



mylider

www.lider-project.eu



# Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources



*A. Gómez-Pérez*

*Universidad Politécnica de Madrid*

*asun@fi.upm.es*

**Acknowledgements: Jorge Gracia (UPM), Victor Rodríguez Doncel (UPM), LIDER Consortium members Nurial Bel (UPF) , Marta Villegas (UPF)**



# Context – Ontology Engineering Group

**Directors:** A. Gómez-Pérez, O. Corcho

**Position:** 8th in the UPM ranking (200 groups)

Founded: 1994

## ➤ **Research Group (30 people)**

### ➤ **Experience on**

1. Ontologies, Semantic Web, Linked Data, Open Data
2. Semantic E-science
3. Multilingualism

## ➤ **ODI Madrid : Madrid Node of the Open Data Institute**

### ➤ **Projects**

- 27 EU projects (7 as coordinator)
- 54 National Projects
- 27 contracts with companies

### ➤ **Standardization activities**

- >25 @ W3C, ISO, OASIS, etc.

### ➤ **Impact of publications H-index (scholar)**

- Asunción Gómez-Pérez (h:50, citations 14852)
- Oscar Corcho García (h: 36, citations 8152)

### ➤ **Services to the Spanish community**

- esDbpedia
- linkeddata.es
- vocab.linkeddata.es



<http://www.oeg-upm.net/>





<https://github.com/oeg-upm>



@oeg-upm

170+ Past Collaborators  
50+ Past Visitors

- This work is licensed under the Creative Commons Attribution – Non Commercial – Share Alike License
- You are free:
  -  to Share — to copy, distribute and transmit the work
  -  to Remix — to adapt the work
- Under the following conditions
  - Attribution — You must attribute the work by inserting
    - “[source <http://www.oeg-upm.net/>]” at the footer of each reused slide
- a credits slide stating: “Linguistic Linked Data: Paving the Way Towards Maximising (Re)Usability of Linguistic Resources ” by A. Gómez-Pérez ”
  - Non-commercial
  - Share-Alike









# Language Resources

Finding and reusing LR in third party applications is manual and time consuming



# Lack of interoperability of Language resources

- Ecosystem of
  - Open and Closed resources
  - Silos of LRs
  - Complementary resources
    - Lexicon, Corpora, Dictionaries, Grammars, ....
  - Heterogeneous formats
    - E.g, for Lexicons: Lexinfo, LMF, LIR, Lemon, ...
  - Several repositories with different metadata and schemas
  - Many APIs and services for querying



Discovery and reuse LR in third party applications is hard, manual and time consuming



# The Open Data stars



★ Make your stuff available on the Web (whatever format) under an open license

★ ★ Make it available as structured data (e.g., Excel instead of image scan or a table)

★ ★ ★ Use non-proprietary formats (e.g., CSV instead of Excel)

★ ★ ★ ★ **Use URIs to identify things, so that people can point at your stuff**

★ ★ ★ ★ ★ **Link your data to other data to provide context**

# Open data formats and publication

In Books

SOV The Web – Proprietary formats  
(Human & Machine readable formats)



As files on the Web following standards  
(XML, HTML, CSV, LMF, etc.)



REAL ACADEMIA ESPAÑOLA

La institución

Obras académicas

Biblioteca y Archivo

Consultas lingüísticas

Inicio » Recursos » Diccionarios » Diccionario de la lengua española

## Diccionario de la lengua española

El *Diccionario de la lengua española (DRAE)* es la obra de referencia de la Academia. La edición actual —la 22.<sup>a</sup>, publicada en 2001— incluye más de 88 000 entradas.



á é í ó ú ü ñ

[Ayuda](#)

**red.**

**Artículo enmendado**

(Del lat. *rete*).

1. f. Aparejo hecho con hilos, cuerdas o alambres trabados en forma de mallas, y convenientemente dispuesto para pescar, cazar, cercar, sujetar, etc.

2. f. Labor o tejido de mallas.

3. f. **redecilla** (ll prenda de malla para el pelo).

4. f. Lugar donde se vende pan y otras cosas que se dan por

A free/o

Macedonia

red

red.

(Del lat. *rete*).

1. f. Aparejo he forma de mallas cazar, cercar, s
2. f. Labor o tej
3. f. **redecilla** (
4. f. Lugar donc entre verjas.



Datasets Organizations About Blog Help Search

Organizations / Ontology Engineering Group ... / Apertium RDF EN-ES

Apertium RDF EN-ES

Followers  
0

Organization



Ontology Engineering Group (UPM)

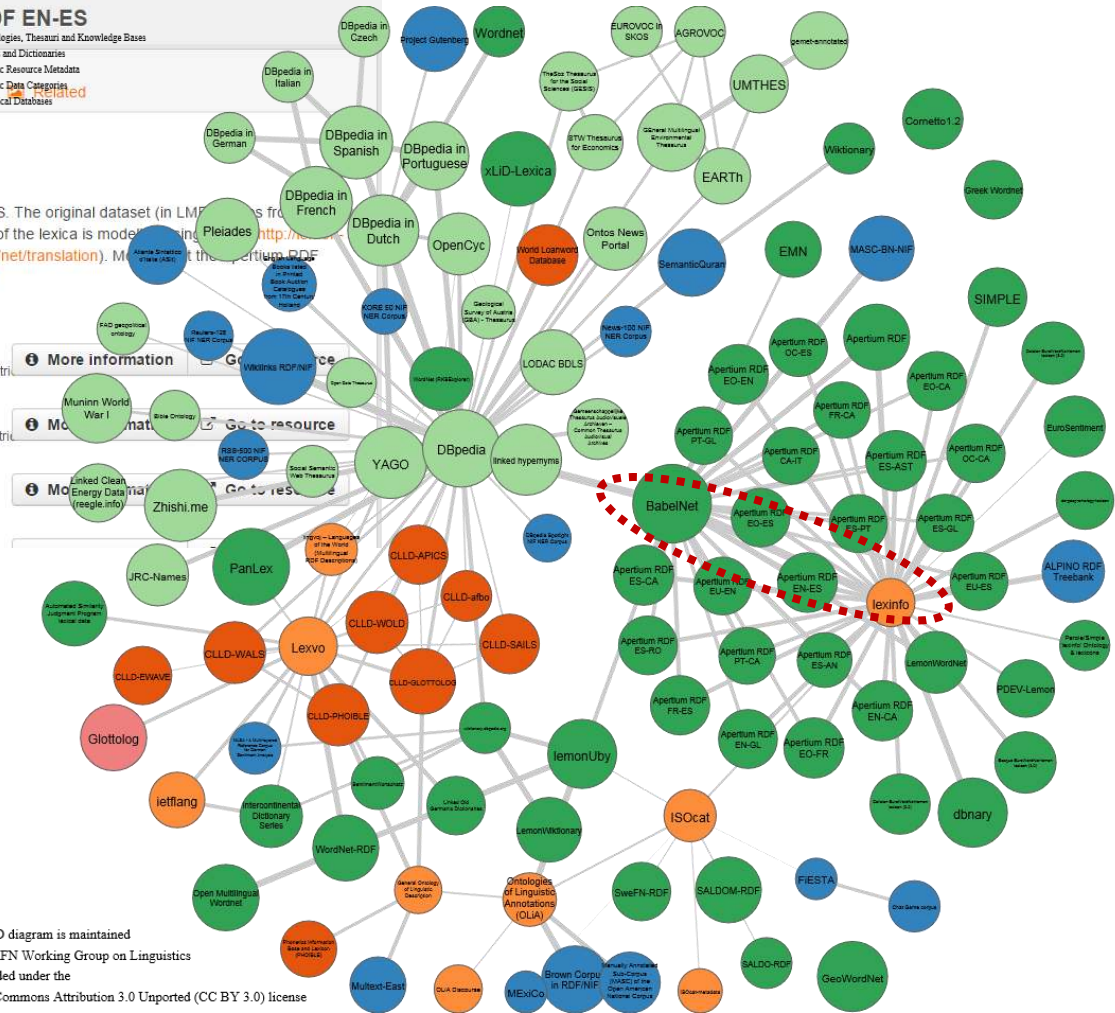
The Ontology Engineering Group (OEG) is based at the

Apertium RDF EN-ES

RDF version of the Apertium bilingual dictionary EN-ES. The original dataset (in LMF format) is from <http://hdl.handle.net/10230/17110>. The RDF version of the lexica is modelled using <http://model.net/> and the translation module (<http://purl.org/net/translation>). More information at the Apertium RDF dictionaries at <http://linguistic.linkeddata.es/apertium/>

Data and Resources

- English lexicon URI**  
This online lexicon not only contains the EN lexical entries...
- Spanish lexicon URI**  
This online lexicon not only contains the ES lexical entries...
- EN-ES translation set URI**  
No description for this resource
- SPARQL endpoint**



The LLOD diagram is maintained by the OKFN Working Group on Linguistics and provided under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license



# Language Resources

**THE NEED OF CONNECTING  
LICENSED LANGUAGE  
RESOURCES**



**“Red”**

**Pronunciation: [red]**

**Grammar category: sustantivo femenino**

**Singular: “red”**

**Plural: “redes”**



**“Red”**

**Etimologi y Del latin “rete”**

**Gender: “f”**

**Definition.: “Conjunto de ordenadores o de equipos informáticos conectados entre sí....”**

**Complementary but not connected**

Wikilengua del español

**“Red”**

**Norm: UNE 21302-131**

**English: network**

**German: Netzwerk**



**“Red”**

**Synonyms: “sistema”, “malla”, “distribución”**

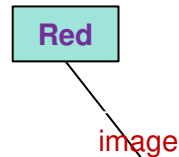
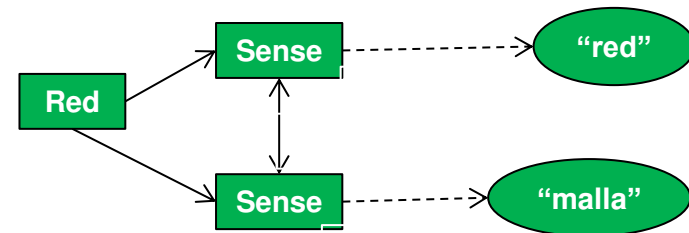
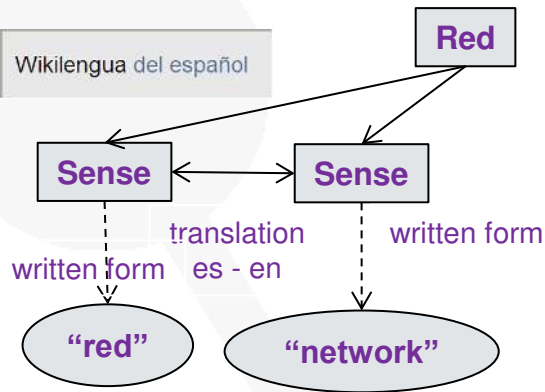
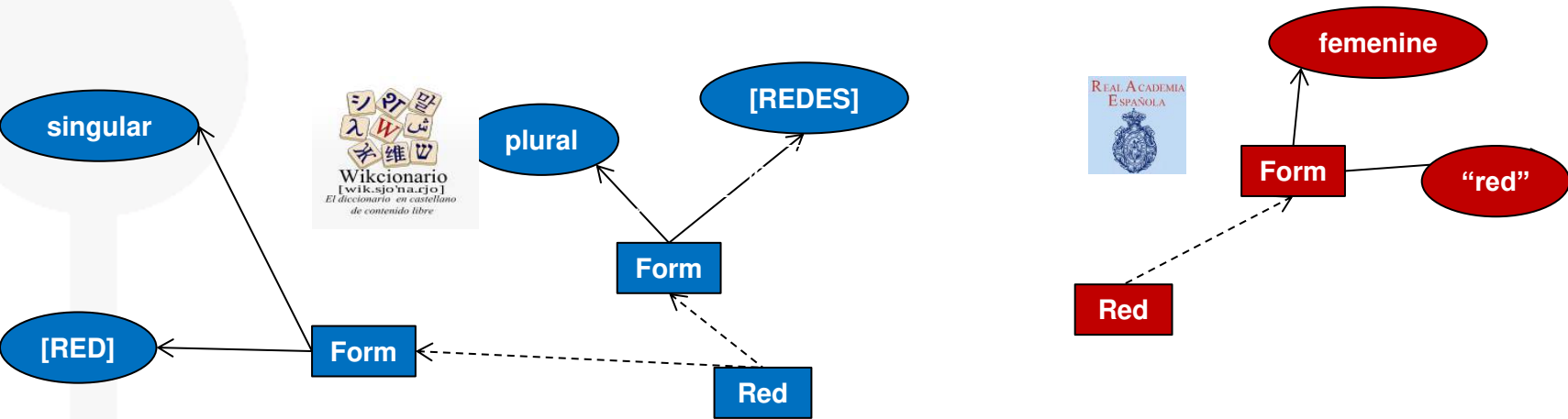
**“Red\_de\_computadores”**

