

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>

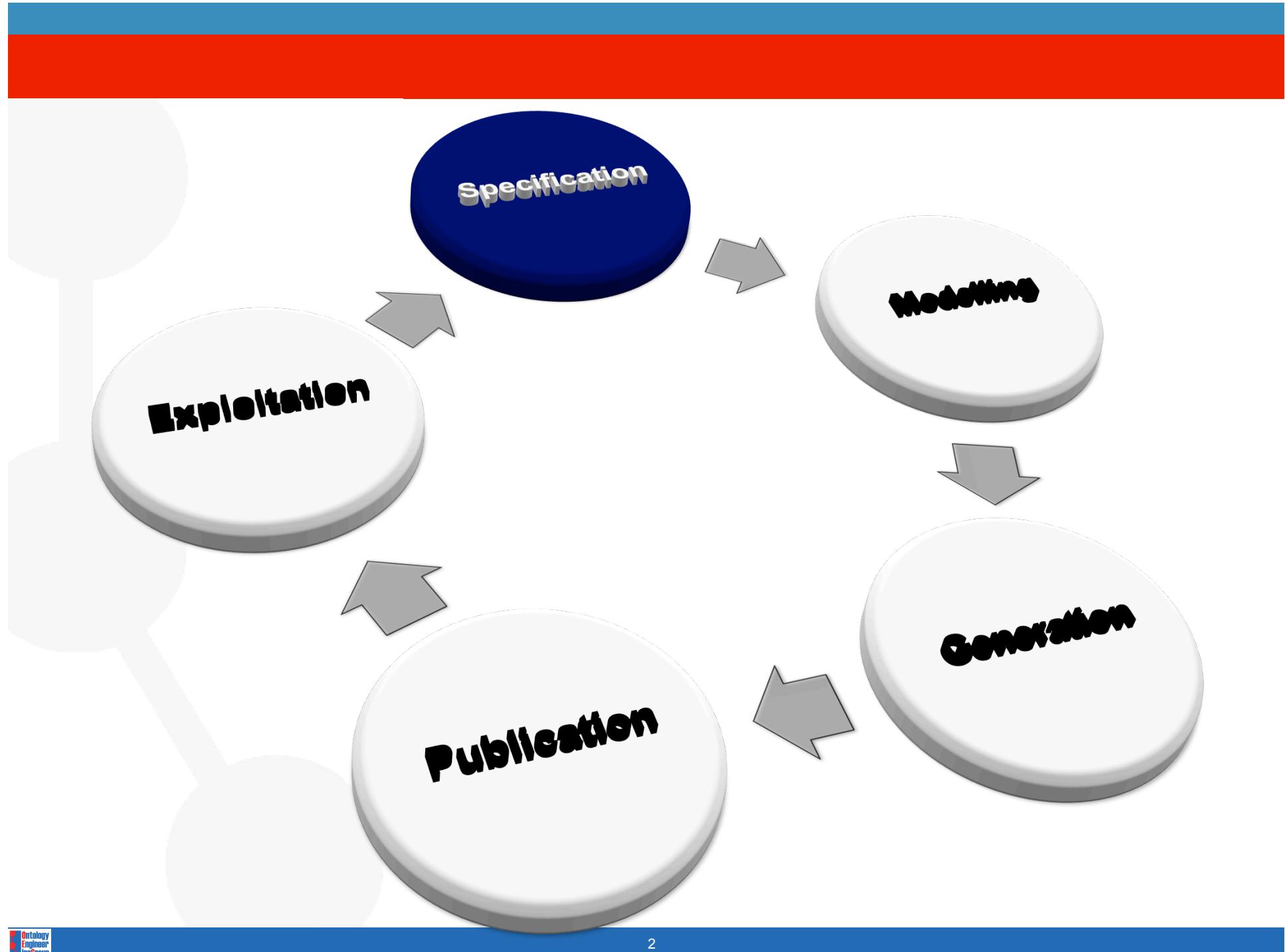
bvillazon@fi.upm.es

Phone: 34.91.3366605, Fax: 34.91.3524819

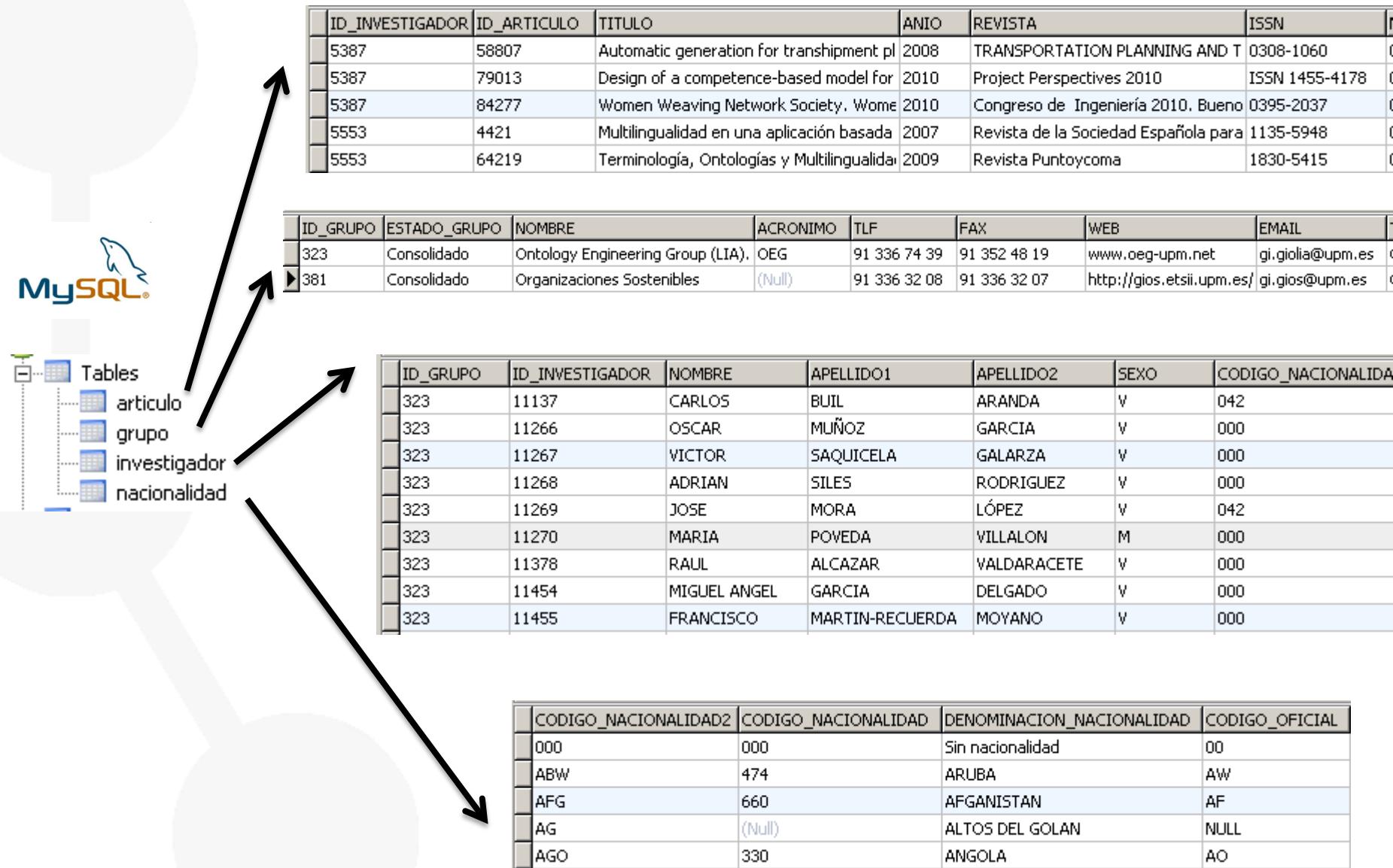
Slides available at: <http://www.slideshare.net/boricles/>

Acknowledgements:

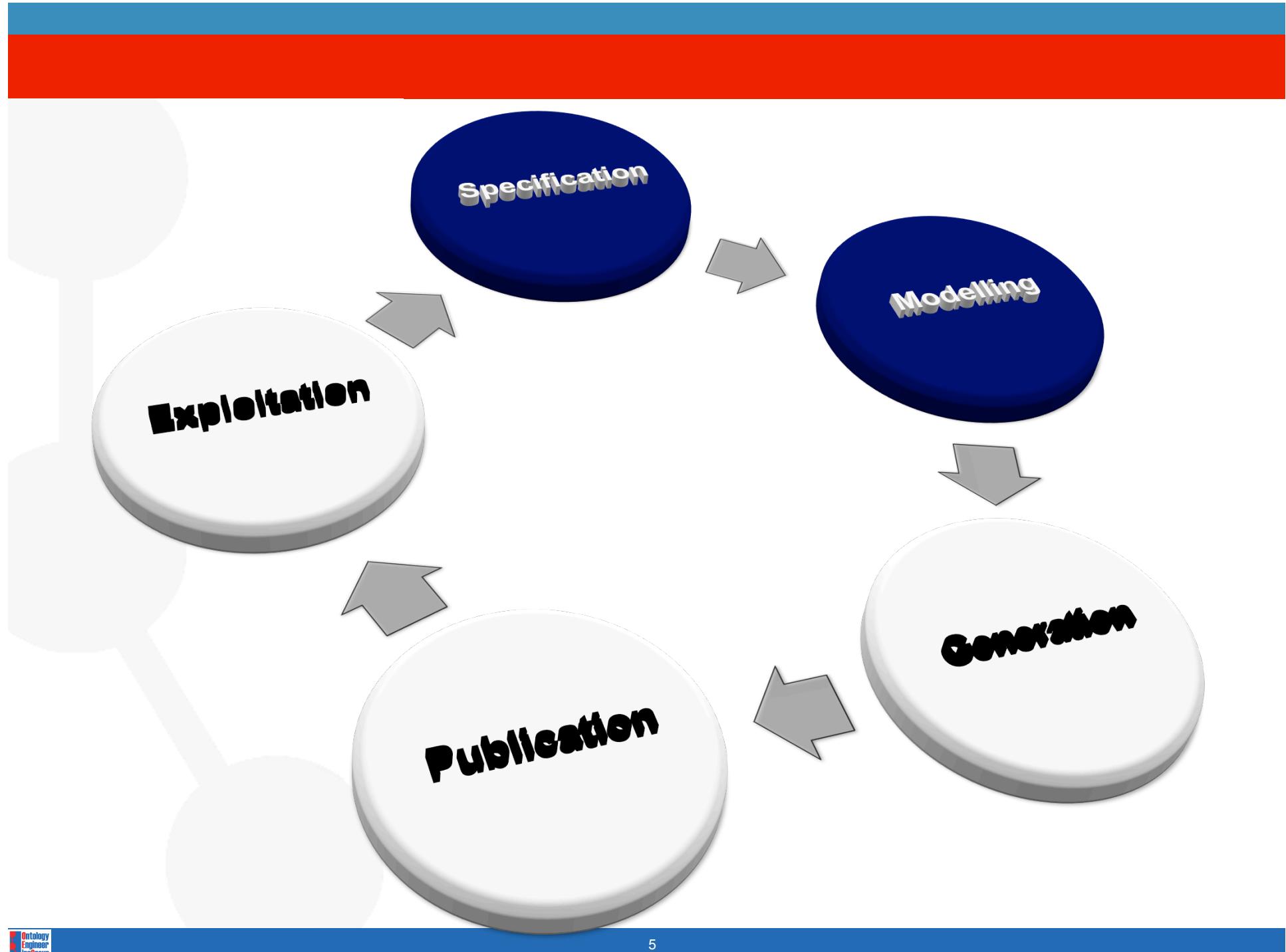
Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0



