



monnet



# Challenges of the multilingual Web of Data and research agenda at OEG

**Jorge Gracia**

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

[jgracia@fi.upm.es](mailto:jgracia@fi.upm.es)

**Talk at OEG**  
**24/10/2012**

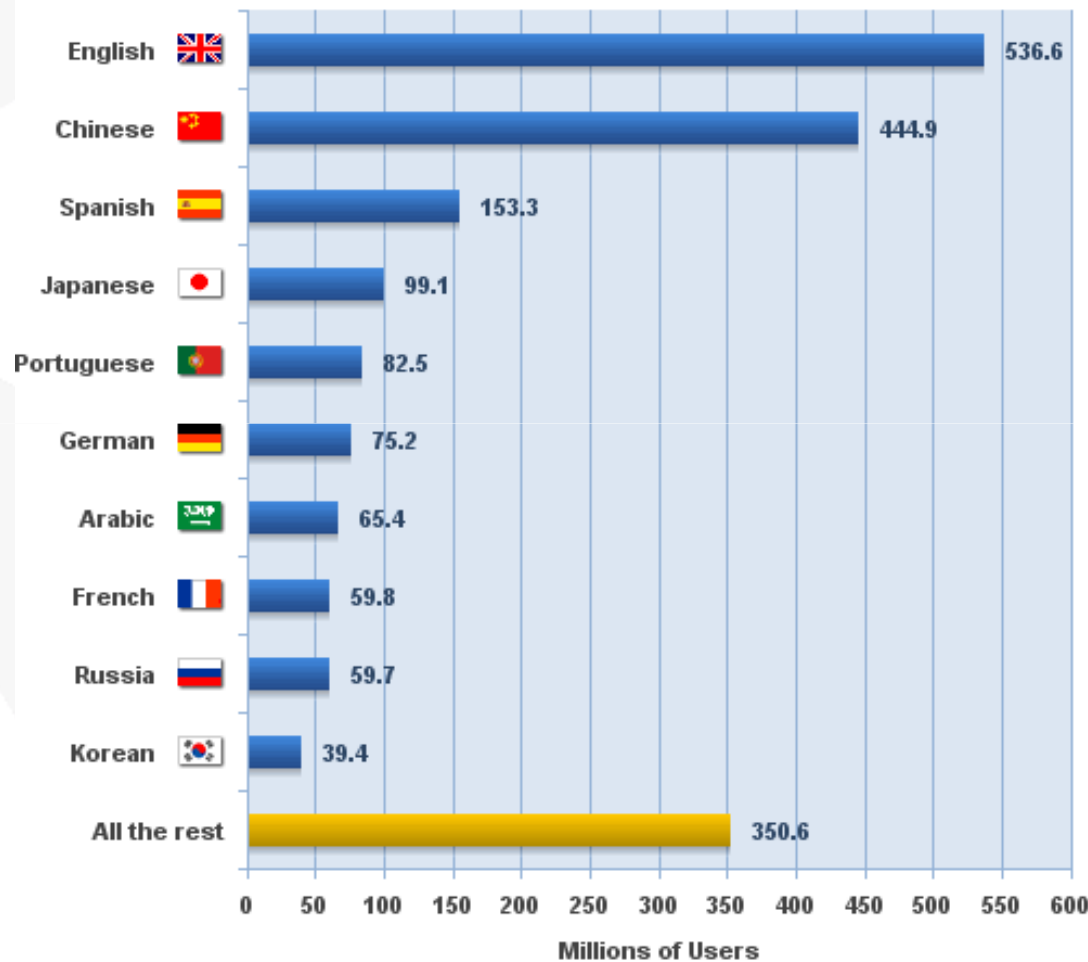


- Multilingualism and the Semantic Web
- Motivating example and research questions
- Architecture and research challenges
- Research agenda at OEG (on Multilingual Web of Data)

- Multilingualism and the Semantic Web
- Motivating example and research questions
- Architecture and research challenges
- Research agenda at OEG (on Multilingual Web of Data)

# The Web is multilingual

Top Ten Languages in the Internet  
2010 - in millions of users



Use of English on the Web: **26.8 %**  
(May 2011)

Source: Internet World Stats - [www.internetworldstats.com/stats7.htm](http://www.internetworldstats.com/stats7.htm)  
Estimated Internet users are 1,966,514,816 on June 30, 2010  
Copyright © 2000 - 2010, Miniwatts Marketing Group

