



# *Multilingualism in Linked Data*



G. Aguado



J. Gracia



A. Gómez-Pérez



E. Montiel-Ponsoda



D. Vila

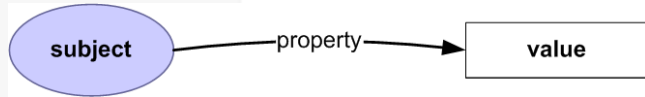
Ontology Engineering Group (OEG)  
Artificial Intelligence Department  
Universidad Politécnica de Madrid (UPM)

**W3C Multilingual Web Workshop**  
**Rome, 12-13 March 2013**



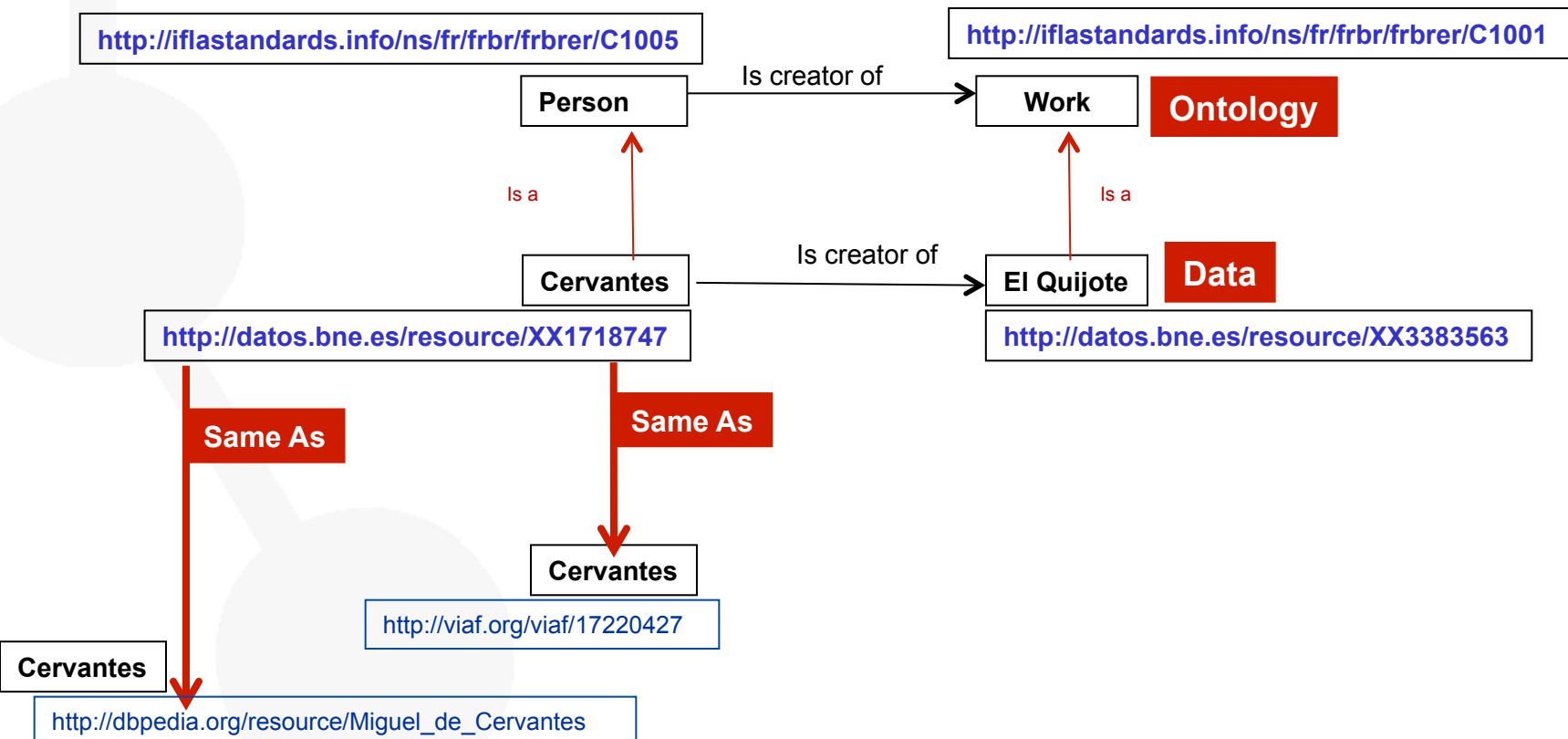
# Foundations: the model, the data, URIs and links