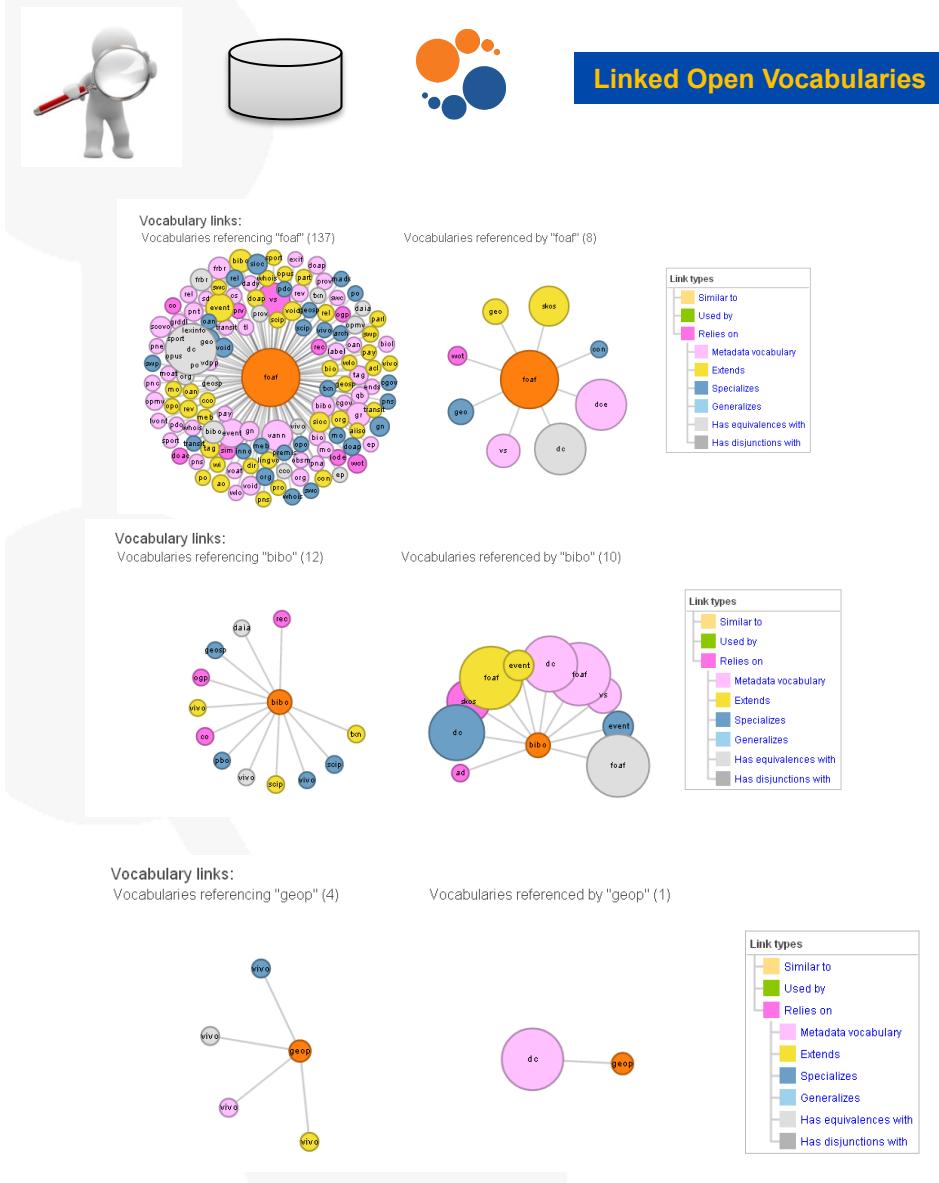
Specification – RDB about Research Groups



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



Modelling



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

FOAF Friend of a Friend vocabulary

BIBO The Bibliographic Ontology

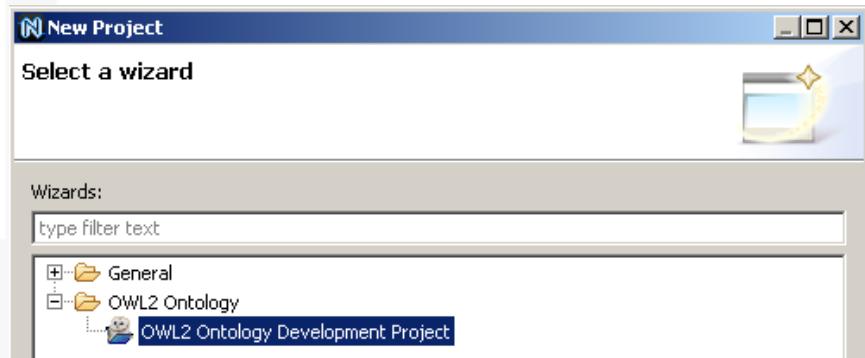
GEOP FAO Geopolitical Ontology

Modelling – NeOn Toolkit

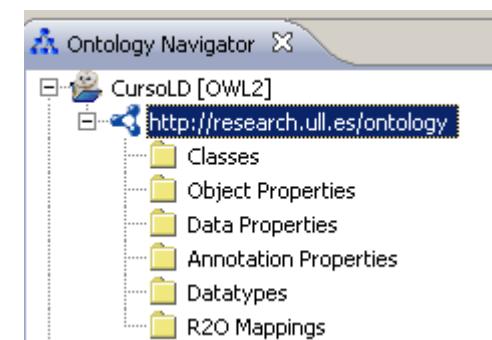
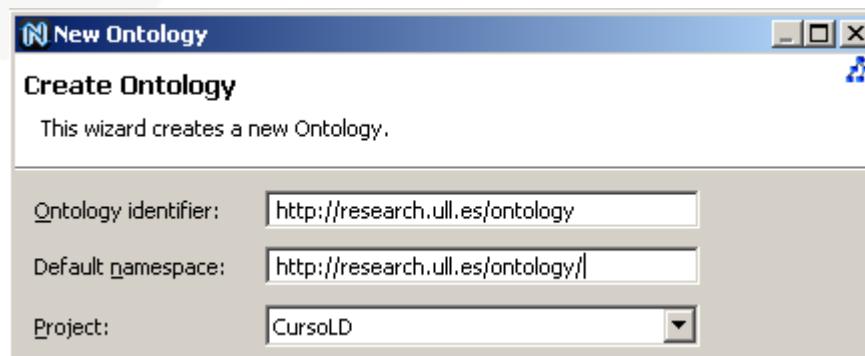