# The Semantic Web is multilingual

## The **Semantic Web**...

- is highly **multilingual**

*495 different language tags in LOD <sup>(1)</sup>*

- although **language tags** are underused

*Semantic sources using language tags: 3% - 7% <sup>(3)(2)</sup>*

- Decreasing proportion of **English labels**

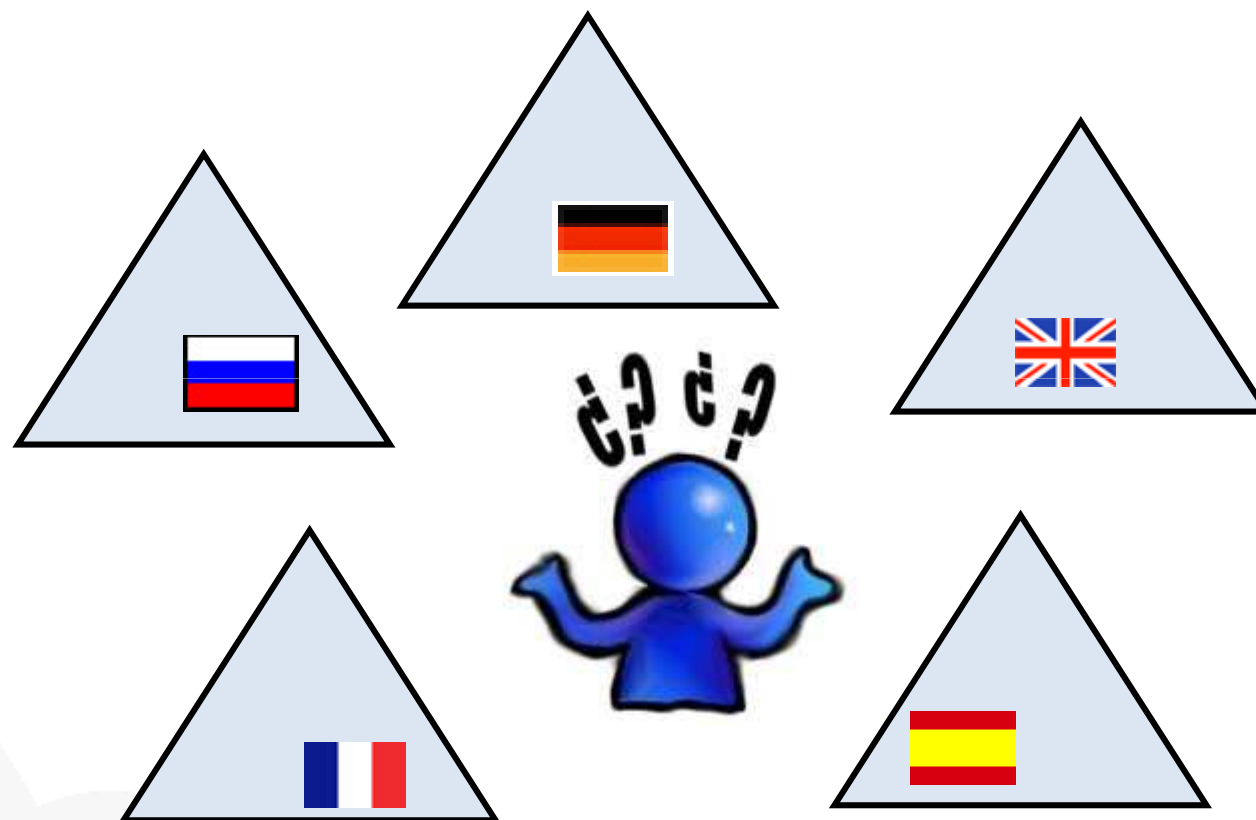
*80% in 2007 <sup>(2)</sup> → 45% in 2011 <sup>(3)</sup>*

(1) Source: LODStats <http://aksw.org/projects/LODStats> (June 2012)

(2) Source: Watson <http://watson.kmi.open.ac.uk/blog/2007/11/20/1195580640000.html> (Nov 2007)

(3) Source: B. Ell et al., "[Labels in the web of data](#)," in Proc. of ISWC'11, pp. 162-176.

