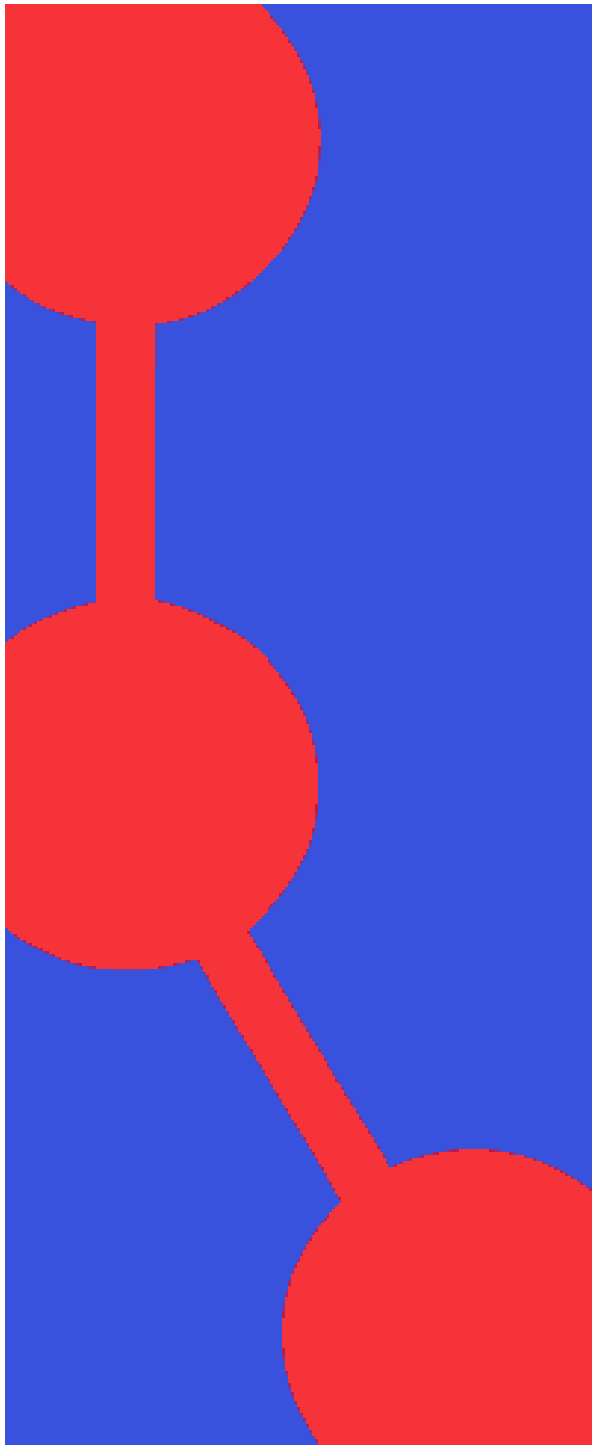


Geographical Linked Data: a Spanish Use Case

Alexander de León, Victor Saquicela, Luis M. Vilches, **Boris Villazón-Terrazas,**
Freddy Priyatna, Oscar Corcho, Carlos Buil, José Mora, and Jean-Paul Calbimonte
Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid
http://www.oeg-upm.net
Phone: 34.91.3366605, Fax: 34.91.3524819

ToC

- Data Integration Group
- Linked Data
- Motivation
- A Process for Publishing Linked Government Data
  - Identification of the data sources
  - Ontology modelling
  - Generation of the RDF data
  - Creation of RDF from geometrical information
  - Alignment of the datasets
  - Data publication and visualization
- Demo
- Future Work

Data Integration Group

- Oscar Corcho
  - Alexander de Leon
  - Victor Saquicela
  - Freddy Priyatna
  - Luis M. Vilches
  - Carlos Buil
  - Jean Paul Calbimonte
  - José Mora
  - Boris Villazón-Terrazas
- http://delicias.dia.fi.upm.es/wiki/index.php/Large-scale_data_integration
- dataintegration@delicias.dia.fi.upm.es

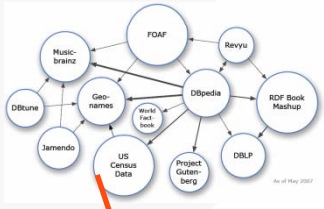
- Data Integration Group
- Linked Data
- Motivation
- A Process for Publishing Linked Government Data
  - Identification of the data sources
  - Ontology modelling
  - Generation of the RDF data
  - Creation of RDF from geometrical information
  - Alignment of the datasets
  - Data publication and visualization
- Demo
- Future Work

Linked Data

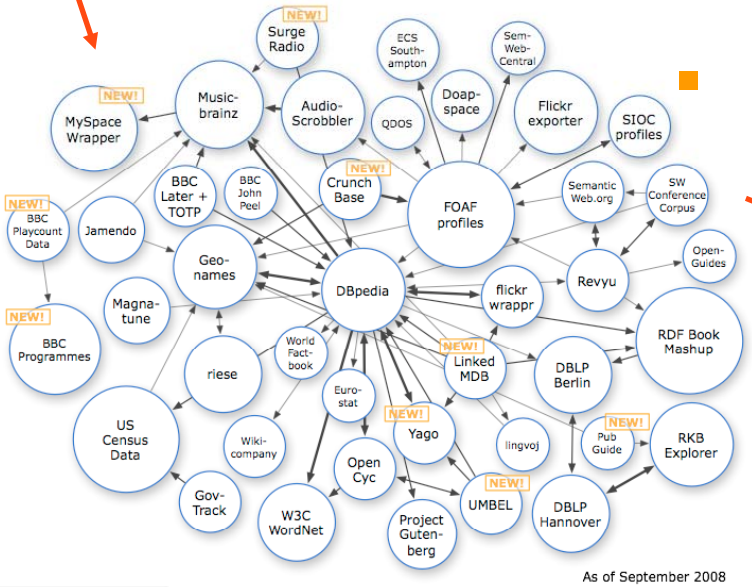
- Linked Data is a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data using URIs and RDF.

- Linked Data Principles
  1. Use URIs as names for things
  2. Use HTTP URIs so that people can look up those names.
  3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
  4. Include links to other URIs, so that they can discover more things.

Linked Open Data evolution

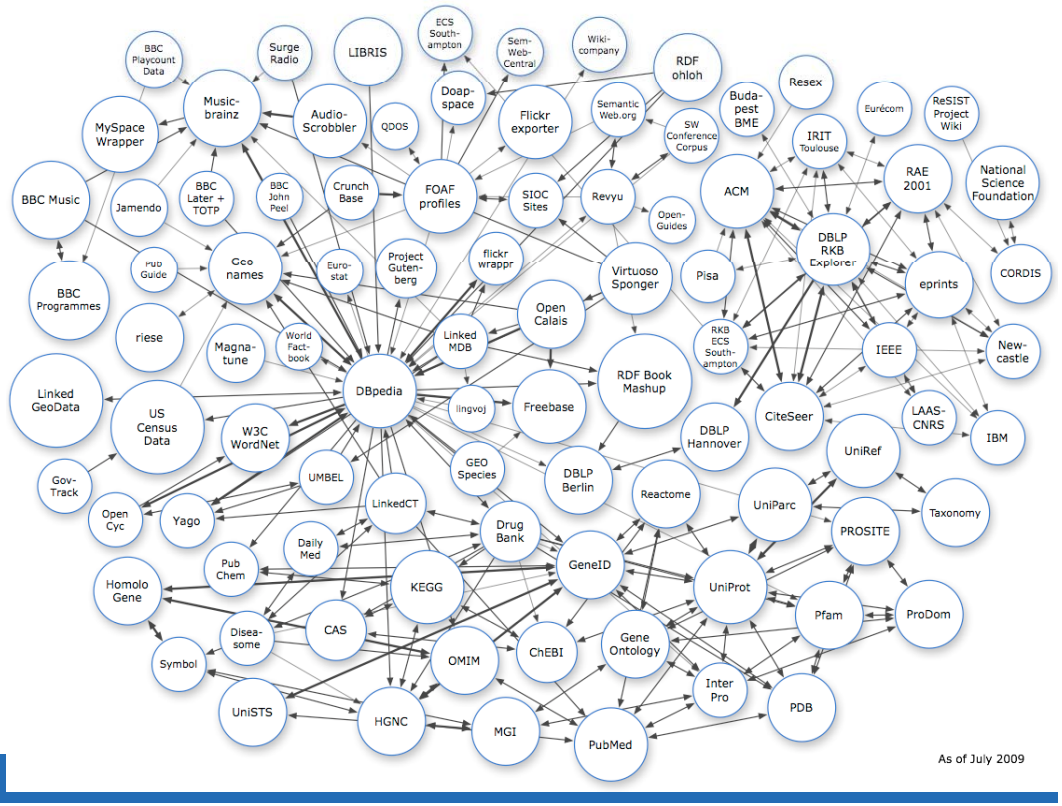


- 2007



As of September 2008

- 2008



As of July 2009

- 2009

- Data Integration Group
- Linked Data
- Motivation
- A Process for Publishing Linked Government Data
  - Identification of the data sources
  - Ontology modelling
  - Generation of the RDF data
  - Creation of RDF from geometrical information
  - Alignment of the datasets
  - Data publication and visualization
- Demo
- Future Work

So does that mean I have to publish my data as Linked Data, now?

- But, why?

- What was your incentive to publish an HTML page in 1990?
  - Share data in documents and because your neighbor was doing it

- So, why should we publish Linked Data in 2010?
  - Share data as data and because your neighbor is doing it

And guess who is starting to publish Linked Data now?

