



Publishing Linked Data from spreadsheets

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>

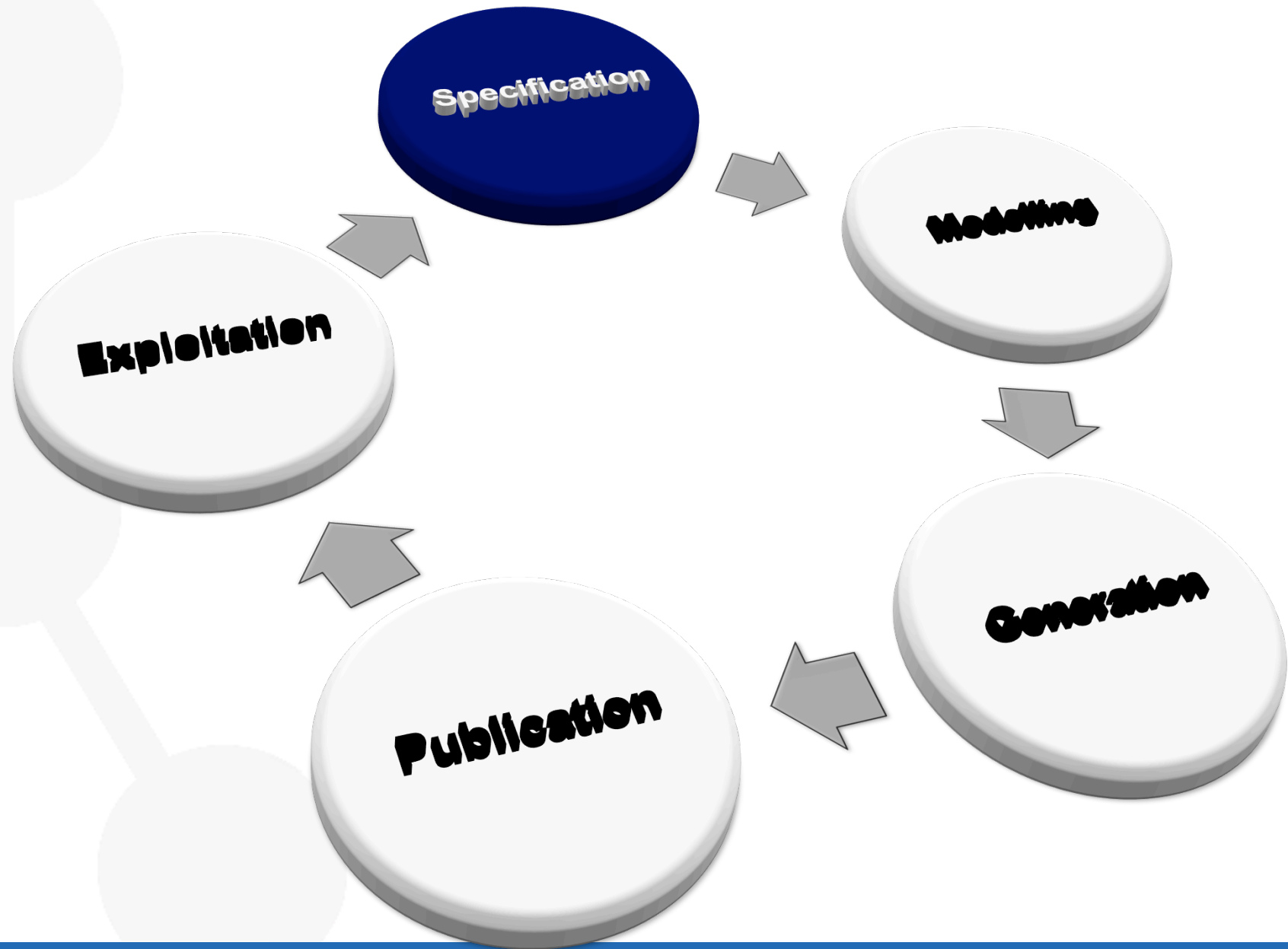
bvillazon@fi.upm.es

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 – Spreadsheet about statistics

- Unemployment

Mercado laboral

Paro registrado

Paro registrado por CCAA y periodo.

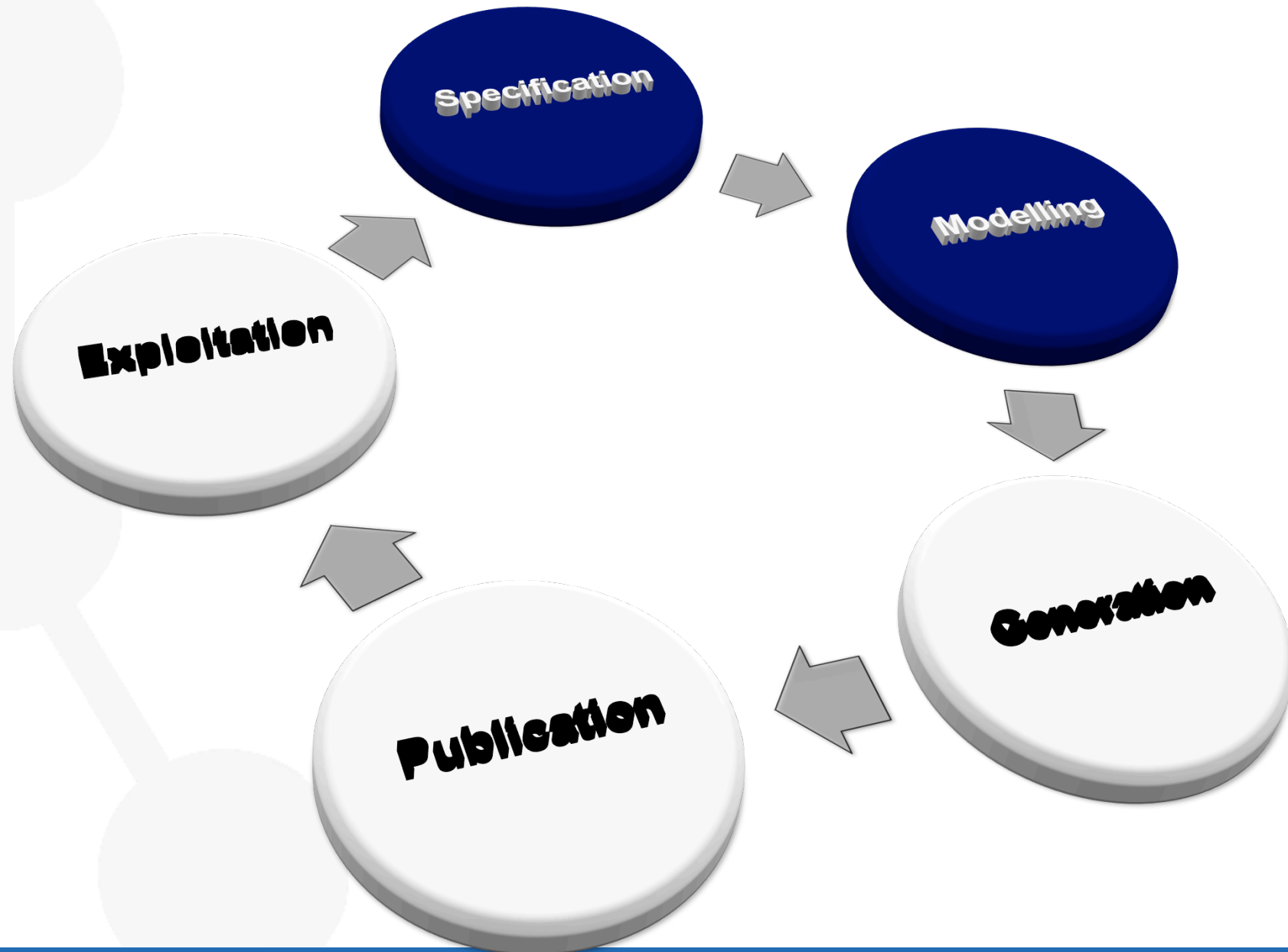
Unidades:miles de personas (media anual)

