

Annotation

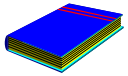
Oscar Corcho, Boris Villazón Terrazas, Asunción Gómez-Pérez
{ocorcho,bvillazon,asun}@fi.upm.es
http://www.oeg-upm.net/

Ontological Engineering Group
Laboratorio de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo sn,
28660 Boadilla del Monte, Madrid, Spain

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 1605
Type of aircraft: non referenced/B

Based on thesauri

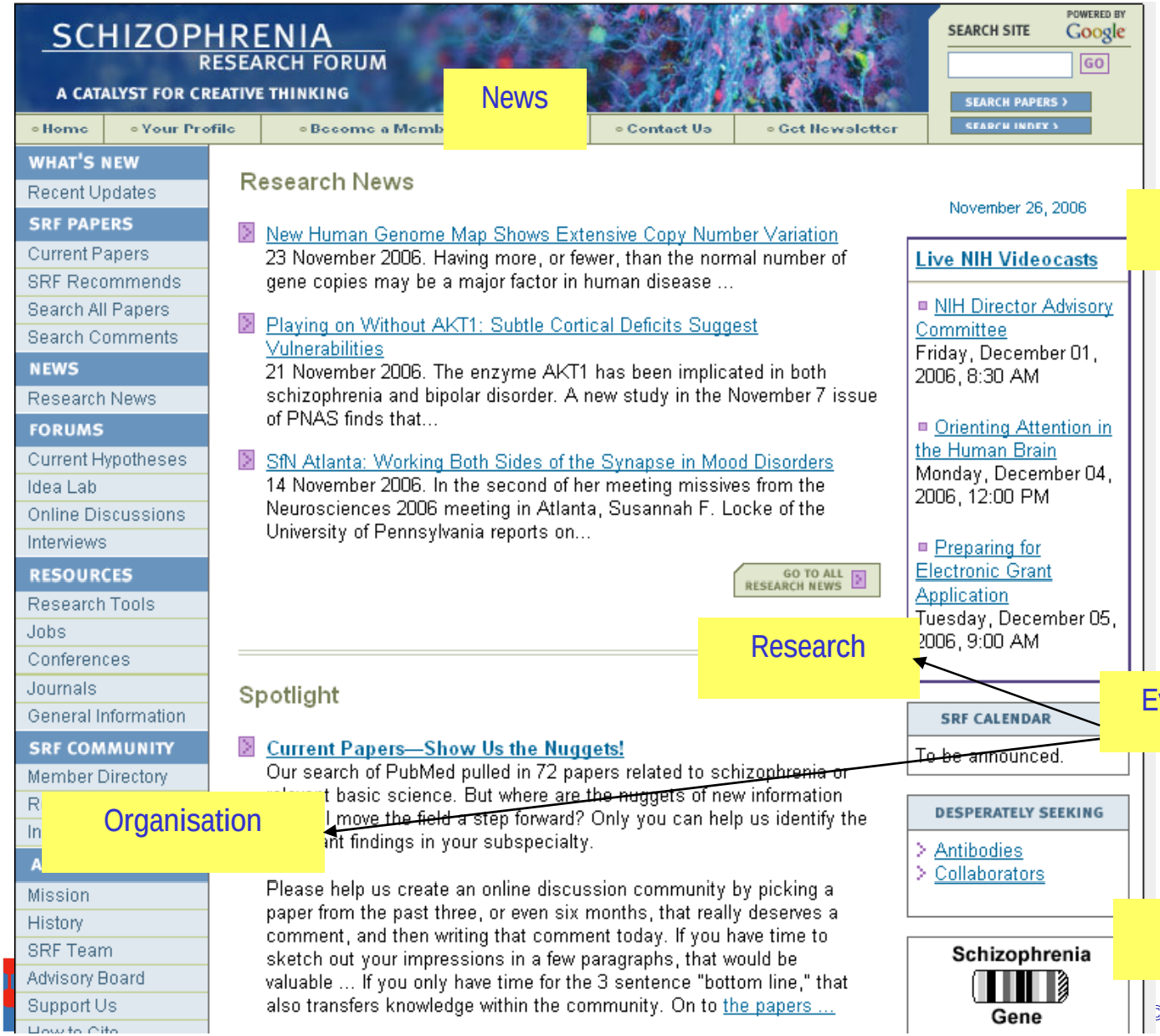
Madrid is a reference to the term with ID 7010413 in the thesaurus, which refers to the city of Madrid in Spain.
Spain is a reference to the term with ID 1000095, which refers to the kingdom of Spain in Europe.
Chicago is a reference to the term with ID 7013596, which refers to the city of Chicago in 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

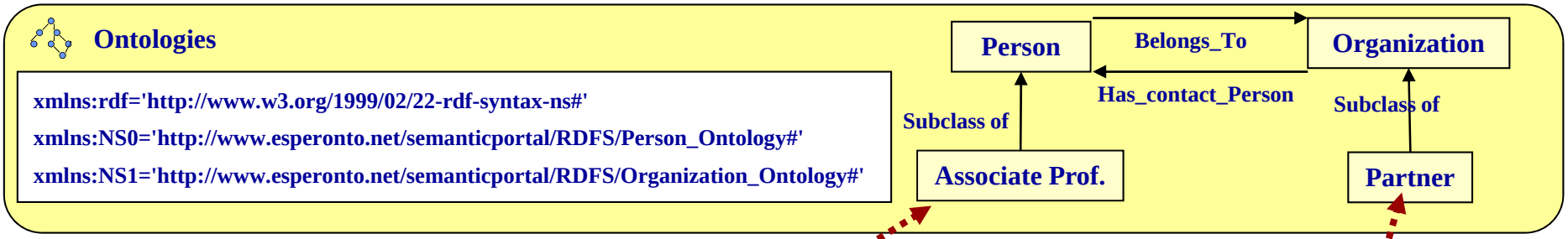
Concept instances relate a part of the document to one or several concepts in an ontology. For example, "Flight details" may represent an instance of the concept `Flight`, and can be named as `AA7615_Feb08_2003`, although concept instances do not necessarily have a name.
Attribute values relate a concept 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)*



Ontologies and Metadata



Instance of

Instance of

Annotation (RDF)

```
<rdf:Description rdf:about='Asunción Gómez-Pérez'>
<rdf:type rdf:resource='Associate Prof'/>
<NS0:Full_Name>A. GomezPerez</NS0:Full_Name>
<NS0:Belongs_To>UPM</NS0: Belongs_To >
<NS0:e-mail>asun@fi.upm.es</NS0:e-mail>
```

```
<rdf:Description rdf:about='UPM'>
<rdf:type rdf:resource='Partner'/>
<NS1:Acronym>UPM</NS1:Acronym>
<NS1:Has_Contact_Person>Asunción Gómez-Pérez
        </NS1:Has_Contact_Person >
```

Web Page

Full Name	Asuncion Gomez-Perez
e-mail	asun@fi.upm.es
Photo	

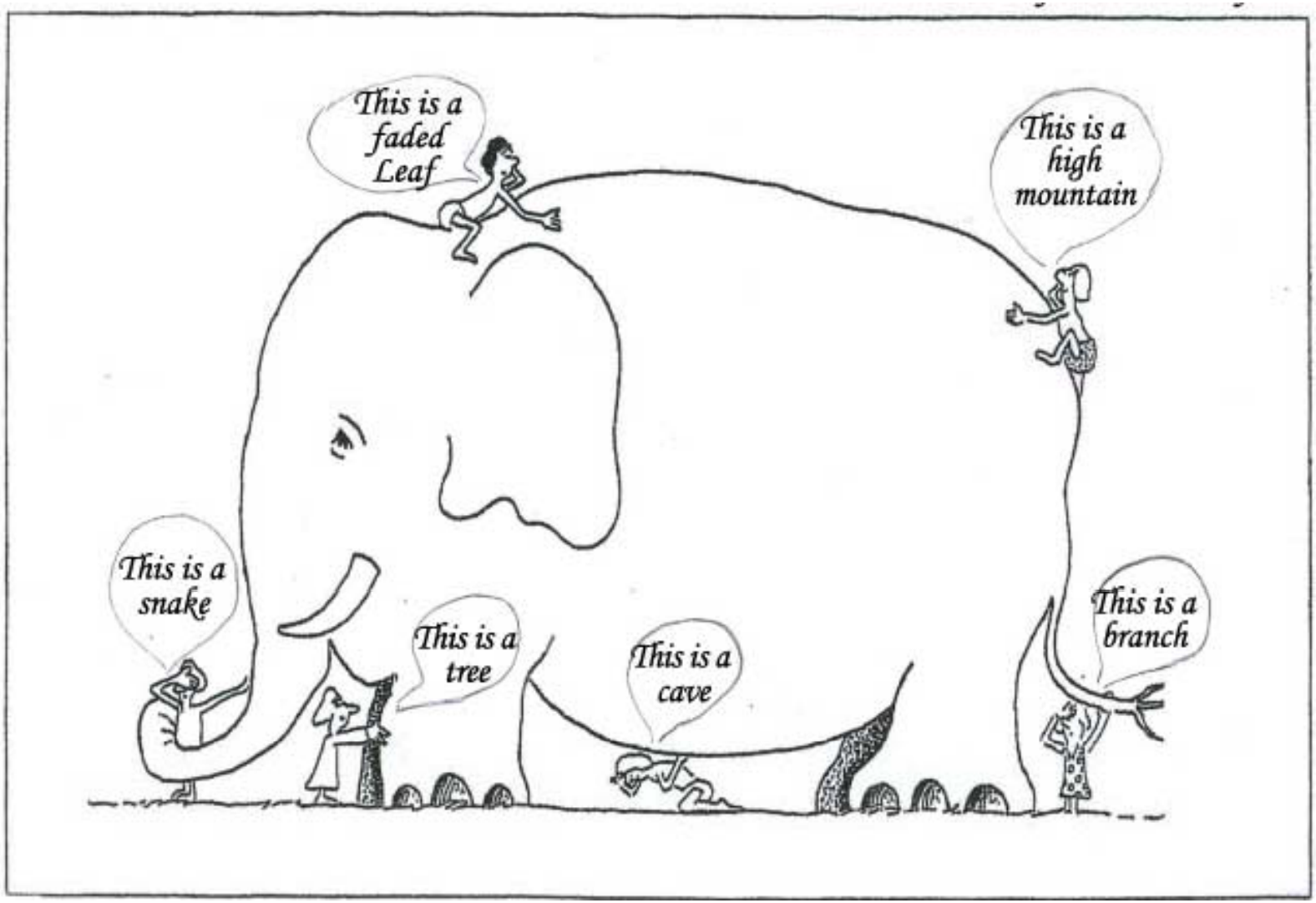
Asunción Gómez-Pérez is contact person UPM (Partner).
Asunción Gómez-Pérez belongs to UPM (Partner).

Full Name	Universidad Politécnica de Madrid
Acronym	UPM
Logo	UPM

UPM has contact person Asunción Gómez-Pérez (Associate Professor).
UPM participates in Esperonto (Project).
UPM team is formed by :

URL http://www.esperonto.net http://www.esperonto.net

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)
  - …

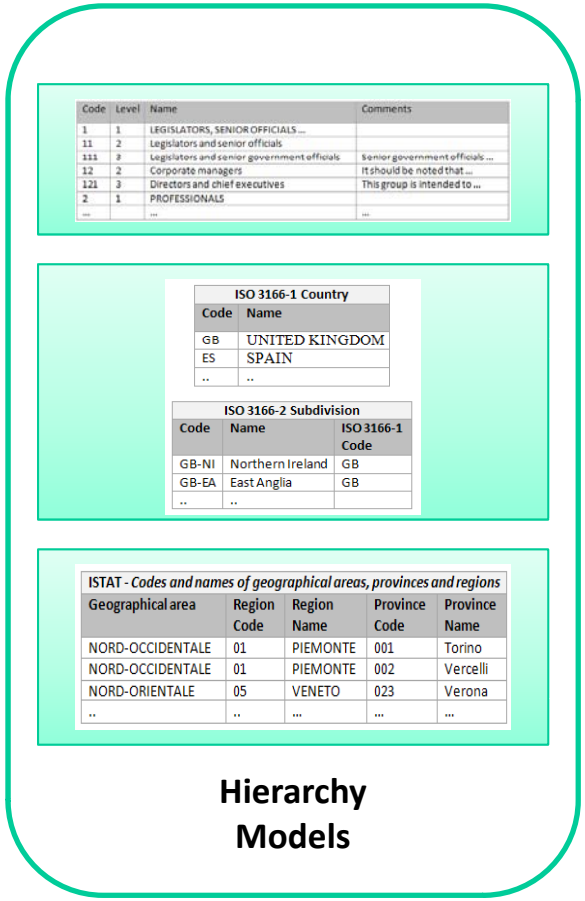
```
<html>
<head><TITLE> Richard Benjamins </TITLE>
<a ONTO="page:Researcher"> </a>
</head>

<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
```

Motivation

Modeled



Code	Level	Name	Comments
1	1	LEGISLATORS, SENIOR OFFICIALS ...	
11	2	Legislators and senior officials	
111	3	Legislators and senior government officials	Senior government officials ...
12	2	Corporate managers	It should be noted that ...
121	3	Directors and chief executives	This group is intended to ...
2	1	PROFESSIONALS	
...		...	...

ISO 3166-1 Country

Code	Name
GB	UNITED KINGDOM
ES	SPAIN
..	..

ISO 3166-2 Subdivision

Code	Name	ISO 3166-1 Code
GB-NI	Northern Ireland	GB
GB-EA	East Anglia	GB
..	..	

ISTAT - Codes and names of geographical areas, provinces and regions

Geographical area	Region Code	Region Name	Province Code	Province Name
NORD-OCCIDENTALE	01	PIEMONTE	001	Torino
NORD-OCCIDENTALE	01	PIEMONTE	002	Vercelli
NORD-ORIENTALE	05	VENETO	023	Verona
..	..	...	...	...

