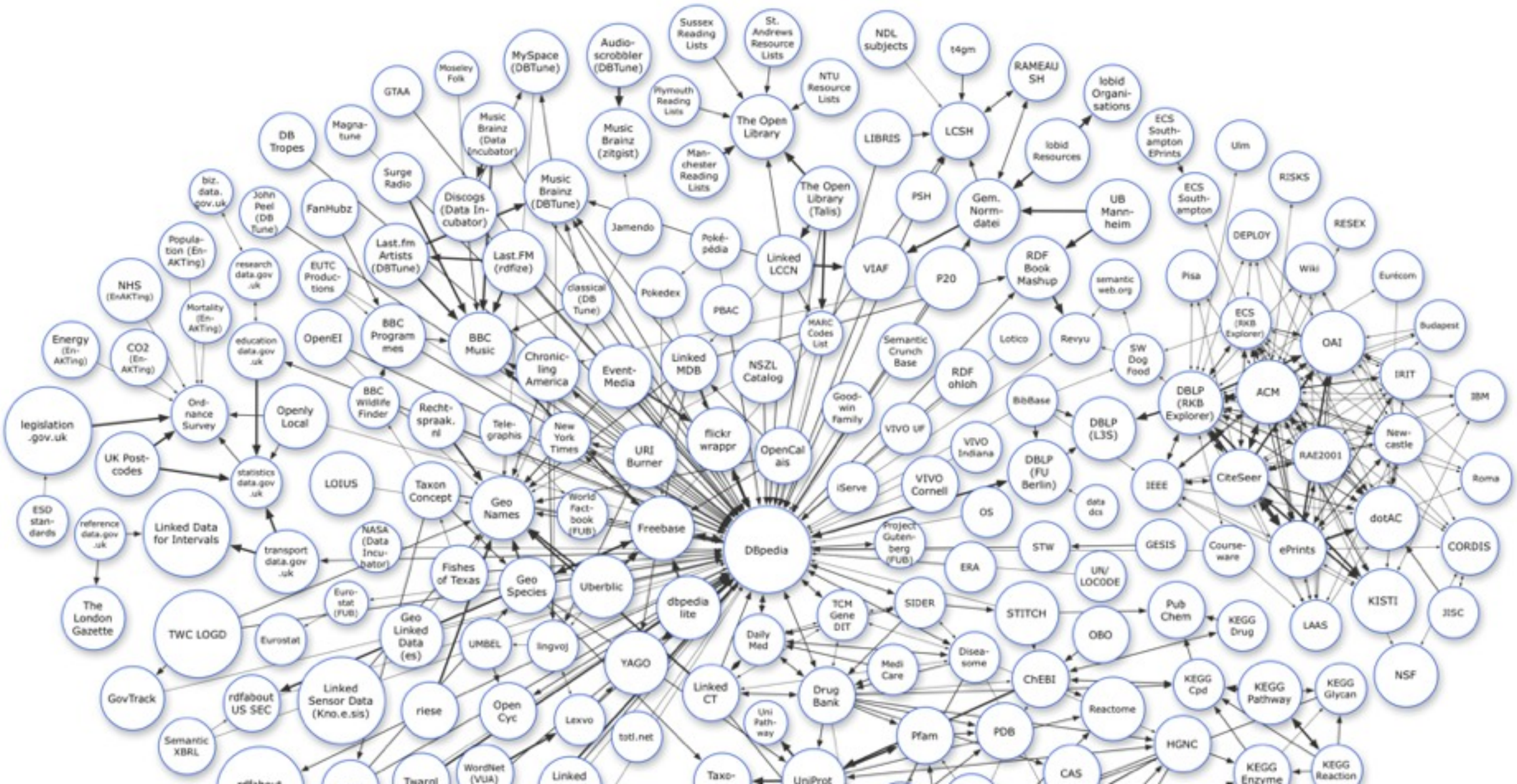


- Taken from data.gov
- Easy publication and integration of data

Application (3/3): Decentralized DBMS

- For the first time, a giant *Web of Data* is emerging!

