Merging Terminology and Knowledge

The power of Eurovoc and IATE through semantic integration

An ENDORSE Workshop Hosted by EU Publications Office
18 April 2024

michael@coreon.com

jochen@coreon.com

www.coreon.com

Presented by

coreon

Jochen Hummel,
CEO

Michael Wetzel,
MD

founded and exited	**TRADOS**		joined industry at	**TRADOS**
founded	twinity		product managed	MultiTerm
chair of	LT-INNOVATE.EU		contributes to	ISO
CEO of	ESTeam		active in	DIN
inventor of	SEMIOX			

About Coreon GmbH

Coreon Multilingual Knowledge System

Fusion of Terminology Management with Knowledge Graph Modelling

to Enable Multilingual AI

- Founded 2012
- Headquarter: Berlin, Germany
- Management owned

*100% focus – nothing else*

Coreon: Multilingual Knowledge System
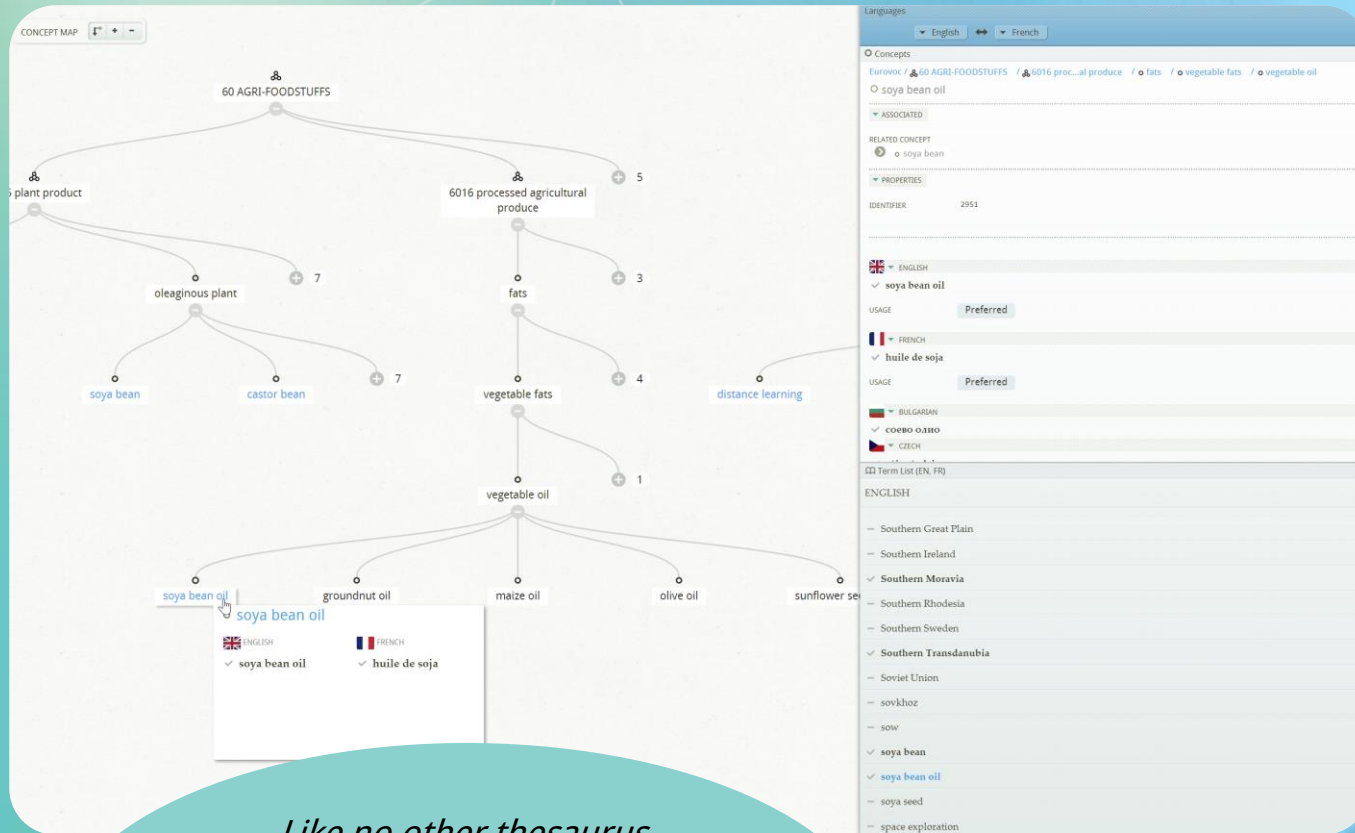
*Like no other thesaurus, taxonomy or terminology tool – Coreon turned out to be the **only comprehensive solution** to implement a multilingual knowledge system.*

Lukas Auer, Liebherr MCCtec, Rostock

- ▸ A fusion of terminology management with knowledge graphs

- ▸ Unlike simple lists or translation-only focused term tools and unlike monolingual, nerdy ontology editors …

- ▸ … most systematic and visual way to control and share large multilingual knowledge graphs

- ▸ Repositories scale from
  - ▸ ~100 to ~200,000+ concepts
  - ▸ monolingual and up to 50+ languages

