

# Semantic Modeling with RDF

DBTech Ext Workshop on Database Modeling and  
Semantic Modeling

Lili Aunimo

# Expected Outcomes

You will learn:

- Basic concepts related to ontologies
  - Semantic model
  - Semantic web
- Basic features of RDF and RDF Schema
- Possible application areas for RDF models
- Ontology authoring using Protégé
- Persistence

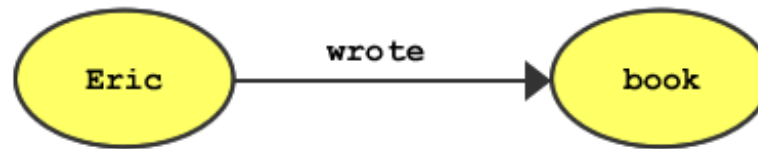
# Database modeling vs. semantic modeling

- Database modeling: conceptual schema, logical schema, physical schema
- Conceptual schema
  - Communicating with the domain expert
  - The magical step between requirements and the conceptual model
- Semantic modeling of data

# Semantic Modeling with RDF

- RDF
  - Resource Description Framework
  - Data model
- W3C suite of specifications
- Semantic web technologies
  - HTML, XML documents
  - Semantic modeling of documents or data?

# The RDF Data Model



# The RDF Data Model

- Directed multigraph
- Serialized in e.g. XML, N-triple
- Resource, property, value – triples
  - Statement
  - Subject – verb - object
- Resources have a URI
  - Global id

# Serializations

(<http://www.dbtechnet.org/kb#Eric>, <http://.../kb#wrote>, "book")

```
<rdf:RDF>
```

```
< http://www.dbtechnet.org/kb#Eric
```

```
  rdf:about="&kb;ClassesProperties_Class1“
```

```
    http://www.dbtechnet.org/kb#wrote="book“/>
```

