



Publishing Linked Data from RDB

Boris Villazón-Terrazas. Oscar Corcho

Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

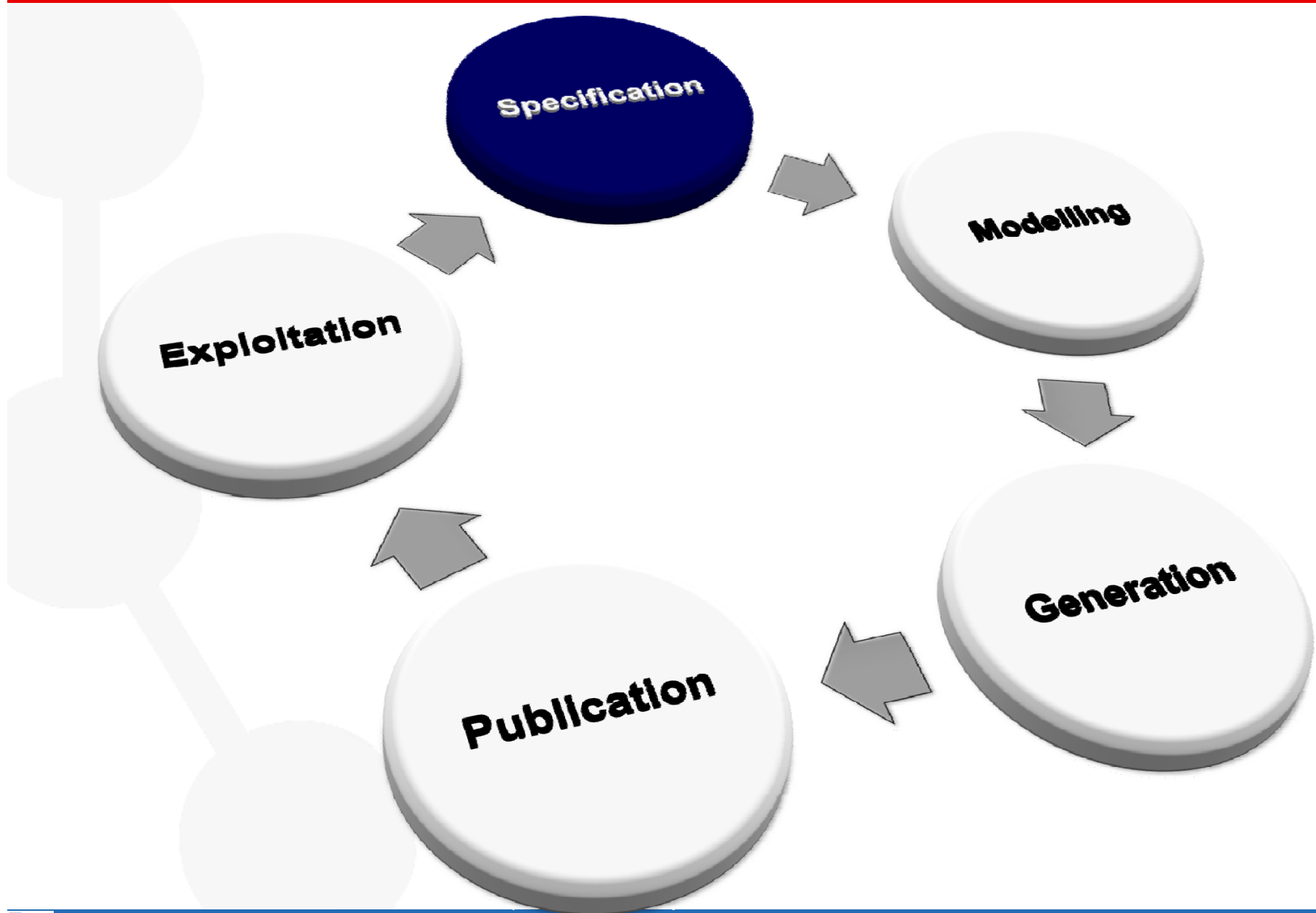
<http://www.oeg-upm.net>

bvillazon@fi.upm.es

Phone: 34.91.3366605, Fax: 34.91.3524819

Slides available at: <http://www.slideshare.net/boricles/>

Acknowledgements: Freddy Priyatna, Jan Schulte, Richard Cyganiak and many others that we may have omitted.



Specification – RDB about Research Groups



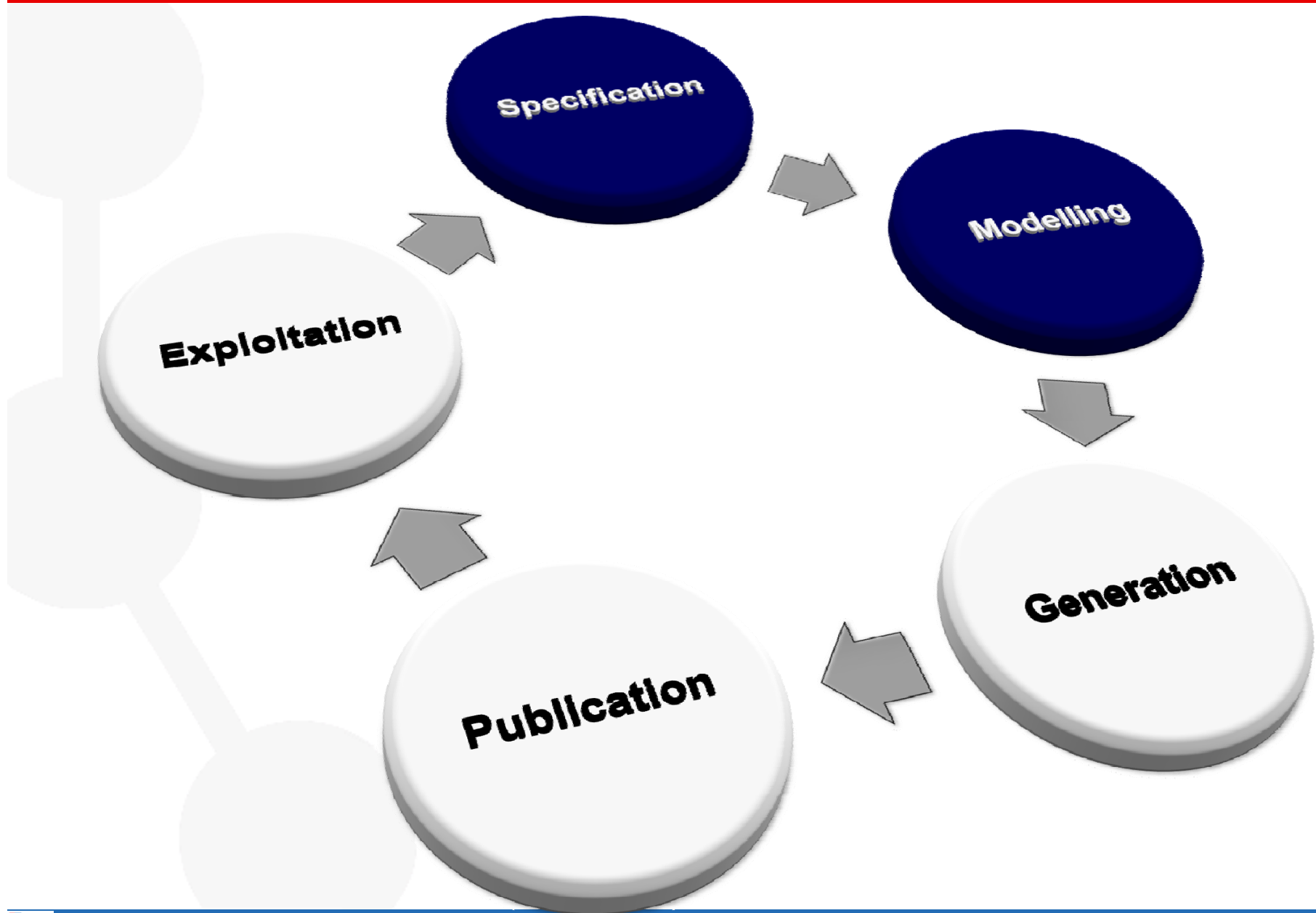
ID_INVESTIGADOR	ID_ARTICULO	TITULO	ANIO	REVISTA	ISSN
5387	58807	Automatic generation for transshipment pl	2008	TRANSPORTATION PLANNING AND T	0308-1060
5387	79013	Design of a competence-based model for	2010	Project Perspectives 2010	ISSN 1455-4178
5387	84277	Women Weaving Network Society. Wome	2010	Congreso de Ingeniería 2010. Bueno	0395-2037
5553	4421	Multilingüedad en una aplicación basada	2007	Revista de la Sociedad Española para	1135-5948
5553	64219	Terminología, Ontologías y Multilingüedad	2009	Revista Puntosycoma	1830-5415

ID_GRUPO	ESTADO_GRUPO	NOMBRE	ACRONIMO	TLF	FAX	WEB	EMAIL
323	Consolidado	Ontology Engineering Group (LIA).	OEG	91 336 74 39	91 352 48 19	www.oeg-upm.net	gi.giolia@upm.es
381	Consolidado	Organizaciones Sostenibles	(Null)	91 336 32 08	91 336 32 07	http://gios.etsii.upm.es/	gi.gios@upm.es

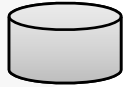
ID_GRUPO	ID_INVESTIGADOR	NOMBRE	APELLIDO1	APELLIDO2	SEXO	CODIGO_NACIONALIDAD
323	11137	CARLOS	BUIL	ARANDA	V	042
323	11266	OSCAR	MUÑOZ	GARCIA	V	000
323	11267	VICTOR	SAQUICELA	GALARZA	V	000
323	11268	ADRIAN	SILES	RODRIGUEZ	V	000
323	11269	JOSE	MORA	LÓPEZ	V	042
323	11270	MARIA	POVEDA	VILLALON	M	000
323	11378	RAUL	ALCAZAR	VALDARACETE	V	000
323	11454	MIGUEL ANGEL	GARCIA	DELGADO	V	000
323	11455	FRANCISCO	MARTIN-RECUERDA	MOYANO	V	000

CODIGO_NACIONALIDAD2	CODIGO_NACIONALIDAD	DENOMINACION_NACIONALIDAD	CODIGO_OFICIAL
000	000	Sin nacionalidad	00
ABW	474	ARUBA	AW
AFG	660	AFGANISTAN	AF
AG	(Null)	ALTOS DEL GOLAN	NULL
AGO	330	ANGOLA	AO

- Base URI
 - `http://research.ull.es`
- TBOX URI
 - `http://research.ull.es/ontology/{class|property}`
- ABOX URI
 - `http://research.ull.es/resource/{resourceType}/{resource}`



Modelling

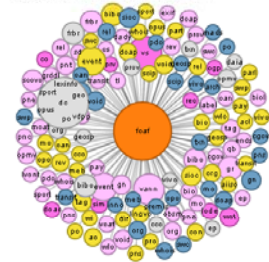


Linked Open Vocabularies

<http://labs.mondeca.com/dataset/lov/>

Vocabulary links:

Vocabularies referencing "foaf" (137)



Vocabularies referenced by "foaf" (8)



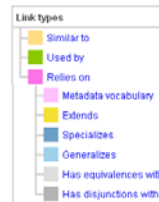
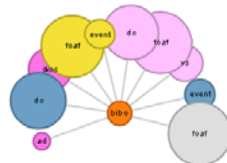
FOAF Friend of a Friend vocabulary

Vocabulary links:

Vocabularies referencing "bibo" (12)



Vocabularies referenced by "bibo" (10)