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr	Unemployr
	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Andalucía	587,5	544	493,7	475,8	465,9	483,9	478,5	462,3	456,5	477,3	492,3	602,9
Aragón	58,5	51,4	45,9	40,6	40	42,5	41,4	39,2	39,6	37,7	35,9	48,2
Principado de Asturias	79,7	72,1	64,3	62,1	58	59,1	60,4	61,6	57,6	53,4	50	52,8
Illes Balears	36,8	33,3	29,8	29,1	30,8	35,4	38,8	39,3	38,3	36,1	37,2	49,1
Canarias	133,7	123,2	109,6	106,6	108,9	112,7	119,4	132,1	133,4	130,4	133,7	175,1
Cantabria	31,8	28,3	24,6	23,3	23,4	24,5	25	25,6	24,2	21,8	20,6	25,4
Castilla y León	145,5	133,7	123,1	120,2	114,9	116,1	116,4	116,9	113,3	110	106,1	123,3
Castilla-La Mancha	108,3	97,6	88,6	87	86,6	88,9	91,4	93,5	94	91,6	91,7	115,5
Cataluña	322,5	272,5	232,5	213,8	222,1	252,5	264	270,4	265,5	260,8	256,4	333,7
Comunitat Valenciana	279,5	240	201,9	169,6	158,9	171	179,8	190,9	197,7	195,9	201,4	277,9
Extremadura	80,1	77,2	71,4	70,5	70,3	74	75,5	78,1	77,3	76,2	74,6	84,5
Galicia	201,7	186,3	168	163,4	159,5	168,4	174,7	176,9	172,1	161,3	150,5	162,5
Comunidad de Madrid	321,6	284	242,1	223,2	218,7	244,7	253,8	248,1	226	216,5	221,7	281
Región de Murcia	59,9	51,9	44,6	44	43,7	45,2	45,7	45,2	44,2	44,4	46,6	69,8
Comunidad Foral de Navarra	27,8	25,1	22,3	21,3	21,1	21,9	21,7	21,7	22,3	21	20,6	25,3
País Vasco	135,1	117,9	104	94,1	88,4	88,9	89,7	89,7	84,8	80,8	75,6	84,6
Rioja, La	12	10,9	9,6	9,1	8,9	9,7	9,5	9,8	10	10,2	10	12,8
Ceuta	4,7	4,4	4,1	4,4	4,7	4,8	5,2	6	6,5	7,4	6,9	7,4
Melilla	5,3	5,6	5,3	5,3	5,5	5,6	6,1	6,2	6,5	6,7	6,9	7,9

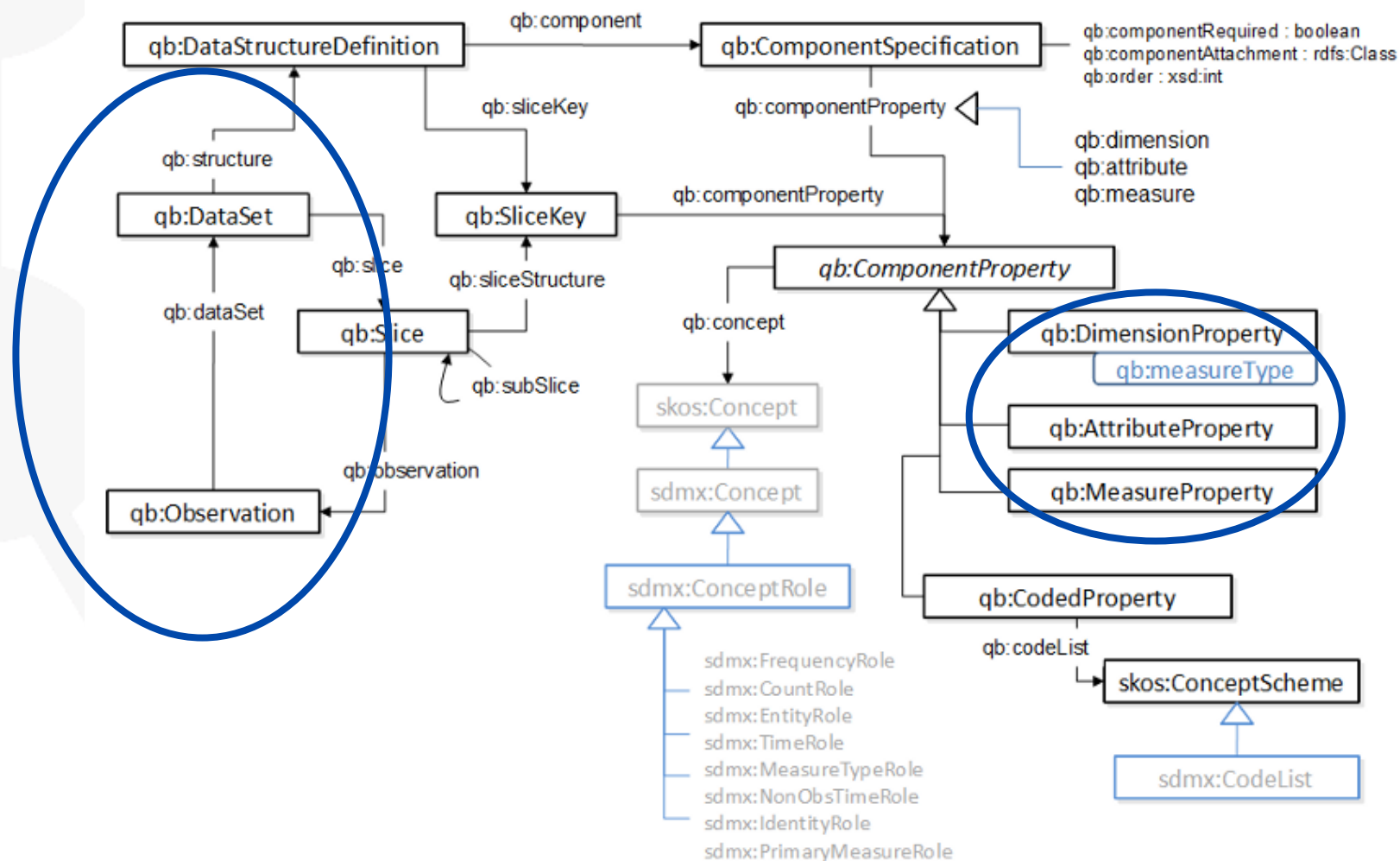
Notas:

