Annotation

Oscar Corcho, Boris Villazón-Terrazas, Asunción Gómez-Pérez
({ocorcho,bvillazon,asun}@fi.upm.es)
Universidad Politécnica de Madrid

Main References

Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. Ontological Engineering. *Springer Verlag. 2003*

Chapter 5: Ontology tools

Corcho, O.
Ontology-based document annotation: trends and open research problems
International Journal of Metadata, Semantics and Ontologies 1(1):47-57. 2006

Handschuh S, Staab S (2003)
Annotation for the Semantic Web. IOS Press

Table of Contents

What is the metadata of this HTML fragment?

Based on Dublin Core

The *contributor* and *creator* is the flight booking service "www.flightbookings.com".
The *date* would be January 1st, 2003, in case that the HTML page has been generated on that specific date.
The *description* would be something like "flight details for a travel between Madrid and Seattle via Chicago on February 8th, 2004".
The document *format* is "HTML".
The document *language* is "en", which stands for English

Flight details

Outbound

Leaving from **Madrid** - Barajas - Spain
on Saturday 08 February 2003 at **11:50**
Arriving in **Chicago** - O'Hare International - United States of America
same day at **14:10**
Airline: American Airlines
Flight No. AA 7615
Type of aircraft: Airbus Industrie A340 All Series PAX/H

Leaving from **Chicago** - O'Hare International - United States of America
on Saturday 08 February 2003 at **16:48**
Arriving in **Seattle** - Seattle/Tacoma International - United States of America
same day at **19:23**
Airline: American Airlines
Flight No. AA 1805
Type of aircraft: non referenced/B

Based on thesauri

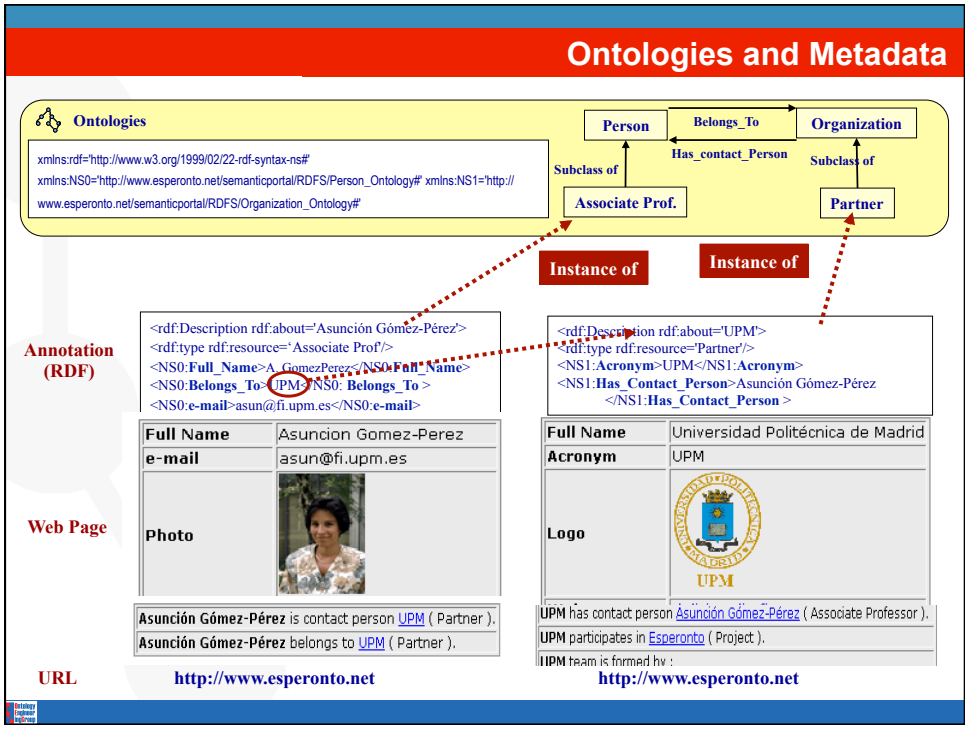
Madrid is a reference to the term with ID 7010413 in the thesaurus, which refers to the city of Madrid in Spain.
Spain is a reference to the term with ID 1000095, which refers to the kingdom of Spain in Europe.
Chicago is a reference to the term with ID 7013596, which refers to the city of Chicago in Illinois, US.
United States of America is a reference to the term "United States" with ID 7012149, which refers to the US nation.
Seattle is a reference to the term with ID 7014494, which refers to the city of Seattle in Washington, US.

Based on ontologies

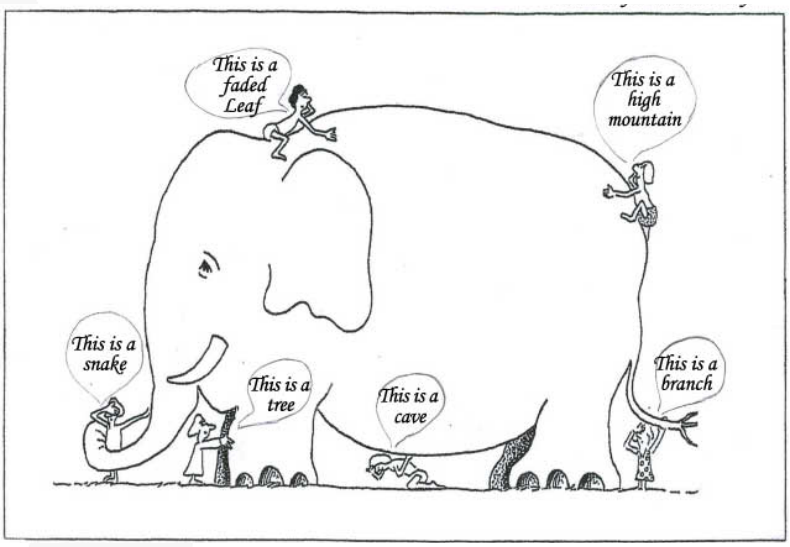
Concept instances relate a part of the document to one or several concepts in an ontology. For example, "Flight details" may represent an instance of the concept `Flight`, and can be named as `AA7615_Feb08_2003`, although concept instances do not necessarily have a name.
Attribute values relate a concept instance with part of the document, which is the value of one of its attributes. For example, "American Airlines" can be the value of the attribute `companyName`.
Relation instances that relate two concept instances by some domain-specific relation. For example, the flight `AA7615_Feb08_2003` and the location `Madrid` can be connected by the relation `departurePlace`

Annotation assert facts using terms *(metadata in RDF)*

Represent terms and their relationships *(ontology in RDFS/OWL)*



An Indian Tale: Blind Men and an Elephant



The early days of annotation in the Web

- The main objective in the early days was the agreement on how to include annotations on the Web
  - (KA)2
  - SHOE (Simple HTML Ontology Extension)
  - …

