



# Ontology-based Access to Legacy Databases

Oscar Corcho, Asunción Gómez-Pérez  
{asun,ocorcho}@fi.upm.es

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

# Motivation



- Migración **Web Profunda** → **Web Semántica**

*Uno de los principales impulsos para la Web Semántica sería la exposición en ella de la gran cantidad de bases de datos relacionales existentes en la Web, de modo que su contenido pueda ser procesado automáticamente.*

*Tim Berners Lee 1998 Design Issues*

*No tiene sentido tratar de anotar manualmente los millones de páginas existentes en la Web. Esas páginas están diseñadas para consumo humano. Para poblar la Web semántica hay que volverse hacia las bases de datos existentes y su contenido.*

*Tim Berners Lee 2006 Entrevista iSIGHT*

- **Explicitación de la semántica** de las bases de datos relacionales

*Los tres problemas de investigación más importantes en Bases de Datos solían ser 'rendimiento', 'rendimiento' y 'rendimiento'; en los próximos años los tres problemas más importantes y que verdaderamente supondrán un desafío para la investigación serán 'la semántica', 'la semántica' y 'la semántica'.*

*Stefano Ceri 2004 SWDB Toronto*

- **Reutilizar recursos no ontológicos y ontológicos** para construir nuevas ontologías.

*Gómez-Pérez, Suárez-Figueroa, D5.3.1. NeOn deliverable*

# Motivation



*Set of  
Standards*



Publish



Generate



# Motivation



**XML**



Different  
sources

Modeled

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

ISO 3166-1 Country	
Code	Name
GB	UNITED KINGDOM
ES	SPAIN
..	..

ISO 3166-2 Subdivision		
Code	Name	ISO 3166-1 Code
GB-NI	Northern Ireland	GB
GB-EA	East Anglia	GB
..	..	..

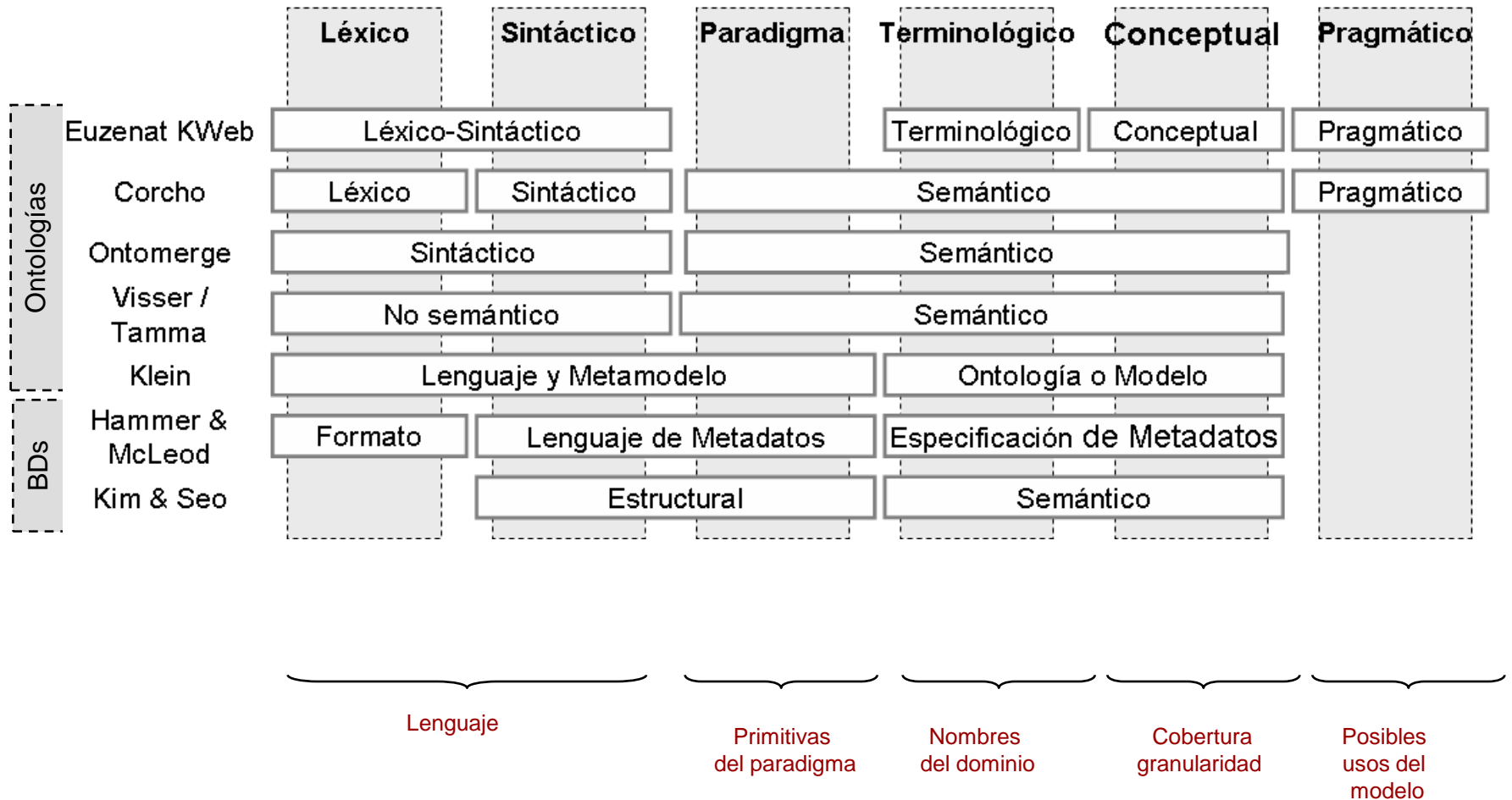
ISTAT - Codes and names of geographical areas, provinces and regions				
Geographical area	Region Code	Region Name	Province Code	Province Name
NORD-OCCIDENTALE	01	PIEMONTE	001	Torino
NORD-OCCIDENTALE	01	PIEMONTE	002	Vercelli
NORD-ORIENTALE	05	VENETO	023	Verona
..	..	...	...	...

Hierarchy  
Models

Generate

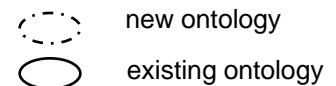
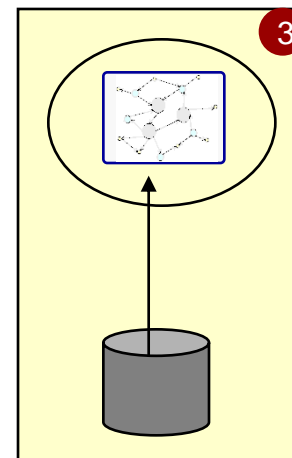
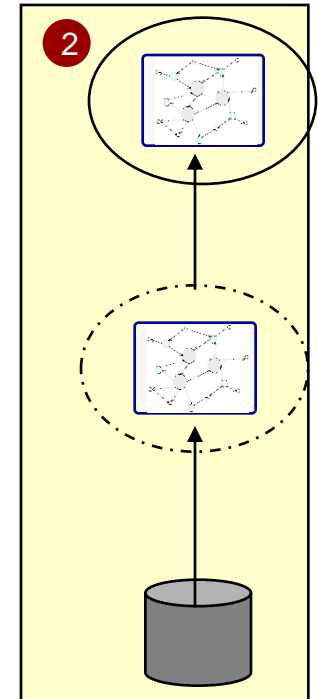
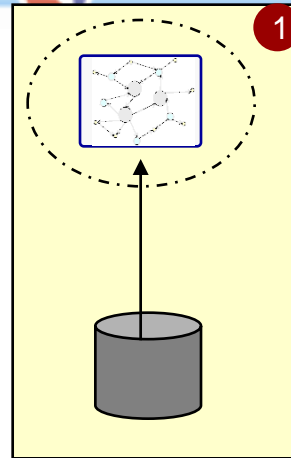