- ▸ Configurable meta data to the needs of customer domains

- ▸ Operated as a SaaS solution

Agenda: Merging Terminology and Knowledge

Method: Eurovoc + IATE in One

Illustration of the Results

Use Cases

Discussion and Outlook
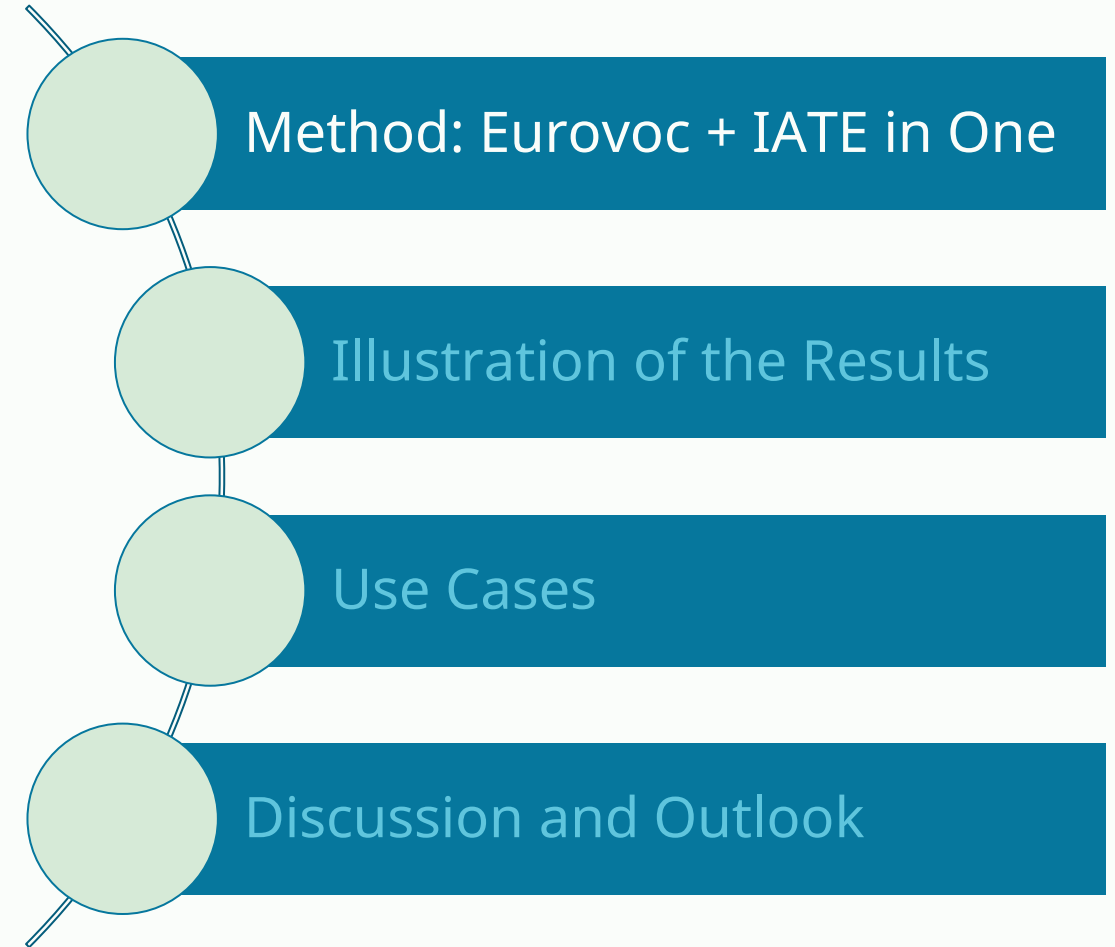
About Eurovoc

- Managed, curated and shared by the Publications Office of the European Union

- Multilingual, multidisciplinary thesaurus covering the activities of the EU

- ~ 7500 concepts

Eurovoc – Samples

About IATE

- Administered by the Translation Centre for the Bodies of the European Union

- The terminology database of the European Union

- 670,000+ concepts, covering 7 million terms

IATE – a Sample

https://iate.europa.eu/entry/result/3542165/en

Eurovoc + IATE: Complementary Resources

- Used for indexing, a meta data "vocabulary" for EU publications and texts
- Hierarchical "top level" taxonomy

= a *knowledge graph*

- For translation and any variant of multilingual text work
- A large, very detailed language resource that (probably?) covers more or less all things ever dealt with by EU

= a *terminology database*

What do they contain?
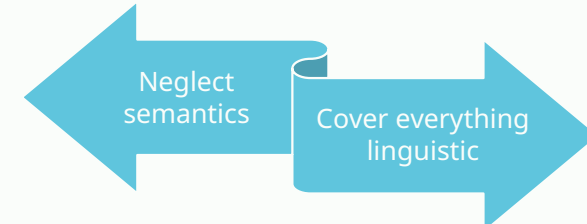
Knowledge Graphs (such as Eurovoc)