Hierarchy Models

Generate

The early days of the Semantic Web

Availability of content

– Including the “Deep Web” and Web Services

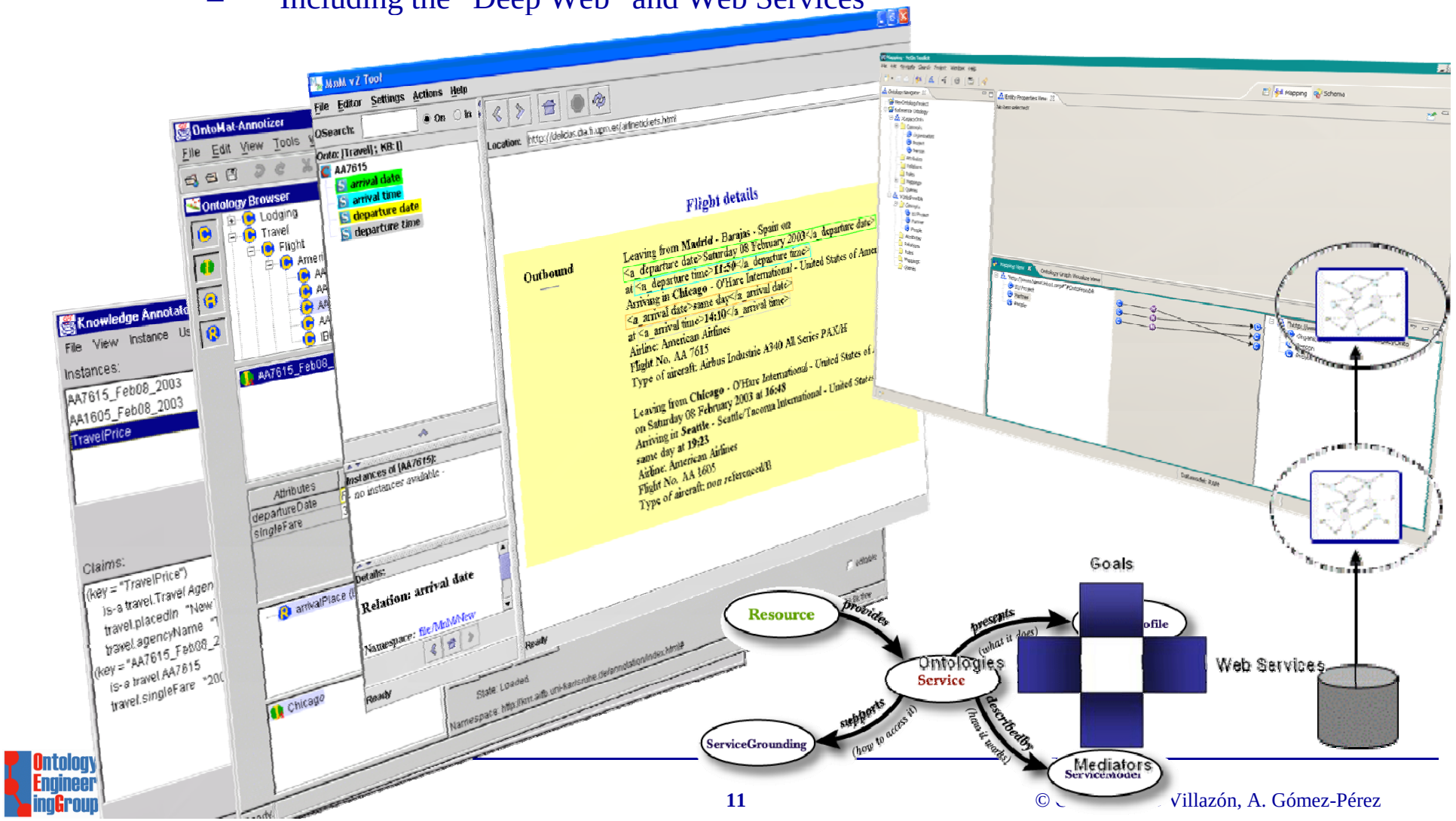


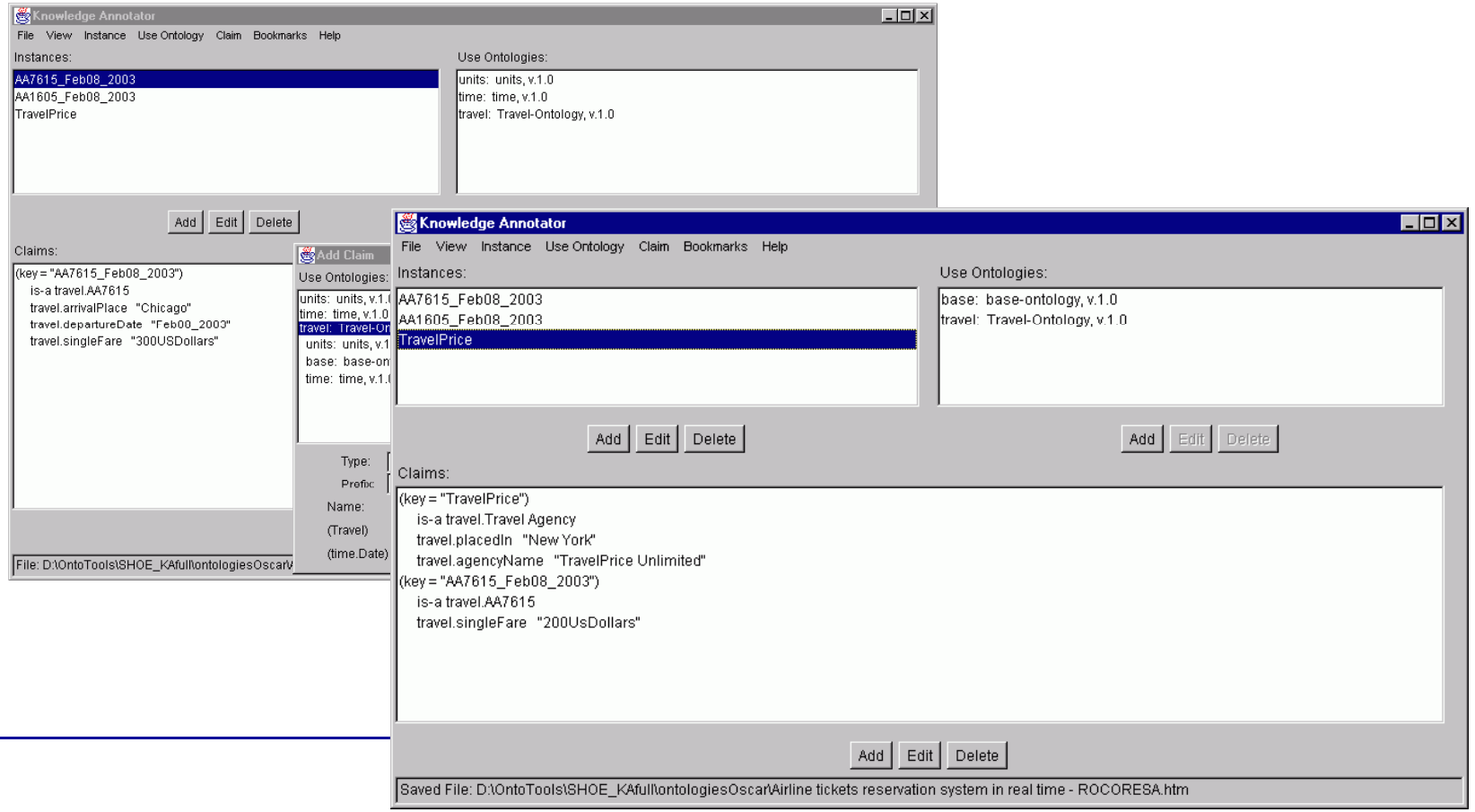
Table of Contents

Web Page Annotation. Dimensions

- **Sources**
  - Source type
    - Text: HTML, XML, PDF, etc.
    - Multimedia: images, video, audio, etc.
    - Web Services
  - Origin
    - Static: files
    - Dynamic: databases and forms
- **Used technologies**
  - Knowledge extraction
    - NLP, IE, Layout
  - Wrapper generation
    - Toolkits, ML, Browsing
- **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
  - Ontology-based

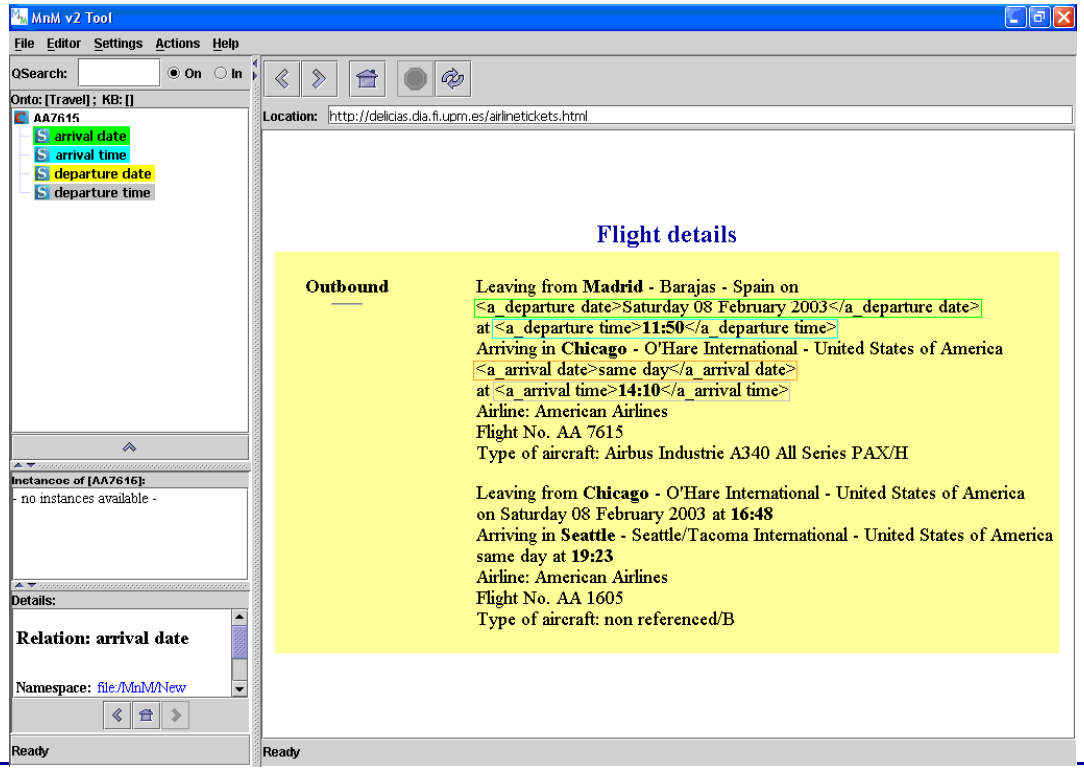
Annotation tools. SHOE Knowledge Annotator

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



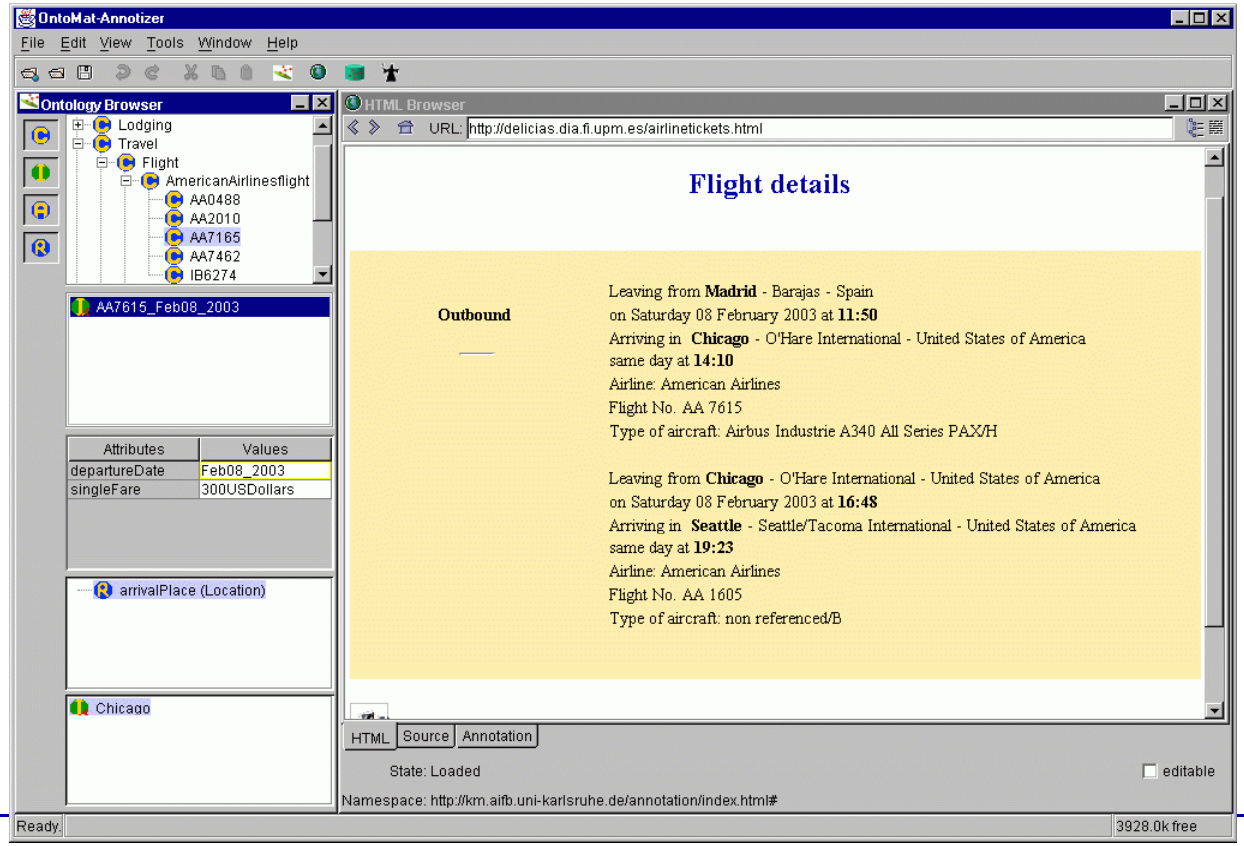
Annotation tools. MnM

- Standalone application
- Manual annotation with drag&drop
- Semi-automatic annotation with information extraction tools (Amilcare)
- OCML, RDF and XML



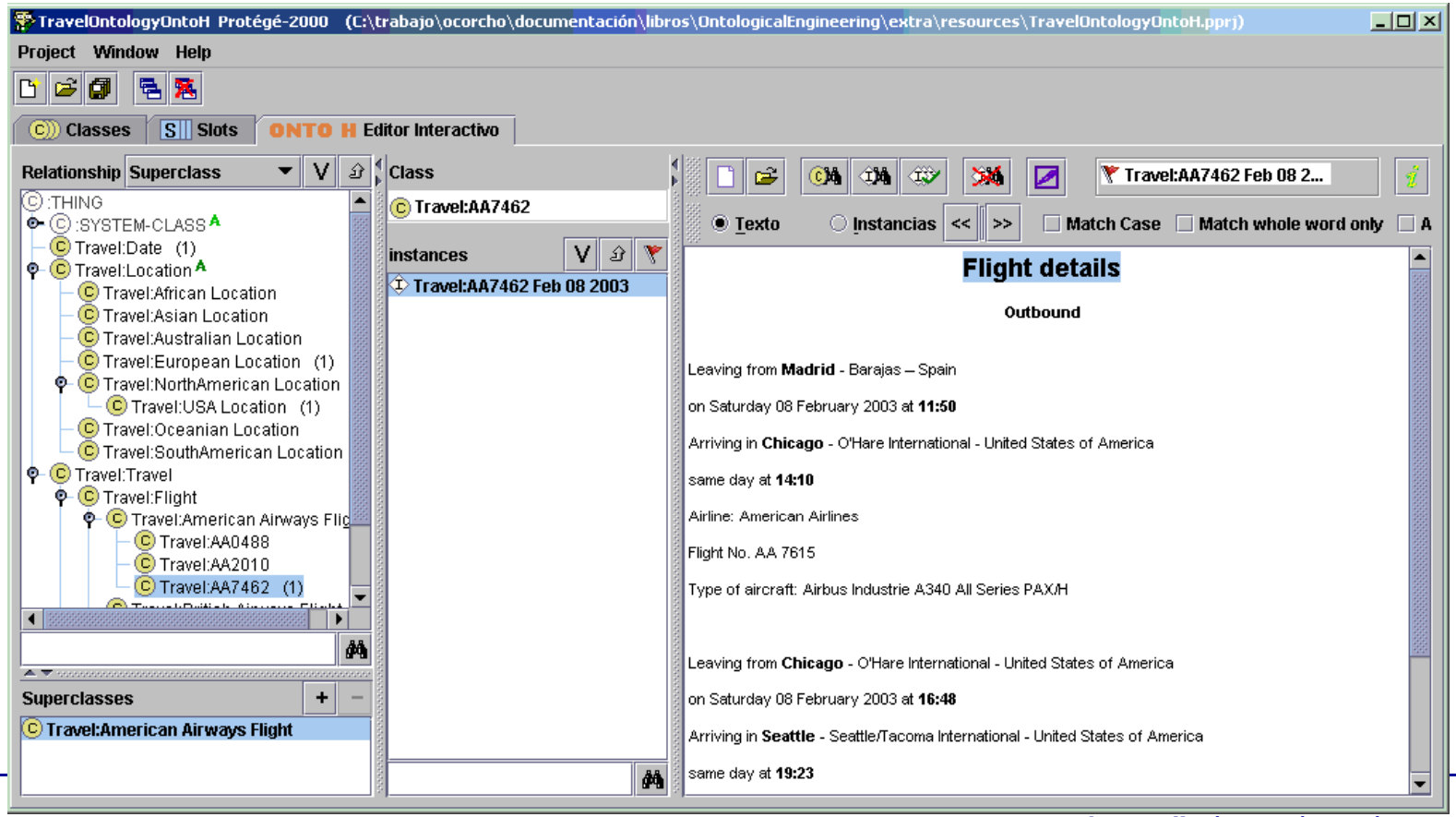
Annotation tools. OntoMat

- **Standalone application**
- **Manual annotation with drag&drop**
- **RDF and OWL**



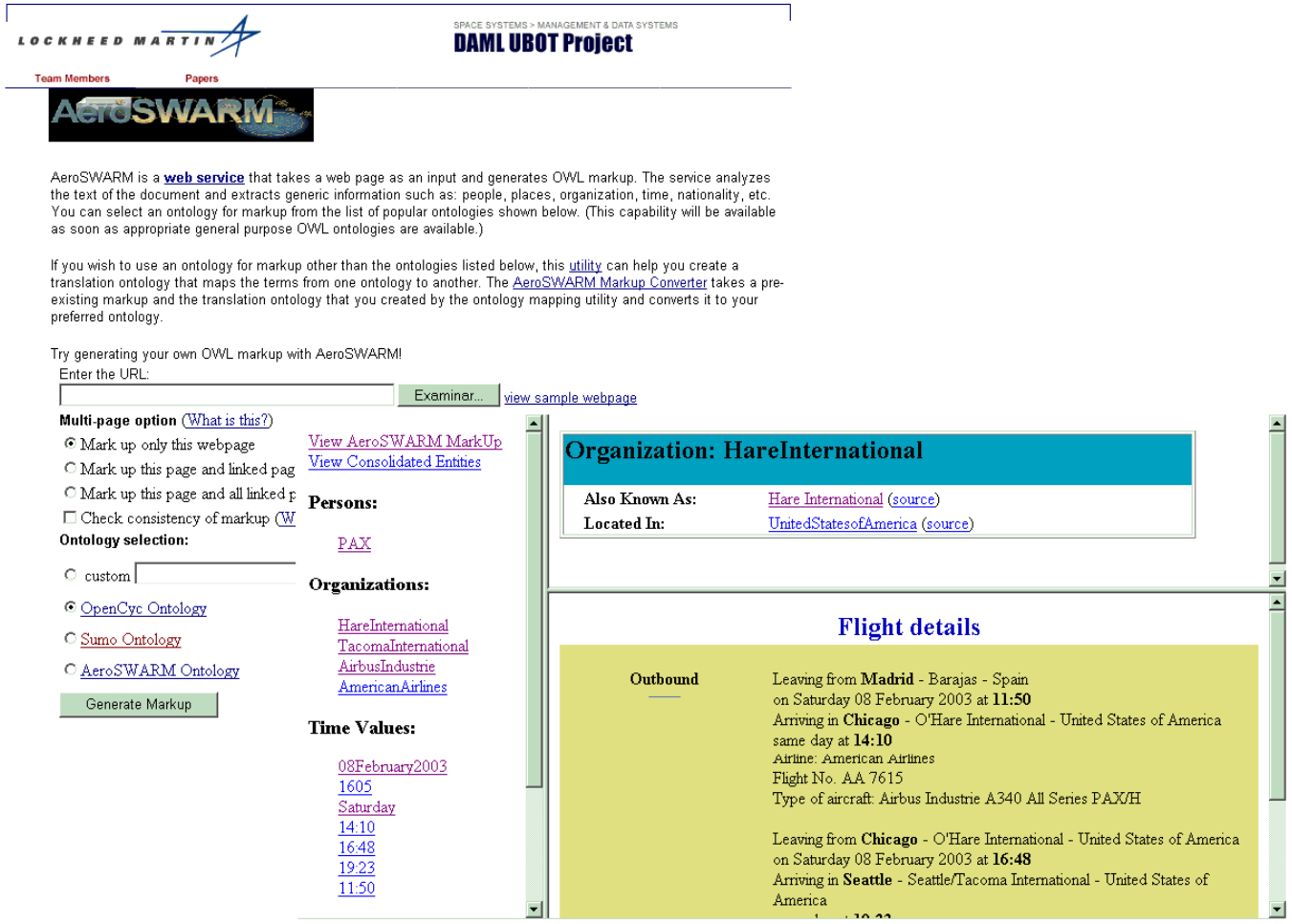
Annotation tools. ONTO-H

- **Protégé plug-in**
- **Manual annotation with drag&drop**
- **On cascade annotation, with annotation rules**