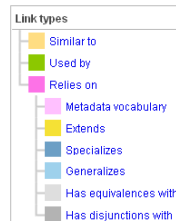
BIBO The Bibliographic Ontology

Vocabulary links:

Vocabularies referencing "geop" (4)



Vocabularies referenced by "geop" (1)

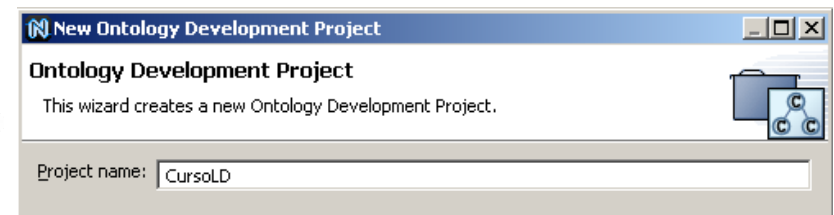
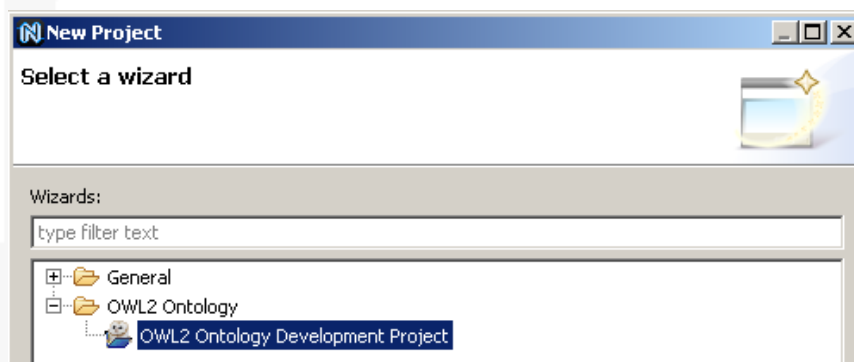


GEOP FAO Geopolitical Ontology

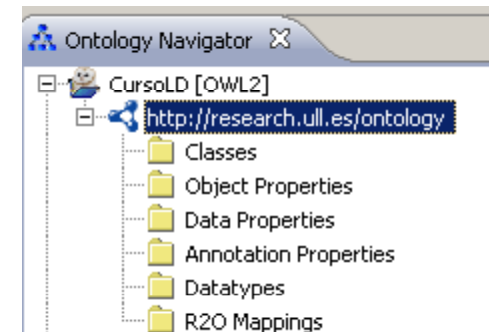
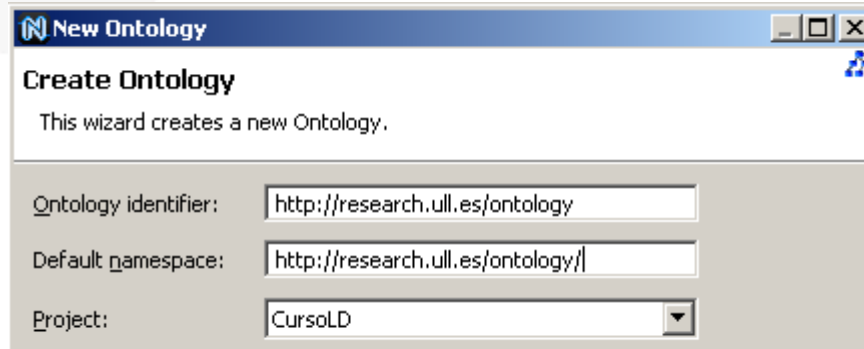


<http://neon-toolkit.org/>

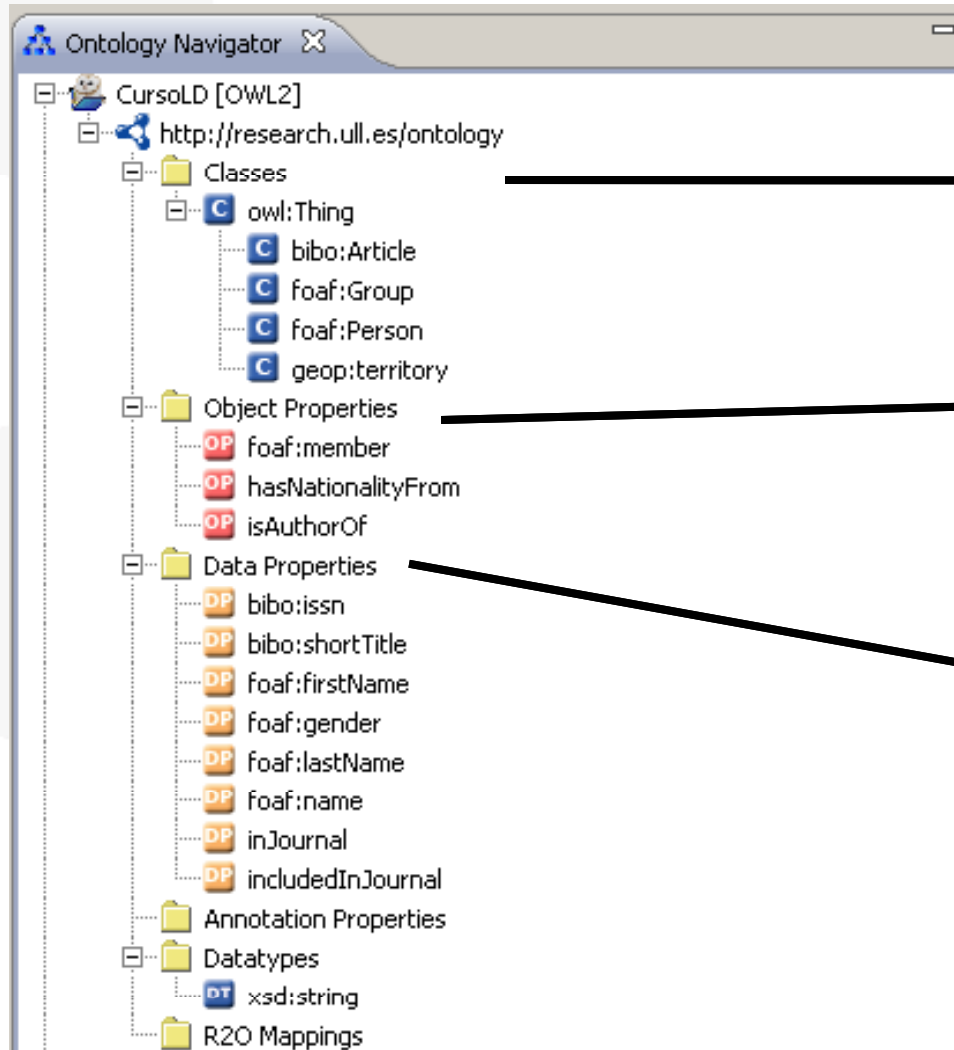
New Project



New Ontology



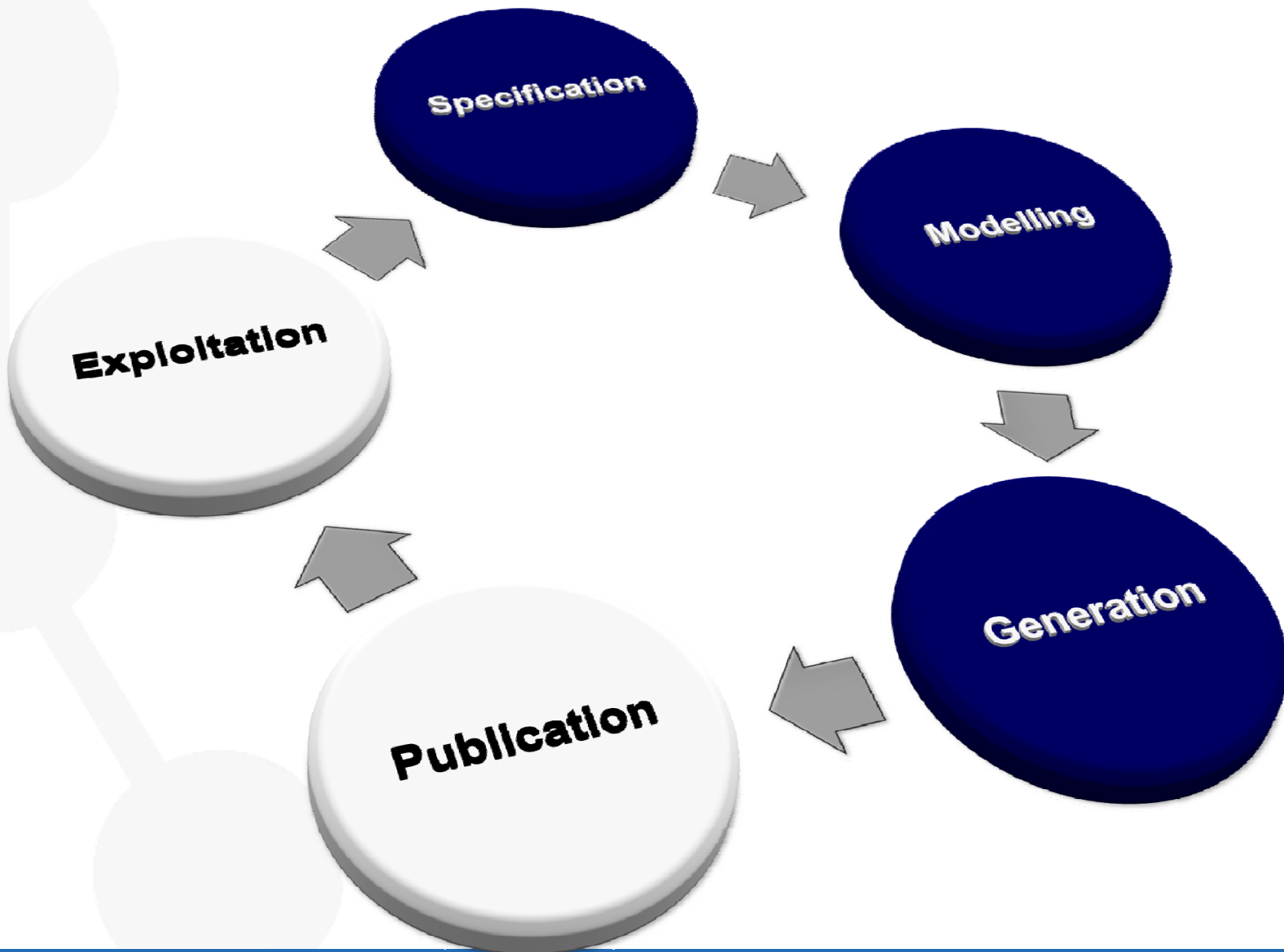
Modelling – Creating some elements



Classes

Object Properties

Datatype Properties



Transformation – RDB2RDF

- A majority of dynamic Web content is backed by relational databases (RDB), and so are many enterprise systems.