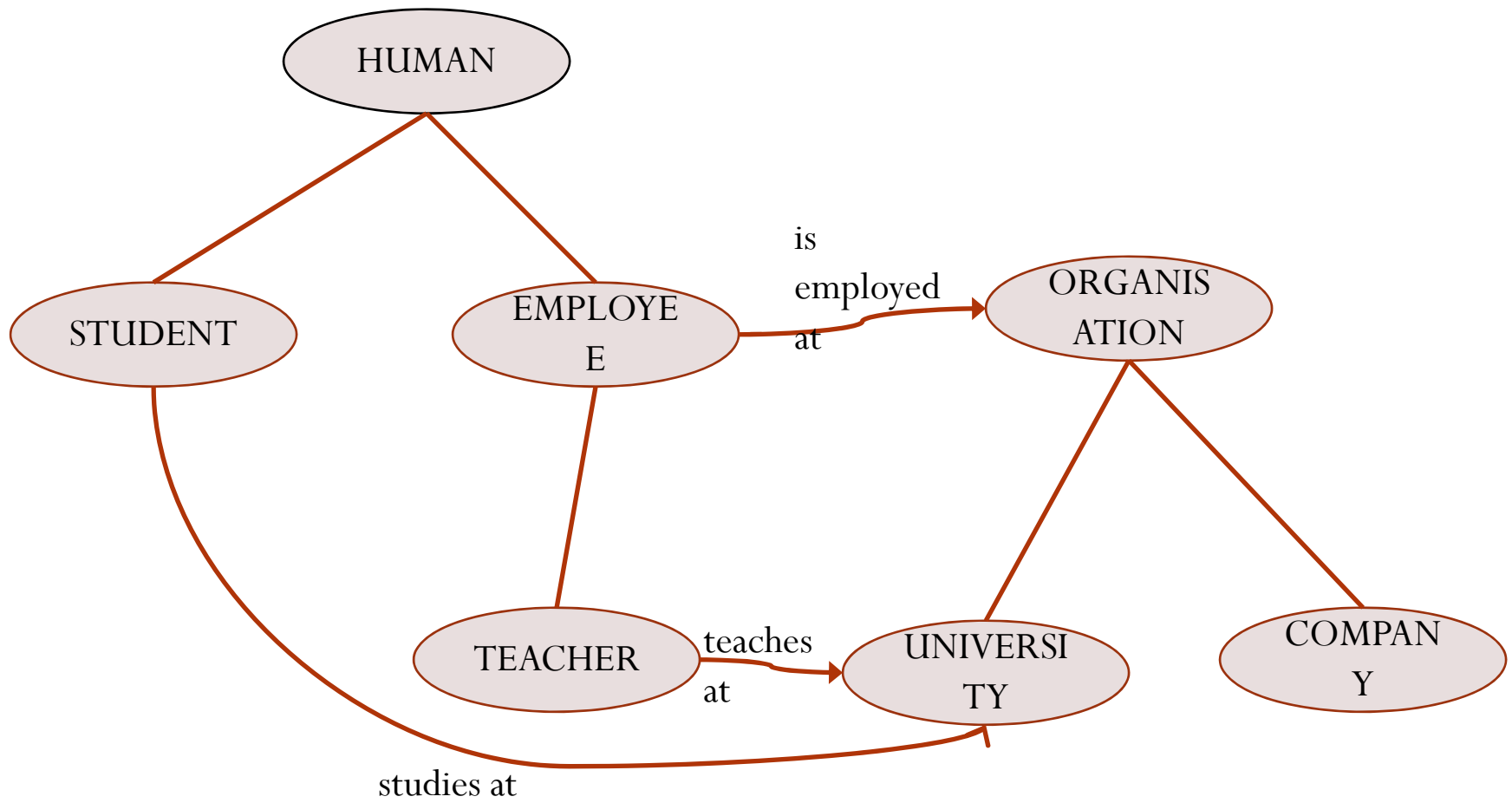
```
</rdf:RDF>
```

# RDF Schema

- Unlike XML schema
- Vocabulary for a RDF model
- Expressed as a RDF model
- Classes and instances (also called data)
  - analogy in object-oriented programming



# RDFS as directed graph



# Serialization

XML:

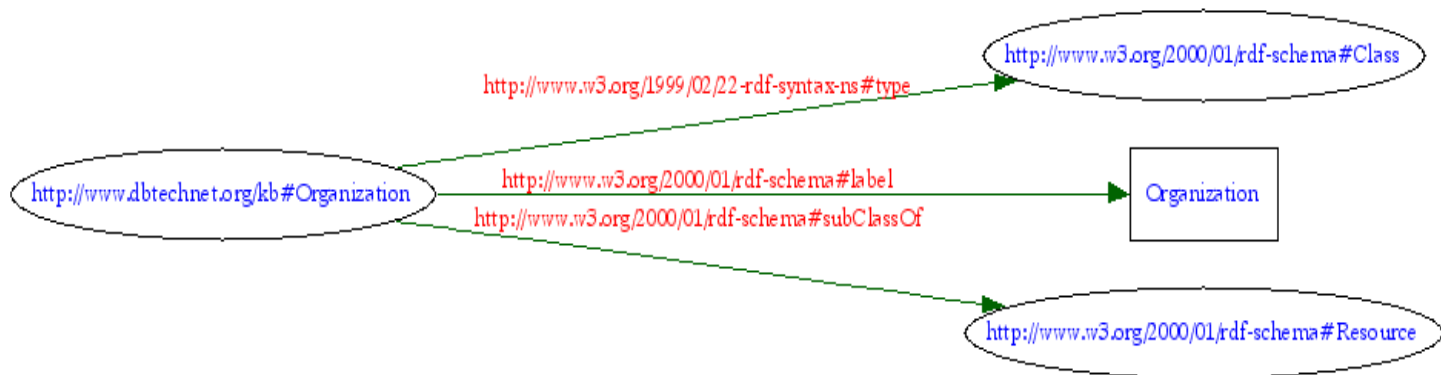
```
<rdfs:Class rdf:about="Organization"  
  rdfs:label="Organization">  
  <rdfs:subClassOf rdf:resource="&rdfs;Resource" />  
</rdfs:Class>
```