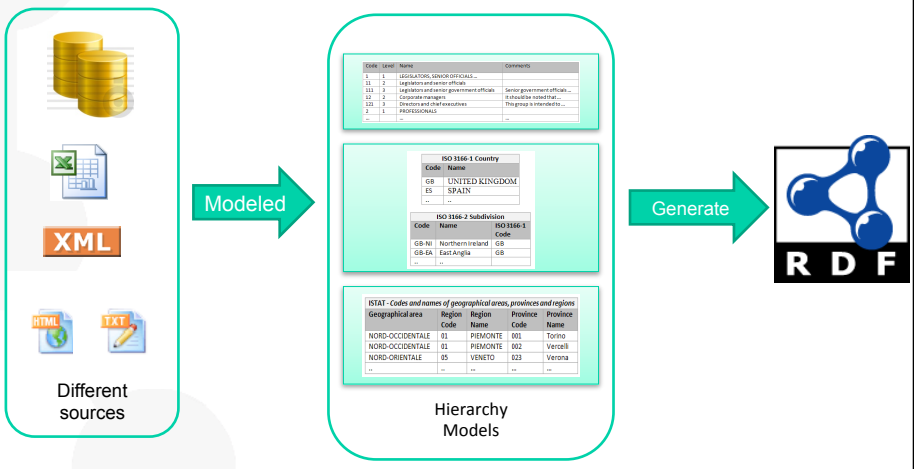
```
<html>
<head><TITLE> Richard Benjamins </TITLE>
<a ONTO="page:Researcher"> </a>
</head>

<H1> <A HREF="pictures/id-rich.gif">
<IMG  align=middle SRC="pictures/richard.gif"></A>
<a ONTO="page[photo=href]"
HREF="http://www.iiia.csic.es/~richard/pictures/richard.gif" ></a>
<a ONTO="page[firstName=body]">Richard</a>
<a ONTO="page[lastName=body]">Benjamins </a>
</h1> <p>

<A ONTO="page[affiliation=body]" HREF="#card">
Artificial Intelligence Research Institute (IIIA)</A> -
<a href="http://www.csic.es/">CSIC</a>, Barcelona, Spain <br>
and <br>
<A ONTO="page[affiliation=body]" HREF="http://www.swi.psy.uva.nl/">
Dept. of Social Science Informatics (SWI)</A>
-
<A HREF="http://www.uva.nl/uva/english/">UvA</A>, Amsterdam, the
Netherlands
```

- Not so far from RDFa

Motivation



The early days of the Semantic Web

Availability of content

- Including the "Deep Web" and Web Services

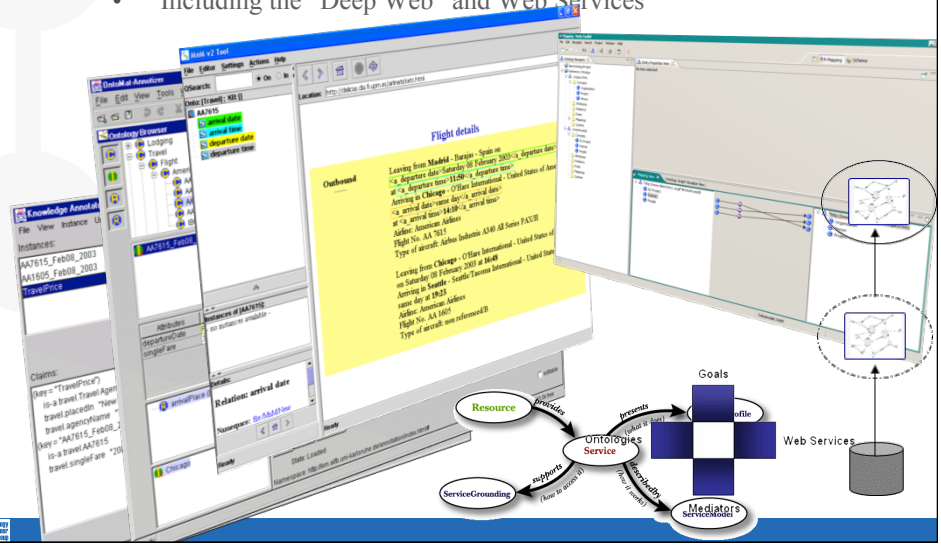


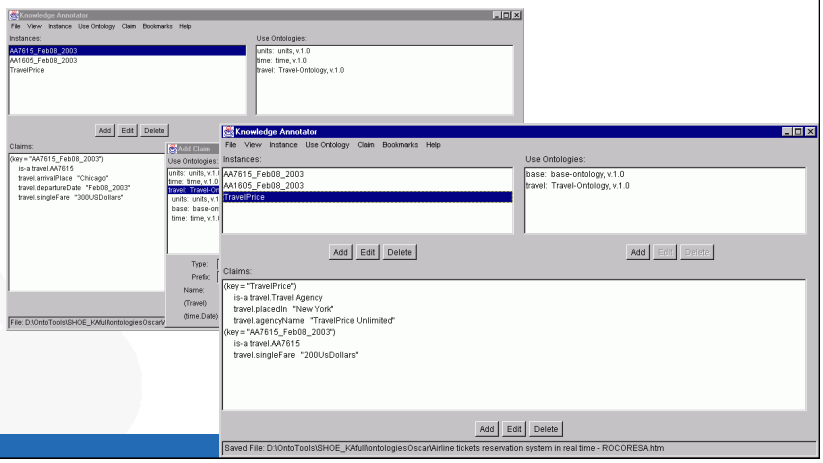
Table of Contents

Web Resource Annotation. Dimensions

- Sources
  - Source type
    - Text: HTML, XML, PDF, etc.
    - Multimedia: images, video, audio
    - Web Services
  - Origin
    - Static: files
    - Dynamic: databases and forms
    - Very dynamic: data streams
- Technologies
  - Natural Language Processing
  - Information Extraction
  - Layout Analysis
  - Clustering
  - Topic identification
  - Wrapper generation
- Annotation Process
  - Maintenance (adaptivity to changes in the sources)
    - Verification
    - Robustness
    - Auto-adaptivity
  - Annotation Supervision
    - Manual
    - Supervised (semi-automatic)
    - Unsupervised (automatic)
- Degree of formality
  - Web 2.0 tagging
  - Dublin-core based
  - Thesauri
  - Ontology-based

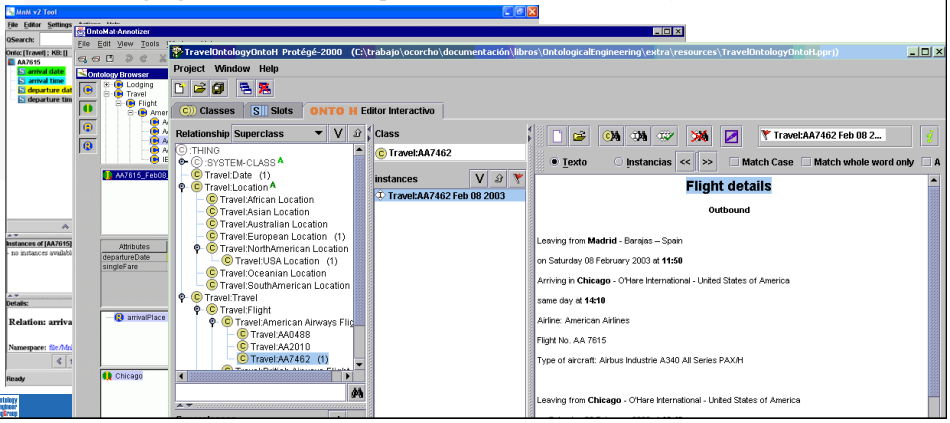
Web 1.0. SHOE Knowledge Annotator

- Standalone application with no Web browser
- Manual annotation
- SHOE



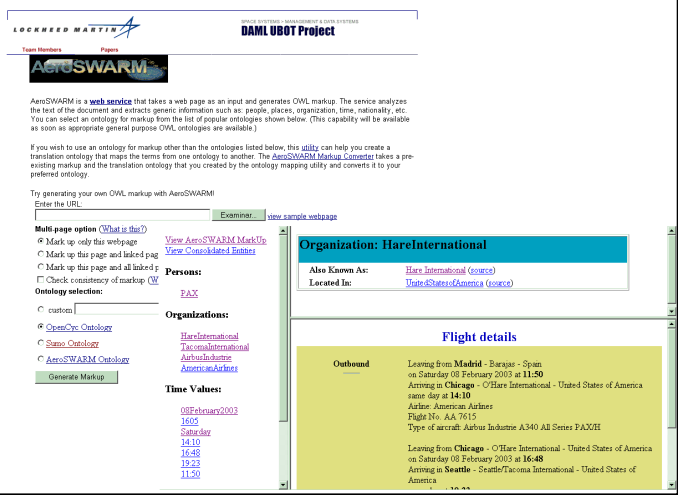
Web 1.0. MnM, OntoMat, ONTO-H, etc.