- UK Government
- US Government
- BBC
- Open Calais
- Freebase
- NY Times
- CNET
- Dbpedia
- ….

- **GeoLinkedData** is an open initiative whose aim is to enrich the Web of Data with Spanish geospatial data.
- It has started off by publishing diverse information sources, such as National Geographic Institute of Spain (IGN).

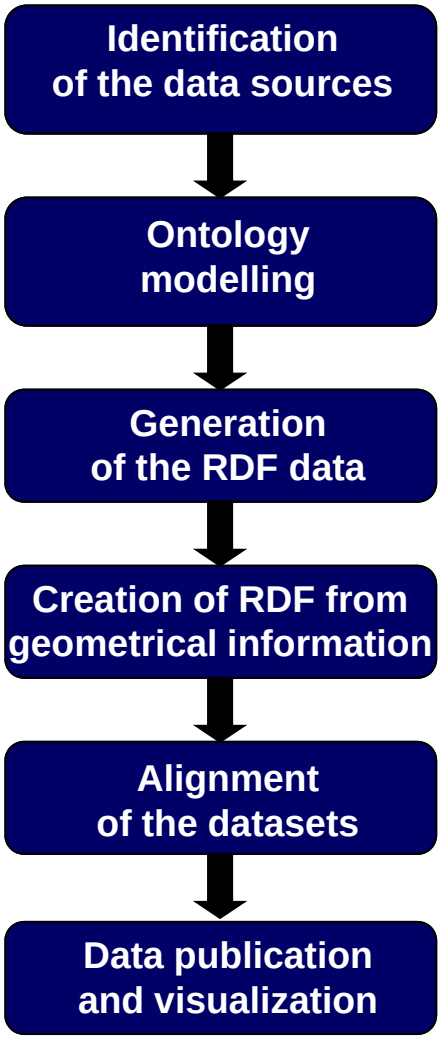
GeoLinked Data

GOBIERNO DE ESPAÑA MINISTERIO DE FOMENTO INSTITUTO GEOGRÁFICO NACIONAL

- http://geo.linkeddata.es
- Recently, National Statistics Institute (INE)

- Data Integration Group
- Linked Data
- Motivation
- A Process for Publishing Linked Government Data
  - Identification of the data sources
  - Ontology modelling
  - Generation of the RDF data
  - Creation of RDF from geometrical information
  - Alignment of the datasets
  - Data publication and visualization
- Demo
- Future Work