➡ *Linked Open Data*



UNIVERSITAS FRIBURGENSIS

Building Block 1: RDF

- Triples!

1:subject, 2:predicate, 3:object

ex.: philippe made idmesh_paper:

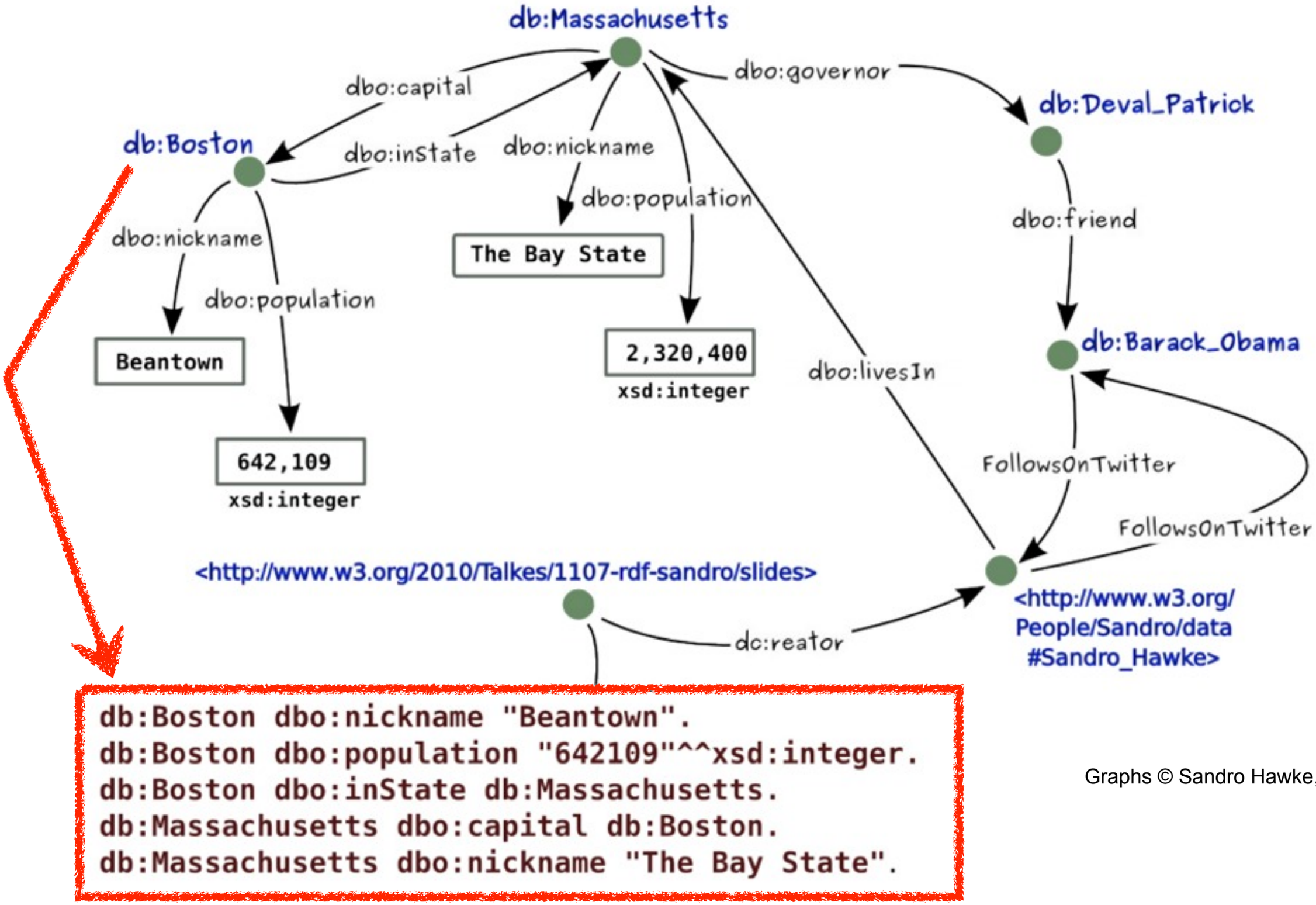
1: http://data.semanticweb.org/person/philippe-cudre-mauroux

2: http://xmlns.com/foaf/0.1/made

3: http://data.semanticweb.org/conference/www/2009/paper/60

- Everything is a URI!
- Literals / XML data types also possible for object
- That's it!