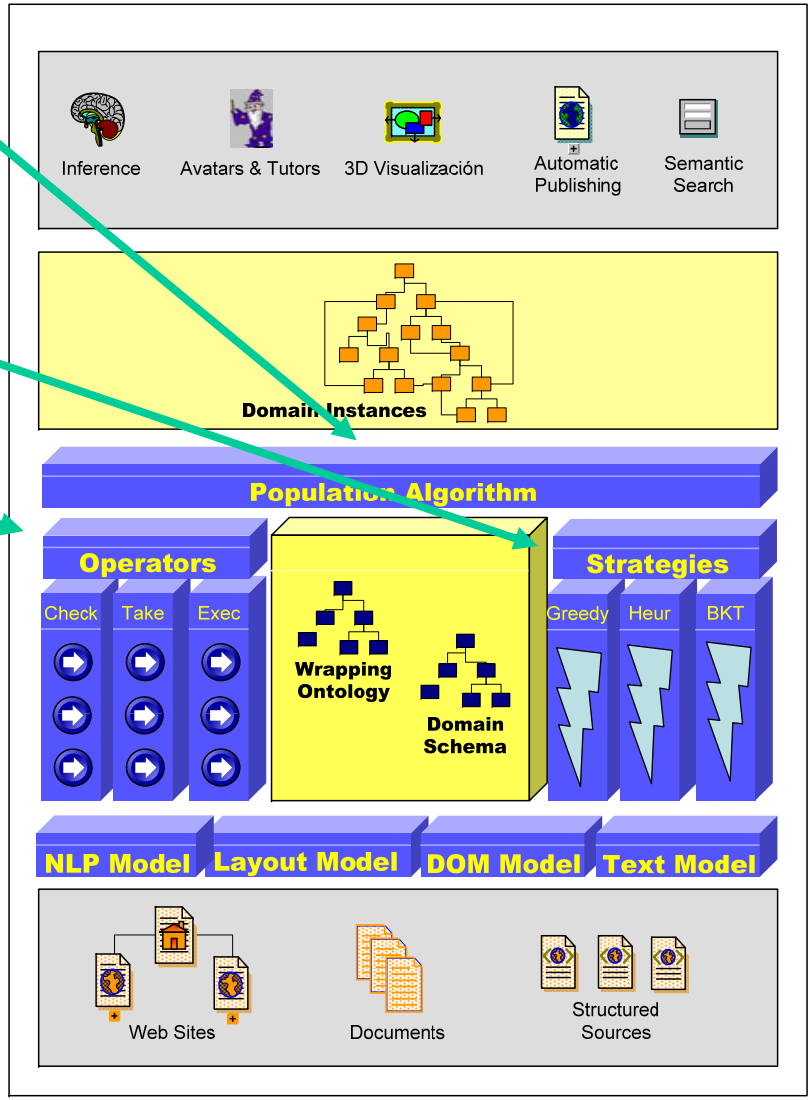
Annotation tools. AeroSWARM

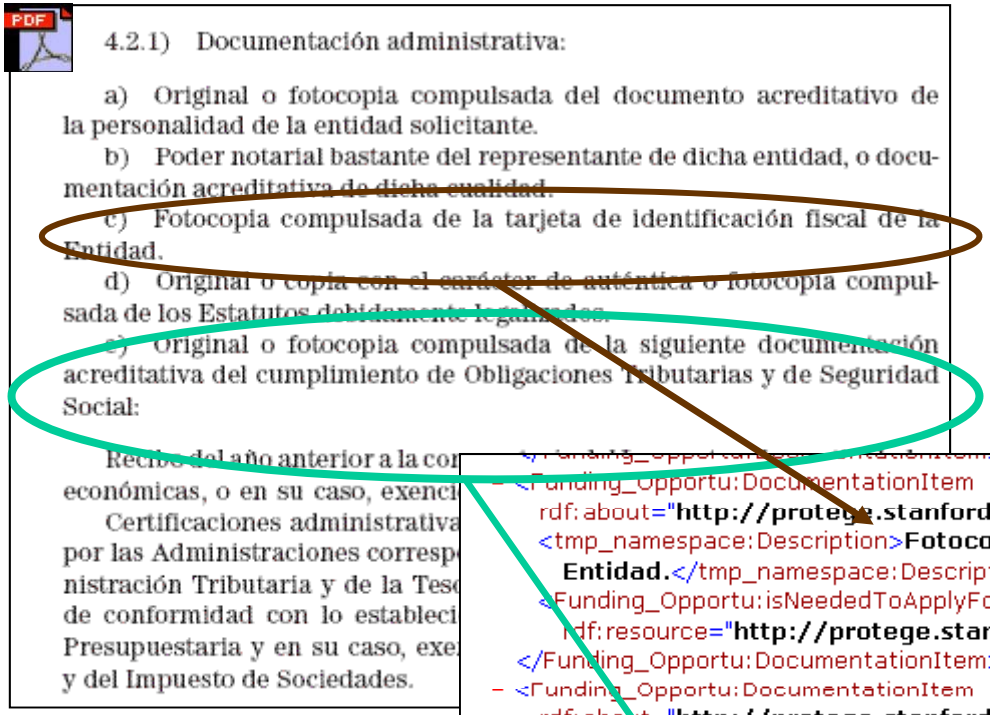
- **Web server for any Web document**
- **Automatic annotation with predefined ontologies: OpenCyc, SUMO and AeroSWARM**
- **RDF**



Annotation tools. Knowledge Parser

- **Semantic-based population**
  - Explicit wrapping knowledge
  - Bootstrapping
- **Strategies**
  - Heuristic
  - Backtracking
  - etc.
- **Operators**
  - In-Row
  - Is-Proper-Name
  - Is-Integer-Greater-Than, etc.
- **Pre-processing**
  - Natural language
  - Layout
  - XML/DOM
  - Plain text





4.2.1) Documentación administrativa:

a) Original o fotocopia compulsada del documento acreditativo de la personalidad de la entidad solicitante.

b) Poder notarial bastante del representante de dicha entidad, o documentación acreditativa de dicha cualidad.

c) Fotocopia compulsada de la tarjeta de identificación fiscal de la Entidad.

d) Original o copia con el carácter de auténtica o fotocopia compulsada de los Estatutos debidamente legalizados.

e) Original o fotocopia compulsada de la siguiente documentación acreditativa del cumplimiento de Obligaciones Tributarias y de Seguridad Social:

Recibo del año anterior a la co
económicas, o en su caso, exenci
Certificaciones administrativa
por las Administraciones corresp
nistración Tributaria y de la Tes
de conformidad con lo estableci
Presupuestaria y en su caso, exe
y del Impuesto de Sociedades.

```
- <Funding_Opportu:DocumentationItem
    rdf:about="http://protege.stanford.edu/tmp_namespace#tmp_Instance_10032" rdfs:label="BOE">
    <tmp_namespace:Description>Fotocopia compulsada de la tarjeta de identificación fiscal de la
      Entidad.</tmp_namespace:Description>
    <Funding_Opportu:isNeededToApplyFor
      rdf:resource="http://protege.stanford.edu/tmp_namespace#tmp_Instance_10004" />
  </Funding_Opportu:DocumentationItem>
- <Funding_Opportu:DocumentationItem
    rdf:about="http://protege.stanford.edu/tmp_namespace#tmp_Instance_10033" rdfs:label="BOE">
    <tmp_namespace:Description>Original o copia con el carácter de auténtica o fotocopia compulsada de
      los Estatutos debidamente legalizados.</tmp_namespace:Description>
    <Funding_Opportu:isNeededToApplyFor
      rdf:resource="http://protege.stanford.edu/tmp_namespace#tmp_Instance_10004" />
  </Funding_Opportu:DocumentationItem>
- <Funding_Opportu:DocumentationItem
    rdf:about="http://protege.stanford.edu/tmp_namespace#tmp_Instance_10034" rdfs:label="BOE">
    <tmp_namespace:Description>Original o fotocopia compulsada de la siguiente documentación
      acreditativa del cumplimiento de Obligaciones Tributarias y de Seguridad Social: Recibo del año
      anterior a la convocatoria del Impuesto sobre actividades económicas, o en su caso, exención
      concedida por el órgano competente. Certificaciones administrativas con el carácter de positivas
      expedidas por las Administraciones correspondientes de la Agencia Estatal de Administración
      Tributaria y de la Tesorería Territorial de la Seguridad Social, de conformidad con lo establecido en
      el artículo 81 de la Ley General Presupuestaria y en su caso, exenciones del Impuesto del Valor
      Añadido y del Impuesto de Sociedades.</tmp_namespace:Description>
```

Semantic wikis

RDF presentation

Rhizome

HTML + RDF page edition

IkeWiki

SemanticMediaWiki

SemPerWiki

SweetWiki

WikSAR

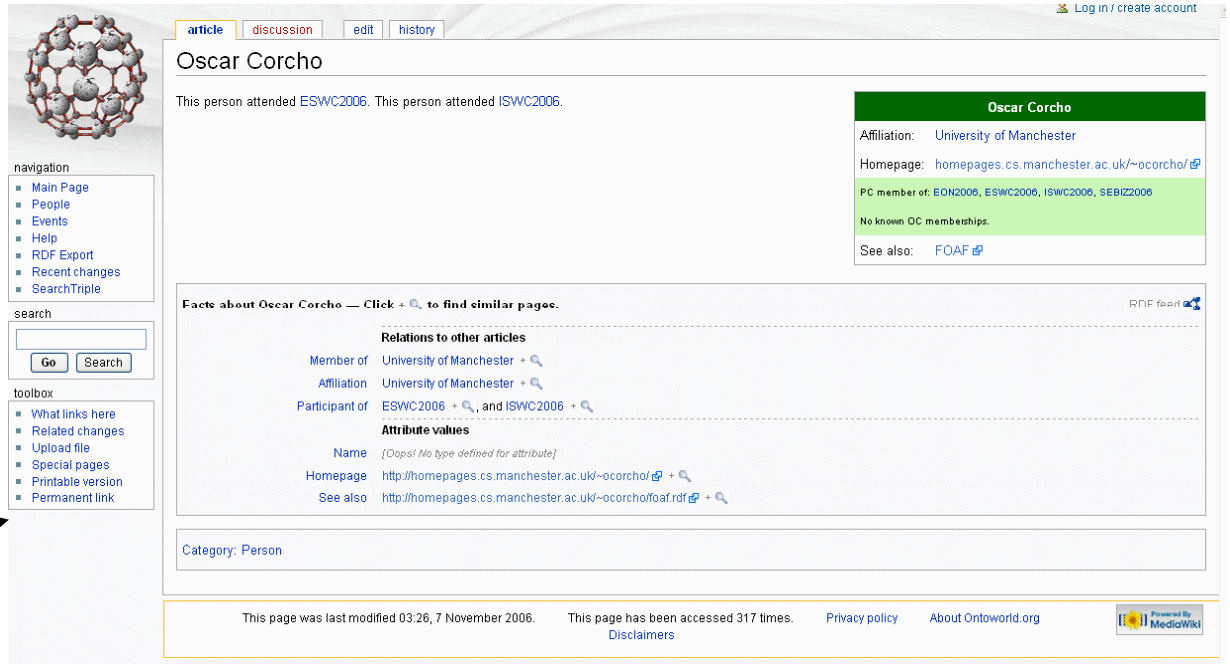
Ontology-based population

OntoWiki

COW

Non-ontology based attribute-value annotation

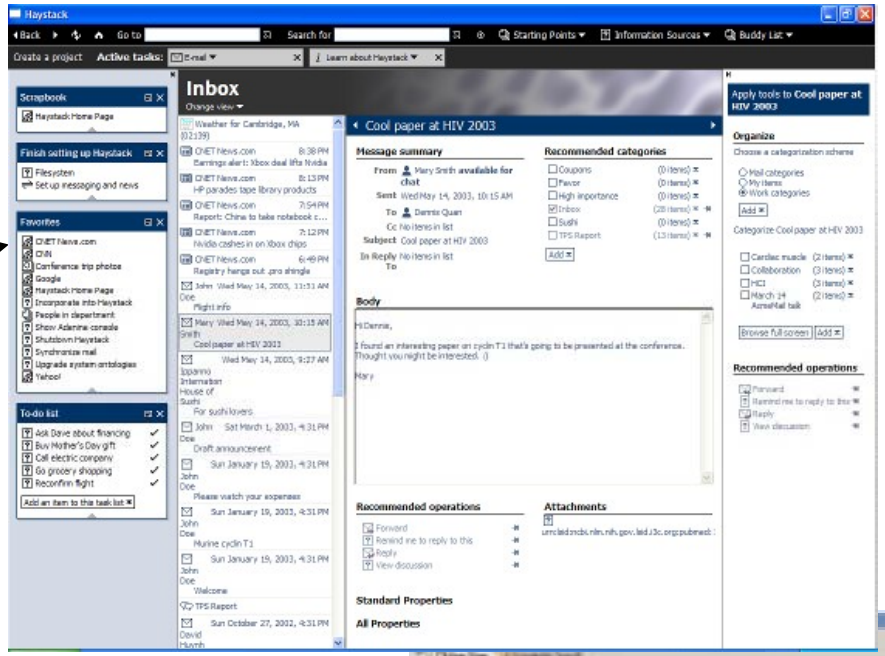
DiamondWiki



SemanticWikipedia

Semantic desktop
(including semantic e-mail)

Haystack
Gnowsis
D-Bin
OpenIris
NEPOMUK



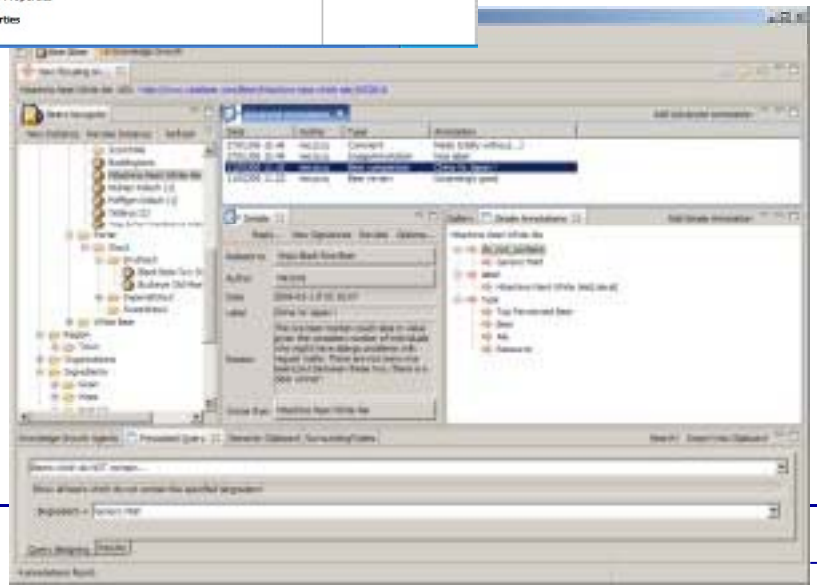
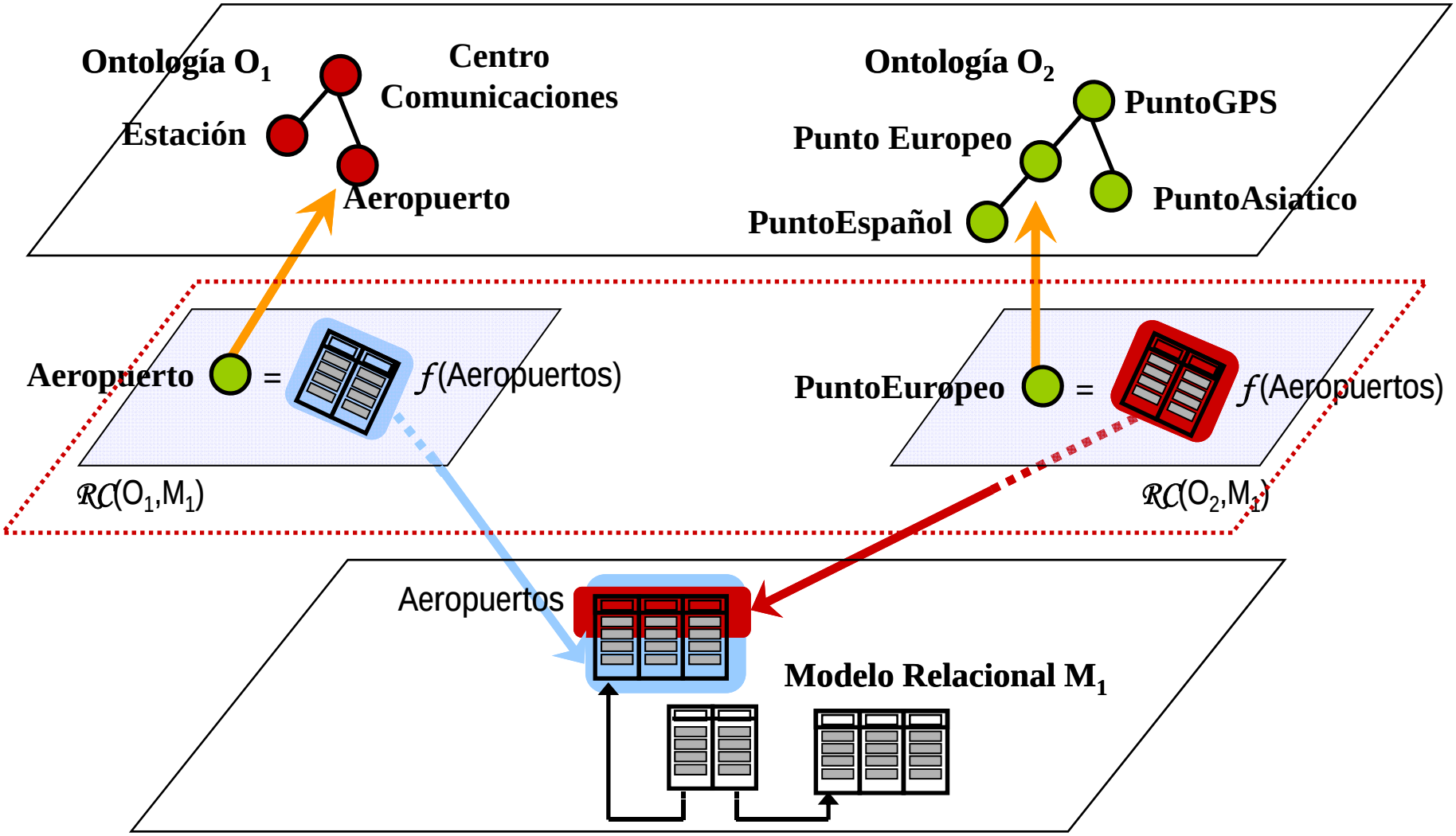
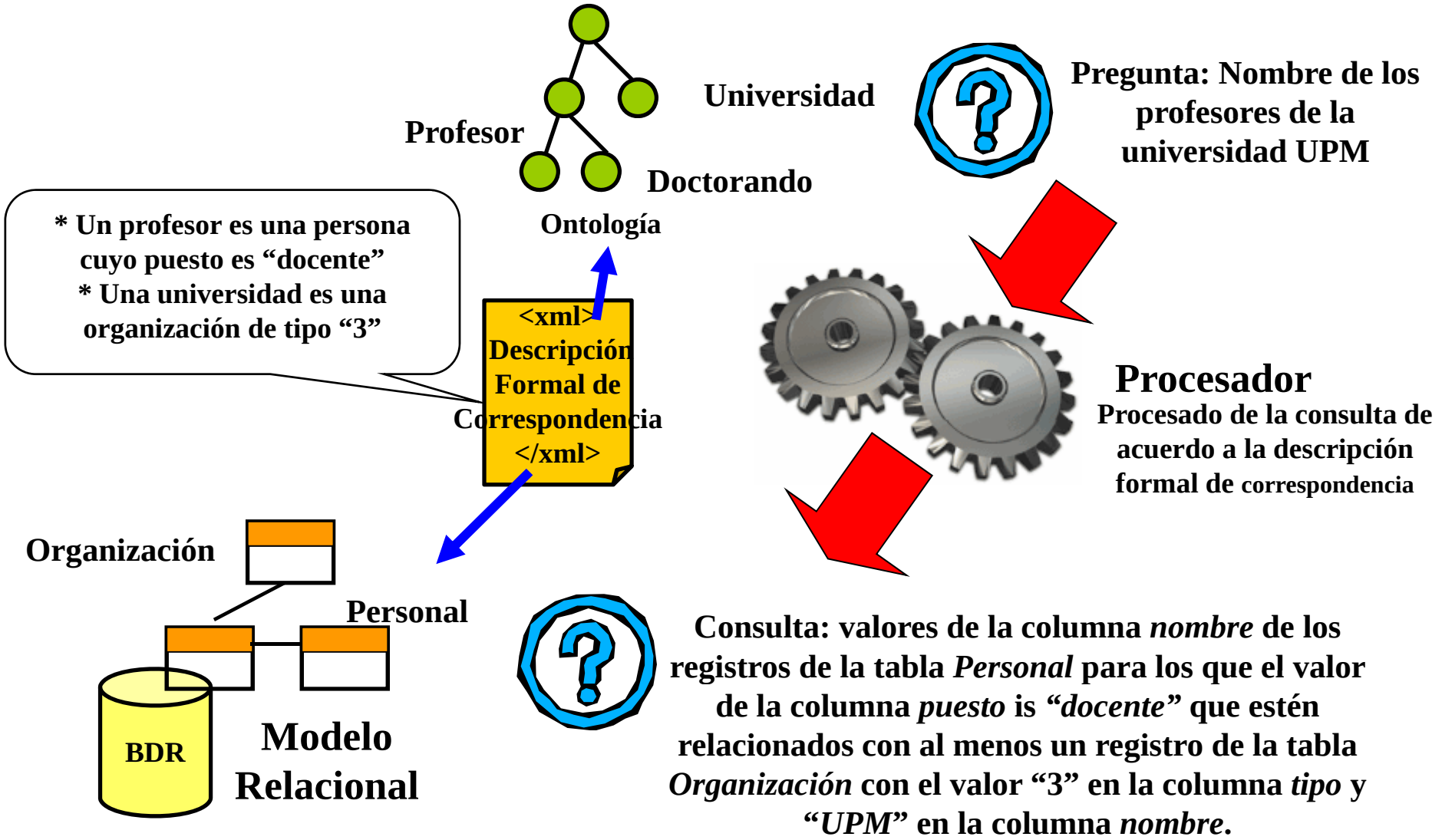