A Process for Publishing Linked Government Data

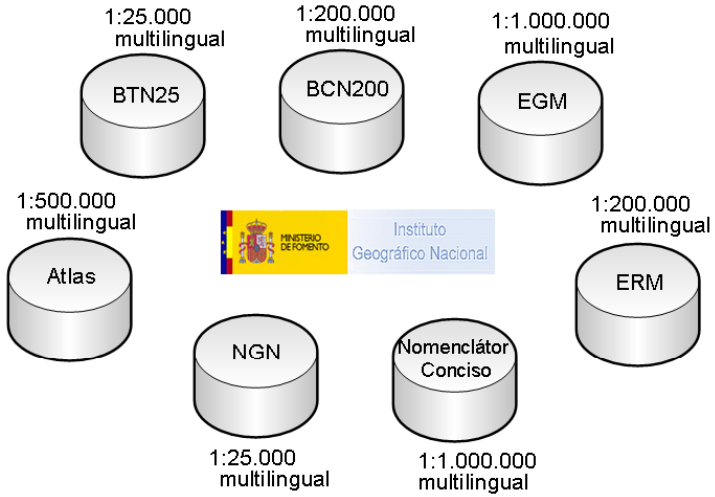


1. Identification of the data sources

IGN

National Geographic Institute of Spain

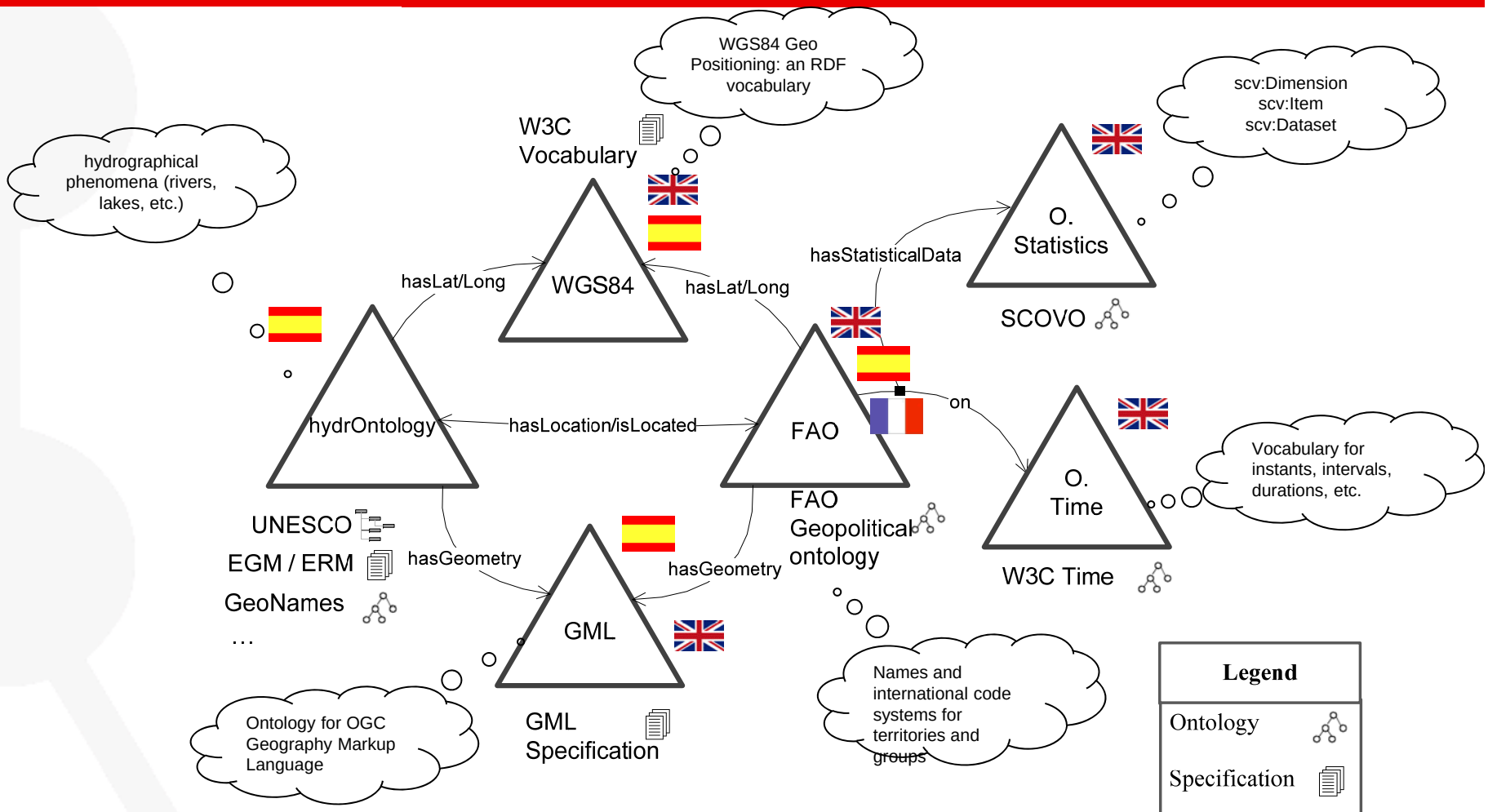
Oracle & MySQL



INE

National Statistic Institute of Spain

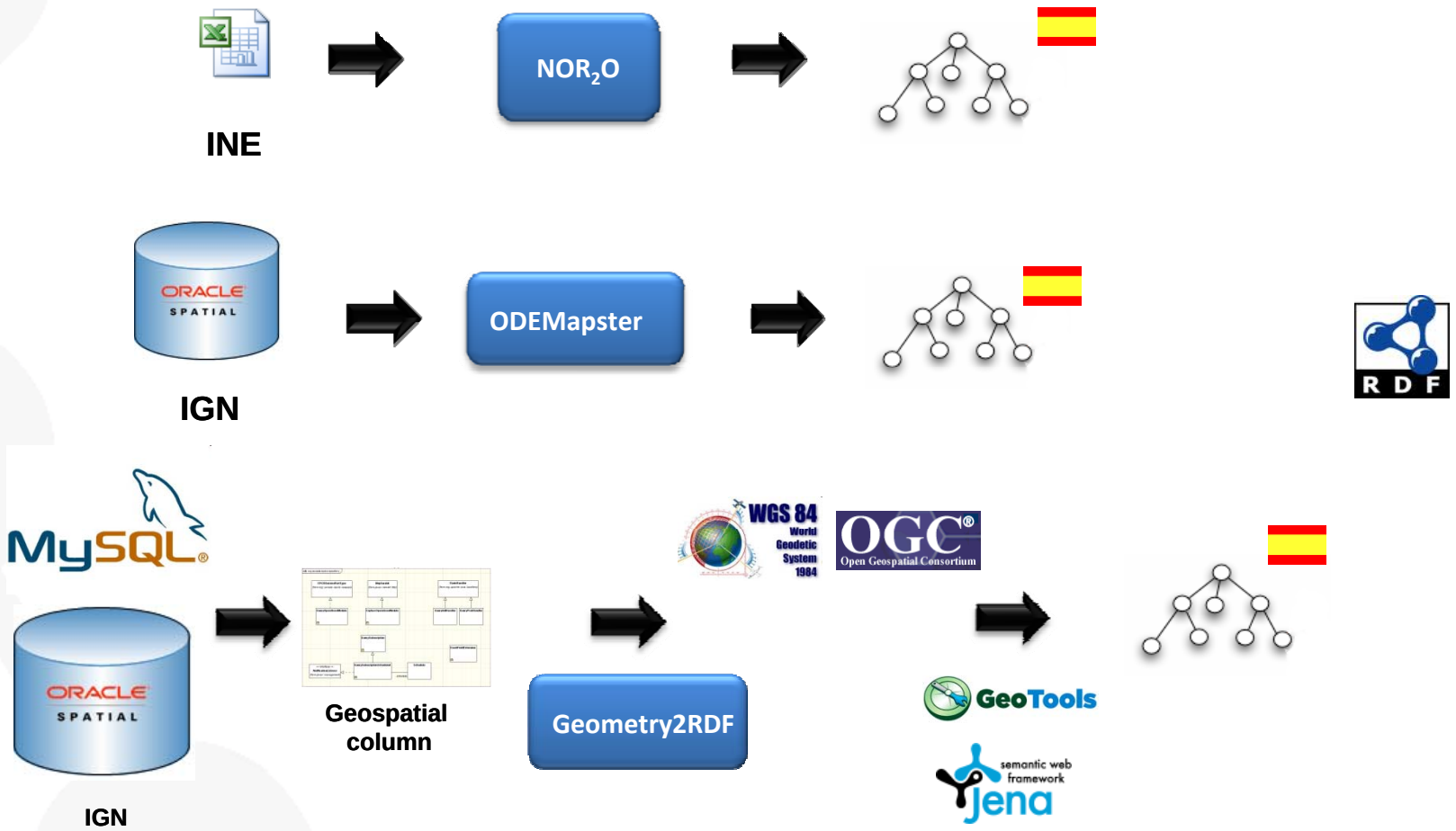
2. Ontology modelling



Classes	33	33
Object Properties	44	44
Data Properties	318	318

Following the INSPIRE
(INfrastructure for SPatial InfoRmation in Europe) recommendation.
hydrOntology,SCOVO, FAO Geopolitcal, WGS84, GML, and Time

3. Generation of the RDF Data



3. Generation of the RDF Data – NOR$_2$O

Industry Production Index

Year

Province

	2009	2008	2007	2006	2005	2004	2003
Total Nacional	3355830	3422239	3336657	3174393	3064129	2942583	2819159
Álava	21988	22316	20870		19638	19779	19638
Albacete	27380	27647	27068	25531	24385	23550	22547
Alicante	136239	142307	140145	133016	123333	113852	111805
Almería	43501	45490	43970	40871	38766	36260	33947
Asturias	71853	73124	72276	70115	68175	67038	65062
Ávila	11455	11708	11434	10900	10611	10319	10211
Badajoz	40874	41358	40168	38045	37052	34972	34866
Illes Balears	91824	93335	91254	88027	87024	85425	75951
Barcelona	467235	477942	469432	444410	436294	417425	397693
Burgos	25567	25891	25372	24504	23733	22882	22159
Cáceres	26307	26494	26064	25039	24846	20596	23140
Cádiz	62817	64505	63338	61691	58986	57138	54402
Cantabria	39611	40393	39560	37690	36561	35649	34017
Castellón	42122	43855	42476	39749	37865	37214	34211
Ciudad Real	32046	33011	31881	30446	29521	29011	26778
Córdoba	48979	50057	49302	47155	45405	43394	41964
Coruña, A	83748	84220	82873	79170	77023	74809	71748
Cuenca	14747	14928	14741	13822	13336	12829	12546
Girona	58404	51467	50108	47169	46827	45145	52482
Granada	60016	62269	61055	57223	54341	50508	49663
Guadalajara	13507	13735	12874	11825	10438	10120	9423
Guipúzcoa	52034	63569	59546	58486	57193	56498	55903
Huelva	26783	27463	27063	25487	24777	24270	22947
Huesca	17837	17109	16694	16025	15390	15078	14283
Jaén	36157	37368	36962	35383	34675	33157	32444
León	33564	34012	33563	32359	31664	30992	30256
Lleida	36920	37638	36065	33956	32739	31515	29605
Lugo	24861	25035	24609	23780	23122	22479	22396
Madrid	511804	519307	503000	478202	456175	436074	407655
Málaga	113362	116583	114547	108713	102382	96087	88257
Murcia	95636	100074	97374	90698	85110	82484	75973
Navarra	43282	43847	43142	41083	40739	39679	38936
Ourense	23304	23711	23526	22843	22452	22118	21560
Palencia	10964	11111	11060	10694	10575	10399	10297

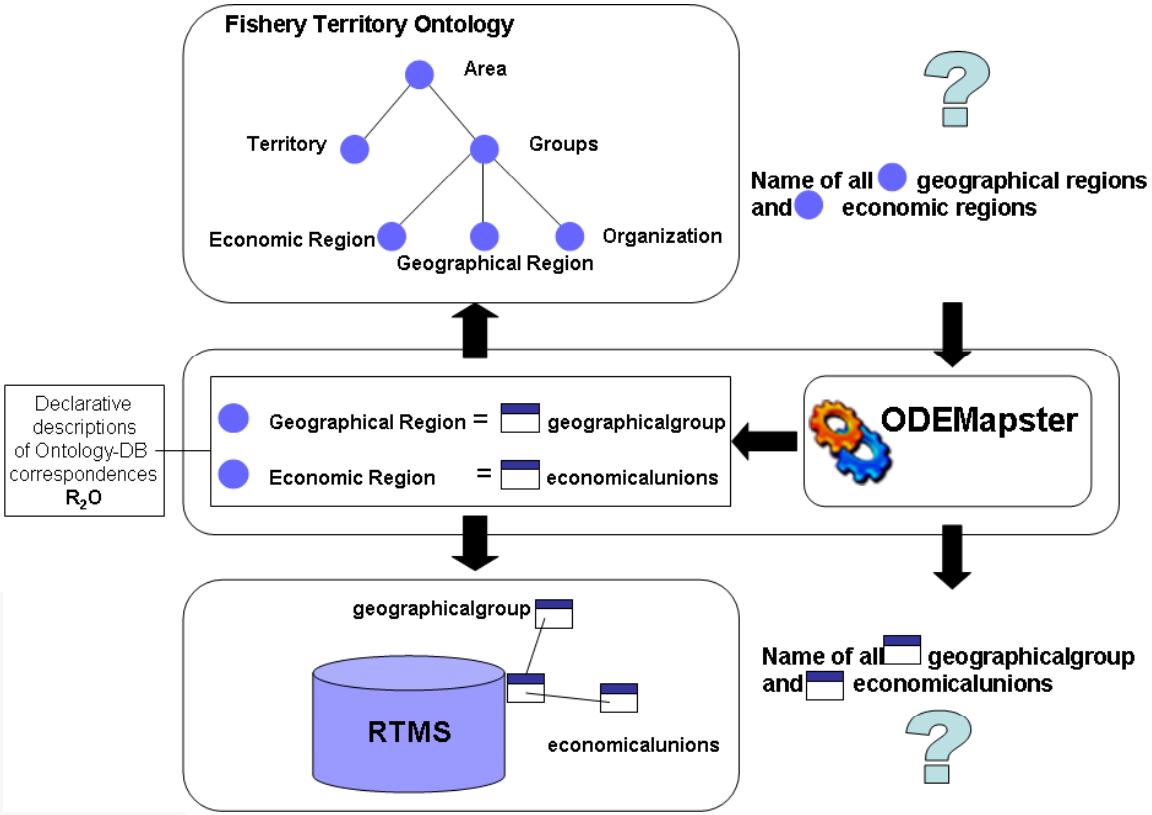
```
<?xml version="1.0" encoding="UTF-8"?>
<!-- <nor xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLo
<Nor type="GenericSpreadSheet" name="Spanish Employment">
  <Schema>
    <SchemaEntities>
      <SchemaEntity name="Location">
        <Attribute name="Name" valueFrom="pcaxis.[A11:A62]" type="string"/>
      </SchemaEntity>
      <SchemaEntity name="Dataset">
        <Attribute name="Name" valueFrom="pcaxis.[B6:H6]" type="string"/>
      </SchemaEntity>
      <SchemaEntity name="Year">
        <Attribute name="Name" valueFrom="pcaxis.[B9:H9]" type="string"/>
      </SchemaEntity>
      <SchemaEntity name="IndustryProductionIndex" type="Nary">
        <Attribute name="hasValue" valueFrom="pcaxis.[B10:H62]" type="string"/>
        <Relation name="inArea" usingSpreadSheetColumn="A" destination="Location"/>
        <Relation name="inPeriod" usingSpreadSheetRow="9" destination="Year"/>
        <Relation name="dataset" usingSpreadSheetRow="6" destination="Dataset"/>
      </SchemaEntity>
    </SchemaEntities>
  </Schema>
  <DataModel>
    <GenericDataModel/>
  </DataModel>
  <Implementation>
    <Spreadsheet type="ms" file="Empresas.xls"/>
  </Implementation>
</Nor>
```