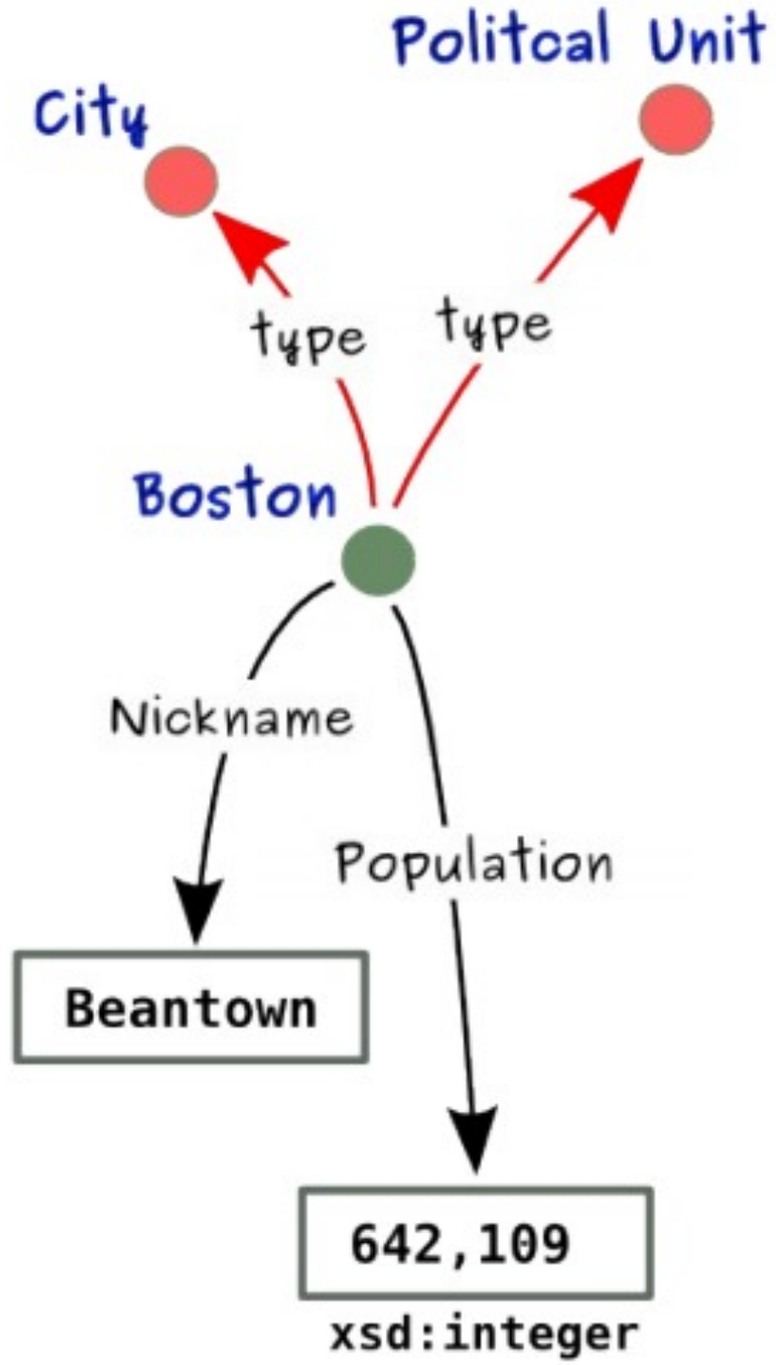
Naturally Forms Distributed Graphs



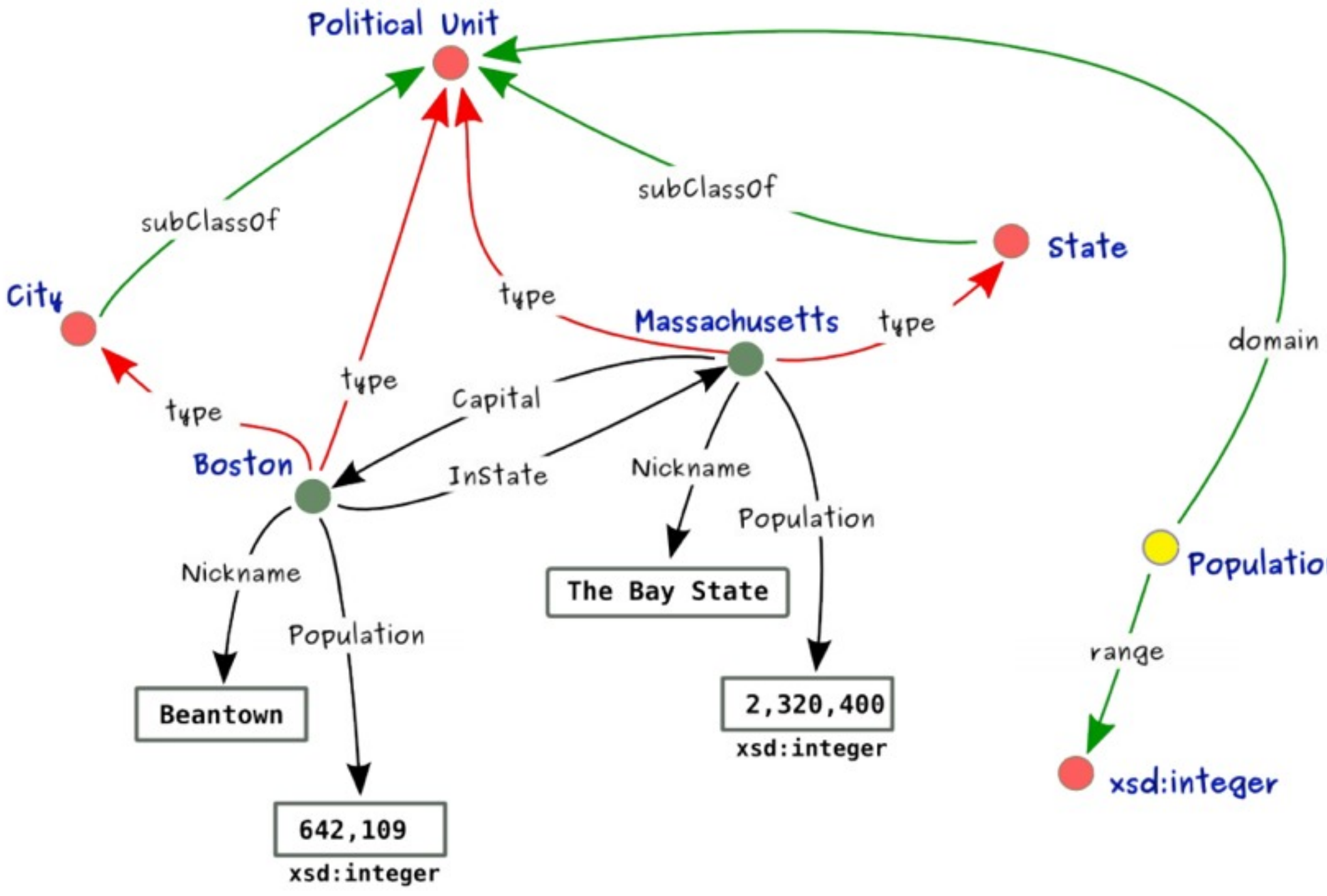
Graphs © Sandro Hawke, W3C

A Bit of Structure

- RDF Schema (RDFS)
  - Like a DB schema but Web-based
    - Discoverable, sharable
  - Defines classes, subclasses, properties, lists, etc.
  - Vocabularies (schemas) can be global (foaf) or local



RDF Schema Example



- Provides structure! easier to check, process

XML Serialization

- Exchange format
- Not meant for humans (ugly!)
- Not meant for DBMS (verbose!)