1.- El INEM ha estimado series de paro registrado para los datos anteriores a mayo de 2005.

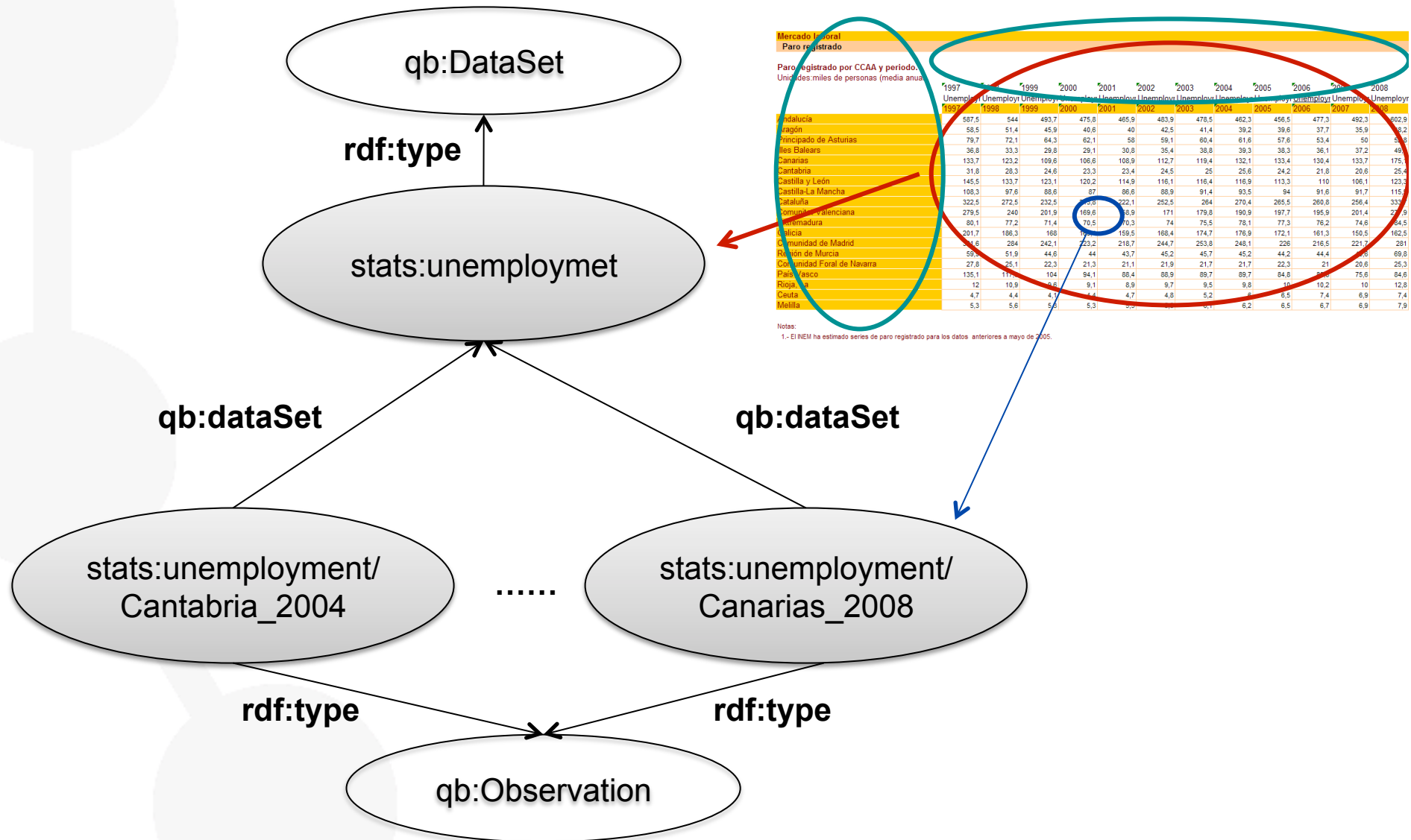
- Base URI
 - `http://stats.ull.es`
- TBOX URI
 - `http://stats.ull.es/ontology/{class|property}`
- ABOX URI
 - `http://stats.ull.es/resource/{resourceType}/{resource}`