<http://neon-toolkit.org/>

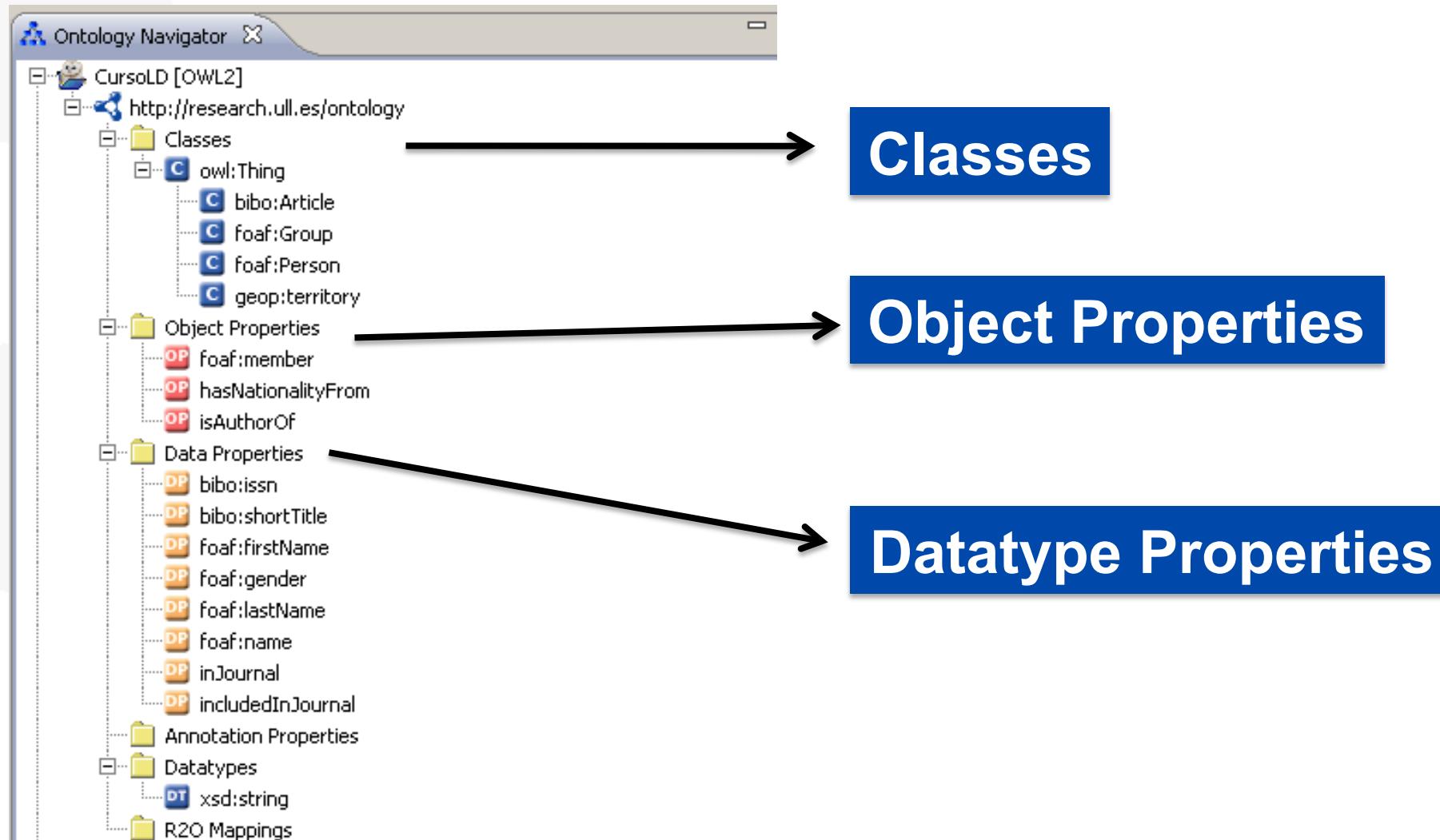
New Project

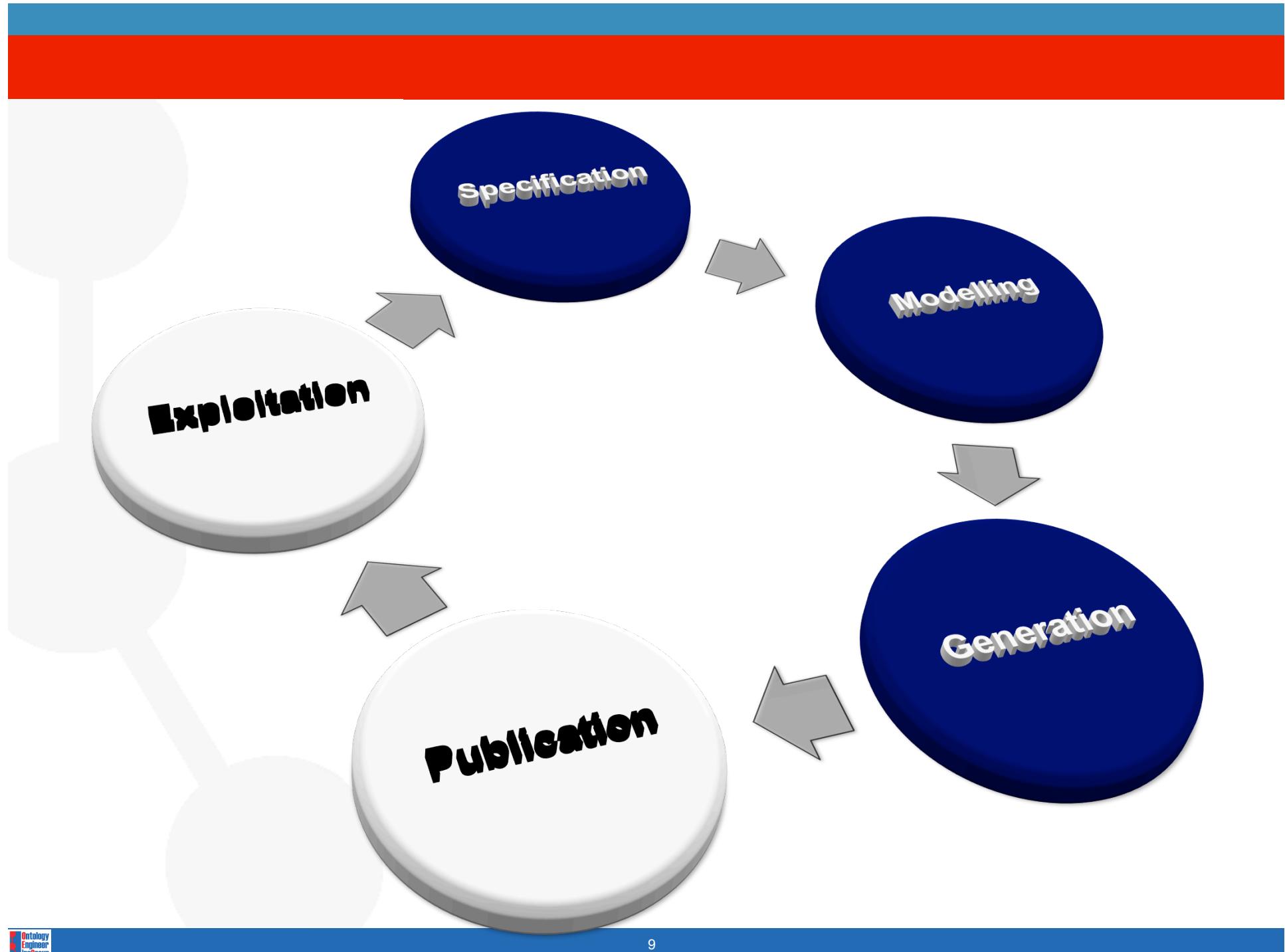


New Ontology



Modelling – Creating some elements





Transformation – RDB2RDF

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



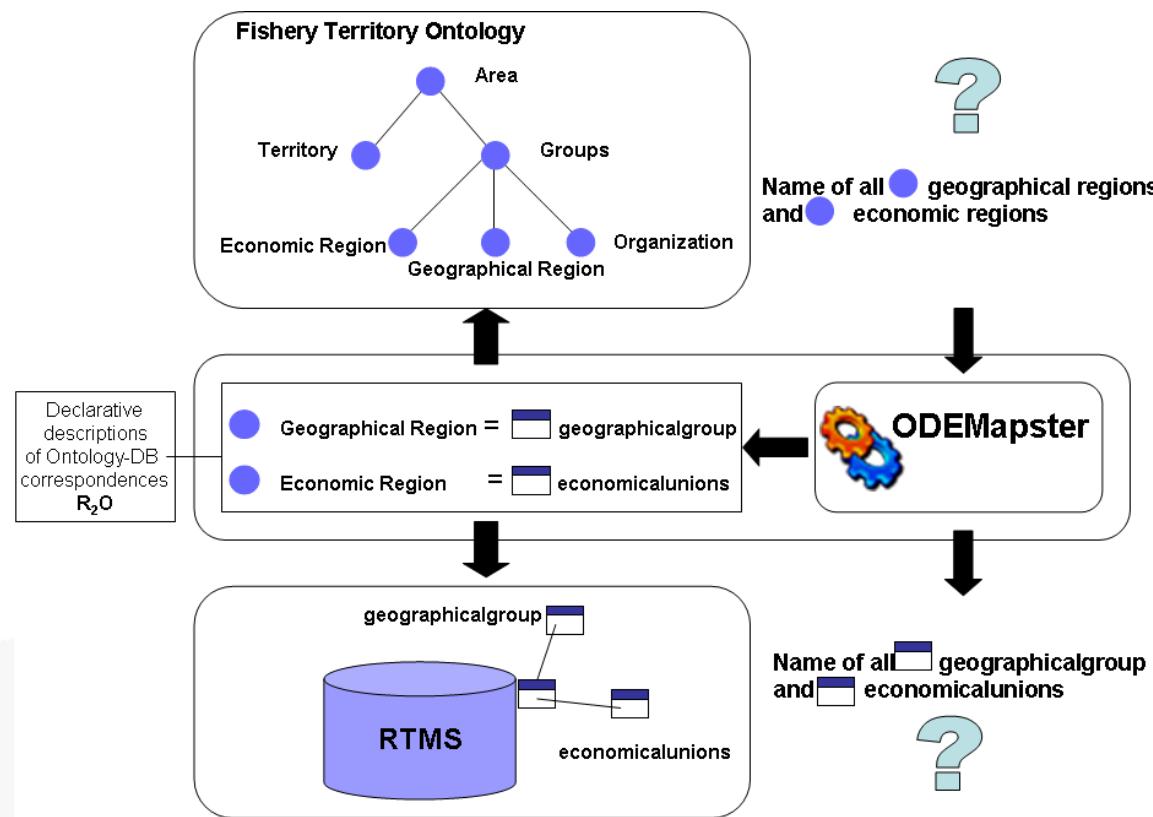
- W3C RDB2RDF Working Group is creating R2RML (RDB2RDF Mapping Language¹⁾)



<http://www.w3.org/TR/r2rml/>

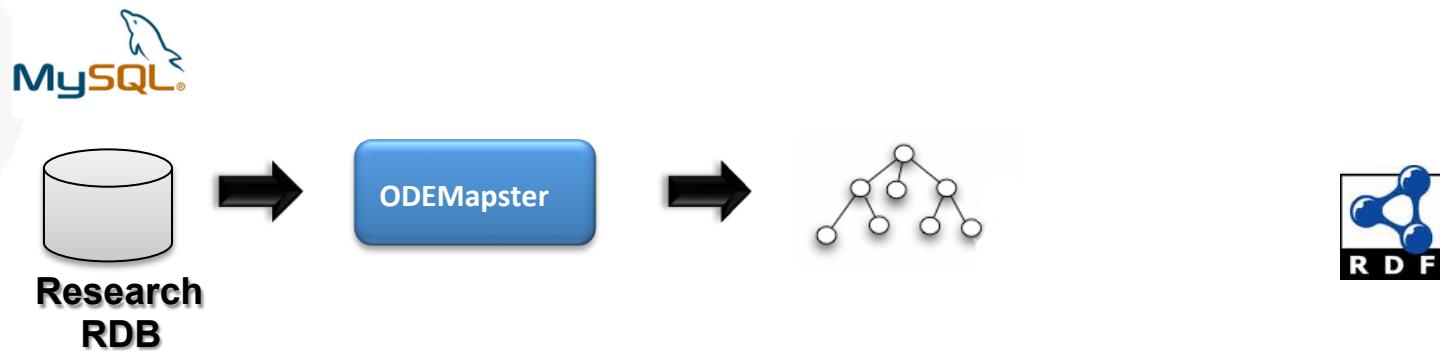
Transformation – R₂O & ODEMastper

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

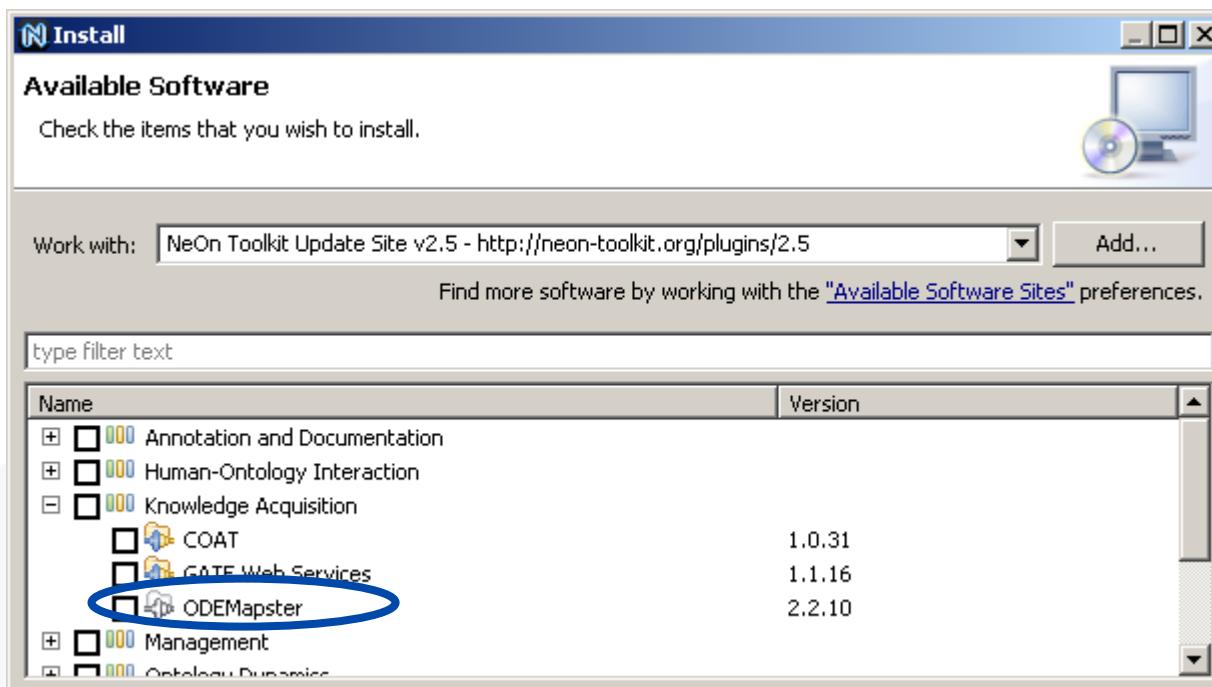


www.oeg-upm.net/index.php/en/downloads/9-r2o-odemaster

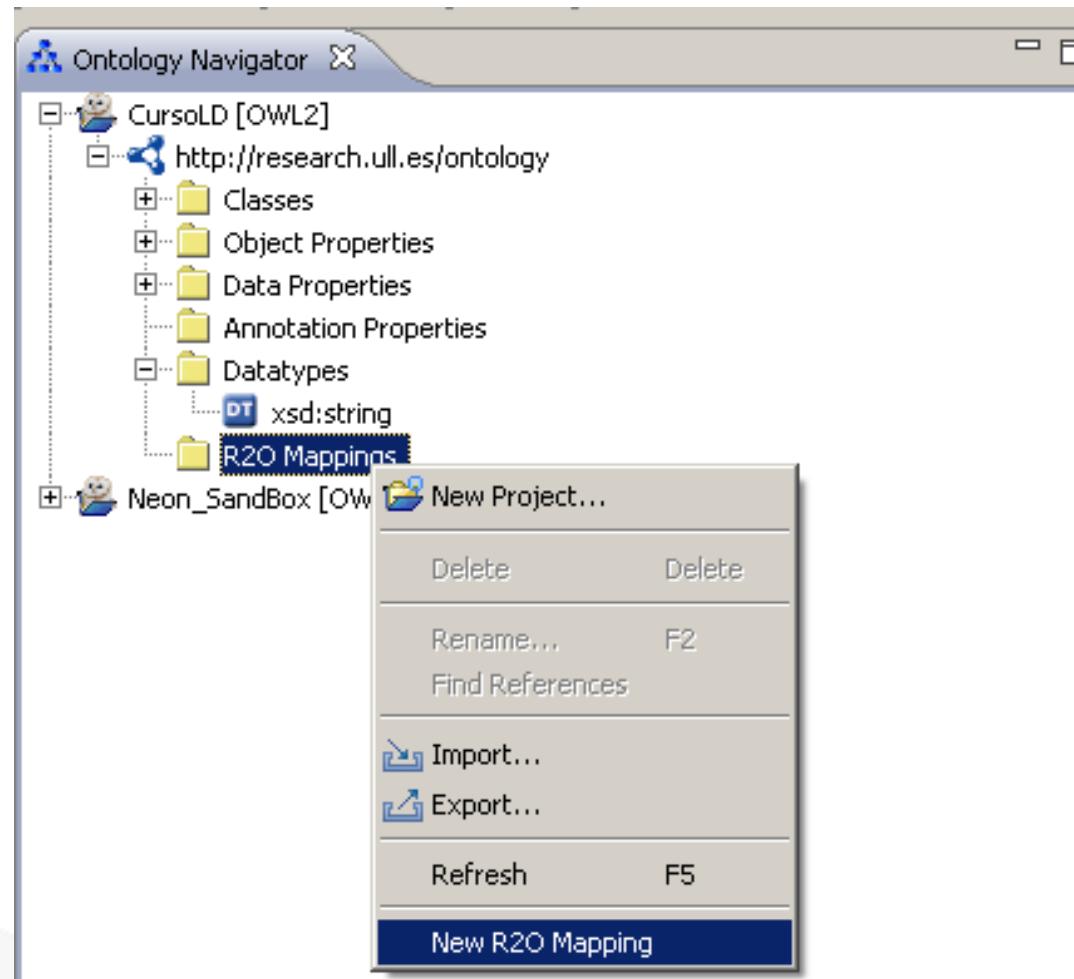
Generation – Transformation - ODEMapster