- Standalone applications
- Manual annotation with drag&drop
- Some features
  - MnM: Semi-automatic annotation with information extraction tools (Amilcare)
  - ONTO-H: cascading annotation rules
- Languages: RDF and OWL, plus other (OCML, XML, etc.)



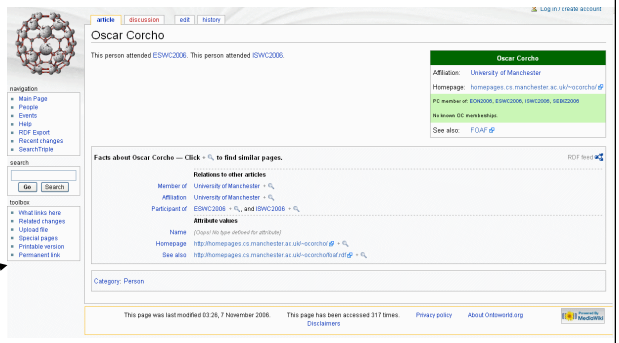
Web 1.0. AeroSWARM

- Web server for any Web document
- Automatic annotation with predefined ontologies: OpenCyc, SUMO and AeroSWARM
- RDF



Web 2.0. Semantic wikis

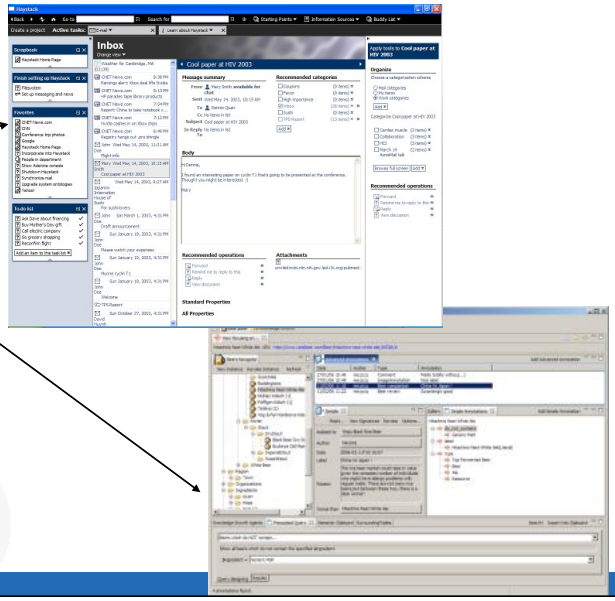
RDF presentation
Rhizome
HTML + RDF page edition
IkeWiki
SemanticMediaWiki
SemPerWiki
SweetWiki
WikSAR
Ontology-based population
OntoWiki
COW
Non-ontology based attribute-value annotation
DiamondWiki



SemanticWikipedia

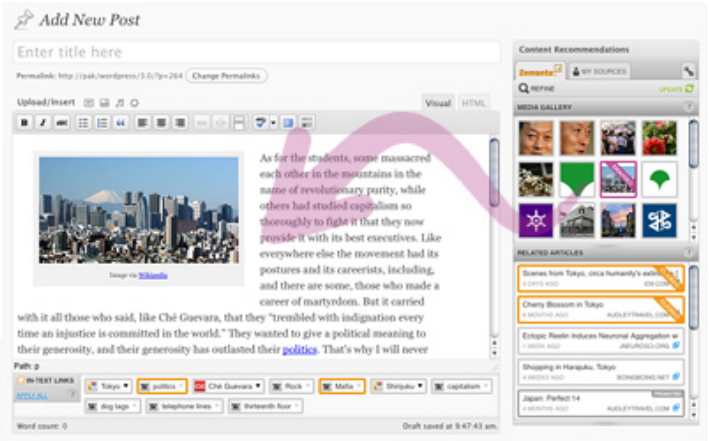
Web 2.0. Semantic desktop (including semantic e-mail)

Haystack
Gnowsis
D-Bin
OpenIrts
NEPOMUK



Web 2.0. Semantics in blogs

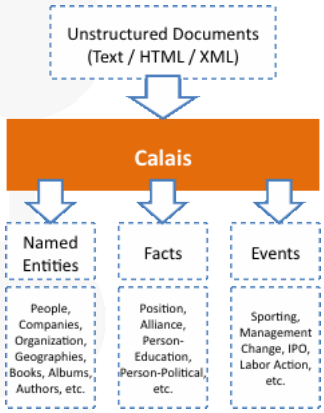
- Zemanta



Web 2.0. Semantics in microblogs

- Not much yet to help on microblog semantic enrichment during creation
- But some work done on semantic characterisation of hashtags

Linked Data. OpenCalais, DBpedia Spotlight



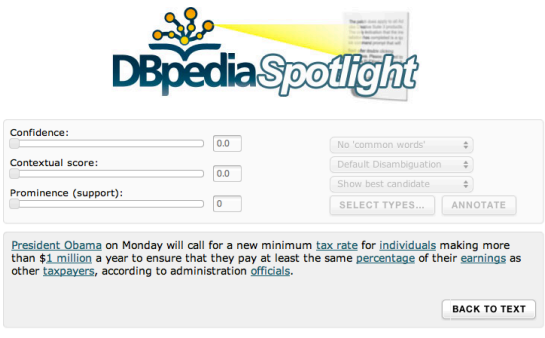
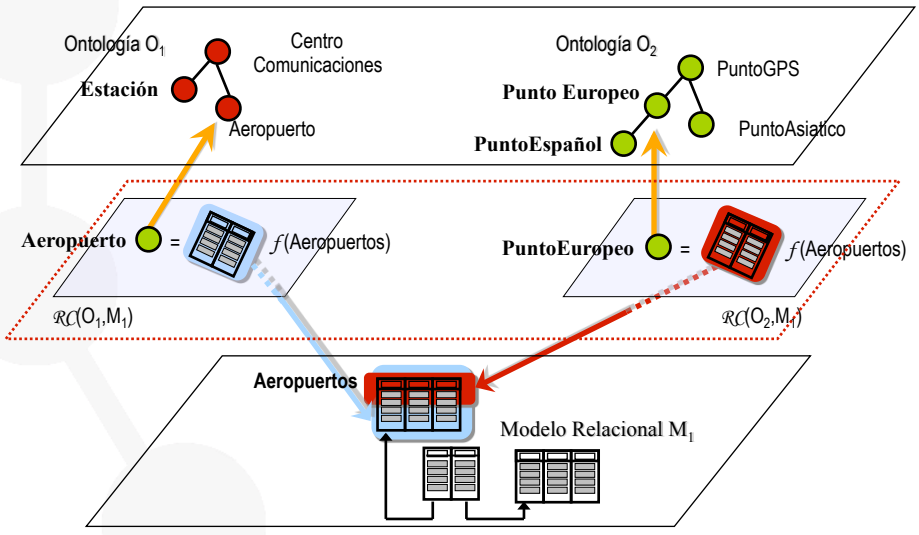
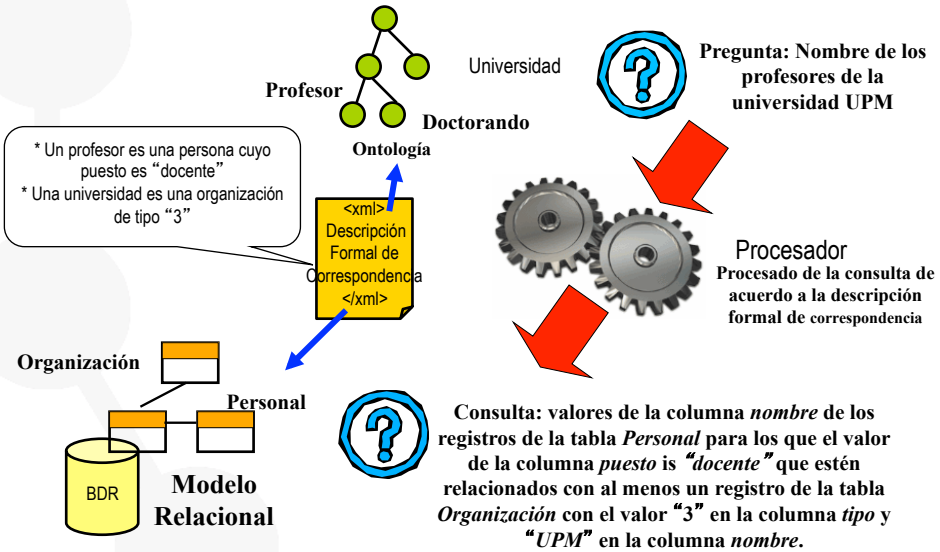