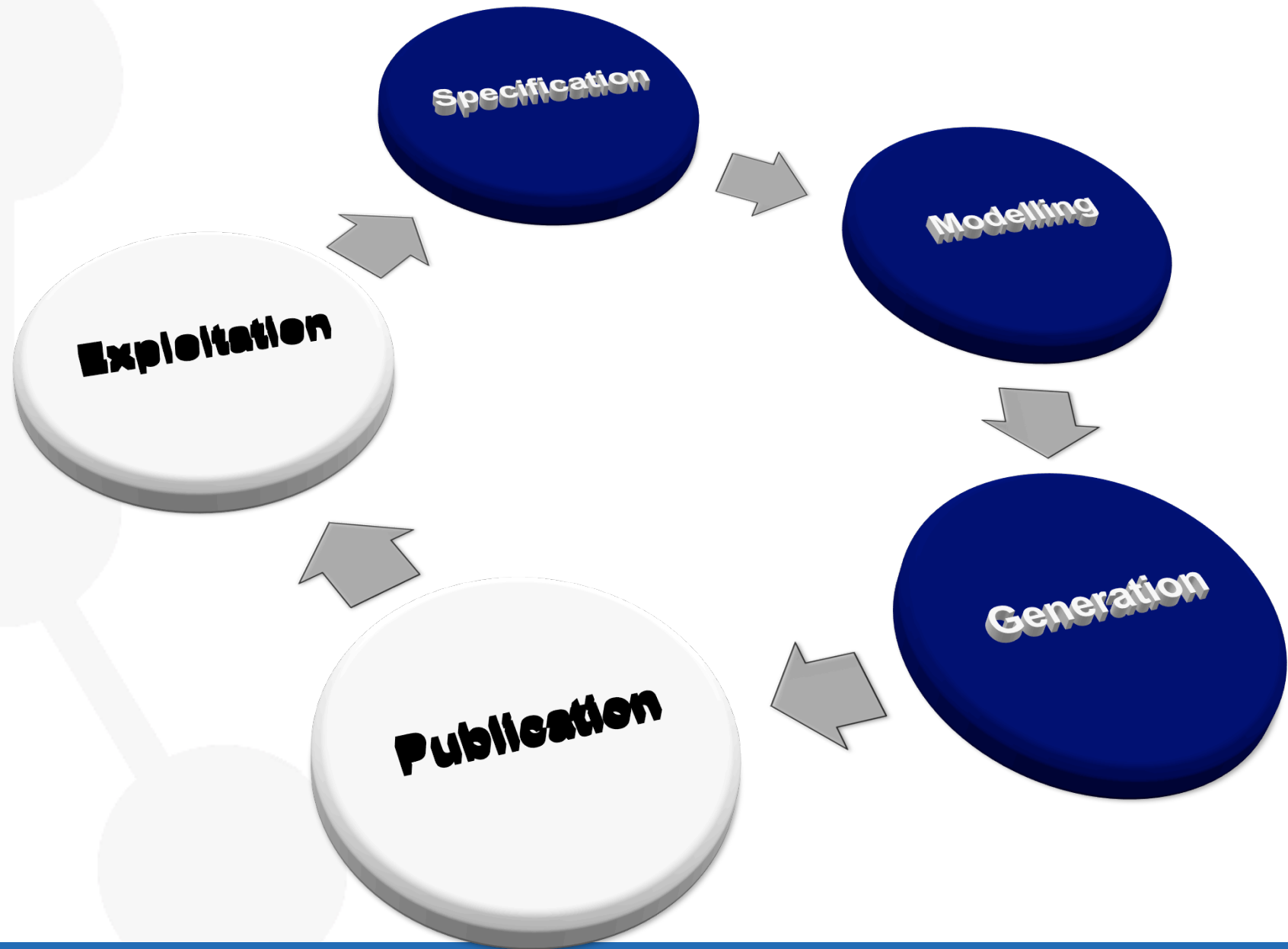


- RDF Data Cube

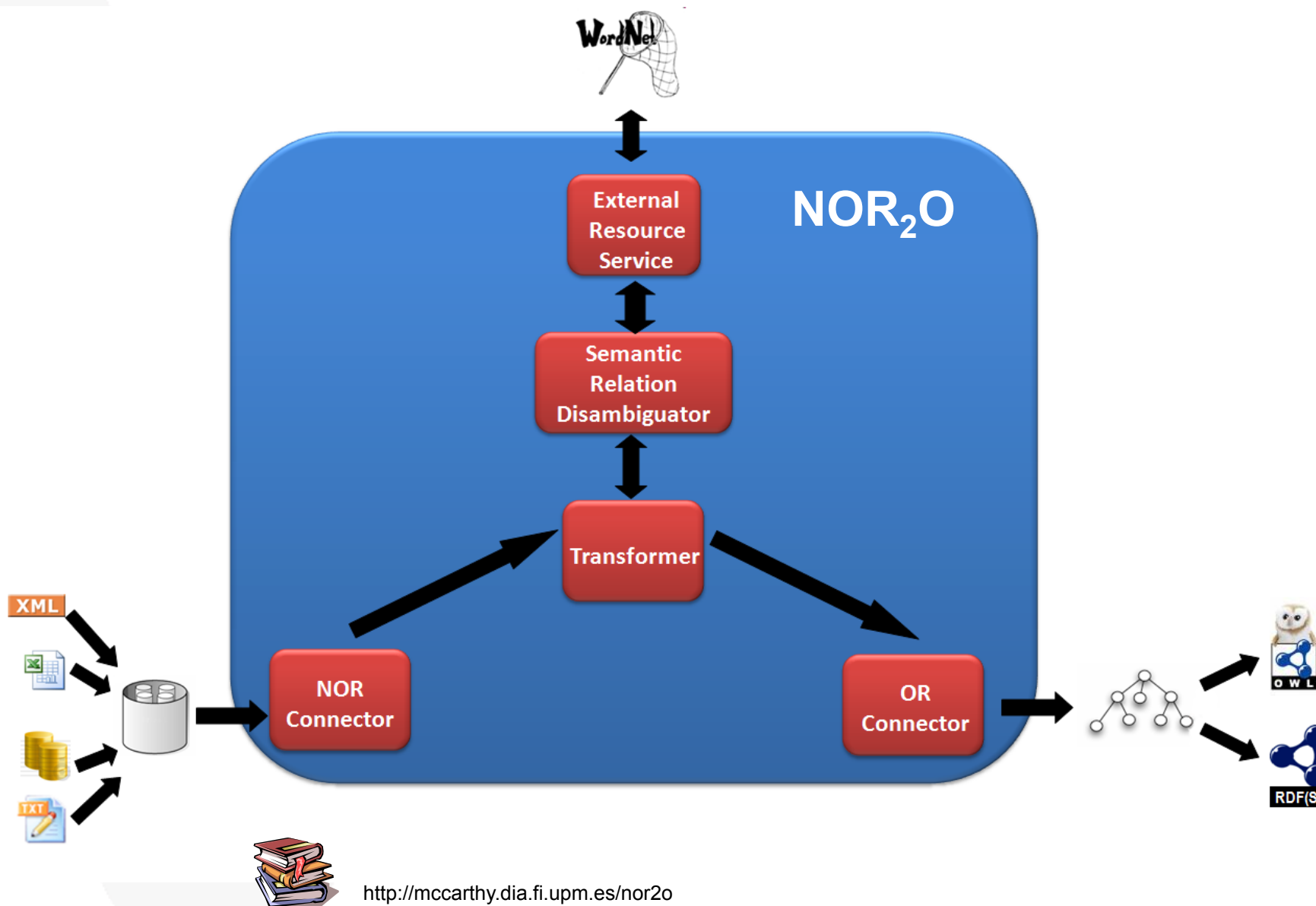


A simple example

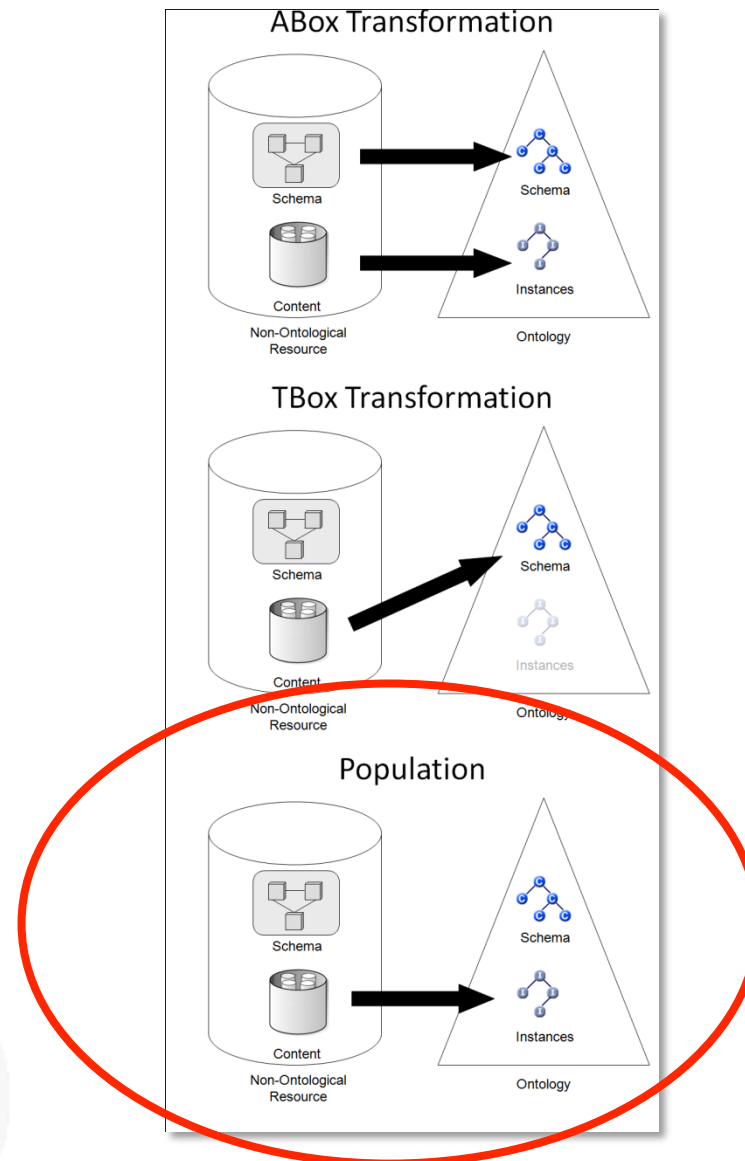




Generation – Transformation – Spreadsheets 2 RDF



Transformation approaches



- 3 configuration files
 - nor.xml, to describe the resource
 - prnor.xml to describe the transformation
 - or.xml to describe the ontology