Table of Contents

Ontology-based view over a relational model (I)



Ontology-based view over a relational model (II)

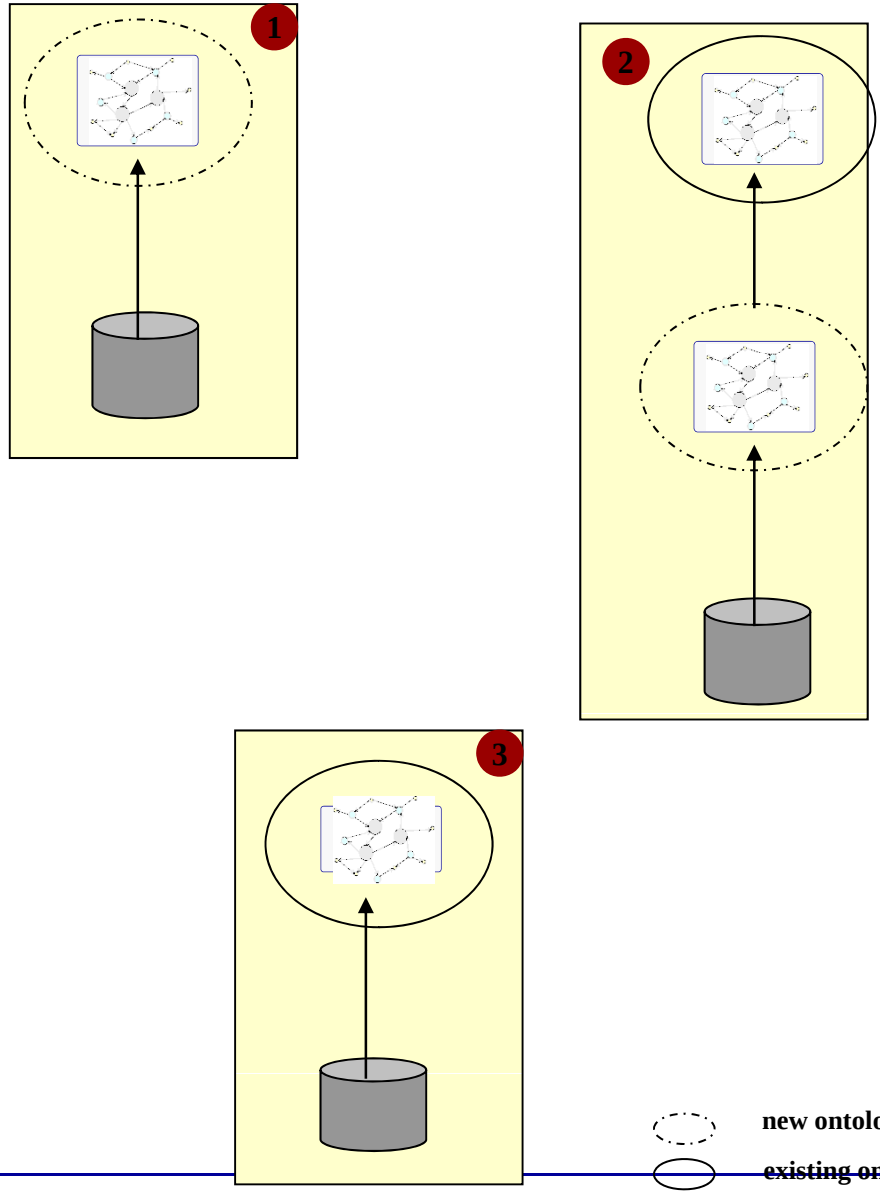


Existing approaches

1. To build a new ontology from a database schema and content (OntoStudio, KaOn Reverse)

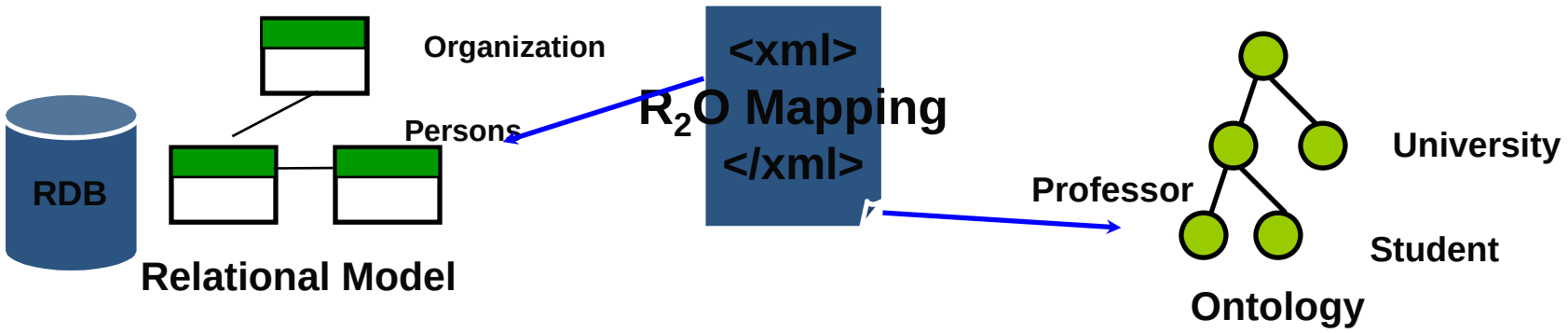
2. To map the ontology created in approach (1) to a legacy ontology (NeOn toolkit UKARL)

3. To map an existing DB to a legacy ontology (NeOn Toolkit UPM)



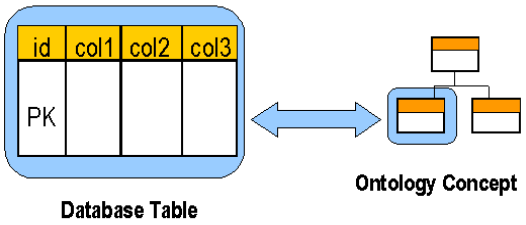
R_2O

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

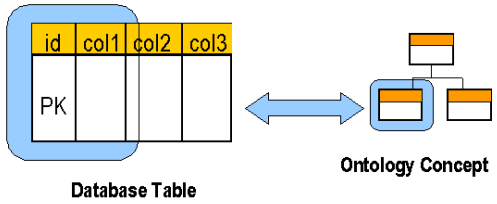


R2O (Relational-to-Ontology) Language

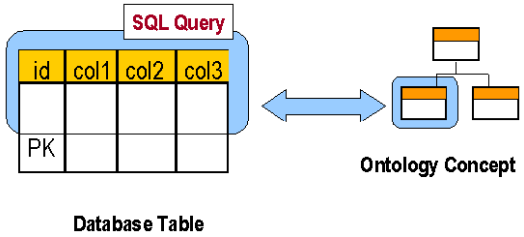
For concepts...



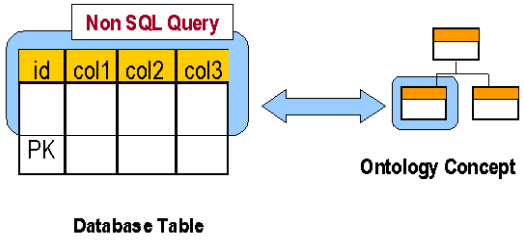
A view maps exactly one concept in the ontology.



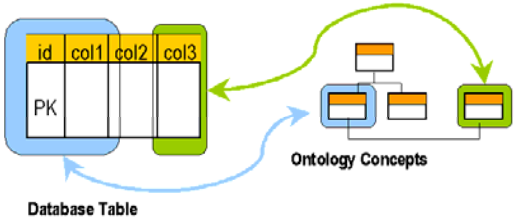
A subset of the columns in the view map a concept in the ontology.



A subset (selection) of the records of a database view map a concept in the ontology.

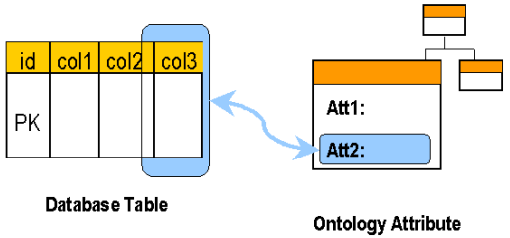


A subset of the records of a database view map a concept in the onto. but the selection cannot be made using SQL.

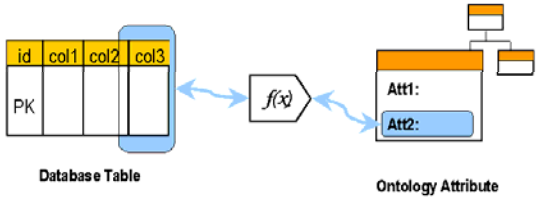


One or more concepts can be extracted from a single data field (not in 1NF).

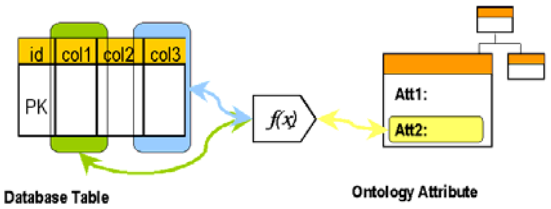
For attributes...



A column in a database view maps directly an attribute or a relation.

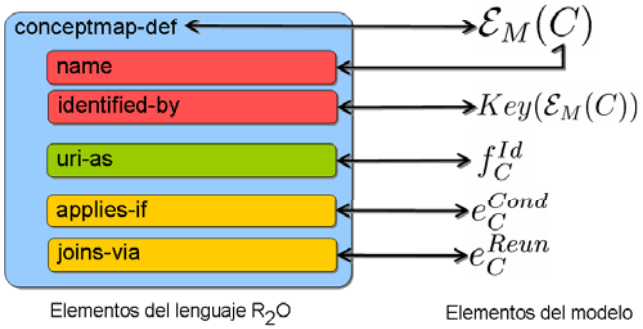


A column in a database view maps an attribute or a relation after some transformation.

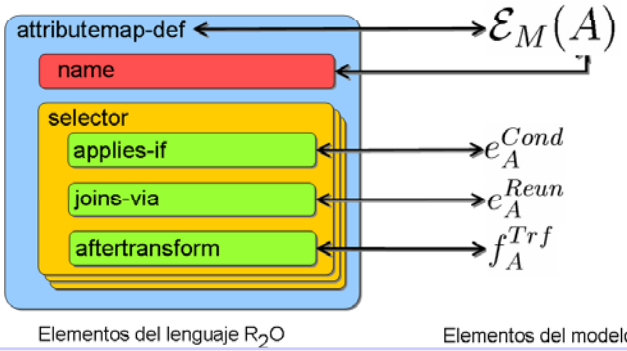


A set of columns in a database view map an attribute or a relation.

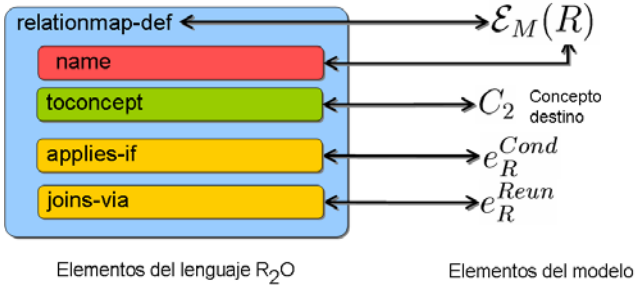
R_2O Basic Syntax




```
<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>
```



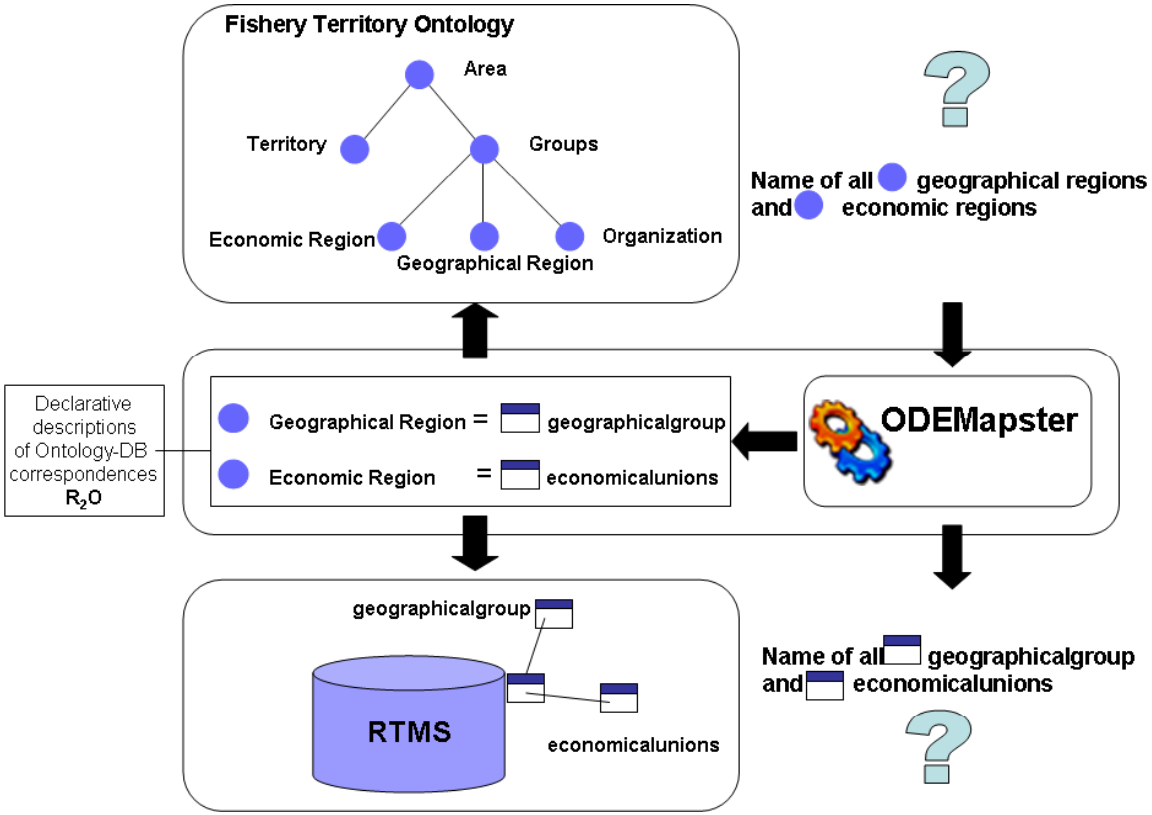

```
<attributemap-def name="http://esperonto/ff#Title">
<aftertransform>
<operation oper-id="constant">
<arg-restriction on-param="const-val">
<has-column>fsb_ajut.titol</has-column>
</arg-restriction>
</operation>
</aftertransform>
</attributemap-def>
```