- 1) Store concepts and their *labels*

- 2) Relate concepts to each other in a hierarchical or associative manner
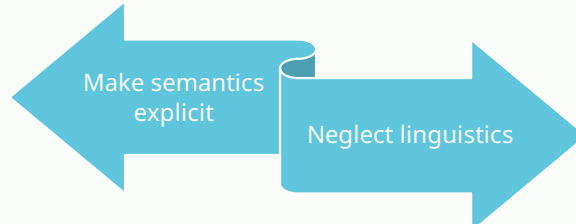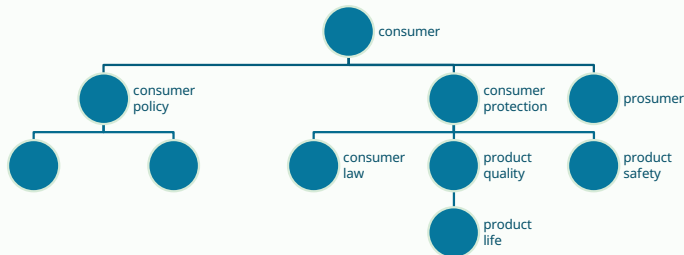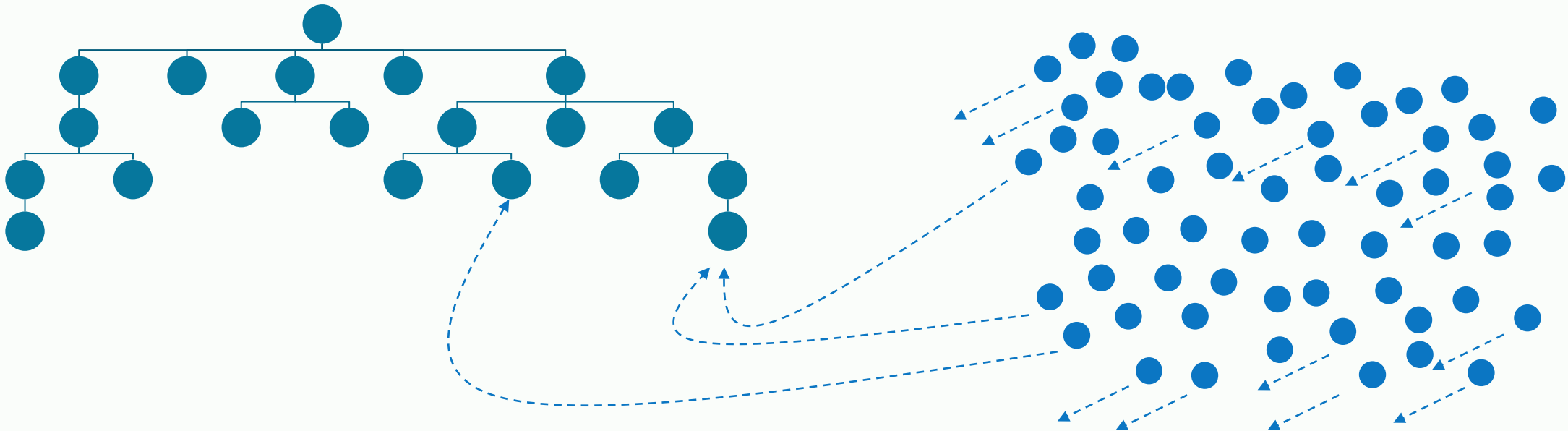

Terminology Resources (such as IATE)

- 1) Store concepts and their *terms*

- 2) Store a lot of descriptive and prescriptive information about the concepts and the terms: Definitions, Notes, Usage Guidance ...
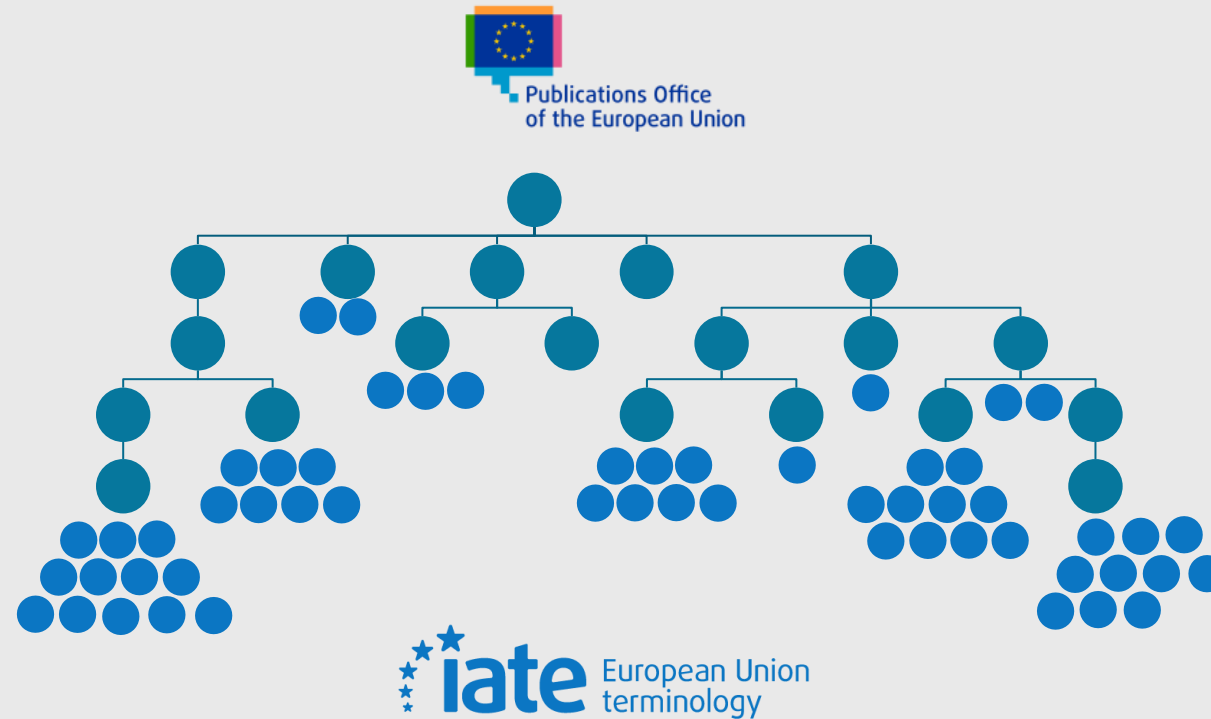
Idea: Hook all IATE Concepts within the Eurovoc Knowledge Graph