# Heterogeneity problem



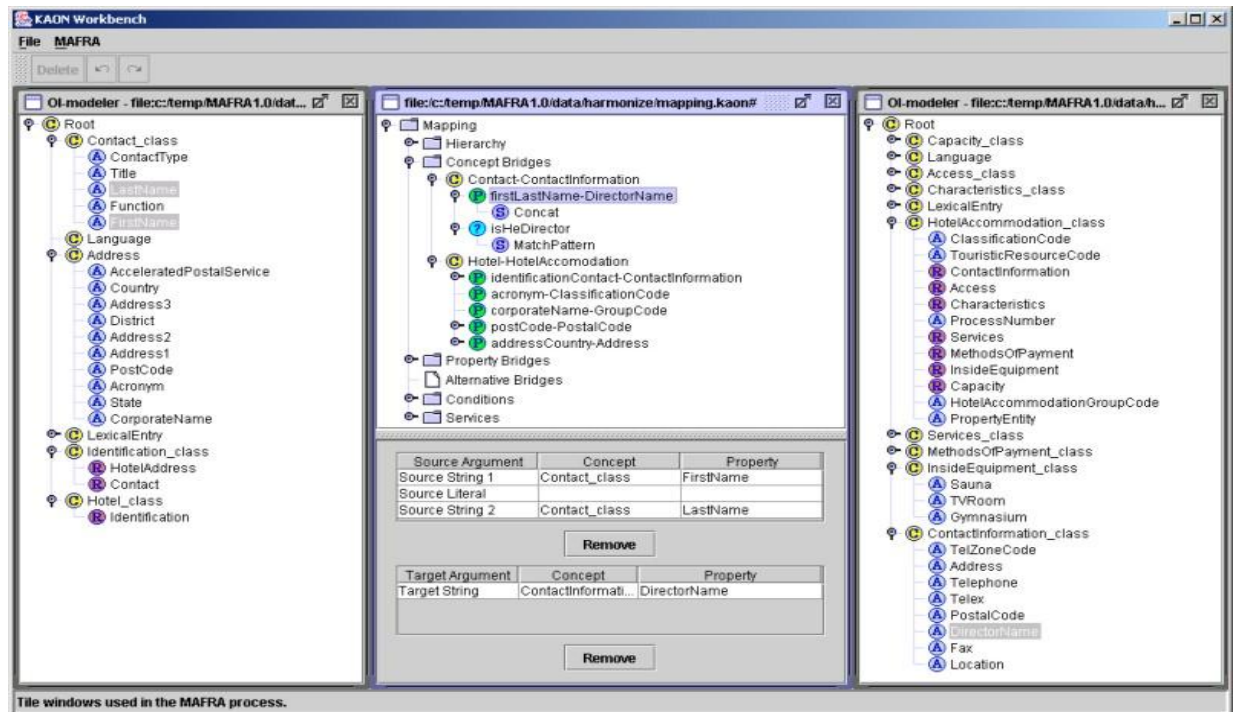
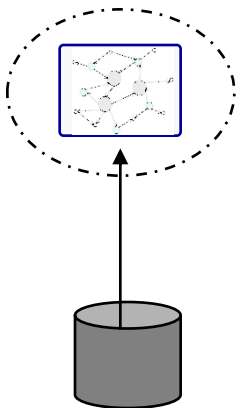
# Existing approaches

1. Construir una **nueva ontología** a partir de 1 esquema y datos de 1 BD (OntoStudio, KaOn Reverse)
2. Mapear la Onto construida en el enfoque 1, con una **ontología de legado** (NeOn toolkit UKARL)
3. Mapear BD existente a una **ontología de legado** (NeOn Toolkit UPM)
  - a) Volcado masivo
  - b) Dirigido por las consultas



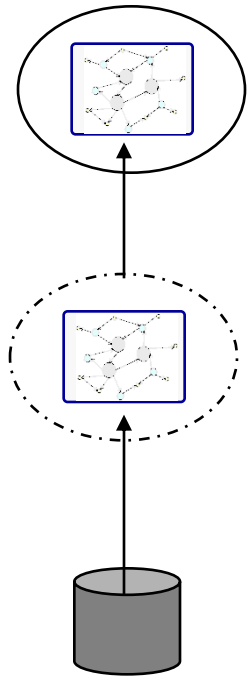
# Existing approaches

1. Construir 1 Ontología a partir de 1 esquema y datos de 1 BD (OntoStudio, KaOn Reverse)

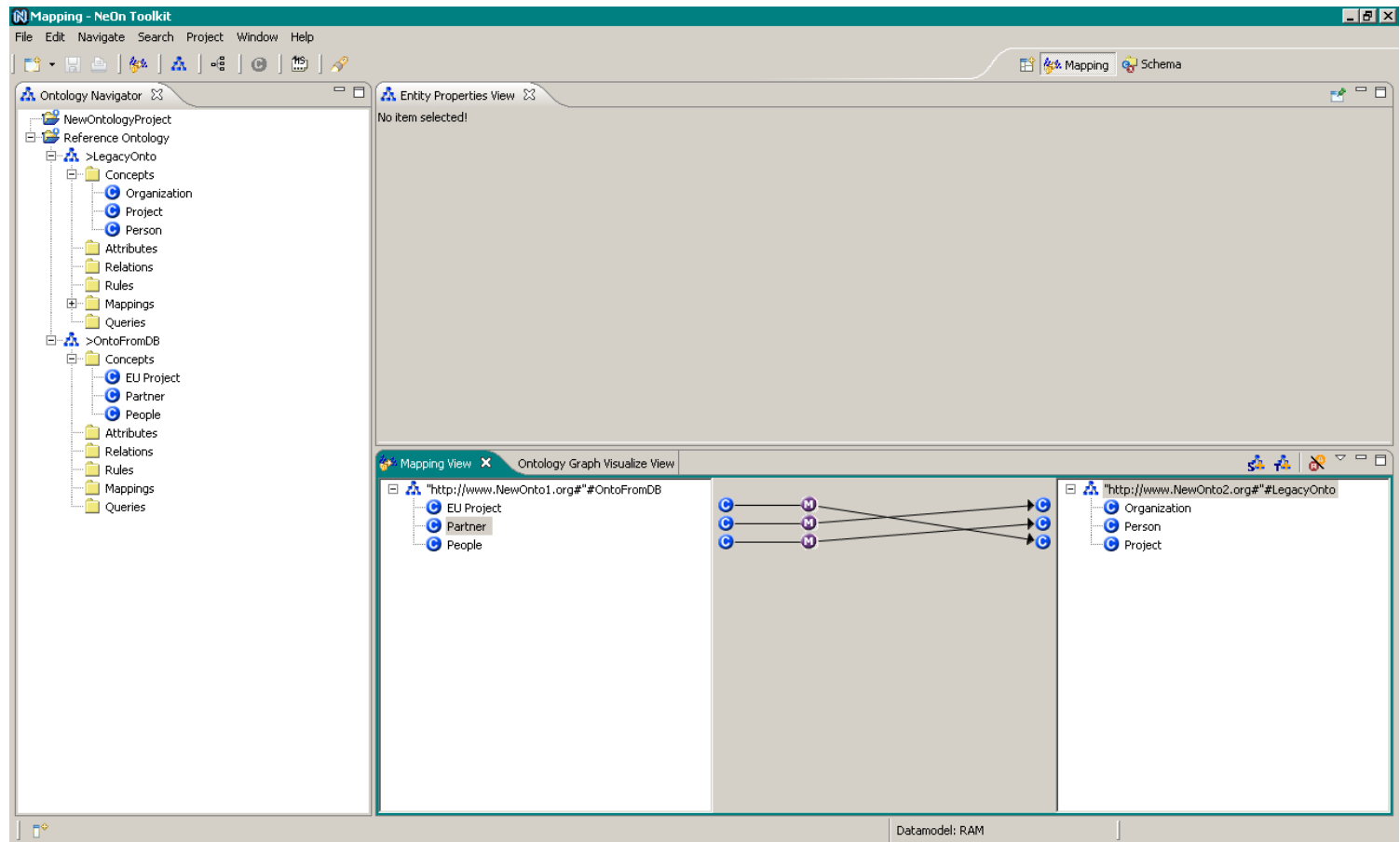


# Existing Approaches

## 2. Mapear la onto construida en el enfoque 1, con una ontologia de legado (NeOn toolkit UKARL)



NeOn  
TOOLKIT



The screenshot shows the NeOn Toolkit Mapping interface. The 'Ontology Navigator' on the left displays a tree structure with 'LegacyOnto' and 'OntoFromDB'. The 'Mapping View' at the bottom shows a mapping between 'http://www.NewOnto1.org/#OntoFromDB' and 'http://www.NewOnto2.org/#LegacyOnto'. The mapping shows 'EU Project', 'Partner', and 'People' from the source ontology mapped to 'Organization', 'Person', and 'Project' in the target ontology.



