

# Terminology and LLOD

Dagmar Gromann, University of Vienna, [dagmar.gromann@univie.ac.at](mailto:dagmar.gromann@univie.ac.at)  
SD-LLD-22, 1 June 2022

# Overview

01

## Terminology

What is terminology and what is it good for?

02

## Benefits of LLOD

How does terminology benefit from LLOD?

03

## Resources

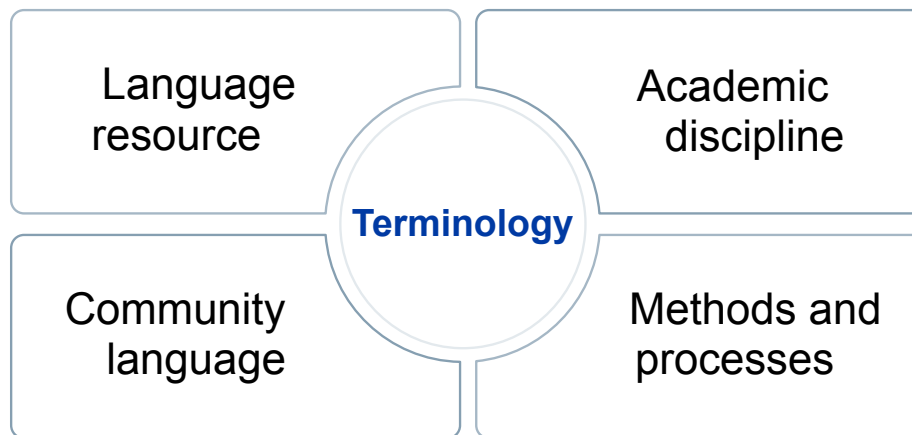
Which resources are there for terminology and LLOD?

04

## Tools

Which tools are there for LLOD and terminology?

# What is terminology?



As language resource:  
*A set of domain-specific terms, concepts  
and their relations in a specific language.*

★ 3548795

information technology and data processing [EDUCATION AND COMMUNICATIONS]

[en](#) [linked data](#)

<b>cs</b>	propojené údaje <b>OBSOLETE</b> propojená data <b>ADMITTED</b> propojitelná data <b>PREFERRED</b>
<b>es</b>	datos enlazados datos vinculados

InterActive Terminology  
for Europe (IATE)  
<https://iate.europa.eu/entry/result/>

# Key aspects of terminology

## Concept-oriented

Represents terms across languages by meaning

## Multilingual

Allows to group synonyms and equivalents

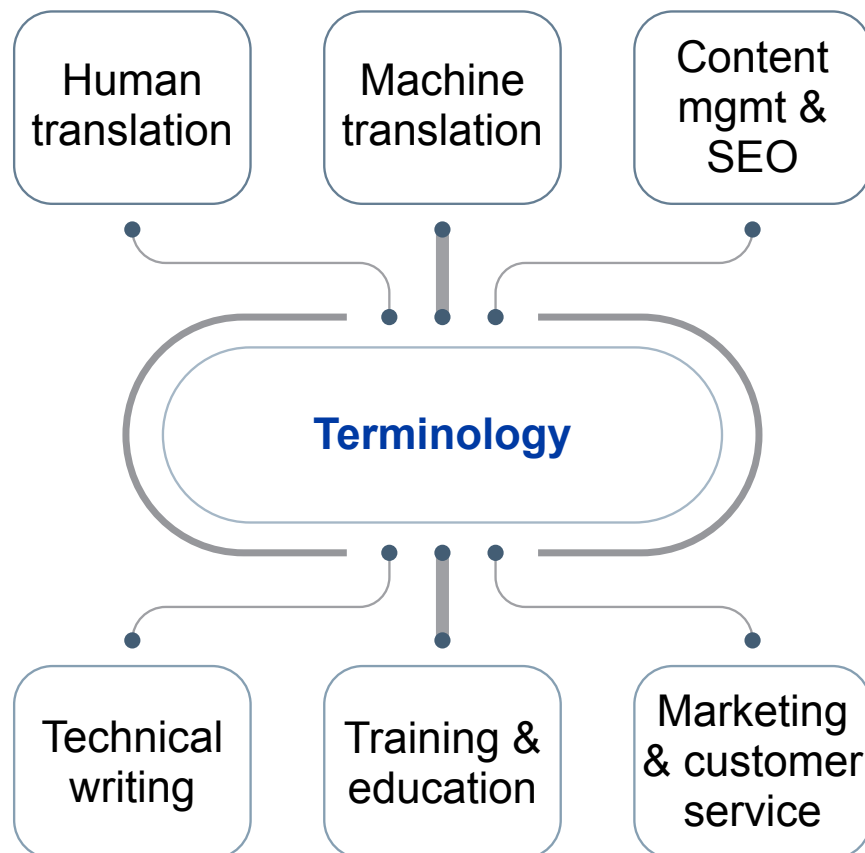
## Domain-specific

Information in a terminology (terms, definitions, etc.) specific to a domain


## Fine-grained

Allows to attach information to the meaning, language and/or term level

# Application areas of terminology



Origination: Created by super at 12/22/2015  
Last modification: terminologist at 01/14/2022  
Subject field: Engine  
Figure:



English (United Kingdom) (EN-GB)  
Definition: Engine with closed gas exchange in which a work cycle consists of four piston strokes with two crankshaft rotations. compression, working, and ejection.  
Process status: proposed

4-stroke engine  
✓  
Origination: Created by ktmlwenglerk at 06/18/2015  
Last modification: super at 12/22/2015  
Term status: preferred term

four-stroke-engine  
✗  
Origination: Created by super at 12/04/2015  
Last modification: ktmlwenglerk at 12/28/2015  
Term status: deprecated term

four-stroke-engine  
✗  
Origination: Created by super at 12/04/2015  
Last modification: ktmlwenglerk at 12/28/2015  
Term status: deprecated term

4-stroke engine  
✗  
Origination: Created by super at 12/04/2015

Example source:

<http://demo.kaleidoscope.at/terminology>



“Bad terminology is the enemy of  
good thinking.”