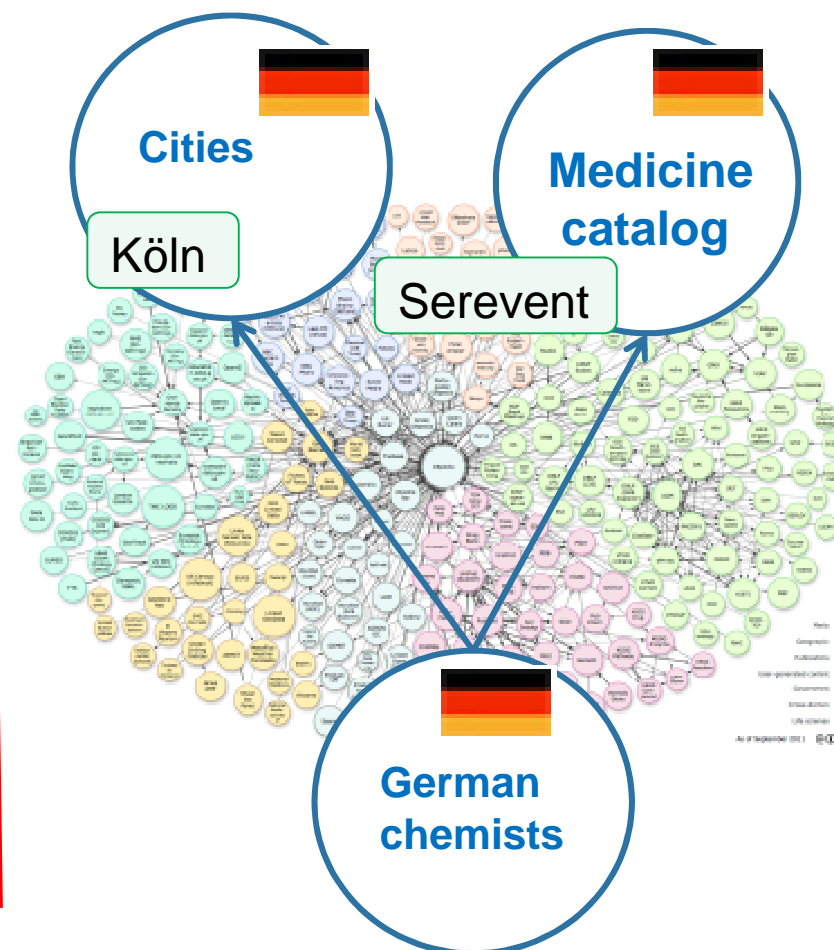
### Towards a Web of monolingual semantic “islands”?



*“How can universal access to the Web of Data be guaranteed to users, independent of the language they speak?” [1]*

- Multilingualism and the Semantic Web
- Motivating example and research **questions**
- Architecture and research challenges
- Research agenda at OEG (on Multilingual Web of Data)