## RDF(S) models (ontologies) and data



**Unique identifiers: URI**  
identify or name a resource

**Equivalence links to other datasets**  
**Same As**



# Sources of information in different languages



## RDF Generation and Linking



## Visualization



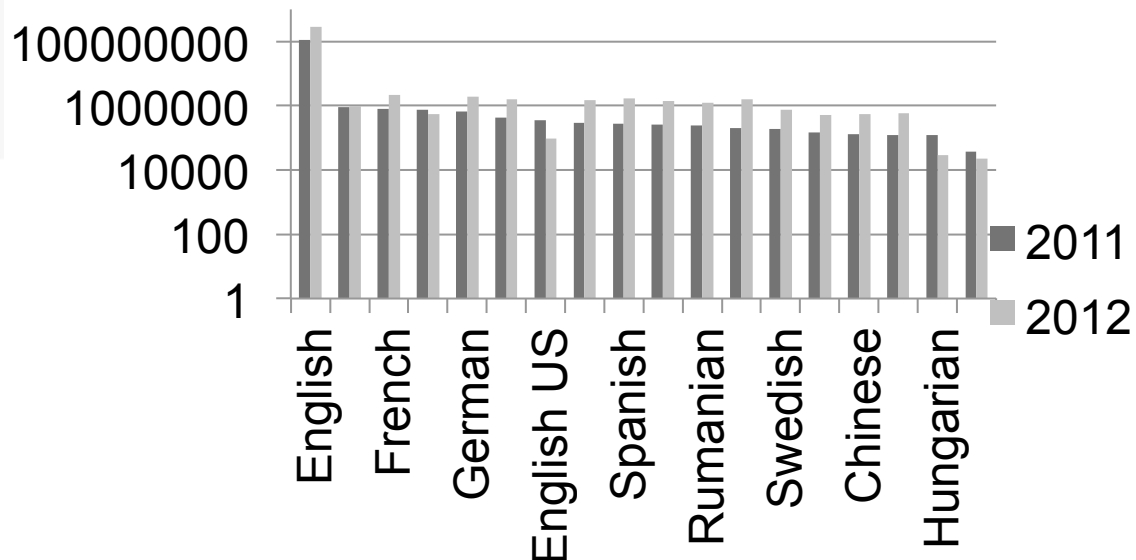
- Analysis of BTC datasets

## 2011

- Analyzed literals: **1,072,386,405**
- Total literals with lang tag: **116,058,734**
- % Literals with lang tag: **10.822 %**
- % Literals tagged as English: **94.68 %**