- W3C RDB2RDF Working Group ¹
 - R2RML: RDB to RDF Mapping Language - <http://www.w3.org/TR/r2rml/>
 - Direct Mapping - <http://www.w3.org/TR/rdb-direct-mapping/>
 - R2RML and Direct Mapping Test Cases - <http://www.w3.org/2001/sw/rdb2rdf/test-cases/>

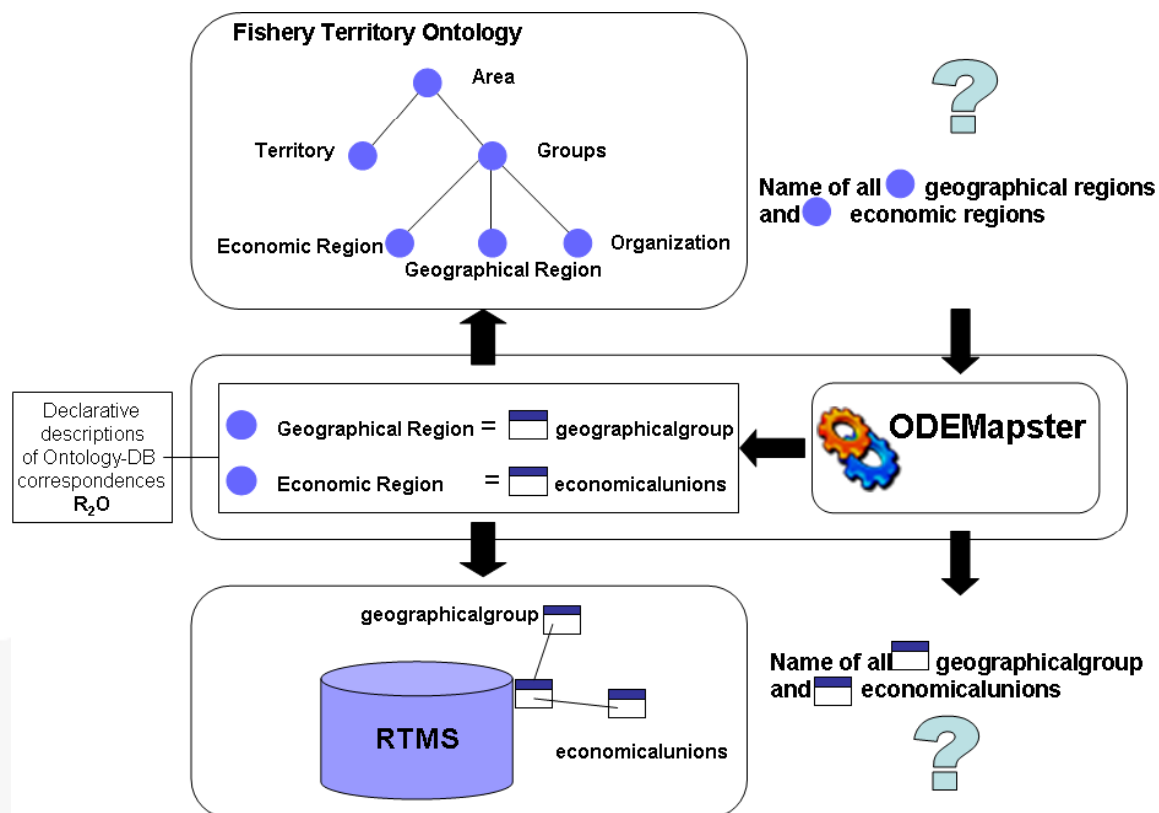
R2RML/Direct Mapping in process to be a W3C Recommendation



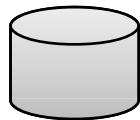
¹ <http://www.w3.org/TR/r2rml/>

Transformation – R2O & ODEMapster

- R₂O is an extensible, fully declarative language to describe mappings between relational database schemas and ontologies.
- The ODEMapster processor generates RDF instances from relational instances based on the mapping description expressed in the R₂O document



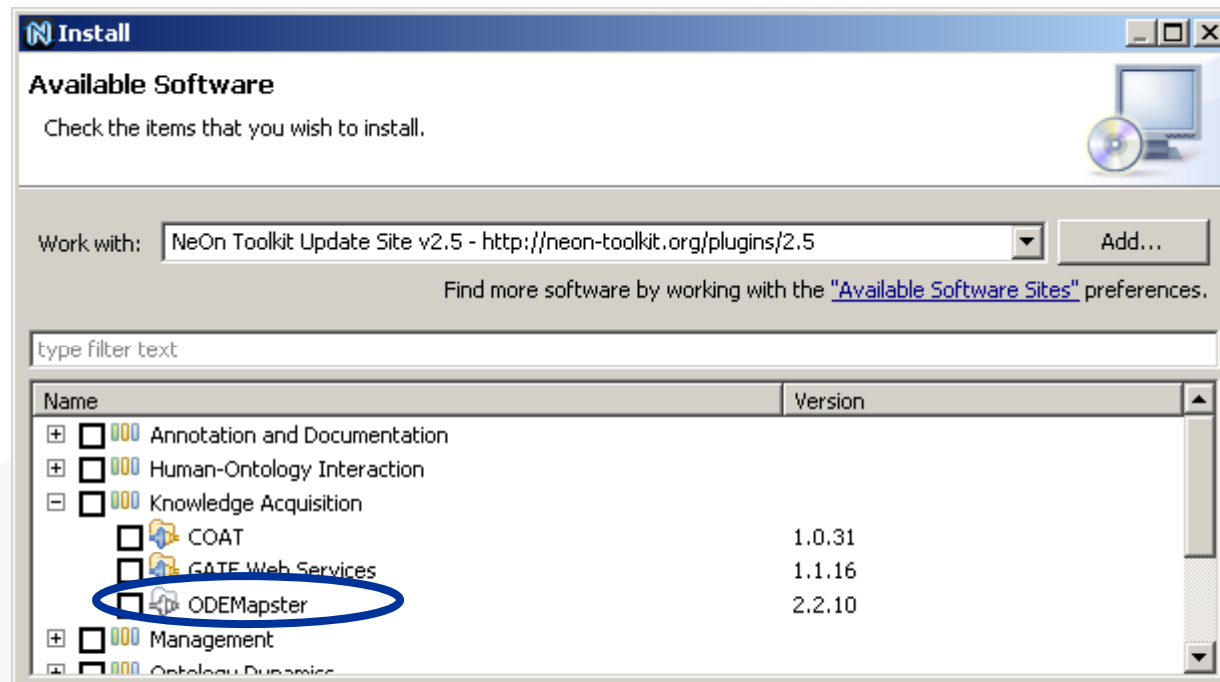
Generation – Transformation - ODEMapster



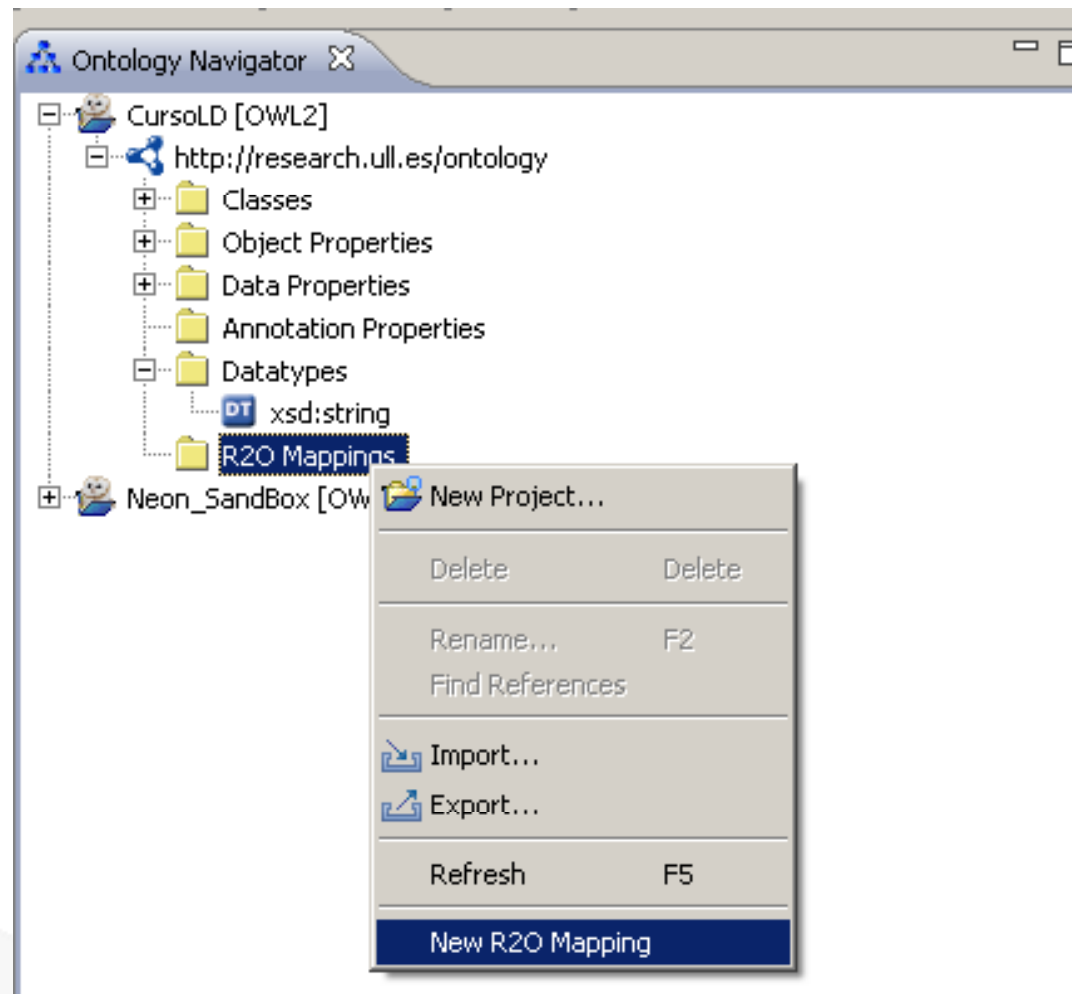
Research
RDB



- Included in the NTK – **currently under revision**



Transformation - ODEMapster



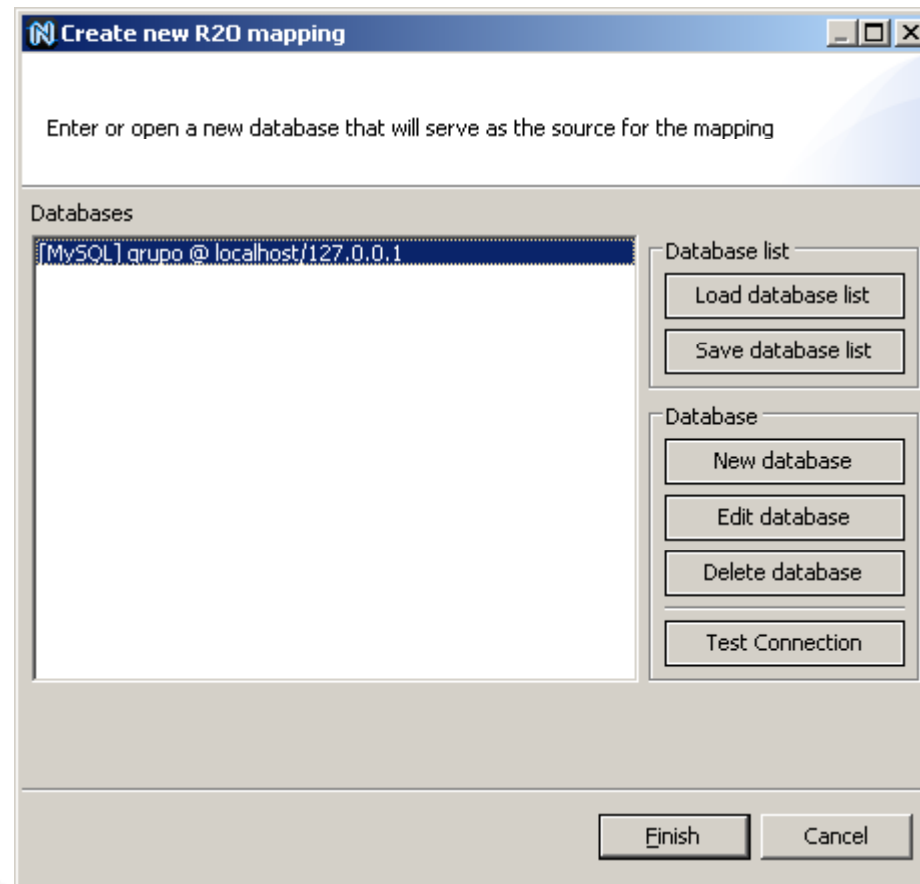
New R2O Mapping – DB connection information