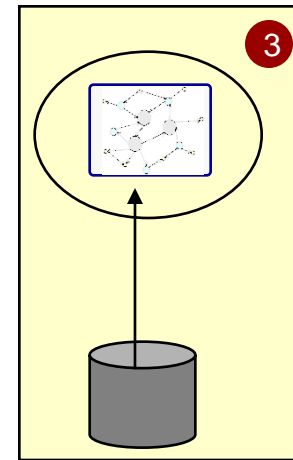


# Existing approaches

## 3. Acceso a contenido de BD usando ontologías de legado

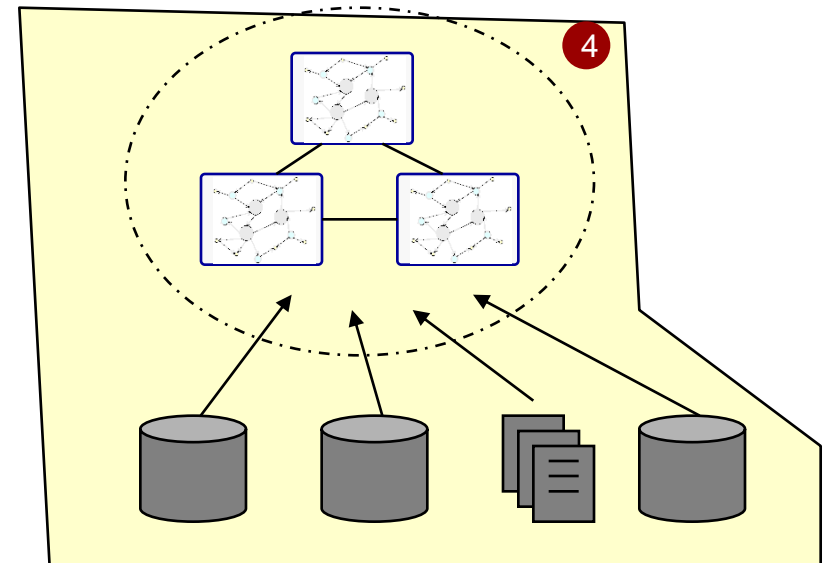
(NeOn toolkit at UPM)

- R20 y ODEMapster
  - Vocado Masivo
  - No migrado masivo
- Fundfinder Case study
- FAO case study

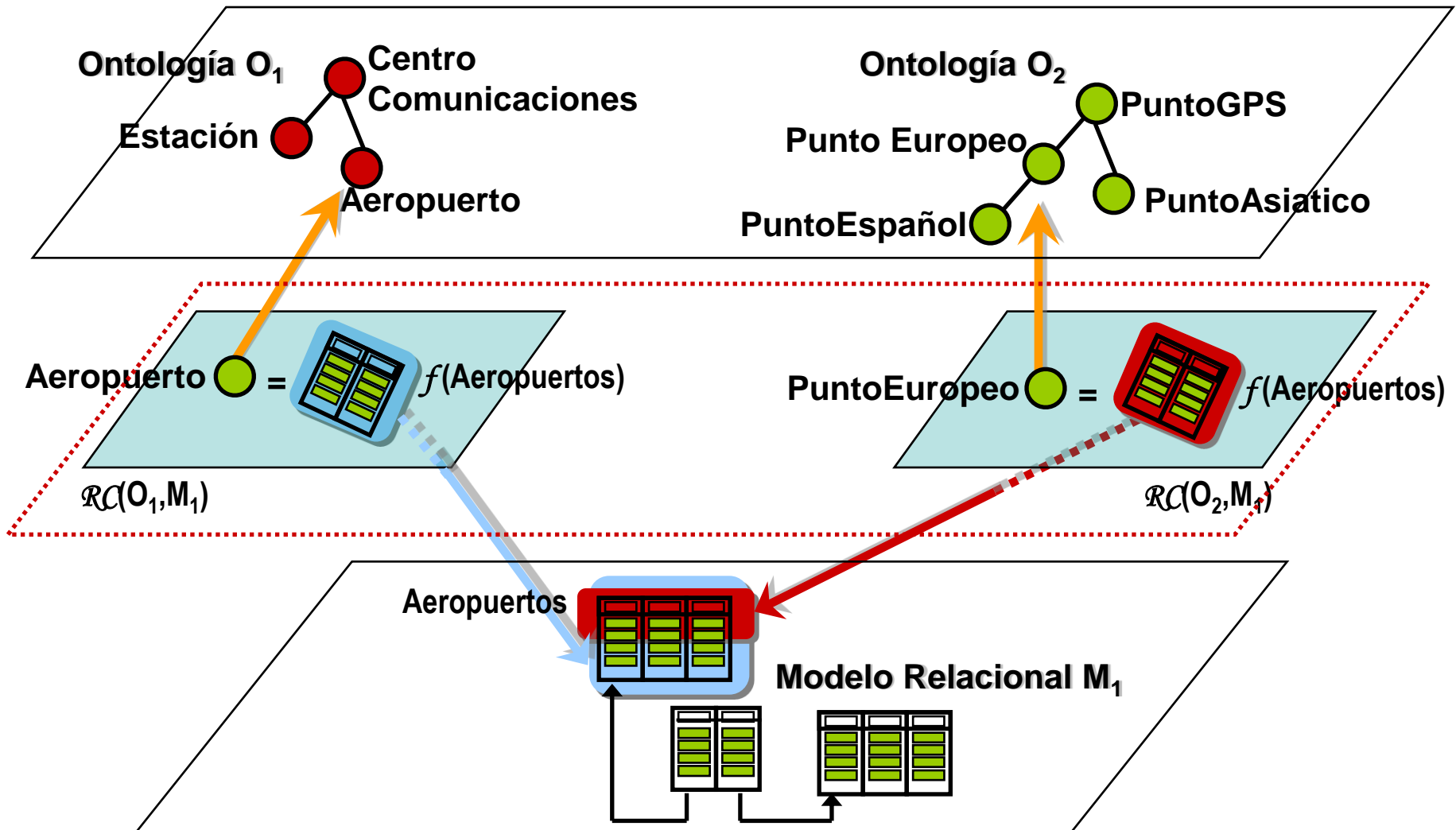


## 4. Construcción de redes de ontologías mediante la reutilización y reingeniería de recursos no ontológicos y ontológicos

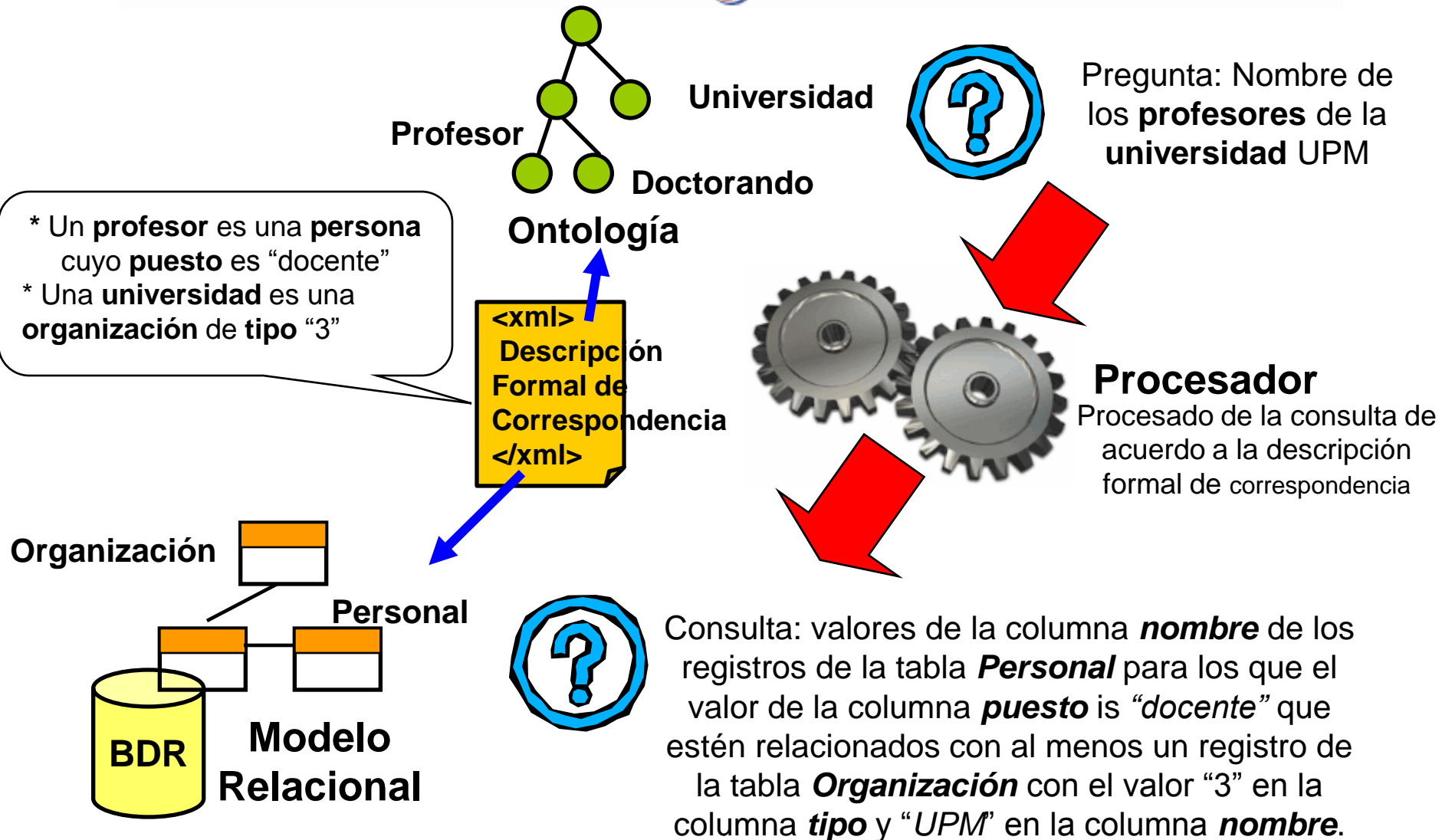
- Seemp case study



# Ontology-based view over a relational model (I)



# Ontology-based view over a relational model (II)



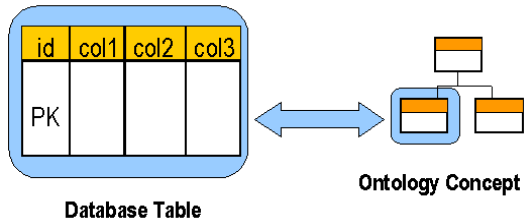
## Upgrading Database content to the semantic Web



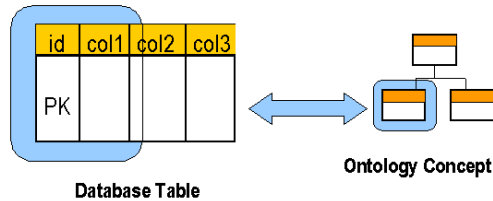
- Integrating information from different DB sources
- Reuse of legacy DBs and legacy ontologies
- R<sub>2</sub>O: Declarative Mapping description language
- ODEMapster: Generic query processor.
  - asking queries to a relational database using ontology terms
  - On demand query answering
  - Batch ontology population
- A well defined method for upgrading and integrating content from heterogeneous sources.

