

# BabelNet: a Multilingual Encyclopedic Dictionary as LOD

Roberto Navigli

DIPARTIMENTO  
DI INFORMATICA

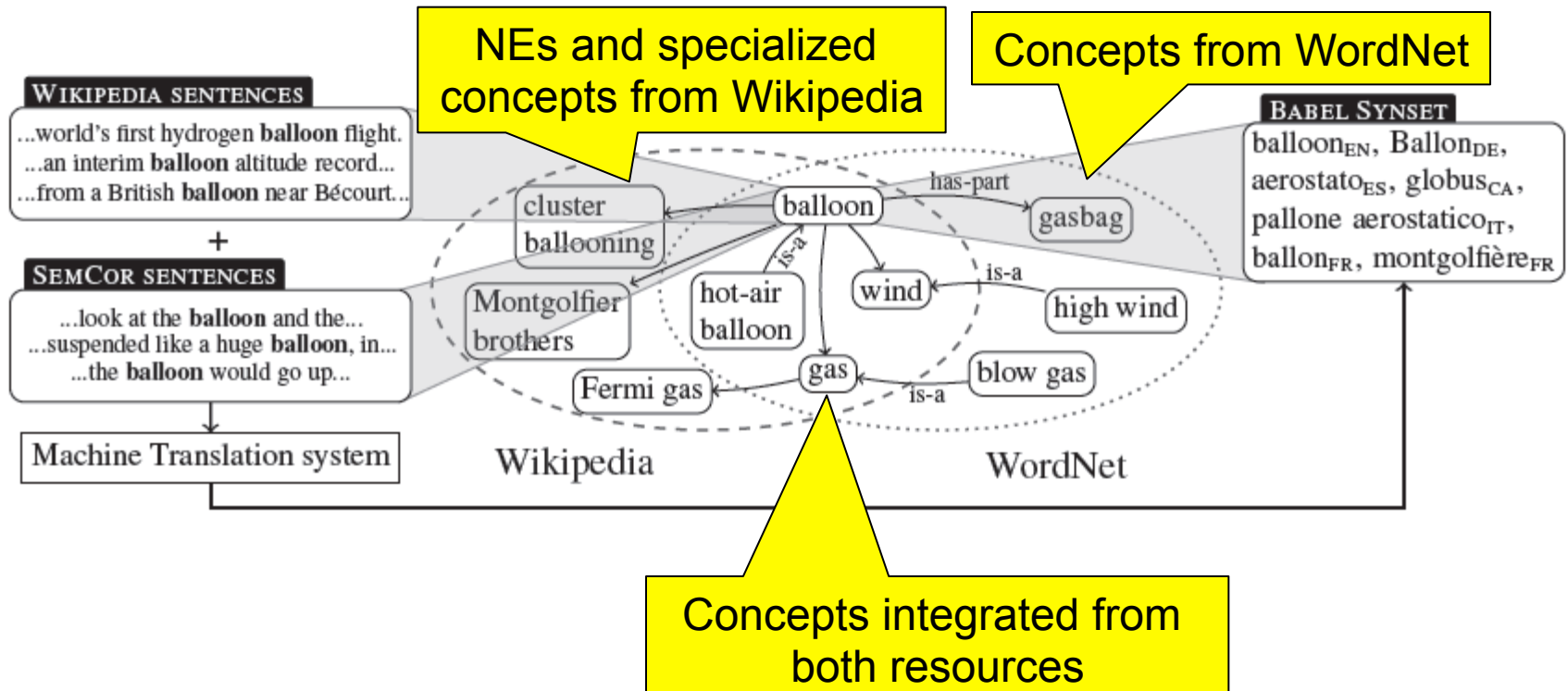


SAPIENZA  
UNIVERSITÀ DI ROMA

Linguistic Computing Laboratory  
<http://lcl.uniroma1.it>

# BabelNet

- A wide-coverage multilingual semantic network including both **encyclopedic** (from Wikipedia) and **lexicographic** (from WordNet) entries



Roberto Navigli and Simone Ponzetto: BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. **Artificial Intelligence**, 193, Elsevier, 2012.