Table of Contents

Ontology-based view over a relational model (I)

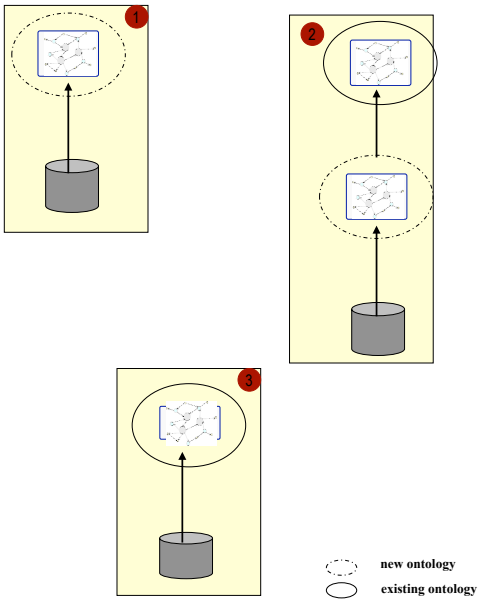


Ontology-based view over a relational model (II)



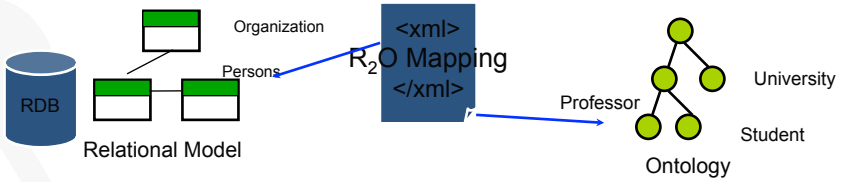
Existing approaches

1. To build a new ontology from a database schema and content (OntoStudio, KaOn Reverse)
2. To map the ontology created in approach (1) to a legacy ontology (NeOn toolkit UKARL)
3. To map an existing DB to a legacy ontology (NeOn Toolkit UPM)

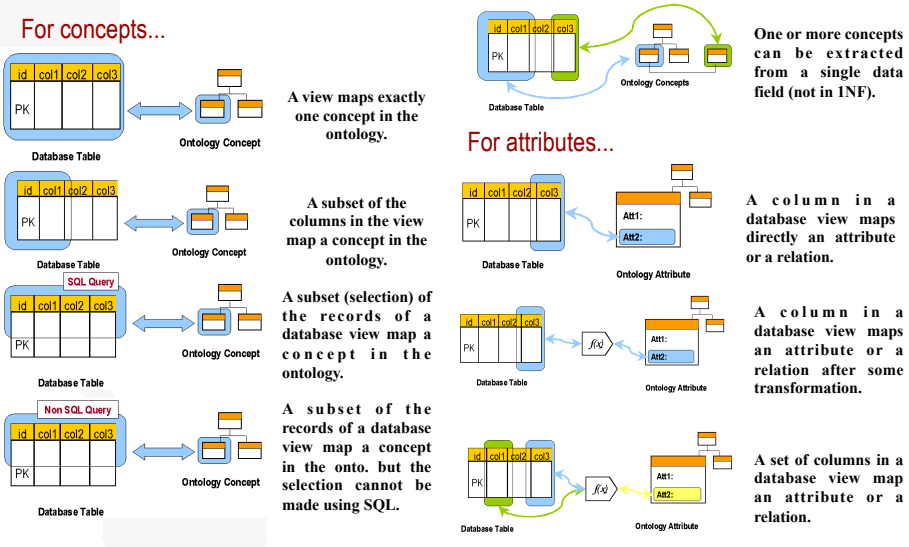


R_2O

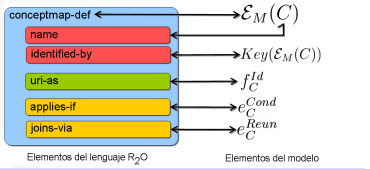
- R_2O is an extensible, fully declarative language to describe mappings between relational database schemas and ontologies.



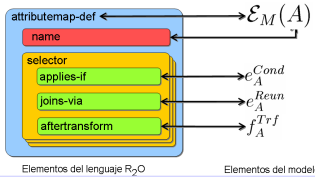
R2O (Relational-to-Ontology) Language



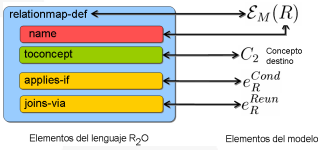
R_2O Basic Syntax




```
<conceptmap-def  name="Customer">
<identified-by> Table key </identified-by>
<uri-as> operation </uri-as>
<applies-if> condition </applies-if>
<joins-via> expression </joins-via>
<documentation>description …</documentation>
<described-by>attributes,relations</described-by>
</conceptmap-def>
```



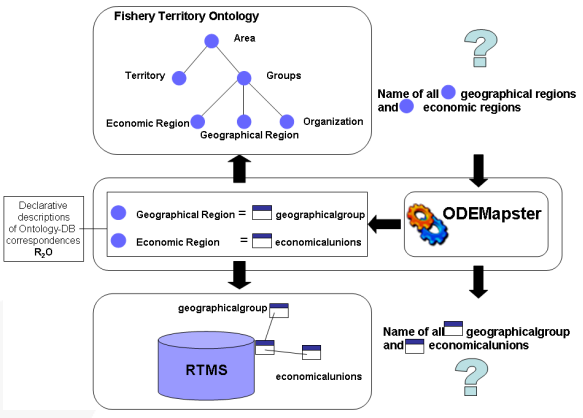

```
<attributemap-def name="http://esperonto/ff#Title">
<aftertransform>
<operation oper-id="constant">
<arg-restriction on-param="const-val">
<has-column>fsb_ajut.titol</has-column>
</arg-restriction>
</operation>
</aftertransform>
</attributemap-def>
```



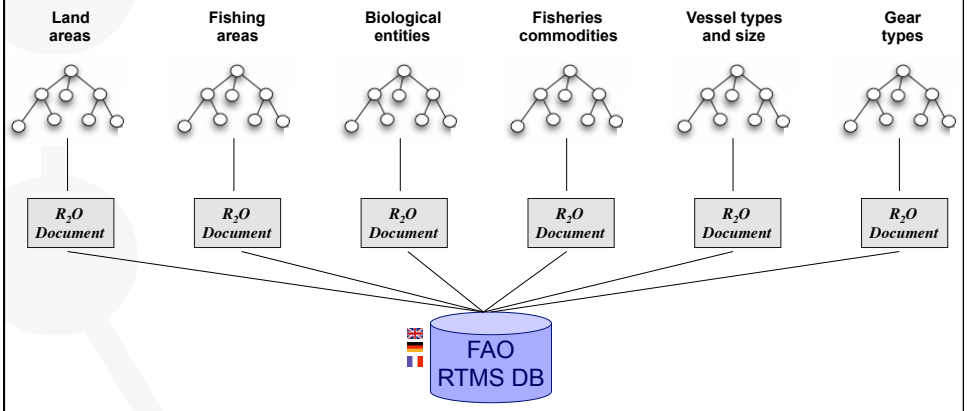

```
<relationmap-def name="http://esperonto/ff#isCandidateFor">
<to-concept name="http://esperonto/ff#FundOpp">
<joins-via>
<operation oper-id="equals">
<arg-restriction on-param="value1">
<has-column>fsb_ajut.id</has-column>
</arg-restriction>
<arg-restriction on-param="value2">
<has-column>fsb_candidate.forFund</has-column>
</arg-restriction>
</operation>
</joins-via>
</relationmap-def>
```