Idea: Hook all IATE Concepts within the Eurovoc Knowledge Graph

One unified, large, linguistically rich knowledge graph

Is there already a link between IATE and Eurovoc?
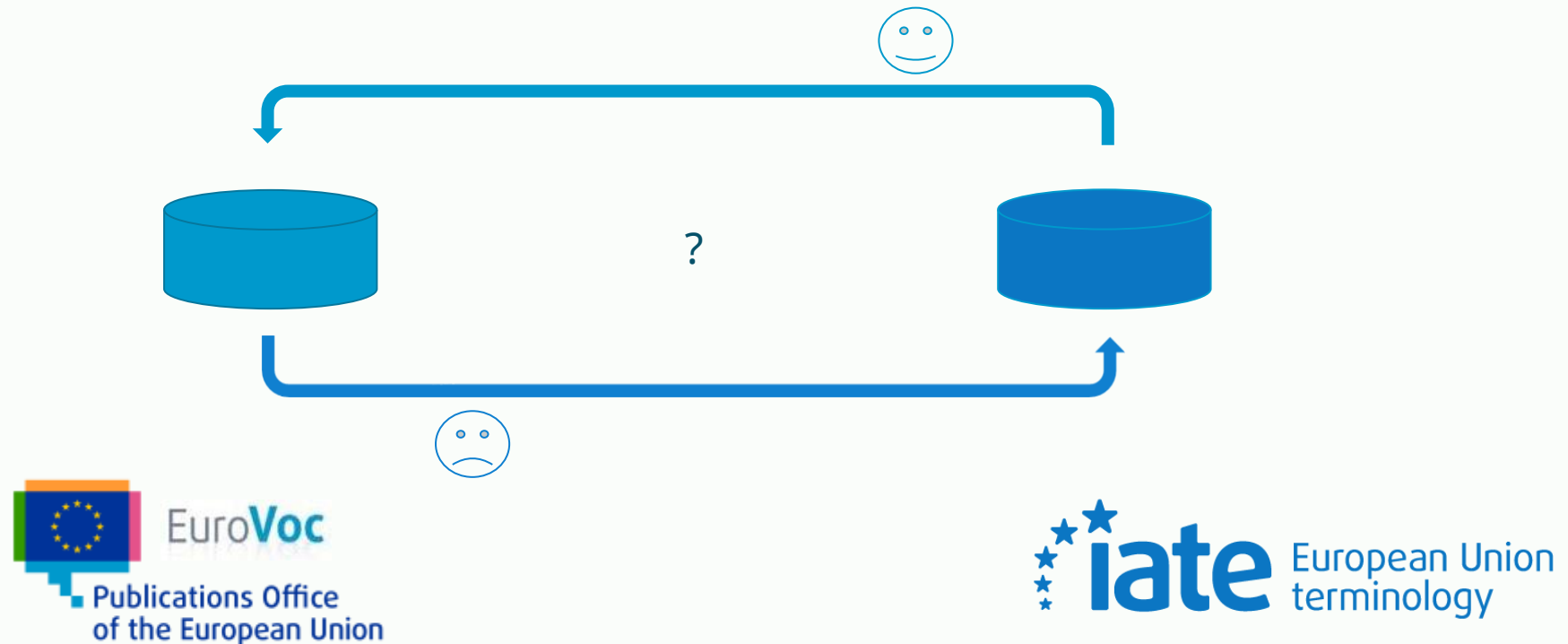
How? Use the "Domain" code as anchor