— **Warren Buffet**

# Main benefits of LLOD for terminology



## Linking terminologies

across domains and  
languages



## New applications

that can exploit multiple  
terminologies seamlessly



## Query terminologies

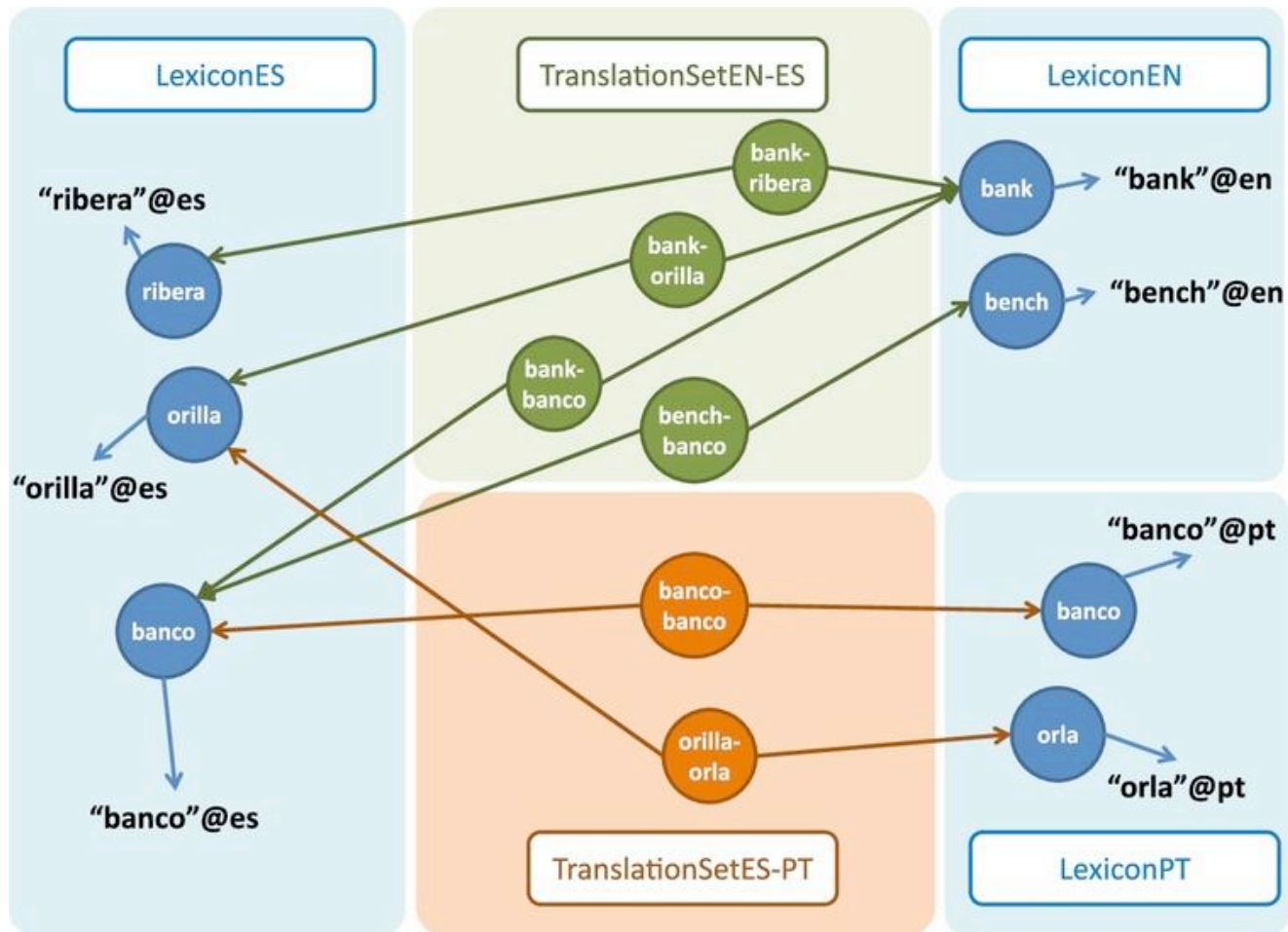
across domains and  
languages

# LLOD resources

<b>Linguistic annotations</b>	<a href="#"><u>Ontologies of Linguistic Annotation (OLiA)</u></a>	<a href="#"><u>General Ontology for Linguistic Description (GOLD)</u></a>	<a href="#"><u>Annohub</u></a>	...
<b>Terminologies</b>	<a href="#"><u>InterActive Terminology for Europe (IATE)</u></a>	<a href="#"><u>Apertium</u></a>	<a href="#"><u>BabelNet</u></a>	...
<b>Representation formats</b>	<a href="#"><u>OntoLex-Lemon</u></a>	<a href="#"><u>Simple Knowledge Organization System (SKOS(-XL))</u></a>		...
<b>Repositories</b>	<a href="#"><u>LLOD cloud</u></a>	<a href="#"><u>LingHub</u></a>	<a href="#"><u>European Language Grid (ELG)</u></a>	...