**Category: redes informáticas**

**Image**



# Integration of Linguistic data



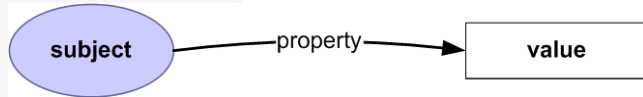


# Language Resources

## The need of connecting licensed LRs

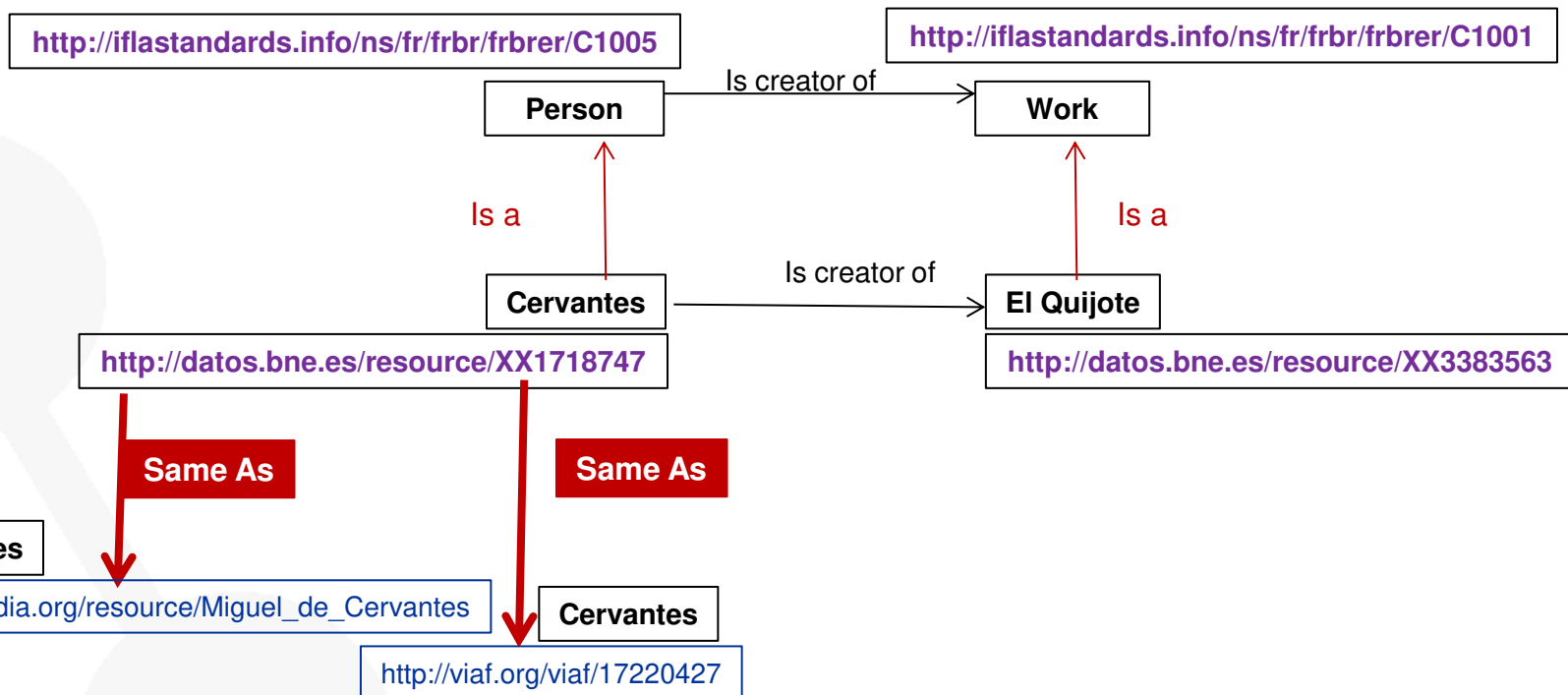
# LINKED DATA FOUNDATIONS

## RDF(S) models



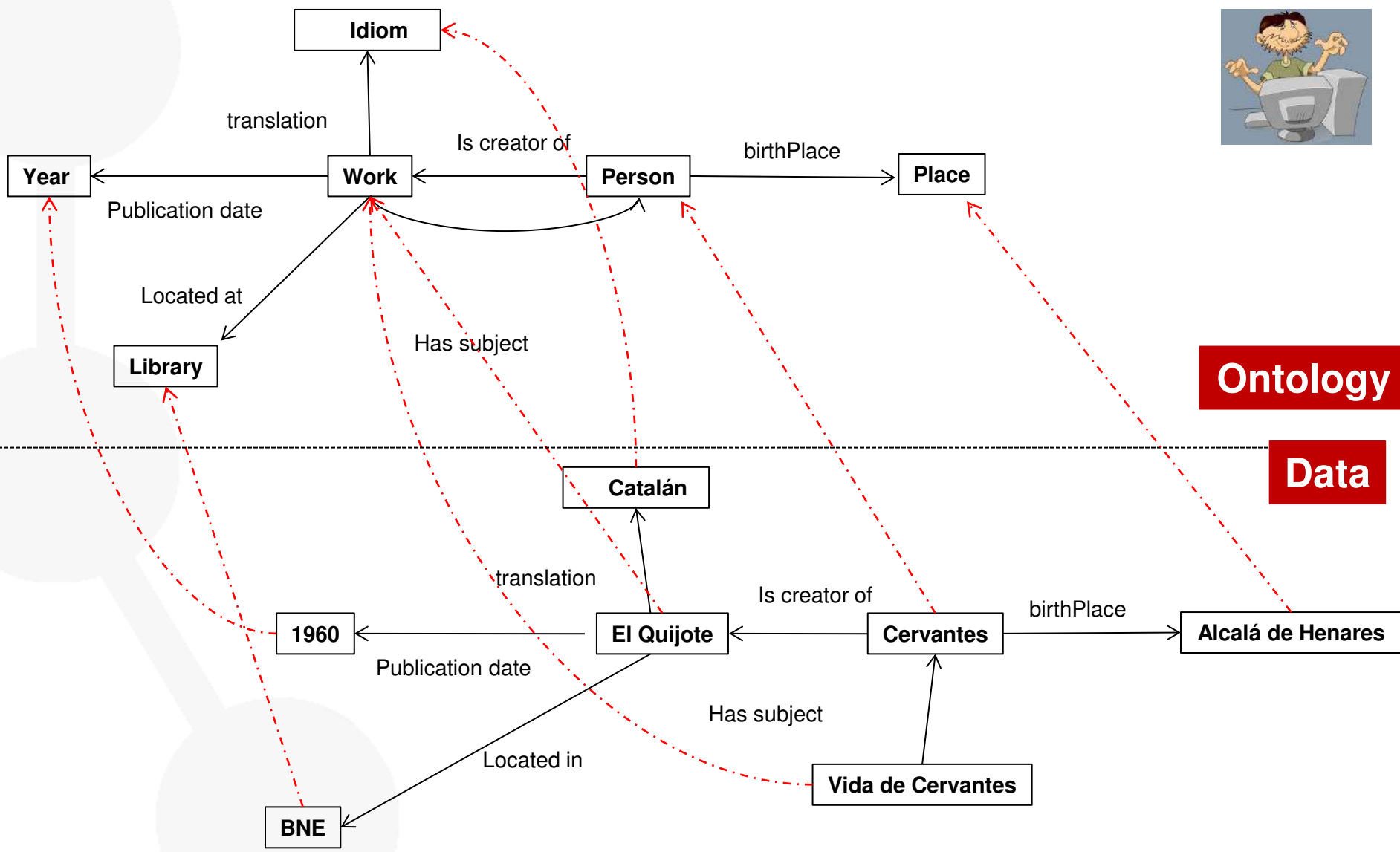
**Unique identifiers: URI**  
identify or name a resource

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





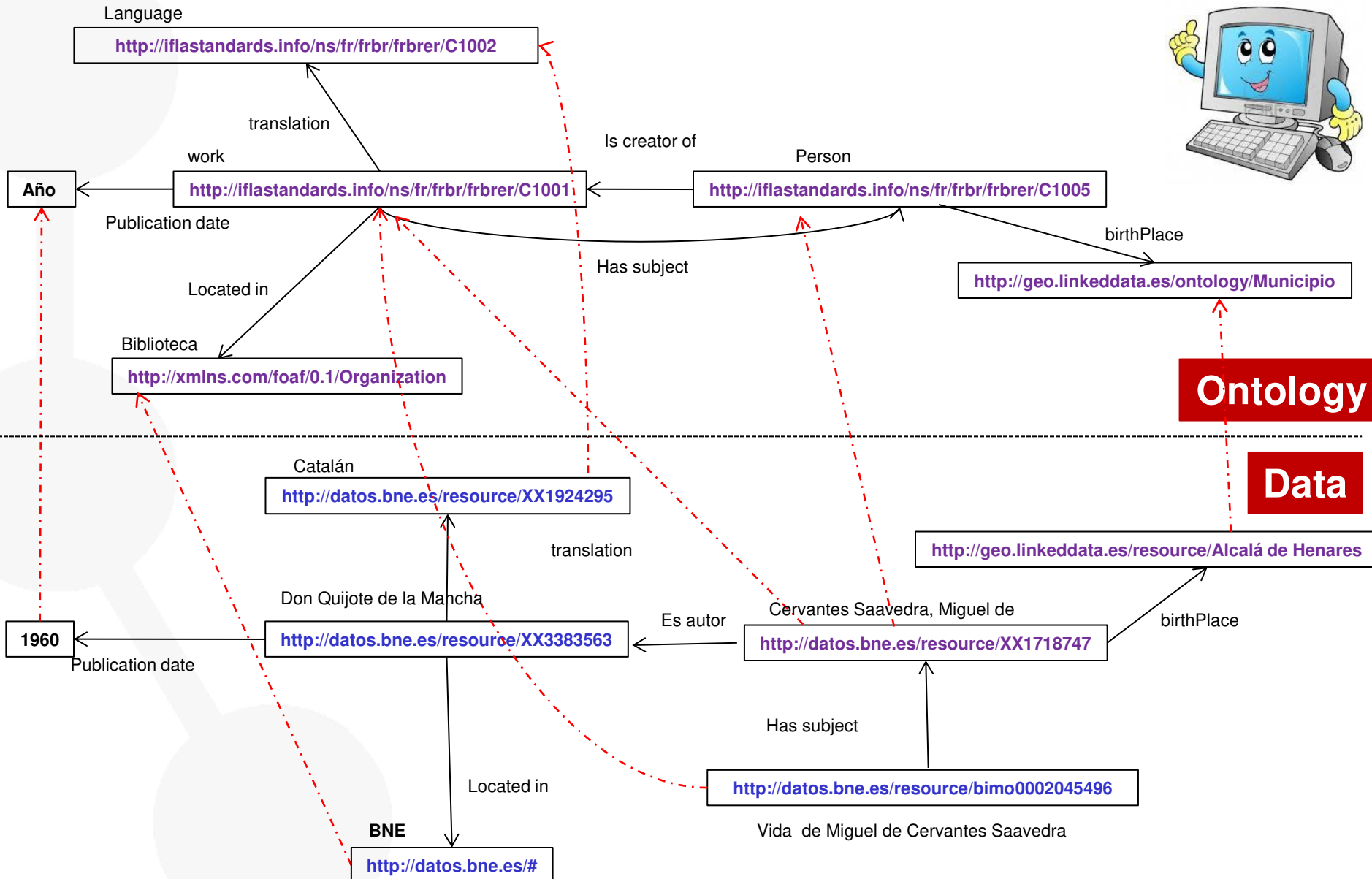
# The model (Ontology) and the data



**Ontology**

**Data**

# Linked data is full of URIs







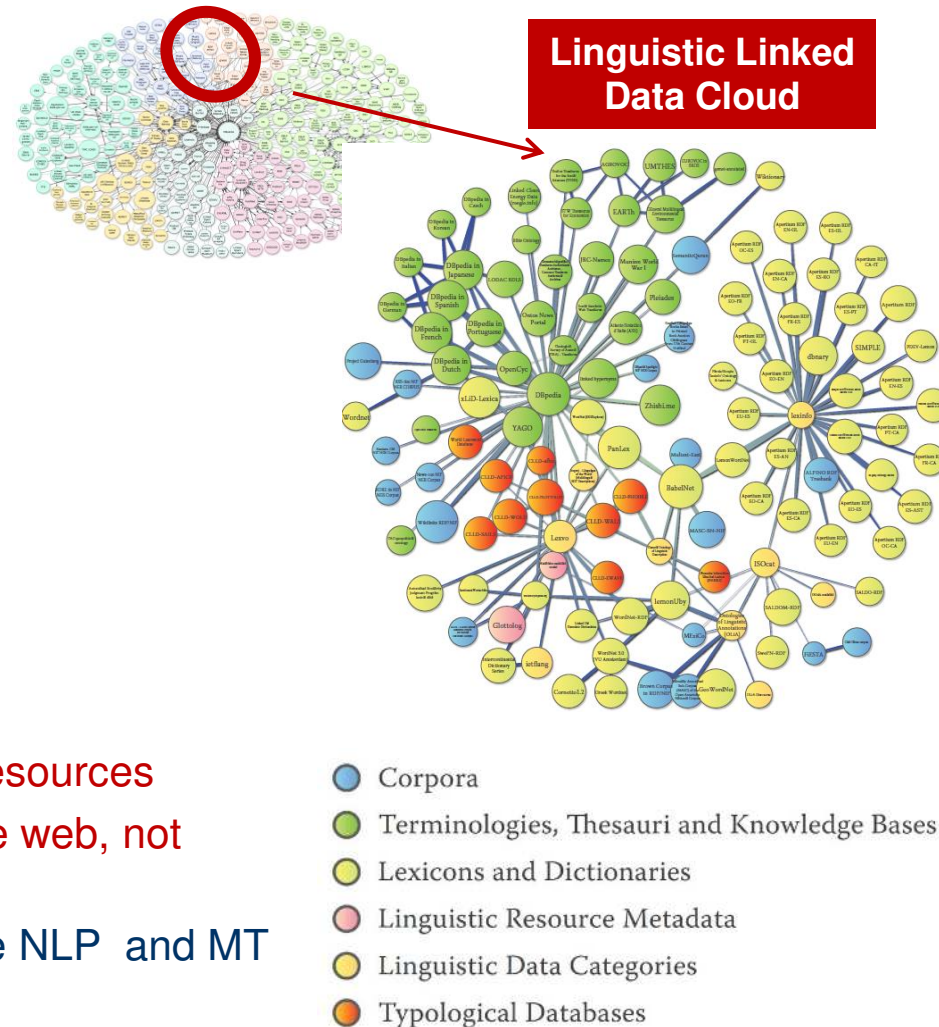
**Language Resources**  
**The need of connecting licensed Language Resources**  
**Linked Data foundations**  
**LINGUISTIC LINKED LICENSED DATA**

## Linked Data interconnects data from resources

- In many domains
- In many languages
- Open and closed License
- Links with other datasets

## Linguistic Linked Data Cloud

- Subset of LOD
- Linguistic domain
- Many type of resources
- Interconnected with other Language Resources
- Enables the lexicalization of data on the web, not necessarily data in the LD format
- Enables a new generation of LD-aware NLP and MT Services



# 3LD



## Linguistic Linked Licensed Data

*Language resources  
such as:*

- Lexica
- Corpora
- Dictionaries
- Grammars ..

*Using **RDF** and  
standard data  
models*



*(vocabularies):*

- Lexica 
- Corpora 
- .....