NOR$_2$O

```
<!-- http://geo.linkeddata.es/resource/Estad%C3%ADstica/Madrid_ipi_2003 -->
<scv:Item rdf:about="Estad%C3%ADstica/Madrid_ipi_2003">
  <rdfs:label xml:lang="es">&#205;ndice de Producci&#243;n Industrial de Madrid en el a&#241;o 2003</rdfs:label>
  <rdf:value rdf:datatype="&xsd;double">407655.0</rdf:value>
  <scv:dimension rdf:resource="A%C3%B1o/2003"/>
  <scv:dataset rdf:resource="Estad%C3%ADstica/%25C3%258DndiceDeProducci%25C3%25B3nIndustrial"/>
  <onto:inProvincia rdf:resource="Provincia/Madrid"/>
</scv:Item>
<!-- http://geo.linkeddata.es/resource/Estad%C3%ADstica/Madrid_ipi_2004 -->
<scv:Item rdf:about="Estad%C3%ADstica/Madrid_ipi_2004">
  <rdfs:label xml:lang="es">&#205;ndice de Producci&#243;n Industrial de Madrid en el a&#241;o 2004</rdfs:label>
  <rdf:value rdf:datatype="&xsd;double">436074.0</rdf:value>
  <scv:dimension rdf:resource="A%C3%B1o/2004"/>
  <scv:dataset rdf:resource="Estad%C3%ADstica/%25C3%258DndiceDeProducci%25C3%25B3nIndustrial"/>
  <onto:inProvincia rdf:resource="Provincia/Madrid"/>
</scv:Item>
```

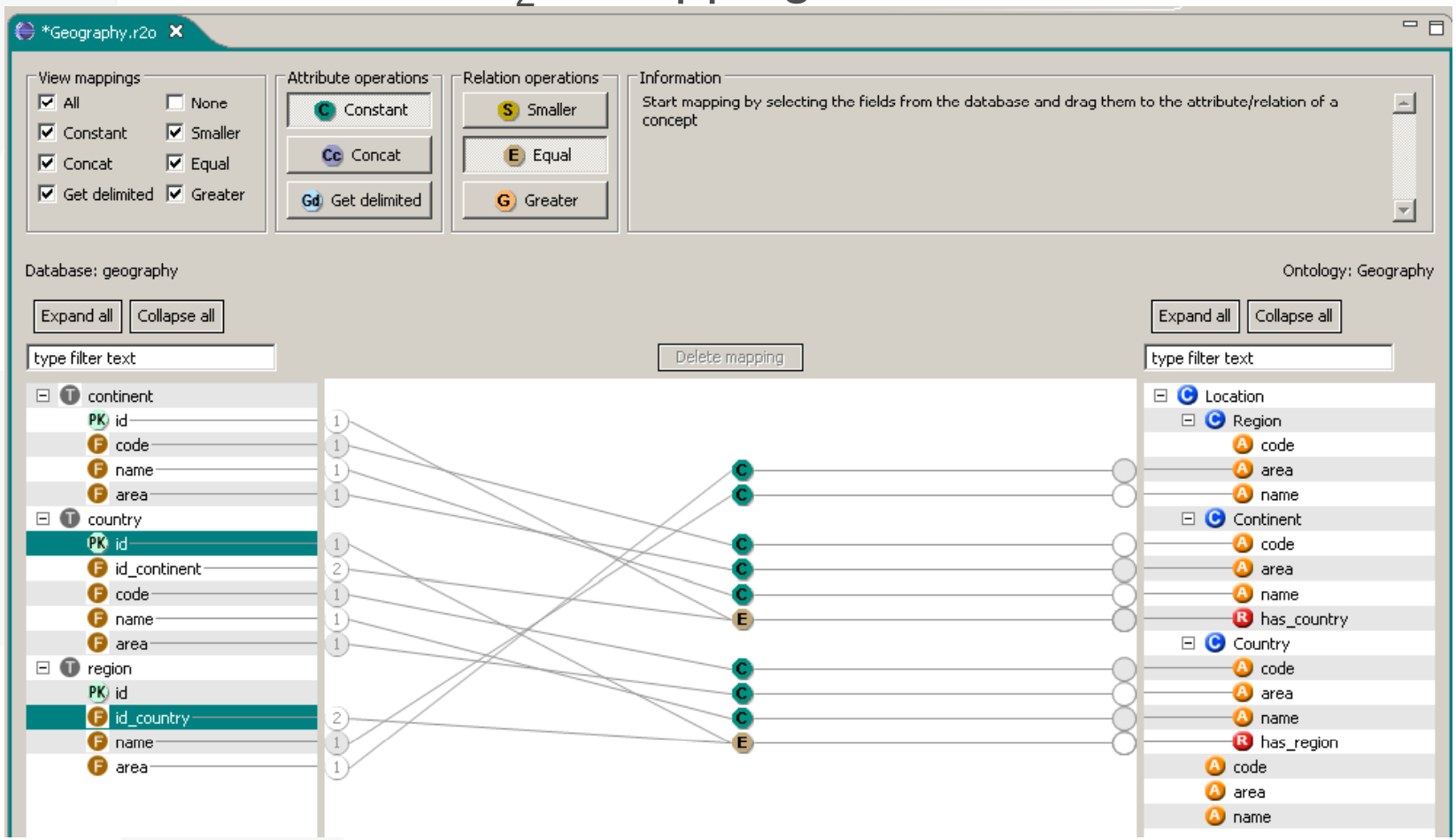
3. Generation of the RDF Data – R_2O & ODEMapster

- R_2O is an extensible, fully declarative language to describe mappings between relational database schemas and ontologies.
- The ODEMapster processor generates Semantic Web instances from relational instances based on the mapping description expressed in the R_2O document



3. Generation of the RDF Data – R_2O & ODEMapster

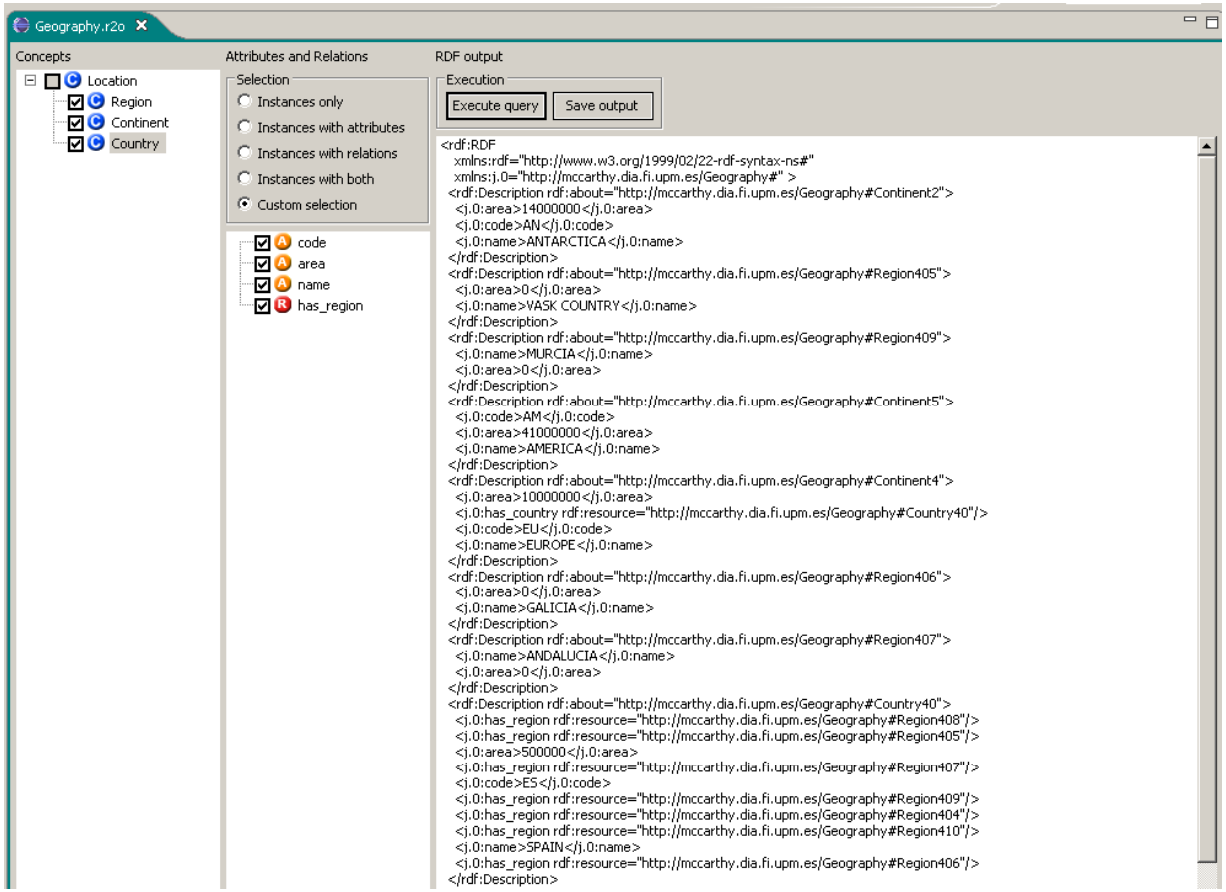
- Creation of the R_2O Mappings



3. Generation of the RDF Data – R_2O & ODEMapster

```
<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>
```

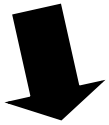
Excerpt of the R_2O document



4. Creation of RDF from geometrical information

NOMBRE	GMLGEOMETRY
Abengibre	(HUGECLOB)
Alatoz	(HUGECLOB)
Albatana	(HUGECLOB)
Balsa de Ves	(HUGECLOB)
Ballestero, El	(HUGECLOB)
Alborea	(HUGECLOB)
Alcadozo	(HUGECLOB)
Alcalá del Júcar	(HUGECLOB)