ODEMapster

- The ODEMapster processor generates Semantic Web instances from relational instances based on the mapping description expressed in the R_2O document
  - Batch process: DB records migrated to the ontology
  - On demand: Querying the DB in terms of ontological terms



FAO Use Case



Land areas	
Concepts	4
Properties	25
Instances	289

Fishing areas	
Concepts	5
Properties	14
Instances	134

Biological entities	
Concepts	5
Properties	21
Instances	11571

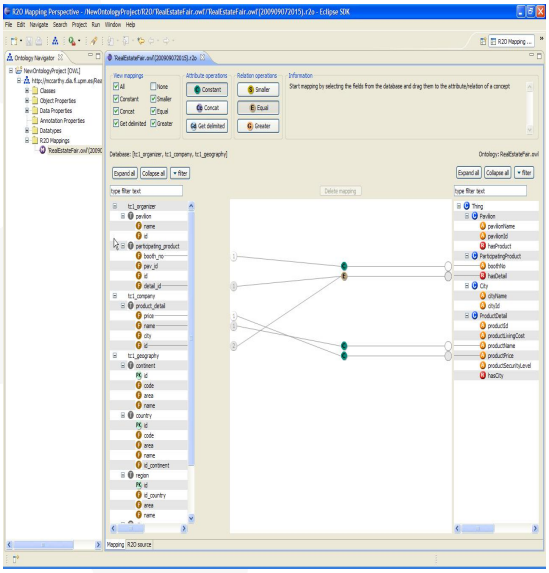
Fisheries commoditie	
Concepts	5
Properties	14
Instances	1380

Vessel types and size	
Concepts	5
Properties	20
Instances	120

Gear types	
Concepts	4
Properties	14
Instances	0

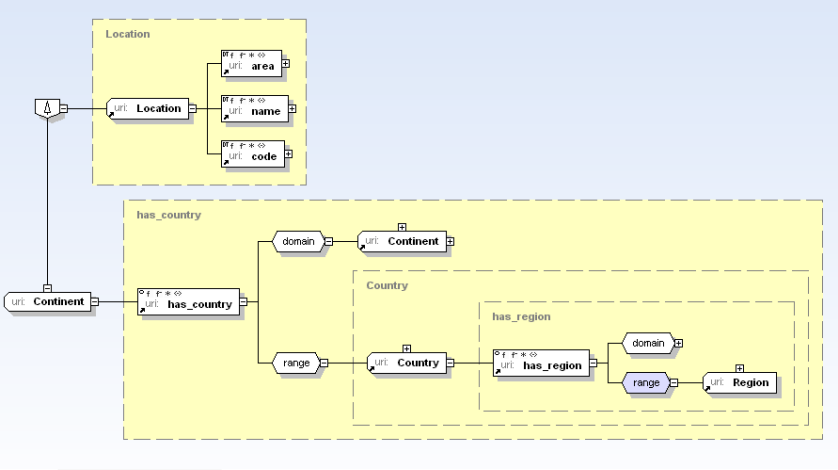
http://www.fao.org/aims/aos/fi/

Mapping Design

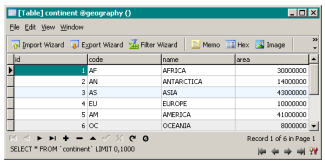


- 3 Mapping Creation Steps
  - Load Ontology
  - Load Database(s)
  - Create mapping
- 2 Usage Modes
  - Online mode (run time query execution)
  - Offline mode (materialized RDF dump)