## A motivating example

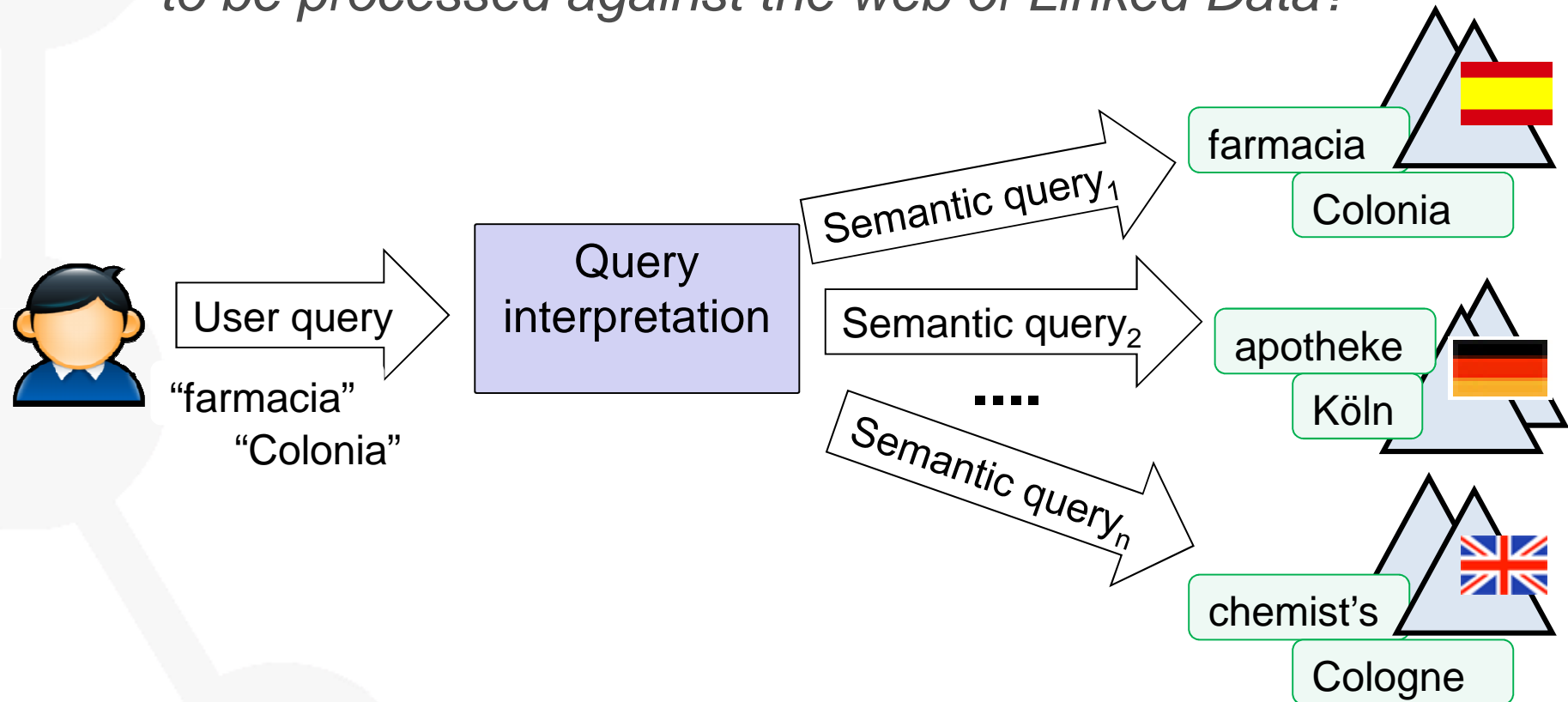


(\* ) Give me duty chemists in Cologne having Beglan



## Some research questions

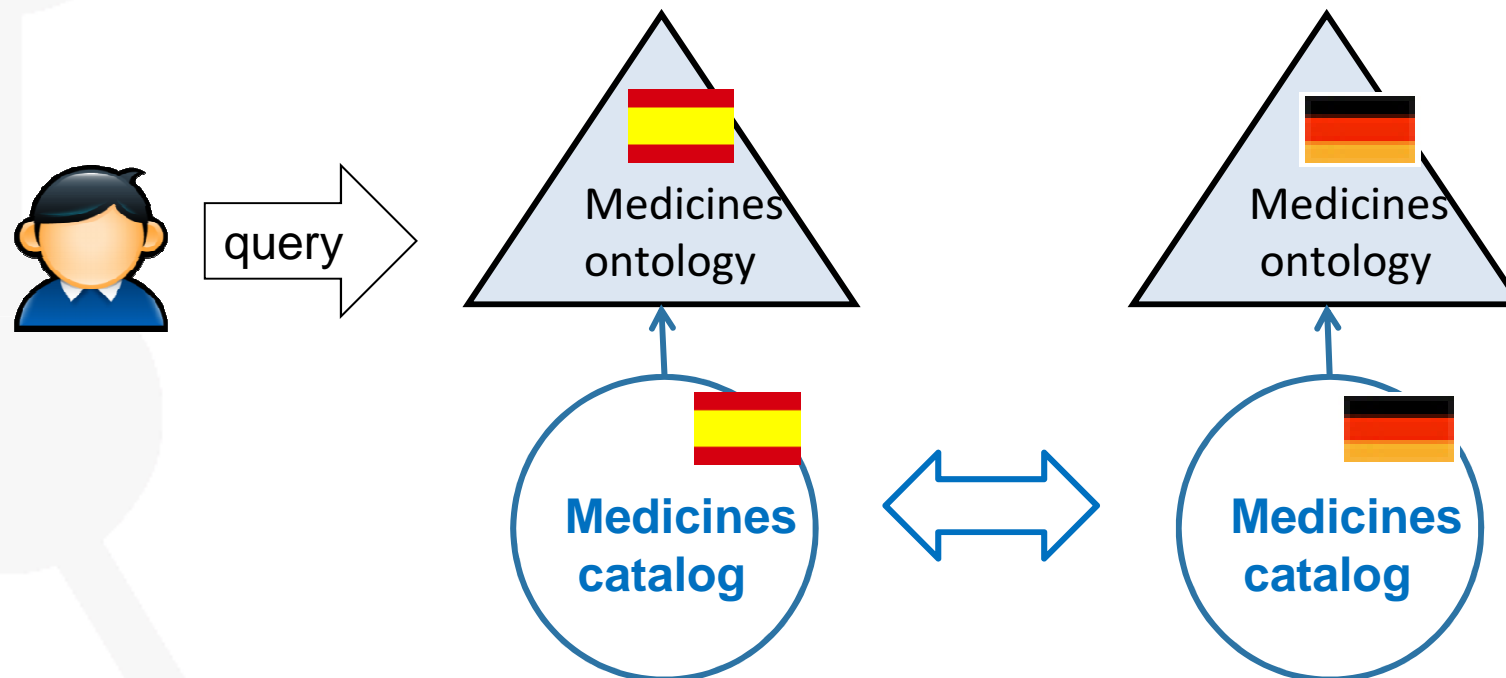
1. How can a user pose *questions* in their own language to be processed against the web of Linked Data?



- Research on multilingual *question answering*, multilingual *query interpretation*, query *federation*, ...

## Some research questions

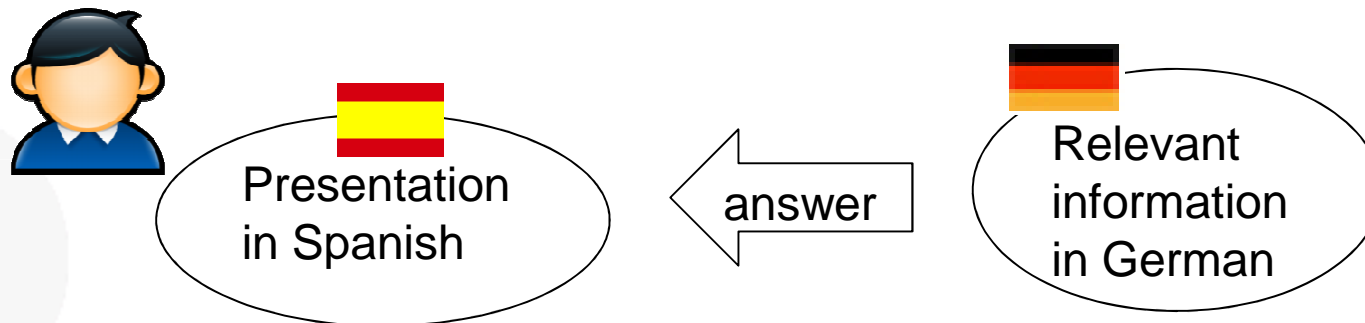
2. How can Linked Data be *delivered* to ontology-based applications localized to different languages?