Oracle STO UTIL package

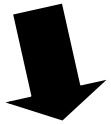


```
SELECT  TO_CHAR(SDO_UTIL.TO_GML311GEOMETRY(geometry))
                AS Gml311Geometry
FROM "BCN200"."BCN200_0301L_RIO" c
WHERE   c.Etiqueta='Arroyo'
```

```
<gml:Polygon srsName="SDO:8223" xmlns:gml="http://www.opengis.net/gml">
  <gml:outerBoundaryIs><gml:LinearRing><gml:coordinates decimal="." cs="," ts=" ">-1.48374108,39.23127677,0 -1.48404
  -1.48480765,39.22531695,0 -1.48544493,39.2231538,0 -1.4858725,39.22212319,0 -1.48608465,39.22141415,0 -1.4860584
  -1.48421128,39.21319056,0 -1.48390698,39.21064575,0 -1.48429166,39.2094524,0 -1.48529268,39.20490629,0 -1.492506
  -1.49590016,39.20383639,0 -1.49905691,39.20405732,0 -1.50256876,39.20446292,0 -1.50338941,39.20453654,0 -1.50483
  -1.51378498,39.2047147,0 -1.51398161,39.20472622,0 -1.5171434,39.20471246,0 -1.51894734,39.20373528,0 -1.5196271
  -1.52344981,39.20108038,0 -1.52699007,39.19901681,0 -1.5284903,39.19815278,0 -1.53151867,39.19619073,0 -1.532086
  -1.5329951,39.19479473,0 -1.53398595,39.19503246,0 -1.53537021,39.19528414,0 -1.53607317,39.19544615,0 -1.536914
  -1.53953498,39.19602121,0 -1.54099994,39.19629185,0 -1.54110394,39.19630216,0 -1.54231647,39.19646151,0 -1.54461
  -1.54957924,39.19633592,0 -1.55131402,39.1964386,0 -1.55382202,39.19666799,0 -1.55798564,39.19690882,0 -1.559800
  -1.56633376,39.19750743,0 -1.5703604,39.19765603,0 -1.57231502,39.19777011,0 -1.57278711,39.19790204,0 -1.573636
  -1.57633787,39.19982863,0 -1.57807687,39.20087703,0 -1.57814542,39.20092292,0 -1.57871557,39.20136237,0 -1.57915
  -1.5805024,39.20233924,0 -1.58088715,39.20278543,0 -1.58107708,39.20313014,0 -1.58122865,39.20365457,0 -1.581447
  -1.5819531,39.20563664,0 -1.58191475,39.20581637,0 -1.58194513,39.20603299,0 -1.58285235,39.20756676,0 -1.584270
```

4. Creation of RDF from geometrical information

```
<gml:Polygon srsName="SDO:8223" xmlns:gml="http://www.opengis.net/gml">
  <gml:outerBoundaryIs><gml:LinearRing><gml:coordinates decimal="." cs="," ts=" ">-1.48374108,39.23127677,0 -1.48404
  -1.48480765,39.22531695,0 -1.48544493,39.2231538,0 -1.4858725,39.22212319,0 -1.48608465,39.22141415,0 -1.4860584
  -1.48421128,39.21319056,0 -1.48390698,39.21064575,0 -1.48429166,39.2094524,0 -1.48529268,39.20490629,0 -1.492506
  -1.49590016,39.20383639,0 -1.49905691,39.20405732,0 -1.50256876,39.20446292,0 -1.50338941,39.20453654,0 -1.50483
  -1.51378498,39.2047147,0 -1.51398161,39.20472622,0 -1.5171434,39.20471246,0 -1.51894734,39.20373528,0 -1.5196271
  -1.52344981,39.20108038,0 -1.52699007,39.19901681,0 -1.5284903,39.19815278,0 -1.53151867,39.19619073,0 -1.532086
  -1.5329951,39.19479473,0 -1.53398595,39.19503246,0 -1.53537021,39.19528414,0 -1.53607317,39.19544615,0 -1.536914
  -1.53953498,39.19602121,0 -1.54099994,39.19629185,0 -1.54110394,39.19630216,0 -1.54231647,39.19646151,0 -1.54461
  -1.54957924,39.19633592,0 -1.55131402,39.1964386,0 -1.55382202,39.19666799,0 -1.55798564,39.19690882,0 -1.559800
  -1.56633376,39.19750743,0 -1.5703604,39.19765603,0 -1.57231502,39.19777011,0 -1.57278711,39.19790204,0 -1.573636
  -1.57633787,39.19982863,0 -1.57807687,39.20087703,0 -1.57814542,39.20092292,0 -1.57871557,39.20136237,0 -1.57915
  -1.5805024,39.20233924,0 -1.58088715,39.20278543,0 -1.58107708,39.20313014,0 -1.58122865,39.20365457,0 -1.581447
  -1.5819531,39.20563664,0 -1.58191475,39.20581637,0 -1.58194513,39.20603299,0 -1.58285235,39.20756676,0 -1.584270
```