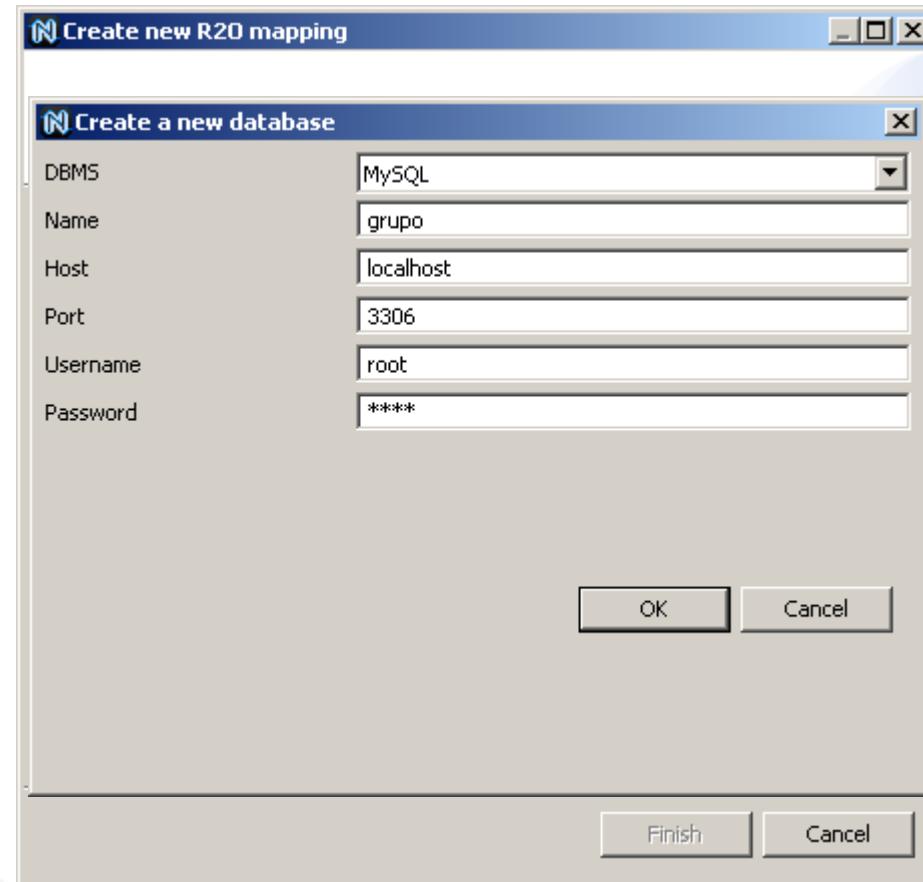
- Included in the NTK – currently under revision



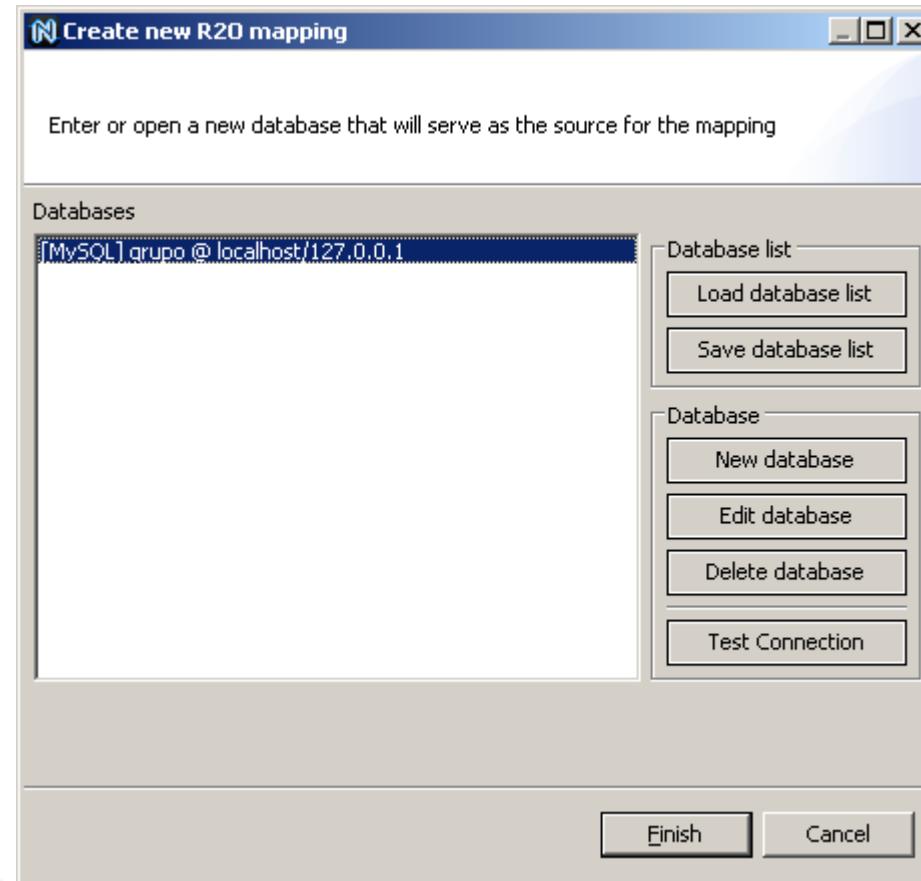
Transformation - ODEMMapster



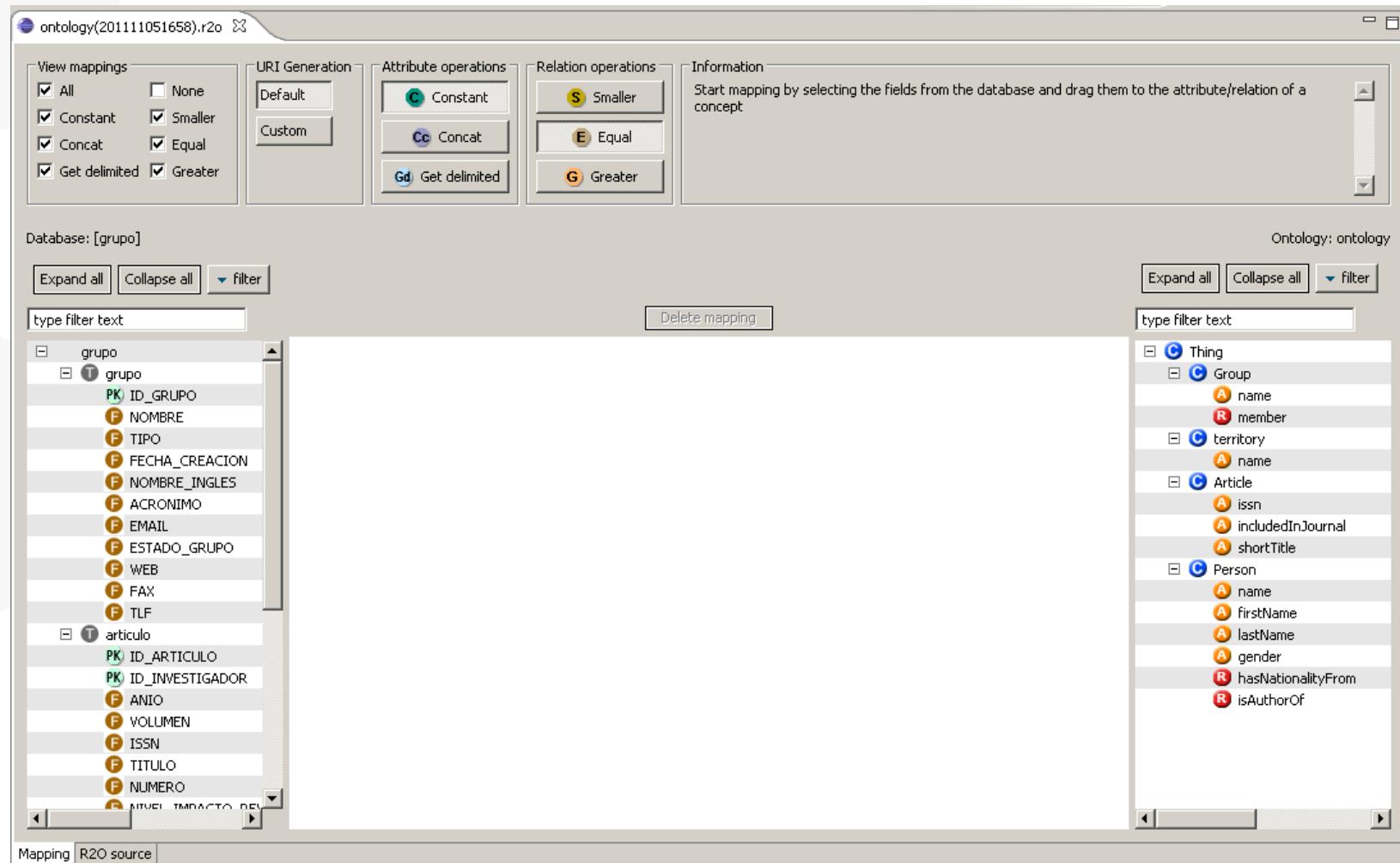
New R2O Mapping – DB connection information



ODEMapster – DB Connections



R2O Mapping Perspective



ODEMapster – Creating Mappings

Database: [grupo]

Expand all Collapse all ▾ filter

type filter text

Delete mapping

grupo

- grupo
 - T grupo
 - PK ID_GRUPO
 - F NOMBRE
 - F TIPO
 - F FECHA_CREACION
 - F NOMBRE_INGLES
 - F ACRONIMO
 - F EMAIL
 - F ESTADO_GRUPO
 - F WEB
 - F FAX
 - F TLF
 - T articulo
 - PK ID_ARTICULO
 - PK ID_INVESTIGADOR
 - F ANIO
 - F VOLUMEN
 - F ISSN
 - F TITULO
 - F NUMERO
 - F NIVEL_IMPACTO_REV