NLP Interchange Format

*Published along with  
a machine-readable  
license.*

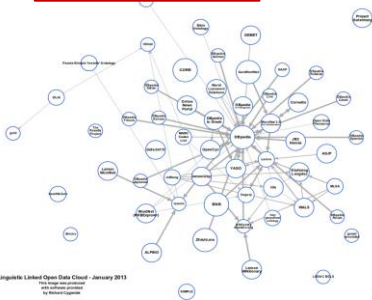
**ODRL**

**Open Digital Rights Language**

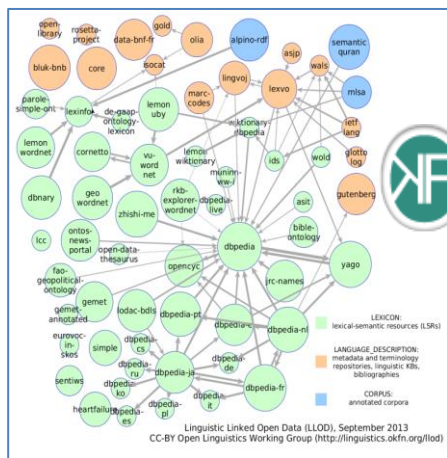


# Linguistic Linked Data Evolution

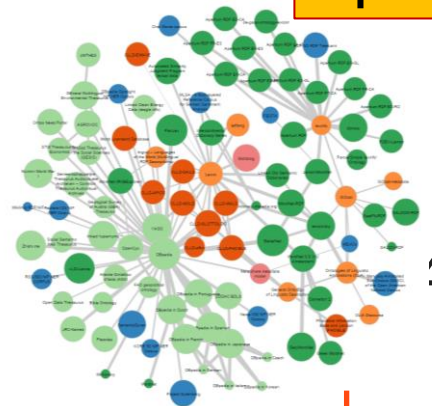
Jan. 2013



Sept. 2013

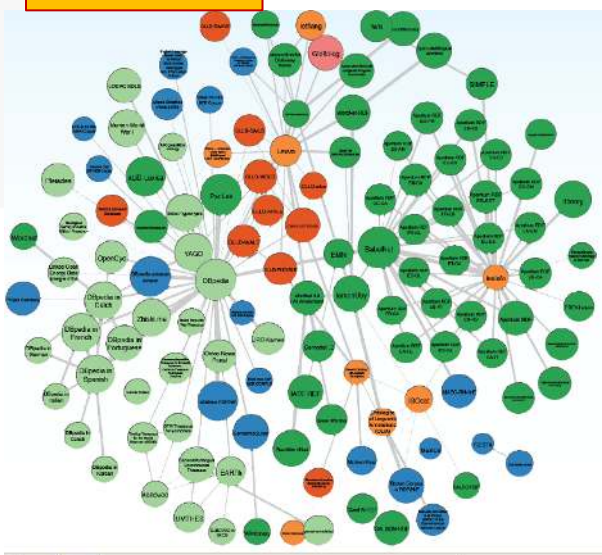


Sept. 2014

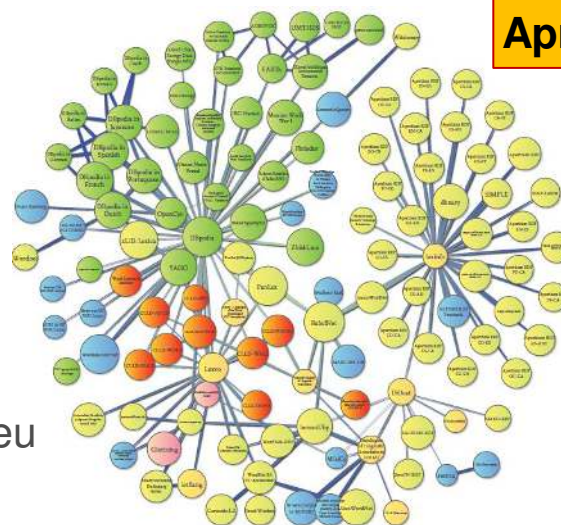


4

Nov. 2015



April. 2015



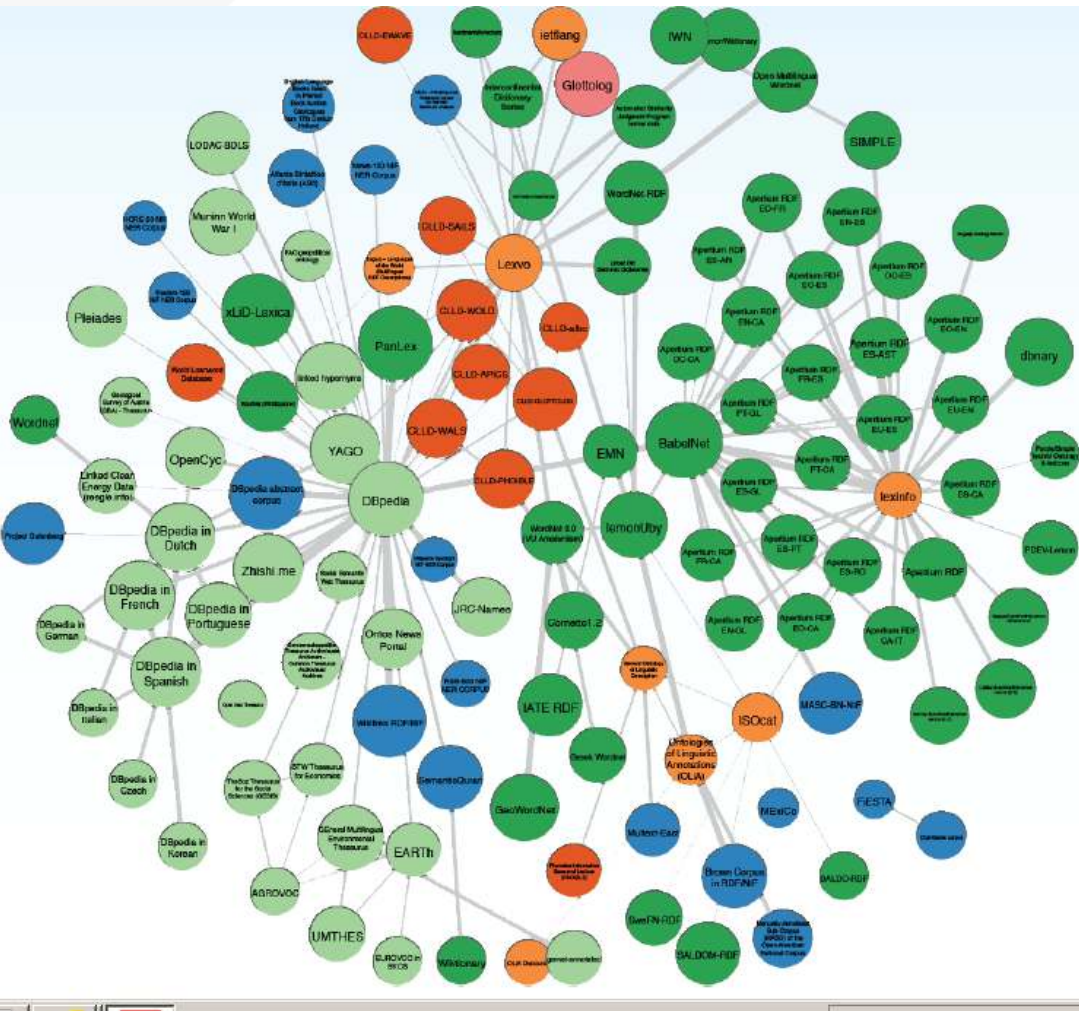
**myl**ider  
www.lider-project.eu

# Linguistic Linked Licensed Data @ Nov 2014



www.lider-project.eu

- Corpora
- Terminologies, Thesauri and Knowledge Bases
- Lexicons and Dictionaries
- Linguistic Resource Metadata
- Linguistic Data Categories
- Typological Databases



## LLOD Cloud in November 2015

- 126 Resources
- 203 Links
- 20 Corpora

Less Centralized: Babelnet, LexVo and LexInfo new hubs

### Criteria for inclusion:

- **Resolvable:** URLs that resolve
- **RDF:** resolve to RDF
- **1000 Triples:** self-explaining
- **Links:** to one resource from the cloud or other 50 links
- **Crawlable:** get the whole resource by crawling
- **Linguistic:** data must be a language resources
- **Registered:** at CKAN





# Linguistic Linked Licensed Data

How do we represent license information?



Lexvo.org - Resource: x  
www.lexvo.org/linkeddata/resources.html

Lexvo.org  
Started FAQ Details Downloads References Legal Co

## Resources

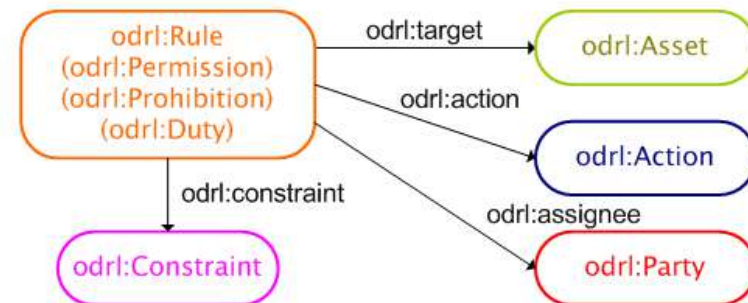
### Download machine-readable data

We offer a machine-readable **RDF** file with information about languages and other entities provided on Lexvo.org. This dump file can be downloaded to your local computer and distributed in accordance with the [Creative Commons Attribution-ShareAlike \(CC-BY-SA\) 3.0 Unported license](#):

- [Lexvo dump \(2013-02-09\)](#)
- [Lexvo dump \(direct link to latest version 2013-02-09\)](#)

We are grateful to a number of people for providing the [original data sources](#) that were used to create Lexvo.org.

Rights expression in ODRL:  
Who [can|cannot|must] act what in which resource how



# Using Policies to govern conditional access to Linked Data

Example of access to Linked Data for a price (15EUR for the dataset or 0.01EUR for a triple thereof)

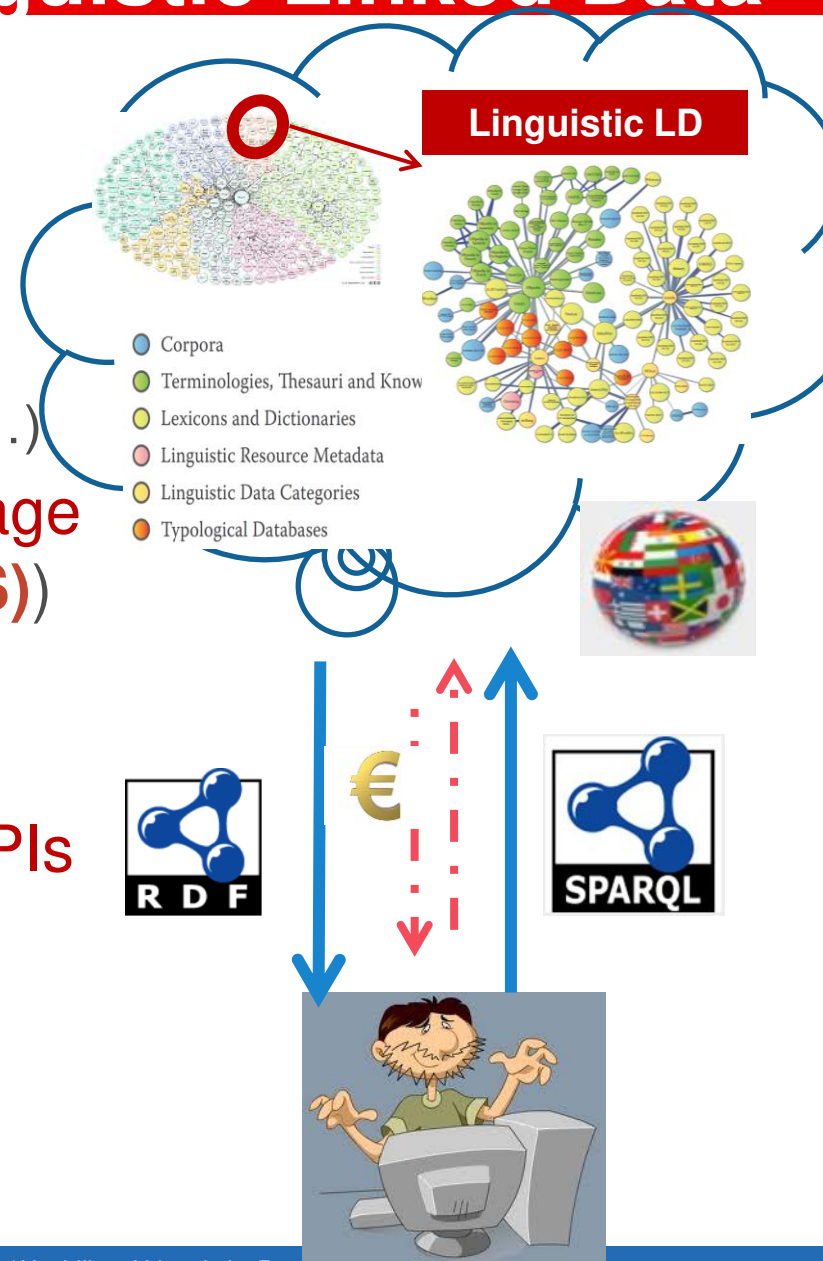
```
@prefix gr:      <http://purl.org/goodrelations/> .
@prefix dcat:   <http://www.w3.org/ns/dcat#> .
<http://salonica.dia.fi.upm.es/ldr/policy/cdaddba4-fc2e-4ee0-a784-e62f1db259bf>
  a
  rdfs:label     "License Offering Paid Linked Data" ;
  odrl:permission [ a
                    odrl:Permission ;
                    odrl:target <http://example.org/dataset/ds01> ;
                    odrl:action odrl:reproduce ;
                    odrl:duty   [ a
                                  odrl:Duty ;
                                  rdfs:label     "Pay" ;
                                  gr:UnitOfMeasurement dcat:Dataset ;
                                  gr:amountOfThisGood "1" ;
                                  odrl:action     odrl:pay ;
                                  odrl:target     "15,00 EUR"
                                ]
                  ] , [ a
                    odrl:Permission ;
                    odrl:action odrl:reproduce ;
                    odrl:target <http://example.org/dataset/ds01> ;
                    odrl:duty   [ a
                                  odrl:Duty ;
                                  rdfs:label     "Pay" ;
                                  gr:UnitOfMeasurement rdf:Statement ;
                                  gr:amountOfThisGood "1" ;
                                  odrl:action     odrl:pay ;
                                  odrl:target     "0,01 EUR"
                                ]
                  ] ..
```




<http://conditional.linkeddata.es>

# Linked Data and Linguistic Linked Data

1. **Agree on vocabularies** for describing
  - Domain datasets
  - LR metadata
  - LR content (Lemon-Ontolex, NIF, ...)
2. Unified and standardized **language** for describing resources ( **RDF(S)** )
3. Unified and standardized **query language** (SPARQL)
4. Standardized **non-proprietary APIs**
5. **Links** to other resources





**Language Resources**  
**The need of connecting licensed Language Resources**  
**Linked Data foundations**  
**Linguistic Linked Licensed Data**

# **OUTCOMES OF THE LIDER PROJECT**

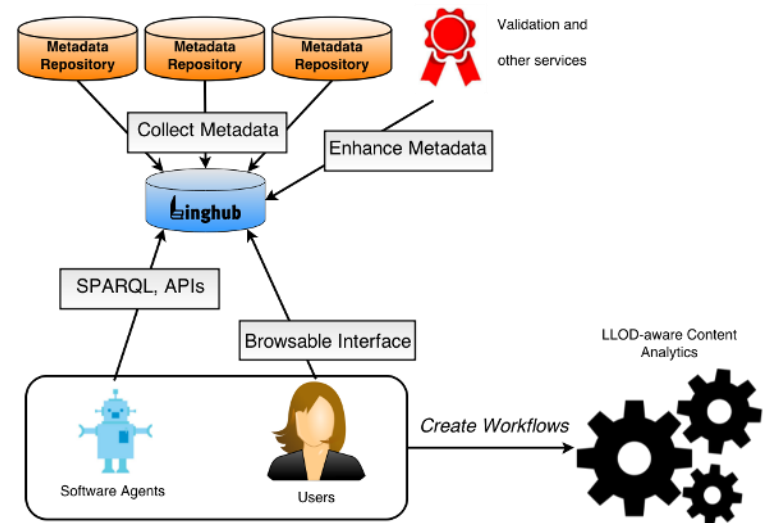
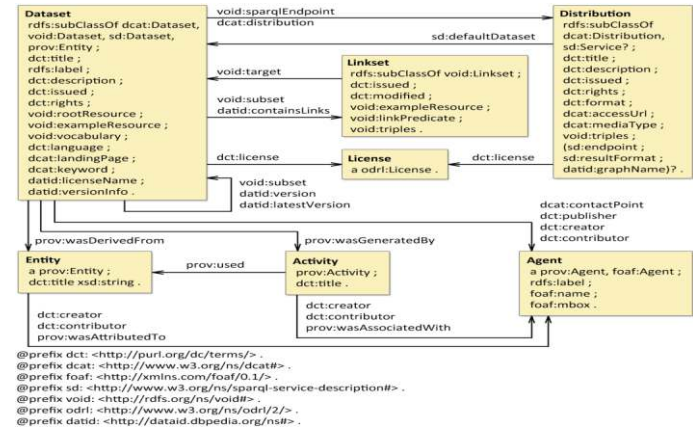
# Discovery LRs metadata with LingHub

## Language resource metadata transformation and aggregation

- Metadata OWL ontology @ LD4LT W3C group
- Exposure of Metashare, Clarin, LREMap, datahub metadata as LD

## LingHub

- Collect and Homogenise metadata
- Enhance metadata
- Helps you to search LRs



myl<sup>ider</sup>

www.lider-project.eu

<http://linghub.lider-project.eu/>

# Conversion of 86 LRs data into LD

## 86 Resources transformed:

- Princeton WordNet,
- BabelNet
- IATE
- Wikipedia Bitaxonomy
- Terminesp (spanish)
- Termcat
- 22 Apertium bilingual dictionaries
- the Brown Corpus
- Wikilinks Corpus
- 27 wordnets from the Open Multilingual Wordnet
- 22 lexical resources for Swedish, including SALDO

## Use of models and guidelines from WP4 standardization activities

- Use of **Ontolex-lemon** Model from the Ontolex group at W3C
- **BPMLOD@W3C guidelines**

## NIF Refinement by the community

## Resources transformed by the community

- LDL 2014
- Summer datathon on LLD



[www.lider-project.eu](http://www.lider-project.eu)

## 8 Guidelines for Linguistic Linked License Data

- Generation of
  - Multilingual Dictionaries (BabelNet)
  - Bilingual Dictionaries
  - Multilingual Terminologies (TBX)
  - WordNets
- Corpus creation using NIF
- NIF-based NLP Web Services
- LLD Exploitation
- LLOD aware services