- IATE entries "mention" Eurovoc node labels as domain codes

How? Use the "Domain" / "subjectField" as the technical link

- Eurovoc W3C SKOS RDF

- IATE ISO TBX


```
Eurovoc #4010 has-narrower IATE #3542165
```

Agenda: Merging Terminology and Knowledge

Method: Eurovoc + IATE in One

Illustration of the Results

Use Cases

Discussion and Outlook
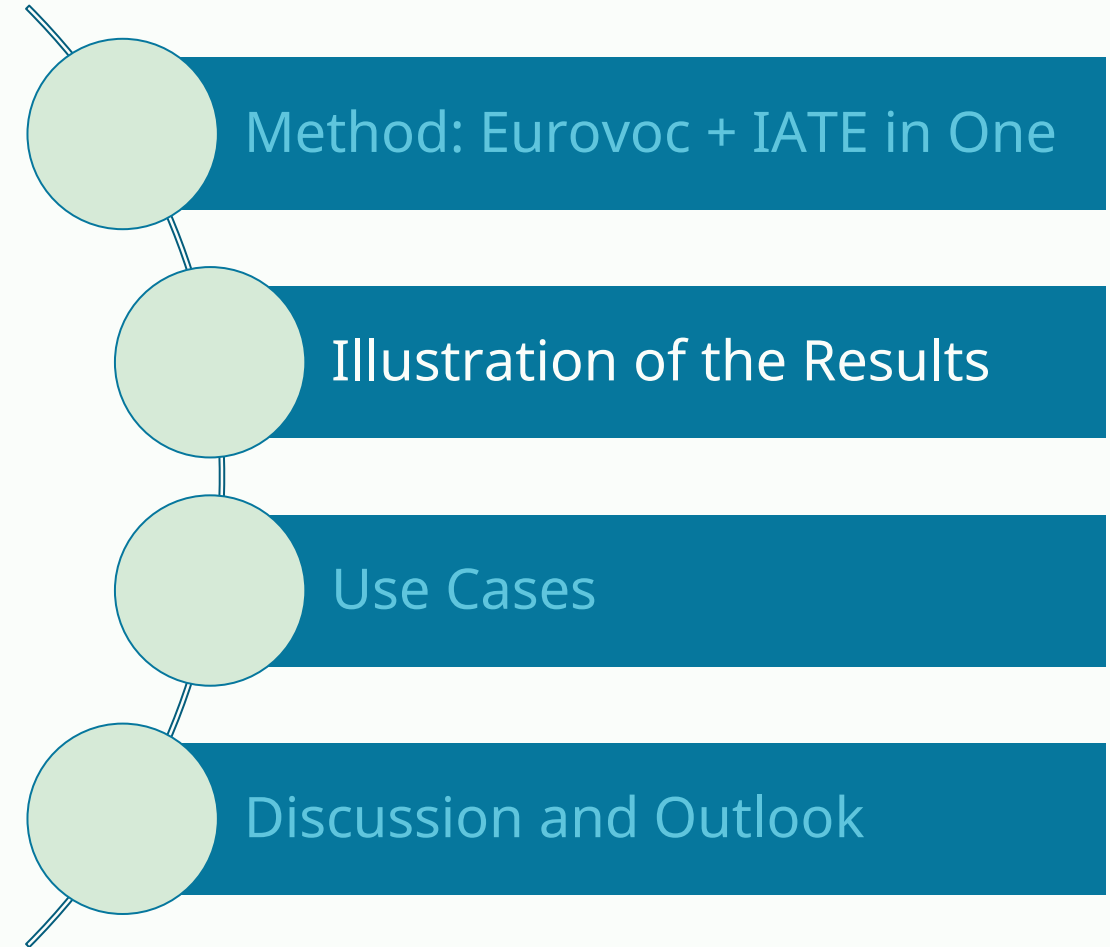
Eurovoc node
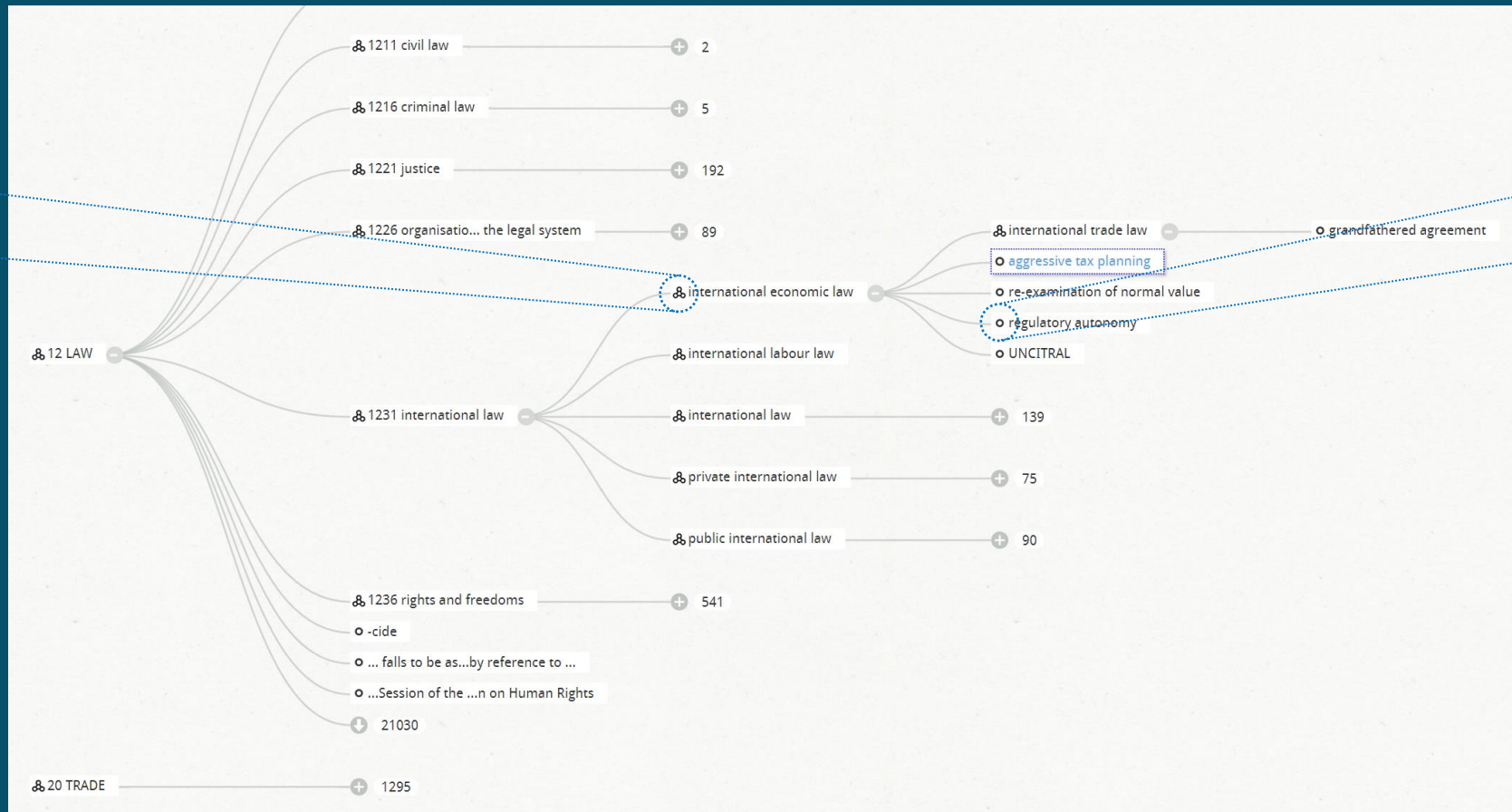
IATE concept