Mapping R2O source

Ontology: ontology

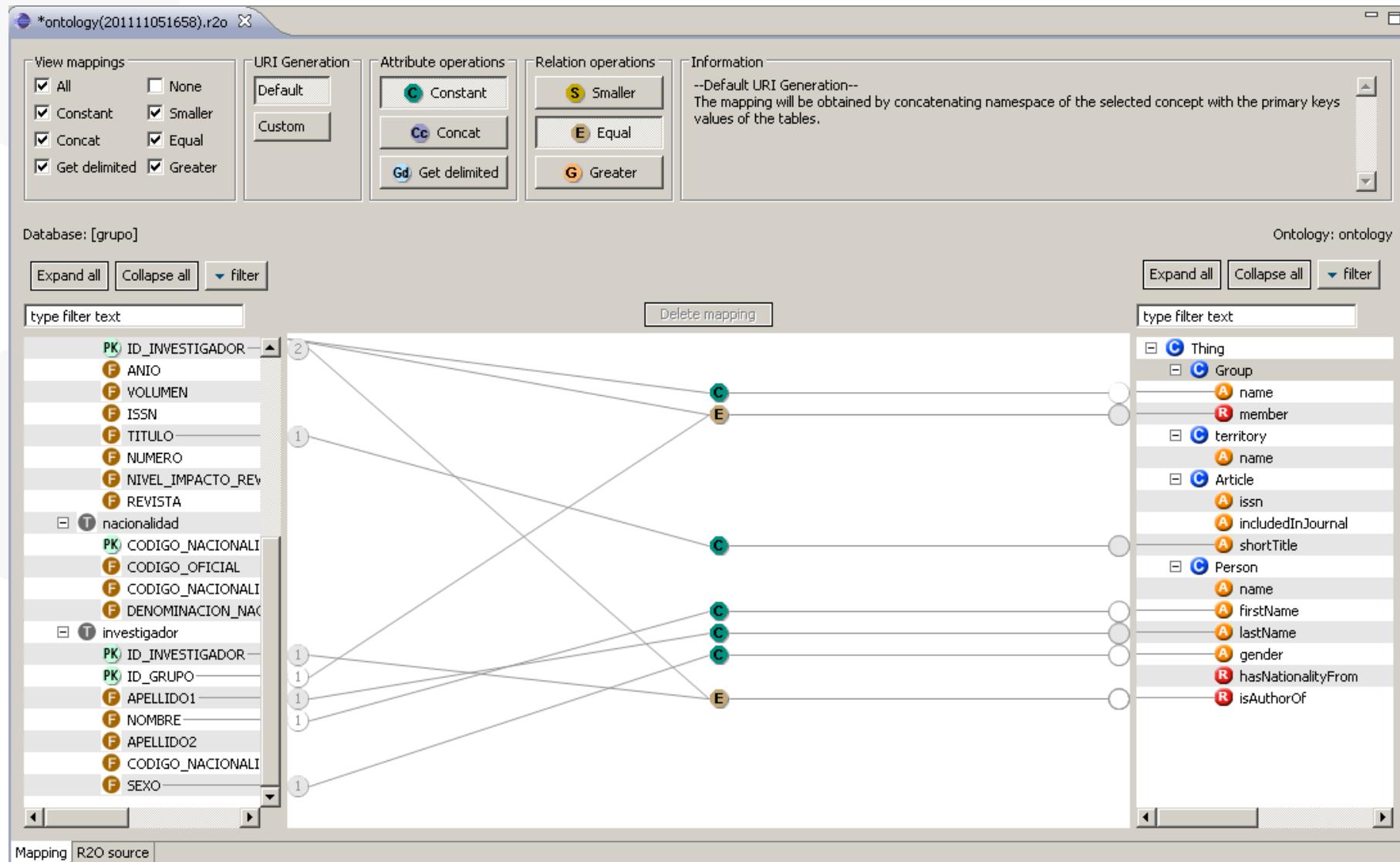
Expand all Collapse all ▾ filter

type filter text

Thing

- Group
 - A name
 - R member
- territory
 - A name
- Article
 - A issn
 - A includedInJournal
 - A shortTitle
- Person
 - A name
 - A firstName
 - A lastName
 - A gender
 - R hasNationalityFrom
 - R isAuthorOf

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|hsq1
jena.mode.type=memory

split_output_per_concept=no
#split_output_per_concept=yes

model.addtypesinstances=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" ?>
<r2o>

    <conceptmap-def name="http://xmlns.com/foaf/0.1/Person" identified-by="person">
        <has-table name="grupo.investigador"/>
        <uri-as encodeURI="true">

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

                </described-by>
            </conceptmap-def>
```

conceptmap-def

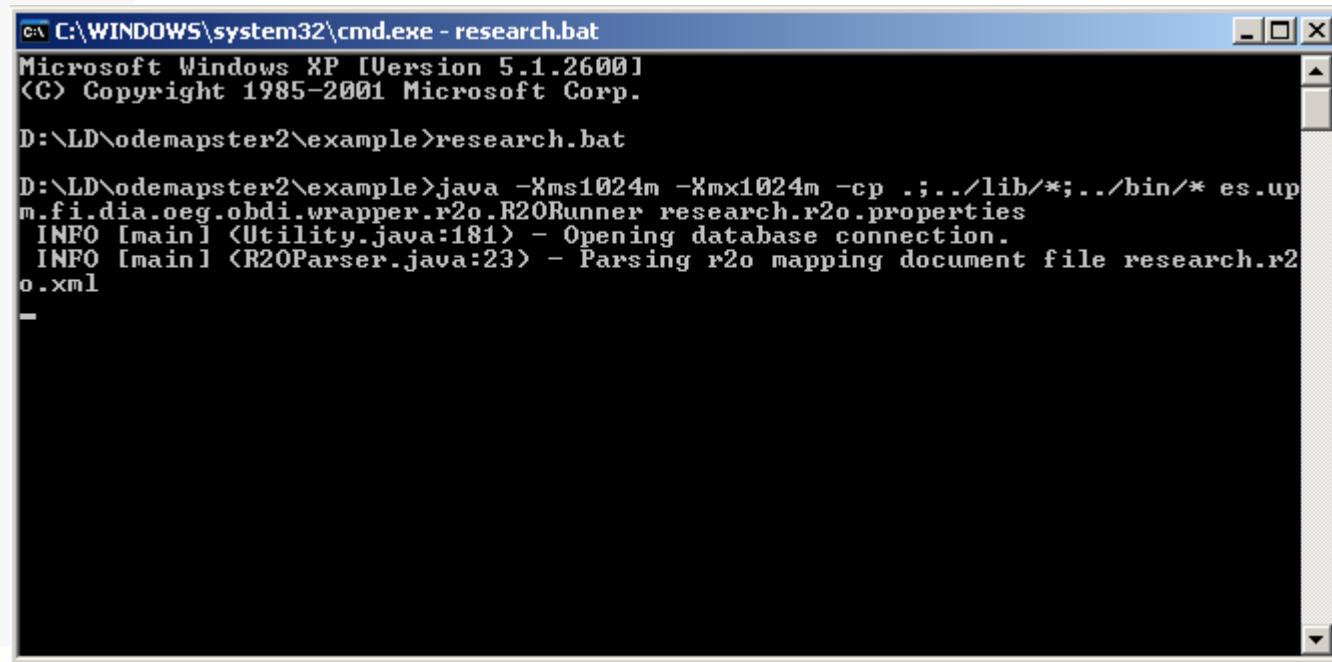
uri-as

attributemap-def

dbrelationmap-def

ODEMapster2 – command line version

- research.bat



The screenshot shows a Microsoft Windows XP command prompt window titled 'cmd C:\WINDOWS\system32\cmd.exe - research.bat'. The window displays the output of running the 'research.bat' script. The script uses Java to run an R2ORunner with properties from 'research.r2o.properties'. It outputs INFO messages about opening a database connection and parsing the mapping document file.

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

D:\LD\odemapster2\example>research.bat

D:\LD\odemapster2\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
-
```

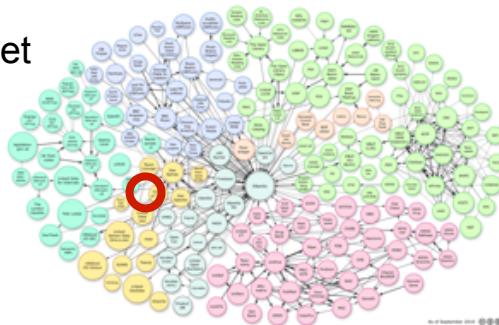
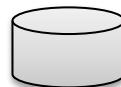
- File generated

```
<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">ESPIÑOZA 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 ESPIÑOZA 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/~macleain/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
```

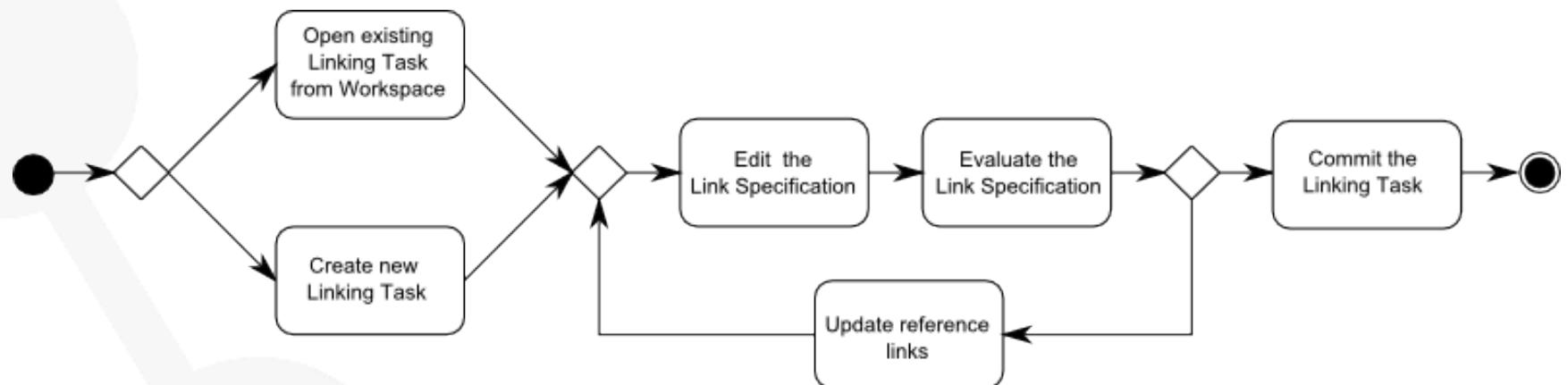
Results:

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

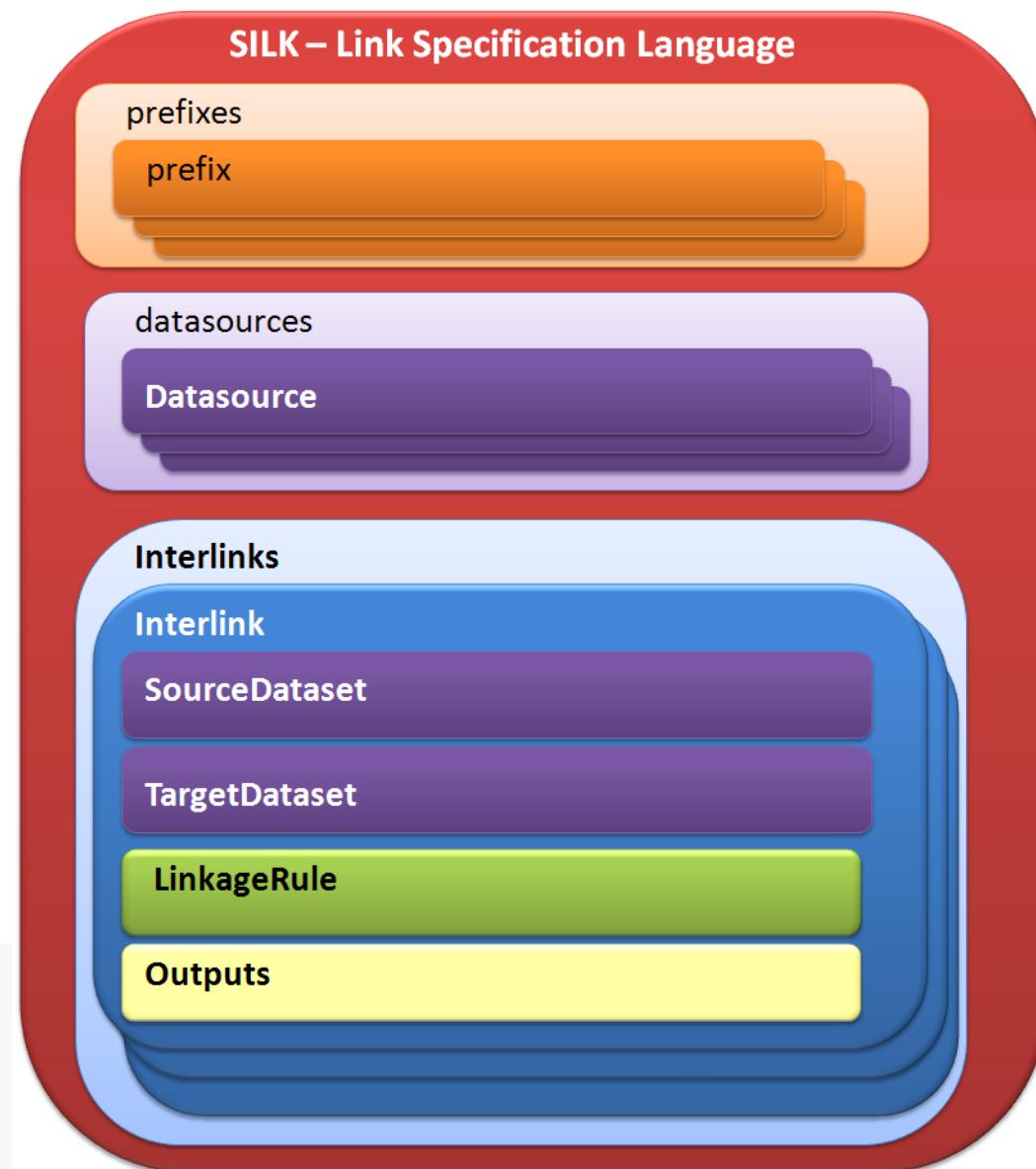
property	hasValue
swc:affiliation ↗	"Universidad Politecnica de Madrid"
swc:holdsRole ↗	< http://data.semanticweb.org/conference/iswc/2010/research/reviewer > ↗
swc:holdsRole ↗	< http://data.semanticweb.org/workshop/cold/2011/pcmember > ↗
swc:holdsRole ↗	< http://data.semanticweb.org/workshop/ocas/2011/pcmember > ↗
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