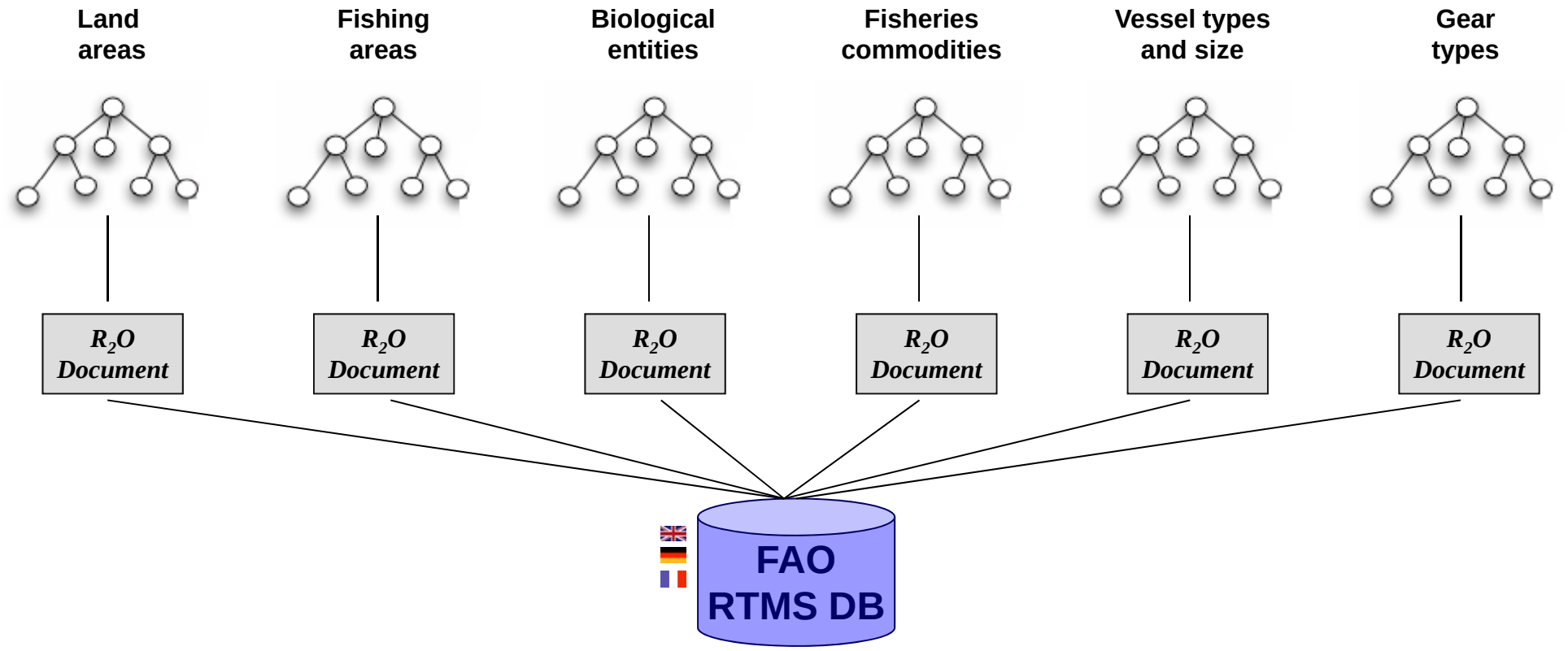

```
<relationmap-def name="http://esperonto/ff#isCandidateFor">
<to-concept name="http://esperonto/ff#FundOpp">
<joins-via>
<operation oper-id="equals">
<arg-restriction on-param="value1">
<has-column>fsb_ajut.id</has-column>
</arg-restriction>
<arg-restriction on-param="value2">
<has-column>fsb_candidate.forFund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

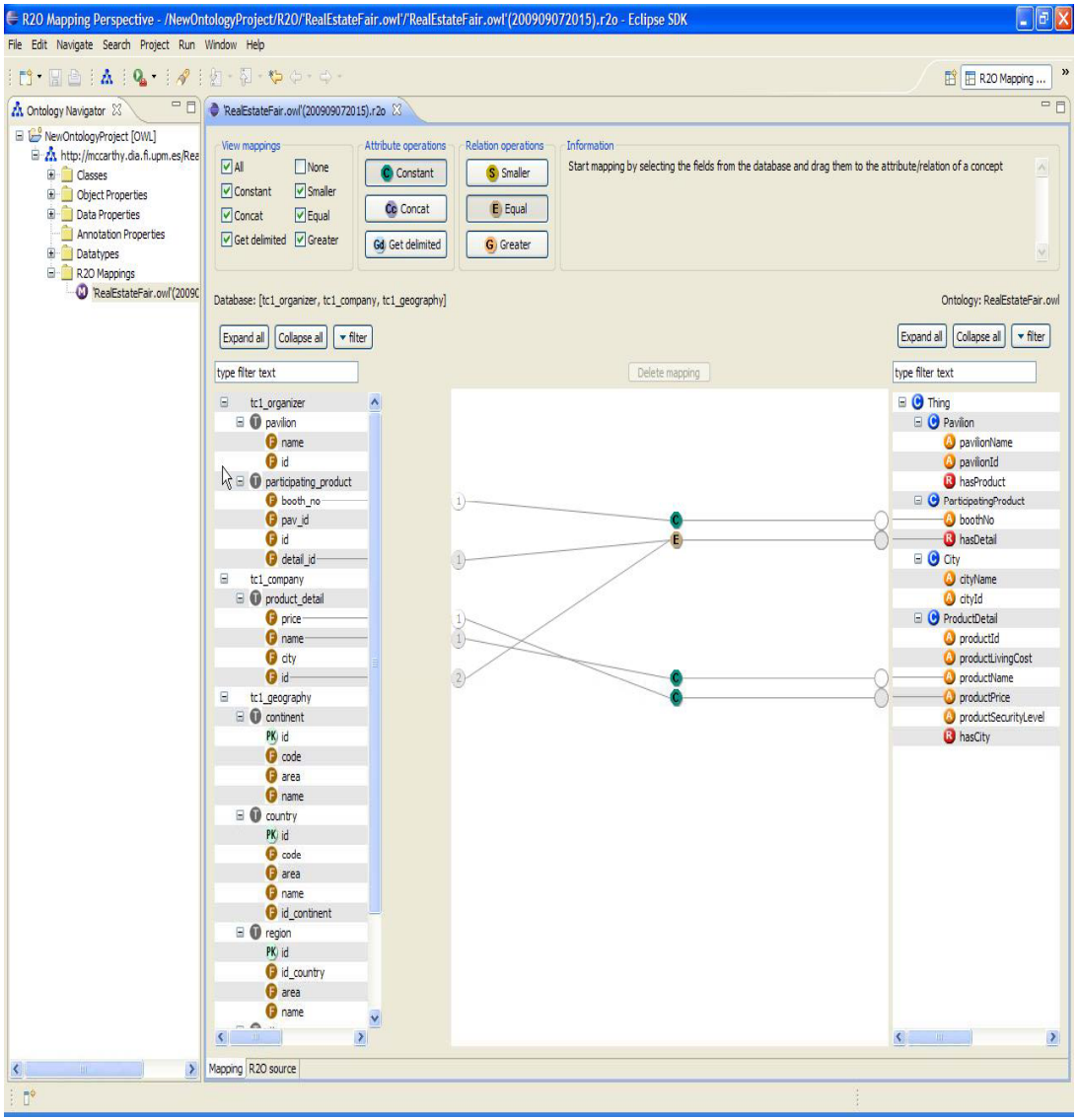
Fisheries commoditie	
Concepts	5
Properties	14
Instances	1380

Vessel types and size	
Concepts	5
Properties	20
Instances	120

Gear types	
Concepts	4
Properties	14
Instances	0

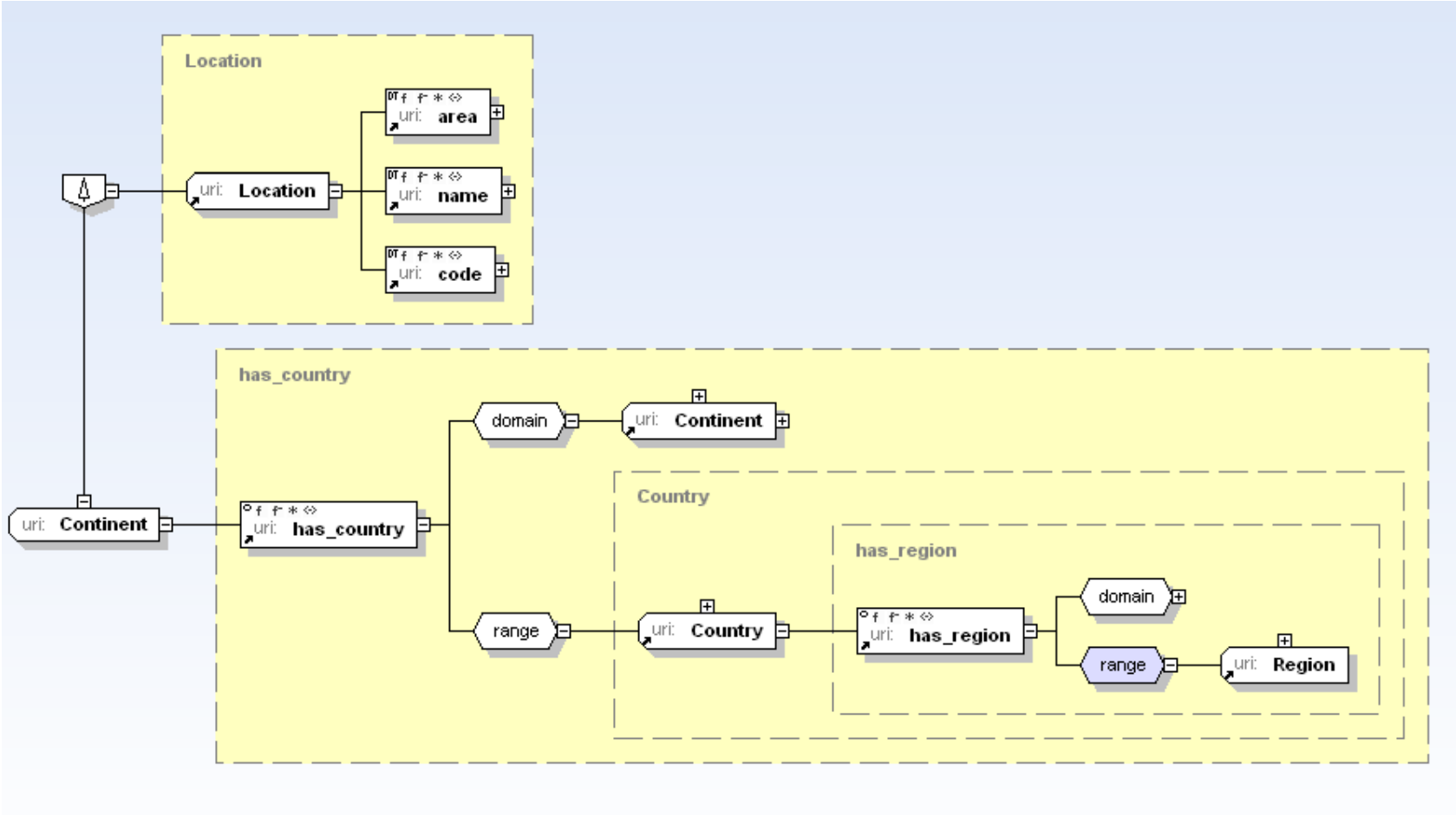
http://www.fao.org/aims/aos/fi/

Mapping Design

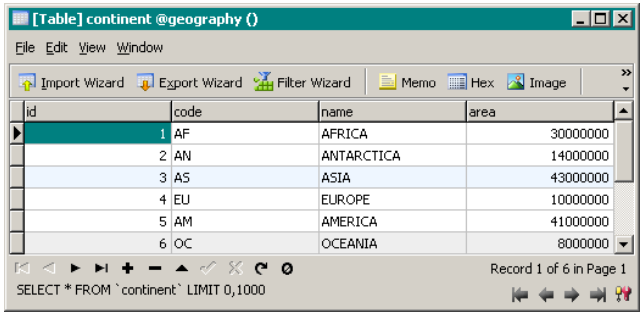


- **3 Mapping Creation Steps**
  - Load Ontology
  - Load Database(s)
  - Create mapping
- **2 Usage Modes**
  - Online mode (run time query execution)