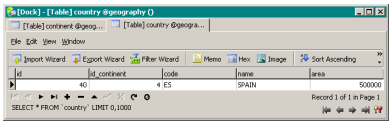
Example: Geography Ontology



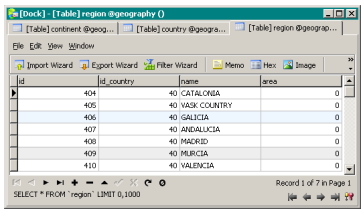
Geography Database



Continent Table

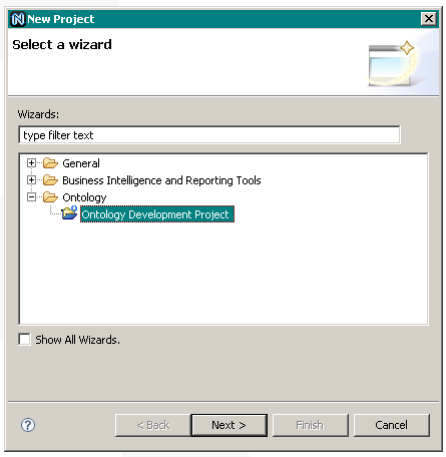


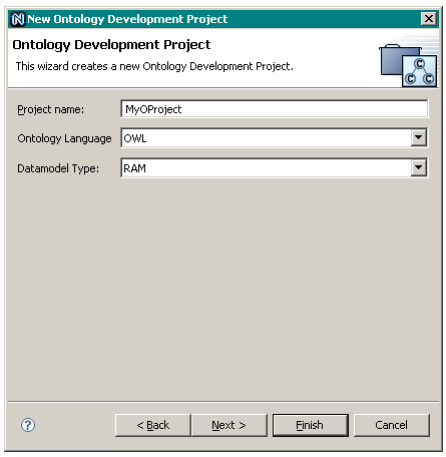
Country Table



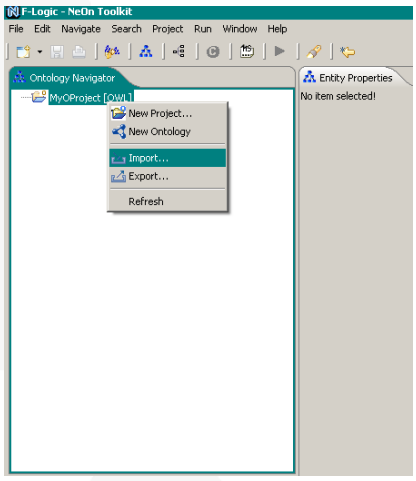
Region Table

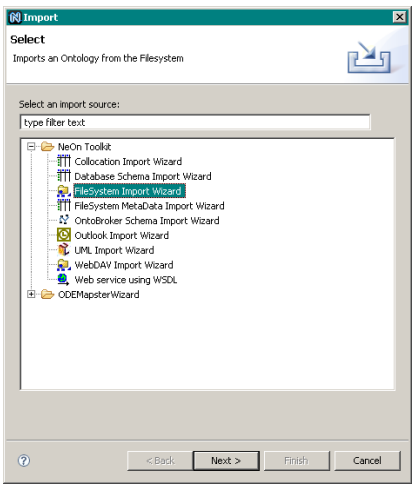
NeOn Toolkit – New Ontology Project



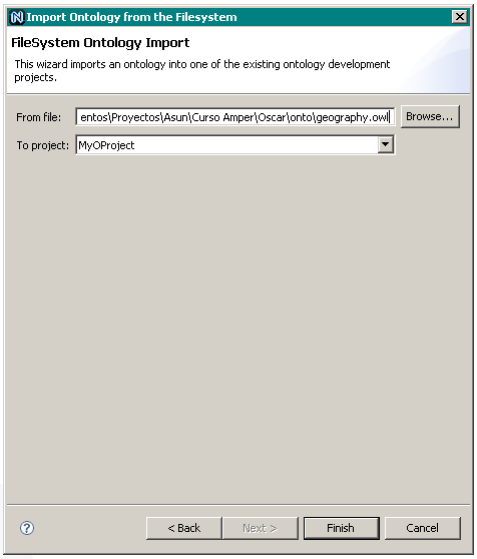


Import Ontology

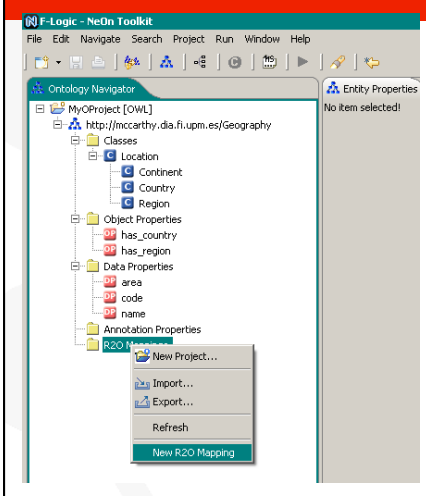


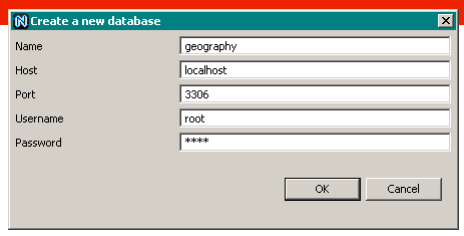


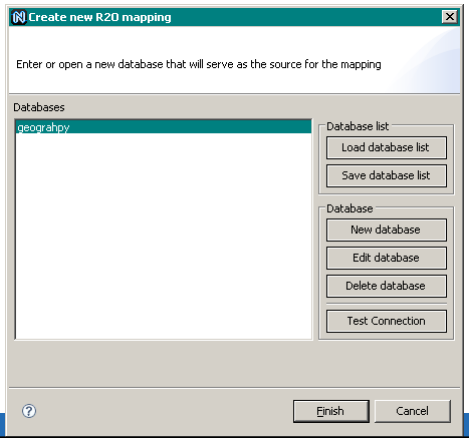
Import Ontology



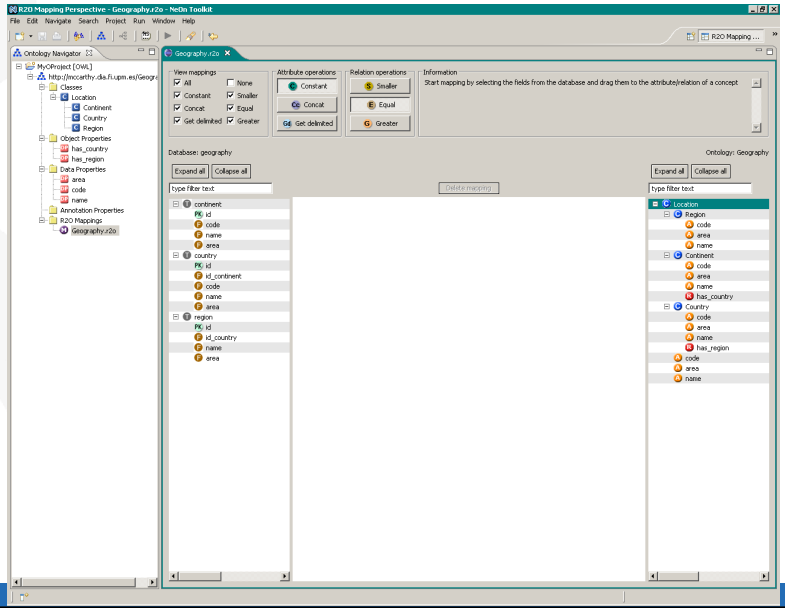
New R_2O Mapping – Database



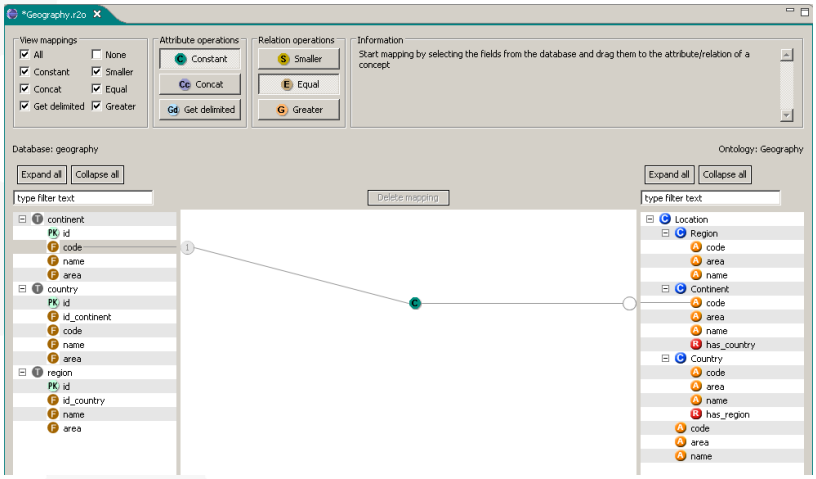




R_2O Mapping Perspective



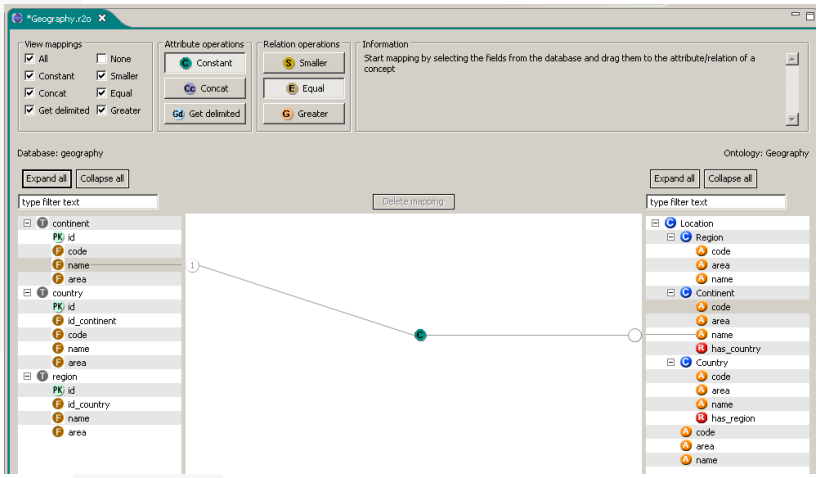
An attribute mapping example



An attribute mapping example – R_2O Code

```
<attributemap-def name="http://mccarthy.dia.fi.upm.es/Geography#code">
  <selector>
    <aftertransform>
      <operation oper-id="constant">
        <arg-restriction on-param="const-val">
          <has-column>continent.code</has-column>
        </arg-restriction>
      </operation>
    </aftertransform>
  </selector>
</attributemap-def>
```

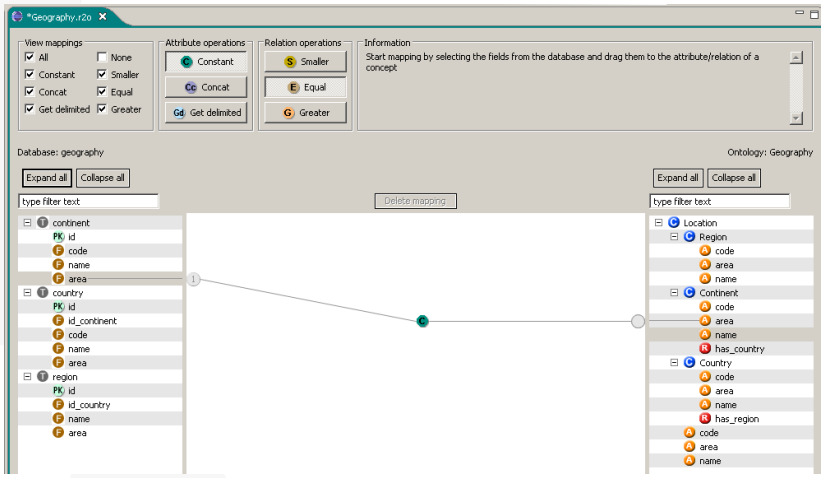
An attribute mapping example



An attribute mapping example – R_2O Code

```
<attributemap-def name="http://mccarthy.dia.fi.upm.es/Geography#name">
  <selector>
    <aftertransform>
      <operation oper-id="constant">
        <arg-restriction on-param="const-val">
          <has-column>continent.name</has-column>
        </arg-restriction>
      </operation>
    </aftertransform>
  </selector>
</attributemap-def>
```