Eurovoc + IATE in One

Knowledge Graph
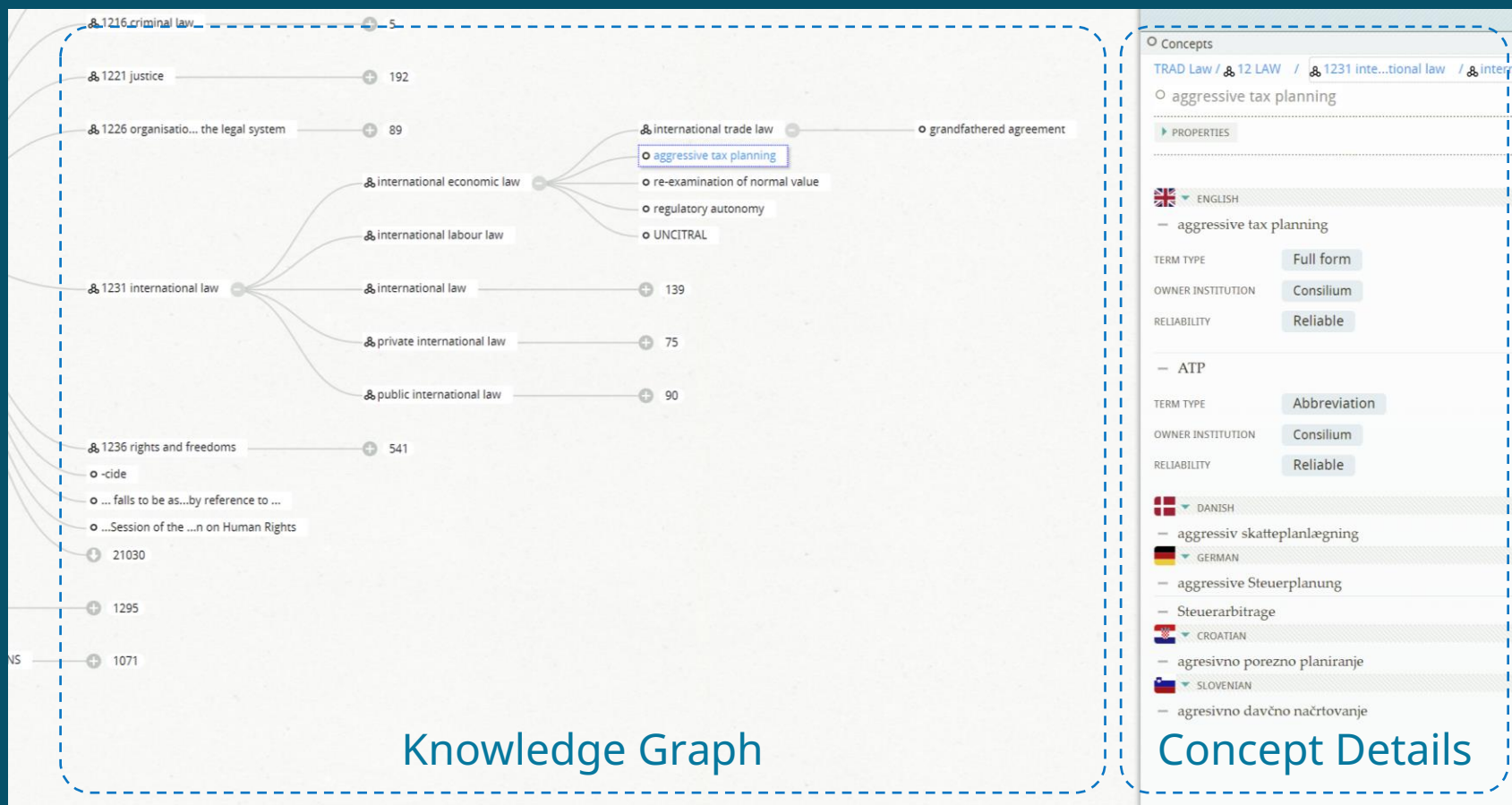
Concept Details

Reading the Details of a IATE Concept while Browsing the Knowledge Graph

Related IATE Concepts now "close" to each other
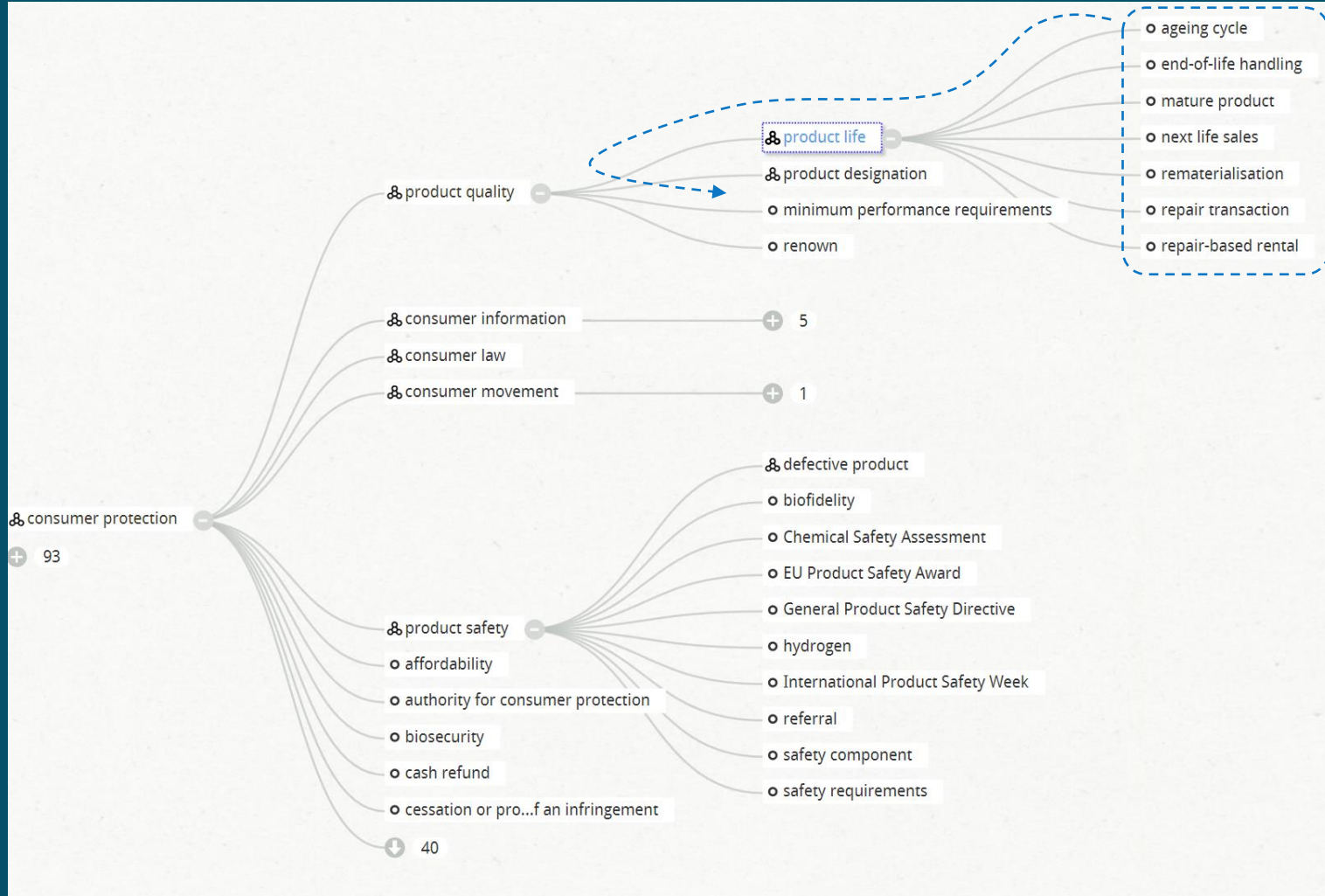
- Visually and technically close, such as ...
- "... all IATE concepts that deal with 'product life' "