# R<sub>2</sub>O (Relational-to-Ontology) Language

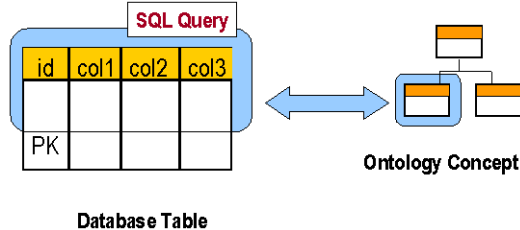
## Casos de mapping cubiertos por el lenguaje para conceptos...



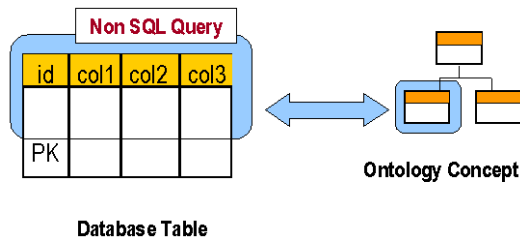
A view maps exactly one concept in the ontology.



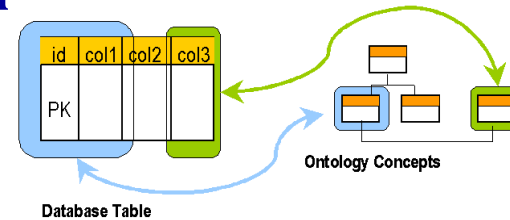
A subset of the columns in the view map a concept in the ontology.



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

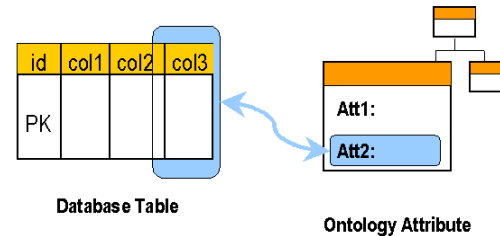


A subset of the records of a database view map a concept in the onto. but the selection cannot be made using SQL.

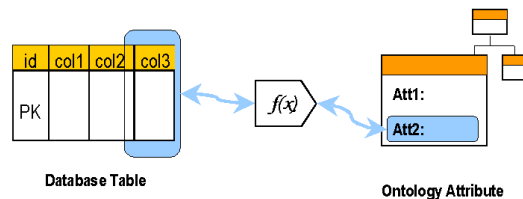


One or more concepts can be extracted from a single data field (not in 1NF).

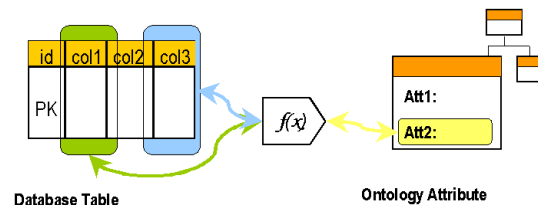
## para atributos...



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



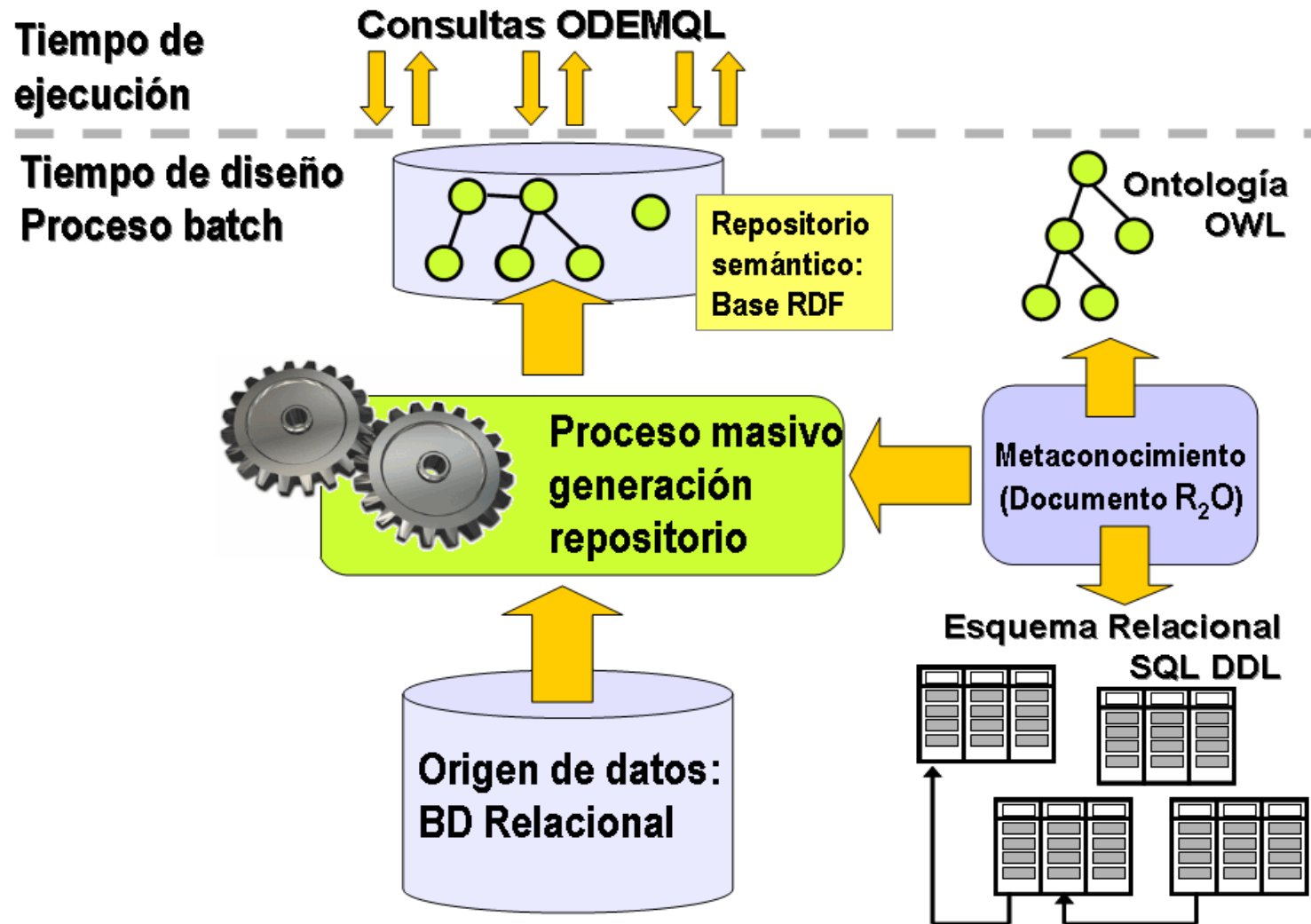
A column in a database view maps an attribute or a relation after some transformation.