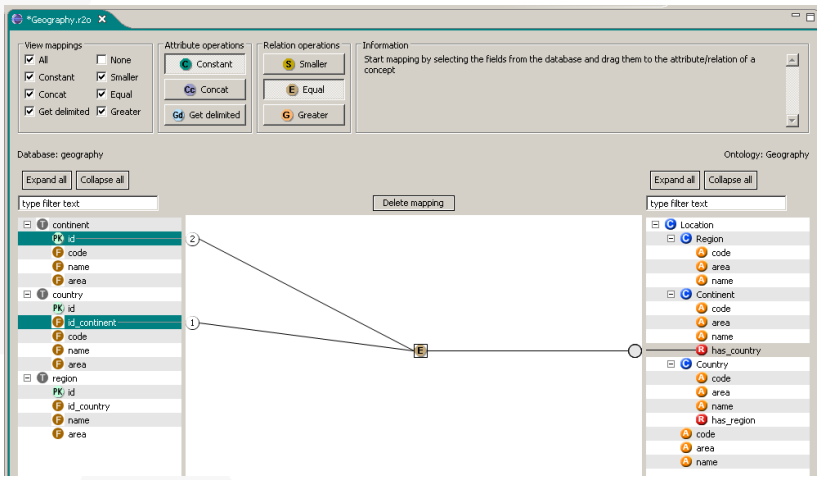
An attribute mapping example



An attribute mapping example – R_2O Code

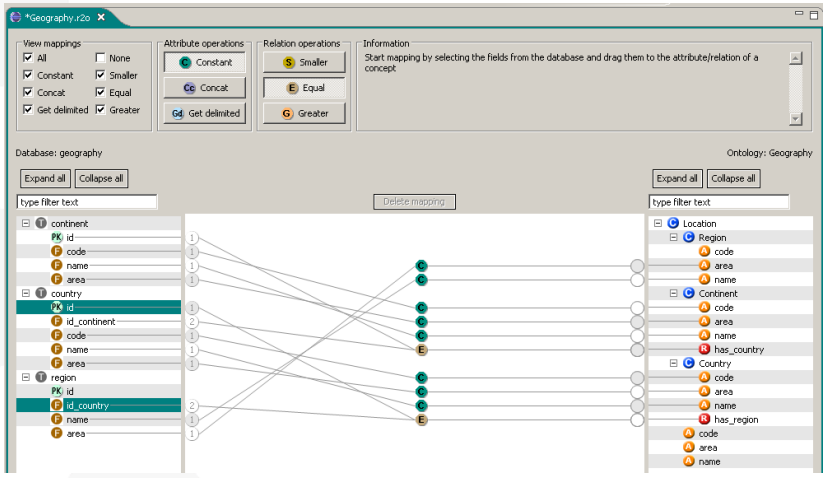
```
<attributemap-def name="http://mccarthy.dia.fi.upm.es/Geography#area">
  <selector>
    <aftertransform>
      <operation oper-id="constant">
        <arg-restriction on-param="const-val">
          <has-column>continent.area</has-column>
        </arg-restriction>
      </operation>
    </aftertransform>
  </selector>
</attributemap-def>
```

A relation mapping example



A relation mapping example – R_2O Code

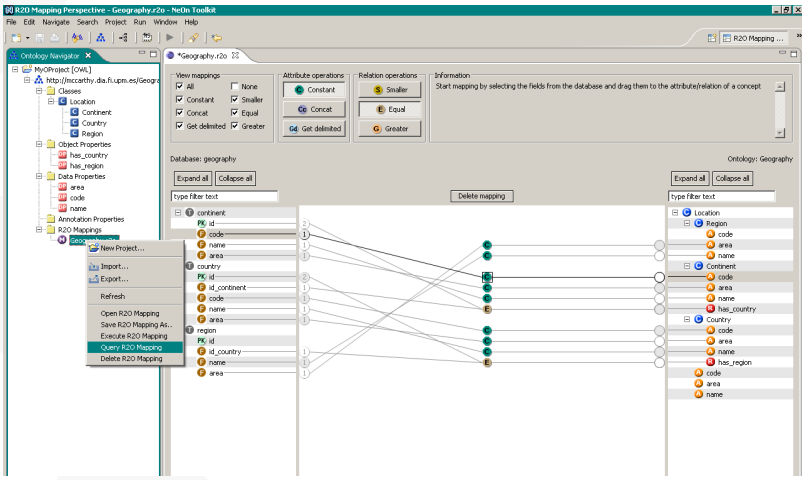
```
<dbrelationmap-def name="http://mccarthy.dia.fi.upm.es/Geography#has_country" toConcept="http://mccarthy.dia.fi.upm.es/Geography#Country">
  <joins-via>
    <condition oper-id="equals">
      <arg-restriction on-param="value1">
        <has-column>continent.id</has-column>
      </arg-restriction>
      <arg-restriction on-param="value2">
        <has-column>country.id_continent</has-column>
      </arg-restriction>
    </condition>
  </joins-via>
</dbrelationmap-def>
```



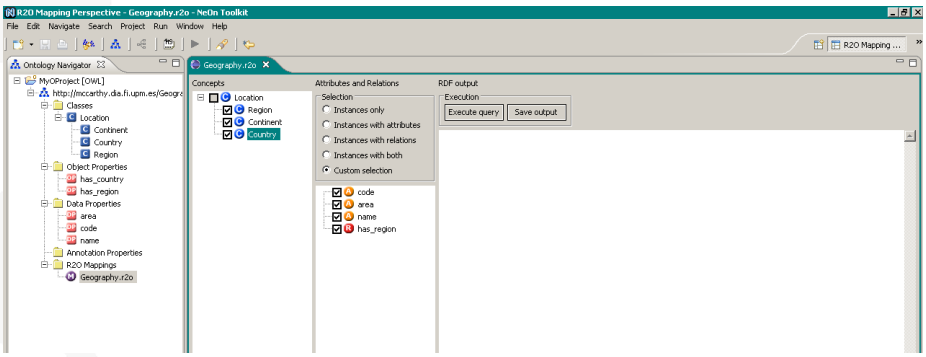
Concept mapping example – uri-as

```
<conceptmap-def name="http://mccarthy.dia.fi.upm.es/Geography#Region">
  <uri-as>
    <operation oper-id="concat">
      <arg-restriction on-param="string1">
        <has-value>http://mccarthy.dia.fi.upm.es/Geography#Region</has-value>
      </arg-restriction>
      <arg-restriction on-param="string2">
        <has-column>region.id</has-column>
      </arg-restriction>
    </operation>
  </uri-as>
  <described-by>
</conceptmap-def>
```