NOR₂O – nor.xml

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Andalucía	587.5	544	493.7	475.8	465.9	463.5	478.5	462.3	455.5	477.3	492.3	602.9
Aragón	43.5	51.4	45.9	46.4	40	42.5	41.4	39.6	39.6	37.7	35.9	46.2
Principado de Asturias	79.7	74.7	67.5	62.1	55	59.1	60.4	61.6	62.3	60.4	50	52.8
Islas Baleares	36.8	33.3	29.8	29.1	30.8	35.4	38.8	39.3	38.3	38.1	37.2	49.1
Canarias	133.7	123.2	109.6	106.6	108.9	112.7	119.4	132.1	133.4	130.4	123.7	175.1
Cantabria	21.8	23.3	24.6	23.3	23.4	24.5	25	25.6	24.2	21.8	20.6	25.4
Castilla y León	145.5	133.7	123.1	120.2	114.9	116.1	116.4	116.9	113.3	110	106.1	123.3
Castilla-La Mancha	108.3	97.6	88.6	87	86.6	88.9	91.4	93.5	94	91.6	91.7	110.5
Cataluña	322.5	272.5	232.5	213.8	222.1	252.5	264	270.4	265.5	260.8	256.4	333
Comunitat Valenciana	279.5	240	201.9	169.6	158.9	171	179.8	190.9	197.7	195.9	201.4	277.5
Extremadura	80.1	77.2	71.4	70.5	70.3	74	75.5	78.1	77.3	76.2	74.6	84.6
Galicia	201.7	186.3	168	163.4	159.5	166.4	174.7	176.9	172.1	161.3	150.5	162
Comunidad de Madrid	321.6	294	242.1	223.2	219.7	244.7	253.8	248.1	228	216.5	221.7	261
Región de Murcia	59.9	51.9	44.6	44	43.7	45.2	45.7	45.2	44.2	44.4	46.6	69.8
Comunidad Foral de Navarra	27.3	25.1	22.3	21.3	21.1	21.9	21.7	21.7	22.3	21	20.6	25.3
País Vasco	135	117.9	104	94.1	88.4	88.9	89.7	89.7	84.8	80.8	76.6	84.6
Rioja	12	15.5	9.6	9.1								
Ceuta	4.7	4.4	4.1	4.4								
Melilla	5.3	5.6	5.3	5.3								

```

<Nor type="GenericSpreadSheet" name="Spanish Employment">
  <Schema>
    <SchemaEntities>
      <SchemaEntity name="Location">
        <Attribute name="Name" valueFrom="pcaxis.[A9:A25]" type="string" />
      </SchemaEntity>
      <SchemaEntity name="Dataset">
        <Attribute name="Name" valueFrom="pcaxis.[B7:M7]" type="string" />
      </SchemaEntity>
      <SchemaEntity name="Year">
        <Attribute name="Name" valueFrom="pcaxis.[B6:M6]" type="string" />
      </SchemaEntity>
      <SchemaEntity name="UnemploymentByLocationInPeriod" type="Nary">
        <Attribute name="hasValue" valueFrom="pcaxis.[B9:M25]" type="string" />
        <Relation name="hasLocation" usingSpreadSheetColumn="A" destination="Location" />
        <Relation name="inPeriod" usingSpreadSheetRow="6" destination="Year" />
        <Relation name="dataset" usingSpreadSheetRow="7" destination="Dataset" />
      </SchemaEntity>
    </SchemaEntities>
  </Schema>
  <DataModel>
    <GenericDataModel />
  </DataModel>
  <Implementation>
    <Spreadsheet type="ms" file="Paro.xls" />
  </Implementation>
</Nor>

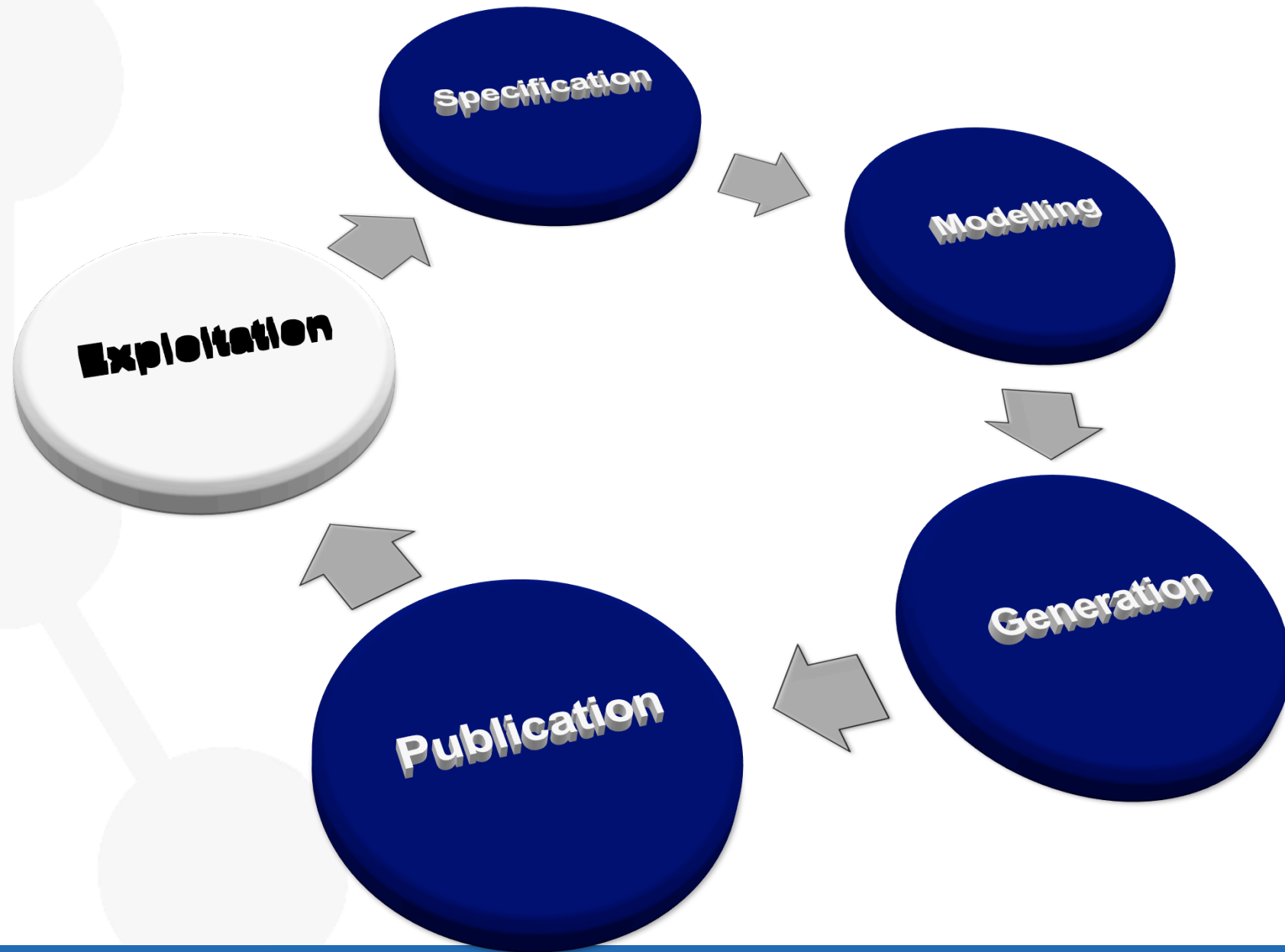
```

- How to transform the resource

```
<Prnor identifier="PR-NOR-CLLD-01" transformationApproach="Population">
  <Class from="UnemploymentByLocationInPeriod" identifier="cube:Observation">
    <ObjectProperty from="hasLocation" to="property:geoArea" />
    <ObjectProperty from="inPeriod" to="dimension:refPeriod" />
    <ObjectProperty from="dataset" to="cube:dataset" />
    <DataTypeProperty from="[hasValue]" to="property:Unemployment" type="http://www.w3.org/2001/XMLSchema#double" />
    <Individual from="UnemploymentByLocationInPeriod" identifier="unemployment/. [hasLocation]._. [inPeriod]" />
  </Class>
  <Class from="Dataset" identifier="cube:DataSet">
    <DataTypeProperty from="[Name]" to="nombre" type="http://www.w3.org/2001/XMLSchema#string" />
    <Individual from="Dataset" identifier="[Name]" />
  </Class>
  <Class from="Year" identifier="Year">
    <DataTypeProperty from="[Name]" to="nombre" type="http://www.w3.org/2001/XMLSchema#string" />
    <Individual from="Year" identifier="year/. [Name]" />
  </Class>
  <Class from="Location" identifier="Region">
    <DataTypeProperty from="[Name]" to="nombre" type="http://www.w3.org/2001/XMLSchema#string" />
    <Individual from="Location" identifier="geoes:ComunidadAut%C3%B3noma/. [Name]" />
  </Class>
</Prnor>
```