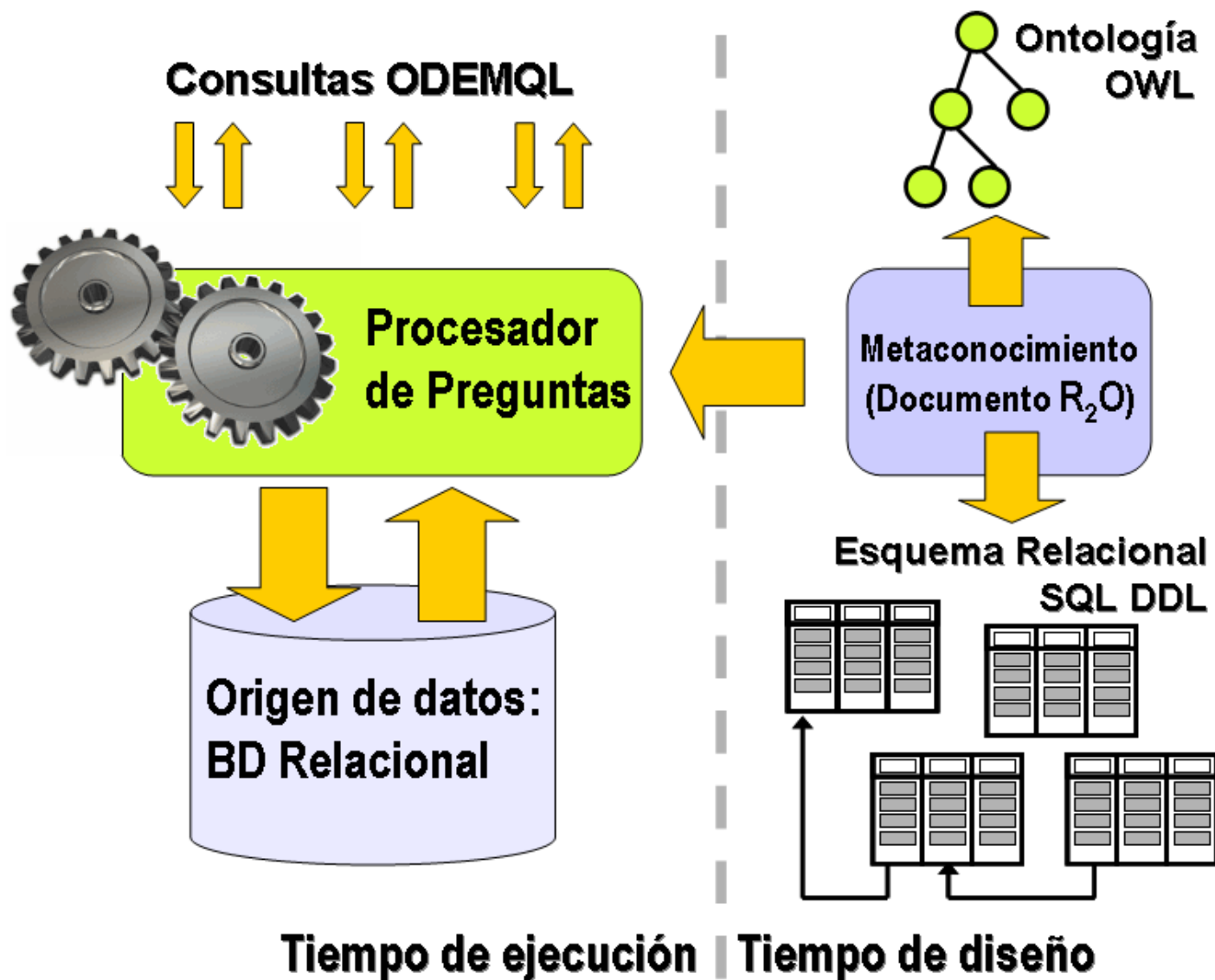
A set of columns in a database view map an attribute or a relation.

# ODEMapster: Volcado masivo

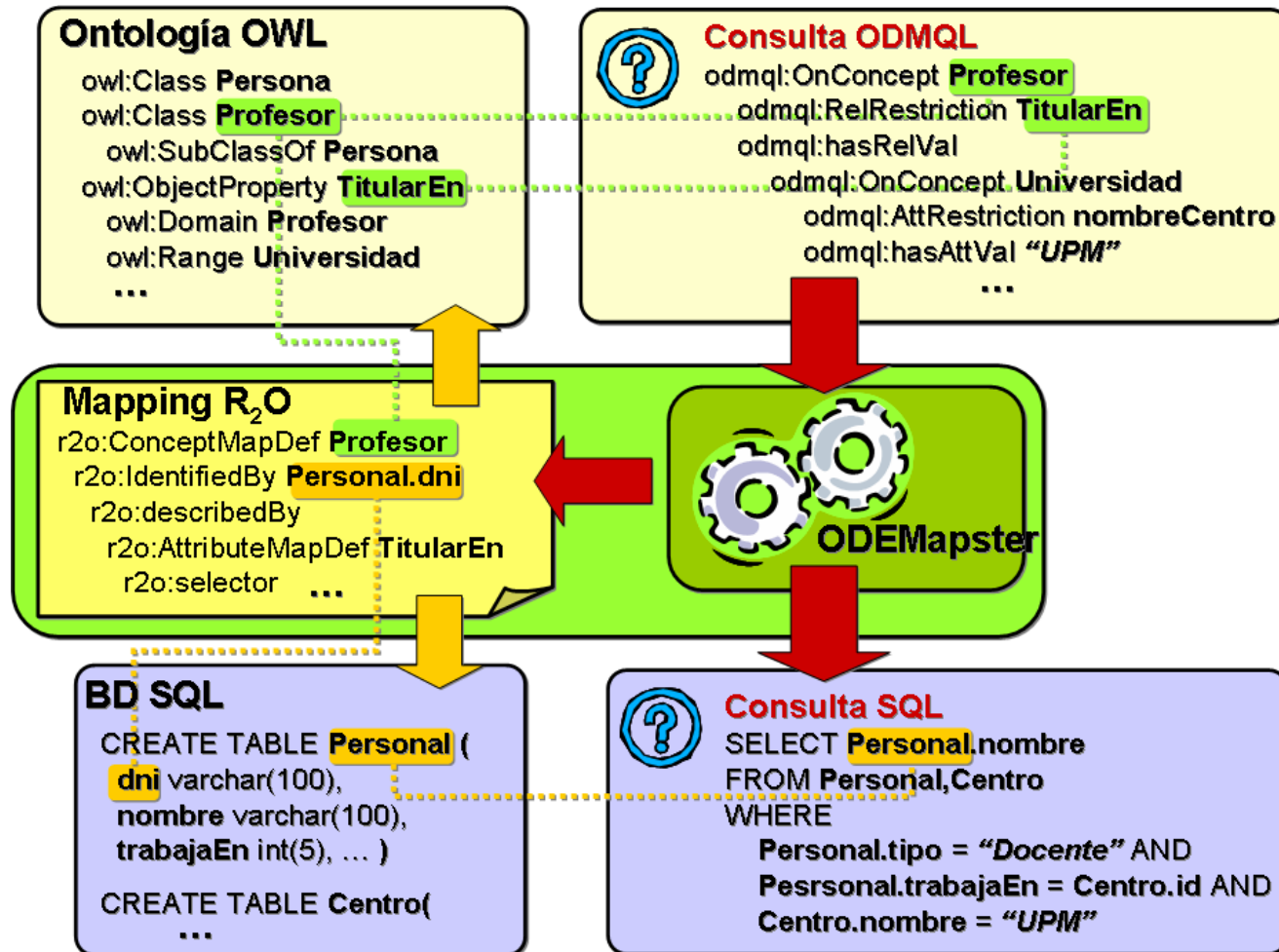
Creación de un repositorio semántico en RDF