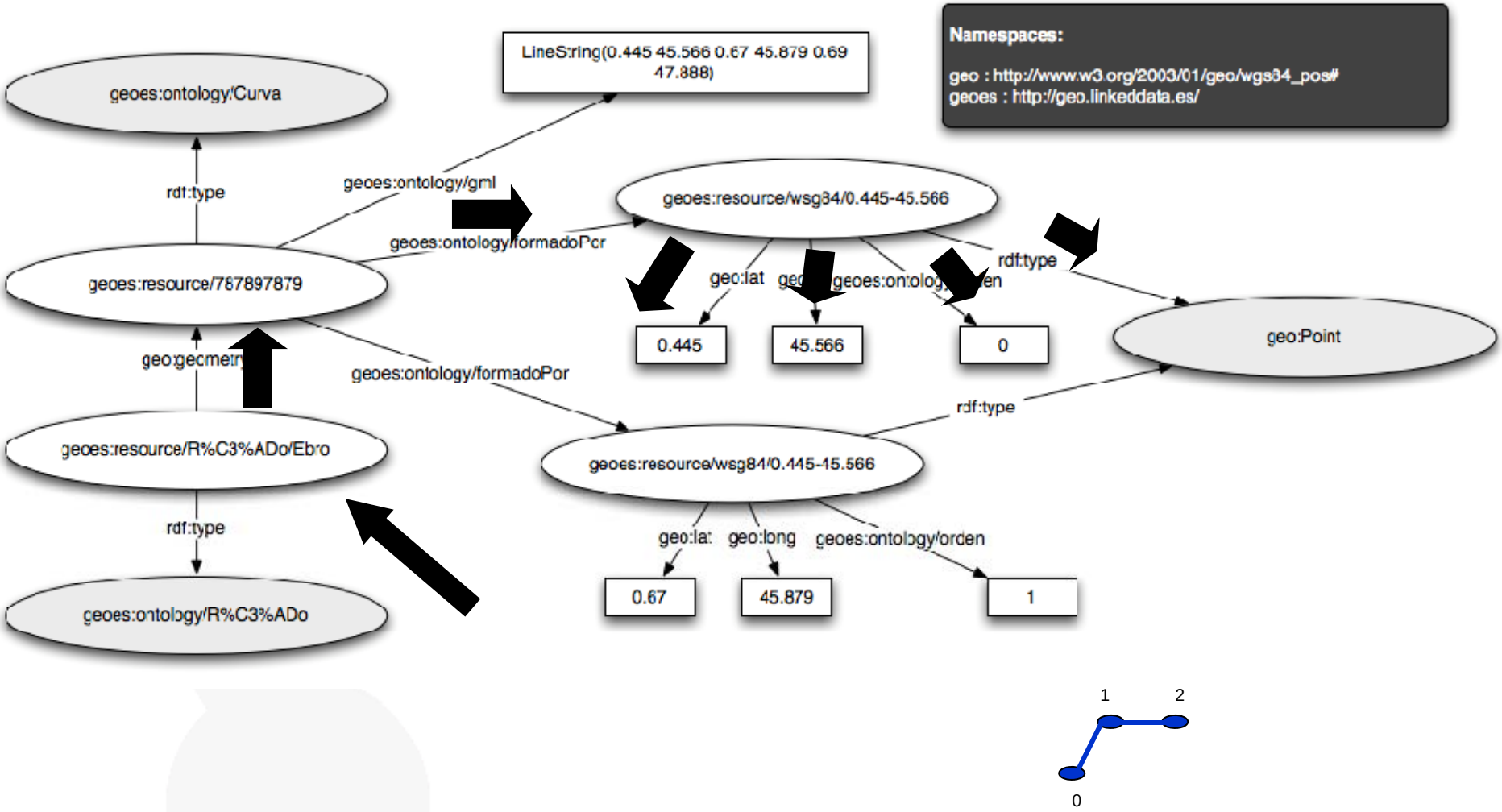


GeoTools

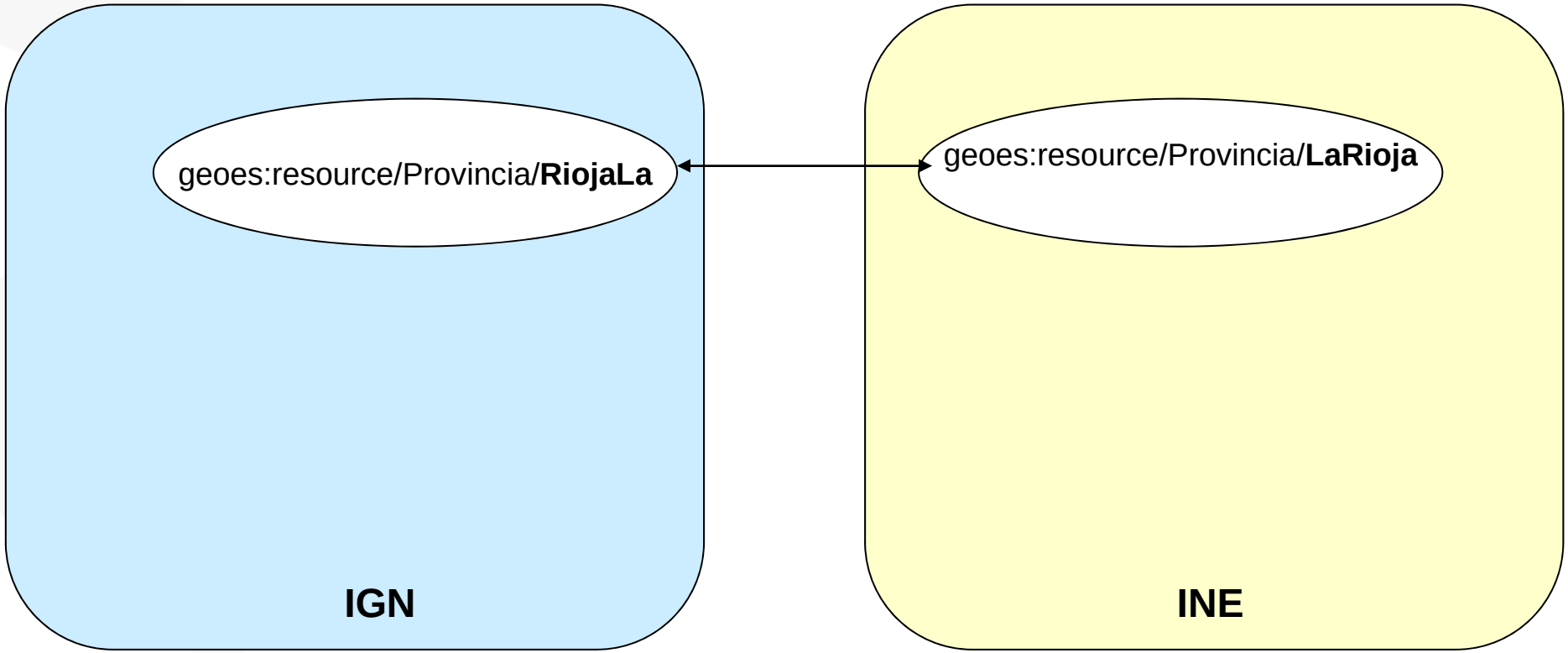
semantic web framework Jena

```
<rdf:Description rdf:about="http://geo.linkeddata.es/resource/wgs84/39.158523176194414_-1.6341427210305737">
  <geontology:order rdf:datatype="http://www.w3.org/2001/XMLSchema#int">205</geontology:order>
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-1.6341427210305737</geo:long>
  <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">39.158523176194414</geo:lat>
  <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#Point"/>
</rdf:Description>
<rdf:Description rdf:about="http://geo.linkeddata.es/resource/wgs84/38.65501872007066_-2.6158910085778153">
  <geontology:order rdf:datatype="http://www.w3.org/2001/XMLSchema#int">6</geontology:order>
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-2.6158910085778153</geo:long>
  <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">38.65501872007066</geo:lat>
  <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#Point"/>
</rdf:Description>
<rdf:Description rdf:about="http://geo.linkeddata.es/resource/wgs84/38.37015012539368_-1.6964507982671462">
  <geontology:order rdf:datatype="http://www.w3.org/2001/XMLSchema#int">141</geontology:order>
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-1.6964507982671462</geo:long>
  <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">38.37015012539368</geo:lat>
  <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#Point"/>
</rdf:Description>
```

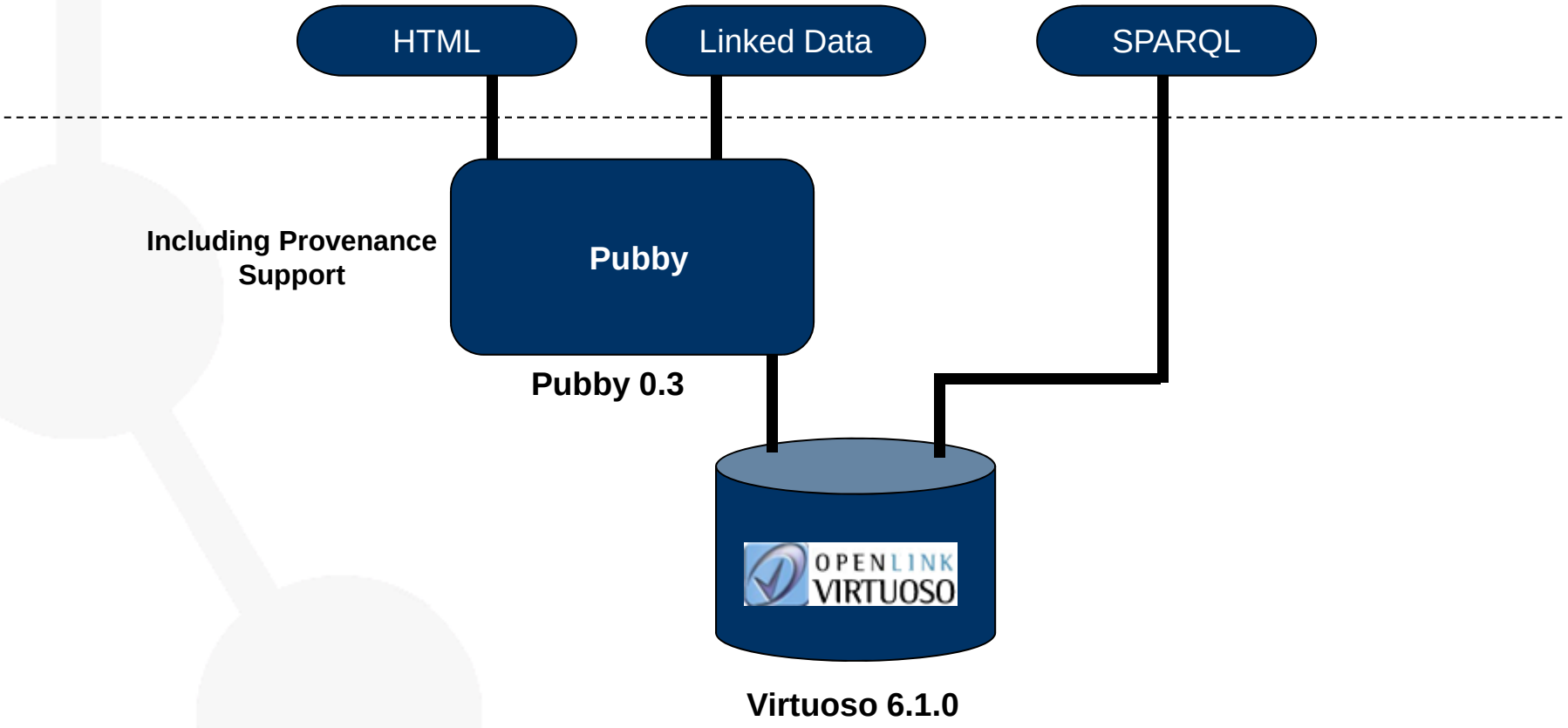
4. Creation of RDF from geometrical information