- ... and more general with 'product quality'

Related IATE Concepts now "close" to each other
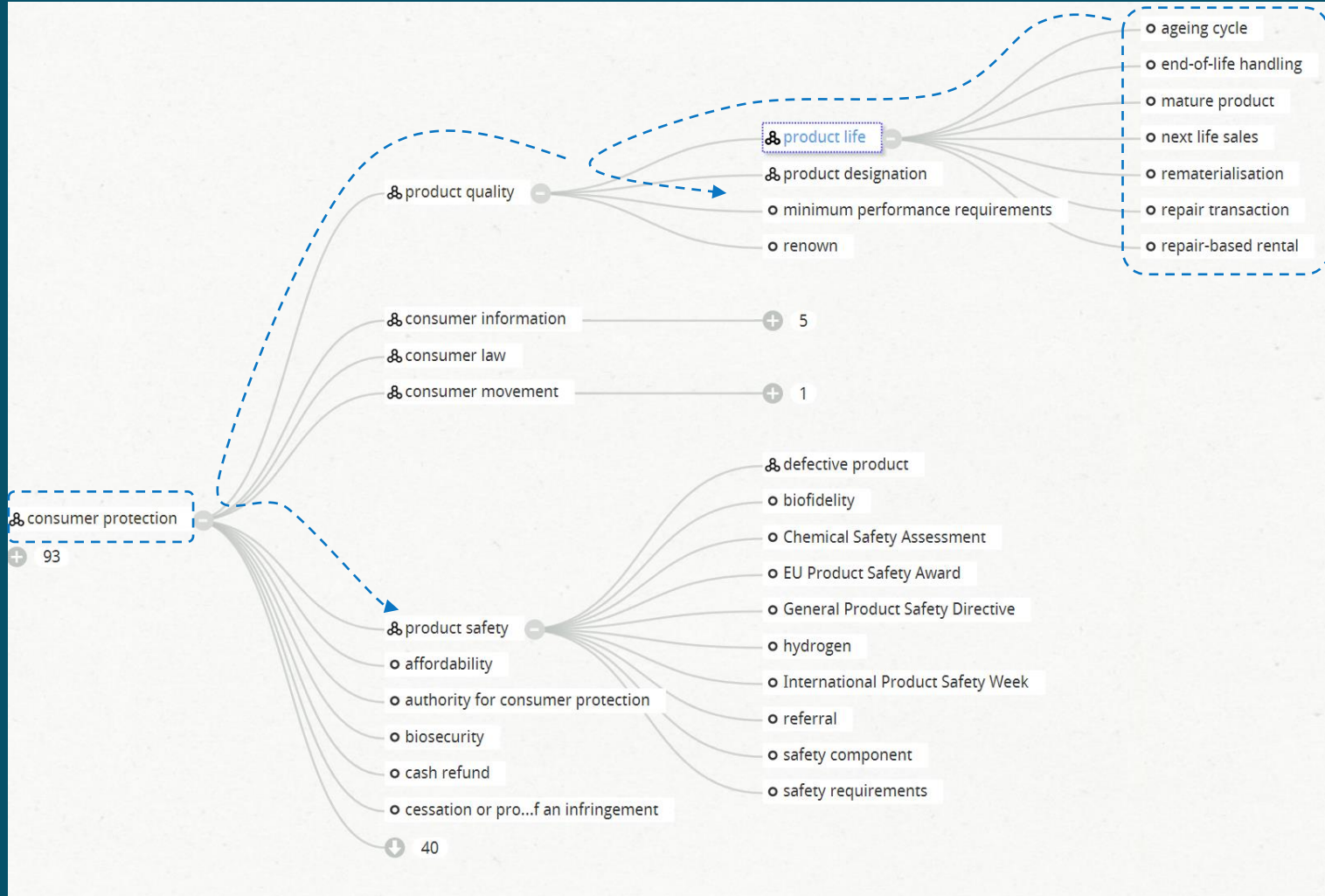
- ... and even with 'product safety'