Querying the Ontology Instances



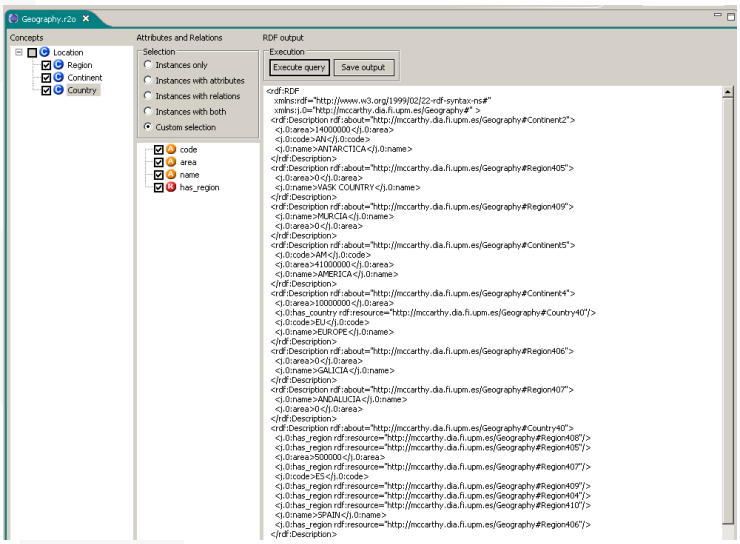
Querying the Ontology Instances



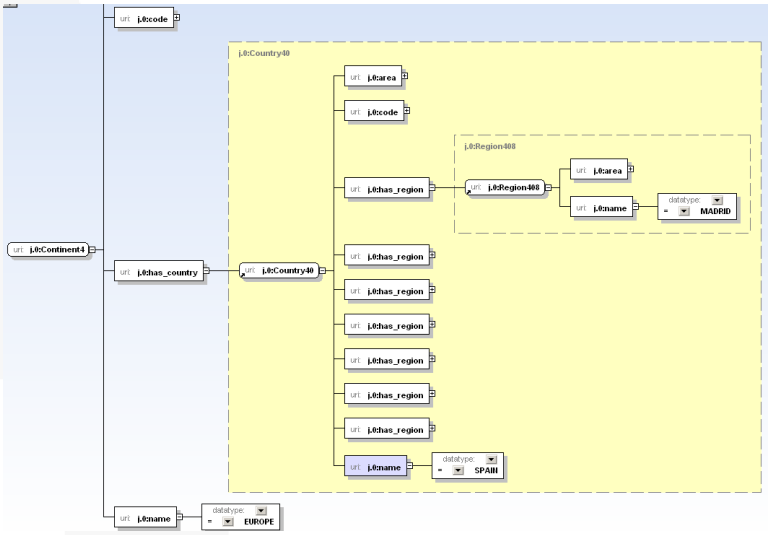
Querying the Ontology Instances – [illegible]

```
<onConcept conceptUri="http://mccarthy.dia.fi.upm.es/Geography#Country">
    <attSelect>
        <onAtt attName="http://mccarthy.dia.fi.upm.es/Geography#code"/>
    </attSelect>
    <attSelect>
        <onAtt attName="http://mccarthy.dia.fi.upm.es/Geography#area"/>
    </attSelect>
    <attSelect>
        <onAtt attName="http://mccarthy.dia.fi.upm.es/Geography#name"/>
    </attSelect>
    <relSelect>
        <onRel relName="http://mccarthy.dia.fi.upm.es/Geography#has_region"/>
    </relSelect>
</onConcept>
```

Retrieving the instances



Instance example



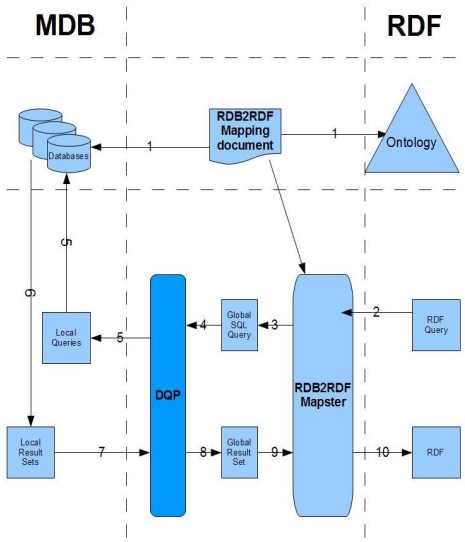
How to embed the ODEMapster Processor

```
public void testFunction() throws Exception
{
    Properties props = new Properties();

    props.setProperty(MapsterConnector.DATABASE_DRIVER, "com.mysql.jdbc.Driver");
    props.setProperty(MapsterConnector.DATABASE_URL, "jdbc:mysql:///rtms_figis");
    props.setProperty(MapsterConnector.DATABASE_USER,"root");
    props.setProperty(MapsterConnector.DATABASE_PWD, "root");
    props.setProperty(MapsterConnector.OUTPUT_FILE_PATH, "c:/develop/space/mapster/examples/jan/output.rdf");
    props.setProperty(MapsterConnector.ONTO_FILE_PATH, "c:/develop/space/mapster/examples/jan/onto.owl");
    props.setProperty(MapsterConnector.R2O_FILE_PATH, "c:/develop/space/mapster/examples/jan/f1.r2o");
    props.setProperty(MapsterConnector.QUERY_FILE_PATH, "c:/develop/space/mapster/examples/jan/fq1.xml");

    MapsterConnector mp = new MapsterConnector();
    mp.setProperties(props);
    mp.process();
}
```

MDB2RDF

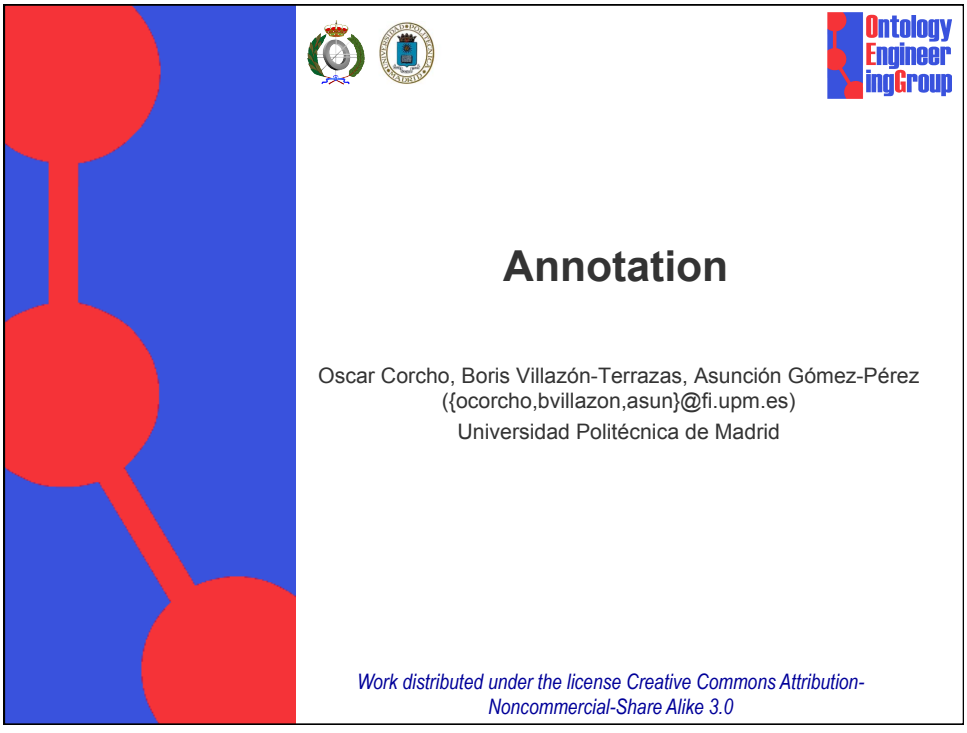


- Mapping defined
- RDF Query Posed
- RDF Query to Global Query
- Global query is sent to DQP
- DQP distributes global queries into local queries and evalute them
- Databases return local results
- Local results sent to DQP
- DQP integrates the results into global result
- Global result is sent to be transformed into RDF
- RDF documents containing RDF instances

63

Other topics not covered today

- Blog annotation tools
  - E.g., Zemanta
- Linguistic-based data annotation
  - E.g., GATE-based tools
- RDF data publishing
  - RDFa
  - Linked Data
- Ontology-based data integration techniques



Annotation

Oscar Corcho, Boris Villazón-Terrazas, Asunción Gómez-Pérez
({ocorcho,bvillazon,asun}@fi.upm.es)
Universidad Politécnica de Madrid