5. Alignment of the datasets



6. Data publication and visualization



http://www4.wiwiss.fu-berlin.de/pubby/

6. Data publication and visualization

Granada at geo.linkeddata.es

http://geo.linkeddata.es/resource/ProvinciaGranada

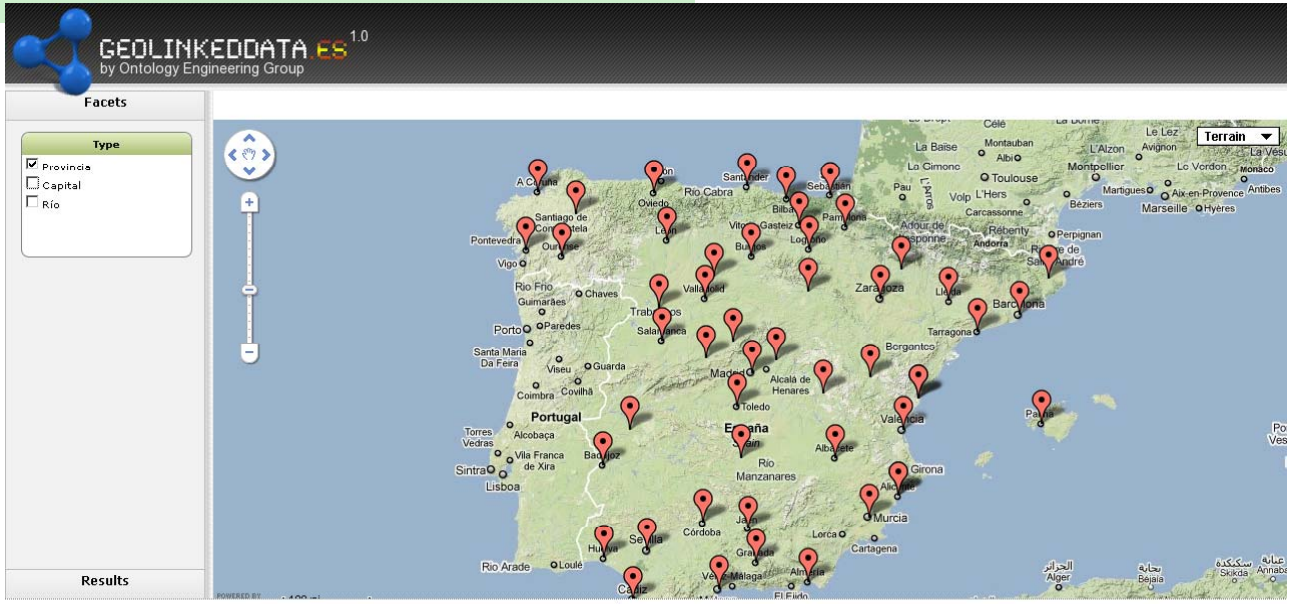
Property	Value
is geoes:esCapitalDe of	<http://geo.linkeddata.es/resource/CapitalGranada>
is geoes:formadaPor of	<http://geo.linkeddata.es/resource/Comunidad_AutonomaAndaluc%C3%ADa>
geo:geometry	<http://geo.linkeddata.es/wgs84/37.1793252820624_-3.59704135514161>
rdfs:label	Granada (xsd:string)
geoes:perteneceA	<http://geo.linkeddata.es/resource/Comunidad_AutonomaAndaluc%C3%ADa>
geoes:tieneCapital	<http://geo.linkeddata.es/resource/CapitalGranada>
rdf:type	geoes:Provincia

Metadata

<http://geo.linkeddata.es/data/ProvinciaGranada>

rdf:type	<http://www.w3.org/2004/03/trix/rdfg-1/Graph>
foaf:primaryTopic	<http://geo.linkeddata.es/resource/ProvinciaGranada>
dcterms:creator	<http://geo.linkeddata.es/provider/ign-e> (more)
dcterms:publisher	<http://geo.linkeddata.es>
dc:rights	La información geográfica digital comprendida en el Equipamiento Geográfico de Referencia Nacional (artículo 1.1 de la Orden FOM/956/2008) así como los Metadatos de los datos geográficos y servicios del IGN-CNIG, no requieren la aceptación de licencia y su uso será, en cualquier caso, libre y gratuito, siempre que se mencione al Instituto Geográfico Nacional como propietario de los datos.
dcterms:spatial	<http://geo.linkeddata.es/resource/Espa%C3%B1a>
prv:createdBy	Anon_0 (more)

expand all



ToC

- Data Integration Group
- Linked Data
- Motivation
- A Process for Publishing Linked Government Data
  - Identification of the data sources
  - Ontology modelling
  - Generation of the RDF data
  - Creation of RDF from geometrical information
  - Alignment of the datasets
  - Data publication and visualization
- Demo
- Future Work