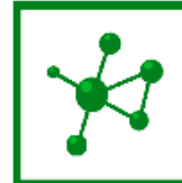
# BabelNet<sup>1.1.1</sup>

A very large multilingual ontology

**means:** more accurate mappings and translations + Wikipedia categories + images + ...

A very large multilingual ontology with 5.5 million concepts • A wide-coverage "encyclopedic dictionary" • Obtained from the automatic integration of WordNet and Wikipedia • Enriched with automatic translations of its concepts • Connected to the Linguistic Linked Open Data cloud!

search explore



PUBLICATIONS DOWNLOADS



BabelNet is an output of the MultiJEDI ERC Starting Grant No. 259234. Concept and application by Roberto Navigli.

BabelNet and its API are licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.



For any commercial use, please contact us.



<http://babelnet.org>

# BabelNet<sup>1.1.1</sup>









A very large multilingual ontology

search · explore · publications · download

Q Type a **term**:  English

(**examples**: plane, apple, star, Italian, bus driver, calcio, drive#n, bus#n#en, horse#en, mela#it)

## Noun

<b>Title:</b> Multilingual Web	•	<b>ID:</b> bn:02289651n	•	<b>Type:</b> Named Entity	 explore
<b>Senses:</b>			<b>Multilingual Web</b>		
<b>Glosses:</b>			The Multilingual Web is an EU funded project whose primary goal is to provide: "Multilingual Web content management: standards and best practices".		
<b>Categories:</b>			European Union and science and technology		
<b>DBPedia:</b>		Multilingual Web			



BabelNet is an output of the MultiJEDI ERC Starting Grant No. 259234. Concept and application by Roberto Navigli.

BabelNet and its API are licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.




For any commercial use, please [contact us](#).





<http://babelnet.org>

(examples: plane, apple, star, Italian, bus driver, calcio, drive#n, bus#n#en, horse#en, mela#it)

## Noun







**Title:** web<sup>1</sup> • **ID:** bn:00080771n • **Type:** Concept  explore



**Senses:**   web<sup>1</sup>

**Glosses:**   an intricate network suggesting something that was formed by weaving or interweaving; "the trees cast a delicate web of shadows over the lawn"

**Title:** web<sup>2</sup> • **ID:** bn:00030977n • **Type:** Concept  explore



**Senses:**   web<sup>2</sup>, entanglement<sup>1</sup>


  embolic,  enchevêtrement,  verstrickung,  aggrovigliamento,  enredo

**Glosses:**   an intricate trap that entangles or ensnares its victim

**Title:** vane<sup>4</sup> • **ID:** bn:00079552n • **Type:** Concept  explore



**Senses:**   vane<sup>4</sup>, web<sup>3</sup>


**Glosses:**   the flattened weblike part of a feather consisting of a series of barbs on either side of the shaft



**Title:** network<sup>1</sup> • **ID:** bn:00057379n • **Type:** Concept  explore








**Senses:**   network<sup>1</sup>, web<sup>4</sup>




  xarxa,  réseau,  netzwerk,  rete,  red

**Glosses:**   an interconnected system of things or people; "he owned a network of shops"; "retirement meant dropping out of a whole network of people who had been part of my life"; "tangled in a web of cloth"

**Title:** World Wide Web<sup>1</sup> • **ID:** bn:00080772n • **Type:** Concept  explore

**Senses:**   World Wide Web<sup>1</sup>, WWW<sup>1</sup>, web<sup>5</sup>

  World Wide Web,  World Wide Web,  World Wide Web,  World Wide Web,  World Wide Web,  World Wide Web

  Webs, WWW, Web, Teranyina mundial, W3,  The Web, World-Wide Web,



# Anatomy of BabelNet

5.5 million Babel synsets if we consider synsets with no English lexicalization

- 6 languages covered (moving to 40+)
- More than 3 million Babel synsets (i.e. concepts and NE)
- More than 26 million word senses:

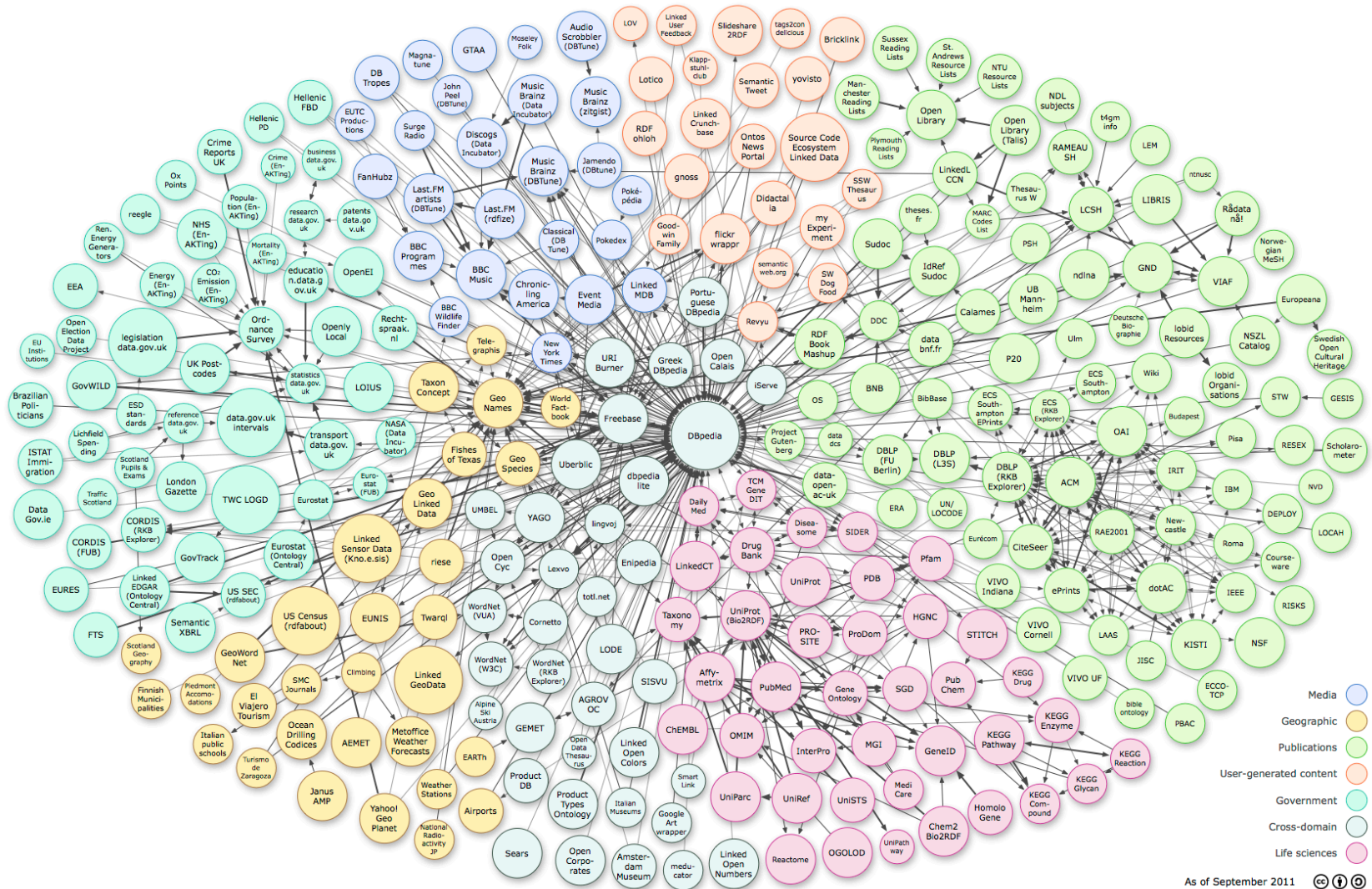
	English	Catalan	French	German	Italian	Spanish	Total
English WordNet	206,978	-	-	-	-	-	206,978
Wikipedia	pages	123,101	524,897	506,892	404,153	349,375	4,863,970
	redirections	105,147	617,379	456,977	217,963	404,009	5,189,524
	translations	3,445,470	2,844,751	2,841,916	3,046,325	3,083,427	15,261,889
WordNet	monosemous	97,876	98,081	97,672	98,475	98,092	490,196
	SemCor	6,852	6,855	6,850	6,856	6,855	34,268
<b>Total</b>	<b>6,550,579</b>	<b>3,778,446</b>	<b>4,091,963</b>	<b>3,910,307</b>	<b>3,773,772</b>	<b>3,941,758</b>	<b>26,046,825</b>

- About 70 million lexico-semantic relations:

	English	Catalan	French	German	Italian	Spanish	Total
WordNet	364,552	-	-	-	-	-	364,552
WordNet glosses	617,785	-	-	-	-	-	617,785
Wikipedia	50,104,898	971,379	5,594,590	5,931,099	3,598,733	3,397,754	69,598,453
<b>Total</b>	<b>51,087,235</b>	<b>971,379</b>	<b>5,594,590</b>	<b>5,931,099</b>	<b>3,598,733</b>	<b>3,397,754</b>	<b>70,580,790</b>



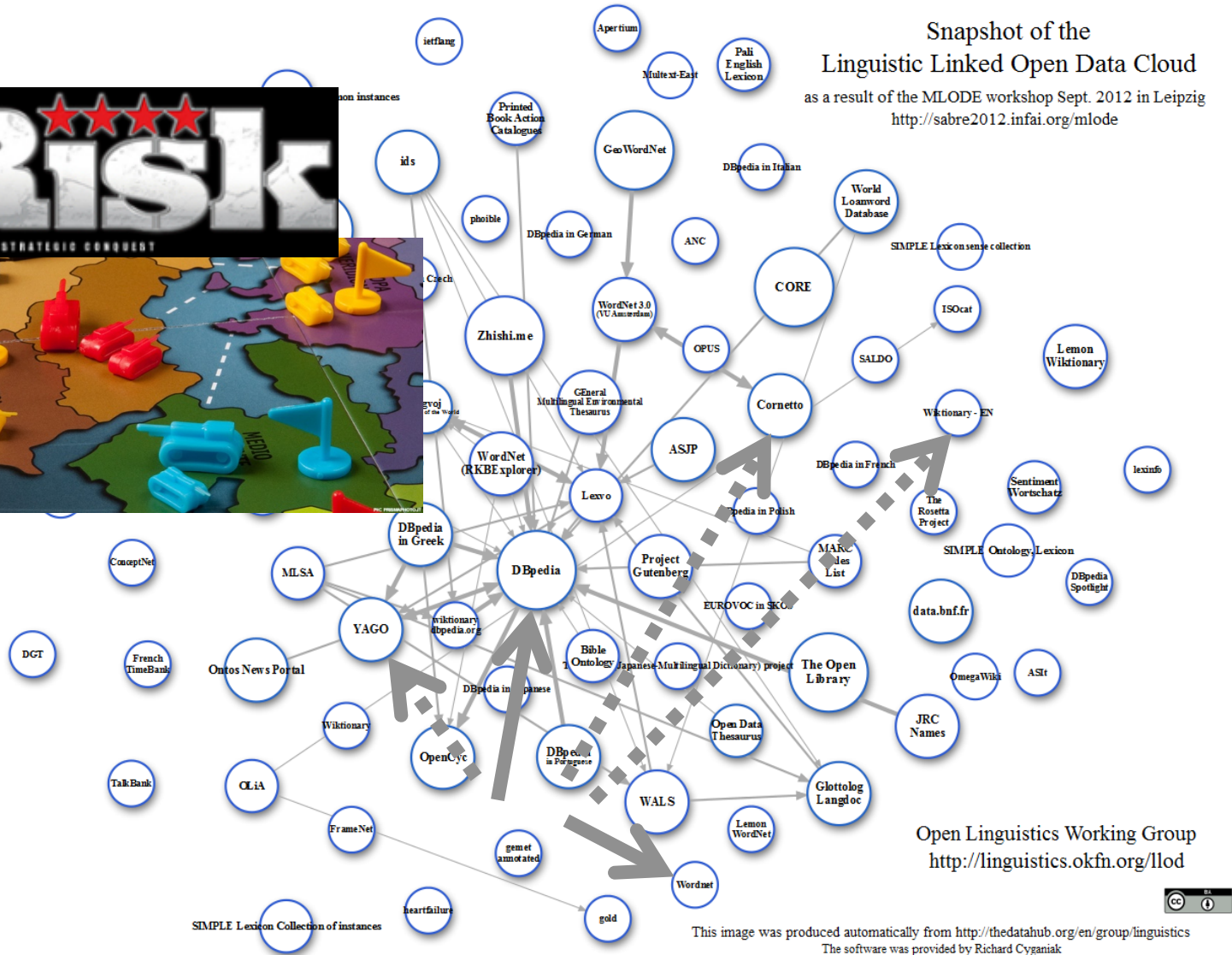
# The Linked Open Data cloud...