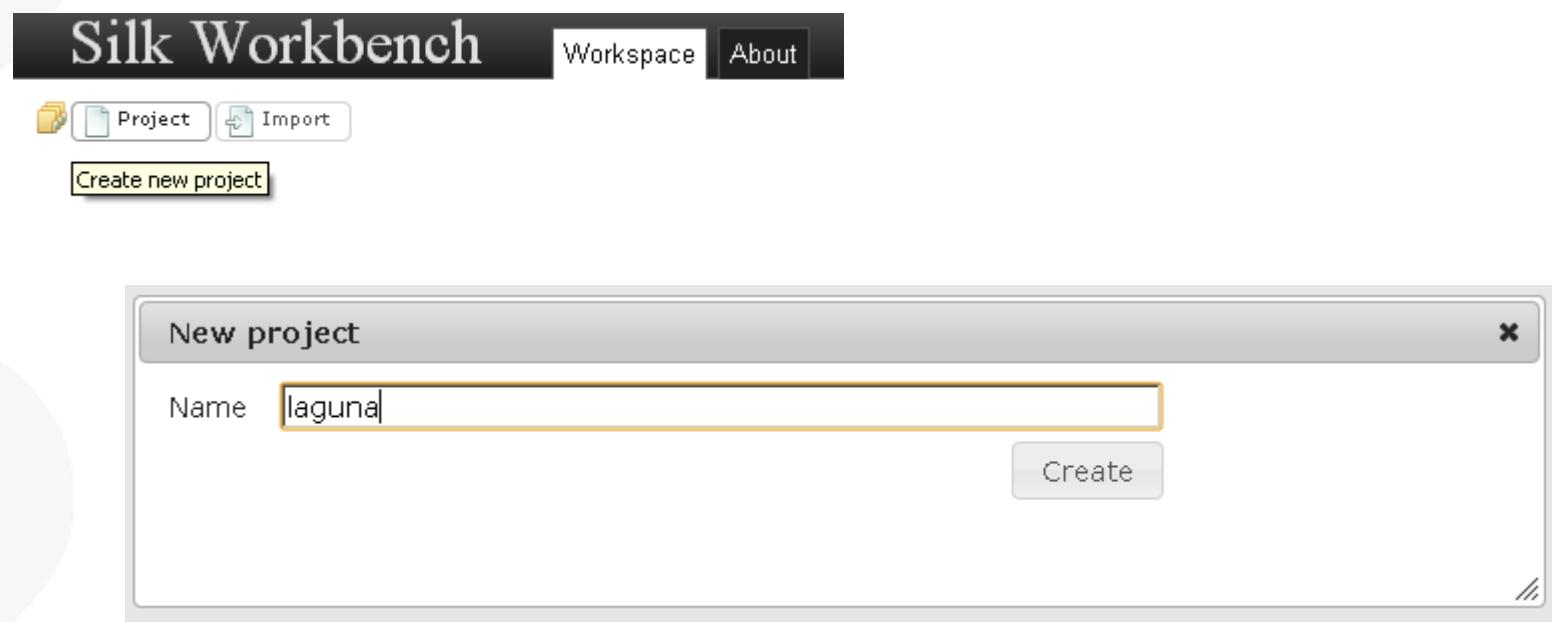
- <http://www4.wiwiiss.fu-berlin.de/bizer/silk/>
 - Copy workbench.war to the webapps directory (tomcat directory)
- Silk Workbench



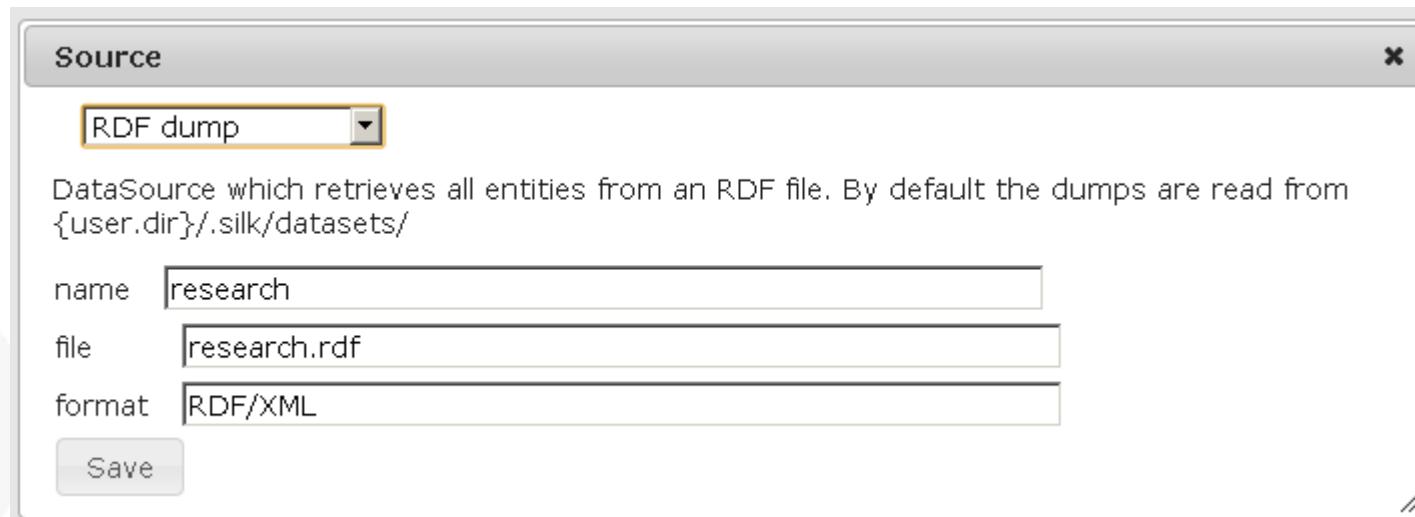
Link Specification Language



- <http://localhost/workbench>



- First source



- Second source

Source ×

SPARQL Endpoint ▼

DataSource which retrieves all entities from a SPARQL endpoint

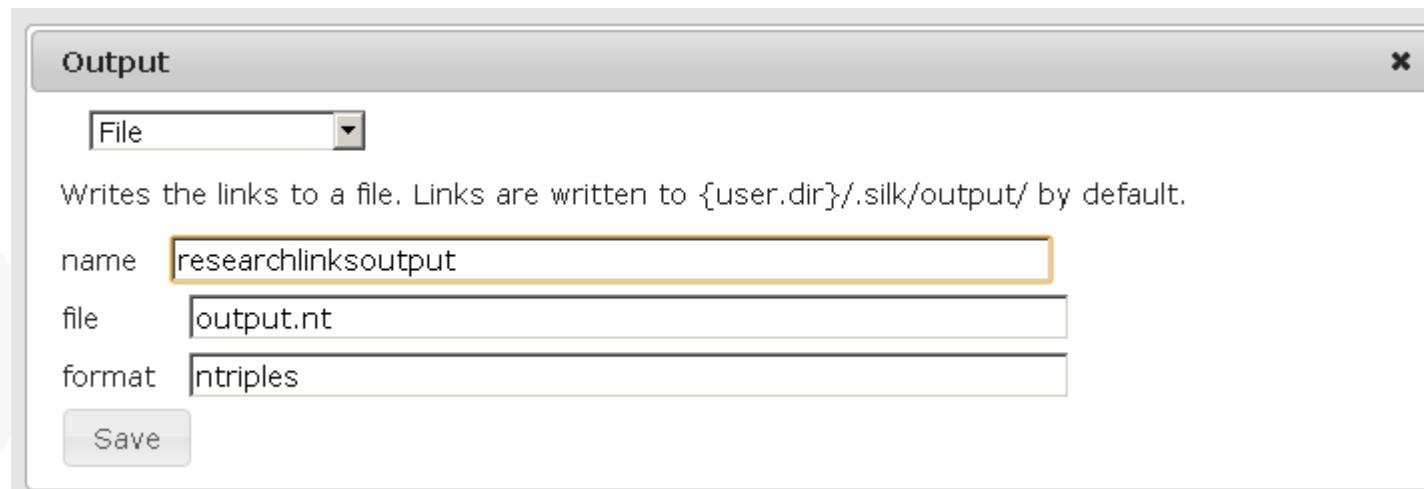
name	data-semanticweb
endpointURI	http://data.semanticweb.org/sparql
login	
password	
graph	
pageSize	1000
entityList	
pauseTime	0
retryCount	3
retryPause	1000

- Add a task

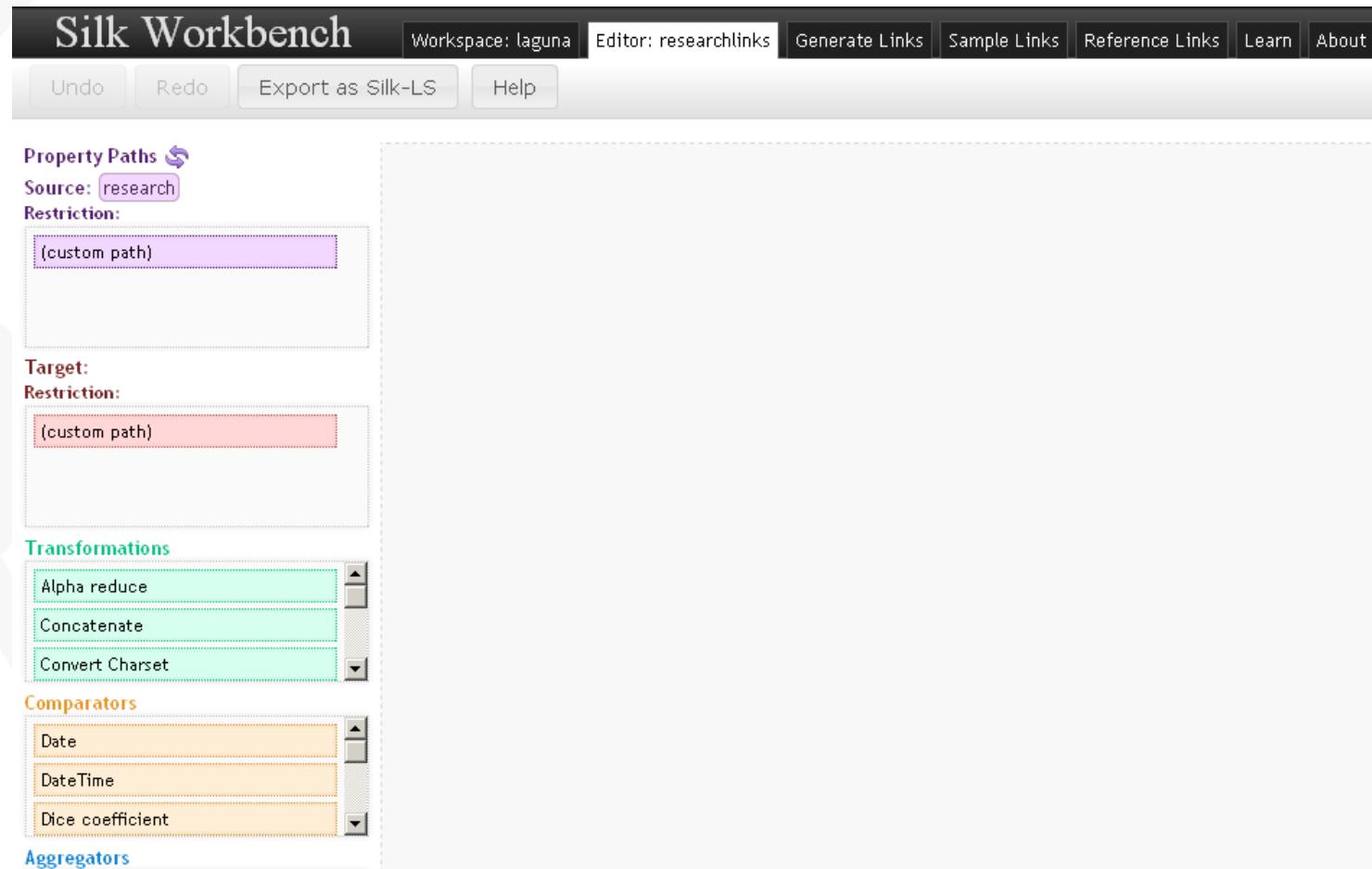
New linking task x

Name	researchlinks
Source	research ▼
Source restriction	?a rdf:type foaf:Person
Target	data-semanticweb ▼
Target restriction	?b rdf:type foaf:Person
Link type	http://www.w3.org/2002/07/owl#sameAs