  - Offline mode (materialized RDF dump)

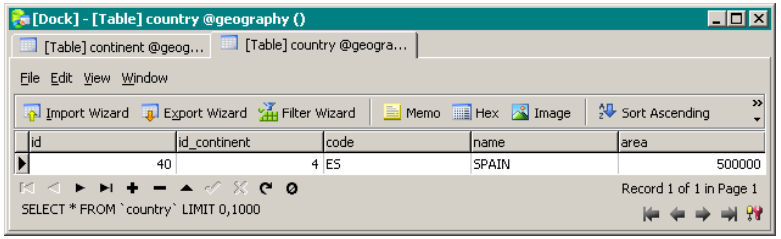
Example: Geography Ontology



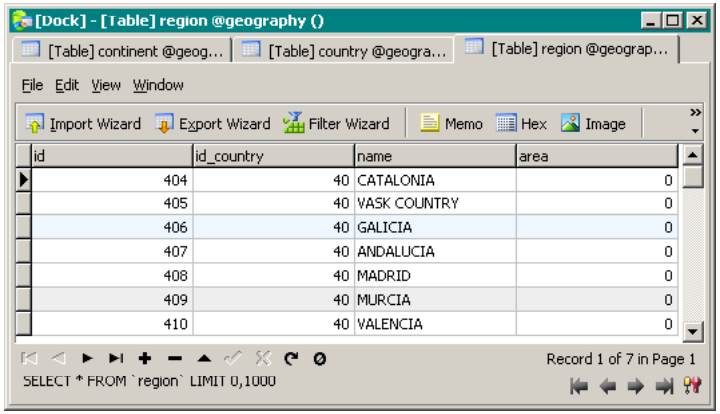
Geography Database



Continent Table

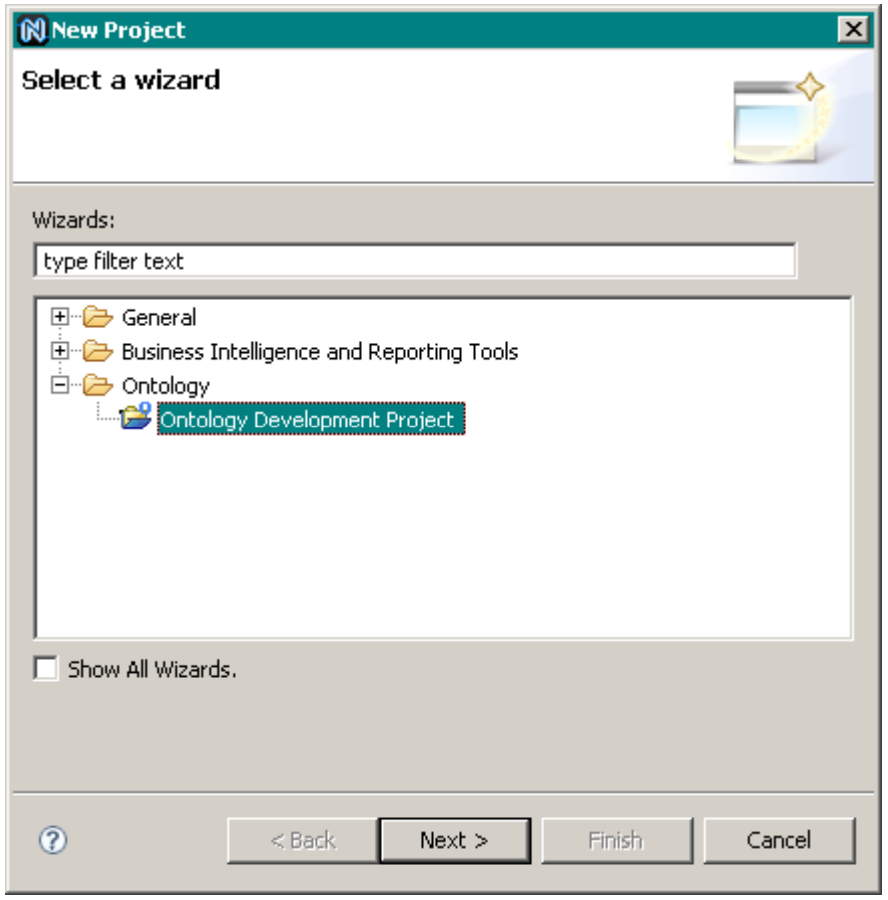


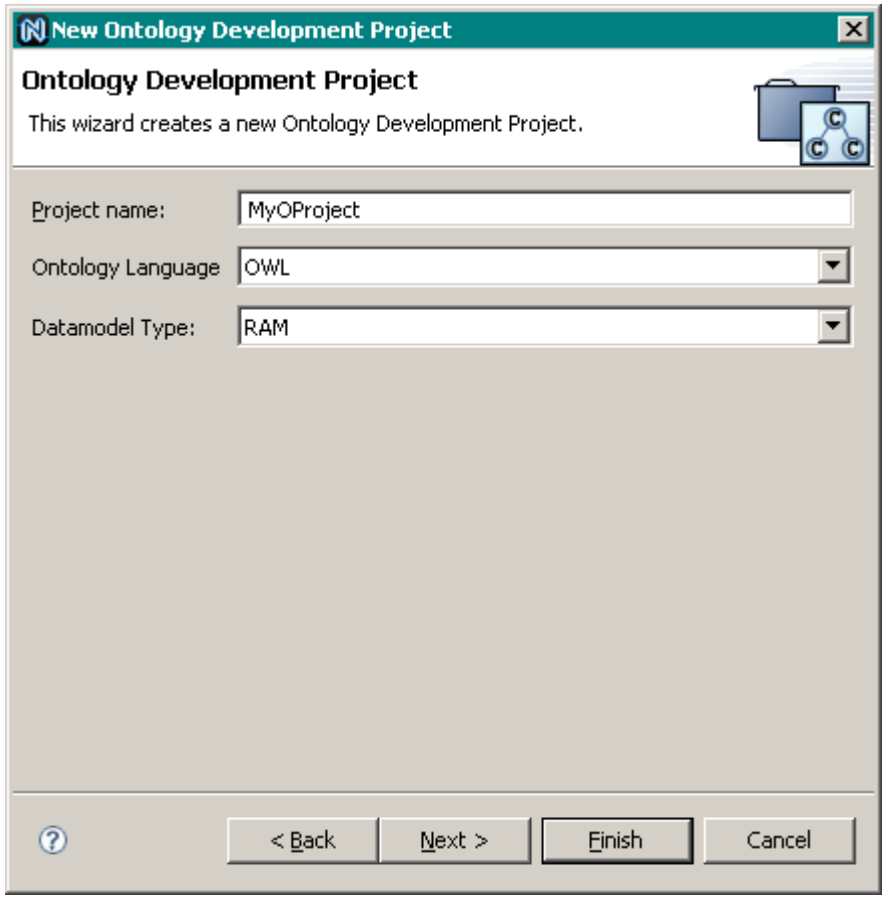
Country Table



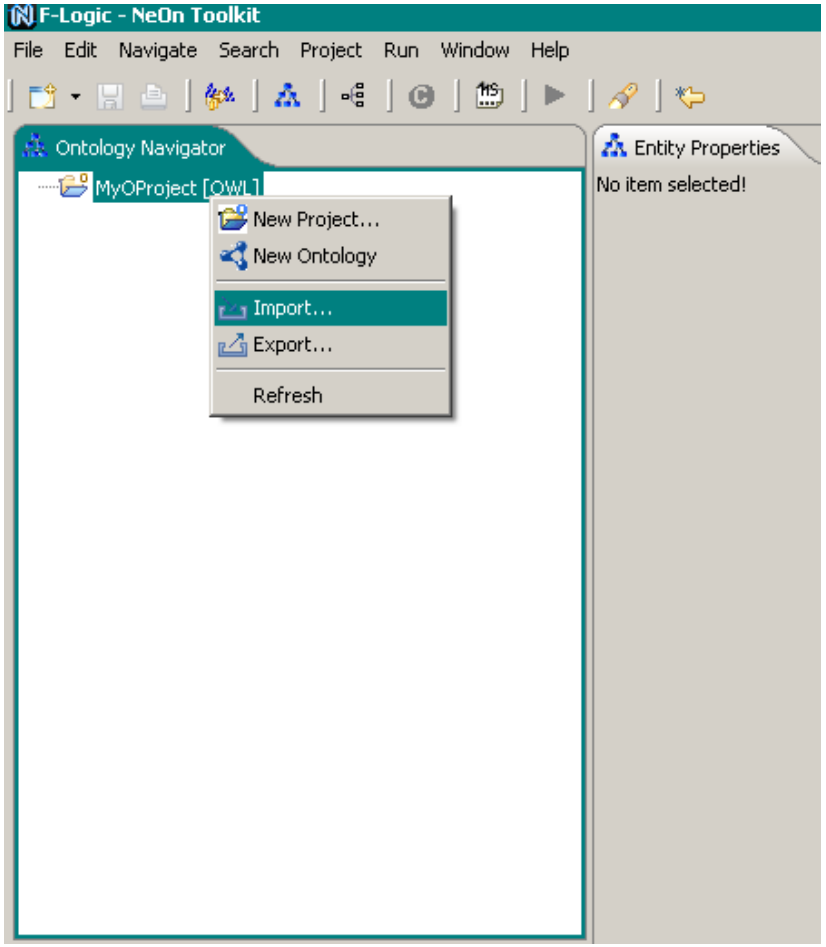
Region Table

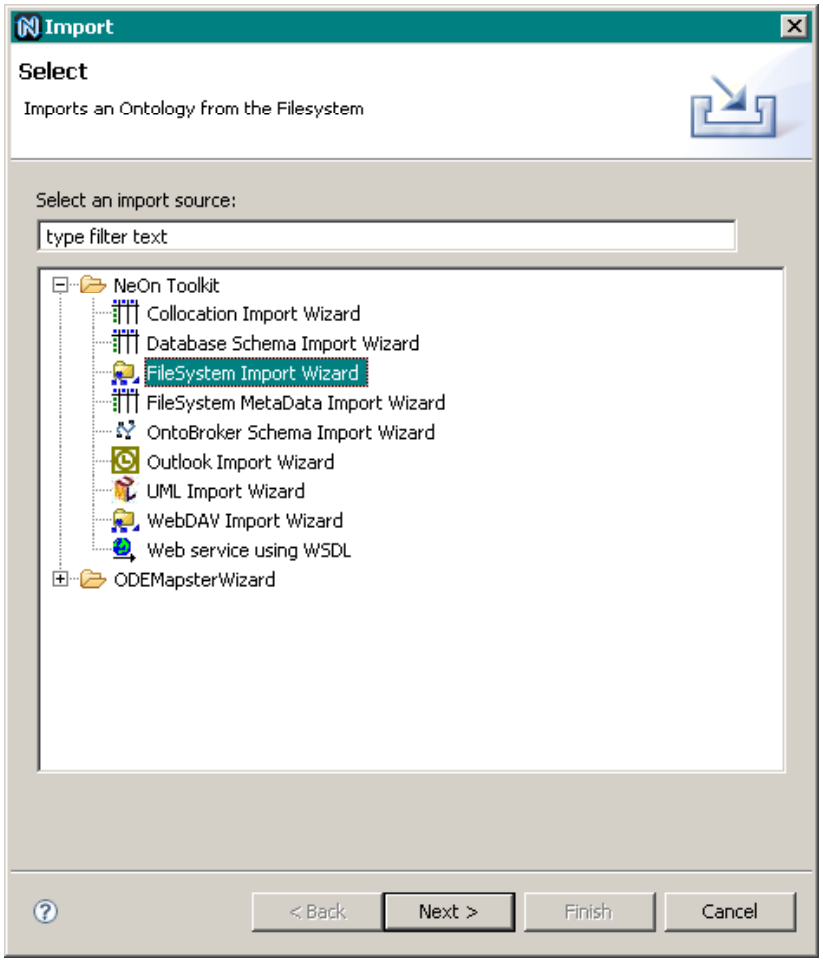
NeOn Toolkit – New Ontology Project



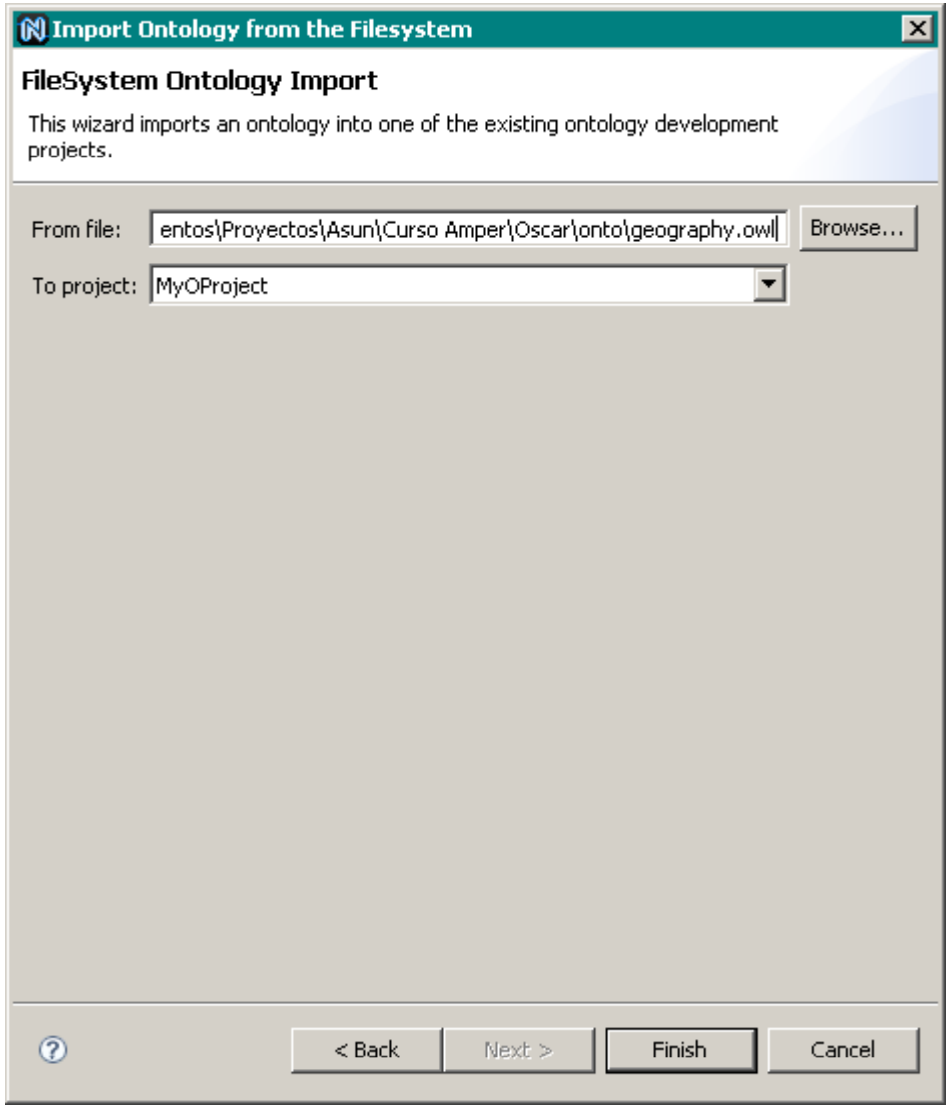


Import Ontology

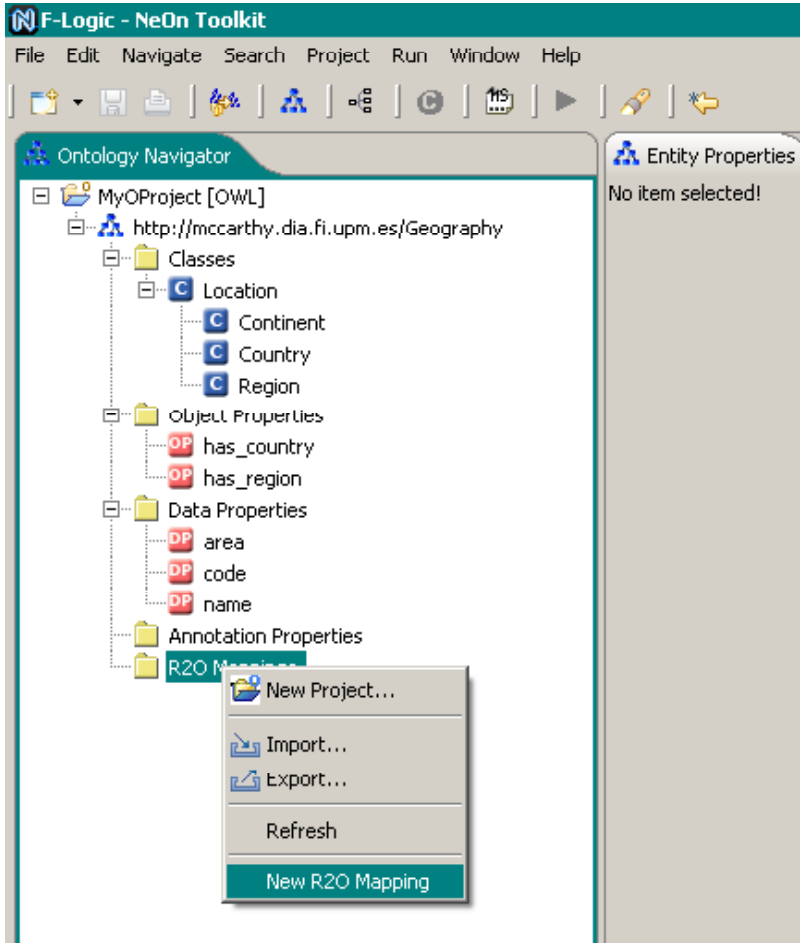


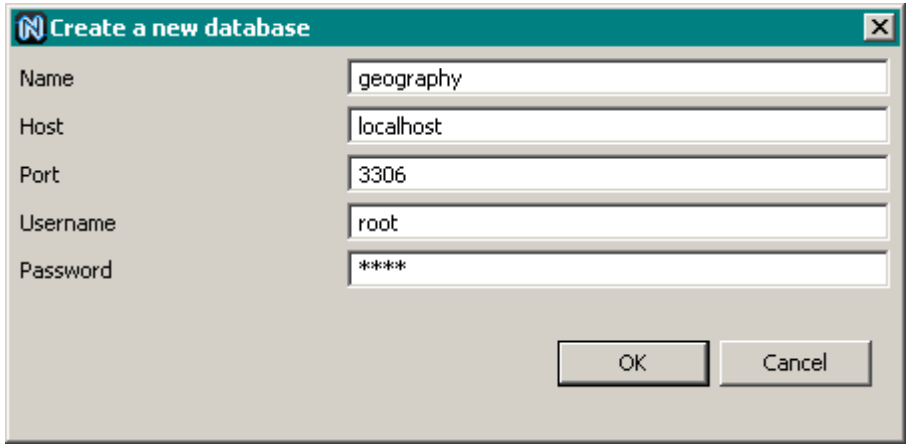


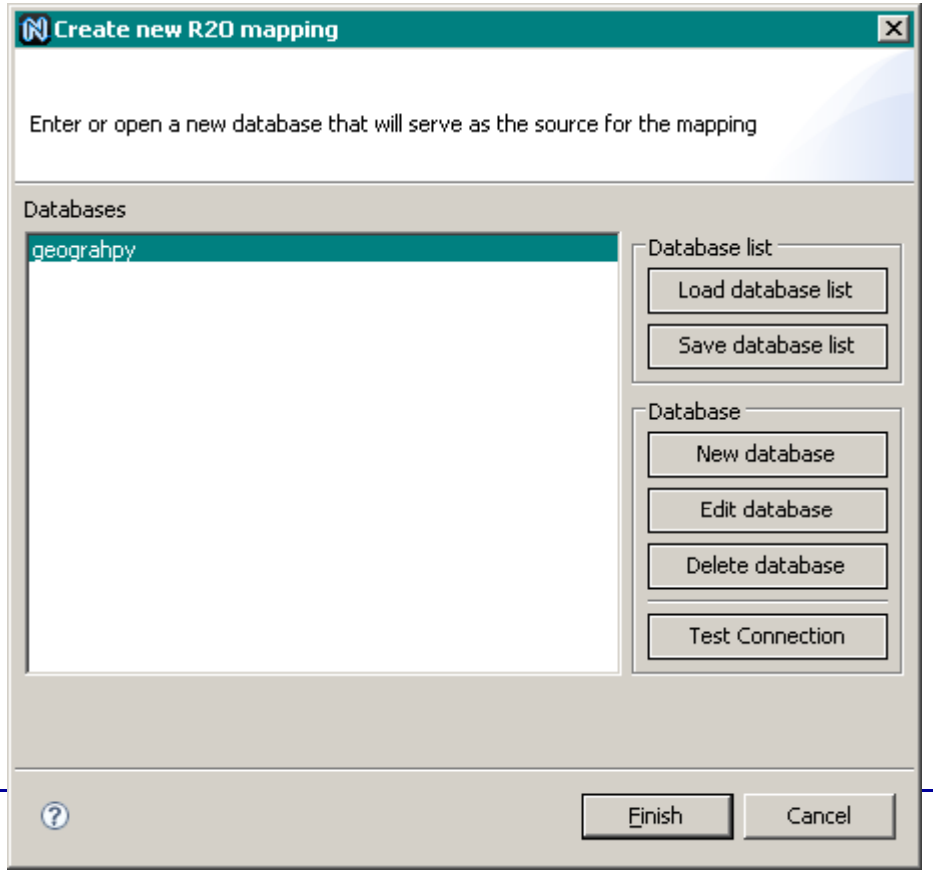
Import Ontology



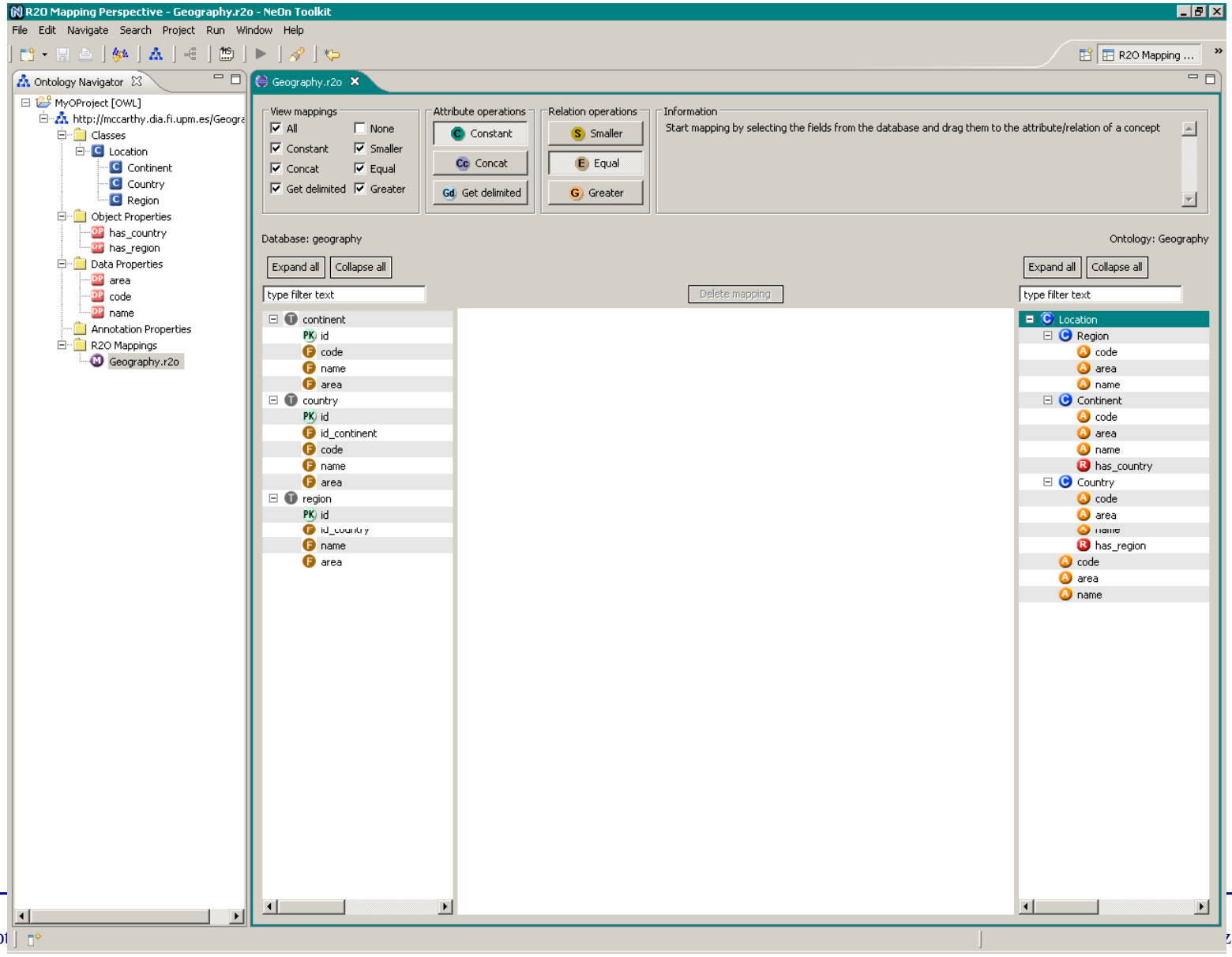
New R_2O Mapping – Database



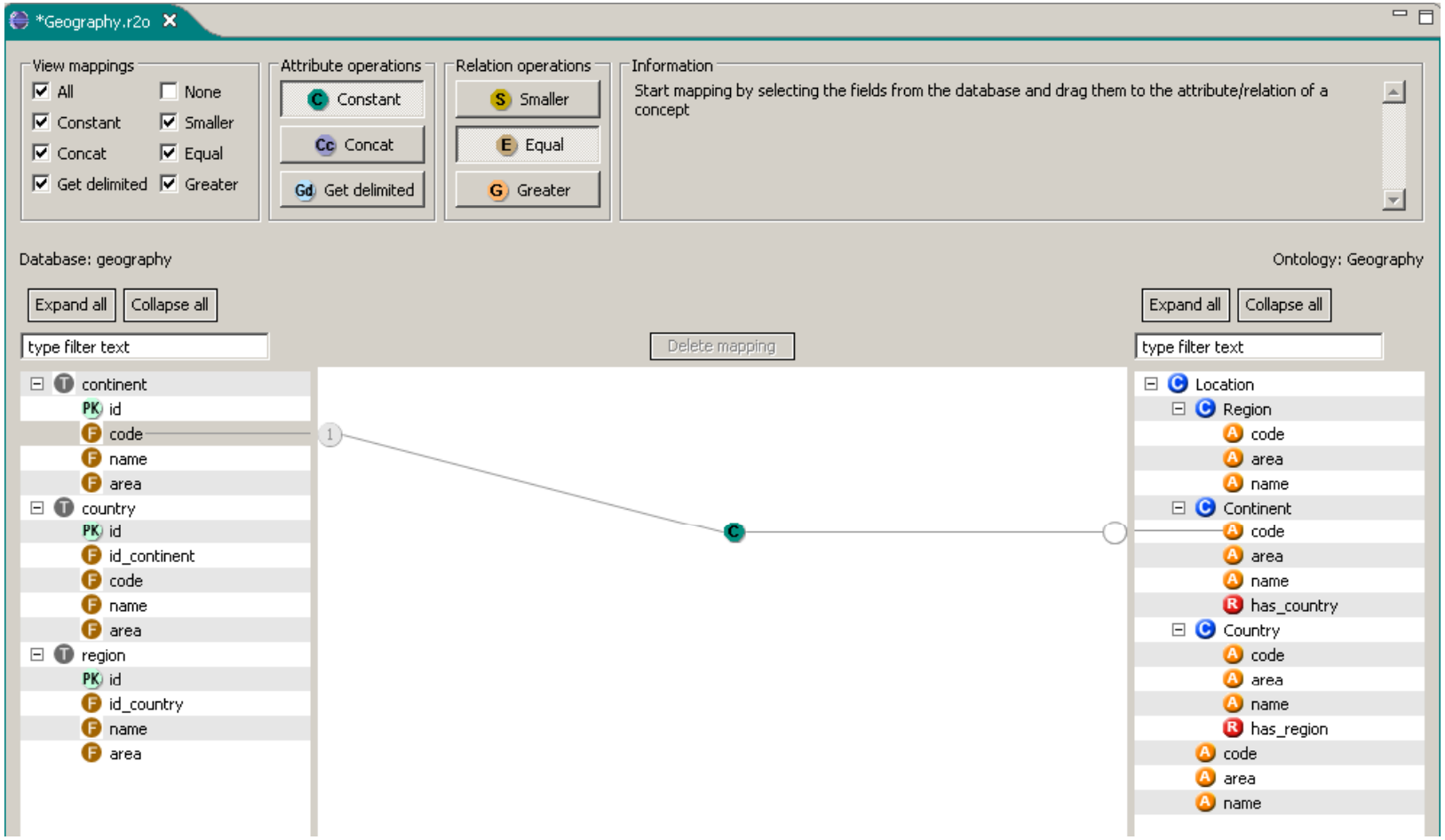




R_2O Mapping Perspective



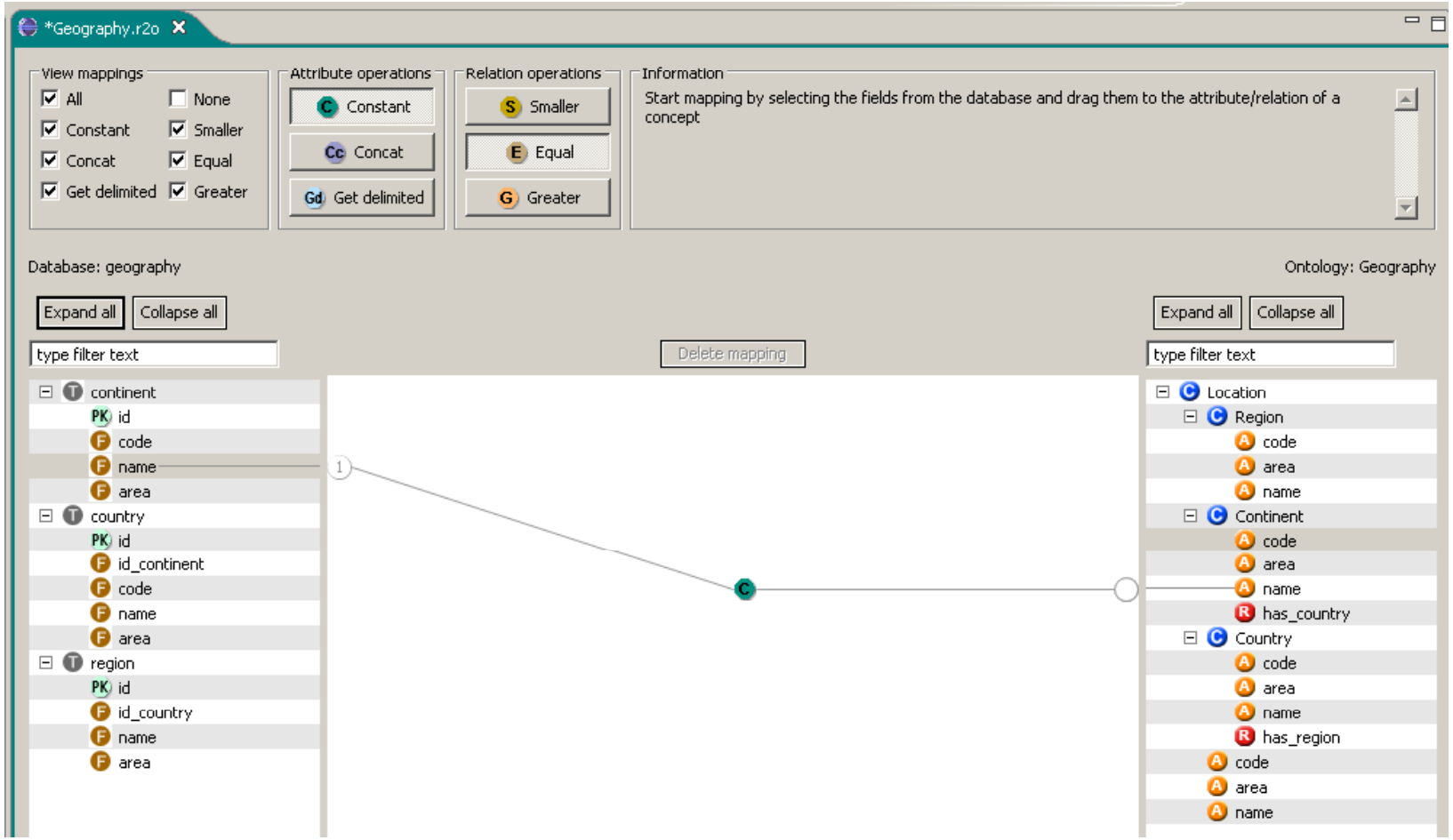
An attribute mapping example



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>
```