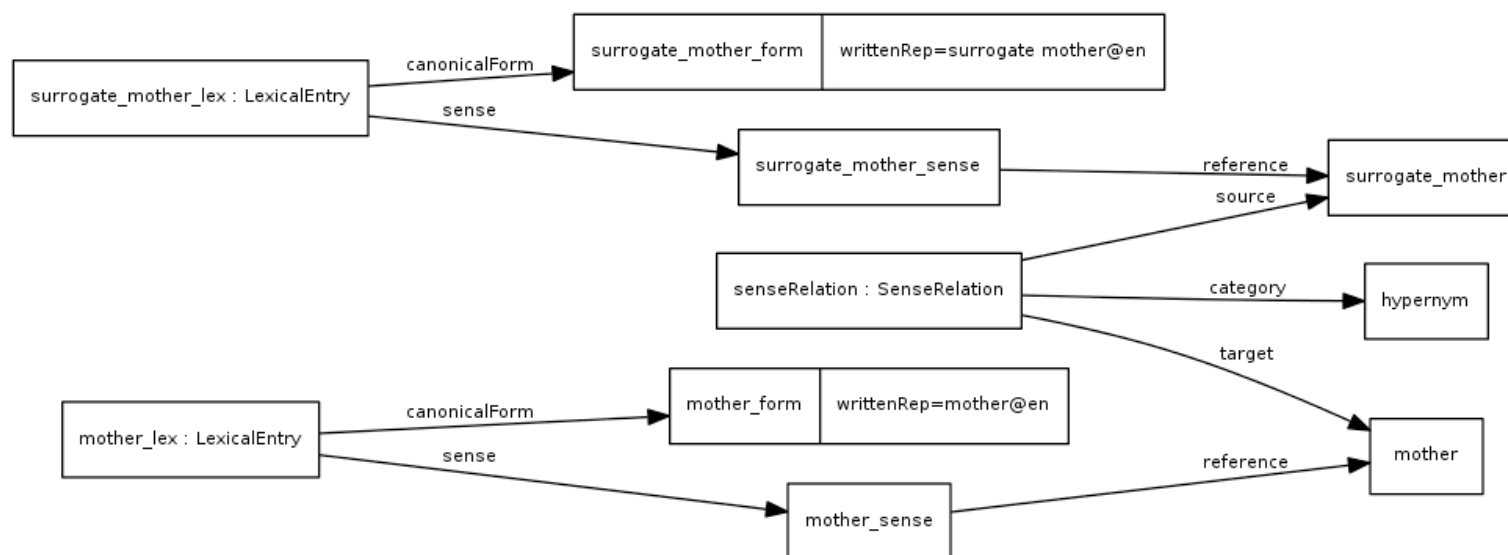


# Terminologies: Apertium example



Source: Gracia et al. (2018)

# Representation: OntoLex-Lemon



```
:surrogate_mother_lex a ontolex:LexicalEntry ;
  ontolex:sense :surrogate_mother_sense ;
  ontolex:canonicalForm :surrogate_mother_form.
```

```
:surrogate_mother_sense ontolex:reference <http://dbpedia.org/ontology/surrogate_mother>.
```

```
:surrogate_mother_form ontolex:writtenRep "surrogate mother"@en .
```

...

Source: <https://www.w3.org/2016/05/ontolex/>

# Representation: What is a concept?

## Ontology element

class, individual, property  
e.g. `ontolex:reference`  
<[http://dbpedia.org/ontology/surrogate\\_mother](http://dbpedia.org/ontology/surrogate_mother)>

## SKOS

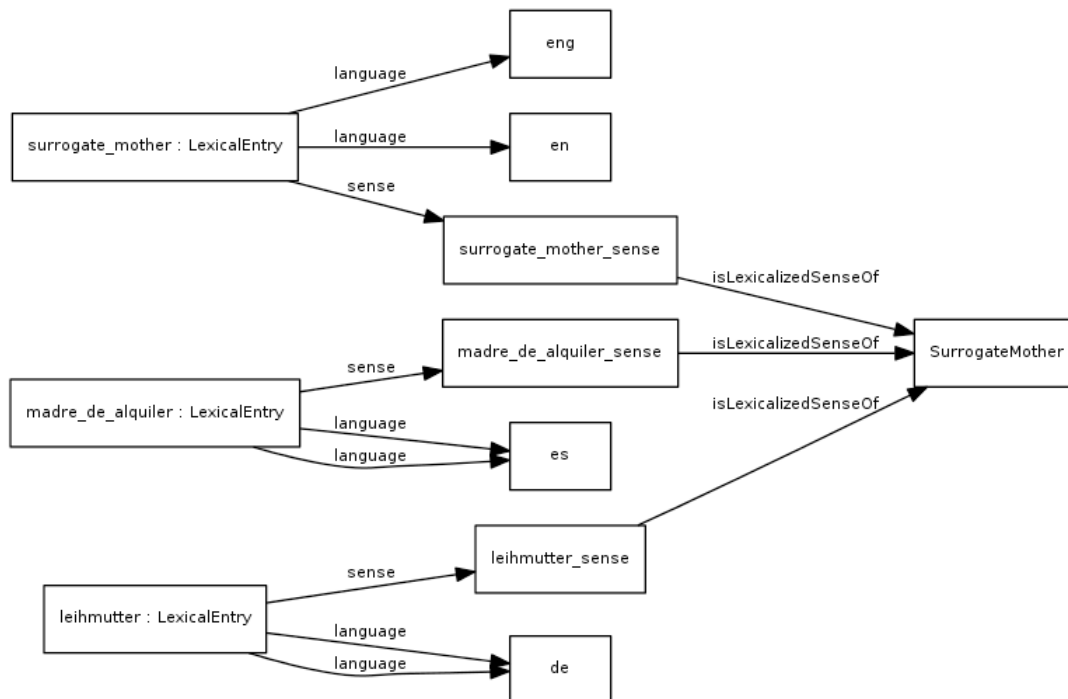
Lexical Concept  
e.g. `ontolex:evokes`  
`:Surrogate_Mother`

## Frames

FrameNet/MetaNet frames  
e.g. on [Framester](#) (Gangemi et al. 2016)

...