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:db="http://dbpedia.org/resource/">
  <rdf:Description rdf:about="http://dbpedia.org/resource/Massachusetts">
    <db:Governor>
      <rdf:Description rdf:about="http://dbpedia.org/resource/Deval_Patrick" />
    </db:Governor>
    <db:Nickname>Bay State</db:Nickname>
    <db:Capital>
      <rdf:Description rdf:about="http://dbpedia.org/resource/Boston">
         <db:Nickname>Beantown</db:Nickname>
      </rdf:Description>
    </db:Capital>
  </rdf:Description>
</rdf:RDF>
```

RDFa

- Embedding RDF information in HTML pages
- Supported by Google, Yahoo, etc.

```
<body>
  <div about="http://dbpedia.org/resource/Massachusetts">The
  Massachusetts governor is
    <span rel="db:Governor">
  <span about="http://dbpedia.org/resource/Deval_Patrick">Deval
  Patrick
  </span>,
    </span>
    the nickname is "<span property="db:Nickname">Bay State</span>",
    and the capital
    <span rel="db:Capital">
  <span about="http://dbpedia.org/resource/Boston">
    has the nickname "<span property="db:Nickname">Beantown</span>".
  </span>
    </span>
  </div>
</body>
```

Building Block 2: SPARQL

- Declarative query language for RDF/S
  - SQL for the Semantic Web!

```
prefix db: <http://dbpedia.org/resource/>
prefix dbo: <http://dbpedia.org/ontology/>