- Research on **discovering relevant data** in other languages, and establishing **cross-lingual links** among sources in different languages

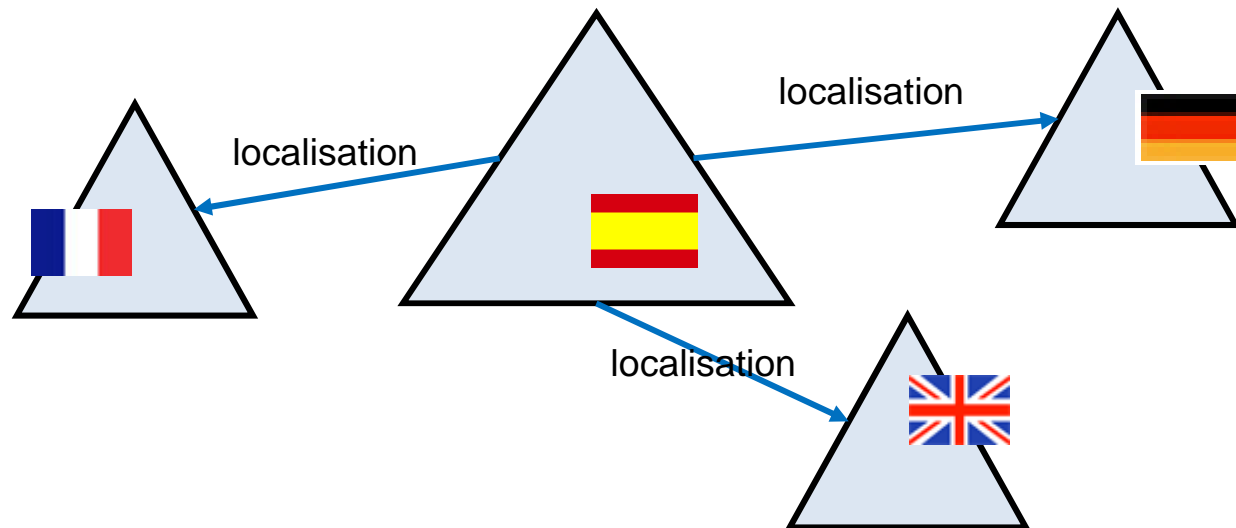
## Some research questions

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



- Research on **adaptation** and localization of **user interfaces** and presentation views to specific linguistic and cultural contexts

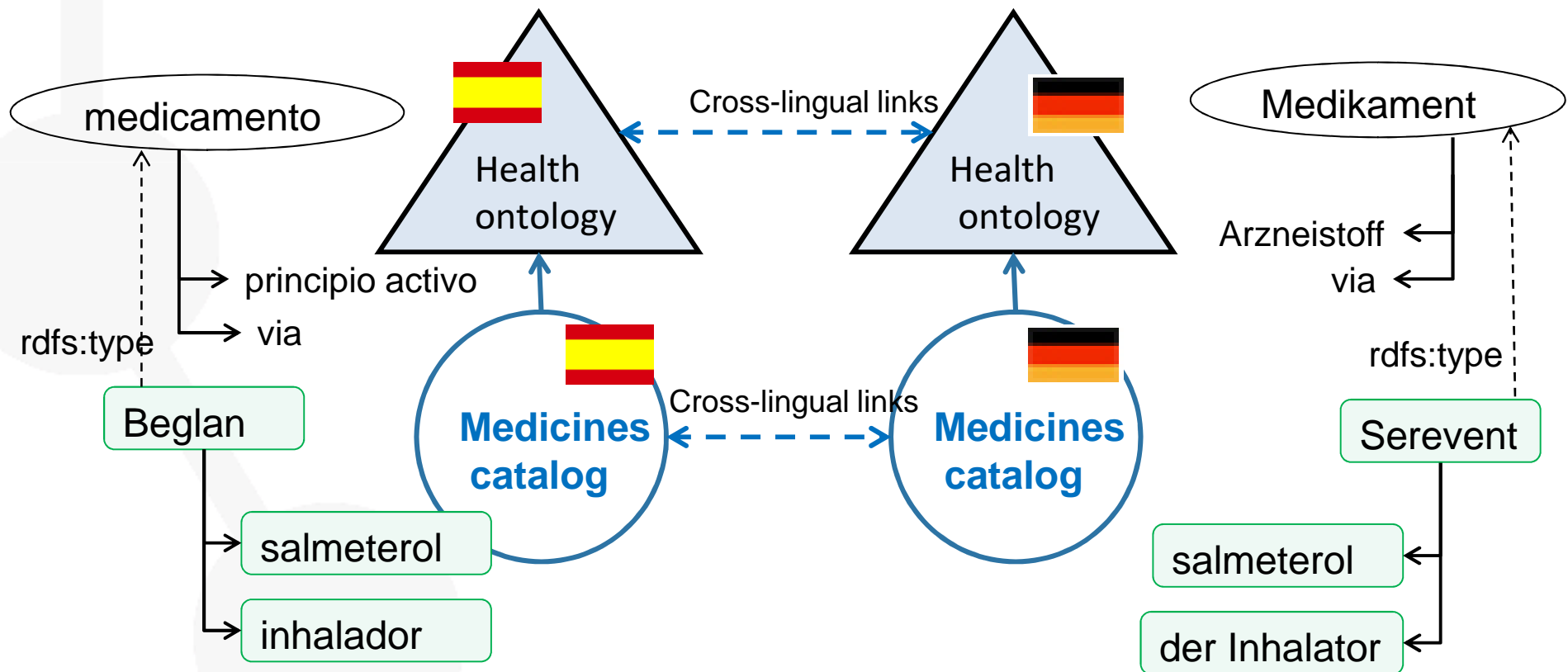
4. How can we *adapt* and *translate* the lexical/terminological layer of an existent ontology into other languages?