N-Triples

(<http://www.dbtechnet.org/kb#Organization>, rdf:type, rdfs:Class)

(<http://.../kb#Organization>, rdfs:label, "Organization")

(<http://.../kb#Organization>, rdfs:subClassOf, rdf:resource="&rdfs;Resource")



# Property, serialization

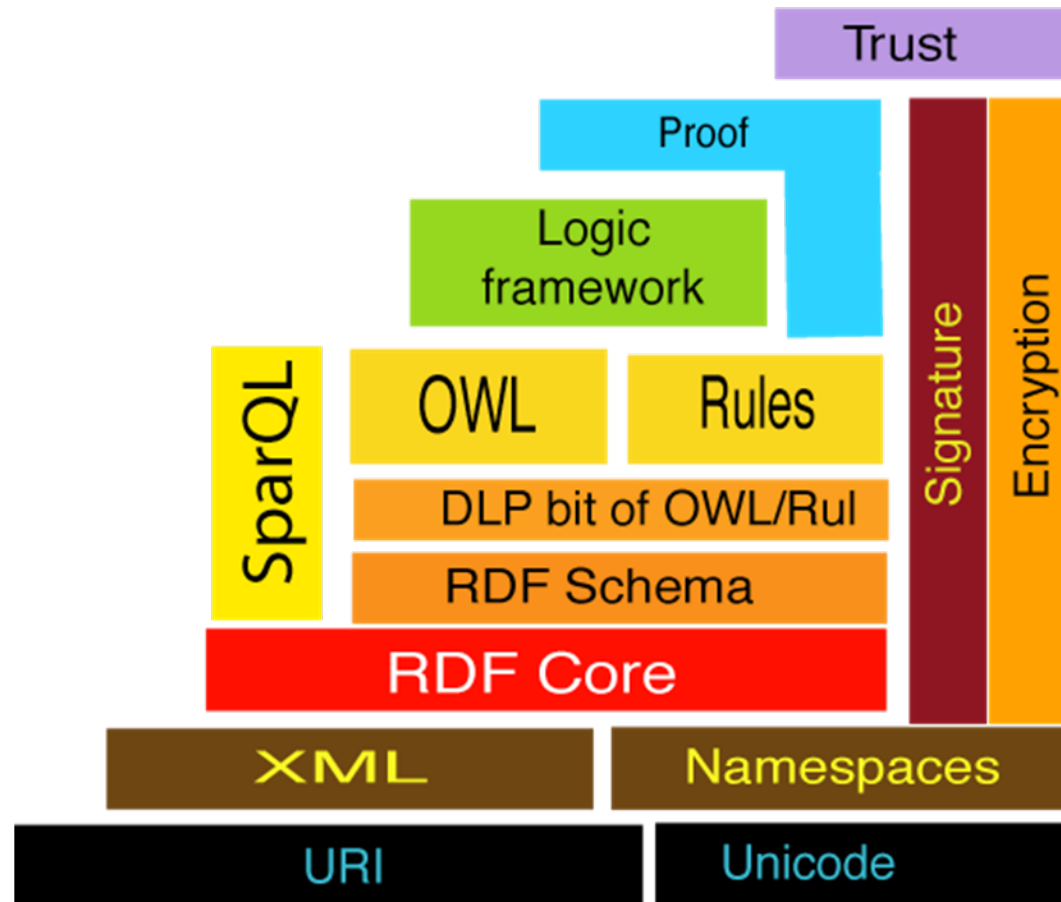
## XML

```
<rdf:Property rdf:about="&kb;teaches_at"      rdfs:label="teaches_at">  
  <rdfs:domain rdf:resource="&kb;Teacher"/>  
  <rdfs:range rdf:resource="&kb;University"/>  
</rdf:Property>
```