<http://www.lider-project.eu/guidelines>

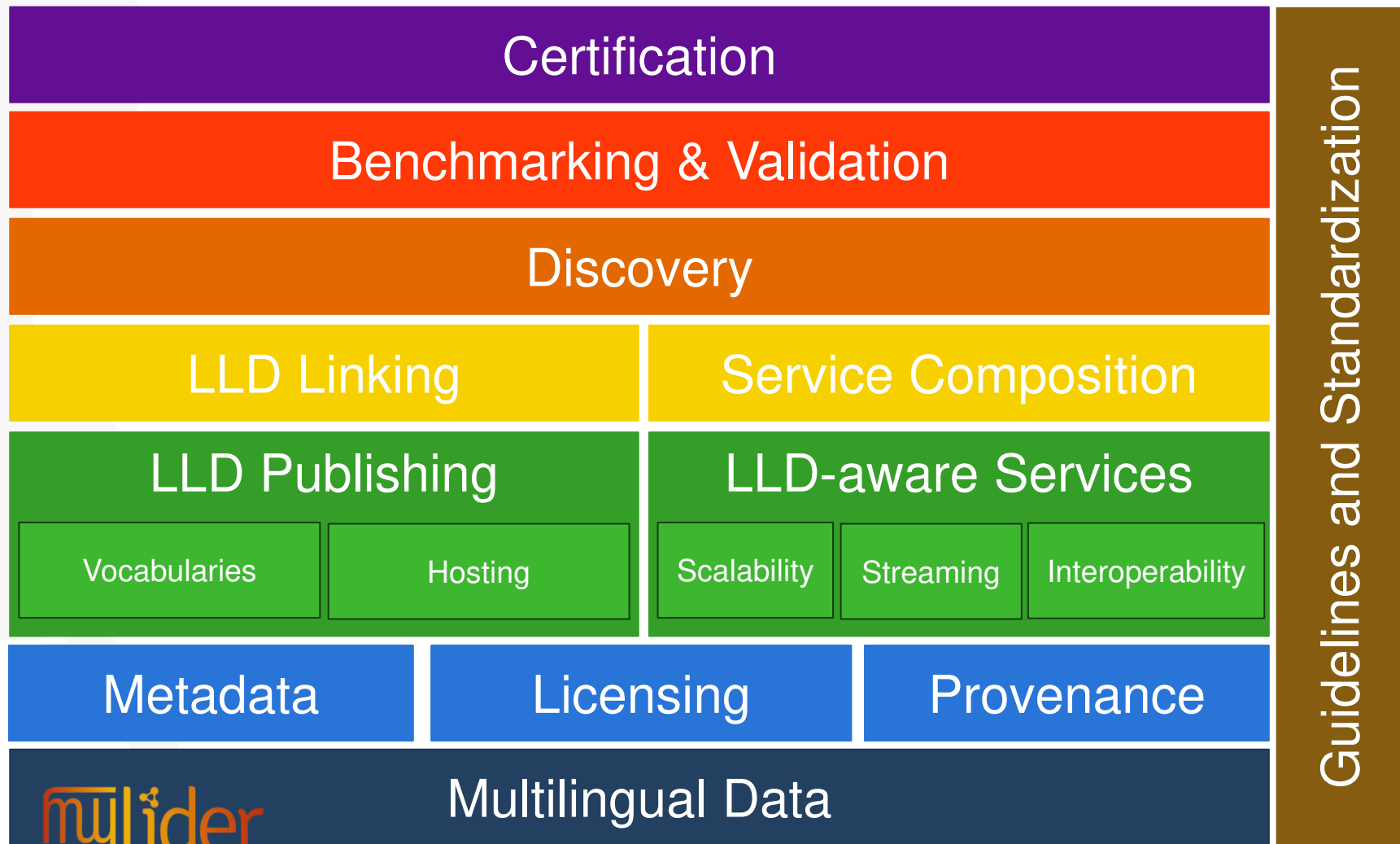
## Best practices for Multilingual Linked Data Publication

- Practices for Naming (URIs)
- Practices for Dereferencing
- Practices for Textual Information
- Practices for Linking
- Practices for Language Identification

The logo for the multilingual Linked Data project, featuring the word 'mlider' in a stylized font where the 'm' and 'l' are connected and the 'i' has a dot above it.

[www.lider-project.eu](http://www.lider-project.eu)





[www.lider-project.eu](http://www.lider-project.eu)

# Reference Architecture and Reference Cards

Ce

Dataset description with DataID

Benchm

Contributing to the Linguistic Linked Open Data Cloud

This card describes how you can contribute to the Linguistic Linked Open Data Cloud

LLD Linking

Documenting metadata of language resources in Datahub

LLD Publishing

This card explains how to add some basic metadata to resources documented in Datahub

Vocabularies

Ho

## Target audience and Scope

The card is intended for people who have a dataset in linked data and wish to document it in Datahub. This card extends the card "Contributing to the Linguistic Linked Open Data Cloud" by adding more details on how to document some basic metadata. See also, as an alternative, the "Dataset description with DataID" card.

Metadata

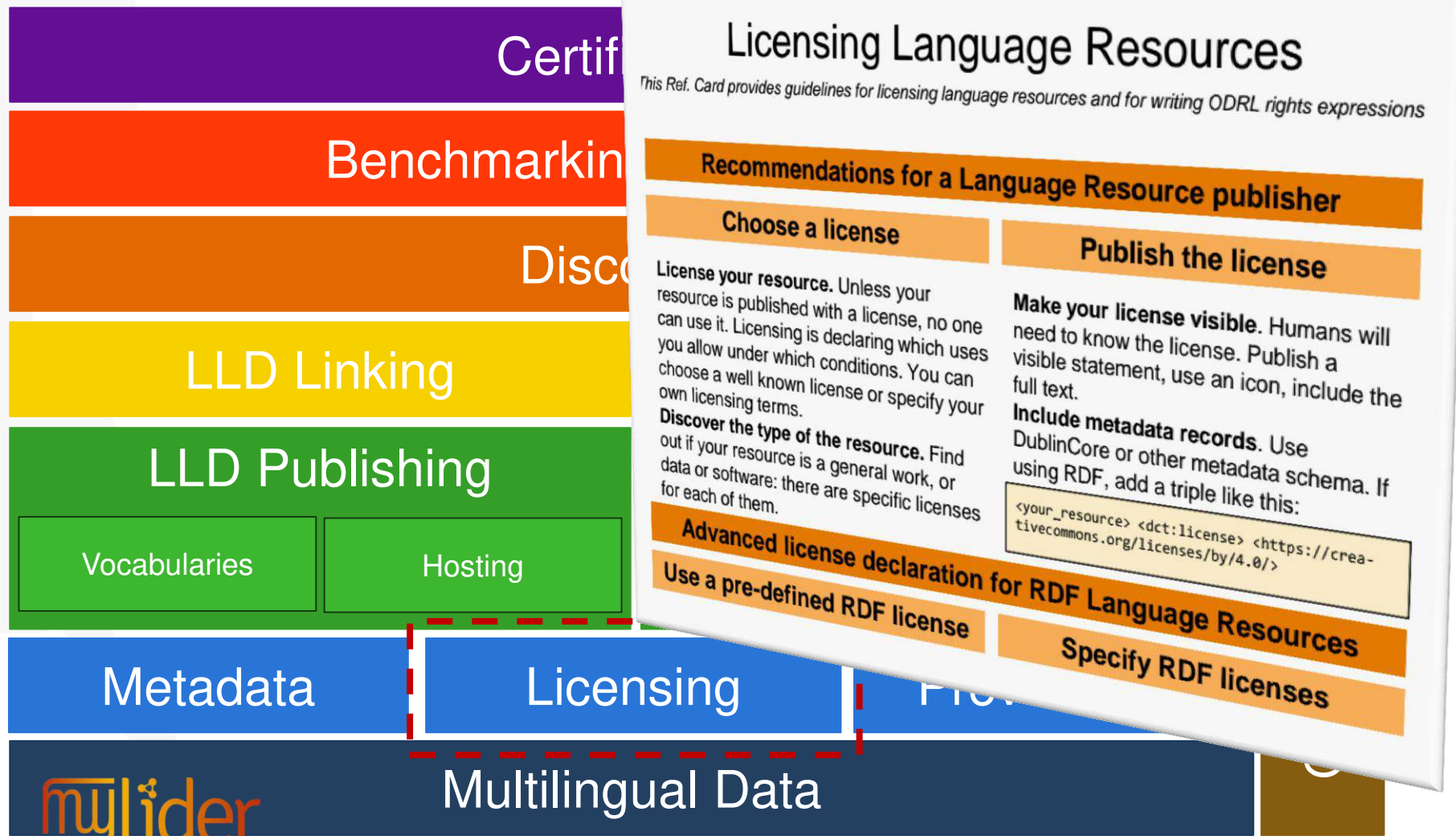
## Registering with Datahub

Go to <http://datahub.io> and follow the steps in the "Contributing to the Linguistic Linked Open Data Cloud" reference card. As result, Datahub automatically creates an RDF metadata file for the dataset, based on the Data Catalogue Vocabulary (DCAT). This metadata file is accessible at [http://datahub.io/dataset/DATASET\\_ID](http://datahub.io/dataset/DATASET_ID) where DATASET\_ID is the identifier of the dataset.

mūlīder

[www.lider-project.eu](http://www.lider-project.eu)

# Reference Architecture and Reference Cards



www.lider-project.eu

# Reference Architecture and Reference Cards



## GUIDELINES

- Multilingual Dictionaries (BabelNet)
  - Bilingual Dictionaries
- Multilingual terminologies (TBX)
  - WordNets

## LLD Linking

## LLD Publishing

Vocabularies

Hosting

Metadata

Licen



Multilingu

[www.lider-project.eu](http://www.lider-project.eu)

## How to publish Linguistic Linked Data

Publishing data on the Web in a way that it can be easily consumed is a daunting task. The Web community has converged on the idea that Linked Data provides the best practices for publishing highly interoperable data. This card explains in detail...

## Linked Data Corpus Creation with NIF

This document describes best practices to follow for the generation of Linked Data text corpora, using the NLP Interchange Format (NIF).

### Target audience

Corpus creators and users seeking to make corpora interoperable and to publish them as linked data. Basic knowledge of RDF is mandatory for conversion. Basic knowledge of linked data and web server access is needed for publication.

### Scope

Conversion of existing corpora into RDF using NIF, as well as creation of linked data corpora from textual data.

### Core concepts

Corpus

We understand a corpus as a collection of documents. Documents contain text represented as strings...

# Reference Architecture and Reference Cards

**Certification**

**Benchmarking**

**Discovery**

**LLD Linking**

**LLD Publishing**

Vocabularies      Hosting

**Metadata**      **Licensing**

**Multilingual Data**

**How to represent cross-lingual LD links**

This card explains the basic mechanisms available to represent links between linked data in different natural languages.

**Target audience**

Data publishers and developers with basic acquaintance of RDF and LD who need to interlink data in different languages in the linked data cloud.

**Scope**

The card review the different types of cross-lingual LD links. It does not discuss the methods/techniques needed to infer the links, but deals with representation mechanisms only. This cards extends the card "How to publish Linguistic Linked Data".

**Core concepts**

Cross-lingual mappings can be established at two different levels: **conceptual level** and **linguistic level**. In the first case, links are

[www.lider-project.eu](http://www.lider-project.eu)

# Reference Architecture and Reference Cards

Certification

Benchmarking & Validation

Discovery

LLD Linking

LLD Publishing

Vocabularies

Hosting

Metadata

Licensing

**mlider**

Multilingual

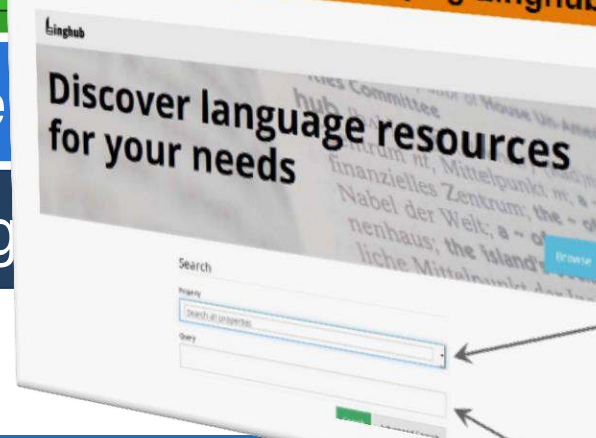
[www.lider-project.eu](http://www.lider-project.eu)

Standardization

## Discovering Language Resources with Linghub

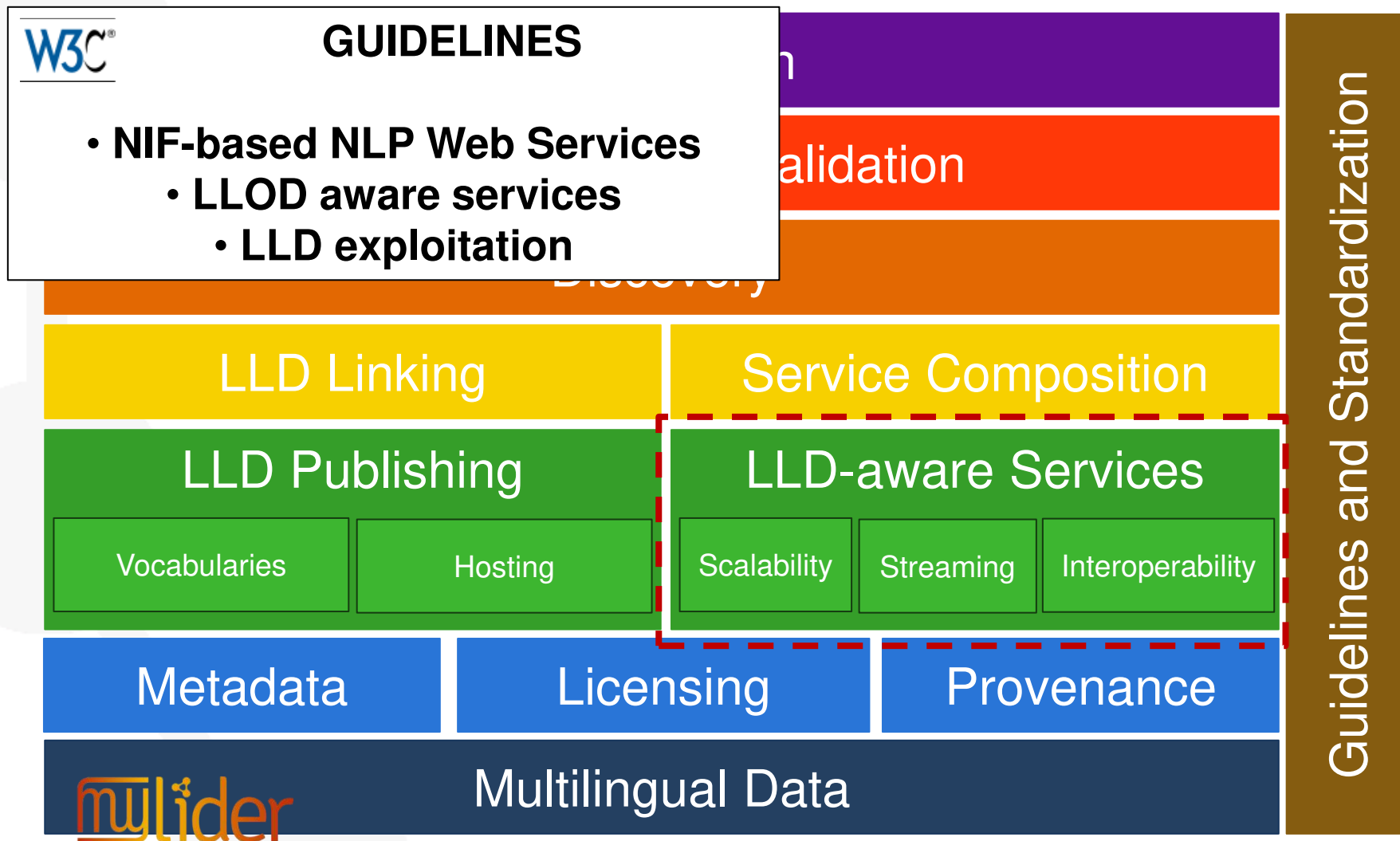
This card describes how you can discover language resources with Linghub. This card is intended for people who need a language resource for a given NLP task.