- Research on *ontology localisation*

## Some research questions

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



### 6. How to *represent* multilingual Linked Data?

- Traditional **annotation** properties for most cases
- **Richer models** for more demanding applications

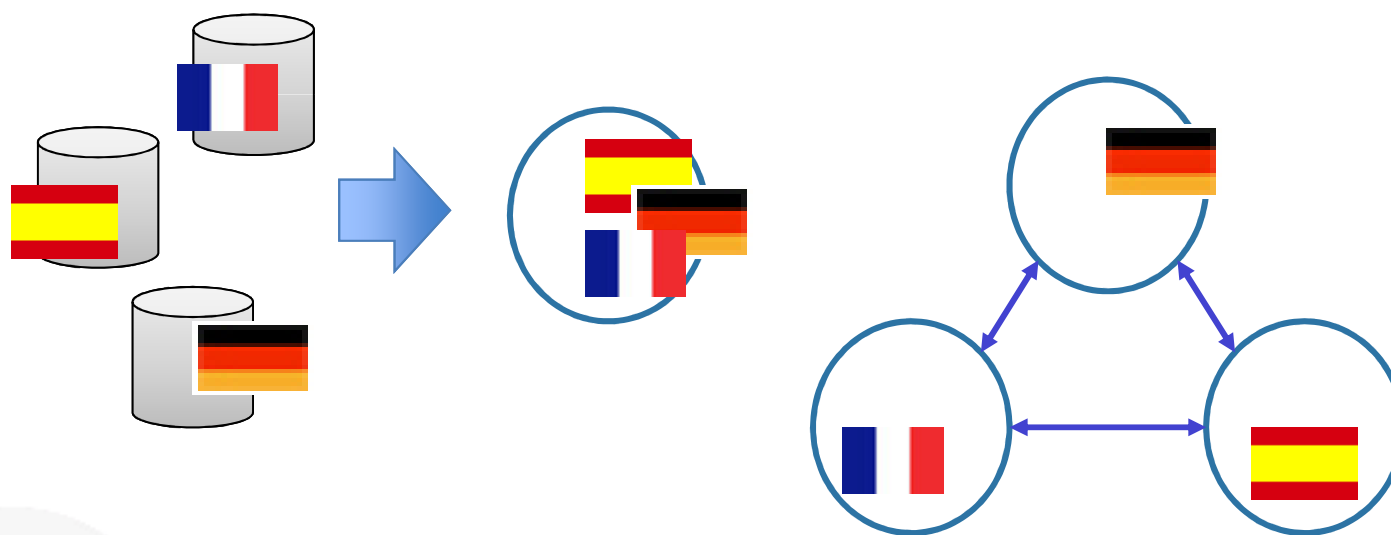
**RDF(S)**

**lemon**

**SKOS-XL**

### 7. How to *generate* multilingual Linked Data?

- Current methods have to be enriched to enable the **generation** of cross-lingual links, as well as the **storage** of multilingual lexical information

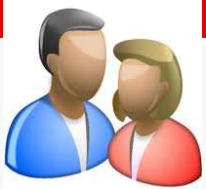


- Multilingualism and the Semantic Web
- Motivating example and research questions
- Architecture and research **challenges**
- Research agenda at OEG (on Multilingual Web of Data)