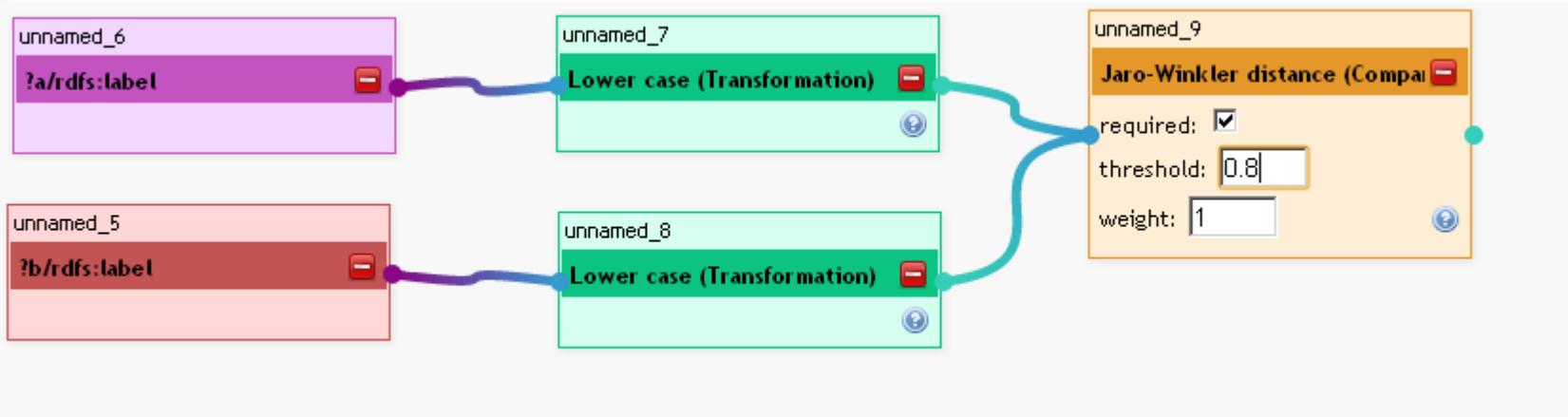
- Add output



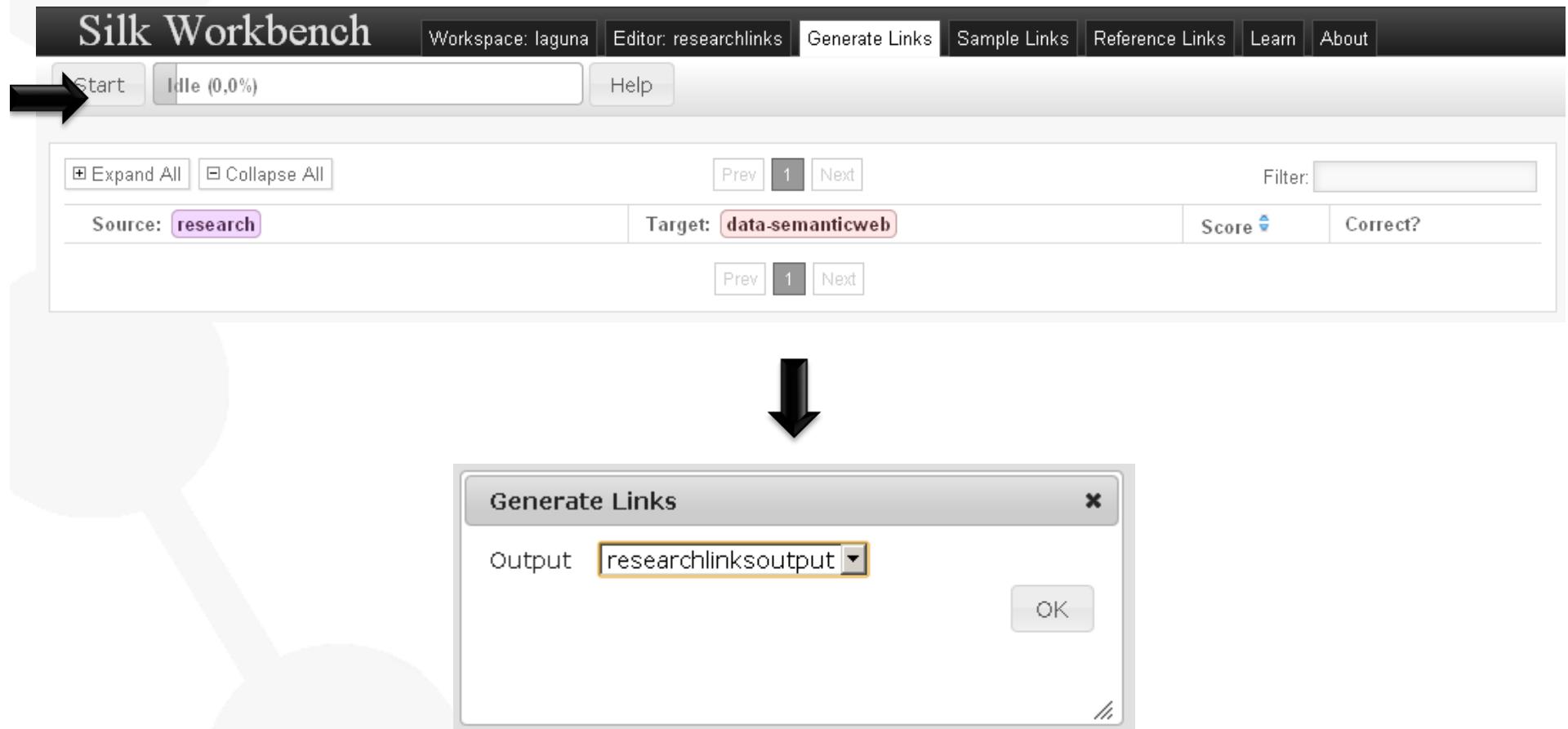
- Edit Linking task



- Edit researchlinks



- Generate links



- Validate links

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

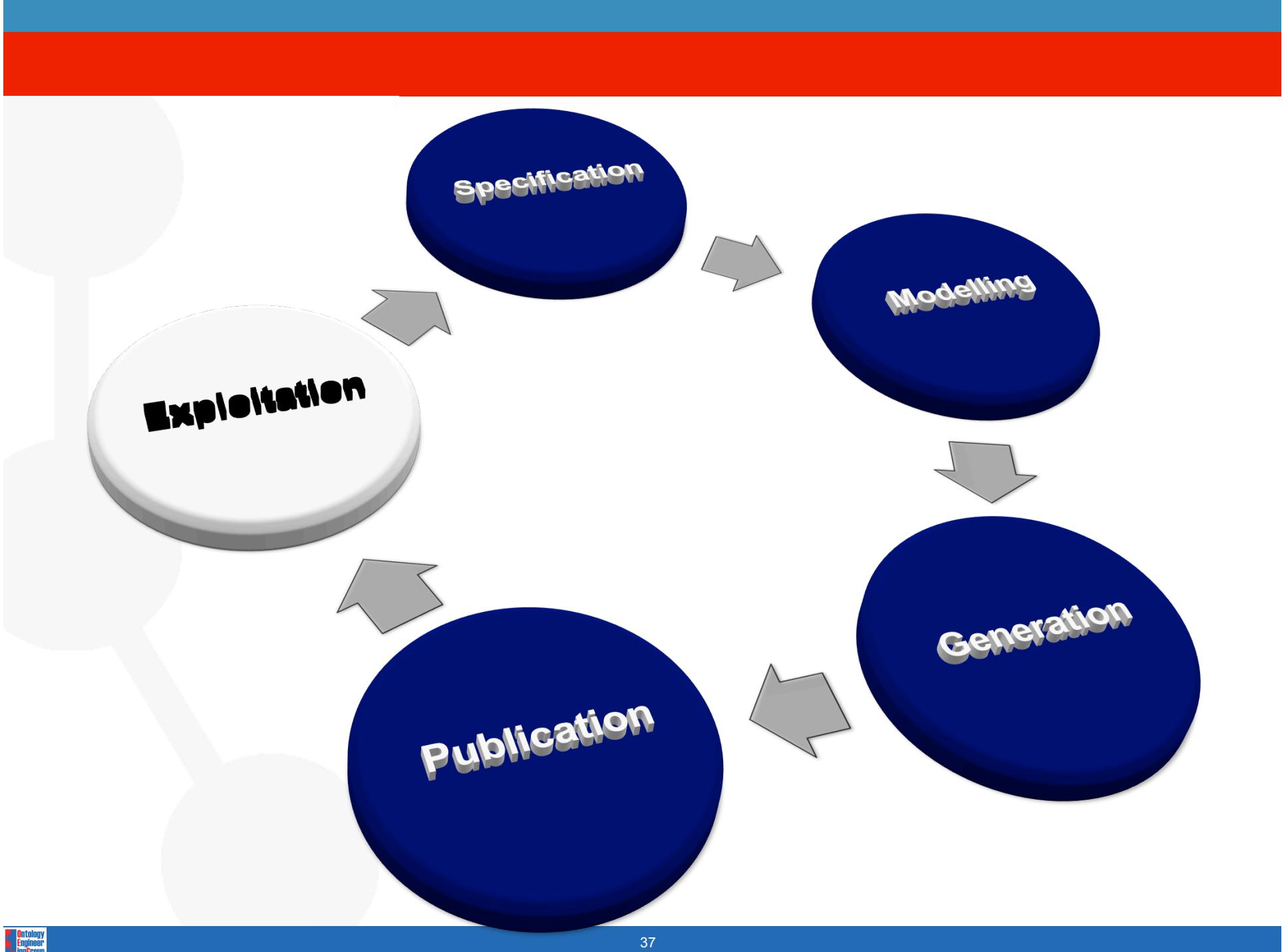
Start Generate Links Task finished in 21s (100,0%) Help

[+] 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%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/taowei-wang	25,9%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/takahira-yamaguchi	31,5%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stratos-idreos	35,3%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stijn-heymans	31,3%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stephan-bloehdorn	31,8%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/steffen-staab	31,3%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stefano-spaccapietra	43,5%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stefan-schllobach	32,7%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/stefan-decker	30,3%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/sofia-angeletou	40,3%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>
▶ esearch.ull.es/resource/Person/miguel-garcia-serrano-gonzalez	http://data.semanticweb.org/person/soeren-auer	47,0%	<input checked="" type="checkbox"/> <input type="button" value="?"/> <input type="button" value="X"/>