The screenshot shows a 'Create new R2O mapping' dialog box with a 'Create a new database' sub-dialog. The sub-dialog contains the following fields:

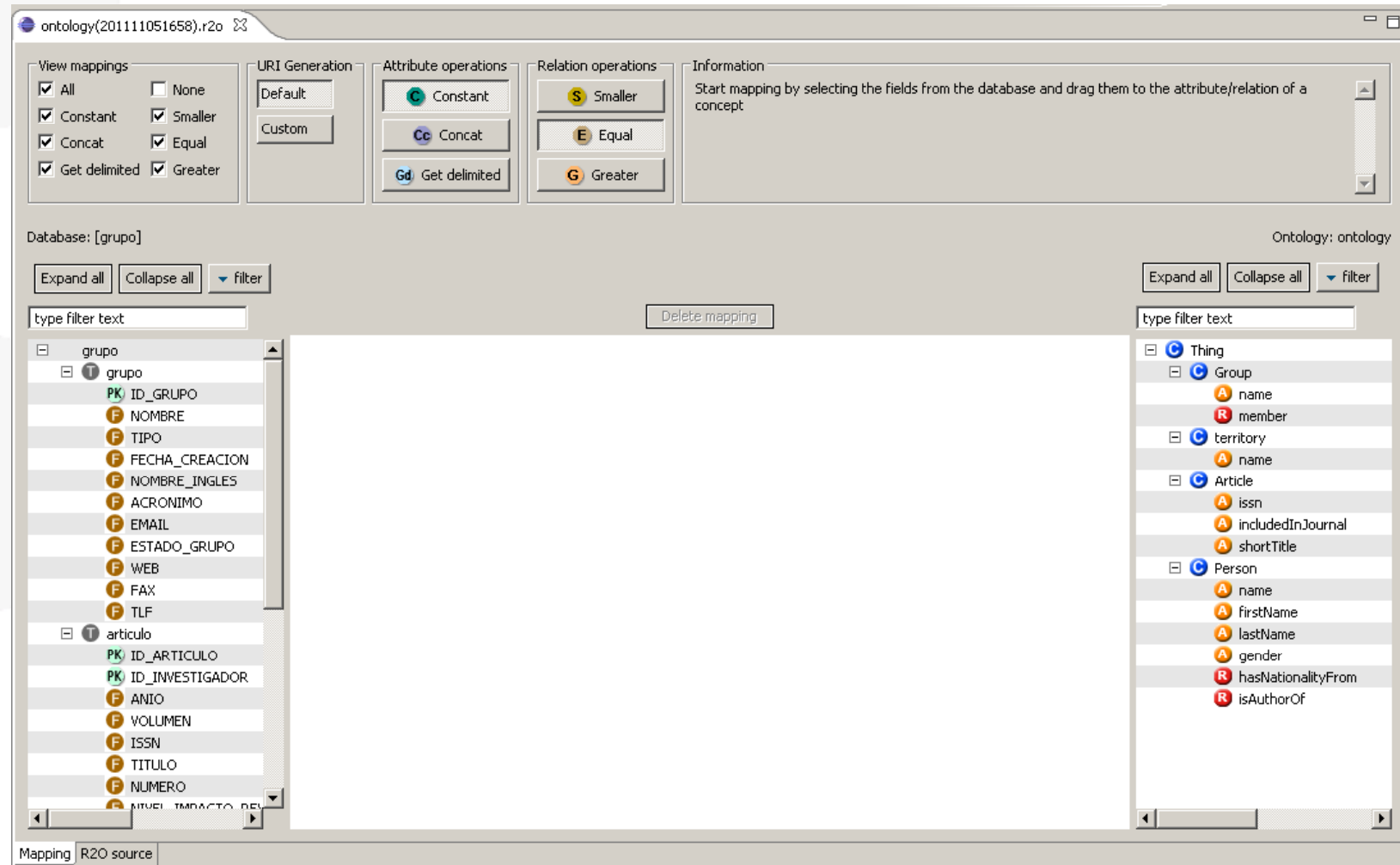
Field	Value
DBMS	MySQL
Name	grupo
Host	localhost
Port	3306
Username	root
Password	****

Buttons: OK, Cancel (for the sub-dialog); Finish, Cancel (for the main dialog).

ODEMapster – DB Connections



R2O Mapping Perspective

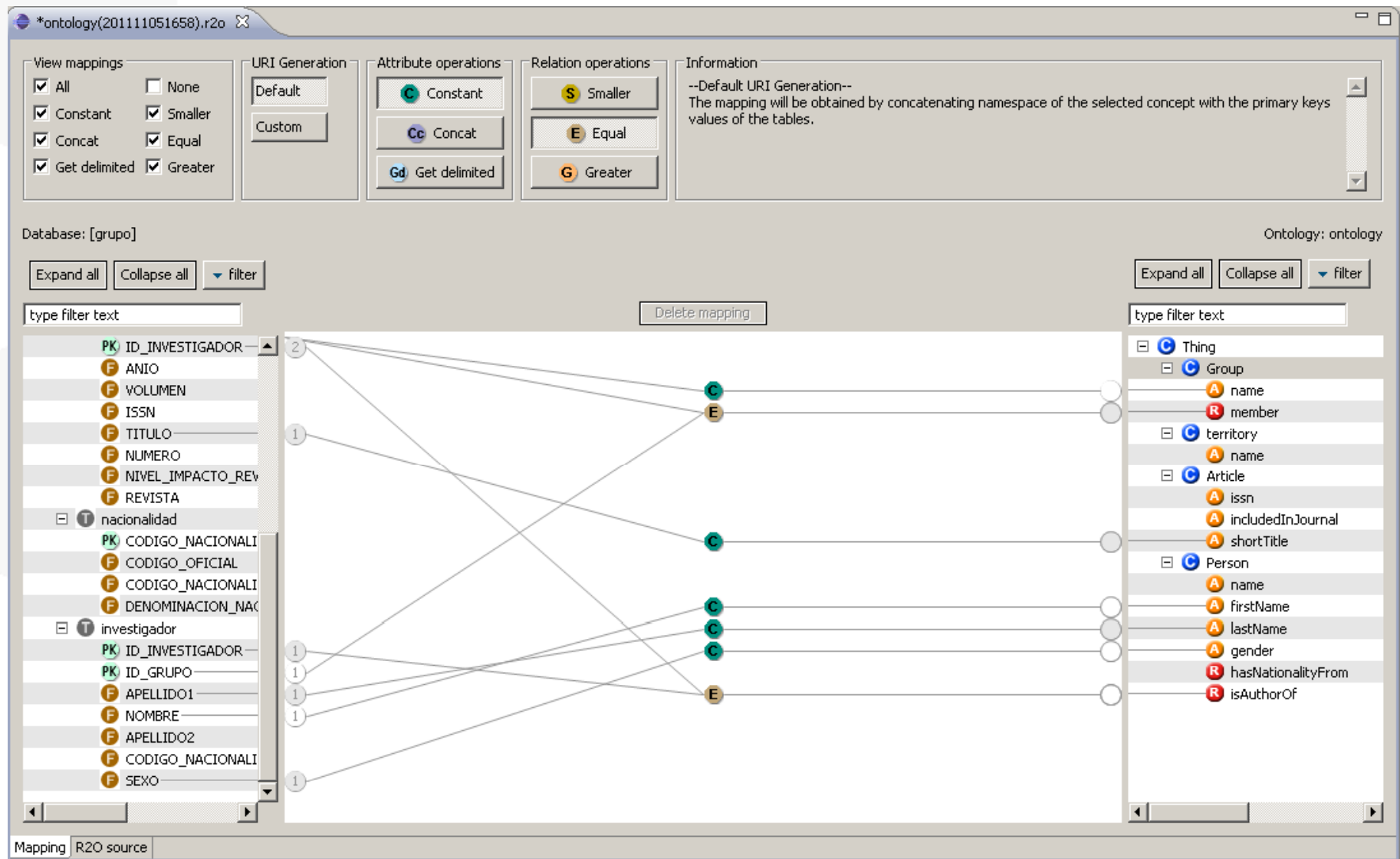


ODEMapster – Creating Mappings

The screenshot displays the ODEMapster application interface for creating mappings between a database and an ontology. The interface is divided into several sections:

- Database: [grupo]** (Left Panel):
 - Buttons: Expand all, Collapse all, filter.
 - Filter input: type filter text.
 - Tree view:
 - grupo
 - PK ID_GRUPO
 - NOMBRE
 - TIPO
 - FECHA_CREACION
 - NOMBRE_INGLES
 - ACRONIMO
 - EMAIL
 - ESTADO_GRUPO
 - WEB
 - FAX
 - TLF
 - articulo
 - PK ID_ARTICULO
 - PK ID_INVESTIGADOR
 - ANIO
 - VOLUMEN
 - ISSN
 - TITULO
 - NUMERO
 - NIVEL_IMPACTO_REV
- Mapping Area** (Center):
 - Contains a single mapping: NOMBRE (from the database).
 - Buttons: Delete mapping.
- Ontology: ontology** (Right Panel):
 - Buttons: Expand all, Collapse all, filter.
 - Filter input: type filter text.
 - Tree view:
 - Thing
 - Group
 - name
 - member
 - territory
 - name
 - Article
 - issn
 - includedInJournal
 - shortTitle
 - Person
 - name
 - firstName
 - lastName
 - gender
 - hasNationalityFrom
 - isAuthorOf
- Bottom Bar:** Mapping | R2O source

ODEMapster – Creating Mappings



ODEMapster2 – command line version

- Odemapster2 folder
- research.r2o.properties

```
r2o.file.path=research.r2o.xml
output.file.path=research.rdf

#RDF/XML|RDF/XML-ABBREV|N-TRIPLE|TURTLE|N3
output.rdflanguage=RDF/XML

#memory|tdb|hsql
jena.mode.type=memory

split_output_per_concept=no
#split_output_per_concept=yes

model.addtypestoinstances=yes

no_of_database=1
database.name[0]=grupo
database.driver[0]=com.mysql.jdbc.Driver
database.url[0]=jdbc:mysql://localhost:3306
database.user[0]=root
database.pwd[0]=root
database.type[0]=mysql
```