# ODEMapster: Dirigido por las consultas



# Lenguajes utilizados por ODEMapster





# Source system description



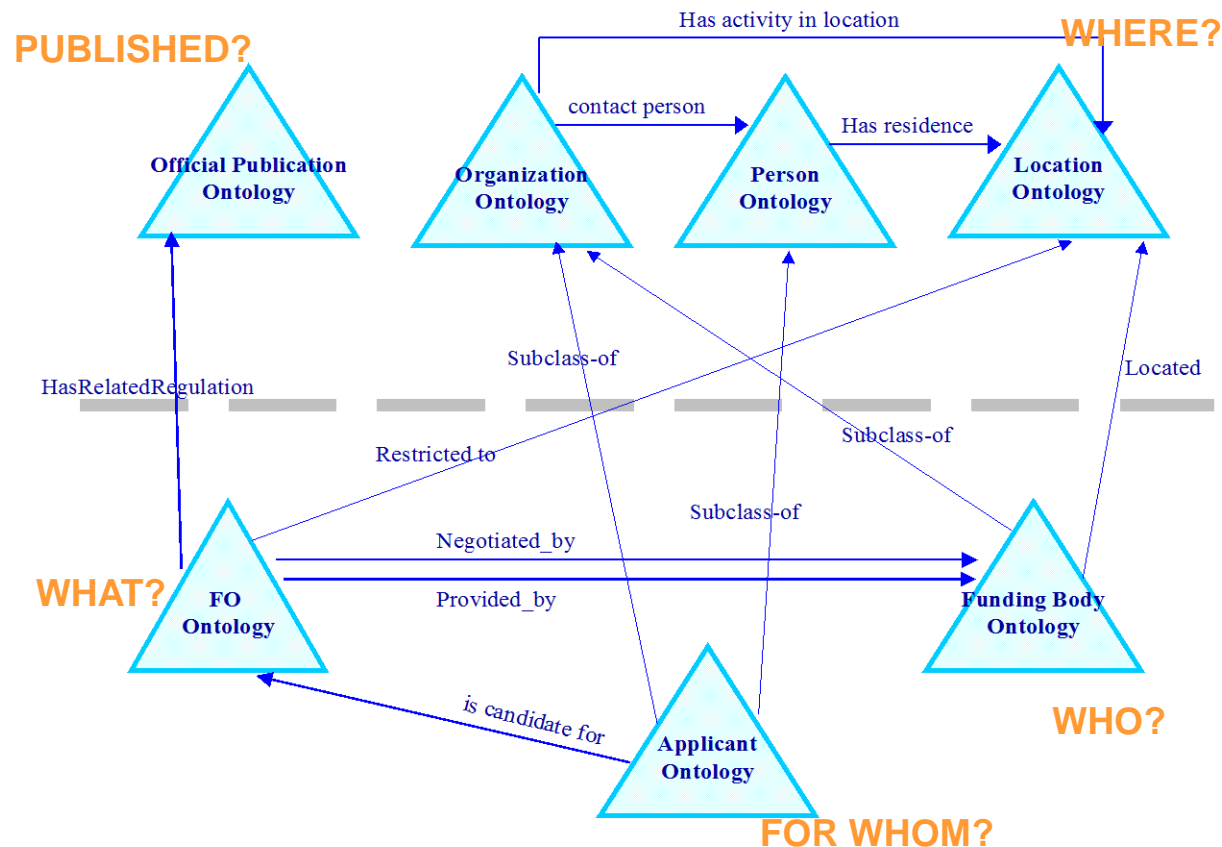
- Source system: FISUB Database (upgraded with R2O)
  - Maintained manually on a daily basis
  - Published on the Web ([www.gencat.net](http://www.gencat.net))
  - Around 300 records, 5 tables (most of the information contained in one of them).
  - Search for fundings by:
    - Dates
    - Sector
    - Subsector
    - Keywords
    - Purpose
- Source system: BOE Web site (Spain's official journal)
  - PDF files published on the web ([www.boe.es](http://www.boe.es))
  - Search by date and number
- Source system: DOGC (Catalonia's official journal)
  - HTML files published on the web ([www.gencat.net/diari](http://www.gencat.net/diari))
  - Search by date and number

# Build the ontologies



## Inter ontology relations.

- **Funding opportunity ontology:**
  - Concepts: 32
  - Instance attributes: 12
  - Subclass-of: 25
  - Ad-hoc relations: 8
- **Funding Body ontology:**
  - Concepts: 7
  - Subclass-of: 5
  - Ad-hoc relations: 2
- **Applicant ontology:**
  - Concepts: 22
  - Instance attributes: 4
  - Subclass-of: 18
  - Ad-hoc relations: 4
- **Official Publication ontology:**
  - Concepts: 9
  - Instance attributes: 6
  - Subclass-of: 7
  - Ad-hoc relations: 1
- **Organization ontology:**
  - Concepts: 6
  - Instance attributes: 10
  - Ad-hoc relations: 6
- **Person ontology:**
  - Concepts: 5
  - Instance attributes: 9
  - Ad-hoc relations: 4
- **Location ontology:**
  - Concepts: 4
  - Instance attributes: 2
  - Subclass-of: 3
  - Ad-hoc relations: 3



# Population example (I)



Funding Opportunities : Consulta de selecció

ID	TITLE	BENEFICIARIES	BEGIN_END	LEG_REF	FUND_C
623	PROFIT - Program:	1. Empreses 2. Agrup	Des de 19 de novembre fins el 19 d	Ordre de 7 de març de 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
614	PROFIT - Program:	1. Pimes 2. Empreses	Des de 19 de novembre fins el 19 d	Ordre de 7 de març de 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
582	PROFIT - Program:	1. Associacions empresar	Des de 19 de novembre fins el 19 d	Ordre de 7 de març de 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
581	PROFIT - Program:	Empreses; Agrupacions i	ernacional Fins al 2 de setembre	Ordre de 13 de juny del 2000 (BOE 148, 21-06-20	SUBVENCIONS:<E
576	PROFIT - Program:	Pimes, Empreses, Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
574	PROFIT - Program:			Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
573	PROFIT - Program:			Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
572	PROFIT - Program:			Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
571	PROFIT - Program:	Pimes; Empreses; Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	Els ajuts que es po

Attribute Mapping with transformation  
(Regular Expression)

Attribute Direct Mapping

Relation Mapping w. Transformation  
(Regular Expression)

Relation Mapping w. Transformation  
(Keyword search)

```
<rdf:RDF>
- <rdf:Description rdf:about="http://cidem.gencat.es/ajut_581">
  <fo:title xml:lang="es">PROFIT - Programa nacional de l'espai</fo:title>
  <fo:deadline>2-setembre-2003</fo:deadline>
  <fo:legalRef rdf:resource="http://official-pub/BOE-148-2000"/>
  <fo:legalRef rdf:resource="http://official-pub/BOE-74-2001"/>
  <fo:legalRef rdf:resource="http://official-pub/BOE-272-2001"/>
  <fo:legalRef rdf:resource="http://official-pub/BOE-261-2002"/>
  <fo:legalRef rdf:resource="http://official-pub/BOE-273-2002"/>
  <fo:provided_by rdf:resource="http://cidem.gencat.es/Administraci? Central_Ministerio de Ciencia y Tecnologia?a"/>
  <rdf:type rdf:resource="http://www.esperanto.net/fundFinder/fundFinder/fundingOpp#Subvention"/>
</rdf:Description>
- <rdf:Description rdf:about="http://cidem.gencat.es/ajut_548">
  <fo:title xml:lang="es">
    Cinqu? programa comunitari d'acci? per a la igualtat d'oportunitats entre homes i dones (2001-2005)
  </fo:title>
```

...

How to  
express this  
using R2O??

# Population example (II)

Funding Opportunities : Consulta de selecció

ID	TITLE	BENEFICIARIES	BEGIN_END	LEG_REF	FUND_C
623	PROFIT - Program:	1. Empreses 2. Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març de 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
614	PROFIT - Program:	1. Pimes 2. Empreses	Des de 19 de novembre fins el 19 d	Ordre de 7 de març de 2000 (BOE 59, 9-03-2000).	Els ajuts que es po
582	PROFIT - Program:	1. Associacions empresar	Des de 19 de novembre fins el 19 d	Ordre de 7 de març de 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
581	PROFIT - Program:	Empreses; Agrupacions i	ernacionals: fins al 2 de setembre	Ordre de 13 de juny del 2000 (BOE 148, 21-06-2000)	SUBVENCIONS:<E
576	PROFIT - Program:	Pimes, Empreses, Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
574	PROFIT - Program:	Pimes; Empreses; Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
573	PROFIT - Program:	1. Pimes 2. Empreses	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
572	PROFIT - Program:	1. Pimes 2. Emprese	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
571	PROFIT - Program:	Pimes; Empreses; Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E
570	PROFIT - Program:	Pimes; Empreses; Agrupa	Des de 19 de novembre fins el 19 d	Ordre de 7 de març del 2000 (BOE 59, 9-03-2000).	SUBVENCIONS:<E

stro: 11 9 de 28

The Operation element defines a transformation based on a regular expression to be applied to the database column for extracting property values

```
<attributemap-def
  name="http://net.esperonto/fundfinder#LegalRef"
  <selector>
    <aftertransform>
      <operation oper-id="constant">
        <arg-restriction on-param="column">
          <has-column>Titol</has-column>
        </arg-restriction>
      </operation>
    </aftertransform>
  </selector>
</attributemap-def-->
```

```
<attributemap
  <selector>
    <aftertra
      <oper
        <arg
          <l
        </ar
        <arg
          <l
        </ar
      </oper
    </selector>
  </attributema
```

```
< relfromatt-def name=" http://net.esperonto/fundfinder#LegalRef">
  <selector>
    <aftertransform>
      <operation oper-id="concat">
        <arg-restriction on-param="string1">
          <has-value>http://official-pub/</has-value>
        </arg-restriction>
        <arg-restriction on-param="string2">
          <has-transform>
            <operation oper-id="apply-regexp">
              <arg-restriction on-param="string">
                <has-column>Referencia Legal</has-column>
              </arg-restriction>
              <arg-restriction on-param="regexp">
                <has-value>(BOE|DOGC|DOCE)\s*(L|S)*\s*(\d*)</has-value>
              </arg-restriction>
              <arg-restriction on-param="pattern">
                <has-value>%1-%2-%3</has-value>
              </arg-restriction>
            </operation>
          </has-transform>
        </arg-restriction>
      </operation>
    </aftertransform>
  </selector>
</ relfromatt-def >
```

```
<rdf:RDF>
  <rdf:Description rdf:about="http://cidem.gencat.es/ajut_581">
    <fo:title xml:lang="es">PROFIT - Programa nacional de l'espai</fo:title>
    <fo:deadline>2-setembre-2003</fo:deadline>
    <fo:legalRef rdf:resource="http://official-pub/BOE-148-2000"/>
    <fo:legalRef rdf:resource="http://official-pub/BOE-74-2001"/>
    <fo:legalRef rdf:resource="http://official-pub/BOE-272-2001"/>
    <fo:legalRef rdf:resource="http://official-pub/BOE-261-2002"/>
    <fo:legalRef rdf:resource="http://official-pub/BOE-273-2002"/>
    <fo:provided_by rdf:resource="http://cidem.gencat.es/Administraci? Central_
    <rdf:type rdf:resource="http://www.esperonto.net/fundFinder/fundFinder/fundin
  </rdf:Description>
  <rdf:Description rdf:about="http://cidem.gencat.es/ajut_548">
    <fo:title xml:lang="es">
      Cinqu? programa comunitari d'acci? per a la igualtat d'oportunitats entre hom
    </fo:title>
```