Reusing GeoLinkedData URIs!

```
<Or name="Unemployment Index" ontologyURI="http://stats.ull.es/resource/"
    ontologyFile="paro.rdf" implementation="OWL" alreadyExist="no" separator="">
  <prefix name="geosec" uri="http://geo.linkeddata.es/resource/" />
  <prefix name="cube" uri="http://purl.org/linked-data/cube#" />
  <prefix name="property" uri="http://stats.ull.es/property/" />
  <prefix name="dimension" uri="http://purl.org/linked-data/sdmx/2009/dimension#" />
  <prefix name="rdf" uri="http://www.w3.org/1999/02/22-rdf-syntax-ns#" />
  <prefix name="year" uri="http://reference.data.gov.uk/id/year/" />
</Or>
```



- Upload the generated file

The screenshot shows the Virtuoso Conductor web interface. The top navigation bar includes tabs for Home, System Admin, Database, Replication, Web Application Server, XML, Web Services, Linked Data, and NNTP. The 'Linked Data' tab is active, and the 'Quad Store Upload' sub-tab is selected. The main content area is titled 'Quad Store Upload'. It contains a form with the following elements:

- ☒ File*: A 'Choose File' button and the text 'No file chosen'.
- ☐ Resource URL*: An empty text input field.
- ☐ Create graph explicitly
- Named Graph IRI*: A text input field containing the URL `http://localhost:8890/graph/ontology/`.
- Buttons: 'Cancel' and 'Upload'.

Black arrows in the image highlight the 'Quad Store Upload' tab and the 'Upload' button.