## 2012

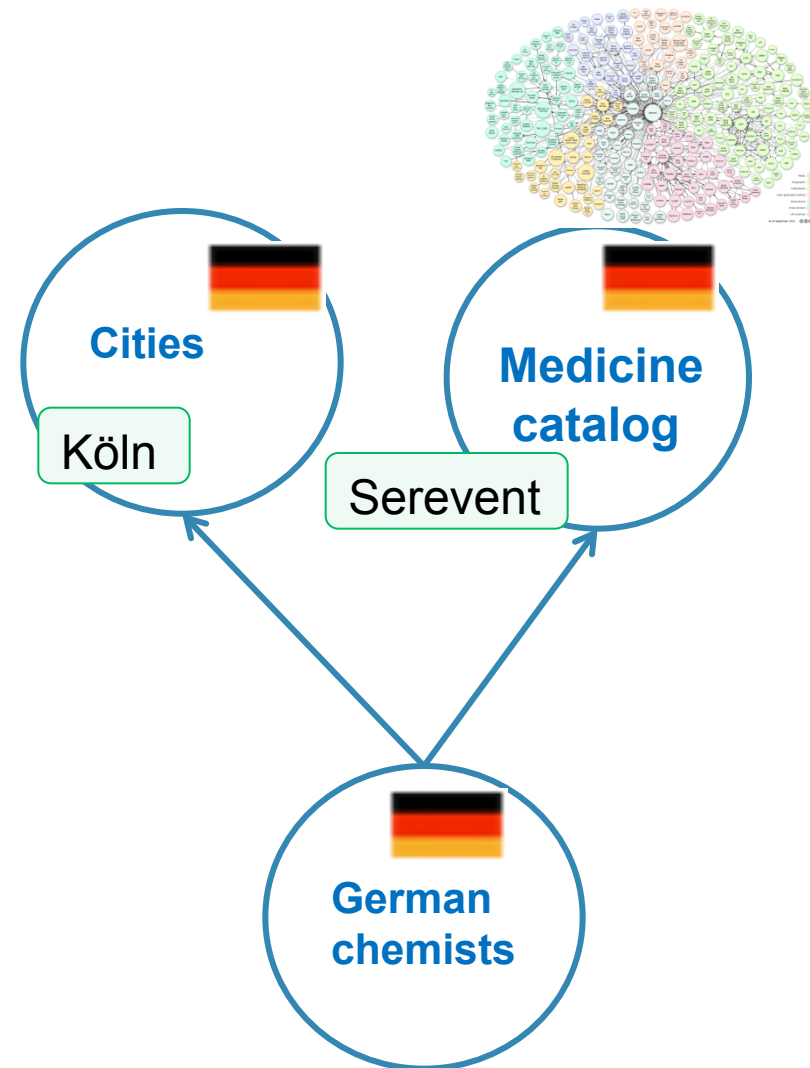
- Analyzed literals: **543,933,327**
- Total literals with lang tag: **304,115,676**
- % Literals with lang tag: **55.91 %**
- % Literals tagged as English: **94.44 %**



# A motivating example for using multilingual LD [1]



(\*) *Give me the duty chemists in Cologne having Beglan*



[1] J. Gracia, E. M. Ponsoda, P. Cimiano, A. G. Pérez, P. Buitelaar, and J. McCrae, "Challenges for the multilingual Web of Data," Journal of Web Semantics

# Multilingualism and the Linked Data Process [2]

Specification

- **Monolingual or multilingual data resources**
  - DB, documents, tables, etc.
  - Linguistic resources: Dictionaries, Lexicons, Thesauri, etc.

Modelling

- **Ontology (TBox URIs)**  
<http://phenomenontology.linkeddata.es/ontology/Municipio>  
<http://iflastandards.info/ns/fr/frbr/frbrer/C1005>

RDF Generation

Links Generation

- **Data (ABox URIs)**  
<http://geo.linkeddata.es/resource/Municipio/Madrid>  
<http://datos.bne.es/resource/XX1718747>

Publication

Exploitation

[2] Villazón-Terrazas, B. et al., Methodological Guidelines for Publishing Government Linked Data. In D. Wood, ed. Linking Government Data. Springer.

# Multilingualism and the Linked Data Process

How can we *adapt* and *translate* the lexical/terminological layer of an existent ontology into other languages?

Specification

Modelling

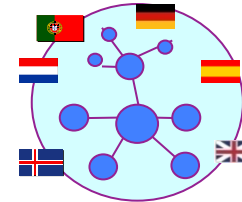
RDF Generation

Links Generation

Publication

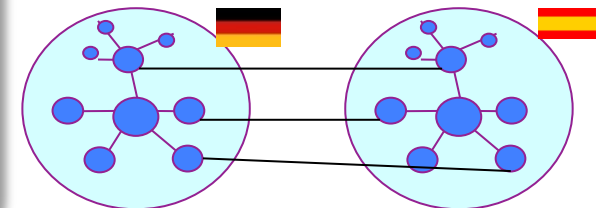
Exploitation

**Multilingual labeling** approach if languages involved share a single view on a certain domain



**Ontology Localization Algorithms**

**Cross-lingual linking** approach if independent monolingual ontologies exist that cover same or similar subject domain (Problems: conceptualization mismatches, or granularity and viewpoint differences)