# Representation: vartrans extension of OntoLex-Lemon



```

:surrogate_mother a ontolex:LexicalEntry;
    dct:language <http://id.loc.gov/vocabulary/iso639-2/eng>,
    http://lexvo.org/id/iso639-1/en> ;
    ontolex:sense :surrogate_mother_sense.
  
```

```

:surrogate_mother_sense ontolex:reference ontology:SurrogateMot
  
```

```

:madre_de_alquiler a ontolex:LexicalEntry;
    dct:language <http://id.loc.gov/vocabulary/iso639-2/es>,
    http://lexvo.org/id/iso639-1/es> ;
    ontolex:sense :madre_de_alquiler_sense.
  
```

```

:madre_de_alquiler_sense ontolex:reference
ontology:SurrogateMother.
  
```

```

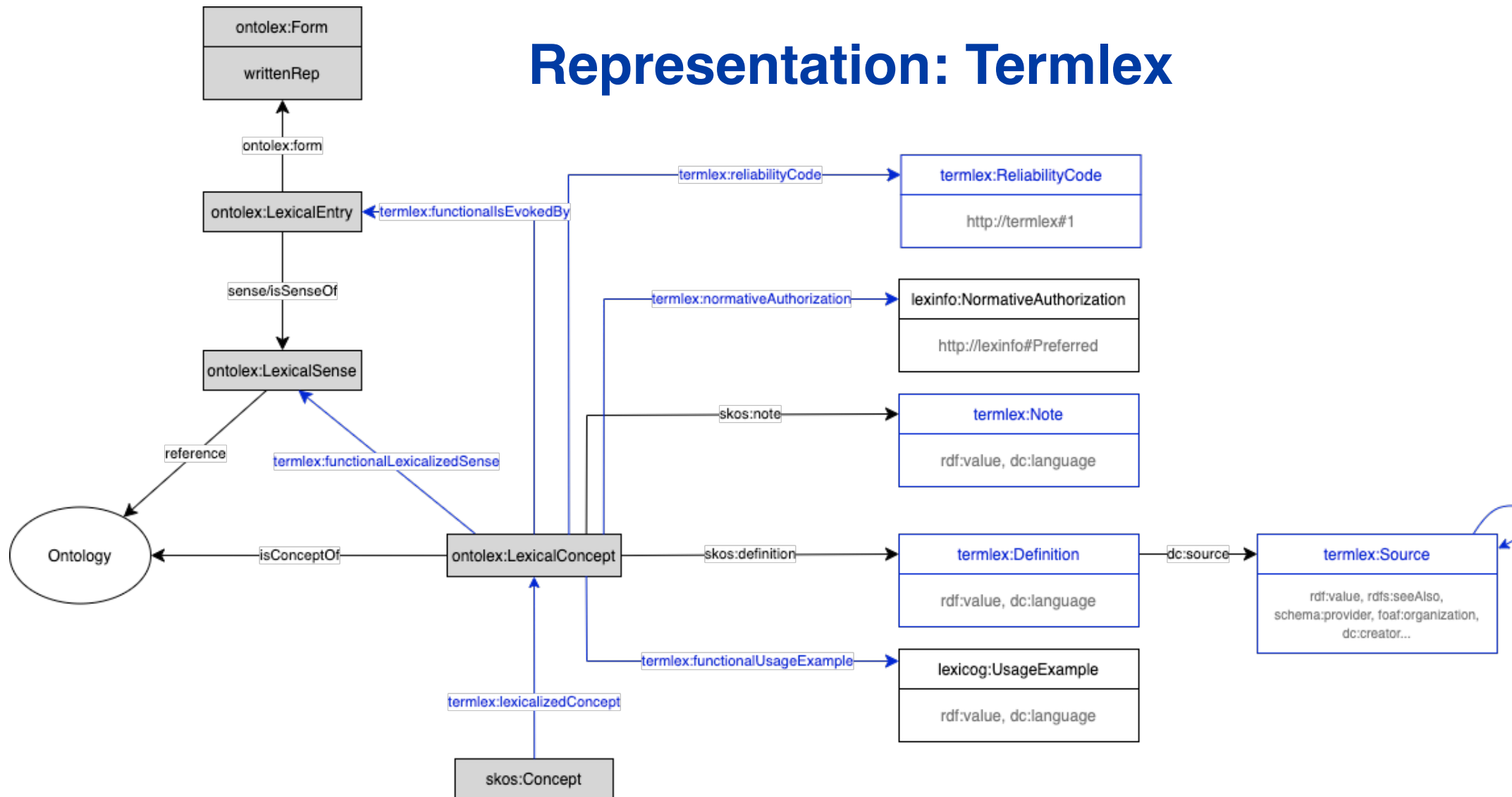
:leihmutter a ontolex:LexicalEntry;
    dct:language <http://id.loc.gov/vocabulary/iso639-2/de>,
    http://lexvo.org/id/iso639-1/de> ;
    ontolex:sense :leihmutter_sense.
  