## N-Triples

```
(http://www.dbtechnet.org/kb#teaches\_at, rdf:type, rdf:property)  
(http://.../kb#teaches\_at, rdfs:label, "teaches at")  
(http://.../kb#teaches\_at, rdfs:domain, http://.../kb#Teacher)  
(http://.../kb#teaches\_at, rdfs:range, http://.../kb#University)
```

# Layers of the Web by Berners-Lee

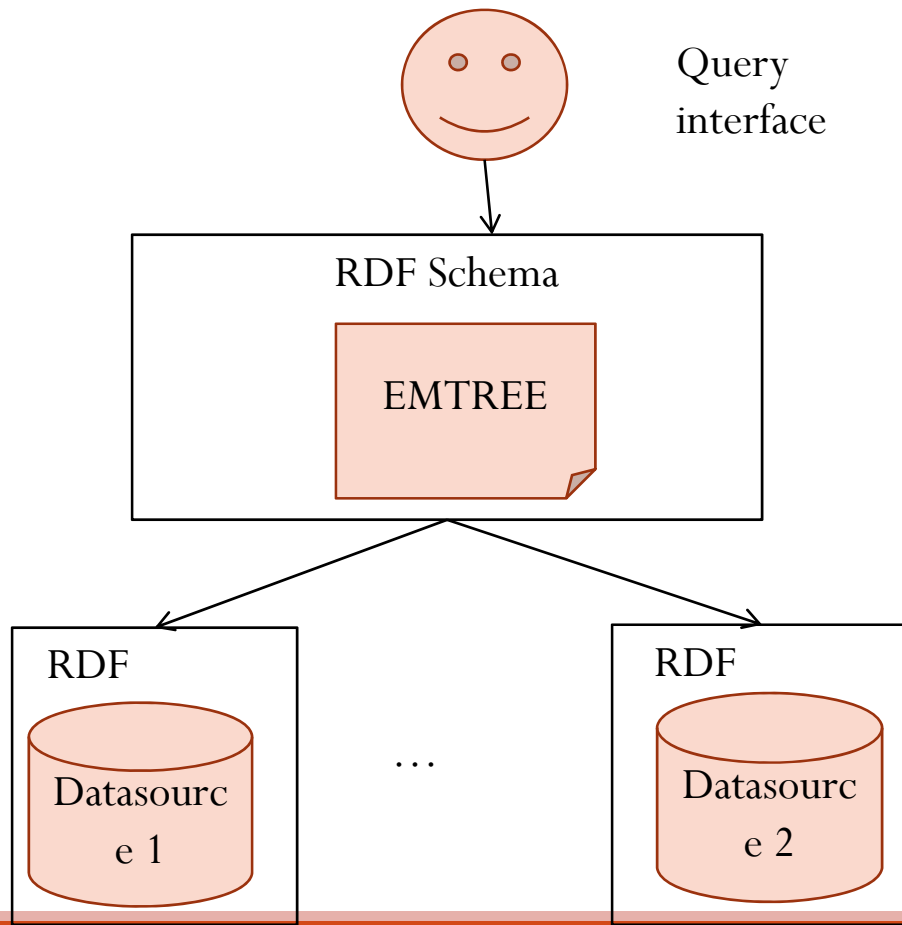


# Motivation

- Systems integration, interoperability, reuse
- Domain modelling for humans and software agents
  - e.g. terminology for database modeling
  - e.g. simulation vs. game
- Data enriched with semantics -> new functionality
- Standardized representation

# Example

- Interdisciplinarity: vertical products to horizontal ones
- Various data sources with different indexing terms and formats
- Syntax: XML
- Semantics
  - Keywords and free text queries not enough
- Homonym, synonym problems etc. .