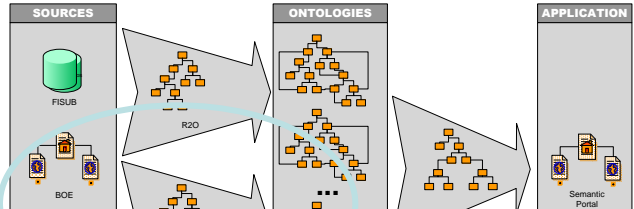
- ConceptMap-Def element: Describes how instances of a concept are extracted from the database.

```
<conceptmap-def name="http://net.esperonto/fundfinder#fundingOpp">  
  <uri-as> ...Description of a transformation to generate the instance uri...  
</uri-as>  
  <applies-if> ...Description of a condition to be verified...  
</applies-if>  
  <described-by> ...Set of attribute descriptions...  
</described-by>  
</conceptmap-def>
```

- AttributeMap-Def element: Describes how values for an attribute are extracted from the database.

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





4.2.1) Documentación administrativa:

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

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

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

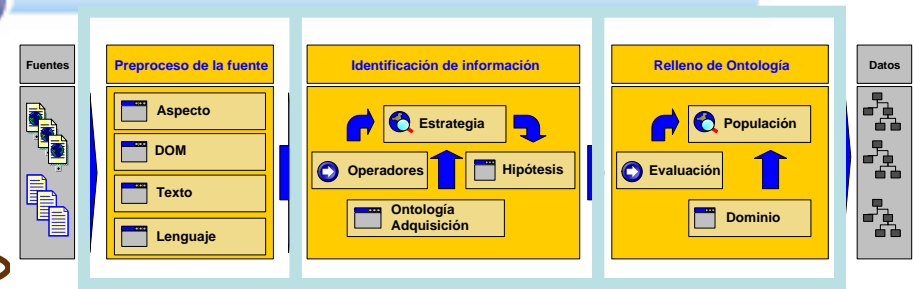
d) **Original o copia con el carácter de auténtica o fotocopia compulsada de los Estatutos debidamente legalizados.**