SELECT ?cap
WHERE { db:Massachusetts dbo:capital ?cap }
```

Building Block 3: OWL

- The Web Ontology Language
  - Very expressive schemas! (ontologies)
  - Description Logics
    - ... and several flavours
  - Example: OWL 2 EL axioms:
    - class inclusion (SubClassOf)
    - class equivalence (EquivalentClasses)
    - class disjointness (DisjointClasses)
    - object property inclusion (SubObjectPropertyOf) with or without property chains, and data property inclusion (SubDataPropertyOf)
    - property equivalence (EquivalentObjectProperties and EquivalentDataProperties),
    - transitive object properties (TransitiveObjectProperty)
    - reflexive object properties (ReflexiveObjectProperty)
    - domain restrictions (ObjectPropertyDomain and DataPropertyDomain)
    - range restrictions (ObjectPropertyRange and DataPropertyRange)
    - assertions (SameIndividual, DifferentIndividuals, ClassAssertion, ObjectPropertyAssertion, DataPropertyAssertion, NegativeObjectPropertyAssertion, andNegativeDataPropertyAssertion)
    - functional data properties (FunctionalDataProperty)
    - keys (HasKey)
  - Inference! ex.: *TransitiveObjectProperty(hasAncestor)* $hasAncestor(x, y) \wedge hasAncestor(y, z) \rightarrow hasAncestor(x, z)$

Research Topic 1: idMesh--Disambiguation of URIs

- The World, modeled as a collection of identifiers
  - Linked Data
  - Semantic Web
  - RESTful services
  - ...

Naming & Decentralization

- The great thing about *unique identifiers* is that there are *so many* to choose from
  - Top-down efforts to impose canonical ontological names (instances) have failed largely
    - Decentralized naming game
    - Soaring dimensions in Web 2.0 / 3.0 contexts

http://semanticweb.org/id/Philippe_Cudre-Mauroux
http://data.semanticweb.org/person/philippe-cudre-mauroux
http://people.csail.mit.edu/pcm/i http://lsirpeople.epfl.ch/cudre/i
http://semanticweb.org/wiki/Special:ExportRDF/Philippe_Cudr%C3%A9-Mauroux
http://tw.rpi.edu/wiki/Special:ExportRDF/Philippe_Cudre-Mauroux
http://wiki.ontoworld.org/index.php/Special:ExportRDF/Philippe_Cudr%C3%A9-Mauroux
http://korrekt.org/index.php/Special:ExportRDF/Philippe_Cudr%C3%A9-Mauroux
http://prauw.cs.vu.nl:8080/sesame/repositories/socionet?query=http%3A%2F%2Fwww.cs.vu.nl%2F%7Epmika%2Fsocionet%23Philippe%2BCudre-Mauroux
http://www.zoominfo.com/Search/PersonDetail?PersonID=402960578 http://www.flickr.com/photos/28735...@N00/
http://www.facebook.com/profile.php?id=1251943... .....

ID JUNGLE

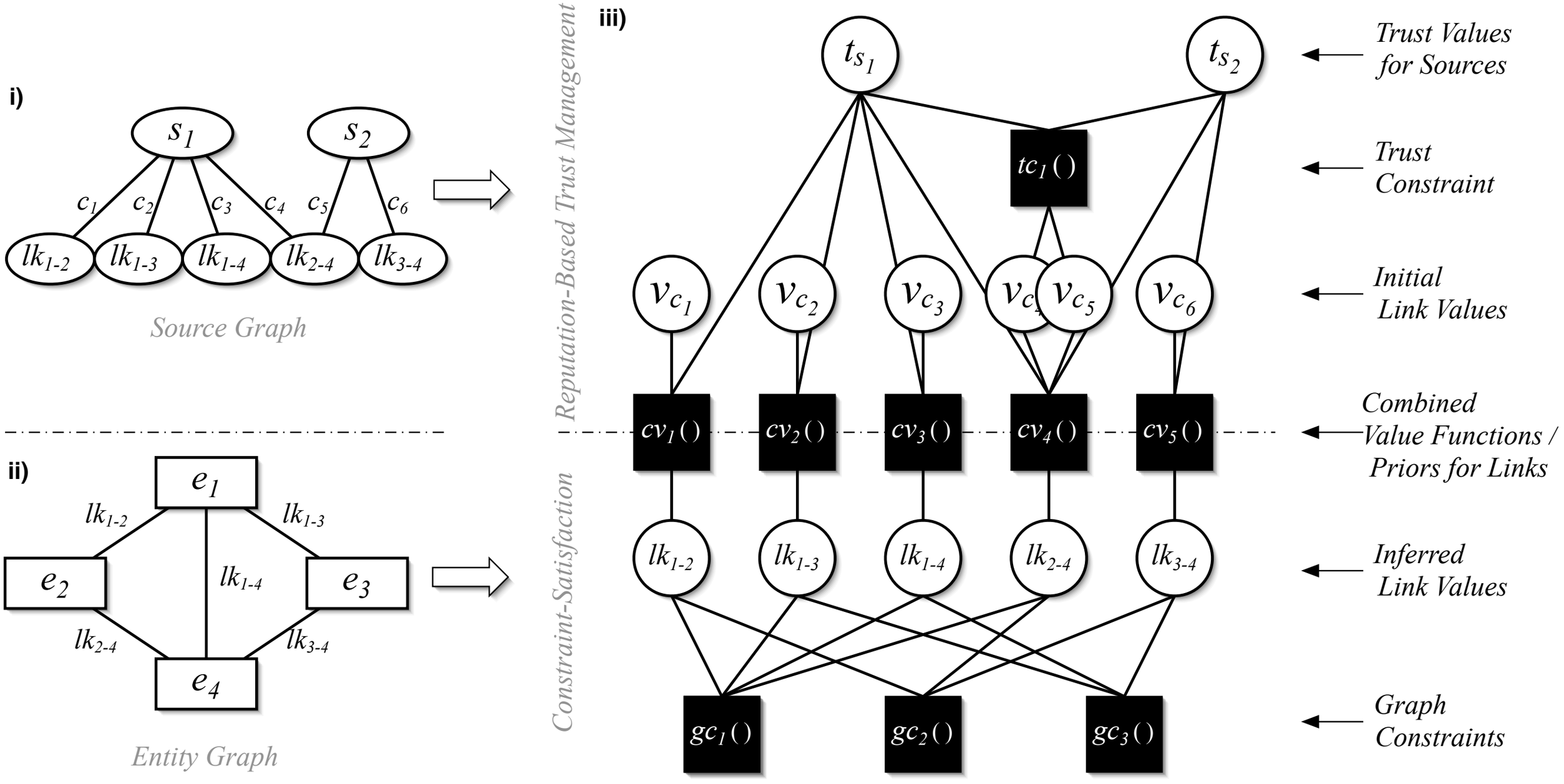
Entity Consolidation

- Urgent need to disambiguate / consolidate URIs
  - URI directories
    - *semantic DNS ?*
  - Pairwise, decentralized mechanisms