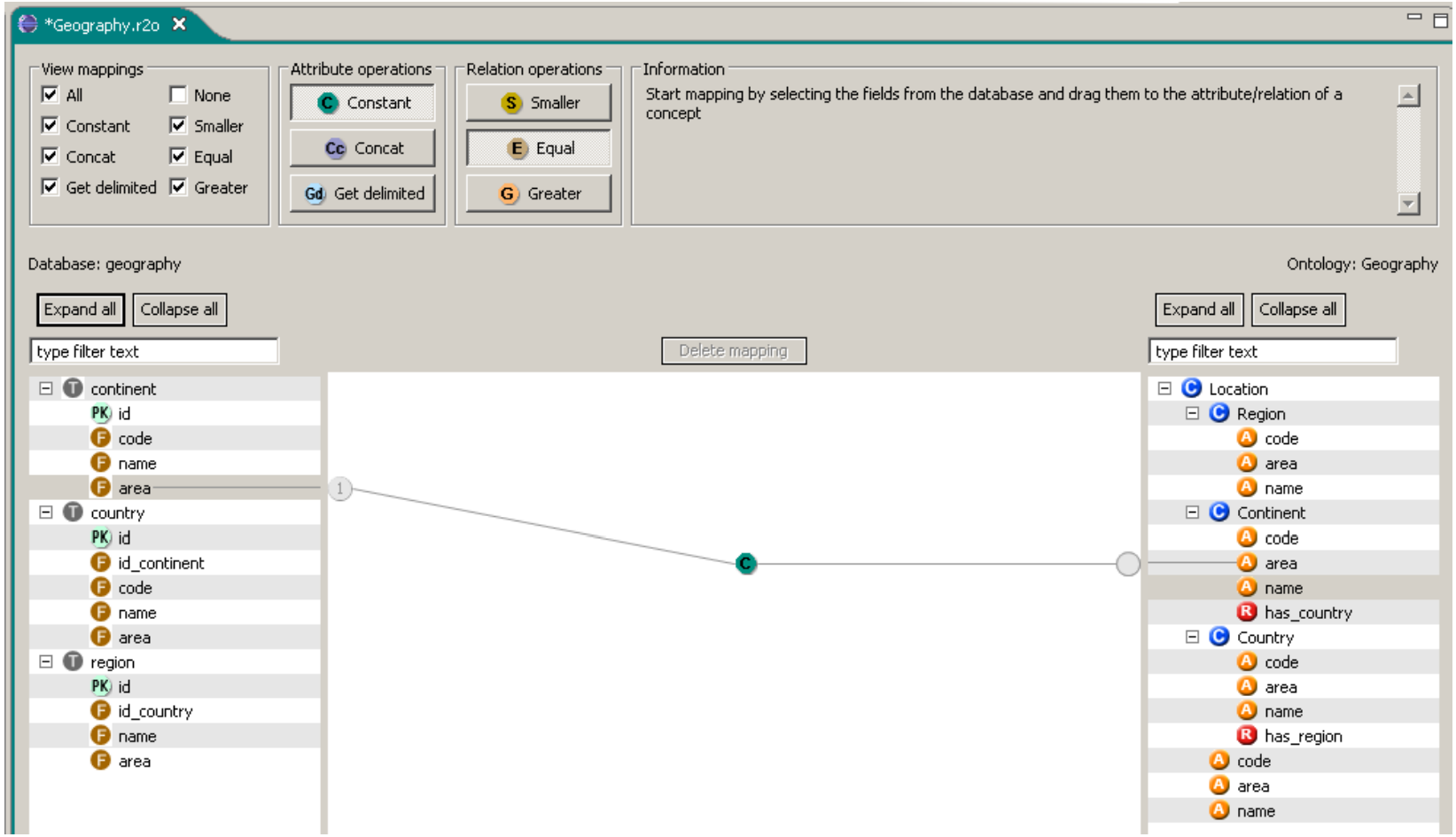
An attribute mapping example



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>
```

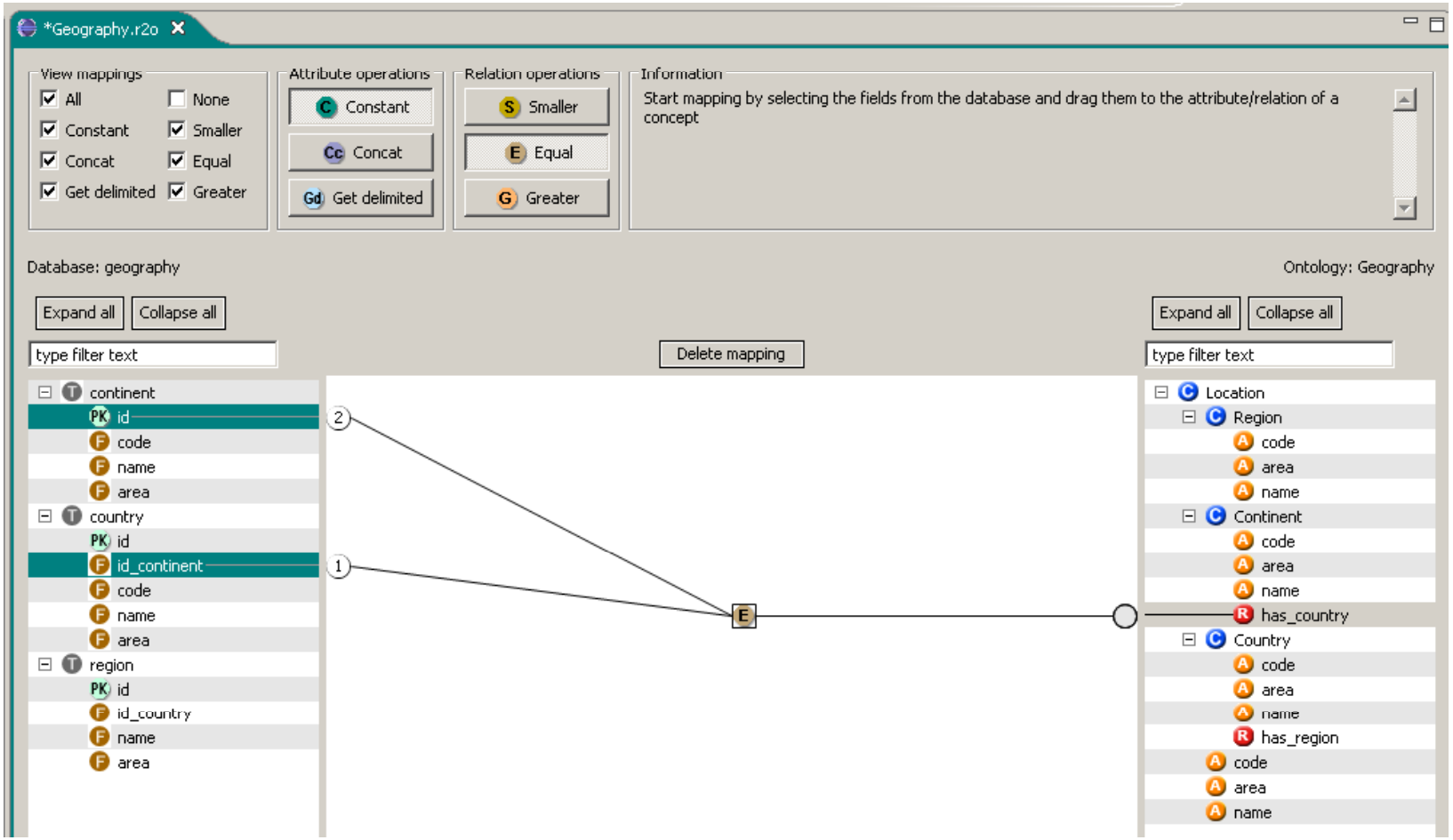
An attribute mapping example



An attribute mapping example – R_2O Code

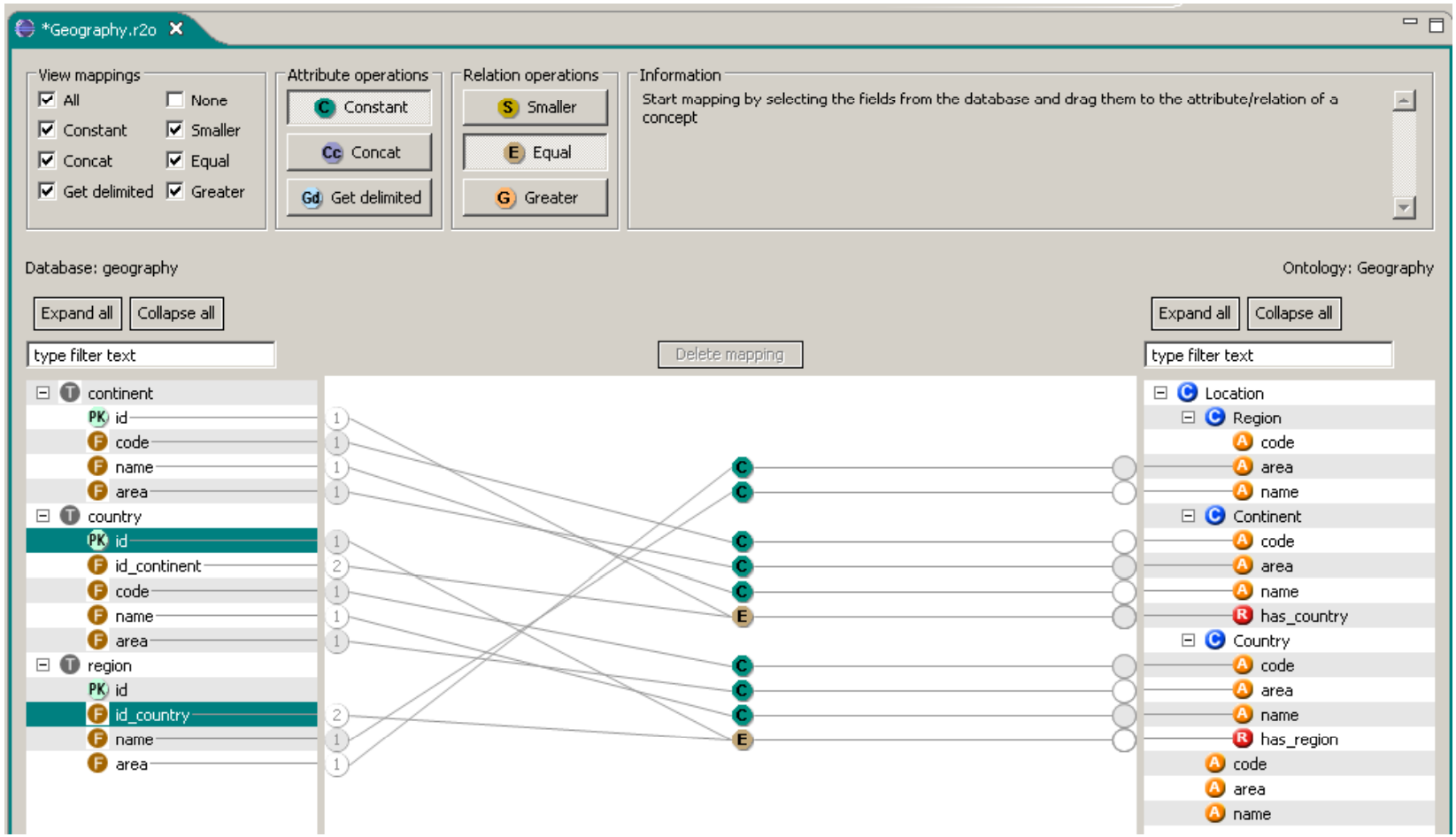
```
<attributemap-def name="http://mccarthy.dia.fi.upm.es/Geography#area">
  <selector>
    <aftertransform>
      <operation oper-id="constant">
        <arg-restriction on-param="const-val">
          <has-column>continent.area</has-column>
        </arg-restriction>
      </operation>
    </aftertransform>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