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

Recibo del año anterior a la convocatoria del Impuesto sobre actividades económicas, o en su caso, exención concedida por el órgano competente. Certificaciones administrativas con el carácter de positivas expedidas por las Administraciones correspondientes de la Agencia Estatal de Administración Tributaria y de la Tesorería Territorial de la Seguridad Social, de conformidad con lo establecido en el artículo 81 de la Ley General Presupuestaria y en su caso, exenciones del Impuesto del Valor Añadido y del Impuesto de Sociedades.

```

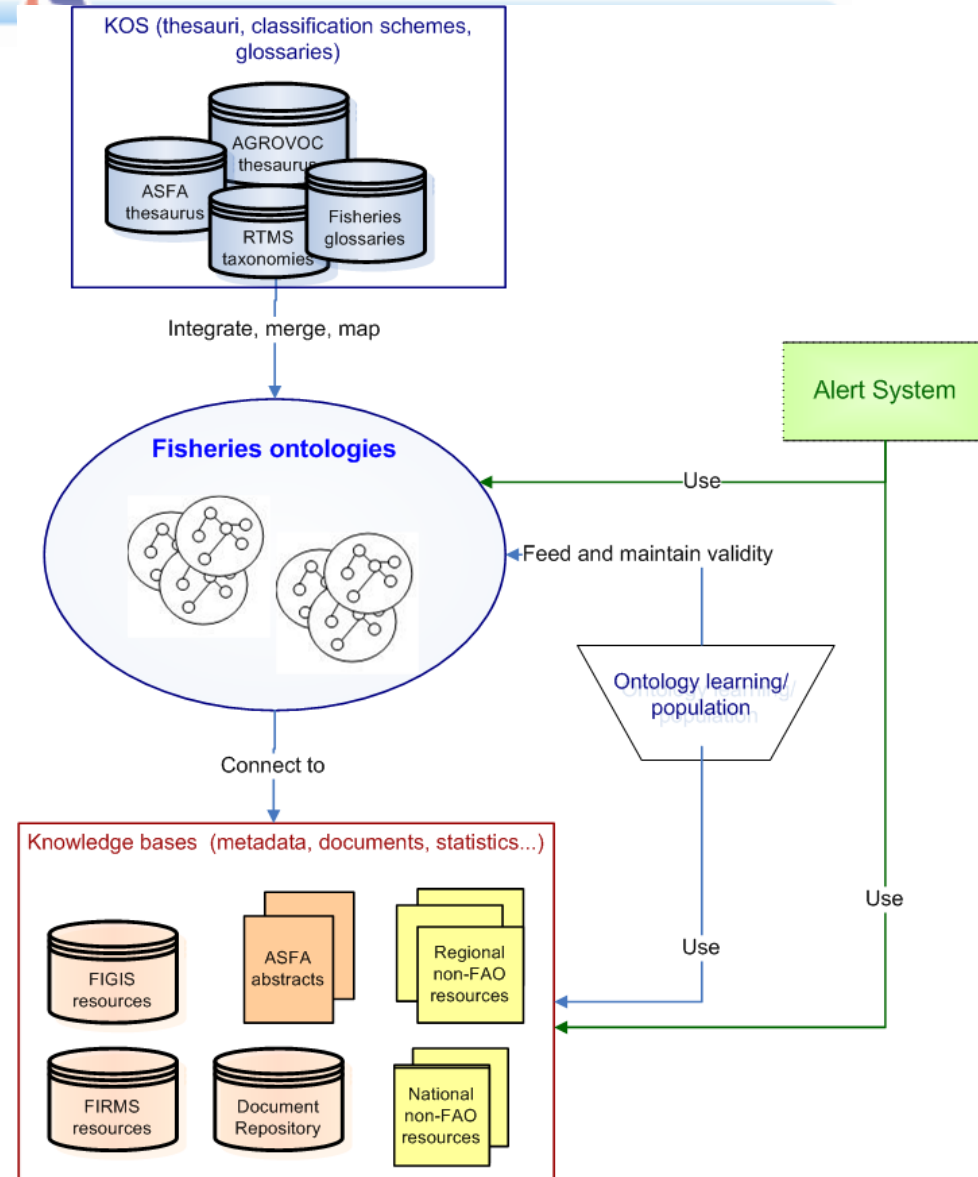
<Funding_Opportu:DocumentationItem
  rdf:about="http://protege.stanford.edu/tmp_namespace#tmp_Instance_10032" rdfs:label="BOE">
  <tmp_namespace:Description>Fotocopia compulsada de la tarjeta de identificación fiscal de la
    Entidad.</tmp_namespace:Description>
  <Funding_Opportu:isNeededToApplyFor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





# FAO Scenario

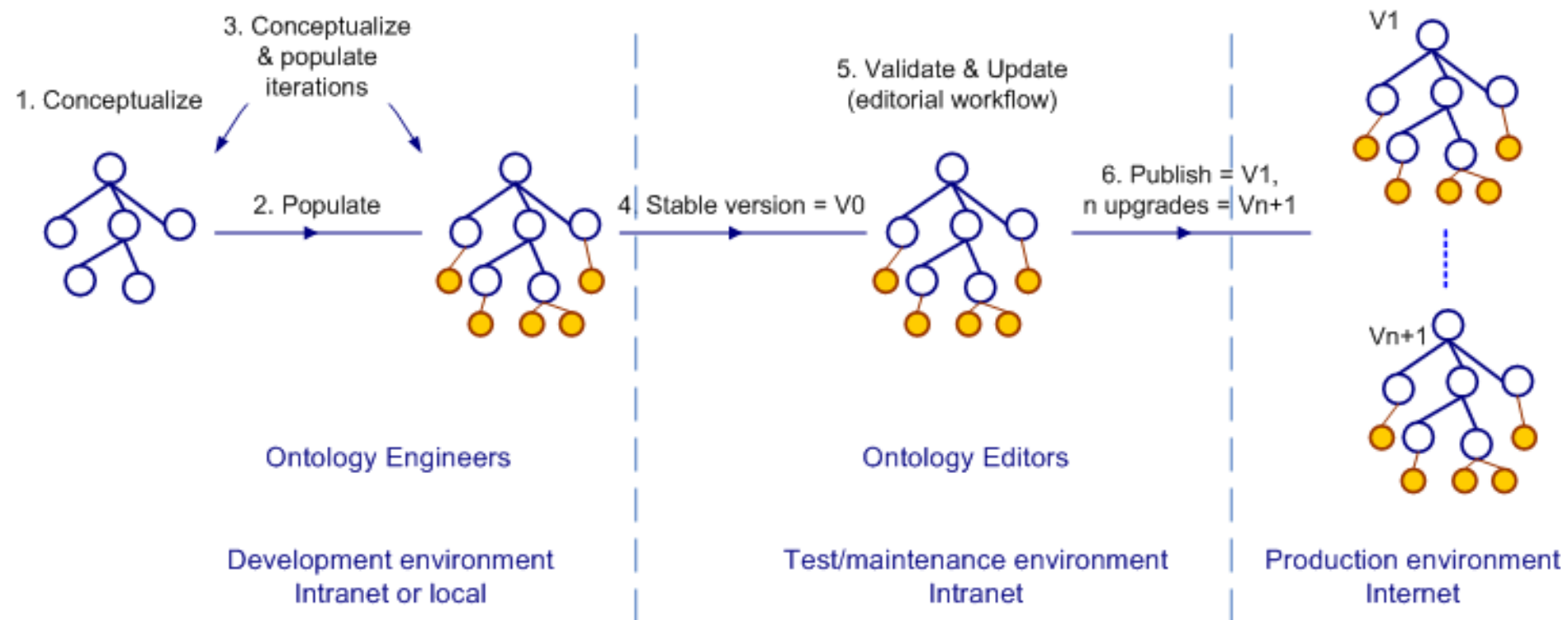
The ontologies produced with this framework will be used by the Food and Agriculture Organization of the United Nations (FAO) in many different large applications such the *Fisheries Stock Depletion Assessment System*.





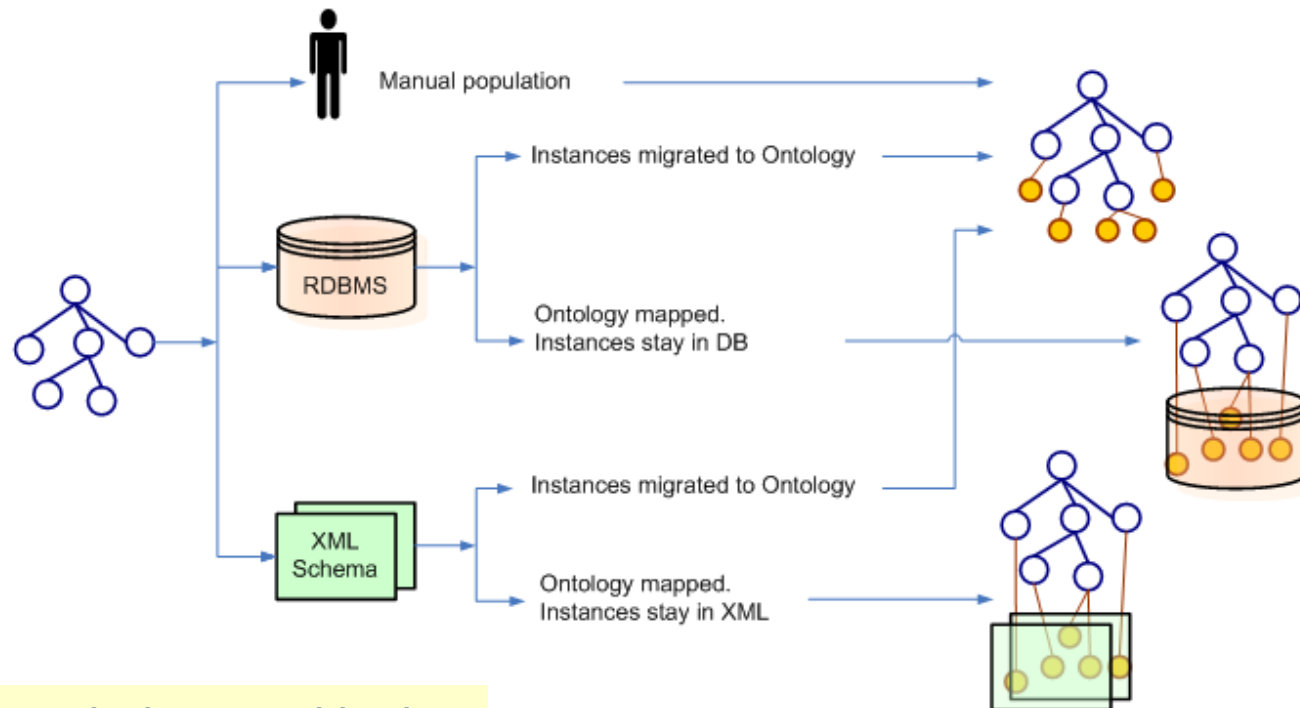
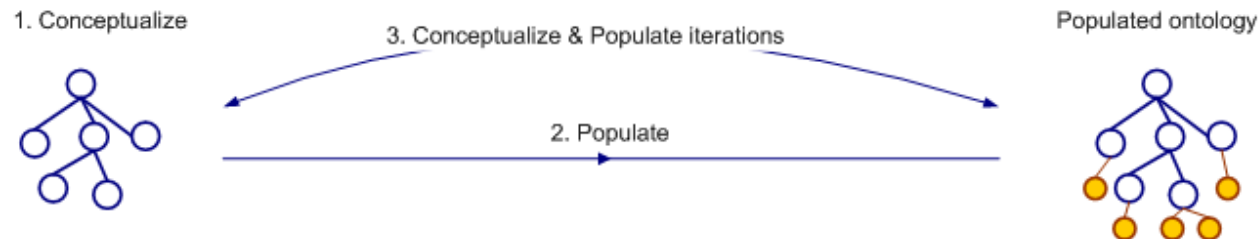
# Fisheries Ontologies Lifecycle

## Fisheries Ontologies Lifecycle



## 2. Ontology Population

Ontology Engineer - major tasks



R2O and X2O are being used in the Ontology population activity

# FAO Case study



Land  
areas



Fishing  
areas



Biological  
entities



Fisheries  
commodities



Vessel types  
and size



Gear  
types



$R_2O$   
Document

$R_2O$   
Document

$R_2O$   
Document

$R_2O$   
Document

$R_2O$   
Document

$R_2O$   
Document



FAO  
FIGIS DB

Land areas	
Concepts	4
Properties	25
Instances	289

Fishing areas	
Concepts	5
Properties	14
Instances	134

Biological entities	
Concepts	5
Properties	21
Instances	11571

Fisheries commodity	
Concepts	5
Properties	14
Instances	1380

Vessel types and size	
Concepts	5
Properties	20
Instances	120

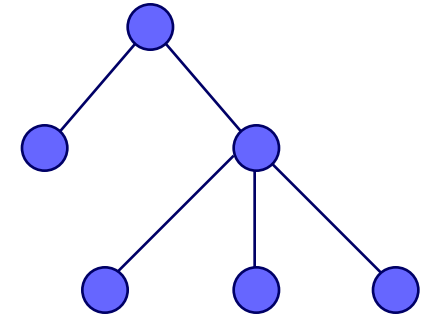
Gear types	
Concepts	4
Properties	14
Instances	0



# Query Driven (on demand process)



FAO Ontologies



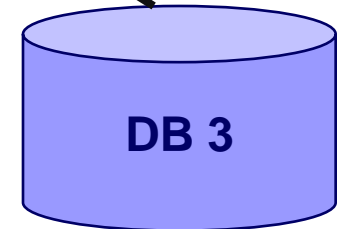
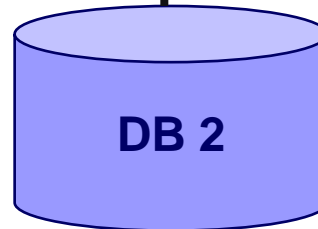
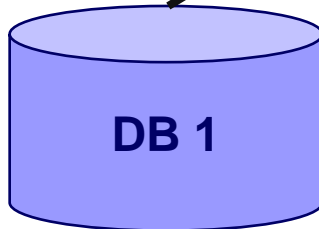
Client



*R<sub>2</sub>O Document*

*R<sub>2</sub>O Document*

*R<sub>2</sub>O Document*



# Conclusions



- Mappings Onto-BD
  - Estudio y caracterización de las diferentes situaciones en el establecimiento de correspondencias entre ontologías y BD
  - Propuesta de un modelo para la definición declarativa de dichas correspondencias
  - Implementación de dicho modelo en un lenguaje formal, declarativo y suficientemente expresivo para describir situaciones complejas : **R<sub>2</sub>O**
  - Definición de un procesador capaz de llevar a cabo la traducción de consultas entre modelos: Procesador **ODEMapster** & Lenguaje **ODEMQL**
  - Modos de ejecución: Dirigido por las consultas y migración masiva