Main sections of an R2O Mapping



ODEMapster2 – command line version

- research.r2o.xml

```
<?xml version="1.0" encoding="UTF-8" ?>
```

conceptmap-def

```
<r2o>
```

```
  <conceptmap-def name="http://xmlns.com/foaf/0.1/Person" identified-by="person">
```

uri-as

```
    <has-table name="grupo.investigador" />
```

```
    <uri-as encodeURI="true">
```

```
      <described-by>
```

attributemap-def

```
        <attributemap-def name="http://xmlns.com/foaf/0.1/firstName">
```

```
        <attributemap-def name="http://xmlns.com/foaf/0.1/lastName">
```

```
        <attributemap-def name="http://xmlns.com/foaf/0.1/gender">
```

dbrelationmap-def

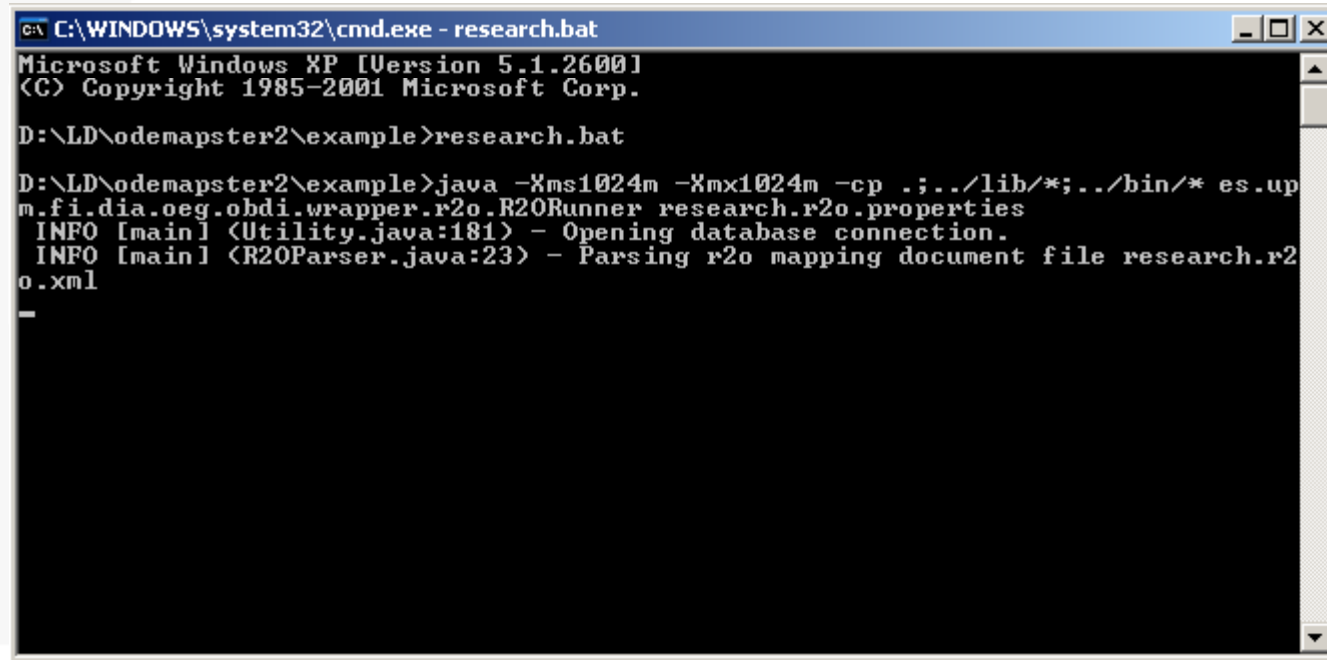
```
        <dbrelationmap-def name="http://research.ull.es/ontology/isAuthorOf" to-concept="article" identified-by="isauthorof">
```

```
      </described-by>
```

```
    </conceptmap-def>
```

ODEMapster2 – command line version

- research.bat



```
C:\WINDOWS\system32\cmd.exe - research.bat
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

D:\LD\odemaster2\example>research.bat

D:\LD\odemaster2\example>java -Xms1024m -Xmx1024m -cp ../lib/*;../bin/* es.up
m.fi.dia.oeg.obdi.wrapper.r2o.R2ORunner research.r2o.properties
INFO [main] (Utility.java:181) - Opening database connection.
INFO [main] (R2OParser.java:23) - Parsing r2o mapping document file research.r2
o.xml
-
```

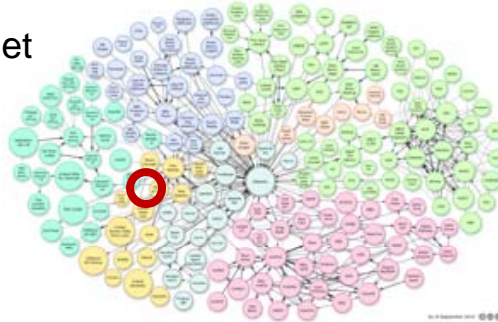
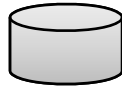
- File generated

```
<rdf:Description rdf:about="http://research.ull.es/resource/Person/jorge_mauricio-espinoza-mejia">
  <j.2:isAuthorOf rdf:resource="http://research.ull.es/resource/Article/a_note_on_ontology_localization"/>
  <j.2:isAuthorOf rdf:resource="http://research.ull.es/resource/Article/discovering_the_semantics_of_user_keywords"/>
  <j.1:gender rdf:datatype="http://www.w3.org/2001/XMLSchema#string">V</j.1:gender>
  <j.1:lastName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ESPINOZA MEJÍA</j.1:lastName>
  <j.1:firstName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">JORGE MAURICIO</j.1:firstName>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">JORGE MAURICIO ESPINOZA MEJÍA</rdfs:label>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
</rdf:Description>
<rdf:Description rdf:about="http://research.ull.es/resource/Person/boris_marcelo-villazon-terrazas">
  <j.2:isAuthorOf rdf:resource="http://research.ull.es/resource/Article/a_pattern-based_method_for_re-engineering_non-ontological_resources_into_ontologies"/>
  <j.1:gender rdf:datatype="http://www.w3.org/2001/XMLSchema#string">V</j.1:gender>
  <j.1:lastName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">VILLAZÓN TERRAZAS</j.1:lastName>
  <j.1:firstName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">BORIS MARCELO</j.1:firstName>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">BORIS MARCELO VILLAZÓN TERRAZAS</rdfs:label>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
</rdf:Description>
<rdf:Description rdf:about="http://research.ull.es/resource/Article/anotacion_de_contenidos_en_la_web_del_futuro">
  <j.0:issn rdf:datatype="http://www.w3.org/2001/XMLSchema#string">0211-2124</j.0:issn>
  <j.0:shortTitle rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Anotación de contenidos en la Web del futuro</j.0:shortTitle>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Article"/>
</rdf:Description>
<rdf:Description rdf:about="http://research.ull.es/resource/Person/cesar_augusto-montana-roa">
  <j.1:gender rdf:datatype="http://www.w3.org/2001/XMLSchema#string">V</j.1:gender>
  <j.1:lastName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MONTAÑA ROA</j.1:lastName>
  <j.1:firstName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">CÉSAR AUGUSTO</j.1:firstName>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">CÉSAR AUGUSTO MONTAÑA ROA</rdfs:label>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
</rdf:Description>
```

Linking - Identify suitable data sets as linking targets



<http://ckan.net>



Semantic Web Dog Food Corpus
<http://data.semanticweb.org/>

Endpoint

<http://data.semanticweb.org/snorql/>

SPARQL:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX swrc-ext: <http://www.cs.vu.nl/~mcakleir/onto/swrc_ext/2005/05#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX ical: <http://www.w3.org/2002/12/cal/ical#>
PREFIX swc: <http://data.semanticweb.org/ns/swc/ontology#>