A relation mapping example



A relation mapping example – R_2O Code

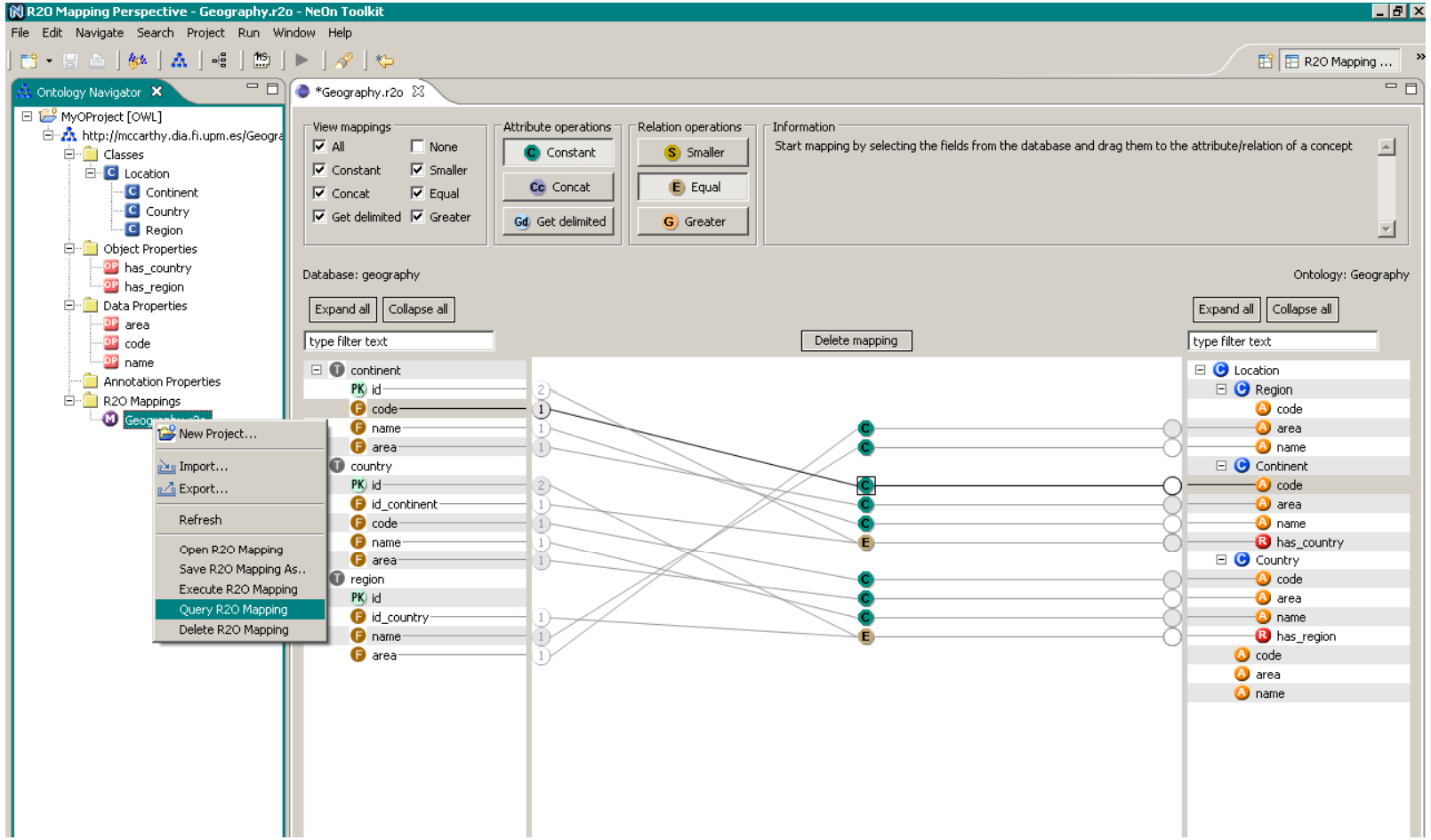
```
<dbrelationmap-def name="http://mccarthy.dia.fi.upm.es/Geography#has_country" toConcept="http://mccarthy.dia.fi.upm.es/Geography#Country">
  <joins-via>
    <condition oper-id="equals">
      <arg-restriction on-param="value1">
        <has-column>continent.id</has-column>
      </arg-restriction>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



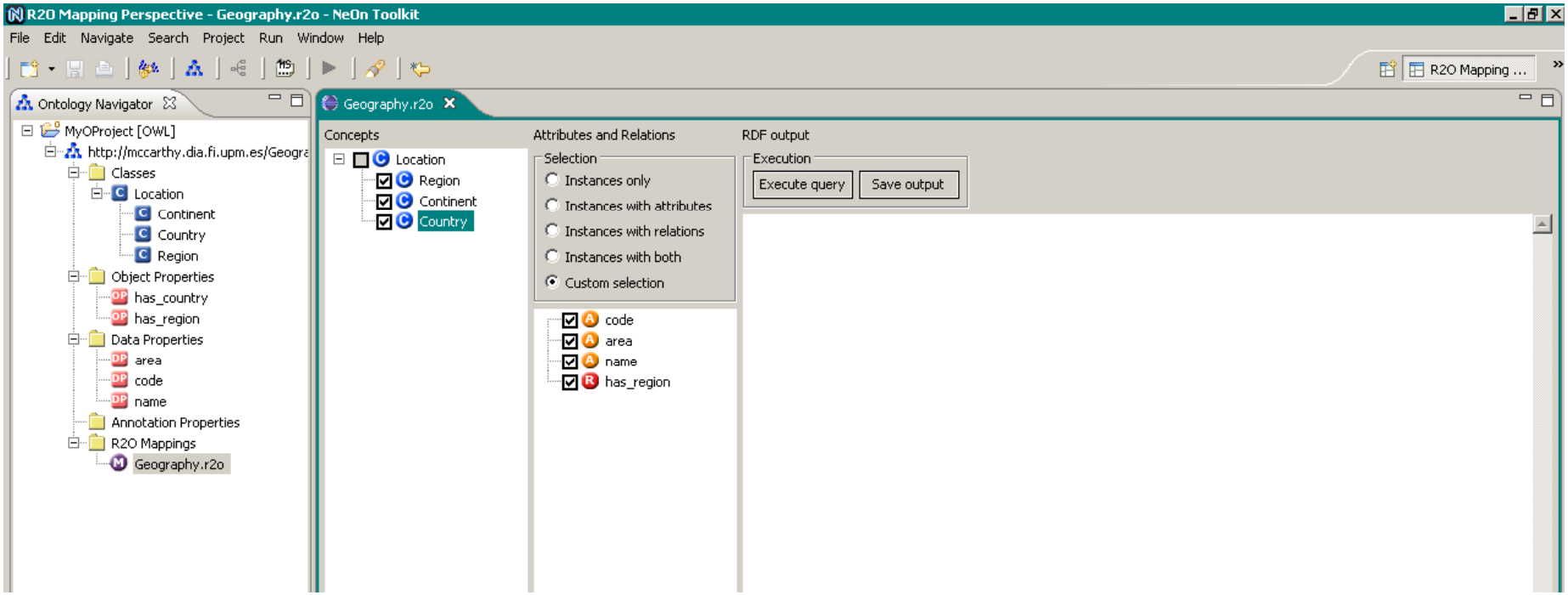
Concept mapping example – uri-as

```
<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



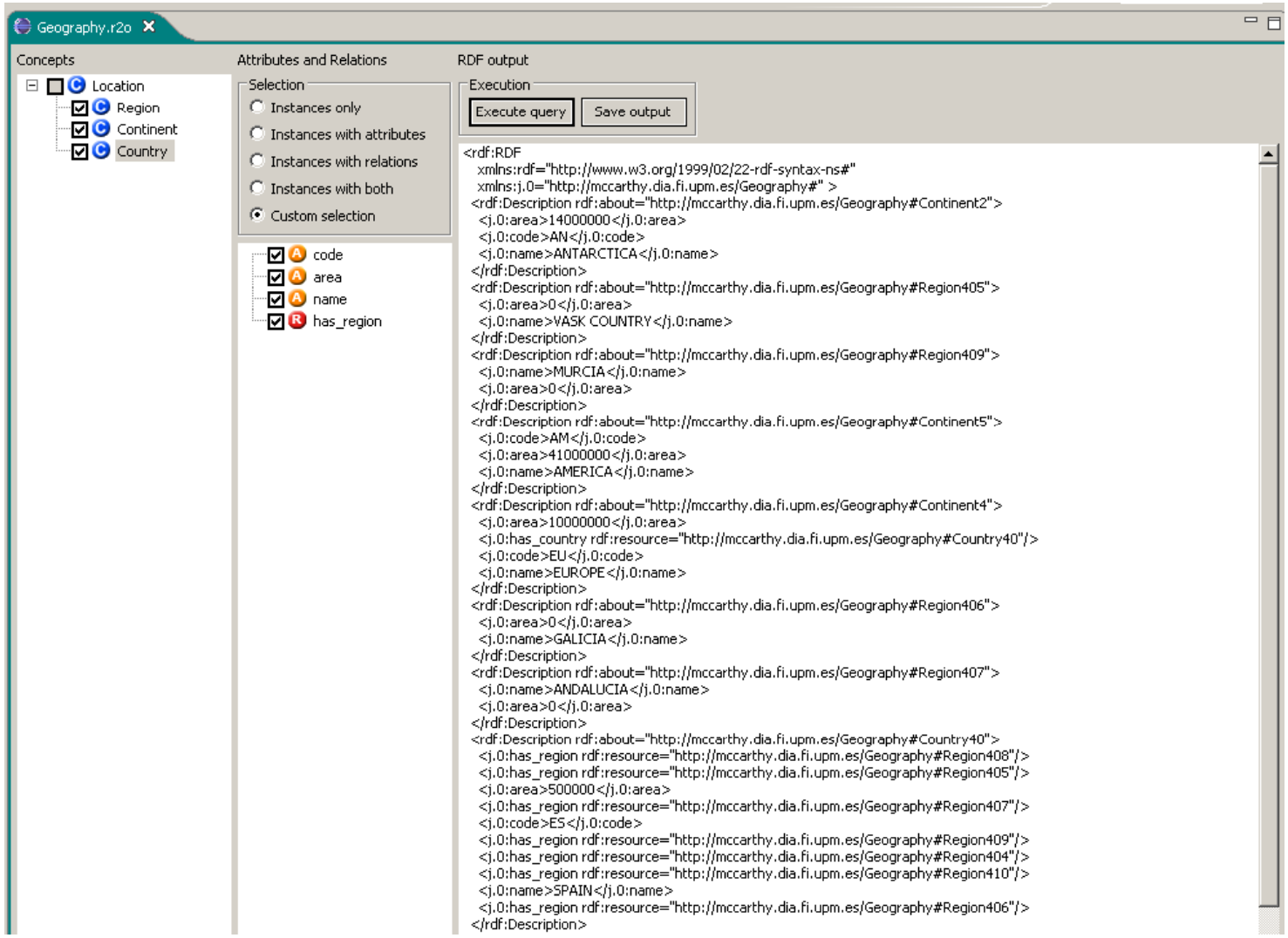
Querying the Ontology Instances



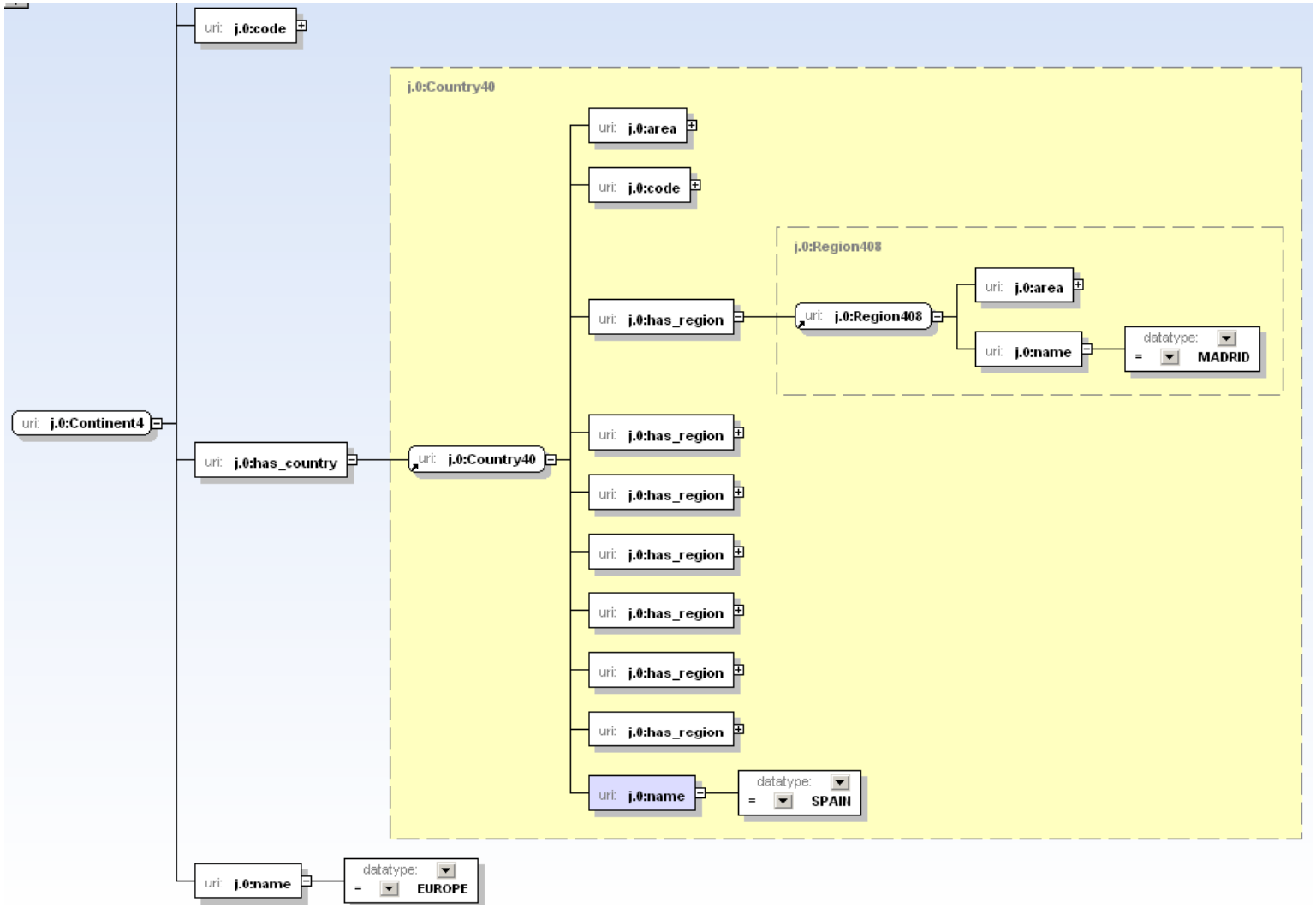
Querying the Ontology Instances – ODQML code

```
<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



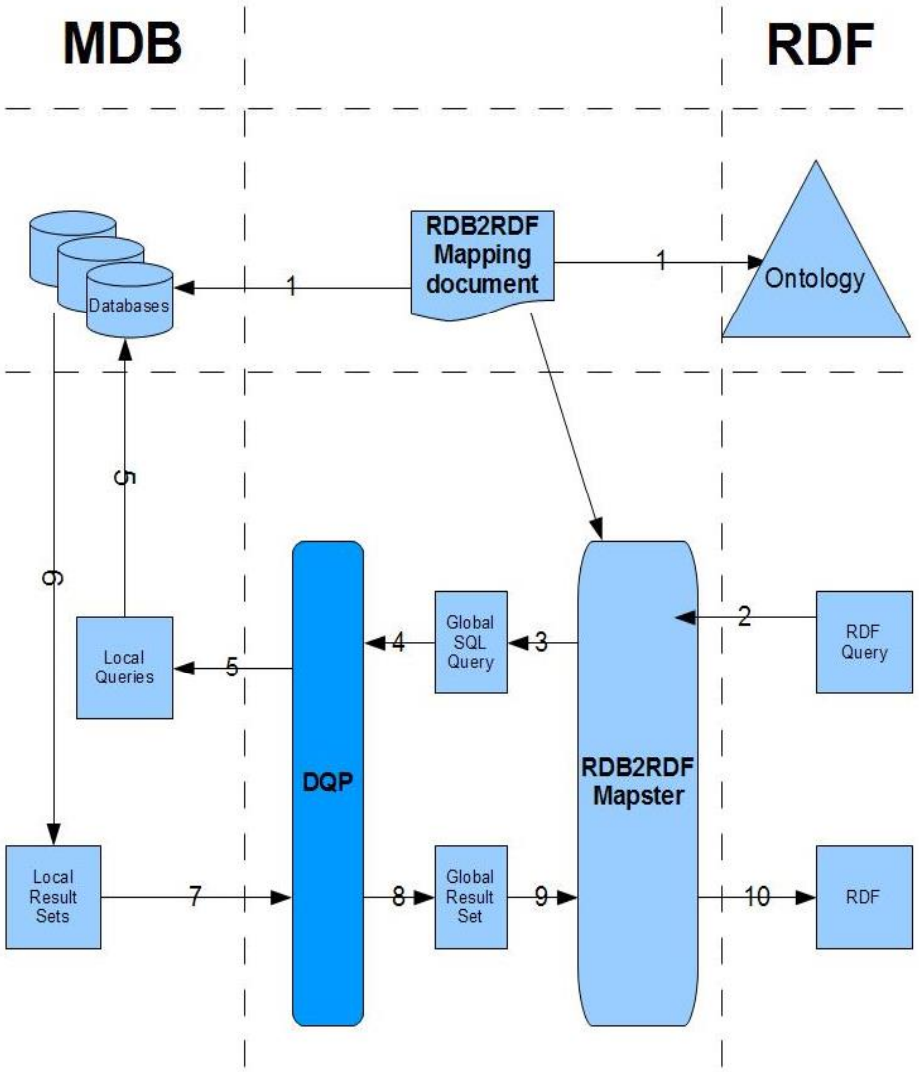
How to embed the ODEMapster Processor

```
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

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**