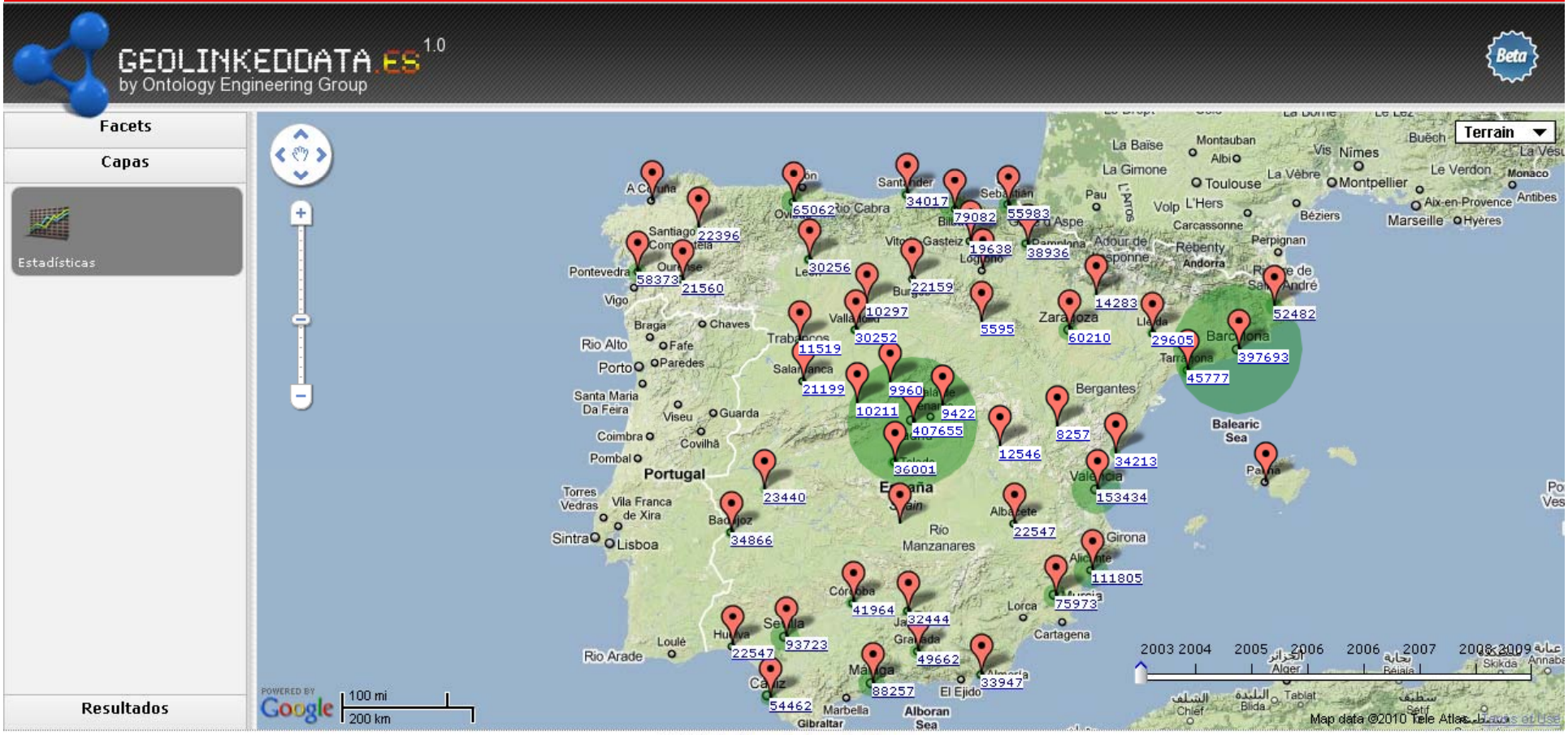
DEMO

http://geo.linkeddata.es/browser

Provinces

Capital of Province

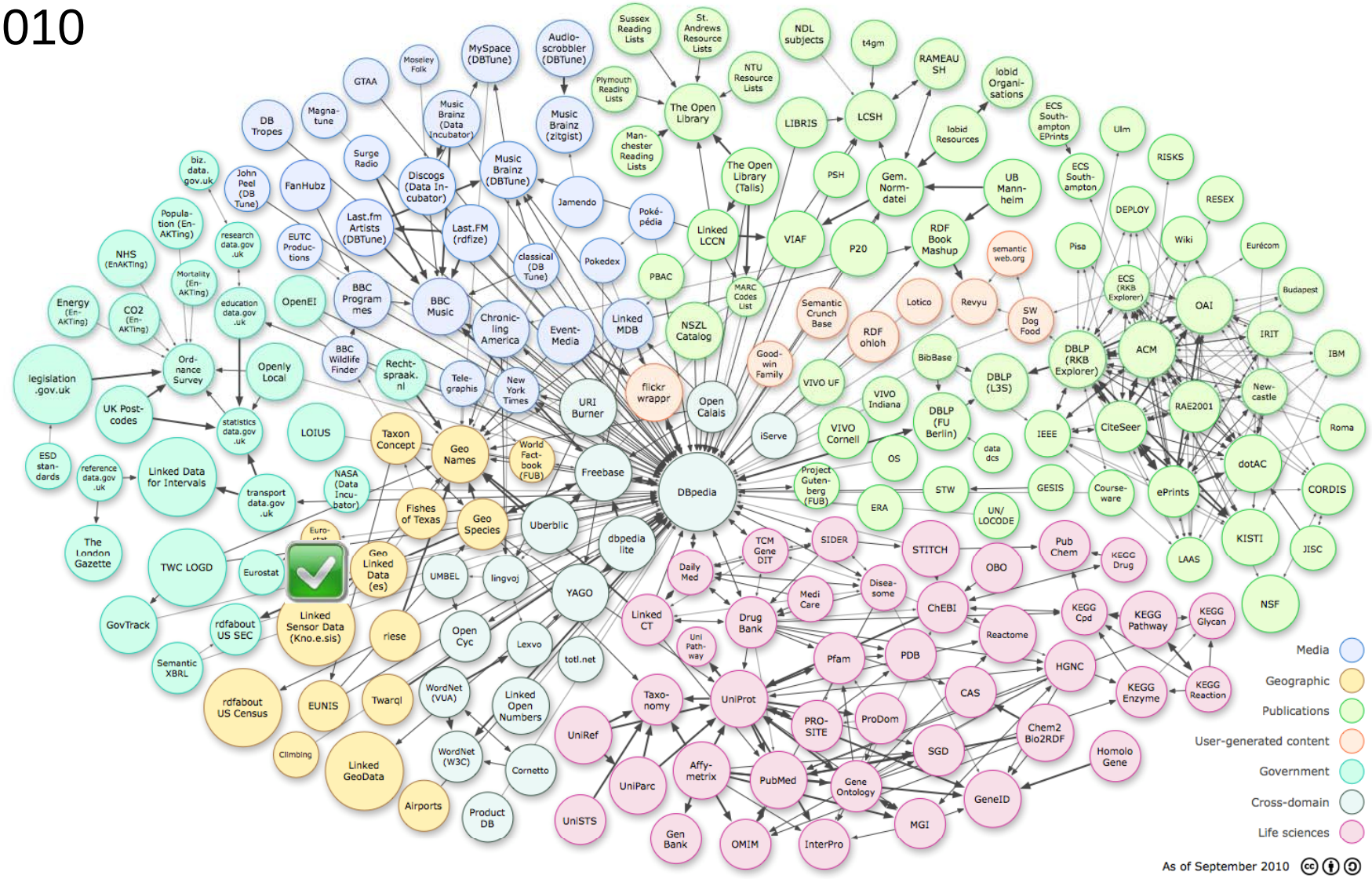
Provinces – Industry Production Index



Beaches

LOD datasets

2010



http://richard.cyganiak.de/2007/10/lod/

- Data Integration Group
- Linked Data
- Motivation
- A Process for Publishing Linked Government Data
  - Identification of the data sources
  - Ontology modelling
  - Generation of the RDF data
  - Creation of RDF from geometrical information
  - Alignment of the datasets
  - Data publication and visualization
- Demo
- Future Work

Future Work

- Data cleansing
- Interlinking with DBpedia and Geonames
- Include other domains
- Improve faceted browser
- Cover complex geometrical information
  - Point
  - LineString
  - Polygons

What we have learned

- Team work can get good results.
- There's no "i" in teamwork.
- Think big, Semantic Web Challenge?

Thank you and ...

coming soon

How to get your data into Sindice and Google with sitemap4rdf

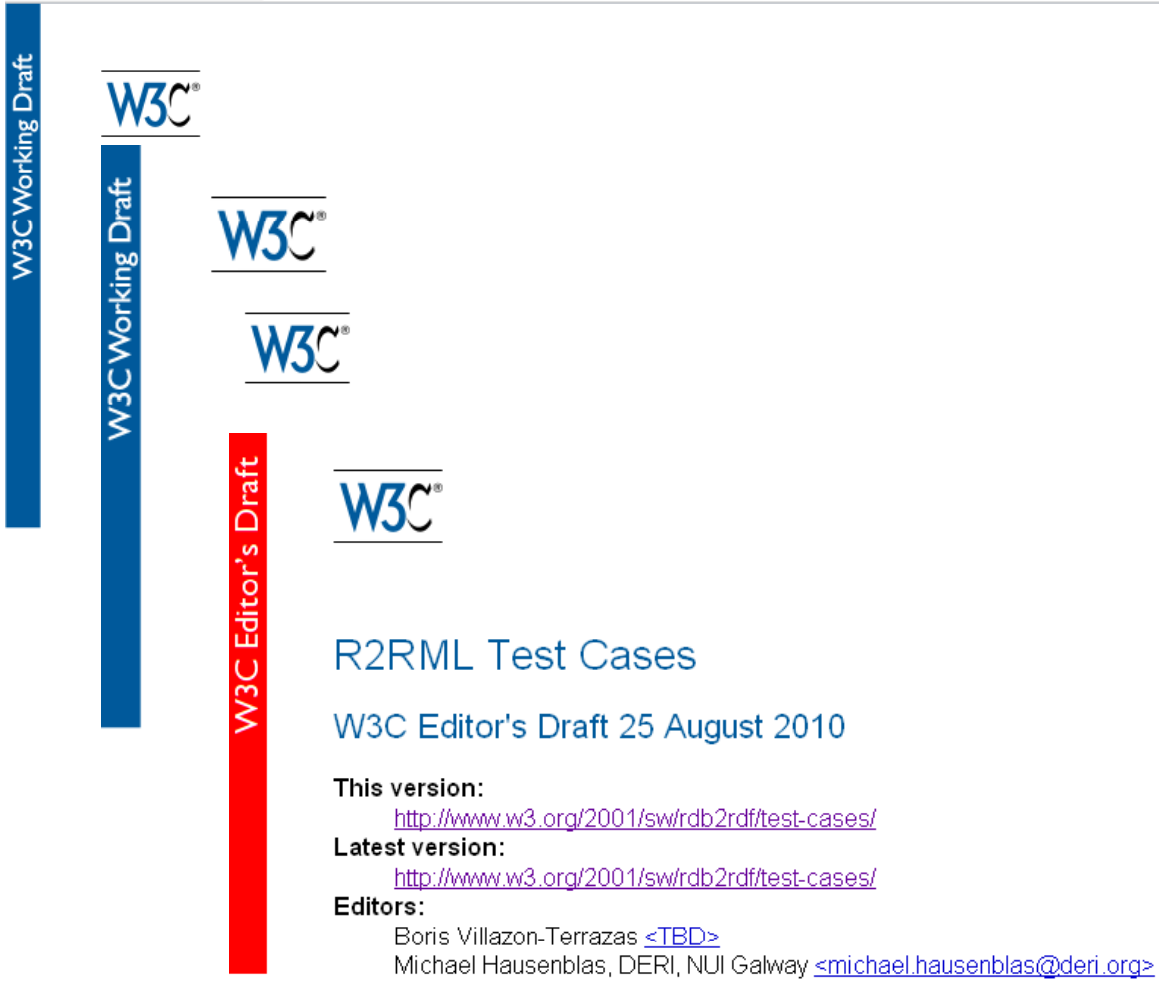
Boris Villazón-Terrazas (OEG), Richard Cyganiak (DERI)

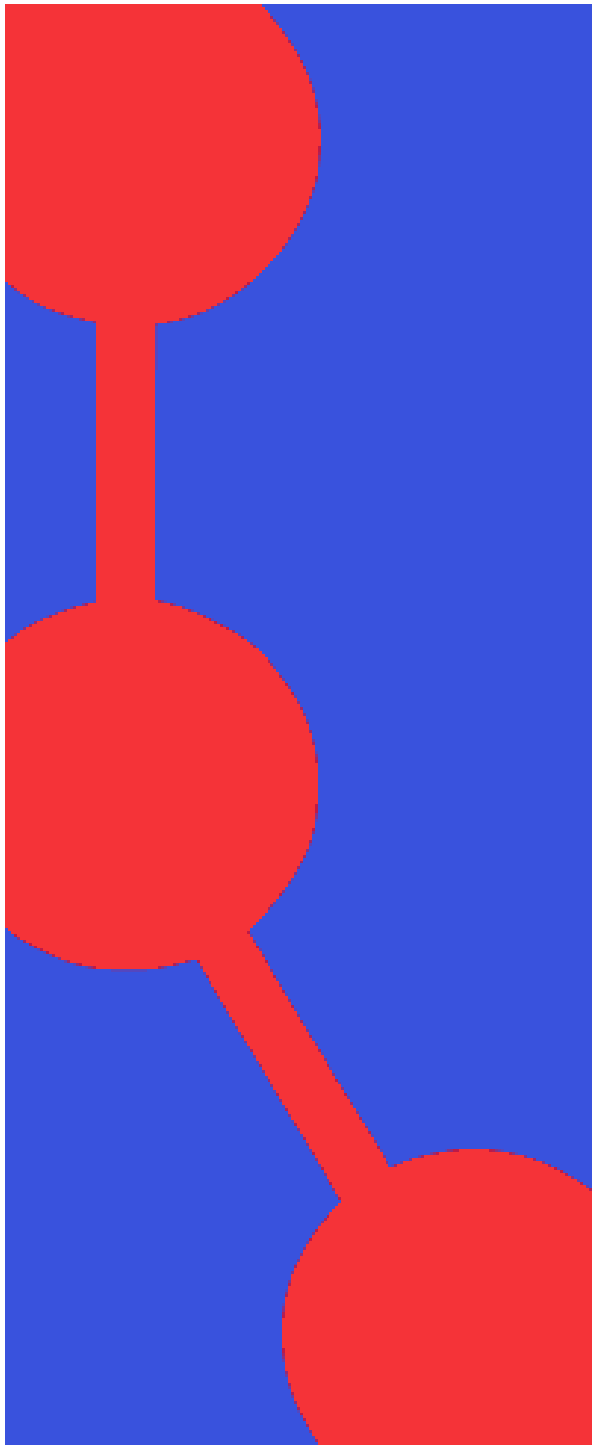
Dataset Dynamics Compendium: A Comparative Study

Jürgen Umbrich (DERI), Boris Villazón-Terrazas (UPM),
Michael Hausenblas (DERI)

Mapping Relational Data to RDF (RDB2RDF)

- **Status of the W3C RDB2RDF work in November 2010**
  - http://www.w3.org/2001/sw/rdb2rdf/status/2010/





Geographical Linked Data: a Spanish Use Case

Alexander de León, Victor Saquicela, Luis M. Vilches, **Boris Villazón-Terrazas,**
Freddy Priyatna, Oscar Corcho, Carlos Buil, José Mora, and Jean-Paul Calbimonte
Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid
http://www.oeg-upm.net
Phone: 34.91.3366605, Fax: 34.91.3524819

but first …

- PhD Thesis