# The \*Linguistic\* Linked Open Data cloud...



Snapshot of the  
Linguistic Linked Open Data Cloud  
as a result of the MLODE workshop Sept. 2012 in Leipzig  
<http://sabre2012.infai.org/mlode>



Open Linguistics Working Group  
<http://linguistics.okfn.org/llod>

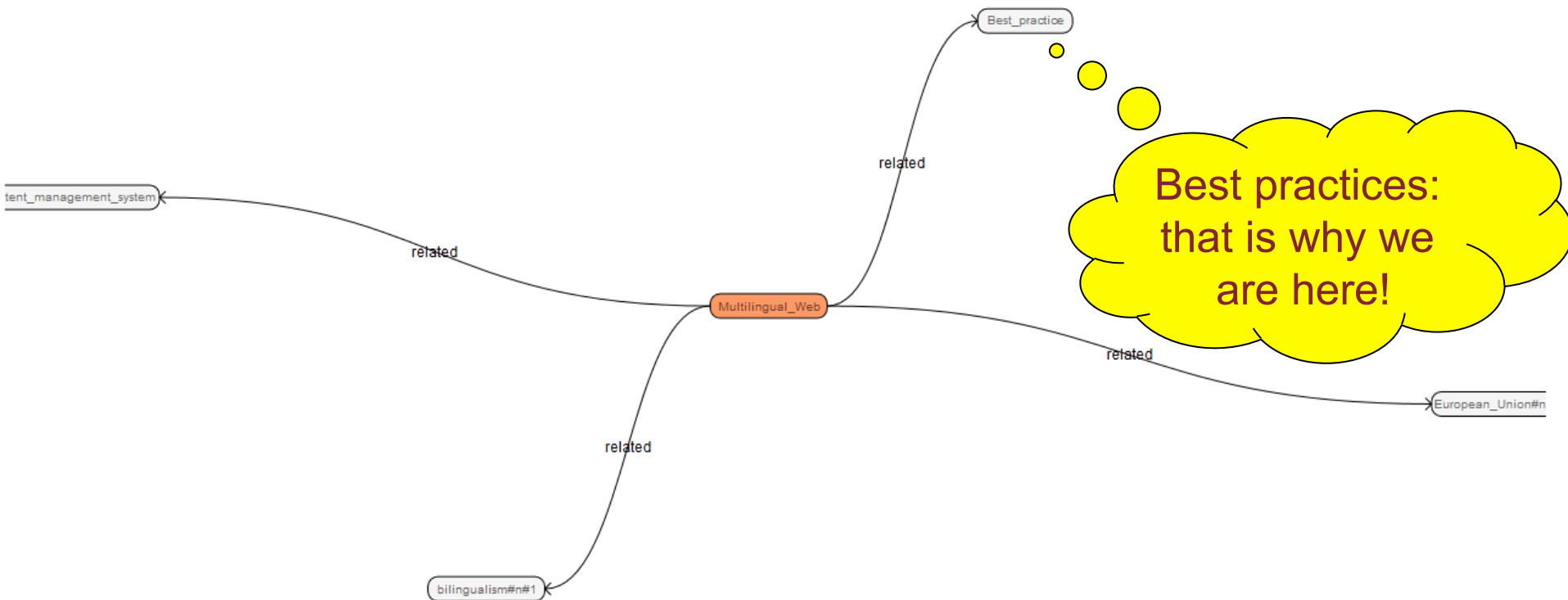


This image was produced automatically from <http://thedatahub.org/en/group/linguistics>  
The software was provided by Richard Cyganiak





# And... centered on the Multilingual Web!



# The contribution of BabelNet to the LOD

- Going to be available in **lemon** Lexicon Model for Ontologies RDF
- A real example of large **multilingual LOD** which could benefit from our discussion today
- Providing **different lexicalizations** for the same concepts in different languages
- Leveraging BabelNet for non-English LOD
- Encyclopedic and lexicographic information available in many languages used for **linking** LD multilingually?



SAPIENZA  
UNIVERSITÀ DI ROMA

**Roberto Navigli**

Linguistic Computing Laboratory

<http://lcl.uniroma1.it>