### Querying Linghub



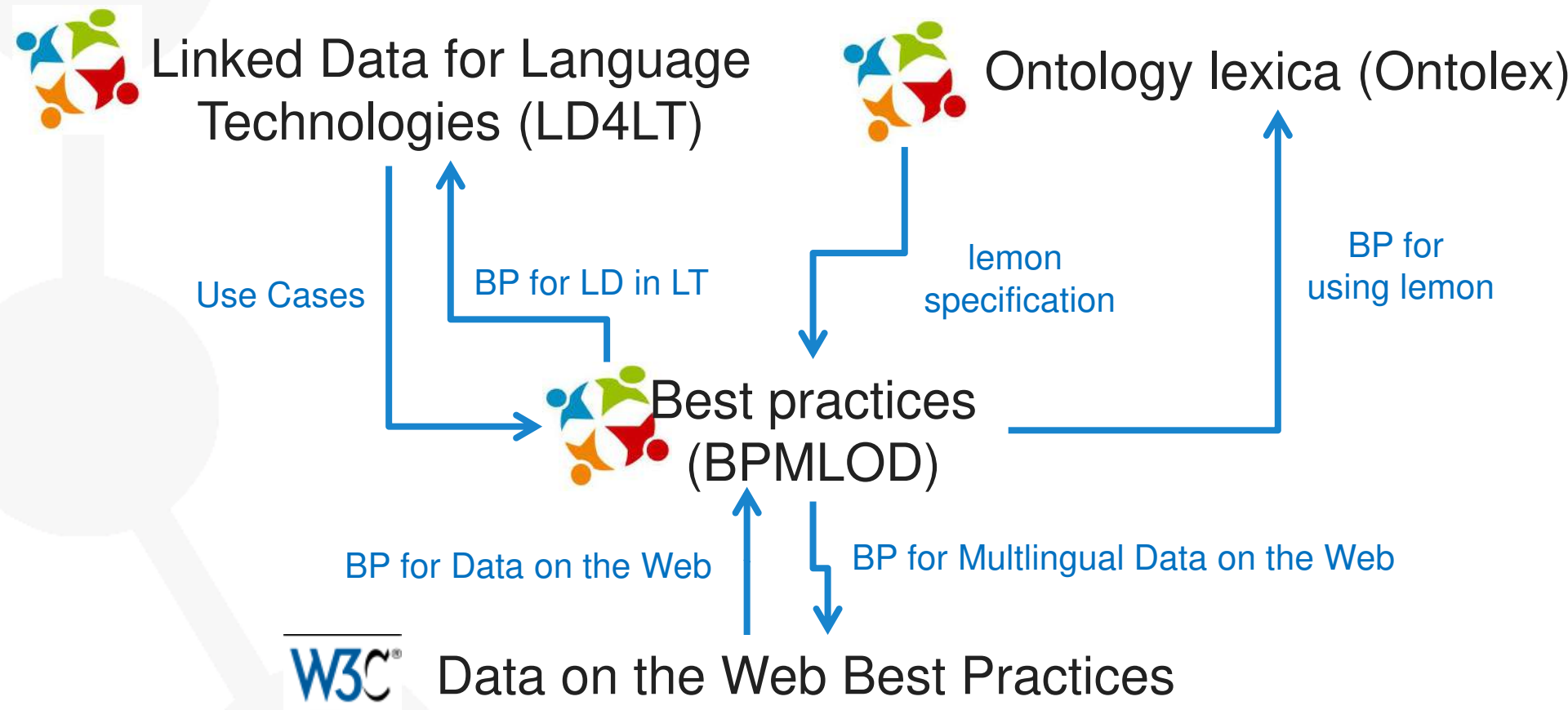
Select a property

# Reference Architecture and Reference Cards



[www.lider-project.eu](http://www.lider-project.eu)

# Relation between W3C groups





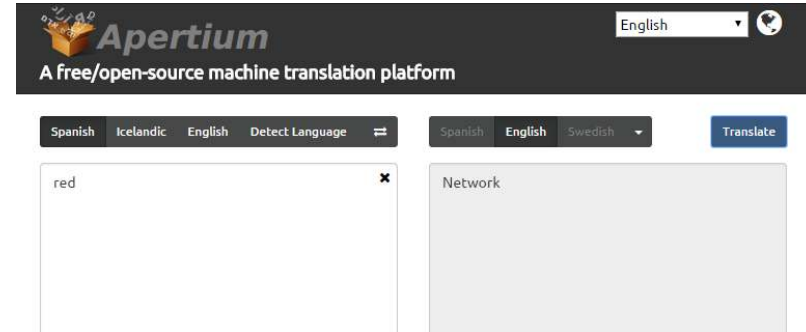
**Language Resources**  
**The need of connecting licensed Language Resources**  
**Linked Data foundations**  
**Linguistic Linked Licensed Data**  
**Outcomes of the LIDER project**

## **THE APERTIUM RDF CASE AND THEIR LINKS**



## Apertium Platform: Machine Translation.

> 40 bilingual dictionaries available in XML.

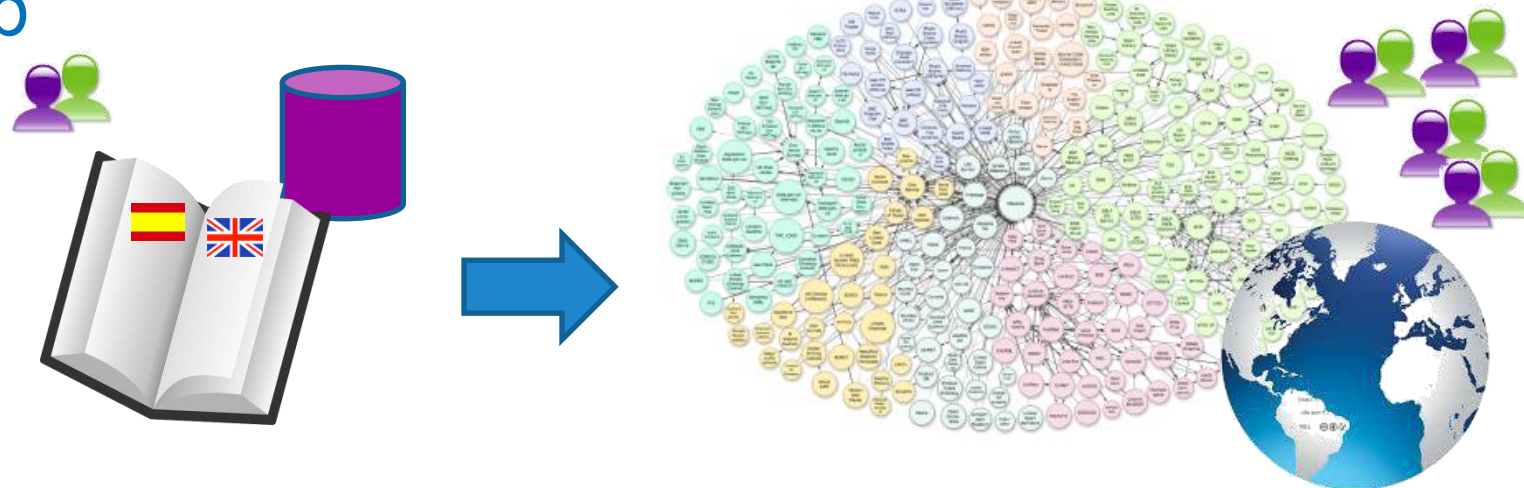


Afrikaans <-> Dutch  
 Breton --> French  
 Catalan <-> Italian  
 Welsh <-> English  
 Danish <-> Norwegian  
 English <-> Catalan  
 English <-> Spanish  
 English <-> Galician  
 Esperanto <-> Catalan  
 Esperanto <-> English  
 Esperanto <-> Spanish  
 Esperanto <-> French  
 Spanish <-> Aragonese  
 Spanish <-> Asturian  
 Spanish <-> Catalan  
 Spanish <-> Galician

Spanish <-> Italian  
 Spanish <-> Portuguese  
 Spanish <-> Romanian  
 Basque --> English  
 Basque --> Spanish  
 French <-> Catalan  
 French <-> Spanish  
 Serbo-Croatian <-> English  
 Serbo-Croatian <-> Macedonian  
 Serbo-Croatian <-> Slovenian  
 Indonesian <-> Malaysian  
 Icelandic <-> Swedish  
 Icelandic --> English  
 Kazakh <-> Tatar  
 Macedonian <-> Bulgarian  
 Macedonian --> English

Norwegian Nynorsk <-> Norwegian  
 Bokmål  
 Occitan <-> Catalan  
 Occitan <-> Spanish  
 Portuguese <-> Catalan  
 Portuguese <-> Galician  
 Northern Sami --> Norwegian  
 Bokmål  
 Swedish <-> Danish  
 .....

GOAL: to expose **translations** contained in bilingual dictionaries as **Linked Data** on the **Web**



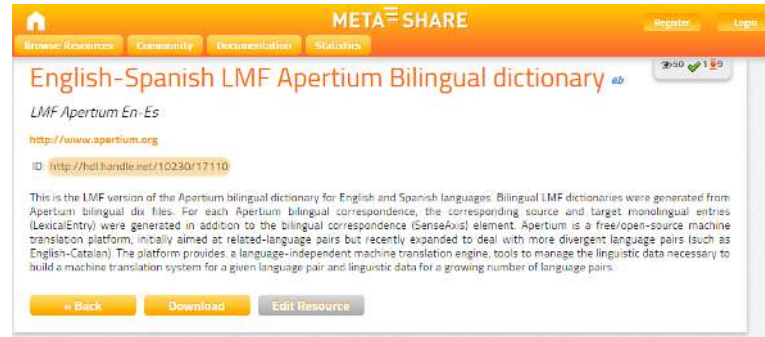
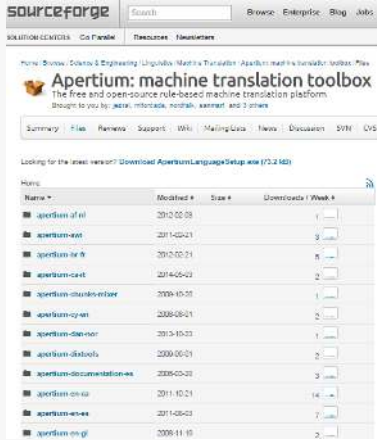
Joint effort by



# Open data formats and publication

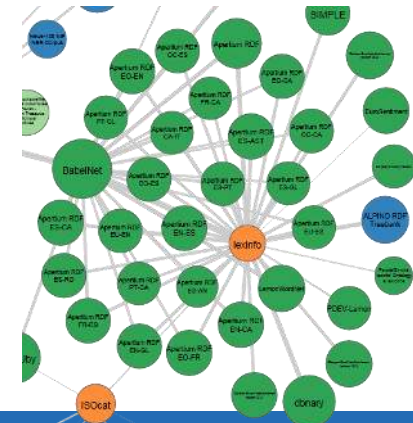
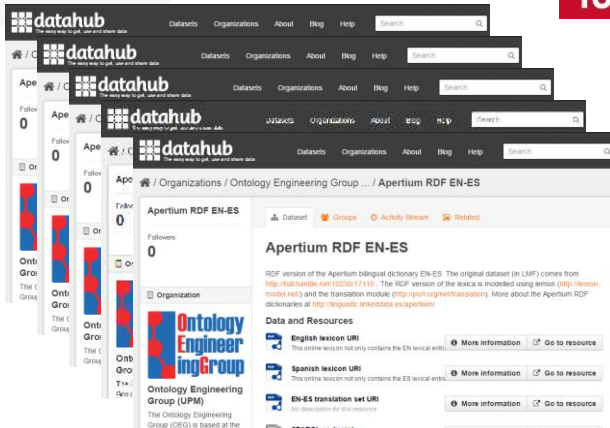
+40 Apertium Open bilingual Dictionaries proprietary format

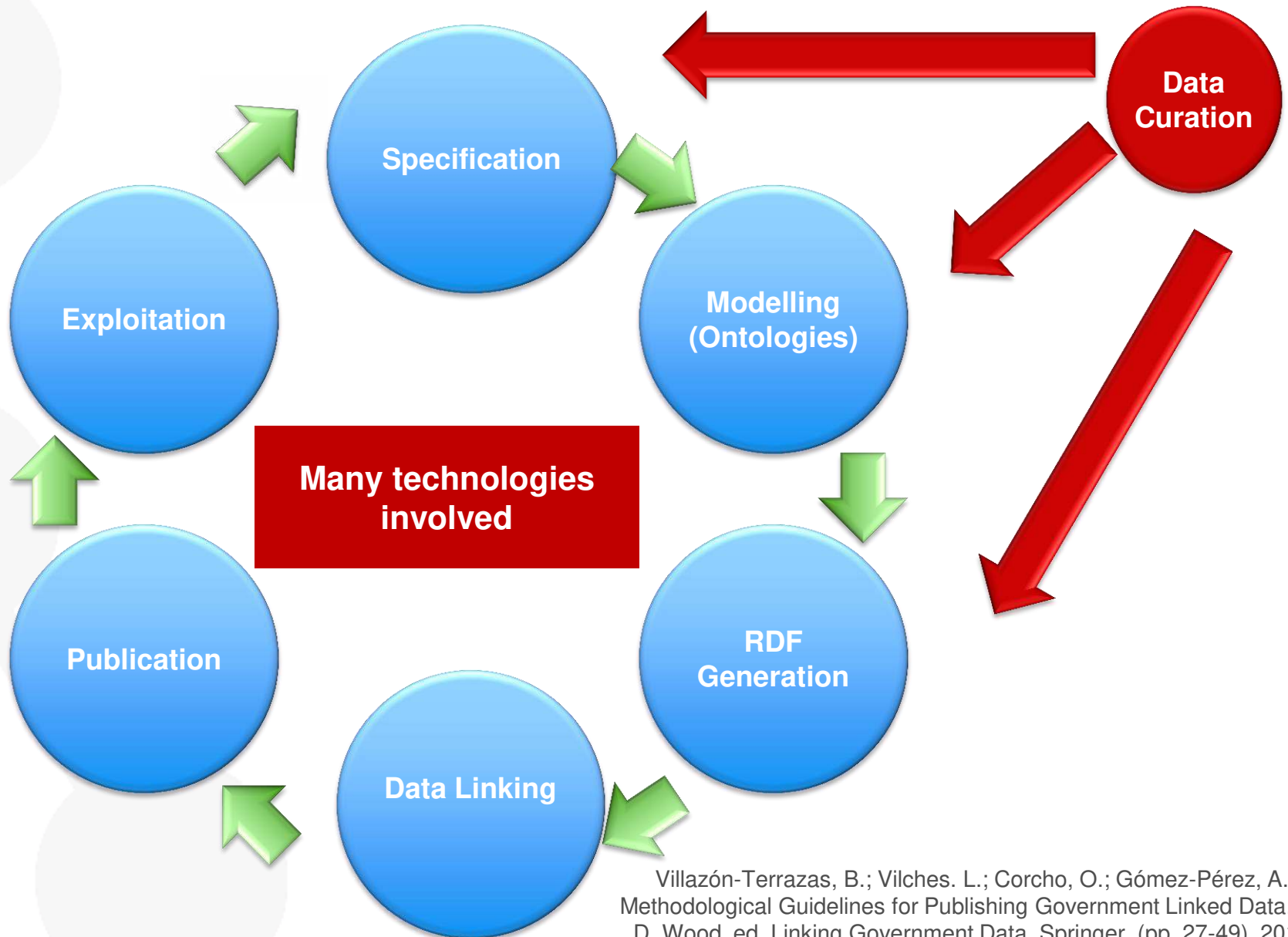
22 Apertium bilingual Dictionaries in LMF



22 Apertium RDF

690K links with LexInfo  
277K links with Babelnet





Villazón-Terrazas, B.; Vilches, L.; Corcho, O.; Gómez-Pérez, A. Methodological Guidelines for Publishing Government Linked Data. In D. Wood, ed. Linking Government Data. Springer. (pp. 27-49). 2011