```

```

:leihmutter_sense ontolex:reference ontology:SurrogateMother.
  
```

Source: <https://www.w3.org/2016/05/ontolex/>

# Representation: Termlex



Source: <https://www.w3.org/community/ontolex/wiki/Terminology>

# LLOD tools for terminologies

## TBX2RDF

Transforming TermBase eXchange (TBX) format to RDF (Cimiano et al. 2015)

## Terme-à-LLOD

Simplifying the Conversion and Hosting of Terminological Resources as Linked Data (Di Buono et al. 2020)

## TermltUp

Extracting terms from corpora and querying them via SPARQL (Martín-Chozas et al. Forthcoming)

## TermltUp Sparql

Endpoint for querying existing terminologies

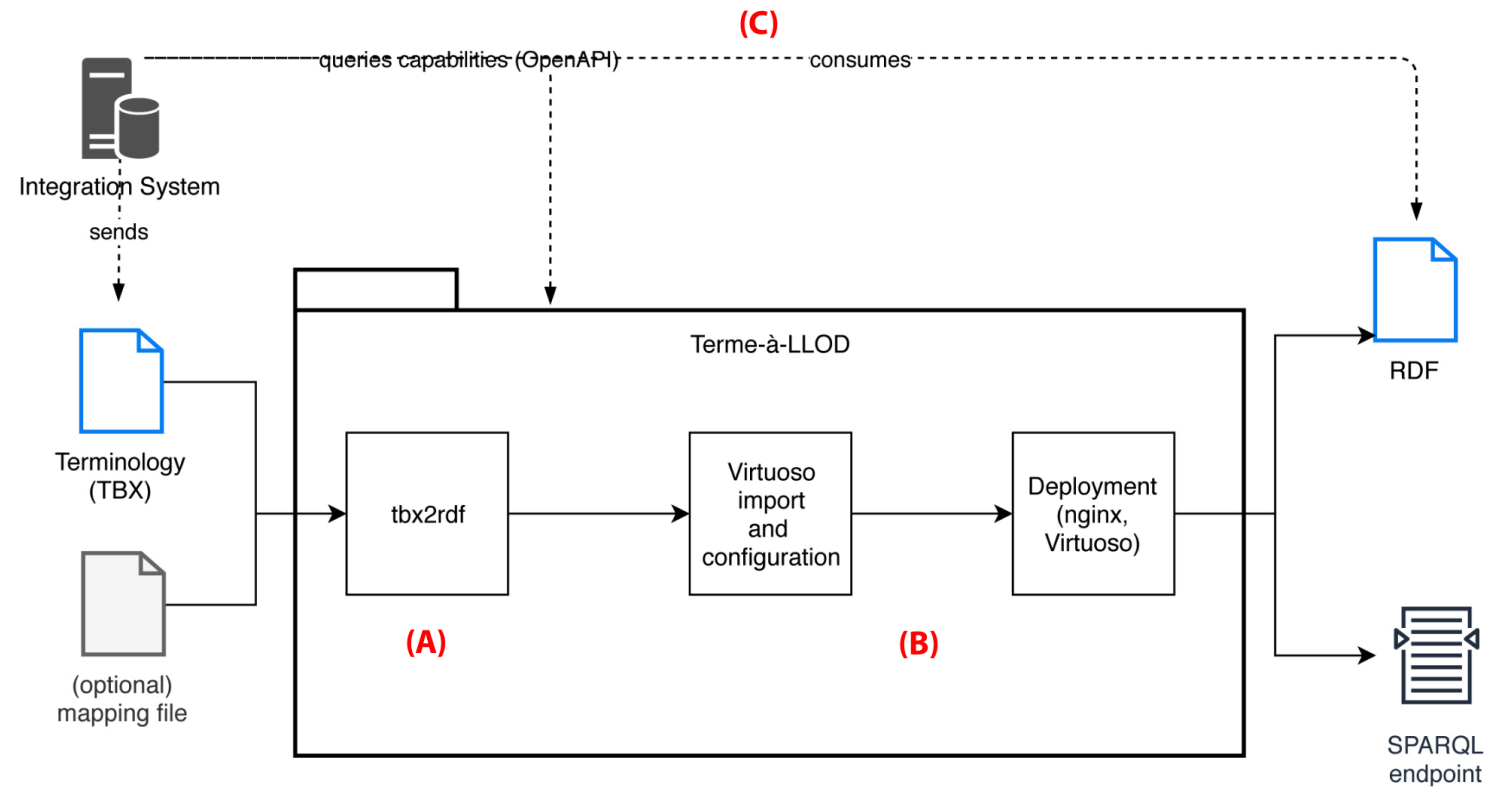
## Text2TCS

Extracts terminological concept maps as graphs and as TBX (soon as RDF)

...