SELECT DISTINCT ?property ?hasValue ?isValueOf
WHERE {
  { <http://data.semanticweb.org/person/boris-villazon-terrazas> ?property ?hasValue }
  UNION
  { ?isValueOf ?property <http://data.semanticweb.org/person/boris-villazon-terrazas> }
}
ORDER BY (!BOUND(?hasValue)) ?property ?hasValue ?isValueOf
```

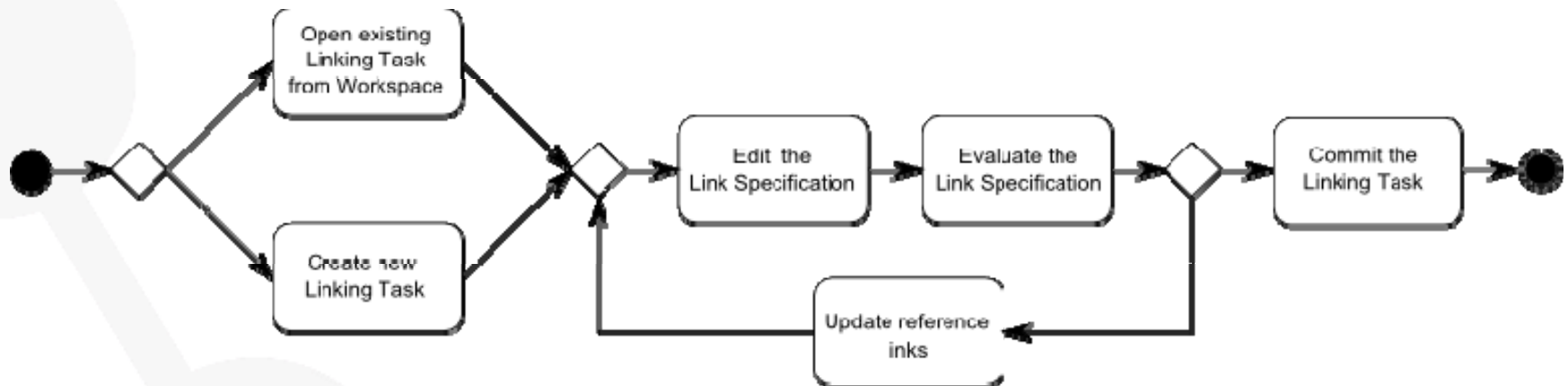
Results:

Description of <http://data.semanticweb.org/person/boris-villazon-terrazas>:

property	hasValue	
swc:affiliation	"Universidad Politecnica de Madrid"	-
swc:holdsRole	< http://data.semanticweb.org/conference/swc/2010/research/reviewer >	-
swc:holdsRole	< http://data.semanticweb.org/workshop/cold/2011/pcmember >	-
swc:holdsRole	< http://data.semanticweb.org/workshop/ocas/2011/pcmember >	-
swrc:affiliation	< http://data.semanticweb.org/organization/ontology-engineering-group-universidad-politecnica >	-
swrc:affiliation	< http://data.semanticweb.org/organization/universidad-politecnica-de-madrid >	-

Linking – Silk framework

- <http://www4.wiwiss.fu-berlin.de/bizer/silk/>
 - Copy workbench.war to the webapps directory (tomcat directory)
- Silk Workbench



Link Specification Language

SILK – Link Specification Language

prefixes

prefix

datasources

Datasource

Interlinks

Interlink

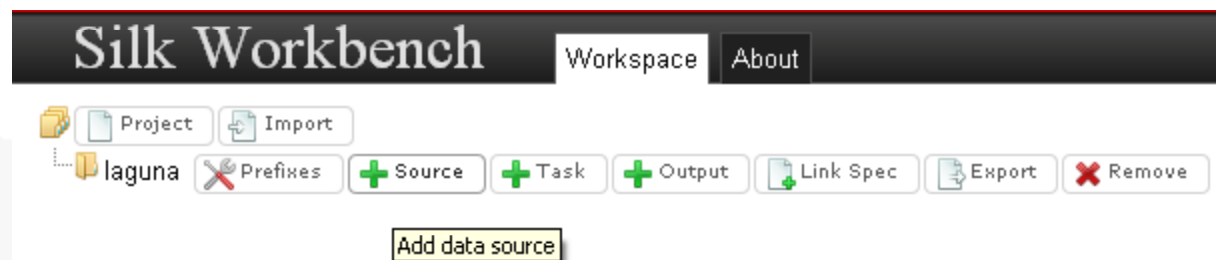
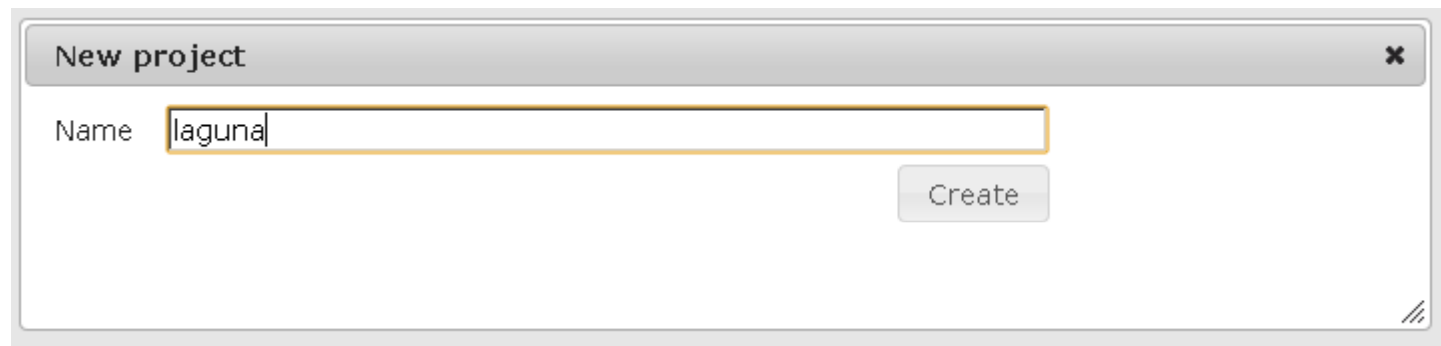
SourceDataset

TargetDataset

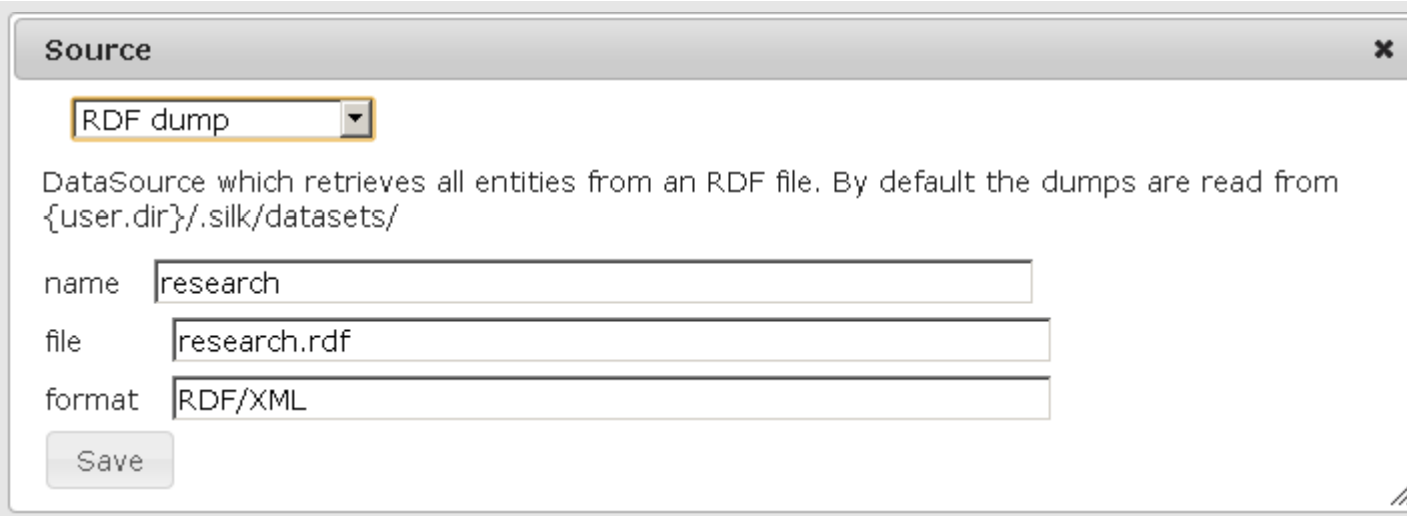
LinkageRule

Outputs

- <http://localhost/workbench>



- First source



Source [X]

RDF dump [v]

DataSource which retrieves all entities from an RDF file. By default the dumps are read from {user.dir}/.silk/datasets/

name

file

format

- Second source

Source

×

SPARQL Endpoint ▾

DataSource which retrieves all entities from a SPARQL endpoint

name	data-semanticweb
endpointURI	http://data.semanticweb.org/sparql
login	
password	
graph	
pageSize	1000
entityList	
pauseTime	0
retryCount	3
retryPause	1000

Save

- Add a task

New linking task [X]

Name	<input type="text" value="researchlinks"/>
Source	<input type="text" value="research"/>
Source restriction	<input type="text" value="?a rdf:type foaf:Person"/>
Target	<input type="text" value="data-semanticweb"/>
Target restriction	<input type="text" value="?b rdf:type foaf:Person"/>
Link type	<input type="text" value="http://www.w3.org/2002/07/owl#sameAs"/>

Create

- Add output

Output [X]

File [v]

Writes the links to a file. Links are written to {user.dir}/.silk/output/ by default.

name

file

format

- Edit Linking task

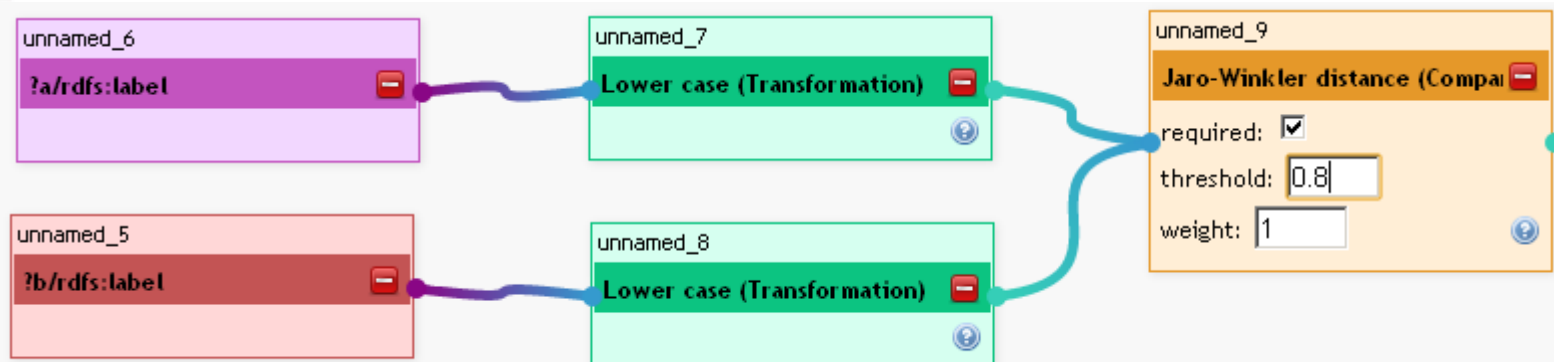
The screenshot displays the Silk Workbench interface. At the top, a dark header bar contains the title 'Silk Workbench' and several tabs: 'Workspace: laguna', 'Editor: researchlinks', 'Generate Links', 'Sample Links', 'Reference Links', 'Learn', and 'About'. Below the header, a light gray bar features buttons for 'Undo', 'Redo', 'Export as Silk-LS', and 'Help'.

The main workspace is divided into a left sidebar and a large central area. The sidebar contains several sections:

- Property Paths** (with a purple icon):
 - Source:** A purple pill-shaped button labeled 'research'.
 - Restriction:** A light purple rectangular box containing the text '(custom path)'.
- Target:** A red rectangular box containing the text '(custom path)'.
- Transformations** (with a green icon): A list of three green rectangular buttons: 'Alpha reduce', 'Concatenate', and 'Convert Charset'. Each button has a small upward or downward arrow on its right side.
- Comparators** (with an orange icon): A list of three orange rectangular buttons: 'Date', 'DateTime', and 'Dice coefficient'. Each button has a small upward or downward arrow on its right side.
- Aggregators** (with a blue icon): This section is currently empty.

The central workspace is a large, empty light gray area, likely for visualizing the linking task configuration.

- Edit researchlinks



- Generate links

The screenshot displays the Silk Workbench application interface. At the top, a dark grey header bar contains the title 'Silk Workbench' and several tabs: 'Workspace: laguna', 'Editor: researchlinks', 'Generate Links' (which is the active tab), 'Sample Links', 'Reference Links', 'Learn', and 'About'. Below the header, a light grey bar features a 'Start' button (indicated by a black arrow pointing to it), a status indicator 'Idle (0,0%)', and a 'Help' button. The main workspace area shows a table with two columns: 'Source' and 'Target'. The 'Source' column contains the value 'research' and the 'Target' column contains 'data-semanticweb'. Above the table, there are controls for 'Expand All', 'Collapse All', and a 'Filter' input field. Below the table, there are 'Prev', '1', and 'Next' navigation buttons. A large black arrow points from the table area down to a 'Generate Links' dialog box. This dialog box has a title bar with a close button (X) and contains an 'Output' label followed by a dropdown menu showing 'researchlinksoutput'. An 'OK' button is located at the bottom right of the dialog box.

- Validate links

Silk Workbench Workspace: laguna Editor: researchlinks Generate Links Sample Links Reference Links Learn About

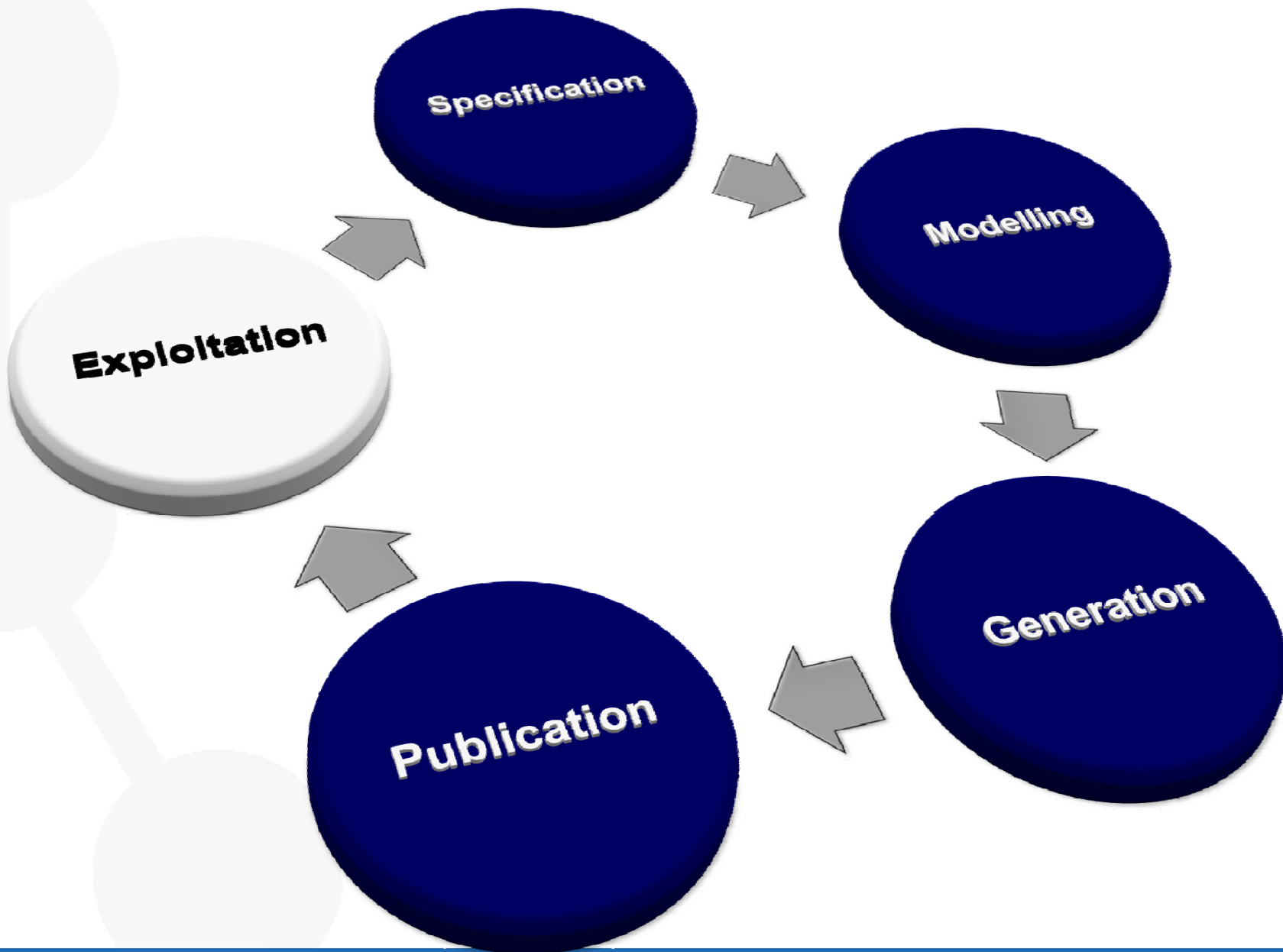
Start Generate Links Task finished in 21s (100,0%) Help

Expand All Collapse All Prev 1 2 3 4 5 6 7 8 9 10 11 Next Filter:

Source: research	Target: data-semanticweb	Score	Correct?
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/thanh-tran	42,9%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/taowei-wang	25,9%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/takahira-yamaguchi	31,5%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stratos-idreos	35,3%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stijn-heyman	31,3%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stephan-bloehdorn	31,8%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/steffen-staab	31,3%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stefano-spaccapietra	43,5%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stefan-schlobach	32,7%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stefan-decker	30,3%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/sofia-angeletou	40,3%	✓ ? ✕
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/soeren-auer	47,0%	✓ ? ✕

- Export