# Approach for Apertium RDF



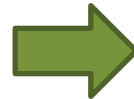
Specification



Apertium LMF



EN-ES



Monolingual  
lexicons

Lexicon ES

Lexicon EN

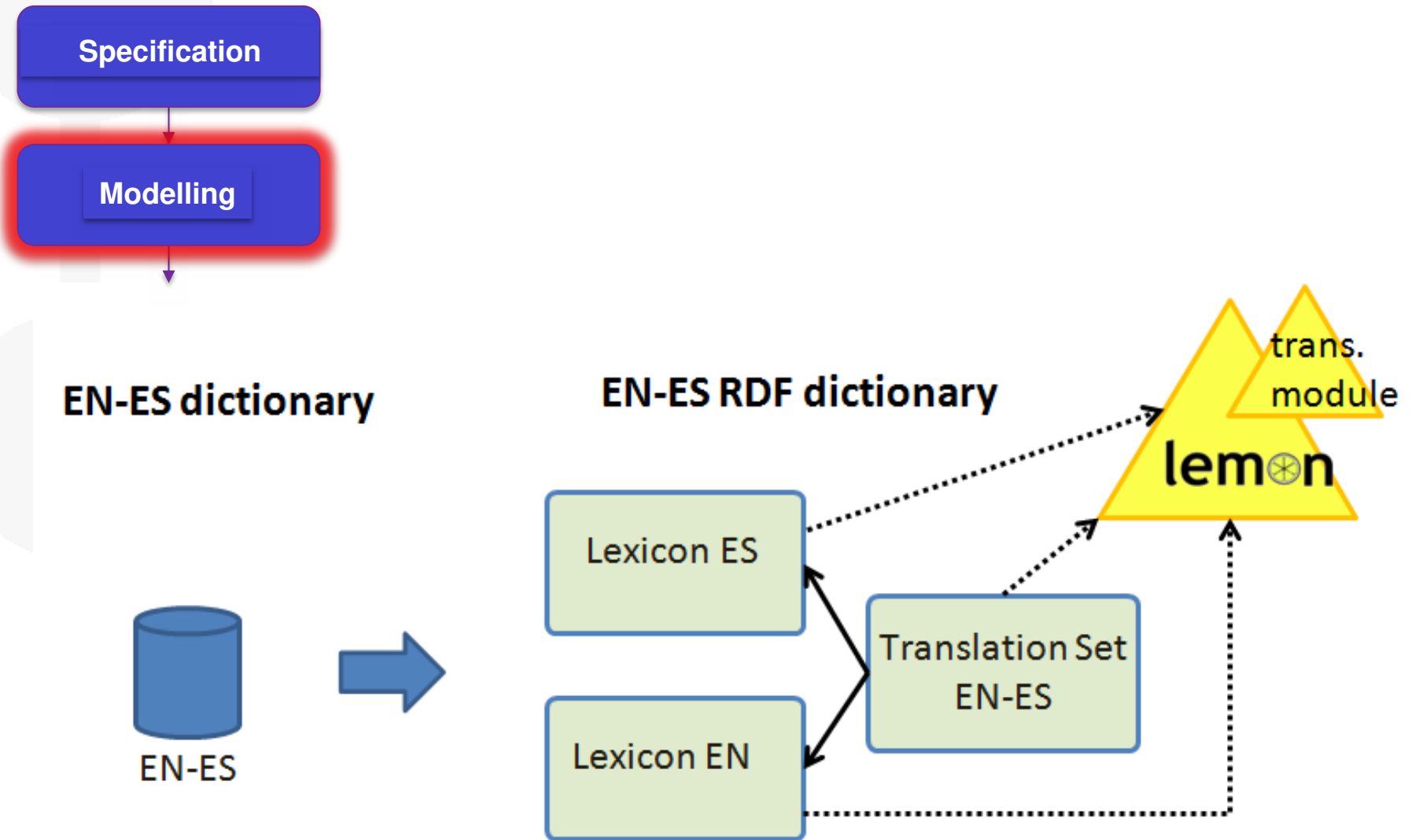
Translation sets

Translation  
Set EN-ES

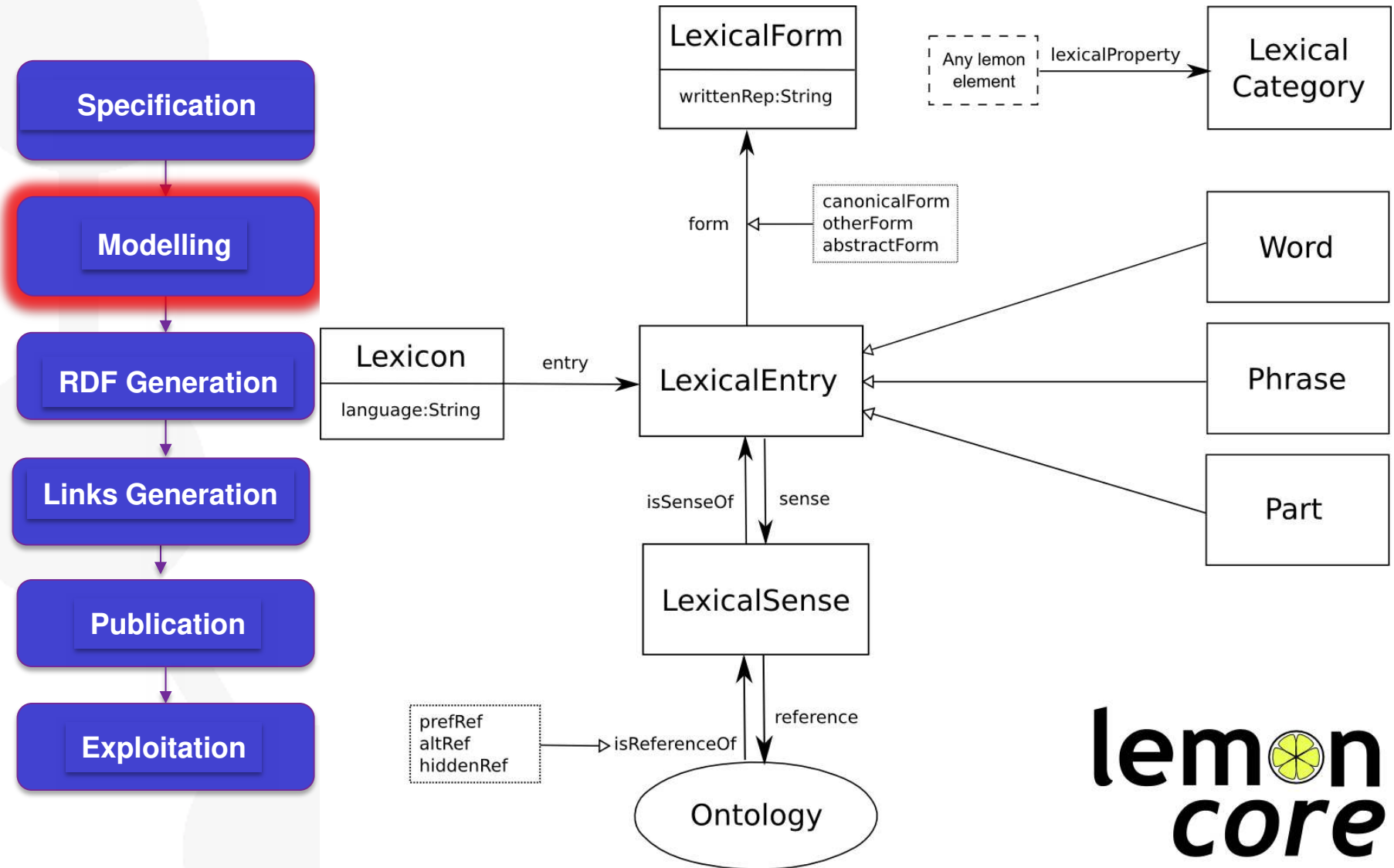
URI Design

```
# Apertium English lexicon:  
http://linguistic.linkeddata.es/id/apertium/lexiconEN  
  
# Apertium Spanish lexicon:  
http://linguistic.linkeddata.es/id/apertium/lexiconES  
  
# Apertium English-Spanish translation set:  
http://linguistic.linkeddata.es/id/apertium/tranSetEN-ES
```