# Term-à-LLOD



Source: di Buono et al. (2020)

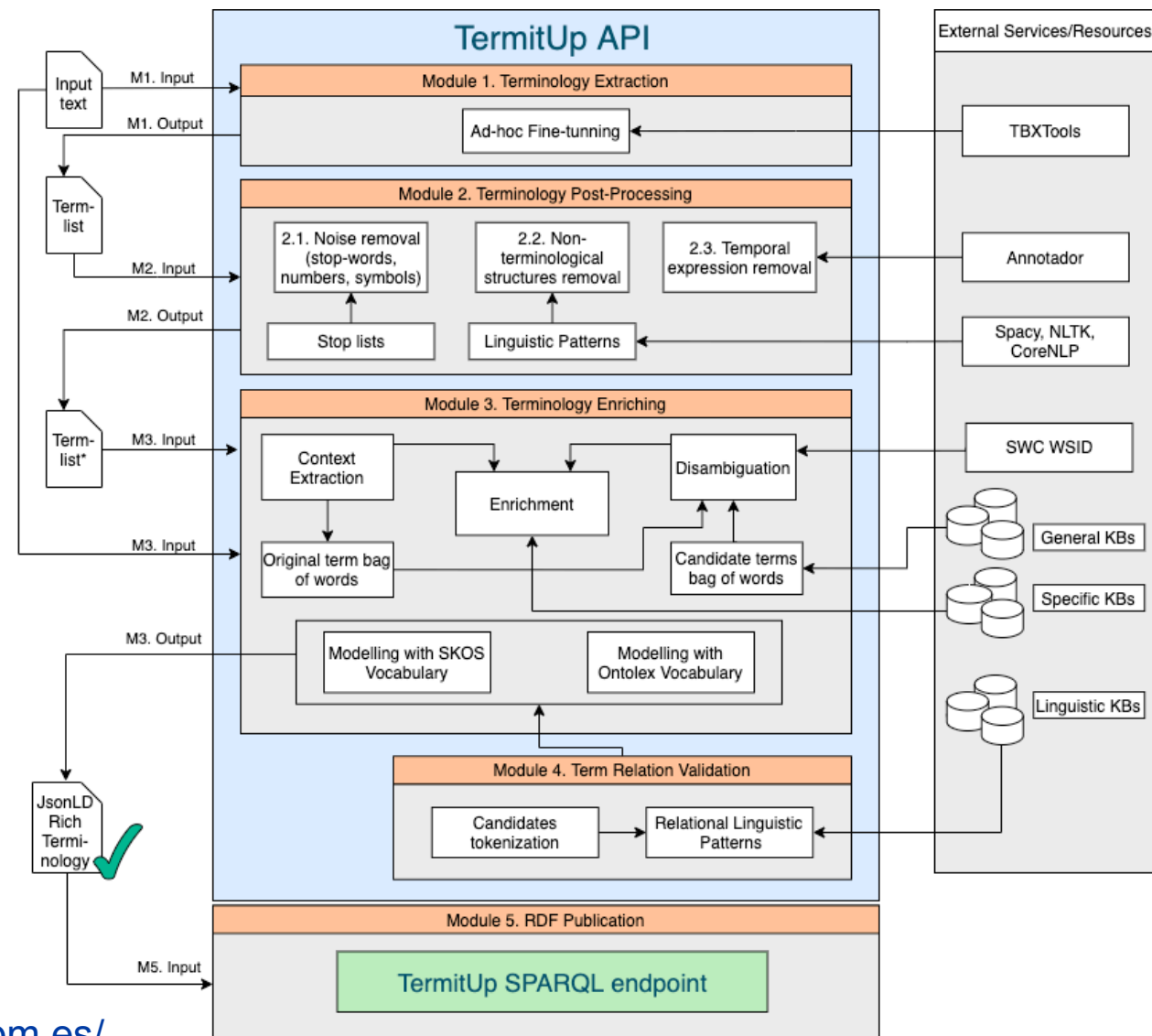


## Term-à-LLOD - Example

<b>Termbases</b>	<b>Lang</b>	<b>Number of Links</b>
GENTERM Pharmaceutical-IATE	English	1380
	Dutch	1084
GENTERM Diseases-IATE	English	22
	Dutch	27
GENTERM Waste management-IATE	English	114
	Dutch	109
GENTERM Solar energy-IATE	English	12
	Dutch	20
GENTERM Printmaking-IATE	English	35
	Dutch	21