```
...
<rdfs:Class rdf:ID="Entity"/>
<rdf:Property rdf:ID="idMeshProperty">
        <rdfs:domain rdf:resource="#Entity" />
        <rdfs:range rdf:resource="#Entity" />
</rdf:Property>
<rdf:Property rdf:ID="LinkConfidence">
        <rdfs:domain rdf:Statement />
        <rdfs:range rdf:datatype="&xsd;decimal" />
</rdf:Property>
<rdf:Property rdf:ID="EquivalentTo">
        <rdfs:subPropertyOf rdf:resource="#idMeshProperty" />
</rdf:Property>
<rdf:Property rdf:ID="NotEquivalentTo">
        <rdfs:subPropertyOf rdf:resource="#idMeshProperty" />
</rdf:Property>
<rdf:Property rdf:ID="Predates">
        <rdfs:subPropertyOf rdf:resource="#EquivalentTo" />
</rdf:Property>
<rdf:Property rdf:ID="Postdates">
        <rdfs:subPropertyOf rdf:resource="#EquivalentTo" />
</rdf:Property>
<rdf:Property rdf:ID="Equidates">
        <rdfs:subPropertyOf rdf:resource="#EquivalentTo" />
</rdf:Property>
```

```
<rdf:Description rdf:about="http://www.epfl.ch/">
        <idMesh: NotEquivalentTo rdf:ID="link0001"
          rdf:resource="http://www.ethz.ch"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.epfl.ch/">
        <idMesh:EquivalentTo rdf:ID="link0002"
          rdf:resource="http://en.wikipedia.org/wiki/EPFL"/>
</rdf:Description>
<rdf:Description rdf:about="#link0002">
        <idMesh:LinkConfidence
          rdf:datatype="&xsd;decimal"> 0.9 </idMesh:LinkConfidence>
</rdf:Description>
```

idMesh: Social Disambiguation



Probabilistic, social inference based on graphs
cf. World Wide Web 2009 paper

UNIVERSITAS FRIBURGENSIS

Research Topic 2: GraphDB

- RDF has a data model that is very different from relational data or XML
  - Urgent need to develop highly scalable, native stores
- GraphDB, a native graph-based RDF store
  - on-the-fly schemas
  - novel graph indexing
  - highly-compressed template-based data
  - dedicated query optimizer
  - extensions for collaborative cloud storage

➡ 2 orders of magnitude faster than Oracle RDF on most queries!

Conclusions

- Semantic Web provides machine-processable data to the Web
- Standardized
- Currently growing exponentially
- Major companies are jumping on the bandwagon
- Many exciting research / R&D opportunities

- To go one step further: invited tutorial @ ISWC

http://people.csail.mit.edu/pcm/SemWebTutorial.html

Thanks for you Attention!

- Q&A