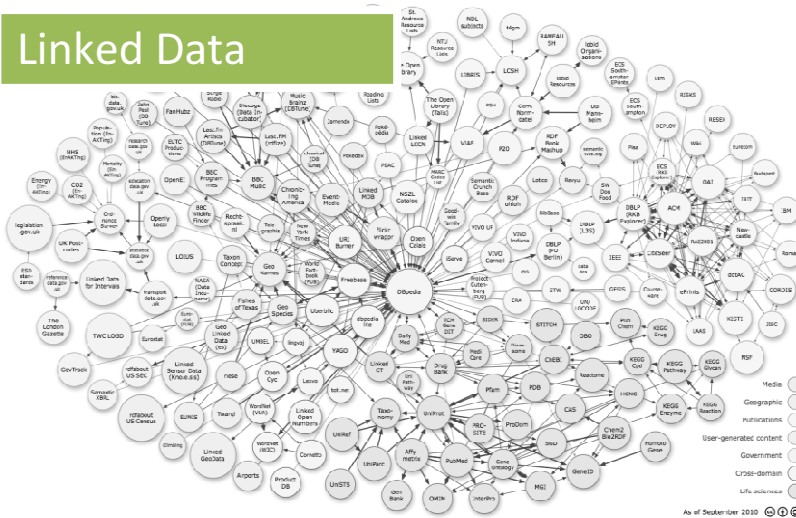
# The Multilingual Web of Data



## Users

## Services for cross-lingual access

## Linked Data



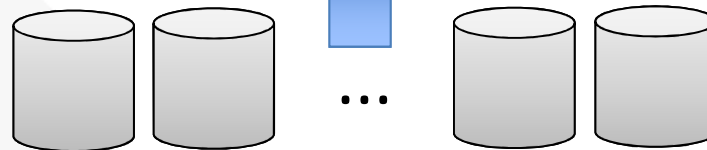
## Multilingual mappings

Multilingual  
linguistic  
information

## Services for cross-lingual linkage

## Services for generating multilingual Linked Data

## Data silos



## Services for translation and ontology localization

Several **challenges** to make this happen:

- Ontology **translation/localisation**
- Cross-lingual **ontology mappings**
- Representation of multilingual **lexical information**
- Cross-lingual **access** and **querying** of Linked Data

- Multilingualism and the Semantic Web
- Motivating example and research questions
- Architecture and research challenges
- Research **agenda** at OEG (on Multilingual Web of Data)

## Research agenda on Multilingual LD at OEG

- Ontology **lexica** representation
- Exploratory **analysis** of the Multilingual Web of Data
- Cross-lingual **ontology matching**
- Cross-lingual **instance matching**
- Ontology **localisation** & ontology translation