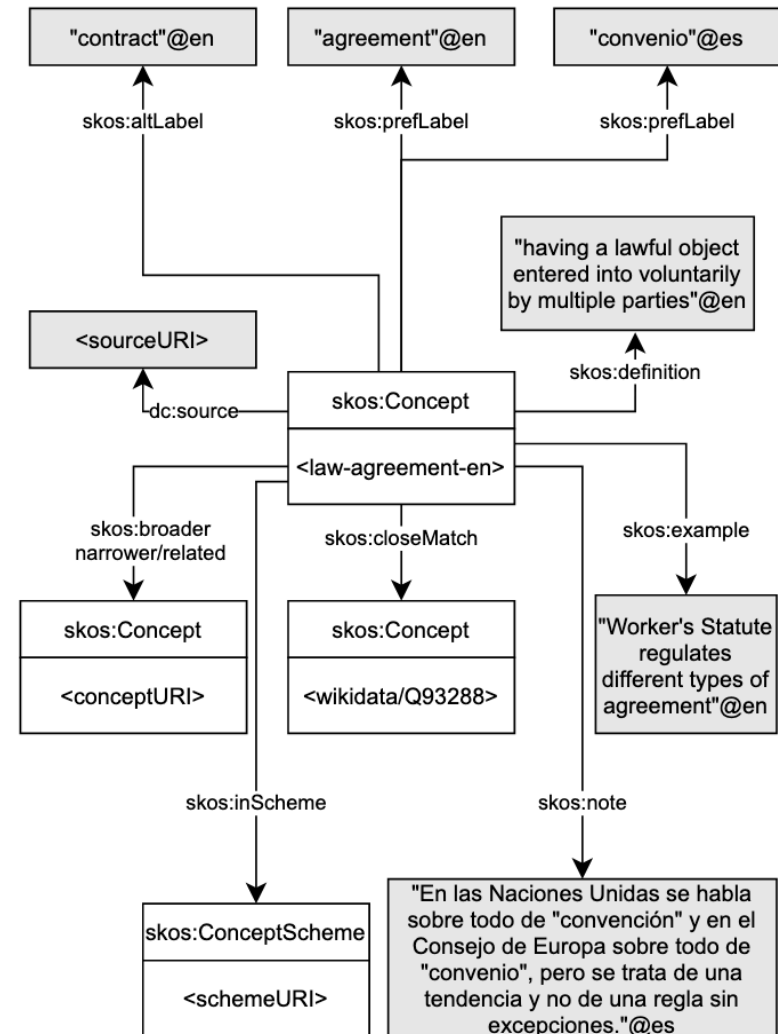
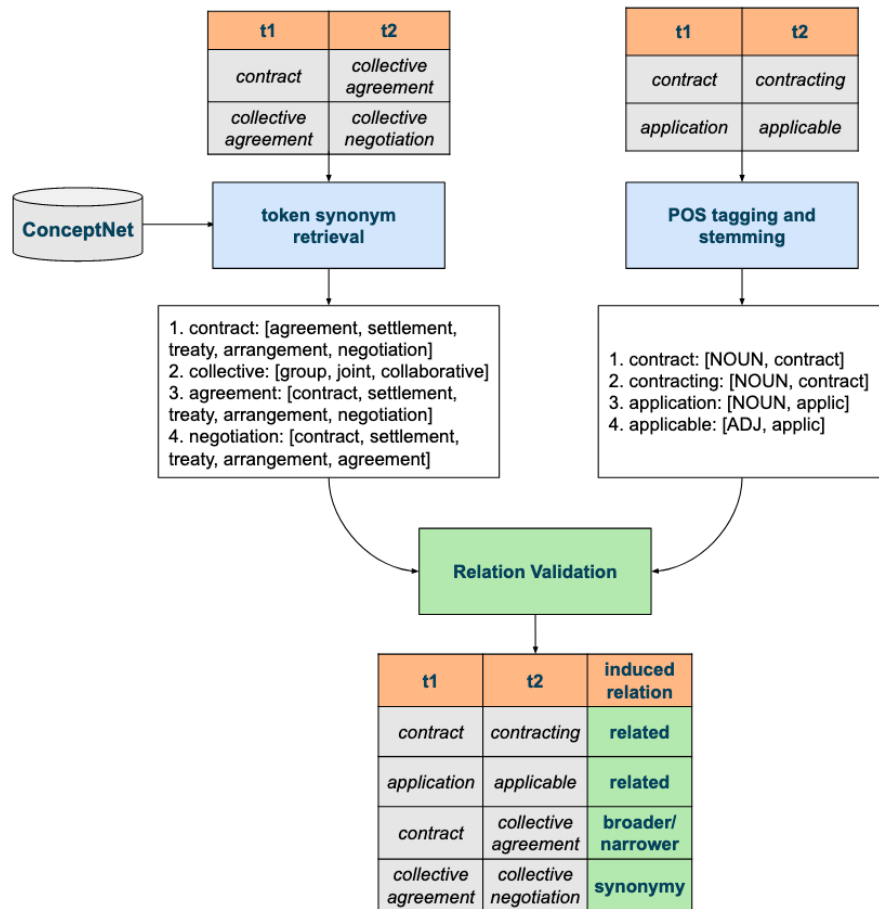
Source: di Buono et al. (2020)

# TermitUp



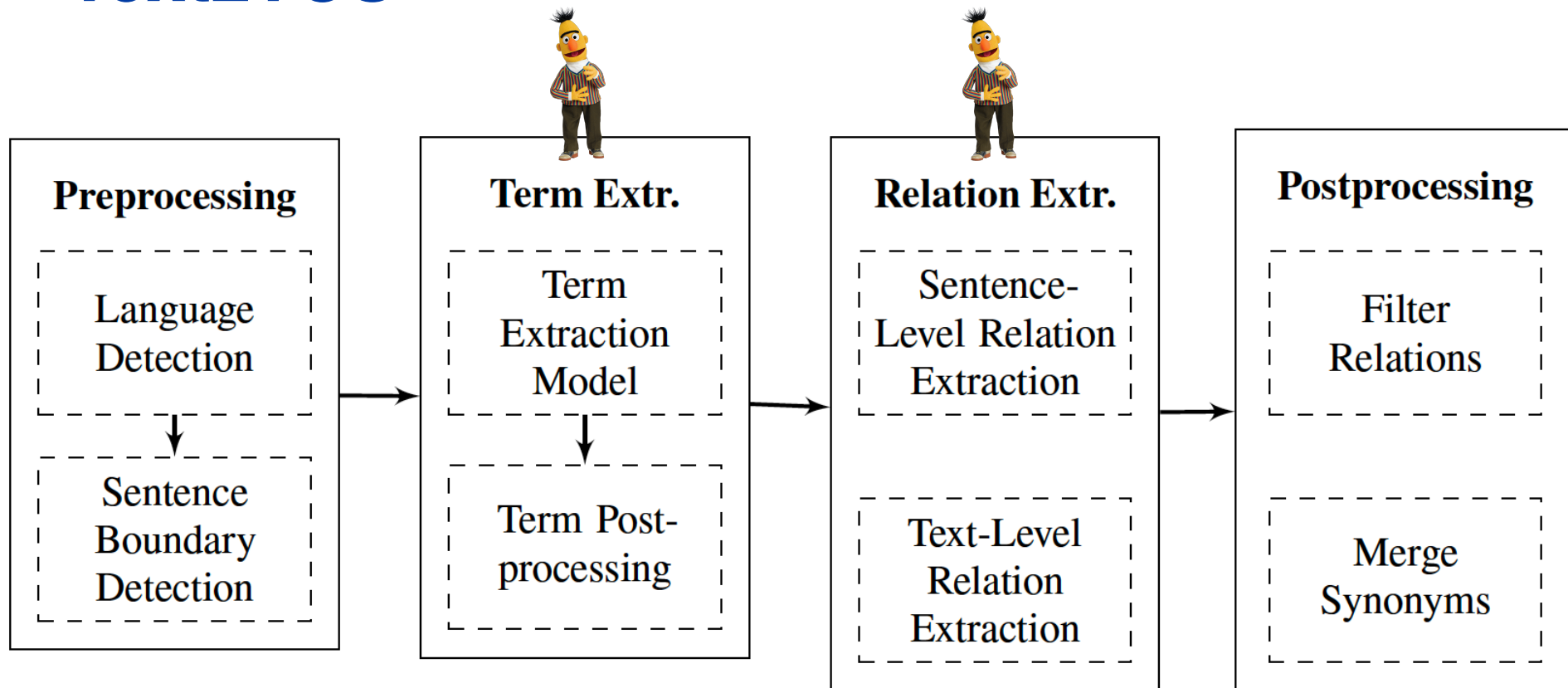
Source: <https://termitup.oeg.fi.upm.es/>