**Cross-lingual Mapping Algorithms**

# Multilingualism and the Linked Data Process

Specification

Modelling

RDF Generation

Links Generation

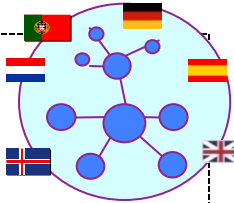
Publication

Exploitation

*How to represent multilingual Linked Data?*

- Traditional **annotation** properties for most cases

```
dbpedia:Miguel_de_Cervantes
  rdfs:label "Miguel de Cervantes"@es .
            "ミゲル・デ・セルバンテス"@ja .
            "미겔 데 세르반테스"@ko .
```



- Richer models** for more demanding applications

**SKOS-XL**

**lemon**

**LIR**

**LexInfo**



# Main issues of cross-lingual linking

Specification

Modelling

RDF Generation

Links Generation

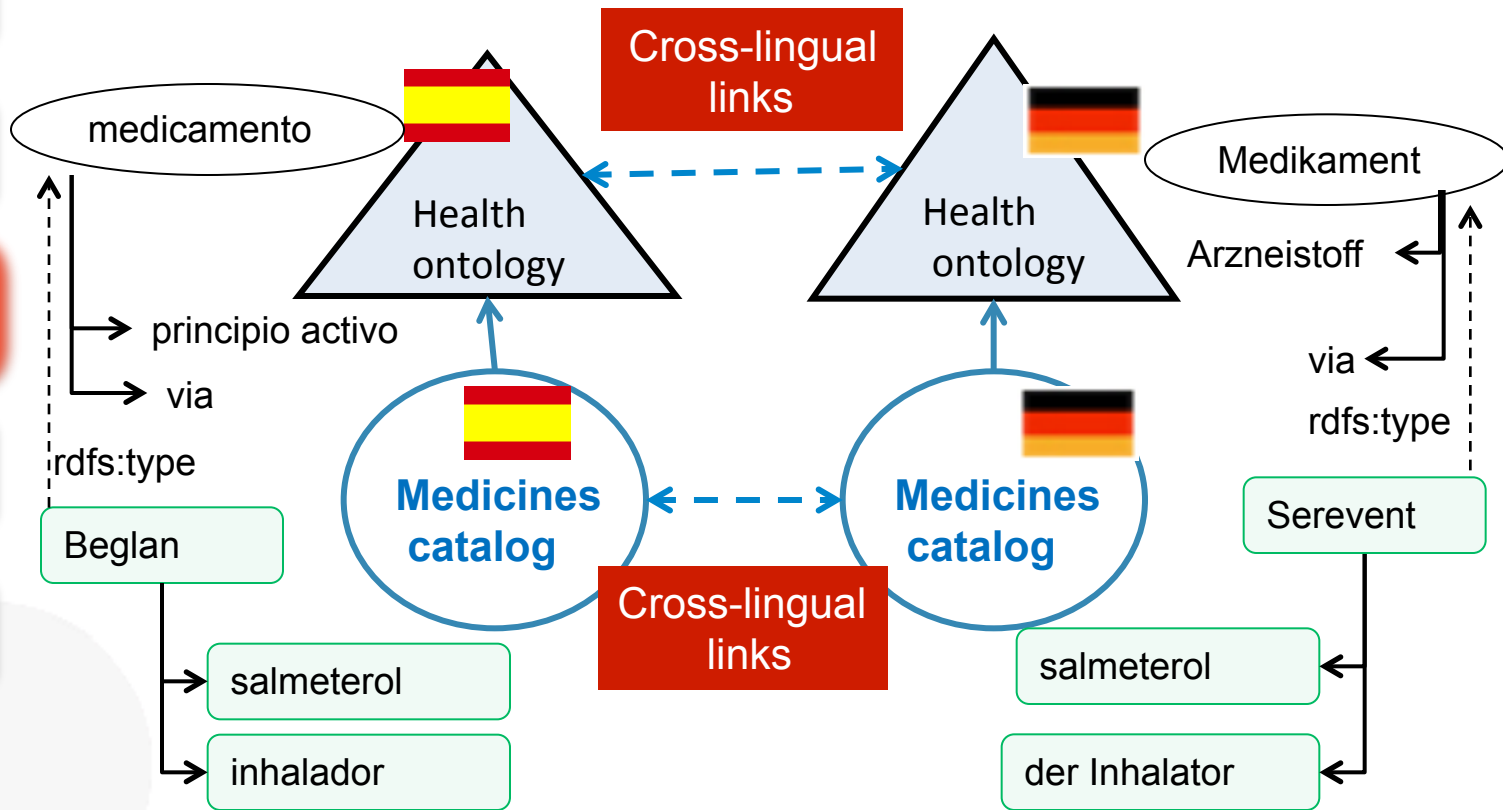
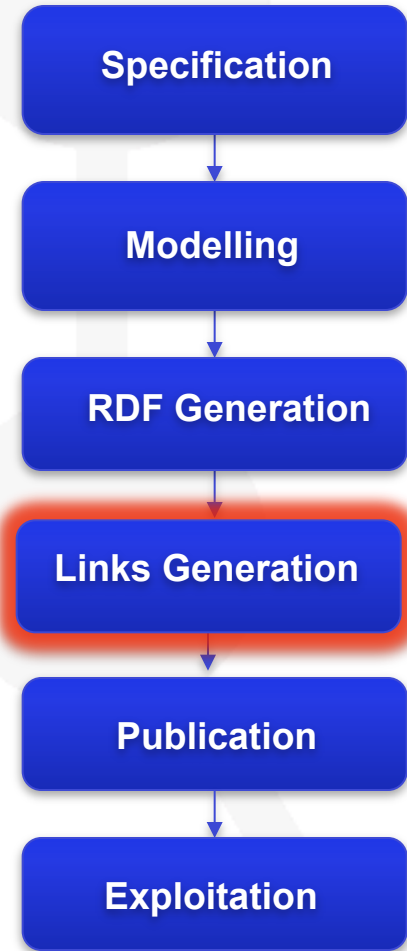
Publication