- <http://stats.ull.es/graph/dataset>

- `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 "...";
```

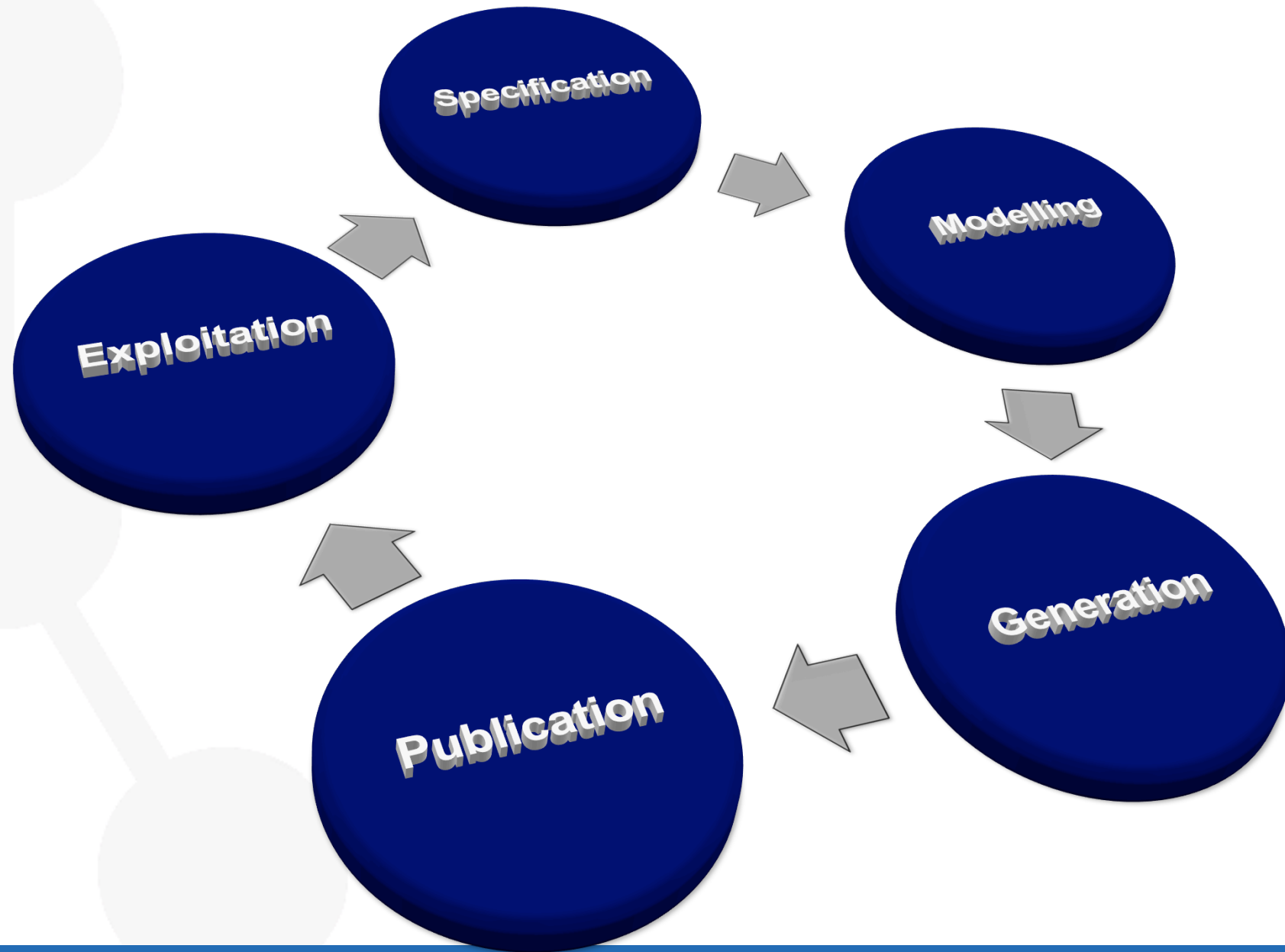
- Simple command line tool
- Sends a SPARQL query to list all URIs
- Generates sitemap

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

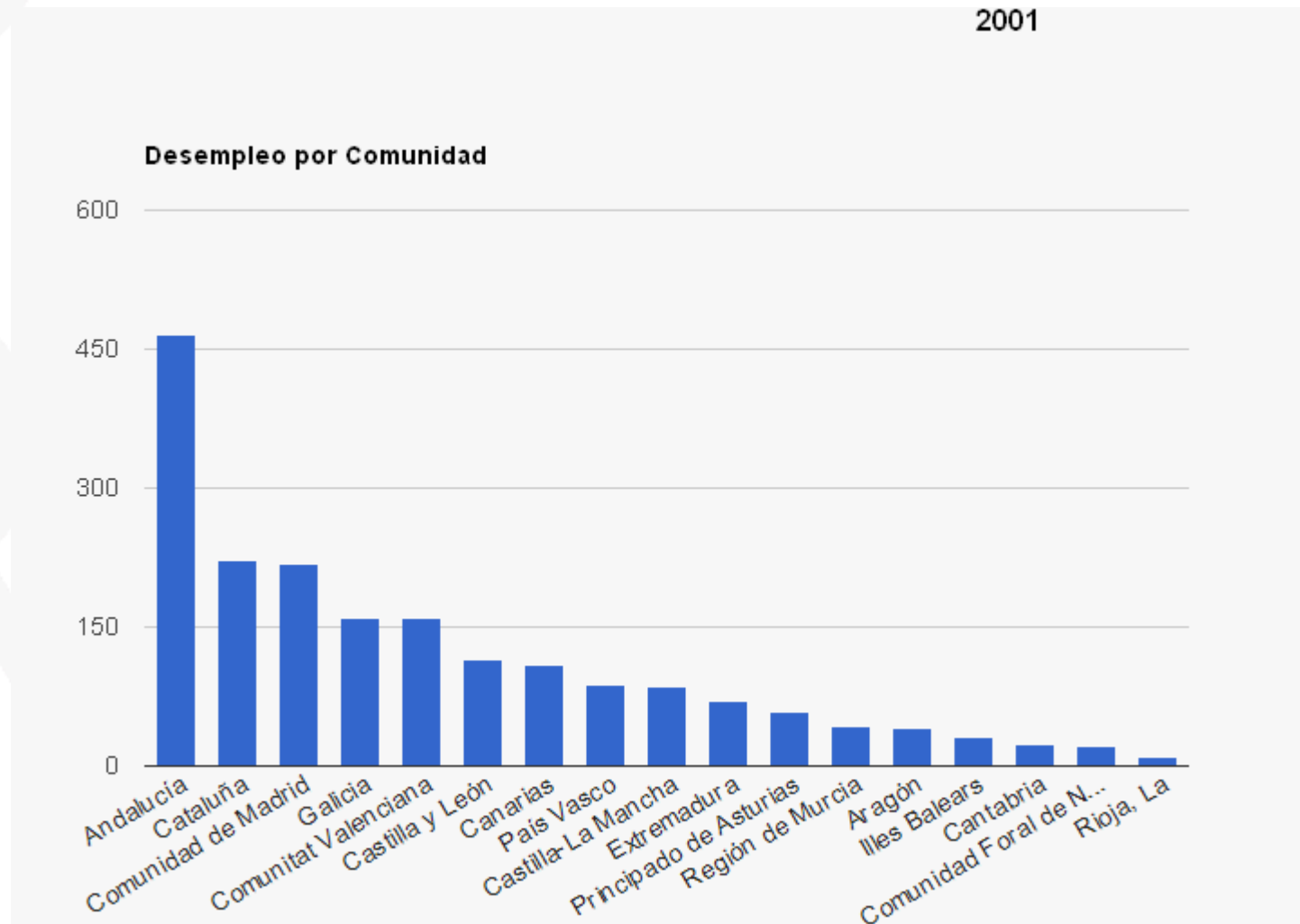
Example:

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

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



<http://localhost/stats/desempleo?year=2001>



Another example

<http://county-rank.data-gov.ie/#County-Donegal>

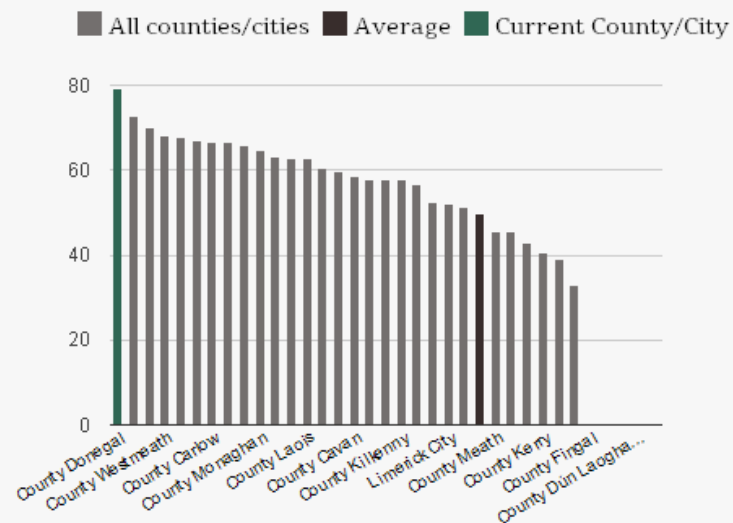