## Querying across data source at Elsevier



# Example GUI

- Aspirin, EMTREE terms -> acetylsalicylic acid”,
- secondary keywords “mortality”, “practice guideline”, “blood clot lysis” and “warfarin”
- Color: document type e.g. article, abstract

**DOPE Browser**

**File**

**Focus Term**

Current term: **acetylsalicylic acid**

Search for term:  **Go!**

**Navigate Thesaurus...**

**Co-occurring Terms**

<input type="checkbox"/> All relevant documents	500	<input type="checkbox"/>
<input type="checkbox"/> analytical, diagnostic and therapeu...	358	<input type="checkbox"/>
<input type="checkbox"/> anatomical concepts	209	<input type="checkbox"/>
<input type="checkbox"/> biological phenomena and functions	221	<input type="checkbox"/>
<input type="checkbox"/> biomedical disciplines, science and art	26	<input type="checkbox"/>
<input type="checkbox"/> chemical, physical and mathematical phe...	85	<input type="checkbox"/>
<input type="checkbox"/> chemicals and drugs	443	<input type="checkbox"/>
<input type="checkbox"/> geographic names	47	<input type="checkbox"/>
<input type="checkbox"/> groups by age and sex	40	<input type="checkbox"/>
<input type="checkbox"/> healthcare	63	<input type="checkbox"/>
<input type="checkbox"/> physician	12	<input type="checkbox"/>
<input checked="" type="checkbox"/> practice guideline	12	<input checked="" type="checkbox"/>
<input type="checkbox"/> hospital	7	<input type="checkbox"/>
<input type="checkbox"/> anesthetist	6	<input type="checkbox"/>
<input type="checkbox"/> emergency ward	5	<input type="checkbox"/>
<input type="checkbox"/> ambulance	4	<input type="checkbox"/>
<input type="checkbox"/> dentist	4	<input type="checkbox"/>
<input type="checkbox"/> general practice	4	<input type="checkbox"/>
<input type="checkbox"/> long term care	4	<input type="checkbox"/>
<input type="checkbox"/> medicare	4	<input type="checkbox"/>
<input type="checkbox"/> coroner	3	<input type="checkbox"/>
<input type="checkbox"/> general practitioner	2	<input type="checkbox"/>
<input type="checkbox"/> health maintenance organization	2	<input type="checkbox"/>
<input type="checkbox"/> hospital management	2	<input type="checkbox"/>

**Term Overlap Display**

**Document List**

Documents 1 to 10 out of 23    [previous](#)    [next](#)

Contents of "blood clot lysis":

1. Antithrombotic Therapy for Acute Myocardial Infarction.  
O'Donnell, C.J.; Ridker, P.M.; Hebert, P.R.; Hennekens, C.H. (1995). Full-length article. Journal of the American College of Cardiology 25 (7), pp. 23S-29S.  
[ <http://linkinghub.elsevier.com/pii/S073510979500105D> ]
2. Can the MADIT Results Be Applied to Myocardial Infarction Patients at Hospital Discharge?.

**Collapse all**    **Clear all**

# Building blocks of the digital library

1. Data converted into a uniform XML format
2. EMTREE ontology converted into RDFS
3. Articles mapped into ontology using concept extraction software
4. Sesame RDF storage and query engine
5. ClusterMap (commercial product) for GUI

# Ontology authoring

1. Determine scope
2. Consider reuse
3. Find out the concepts. Typically substantives (future resources) and verbs (future properties).
4. Define hierarchy
5. Define properties
6. Test the model by populating it with instances.
7. Check for anomalies

# Protégé

- Open source, java ontology authoring tool
  - OWL
  - Plug-ins
  - Version 3.4
    - OWL (1.0) and RDFS support
    - SPARQL support
    - JDBC
- vs. version 4  
only OWL 2.0  
no SPARQL yet  
no JDBC yet

# Persistence

- Frameworks: Jena, Sesame
- Protégé JDBC
- Oracle, Pyrrho

# Summary

- Semantic modeling
- RDF and RDFS basics
- Usage examples
- Authoring using Protégé
- Persistence