# Modelling: for each bilingual dictionary....



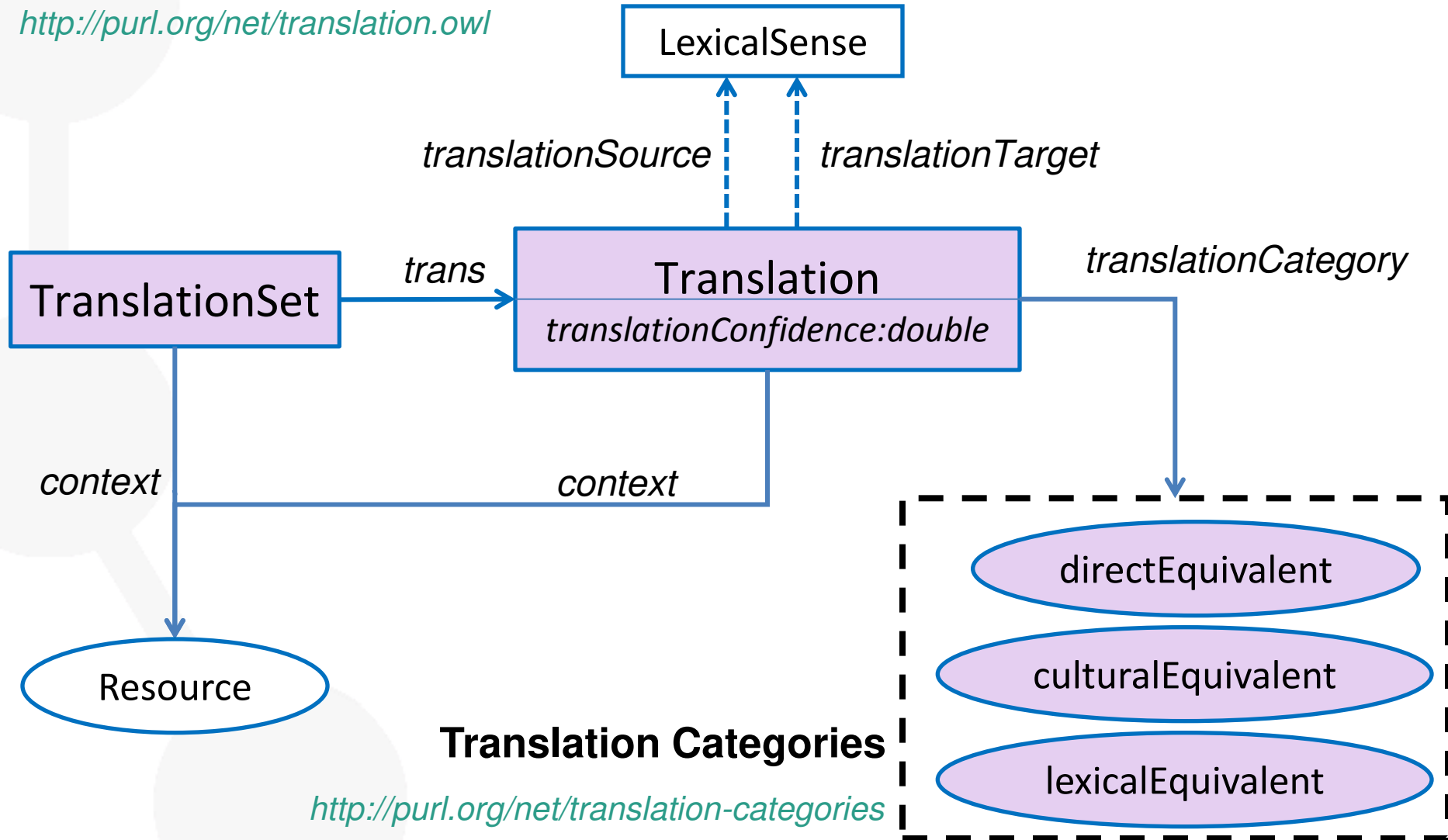
# Modeling lexicons with Lemon



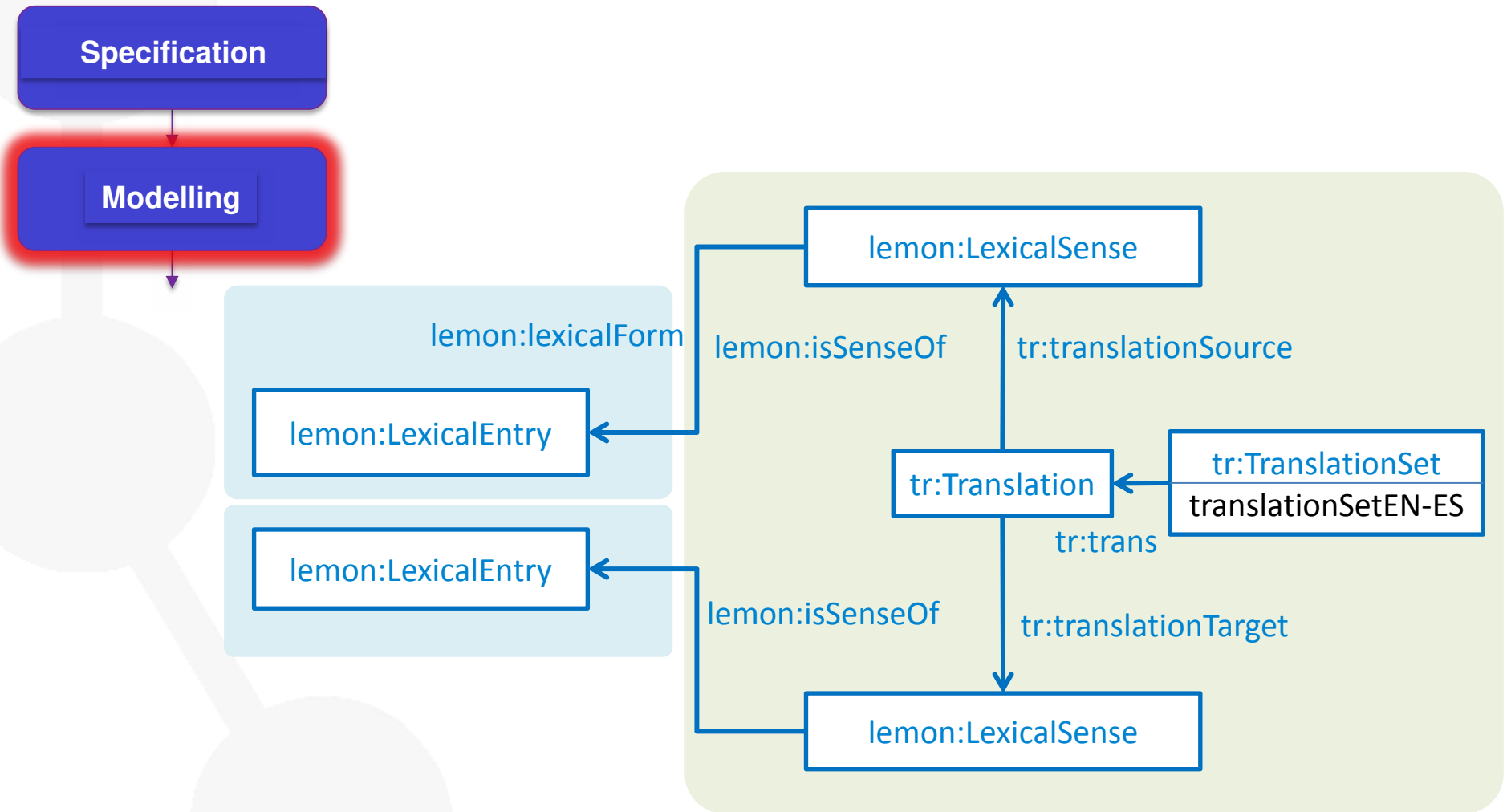


## Translation Module

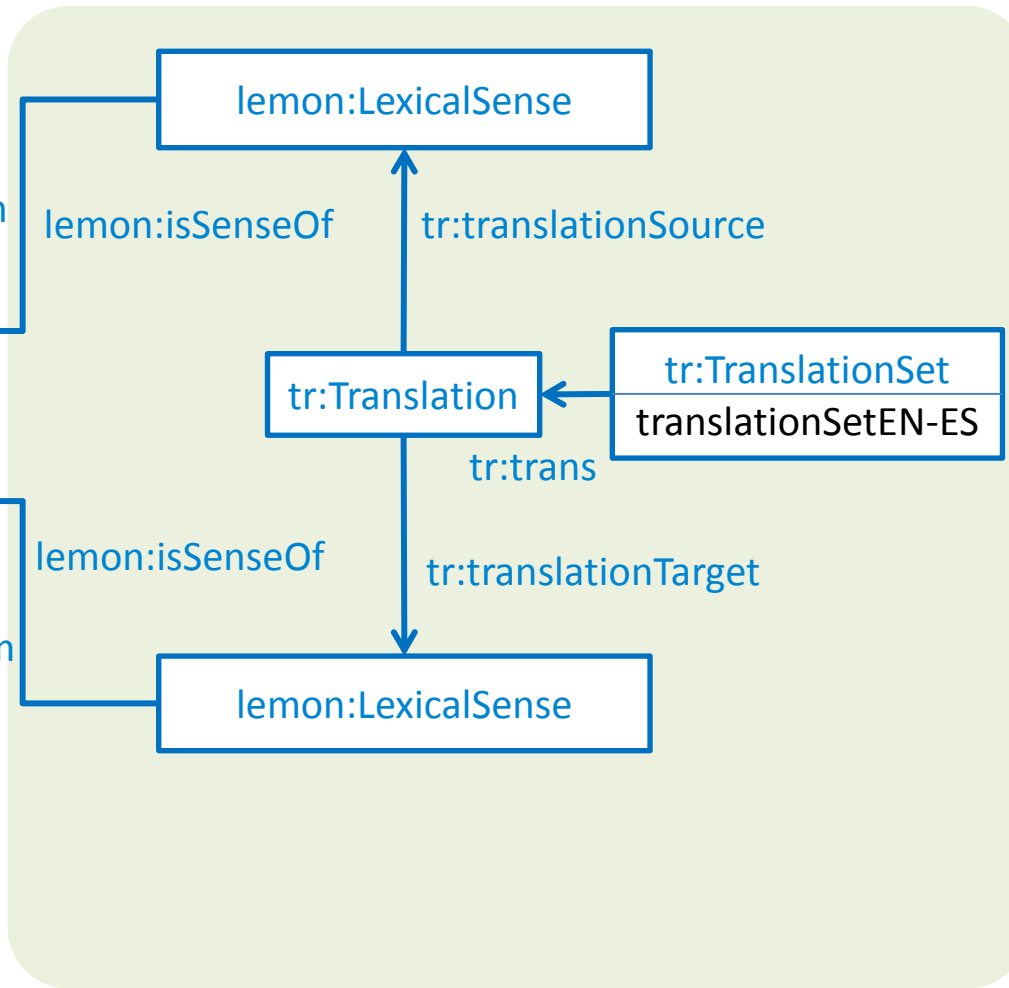
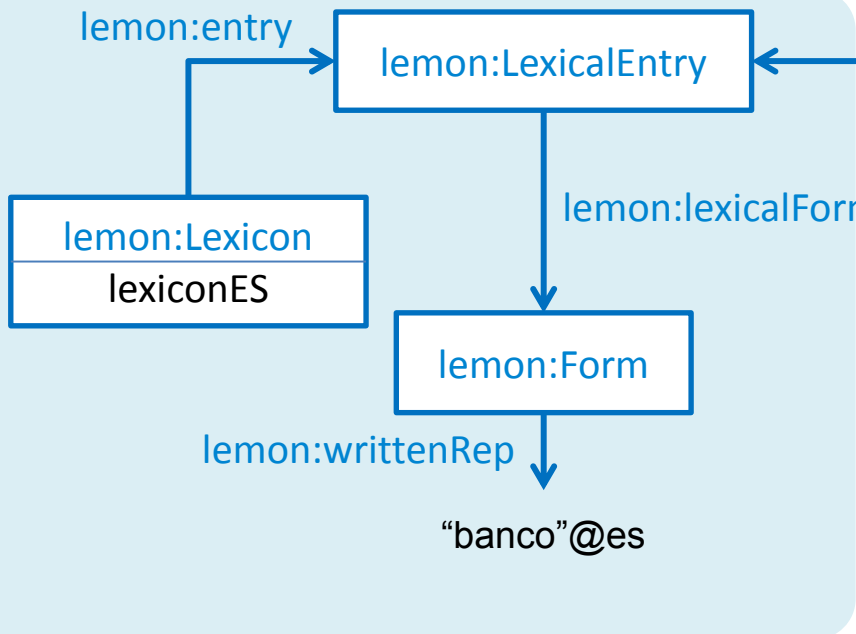
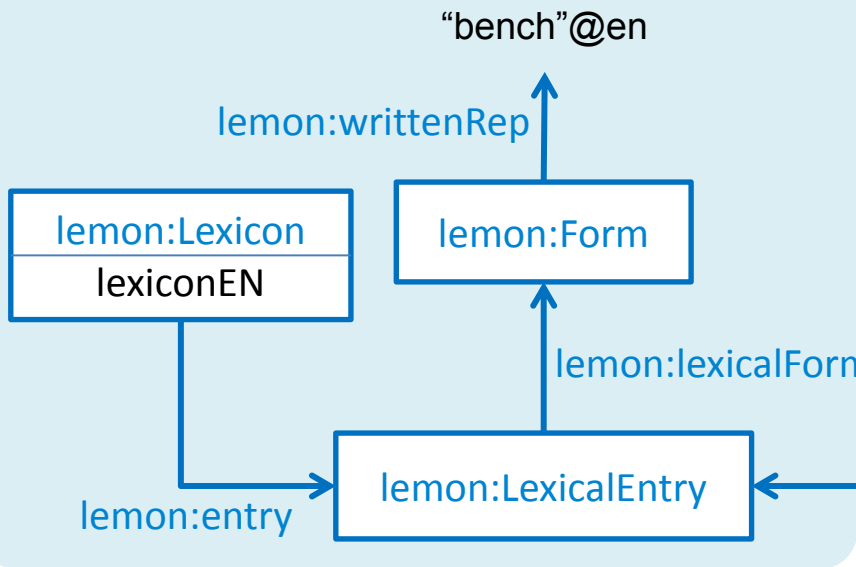
<http://purl.org/net/translation.owl>



# Modeling lexicons with Lemon



# Translation example



Specification

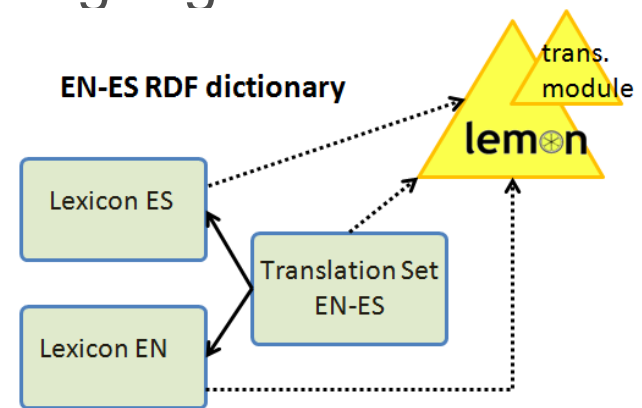
Modelling

RDF  
Generation

RDF generation using Open Refine

For each bilingual dictionary:

1. Lexicon of the source language
2. Lexicon of the target
3. Translations



```
apertium:lexiconEN a lemon:Lexicon ;
dc:source <http://hdl.handle.net/10230/17110> .
...
apertium:lexiconEN lemon:entry apertium:lexiconEN/bench-n-en .

apertium:lexiconEN/bench-n-en a lemon:LexicalEntry ;
lemon:lexicalForm apertium:lexiconEN/bench-n-en-form ;
lexinfo:partOfSpeech lexinfo:noun .

apertium:lexiconEN/bench-n-en-form a lemon:Form ;
lemon:writtenRep "bench"@en .
```



# Apertium RDF: 22 generated datasets

Specification

Modelling

RDF  
Generation

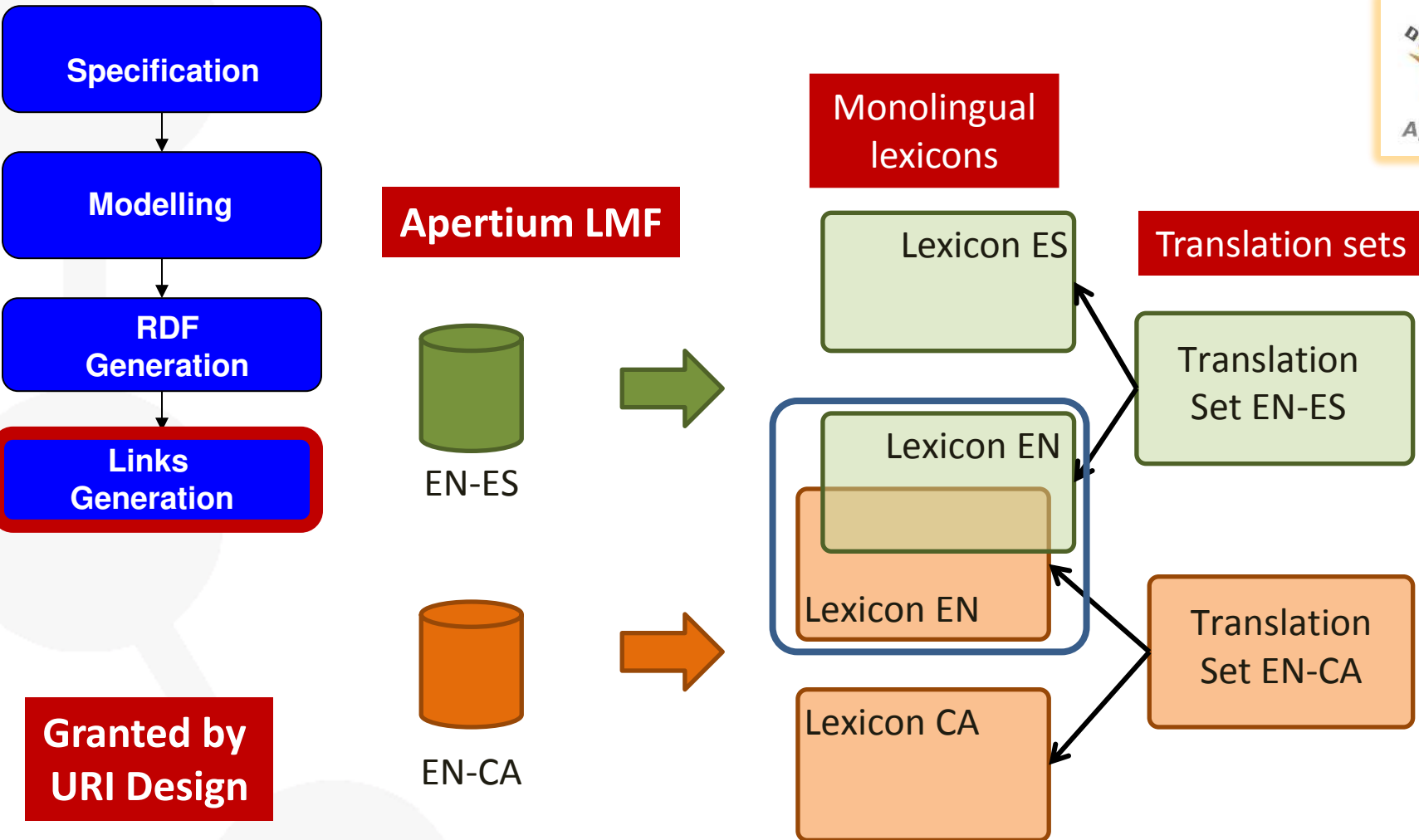
Lang. pair	# triples	# trans.
CA-IT	180,851	7,869
EN-CA	759,601	33,029
EN-ES	576,316	25,830
EN-GL	425,117	20,034
EO-CA	426,301	19,964
EO-EN	617,772	31,474
EO-ES	380,198	17,212
EO-FR	726,281	35,791
ES-AN	71,997	3,110
ES-AST	825,54	36,096
ES-CA	730,501	31,291

Lang. pair	# triples	# trans.
ES-GL	206,284	8,985
ES-PT	279,245	12,054
ES-RO	400,366	17,318
EU-ES	262,336	11,838
EU-EN	265,466	13,089
FR-CA	152,002	6,550
FR-ES	495,614	21,475
OC-CA	346,346	15,983
OC-ES	317,162	14,561
PT-CA	163,149	7,111
PT-GL	234,065	10,144





# Linking with other Apertium RDF



# Apertium English lexicon:

<http://linguistic.linkeddata.es/id/apertium/lexiconEN>

LexiconES

“ribera”@es

ribera

orilla

“orilla”@es

banco

“banco”@es

TranslationSetEN-ES

bank-ribera

bank-orilla

bank-banco

bench-banco

TranslationSetES-PT

banco-banco

orilla-orla

LexiconEN

bank

“bank”@en

bench

“bench”@en

“banco”@pt

banco

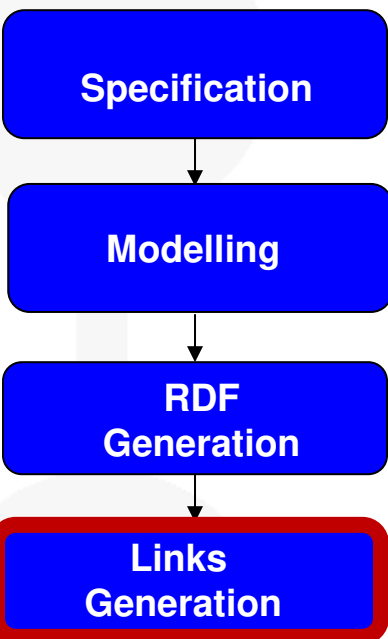
orla

“orla”@pt

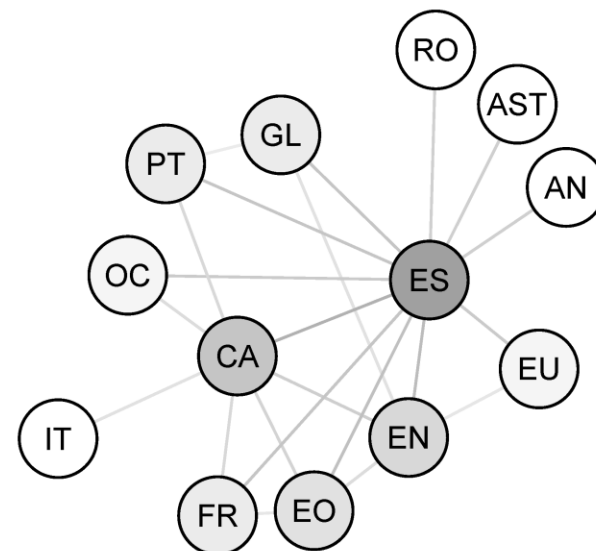
LexiconPT



# Added value of linking Apertiums



- Apertium RDF Graph



- **Dijkstra** algorithm to choose shortest path
- Indirect translations for “bank” EN-> ES -> PT

Pivot translation written repres.	Indirect translation written repres.
"banco"@es	"banco"@pt
"orilla"@es	"orla"@pt





# Links between Apertiums RDF and links with Babelnet

## Direct translations for “bank”@en

Specification

Modelling

RDF  
Generation

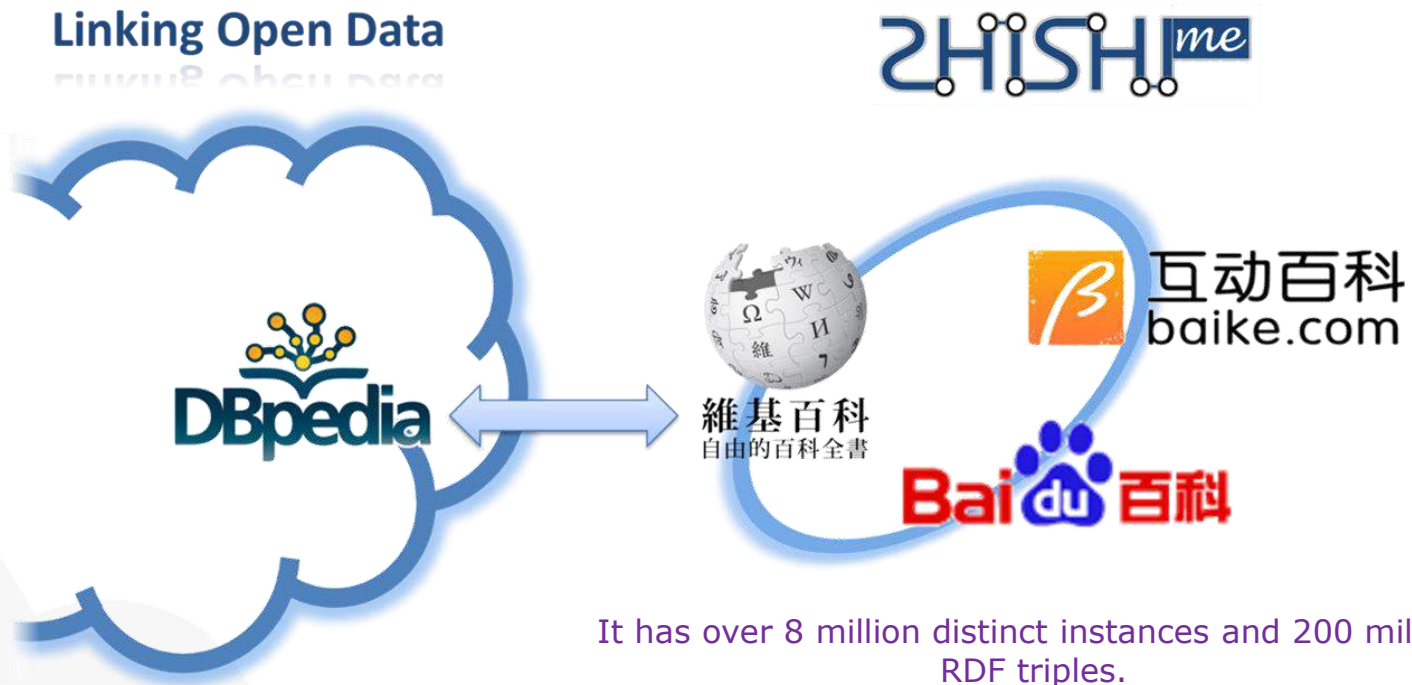
Links  
Generation

Translated Written Repr.	BabelSynset	BabelNet gloss
"banco" @es	<a href="http://babelnet.org/rdf/s00008371n">http://babelnet.org/rdf/s00008371n</a>	“A building in which the business of banking transacted”
"banco" @es	<a href="http://babelnet.org/rdf/s00008366n">http://babelnet.org/rdf/s00008366n</a>	“An arrangement of similar objects in a row or in tiers”
"banco" @es	<a href="http://babelnet.org/rdf/s15346085n">http://babelnet.org/rdf/s15346085n</a>	“An ocean bank, sometimes referred to as a fishing bank or simply bank, ...”
...	...	...
"orilla" @es	<a href="http://babelnet.org/rdf/s00008363n">http://babelnet.org/rdf/s00008363n</a>	“Sloping land (especially the slope beside a body of water)”
"ribera" @es	<a href="http://babelnet.org/rdf/s00008363n">http://babelnet.org/rdf/s00008363n</a>	“Sloping land (especially the slope beside a body of water)”