Exploitation

- How to **discover** cross-lingual links ?
- How to **represent** cross-lingual links?
- How to **store** and reuse cross-lingual links?

# Multilingualism and the Linked Data Process

How to *discover correspondences* between ontologies and between LD expressed in different natural languages?



# Cross-lingual Link Discovery

Specification

Modelling

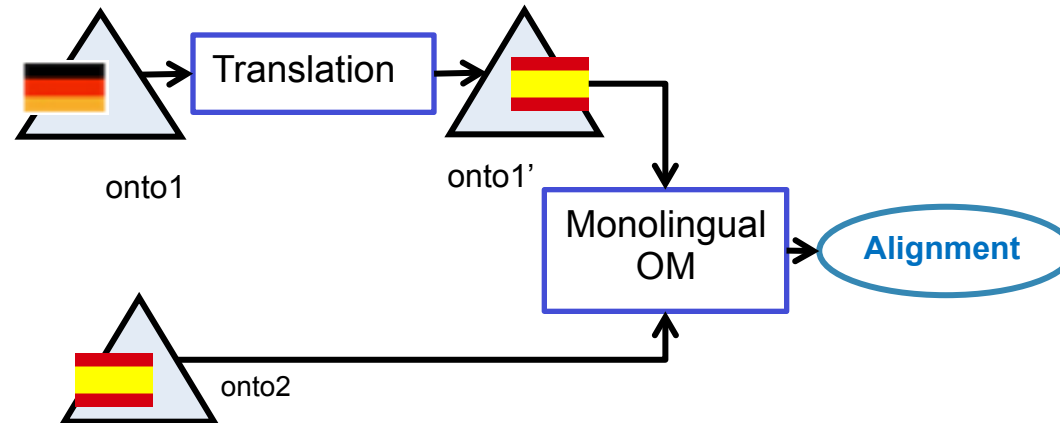
RDF Generation

Links Generation

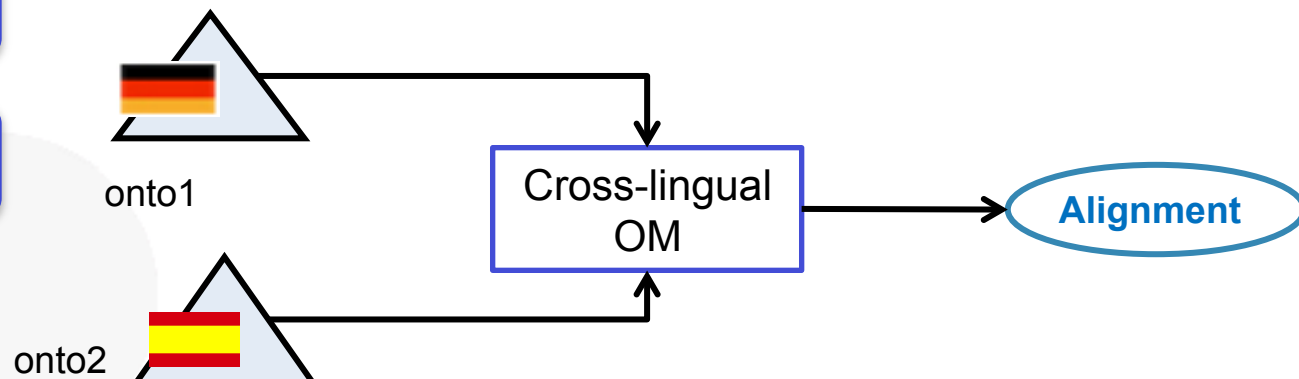
Publication

Exploitation

1. Projecting lexical content of the ontology into a **common language**, then applying traditional OM techniques



2. Comparing ontology entities directly by means of **cross-lingual semantic measures** (see CIDER-CL)



Specification

Modelling

RDF Generation

Links Generation

Publication

Exploitation

- Links can be discovered:
  - runtime -> need of scalable techniques
  - offline -> need of storage methods
- Storage
  - Following Linked Data principles
  - Links can be stored **jointly** to some of the data sources that they relate (e.g., during LD generation)
  - Links can be stored in **separate repositories** to be accessed by semantic applications (e.g., for CL-Question Answering)

# Multilingualism and the Linked Data Process

How can a user pose *questions* in their own language to be processed against the web of Linked Data?

Specification

Modelling

RDF Generation

Links Generation

Publication

Exploitation



“Colonia”

“farmacia”

Semantic query

1. Multilingual query interpretation
2. Query federation, ...



How should the results of a semantic query be *adapted* to the linguistic and cultural background of a user?

1. Adaptation and localization of user interfaces
2. Natural language generation
3. Presentation views to specific linguistic and cultural contexts

# Services for the Multilingual Web of Data



Users

Services for cross-lingual access



Services for cross-lingual linkage

Services for generating multilingual Linked Data

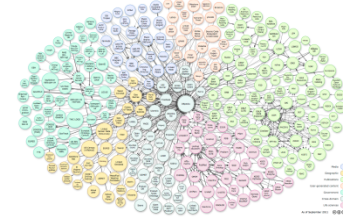
Data silos



Services for translation and ontology localization



Thanks for your attention !



- Ontology **lexica** representation  
Elena Montiel, Lupe Aguado
- **Lexico-syntactic patterns**  
Elena Montiel, Lupe Aguado
- Ontology **localisation** (translation)  
Elena Montiel, Jorge Gracia, Asun Gomez-Perez
- Exploratory **analysis** of the Multilingual Web of Data  
Daniel Vila, Asun Gómez-Pérez, Jorge Gracia
- Cross-lingual **ontology and Instance matching**  
Jorge Gracia, Daniel Vila
- Query **federation**  
Oscar Corcho