```
▼<Project>
  ▼<Config>
    ▶<Prefixes>...</Prefixes>
  </Config>
  ▼<SourceModule>
    ▼<Tasks>
      ▼<SourceTask>
        ▶<DataSource type="file" id="research">...</DataSource>
      </SourceTask>
      ▼<SourceTask>
        ▶<DataSource type="sparqlEndpoint" id="data-semanticweb">...</DataSource>
      </SourceTask>
    </Tasks>
  </SourceModule>
  ▼<LinkingModule>
    ▼<Tasks>
      ▼<LinkingTask>
        ▼<LinkSpecification>
          ▼<Interlink id="researchlinks">
            <LinkType>owl:sameAs</LinkType>
            ▼<SourceDataset dataSource="research" var="a">
              <RestrictTo>?a rdf:type <http://xmlns.com/foaf/0.1/Person> .</RestrictTo>
            </SourceDataset>
            ▼<TargetDataset dataSource="data-semanticweb" var="b">
              <RestrictTo>?b rdf:type <http://xmlns.com/foaf/0.1/Person> .</RestrictTo>
            </TargetDataset>
          </LinkSpecification>
          ▼<LinkageRule>
            ▼<Compare weight="1" threshold="0.95" required="true" metric="jaroWinkler" id="ID5">
              ▼<TransformInput function="lowerCase" id="ID3">
                <Input path="?a/rdfs:label" id="ID1"/>
              </TransformInput>
              ▼<TransformInput function="lowerCase" id="ID4">
                <Input path="?b/rdfs:label" id="ID2"/>
              </TransformInput>
            </Compare>
          </LinkageRule>
          <Filter/>
          <Outputs></Outputs>
        </LinkingTask>
      </Tasks>
    </LinkingModule>
  </Project>
```

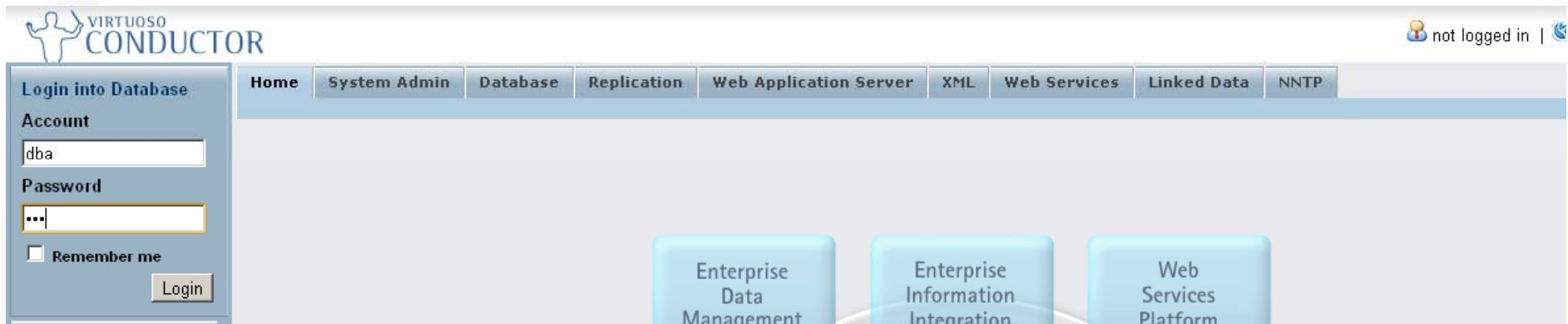


Publication - Virtuoso Open-source edition

- <http://sourceforge.net/projects/virtuoso/files/virtuoso/6.1.4/virtuoso-opensource-win32-20111101.zip/download>
- Unzip to C:\software
- ODBC Registration:
 - Set up VIRTUOSO_HOME variable pointing out to virtuoso-opensource directory:
 - `cd %VIRTUOSO_HOME%\lib`
 - `regsvr32 virtodbc.dll`
- Creating a Windows Service
 - `cd %VIRTUOSO_HOME%\database`
 - `SET PATH=%PATH%;%VIRTUOSO_HOME%\bin;%VIRTUOSO_HOME%\lib`
 - `virtuoso-t -? //to verify`
 - `virtuoso-t +service screate +instance "Instance Name" +configfile virtuoso.ini`
 - `virtuoso-t +service list //to verify`
 - `virtuoso-t -l "Instance Name" +service start //start the service`

Virtuoso - Conductor

- <http://localhost:8890/conductor>



- Upload the generated files

The screenshot shows the Virtuoso Conductor web interface. The top navigation bar includes tabs: Home, System Admin, Database, Replication, Web Application Server, XML, Web Services, Linked Data, and NNTP. Below this is a secondary navigation bar with tabs: SPARQL, Sponger, Statistics, Graphs, Schemas, Namespaces, Access Control, Views, Quad Store Upload, and another unlabeled tab. The 'Quad Store Upload' tab is selected. The main content area is titled 'Quad Store Upload' and contains a form with the following elements:

- ☒ File* (selected) with a 'Choose File' button and the text 'No file chosen'.
- ☐ Resource URL* with an empty text input field.
- ☐ Create graph explicitly.
- Named Graph IRI* with a text input field containing 'http://localhost:8890/graph/ontology/'.
- Buttons for 'Cancel' and 'Upload'.

On the left side of the interface, there is a sidebar with links: Interactive SQL (ISQL), WebDAV Browser, Virtuoso Start Menu, Documentation (web), Tutorials (web), Virtuoso Web Site, and OpenLink Software. At the bottom of the sidebar, it says 'Version: 06.01.3127' and 'Build: Nov 1 2011'. In the top right corner, there is a 'logc' link with a user icon.

- Ontology: <http://research.ull.es/graph/ontology-research.owl>
- Dataset: <http://research.ull.es/graph/dataset-research.rdf>
- Links: <http://research.ull.es/graph/links-output.nt>

- <http://localhost:8890/sparql>

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where { [] a ?Concept } LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout: milliseconds *(values less than 1000 are ignored)*

Options: ☒ Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))



Run Query

Reset

- Now you can play a bit with SPARQL ... ;)

- VOiD description
 - void.ttl

```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ww: <http://vocab.org/waiver/terms/norms> .
@prefix sd: <http://www.w3.org/ns/sparql-service-description#> .
@prefix : <#> .

<> a void:DatasetDescription ;
    dcterms:title "A void Description of the ....." ;
    dcterms:creator <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/phd/39-bvillazon> ;
    foaf:primaryTopic :ResearchGroupsLaguna .

:ResearchGroupsLaguna rdf:type void:Dataset ;
    foaf:homepage <http://research.ull.es/> ;
    dcterms:title "...";
    dcterms:description "...";
    void:subset :VoID ;
    void:subset :RGL .

:VoID
    a sd:Graph, void:Dataset ;
    dcterms:title "SD+VoID triples" ;
    void:dataDump <http://mccarthy.dia.fi.upm.es/data/void.ttl> ;
    void:triples 53 .

:RGL
    a sd:Graph, void:Dataset ;
    dcterms:title "...";
    dcterms:description "...";
```

Metadata Publication – CKAN.net / thedatahub.org

Title

Url

Author

Home

Groups

At ☐ Delete

Lic ☐ Delete

Descri ☐ Delete

Maint: **Tags**

New key with value

New key with value

New key with value

New key with value

You can use [Markdown](#)

Enable effective discovery

- Sindice: the best RDF search engine

The screenshot shows the Sindice website interface. At the top, there's a navigation bar with links: Home, About, Search, Submit, Forum, and Dev. The main header features the Sindice logo and the tagline 'THE SEMANTIC WEB INDEX'. Below this, a section titled 'Sindice - Data Web Services' describes the service as a place where billion pieces of reusable information can be found across hundreds of millions of web pages. A 'LEARN MORE' button is provided. To the right, a search box is visible with tabs for 'Search', 'Submit', and 'Inspector Tool'. The search box includes radio buttons for 'term', 'property', 'advanced', and 'SIG.MA', a text input field for keywords or URIs, and a 'SEARCH' button. Below the search box, it states 'Searching on about 107.86 million documents.' The main content area is divided into several sections: 'LATEST DATA' showing a list of recent updates with timestamps and links; 'HIGHLIGHTS' featuring a section for 'Anything to Triples Any23' with an 'Introduction' button; 'SINDICE TWEET' displaying a tweet from June 21, 2010; and 'SINDICE BLOG' with articles such as 'Sindice planned downtime this weekend' (dated JUN 09, 2010), 'Any23 v0.4.0 Released' (dated MAY 27, 2010), and 'Any23 v0.3.0 Released' (dated APR 23, 2010).

- Sitemap Protocol
 - Used by web crawlers
 - Efficiently find all your content & discover what has been updated

<http://sitemaps.org/>



A sitemap file contains information regarding one or more URLs on your Web site. The information that is stored there helps search engines better spider your website.

Sitemap.xml example

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://yoursite/</loc>
  </url>
  <url>
    <loc>http://yoursite/products/53546</loc>
  </url>
  <url>
    <loc>http://yoursite/products/98421</loc>
  </url>
  <url>
    <loc>http://yoursite/products/41003</loc>
    <lastmod>2010-06-24</lastmod>
    <changefreq>daily</changefreq>
  </url>
</urlset>
```

Optional parts

Sitemap Protocol: Huge sitemaps

- Gzip-compress your sitemap
- Limit: 50k URLs or 10MB
 - split into multiple sitemap files
 - add a *sitemap index file*

Sitemap Protocol: Discovery

- Publish the sitemap file
- Add a line to `http://yoursite/robots.txt`
 - Web site owners use the `/robots.txt` file to give instructions about their site to web robots; this is called *The Robots Exclusion Protocol*.