# Zhishi.me Lexicalization and linking

- Zhishi.me (<http://zhishi.me>)
- First effort to publish large scale Chinese semantic data as Linked Open Data

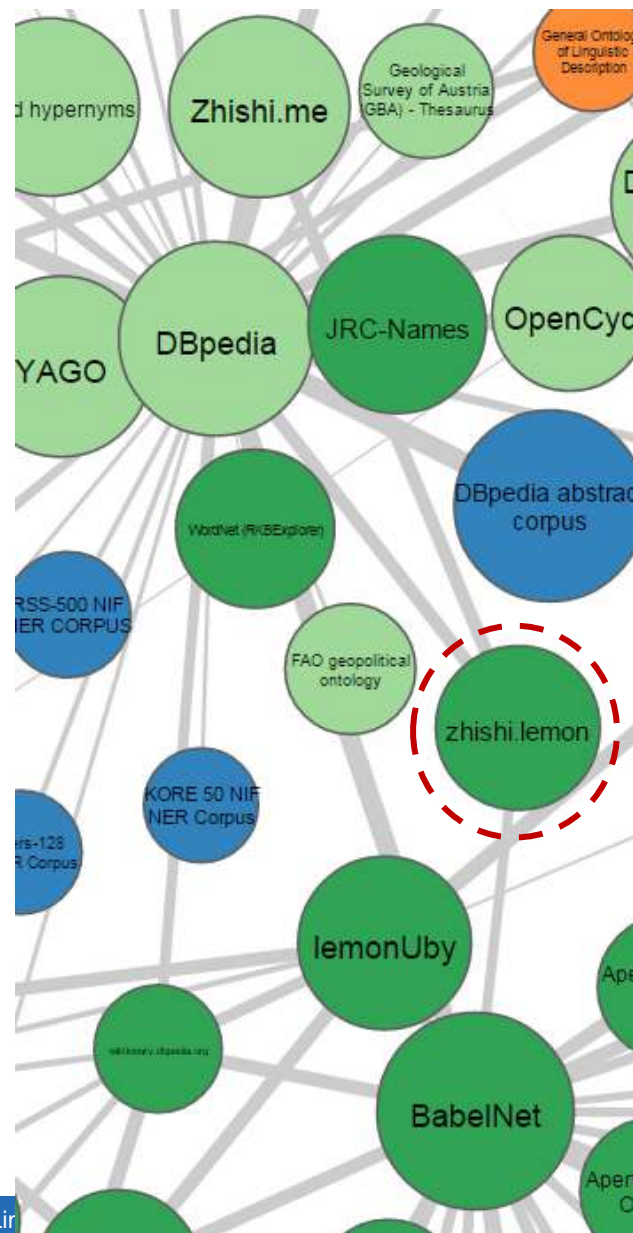


It has over 8 million distinct instances and 200 million RDF triples.

Acknowledgements: Zhijia Fang and Haofen Wang  
East China University of Science and Technology (ECUST)

## Lexicalisation of Zhishi.me

- ECUST (China) and UPM (Spain) collaboration
- New dataset created Zhishi.lemon with:
  - The lexical layer of Zhishi.me in pinyin, traditional, and simplified Chinese
  - Translations into English and Spanish
  - Links to DBpedia, esDBpedia, and BabelNet
- For the links to BabelNet, a disambiguation algorithm based on the *categories* was used
- Initially, 16.000 links to BabelNet were created. More to come in the future



- Apertium data on the Web following **SW** standards
- Common **entry point** for all the Apertium dictionaries
- Unique graph that contains direct and indirect **translations**
- **Linked with BabelNet**



**Language Resources**  
**The need of connecting licensed Language Resources**  
**Linked Data foundations**  
**Linguistic Linked Licensed Data**  
Outcomes of the LIDER project  
The apertium RDF case and their links

# **LINGUISTIC LINKED LICENSED DATA IN USE**



# Multilingual search: retrieve an author

- Servantes Saavedra, Migel
- Therbantes, Minkel nte
- **Cervantes di Saavedra, Michele**
- سرفنتس ساقدرا، میجیل دی
- Zerbantes eta Saabedra, Mikel
- Θερβάντες, Μιγκέλ ντε
- Cervantes
- Sirfantis Saafedrā, Mīgīl dī
- Сервантес Сааведра, Мигель де
- Sewantisi Saweidela, Migai'er de
- 塞万提斯·萨维德拉, 米盖尔 德
- Cervantes, Miguel de



Buscar una persona, grupo, entidad, título de una obra o un evento. Ej: Fernando de Alva, Oñá, Grapes of Wrath, Venice

## Cervantes Saavedra, Miguel de (1547-1616)

Miguel de Cervantes Saavedra fue un soldado, novelista, poeta y dramaturgo español. Supóñese que nació el 29 de septiembre de 1547 en Alcalá de Henares y murió el 22 de abril de 1616 en Madrid, pero fue enterrado el 23 de abril según se anunció en forma pública en la catedral de su ciudad natal. Fue considerado la máxima figura de la literatura española. Fue un escritor muy versátil, sobre todo por haber escrito 11 ingeniosos trabajos. Ocho de la novela, que incluye obras como el ingenioso hidalgo don Quijote de la Mancha, que resultó más exitosa que ninguna otra novela moderna, y una de las mejores obras de la literatura universal. Se le ha dado el sobrenombre de Príncipe de los Ingenios. [Ver más información de Wikipedia](#)



En la BNE

Autor de  
149  
Obras

Tema en  
116  
Obras

Participa en  
158  
Obras

### Sus obras


### Obras sobre la persona


## DATOS·BNE·ES <sup>\*beta</sup>

Inicio

Inicio Obras Temas Ayuda

سافران، میچیل دی

塞万提斯·萨维德拉, 米盖尔·德

Cervantes di Saavedra, Michele

### Filtrar por tipo (2 resultados)

1 Obras 1 Personas

L 3 Ediciones

2 de 2

## Cervantes Saavedra, Miguel de (1547-1616)

Miguel de Cervantes Saavedra fue un soldado, novelista, poeta y dramaturgo español. Supóñese que nació el 29 de septiembre de 1547 en Alcalá de Henares y murió el 22 de abril de 1616 en ...



Il comune di Napoli e il Comitato promotore delle onoranze a Cervantes si pregiano invitare la E.V. alle cerimonie che avranno luogo il 2 maggio 1959 in occasione dello scoprimento della lapide ...

2 maggio 1959

Obra

### Obras destacadas

Don Quijote de la Mancha

- Search a book in Spanish about the “Athénée de Madrid”



# Linguistic Linked Data allows Multilingual data integration



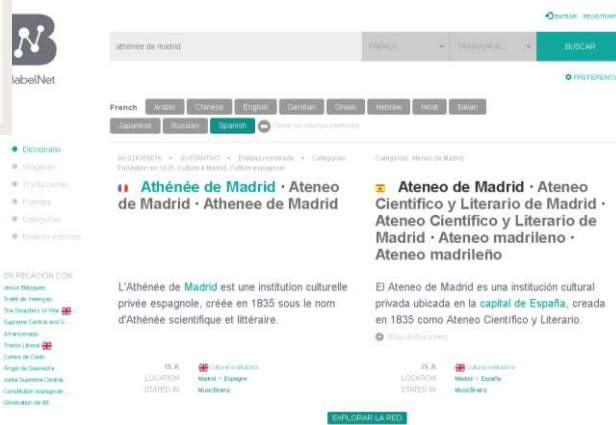
data.bnf.fr

Multilingual complementary information



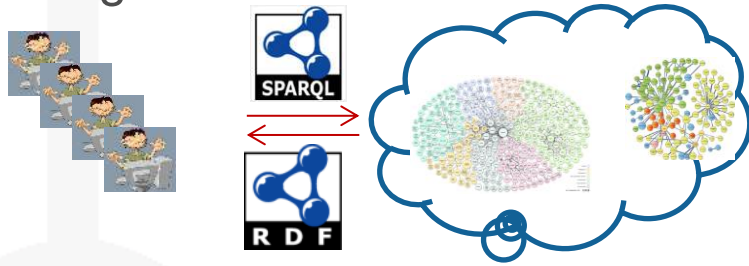
search results for "Athénée de Madrid"

There were no results matching the query. You can also search Athénée de Madrid in :  
Le Catalogue général de la BnF  
La base BnF Archives et Manuscrits  
La bibliothèque numérique Gallica

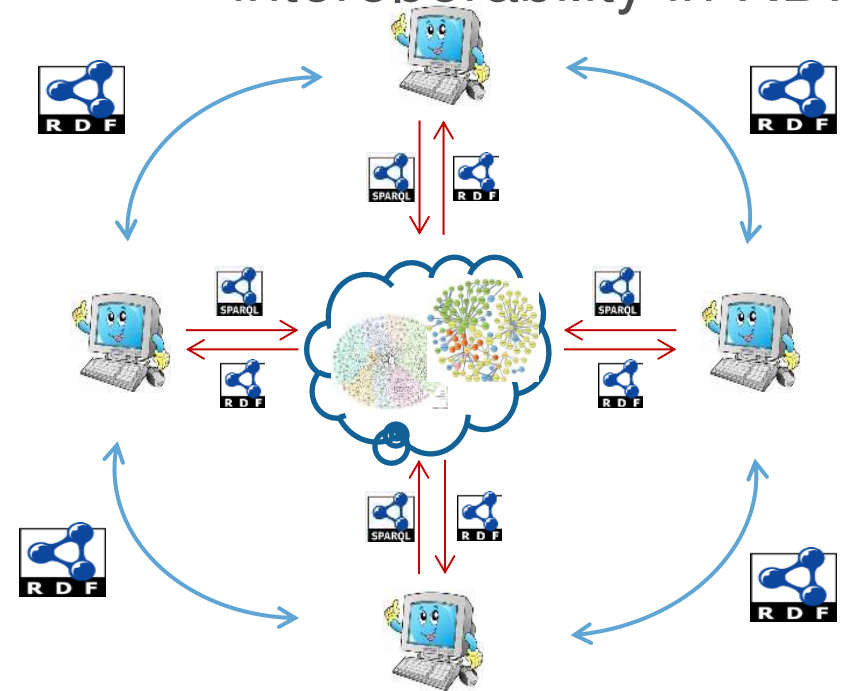


# Uses and Users of Linked Data

1. **Programmers** built applications making queries in SPARQL and get RDF



3. **Machine – Machine** data exchange and semantic interoperability in RDF

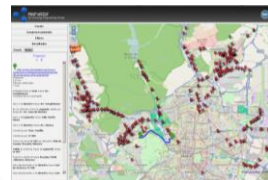


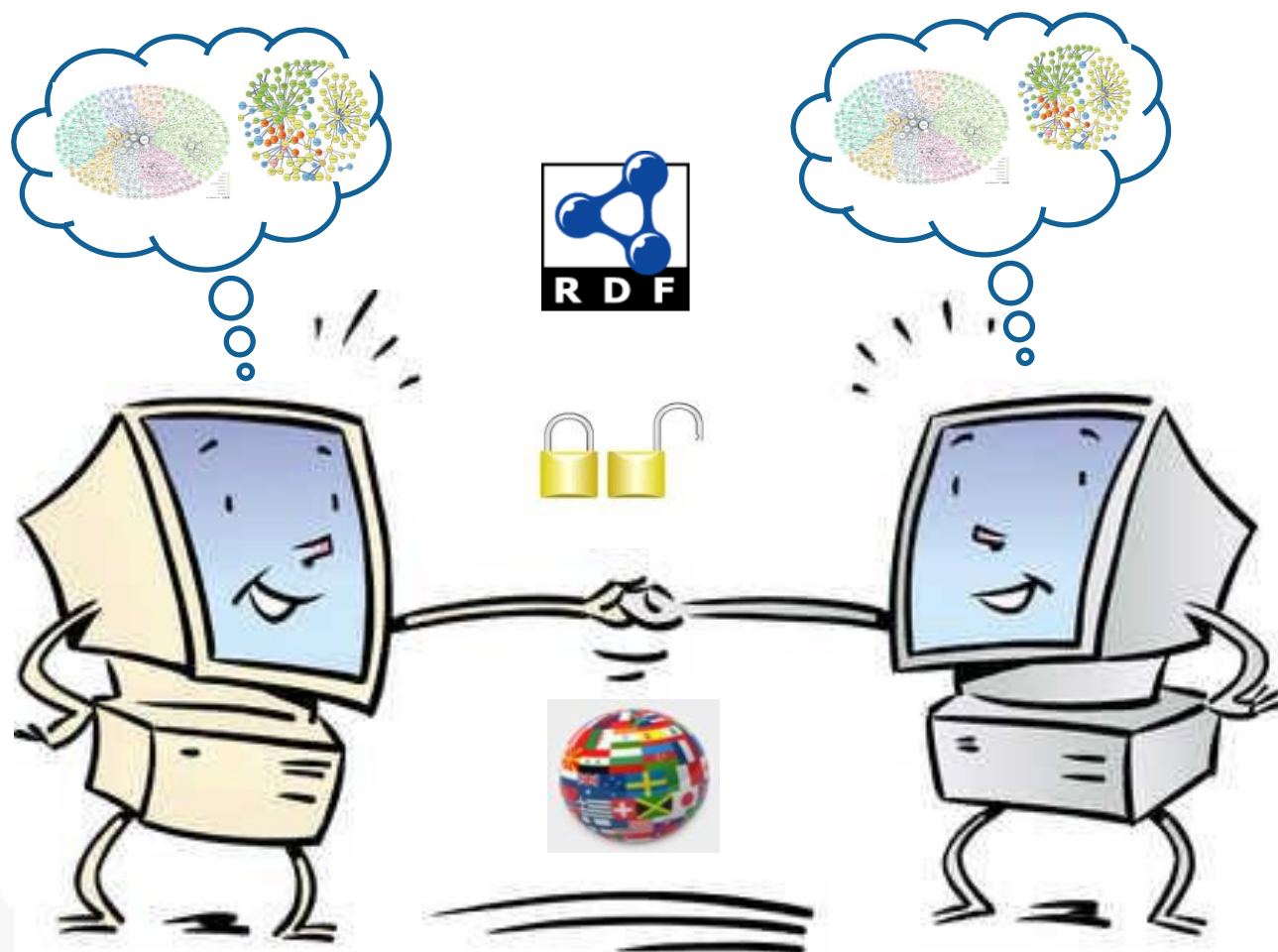
2. **Citizens/Users** access LD through a user interface (they do not see RDF)

**Culture**

**Geographical**

**Smart Cities**







Thanks for your attention !