- 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>
            ▼<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>
```

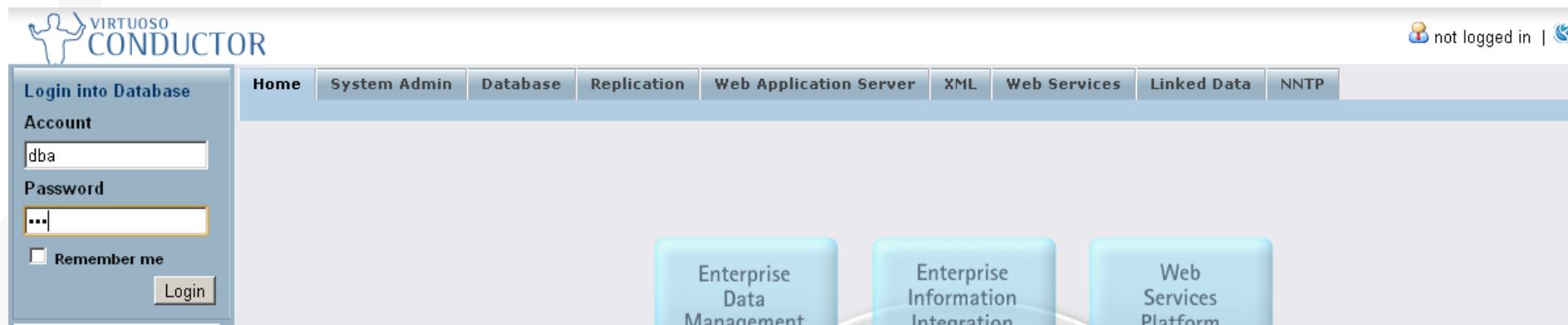


Publication - Virtuoso Open-source edition

- <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>



Virtuoso - Conductor

- Upload the generated files

VIRTUOSO CONDUCTOR

Home System Admin Database Replication Web Application Server XML Web Services Linked Data NNTP
SPARQL Sponger Statistics Graphs Schemas Namespaces Access Control Views Quad Store Upload

Quad Store Upload

File* Choose File No file chosen

Resource URL*

Create graph explicitly

Named Graph IRI* http://localhost:8890/graph/ontology

Cancel Upload

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

- <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](#))



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

Metadata publication – VOiD

- 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 wv: <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.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.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 AEMET metereological dataset

Url http://thedatahub.org/dataset/aemet

Author Ontology Engineering Group, Facultad de Informática, Universidad de Málaga

Home Groups

At links:geolinkeddata 51 Delete

Lic namespace http://aemet.linkeddata.es/resource/ Delete

Descri triples 21324148 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 homepage. At the top, there's a navigation bar with links for Home, About, Search, Submit, Forum, and Dev. Below the navigation is a search bar with tabs for Search, Submit, and Inspector Tool. The search field contains "Search: term property advanced SIGMA". A green button labeled "SEARCH" is next to it. Below the search bar, there's a message about consuming RDF and Microformats from web pages. A "LEARN MORE" button is present. On the left, there's a "LATEST DATA" section with a list of RDF triples from slideshare.net. In the center, there's a "HIGHLIGHTS" section featuring "Anything to Triples Any23" with a sub-section "Introduction". On the right, there's a "SINDICE TWEET" section with a tweet from June 21, 2010, and a "SINDICE BLOG" section with posts about downtime, Any23 releases, and a beta announcement.

Sindice - Data Web Services

Billion pieces of reusable information can already be found across hundreds of millions web pages which embed RDF and Microformats. Start consuming this data today with Sindice Data Web services.

LEARN MORE →

LATEST DATA

14:03:16 (sigma) 136 triples http://www.slideshare.net/andreluizbecker
14:03:15 (sigma) 136 triples http://www.slideshare.net/andreluizbecker
14:03:14 (sigma) 136 triples http://www.slideshare.net/andreluizbecker
14:03:13 (sigma) 136 triples http://www.slideshare.net/andreluizbecker

HIGHLIGHTS

Anything to Triples Any23

Introduction

Any23

SINDICE TWEET

Mon Jun 21 16:57:31 2010 Sindice at #semtech2010 , visit us at the DERI Booth Wed and Thurs. Plus see our presentation Thurs http://bit.ly/cKOVYa

SINDICE BLOG

Sindice planned downtime this weekend JUN 09, 2010

HiDue to an expansion of one of our datacentres (and the electrical work that this implies), Sindice and related services such as sig.ma w...
(More →)

Any23 v0.4.0 Released MAY 27, 2010

Dear All, the Sindice FBK team is proud to announce the Any23 0.4.0 release. (More →)

Any23 v0.3.0 Released APR 23, 2010

Dear All, we're pleased to announce the Any23 0.3.0 releasePlease keep in mind this is a beta, so everybody using Any23 is a developer. (More →)

Enable effective discovery

- 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 URLs
- 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 Sindice website's main navigation bar at the top, featuring links for Home, About, Search, Submit (highlighted in green), Forum, and Dev. Below the navigation, there's a logo for 'indice THE SEMANTIC WEB INDEX'. A section titled 'Share your Web Data presence with Sindice' contains text about pinging semantically enabled pages and guidelines for indexing datasets. To the right, there are two forms: a 'Ping Form' for individual pages and a 'Submit Sitemap form'. Both forms include fields for email and sitemap URLs, and a 'Submit' button.

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](#)

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

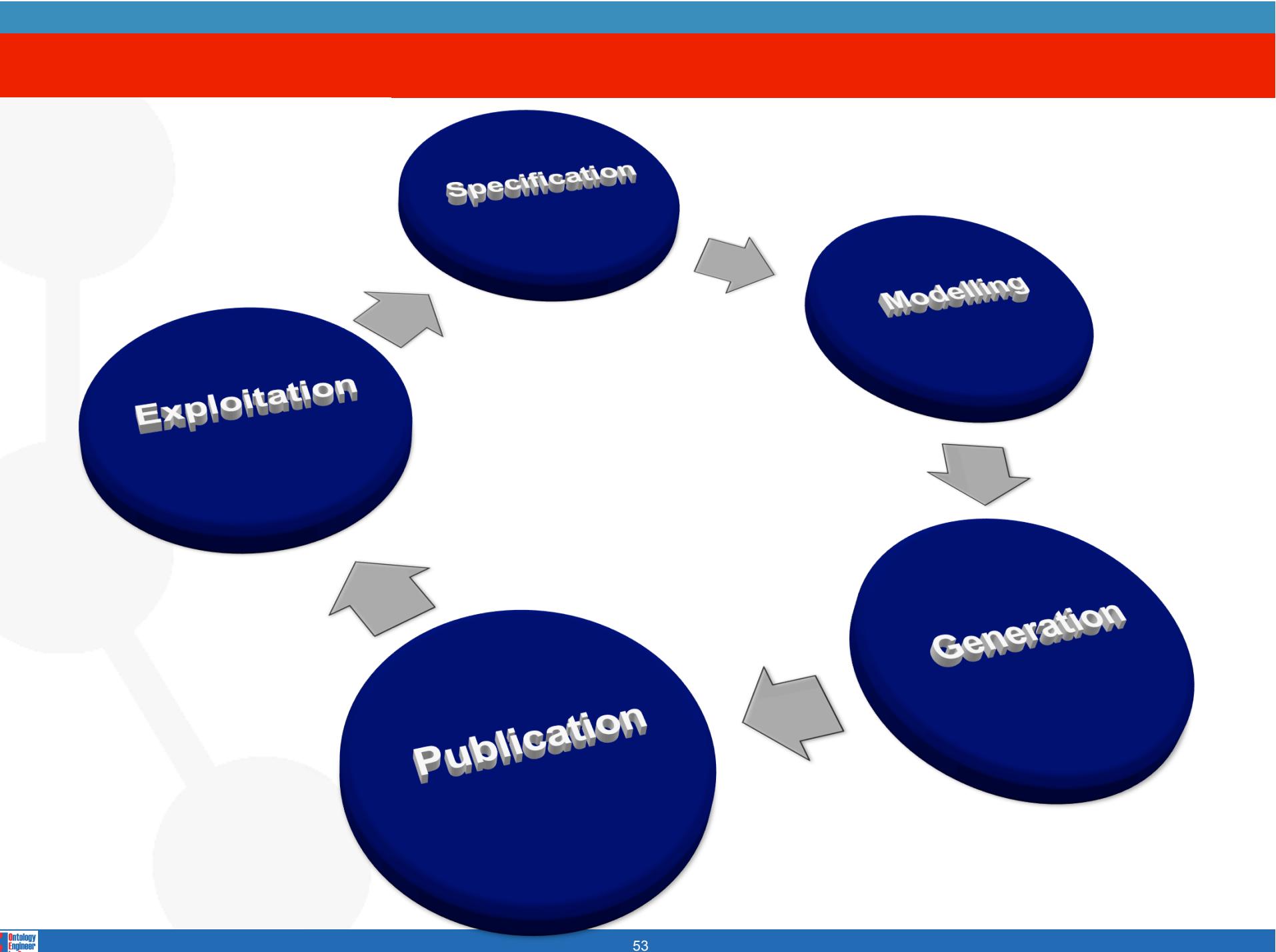
[Submit a Sitemap](#) Show submissions: [By me \(4\)](#) - [All \(4\)](#)

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

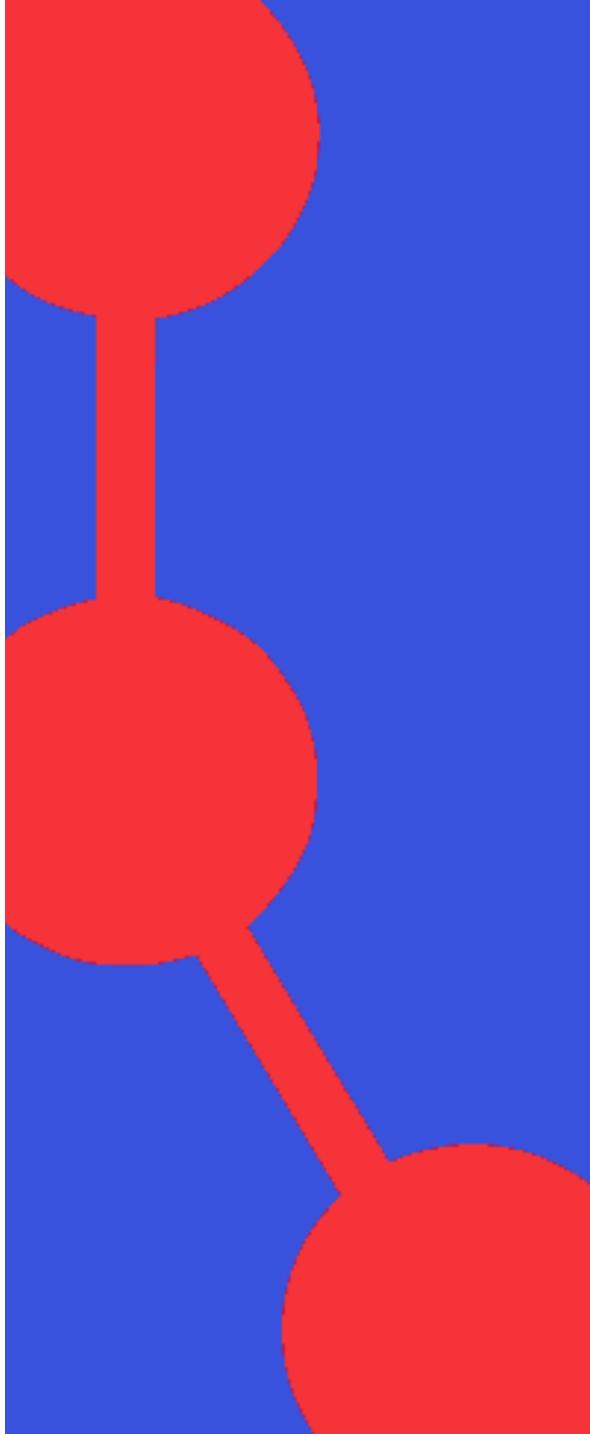
[Delete](#) [Resubmit](#)

[Download this table](#)
[Download data for all sites](#)

Help with:
[Creating and submitting Sitemaps](#)
[Sitemap errors](#)
[Sitemap details](#)
[Sitemaps for multiple sites](#)







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:

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0