# Research agenda on Multilingual LD at OEG

- Ontology **lexica** representation

Elena Montiel, Lupe Aguado



lemon

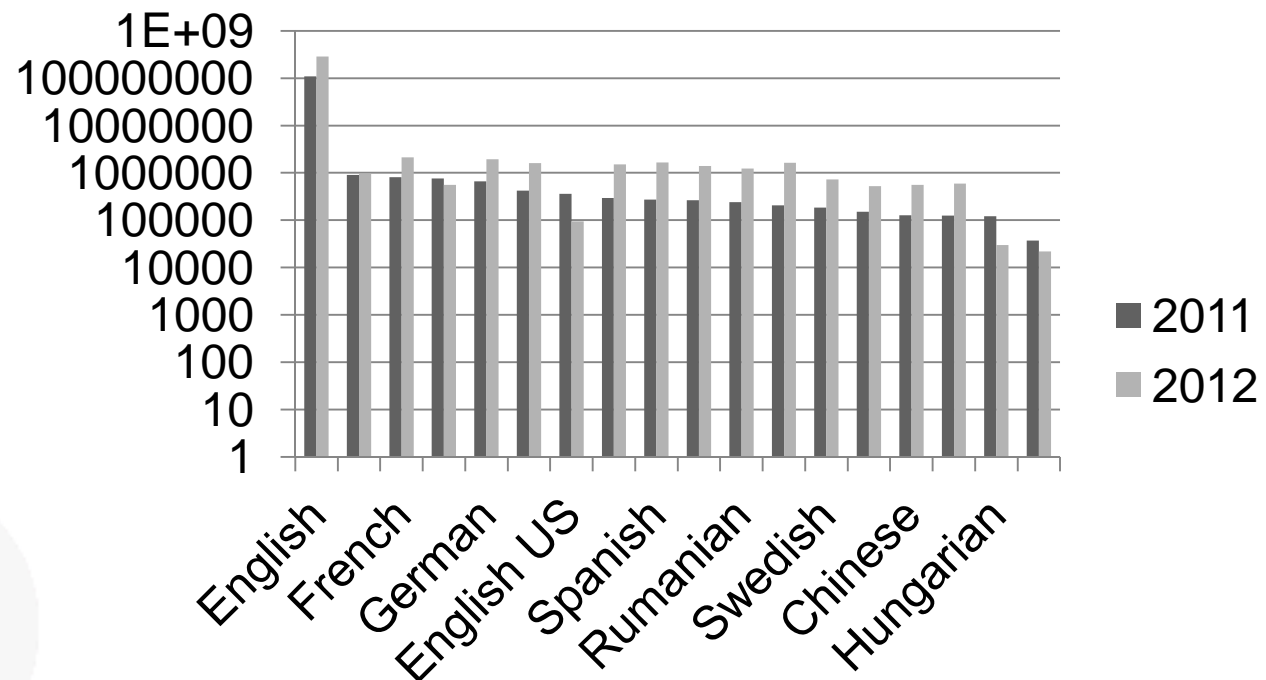


Collaboration with  
Bielefeld, DERI & DFKI

# Research agenda on Multilingual LD at OEG

## ■ Exploratory **analysis** of the Multilingual Web of Data

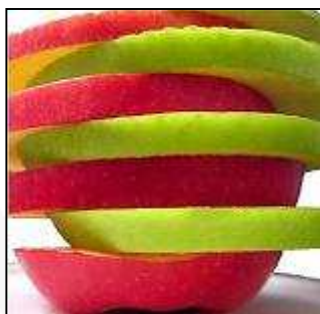
Daniel Vila, Asun Gómez-Pérez, Jorge Gracia



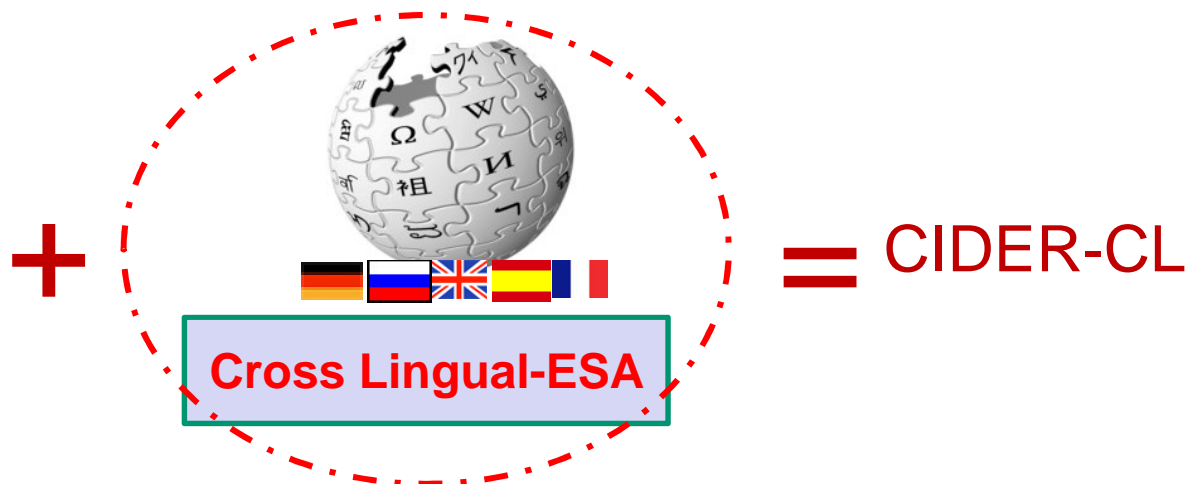
# Research agenda on Multilingual LD at OEG

## ■ Cross-lingual **Ontology** matching

Jorge Gracia, Kartik Asooja



CIDER



# Research agenda on Multilingual LD at OEG

## ■ Cross-lingual Instance matching

Daniel Vila, Jorge Gracia



Collaboration with





# Research agenda on Multilingual LD at OEG

## ■ Ontology **localisation** & translation

Elena Montiel , Mauricio Espinoza, Kartik Asooja



Collaboration with  
Bielefeld & DERI

# Research agenda on Multilingual LD at OEG

Some missing pieces:

- Cross-lingual **access** and **querying** (Q/A, query interpretation, semantic grounding, etc.)
- **Storage** of cross-lingual links
- Multilingual LD **generation**
- Multilingual data **visualization**
- Creation of an ecosystem of Multilingual LD **Web services**
- ...

- [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*, vol. 11, pp. 63-71, Mar. 2012.
- [2] J. McCrae, G. A. de Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gómez-Pérez, J. Gracia, L. Hollink, E. Montiel-Ponsoda, D. Spohr, and T. Wunner, "[Interchanging lexical resources on the semantic web](#)," *Language Resources and Evaluation*, vol. 46, 2012.
- [3] E. Montiel-Ponsoda, J. Gracia, G. A. de Cea, and A. Gómez-Pérez, "[Representing translations on the semantic web](#)," in *Proc. of 2nd Workshop on the Multilingual Semantic Web, at ISWC'11, Bonn, Germany, ISSN 1613-0073*, vol. 775. CEUR-WS, Oct. 2011, pp. 25-37