The indicators displayed are the ones in which the County/City has very good and bad performance

1

MOTOR TAXATION

Number of Motor Tax Transactions

Percentage of motor tax transactions which are dealt with over the counter



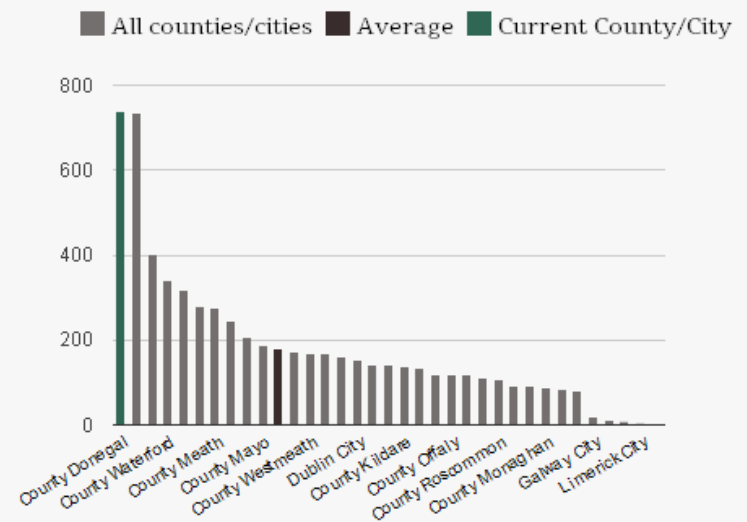
Place your mouse over each bar to get more information.

1

PLANNING APPLICATIONS - DECISION MAKING

Individual Houses

Number of decisions in Column A which were within 8 weeks



Place your mouse over each bar to get more information.





Publishing Linked Data from spreadsheets

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