# TermitUp - Example



Source: <https://termitup.oeg.fi.upm.es/>

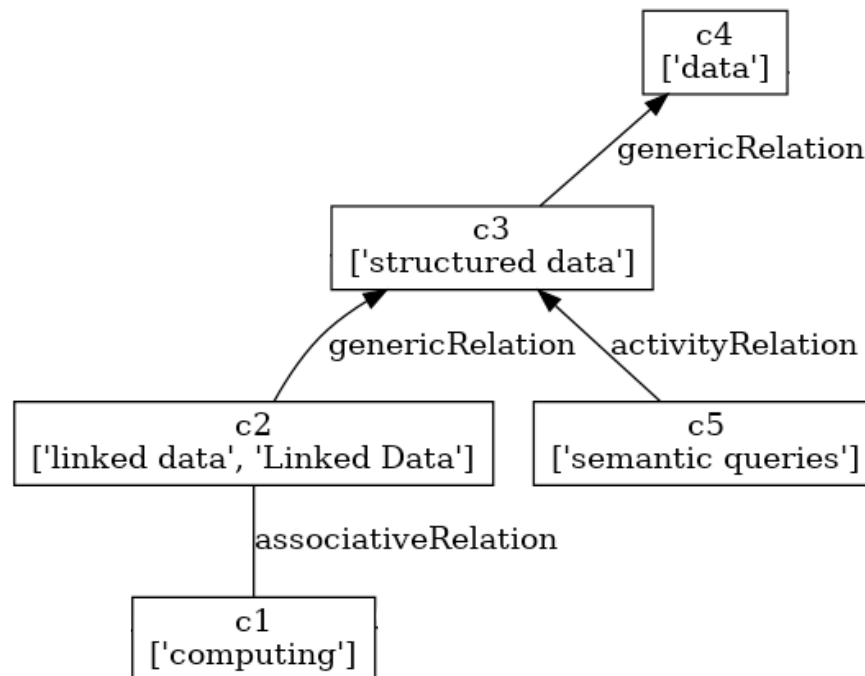
# Text2TCS



Source: <https://text2tcs.univie.ac.at/>

# Text2TCS example

In computing, linked data (often capitalized as Linked Data) is structured data, which is interlinked with other data so it becomes more useful through semantic queries



Source: <https://live.european-language-grid.eu/catalogue/tool-service/8122>

# Next steps and open issues

## Next steps

Consistent and commonly accepted representation of terminologies as LLOD resources + considerably more persistent repositories and tools

## Open issues

Very good use cases for multilingual querying and benefits of LLOD-terminologies

# References

- Abromeit, F., Fäth, C., & Glaser, L. (2020). Annohub—Annotation Metadata for Linked Data Applications. In Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020) (pp. 36-44).
- Cimiano, P., McCrae, J. P., Rodríguez-Doncel, V., Gornostay, T., Gómez-Pérez, A., Siemoneit, B., & Lagzdins, A. (2015). Linked terminologies: applying linked data principles to terminological resources. In Proceedings of the eLex 2015 Conference (pp. 504-517).
- Chiarcos, C., & Sukhareva, M. (2015). OIa—ontologies of linguistic annotation. *Semantic Web*, 6(4), 379-386.
- Di Buono, M. P., Cimiano, P., Elahi, M. F., & Grimm, F. (2020). Terme-a-llod: Simplifying the conversion and hosting of terminological resources as linked data. In Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020) (pp. 28-35).
- Gangemi, A., Alam, M., Asprino, L., Presutti, V., & Recupero, D. R. (2016). Framester: A wide coverage linguistic linked data hub. In European Knowledge Acquisition Workshop (pp. 239-254). Springer, Cham.
- Gracia, J., Villegas, M., Gomez-Perez, A., & Bel, N. (2018). The apertium bilingual dictionaries on the web of data. *Semantic Web*, 9(2), 231-240.
- Martín-Chozas, P., Vázquez-Flores, K., Calleja, P., Montiel-Ponsoda, E., & Rodríguez-Doncel, V. (Forthcoming) [TermitUp: Generation and Enrichment of Linked Terminologies](#). *Semantic Web*, Special Issue on Latest Advancements in Linguistic Linked Data.