Sitemap: `http://yoursite/sitemap.xml`

- Simple command line tool
- Sends a SPARQL query to list all URIs
- Generates sitemap

```
sitemap4rdf http://yoursite/sparql http://yoursite/resource/
```

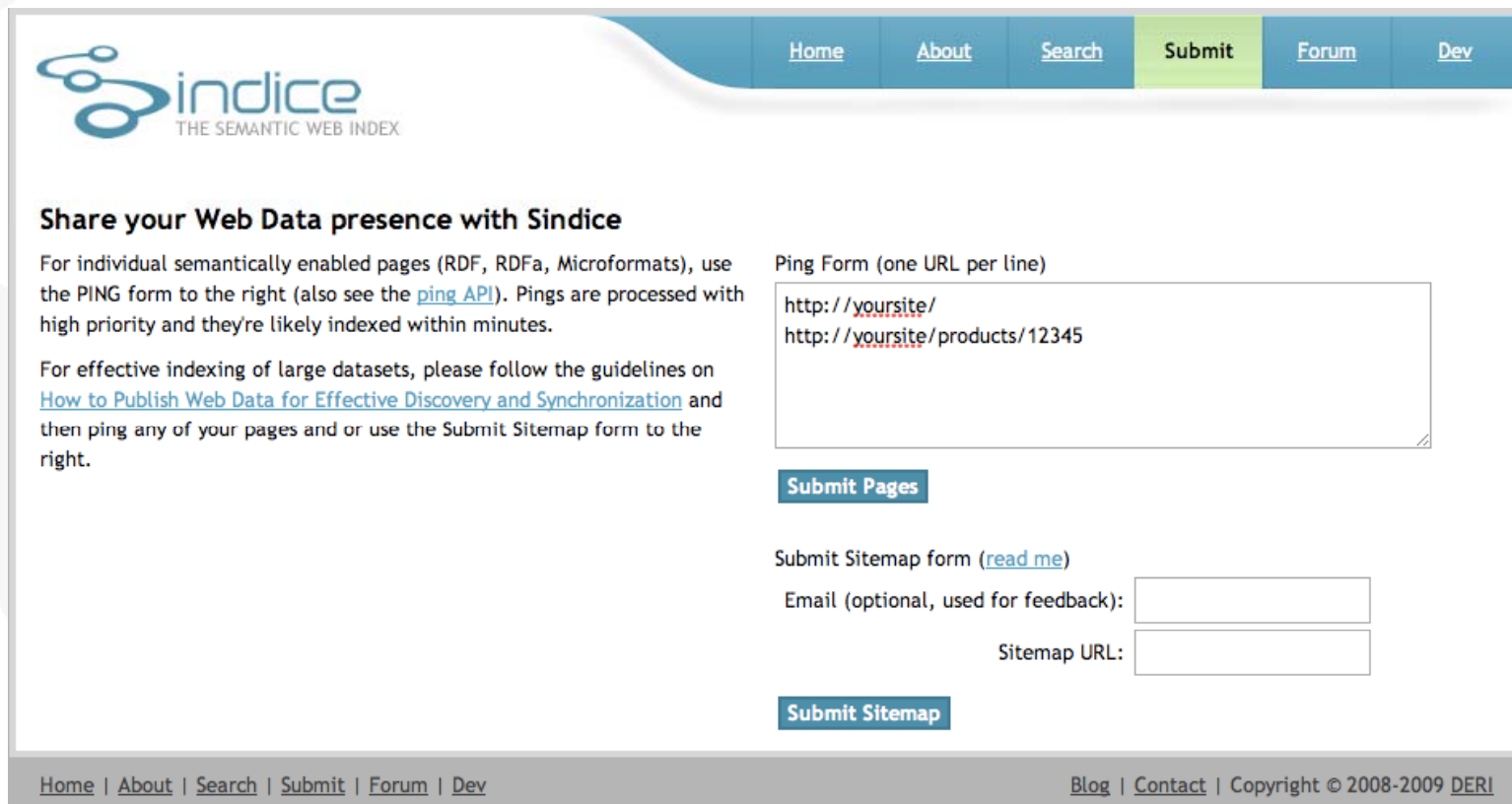
Example:

```
sitemap4rdf http://localhost:8890/sparql http://research.ull.es/
```

- run sitemap4rdf specifying the **SPARQL endpoint** and the **prefix of the URLs** to include in the Sitemap

Submit the sitemap location - Sindice

- <http://sindice.com/main/submit>



The screenshot shows the 'Submit' page of the Sindice website. The header features the 'sindice' logo and a navigation bar with links: Home, About, Search, Submit (highlighted), Forum, and Dev. The main content area is titled 'Share your Web Data presence with Sindice' and provides instructions for submitting individual pages and large datasets. It includes a 'Ping Form' with a text area containing two example URLs. Below the form is a 'Submit Pages' button. Further down is a 'Submit Sitemap form' section with a '(read me)' link, an 'Email' field, and a 'Sitemap URL' field, followed by a 'Submit Sitemap' button. The footer contains a repeat of the navigation links and a copyright notice for DERI.

sindice
THE SEMANTIC WEB INDEX

[Home](#) [About](#) [Search](#) **Submit** [Forum](#) [Dev](#)

Share your Web Data presence with Sindice

For individual semantically enabled pages (RDF, RDFa, Microformats), use the PING form to the right (also see the [ping API](#)). Pings are processed with high priority and they're likely indexed within minutes.

For effective indexing of large datasets, please follow the guidelines on [How to Publish Web Data for Effective Discovery and Synchronization](#) and then ping any of your pages and or use the Submit Sitemap form to the right.

Ping Form (one URL per line)

<http://yoursite/>
<http://yoursite/products/12345>

Submit Pages

Submit Sitemap form ([read me](#))

Email (optional, used for feedback):

Sitemap URL:

Submit Sitemap

[Home](#) | [About](#) | [Search](#) | [Submit](#) | [Forum](#) | [Dev](#) [Blog](#) | [Contact](#) | Copyright © 2008-2009 [DERI](#)

Submit the sitemap location - Google

- <https://www.google.com/webmasters/tools/>

Google webmaster tools

geo.linkeddata.es

[« Back to Home](#)

[Dashboard](#)

[Messages](#)

[Site configuration](#)

Sitemaps

[Crawler access](#)

[Sitelinks](#)

[Change of address](#)

[Settings](#)

[+ Your site on the web](#)

[+ Diagnostics](#)

[+ Labs](#)

Help with:

[Creating and submitting Sitemaps](#)

[Sitemap errors](#)

[Sitemap details](#)

[Sitemaps for multiple sites](#)

Sitemaps

Submit a Sitemap to tell Google about pages on your site we might not otherwise discover.

Submitted URLs

249,318

116,514 URLs in web index

[+ Submit a Sitemap](#)

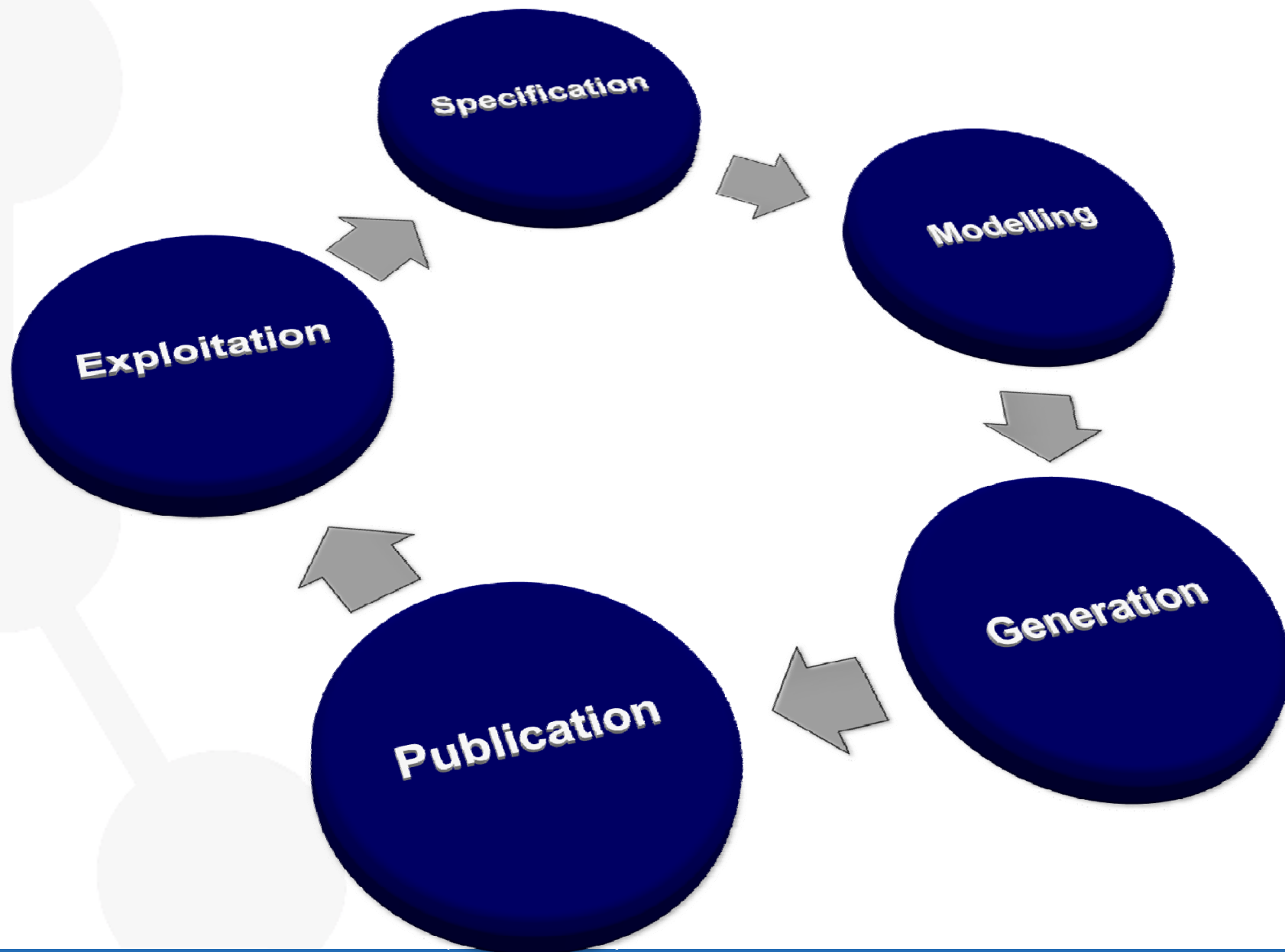
Show submissions: **By me (4)** - [All \(4\)](#)

	Sitemap	Status	Type	Downloaded	URLs submitted	URLs in web index
<input type="checkbox"/>	sitemap.xml.gz	✓	Sitemap	Sep 23, 2010	50,000	42,075
<input type="checkbox"/>	sitemap1.xml.gz	✓	Sitemap	Sep 19, 2010	50,000	7,123
<input type="checkbox"/>	sitemap2.xml.gz	✓	Sitemap	Sep 21, 2010	24,659	9,059
<input type="checkbox"/>	sitemap_index.xml	✓	Index	Sep 24, 2010	124,659	58,257

☐ [Delete](#) [Resubmit](#)

[Download this table](#)

[Download data for all sites](#)



Thank
You!!



Publishing Linked Data from RDB

Boris Villazón-Terrazas, Oscar Corcho

Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

<http://www.oeg-upm.net>

bvillazon@fi.upm.es

Phone: 34.91.3366605, Fax: 34.91.3524819

Slides available at: <http://www.slideshare.net/boricles/>

Acknowledgements: Freddy Priyatna, Jan Schulte, Richard Cyganiak and many others that we may have omitted