(since all in the branch of 'consumer protection')
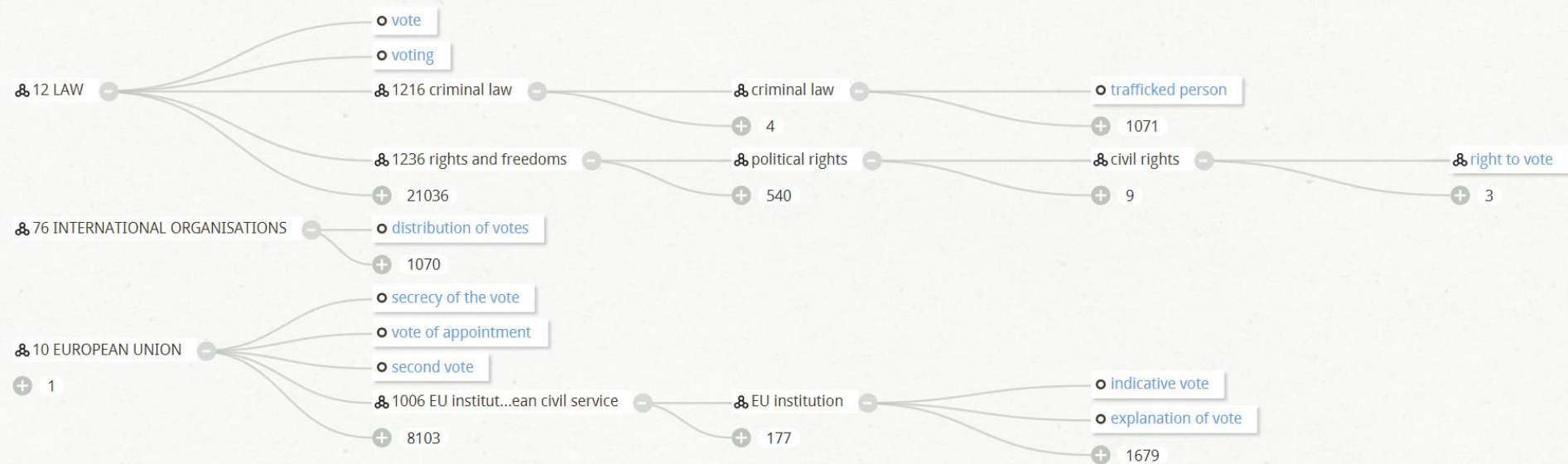
Search for "vote":
understand in which branch the hits from IATE reside

Filter, Search, Export – Re-use by Branches
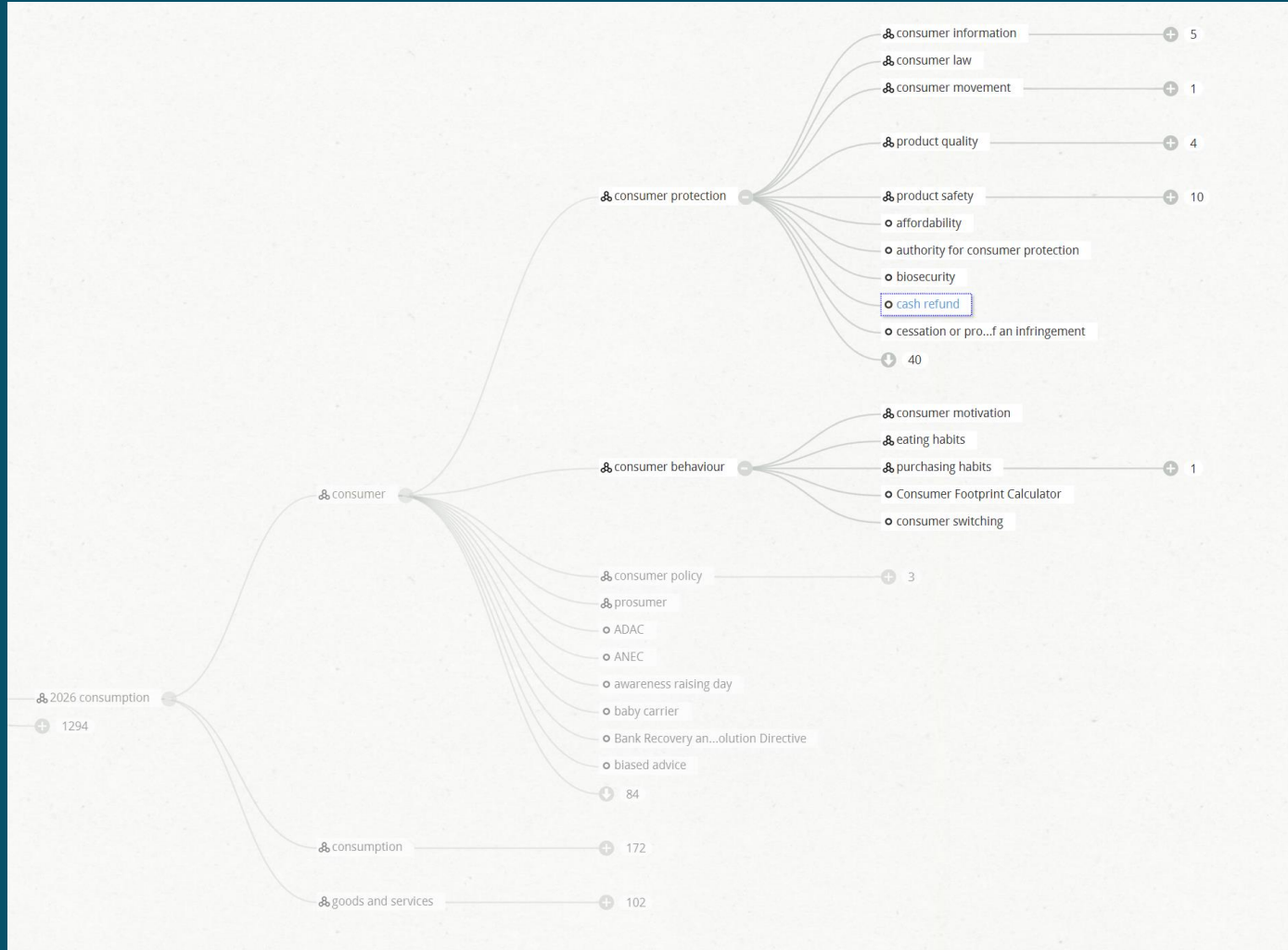
- "Get me all concepts under 'consumer'"
- "Export all En, Fr, It terms that are in the branch of 'consumer behaviour' "
- "Get me translations for terms in branches <x> and <y> but not <z> ..."

Last but not least: In any EU Language
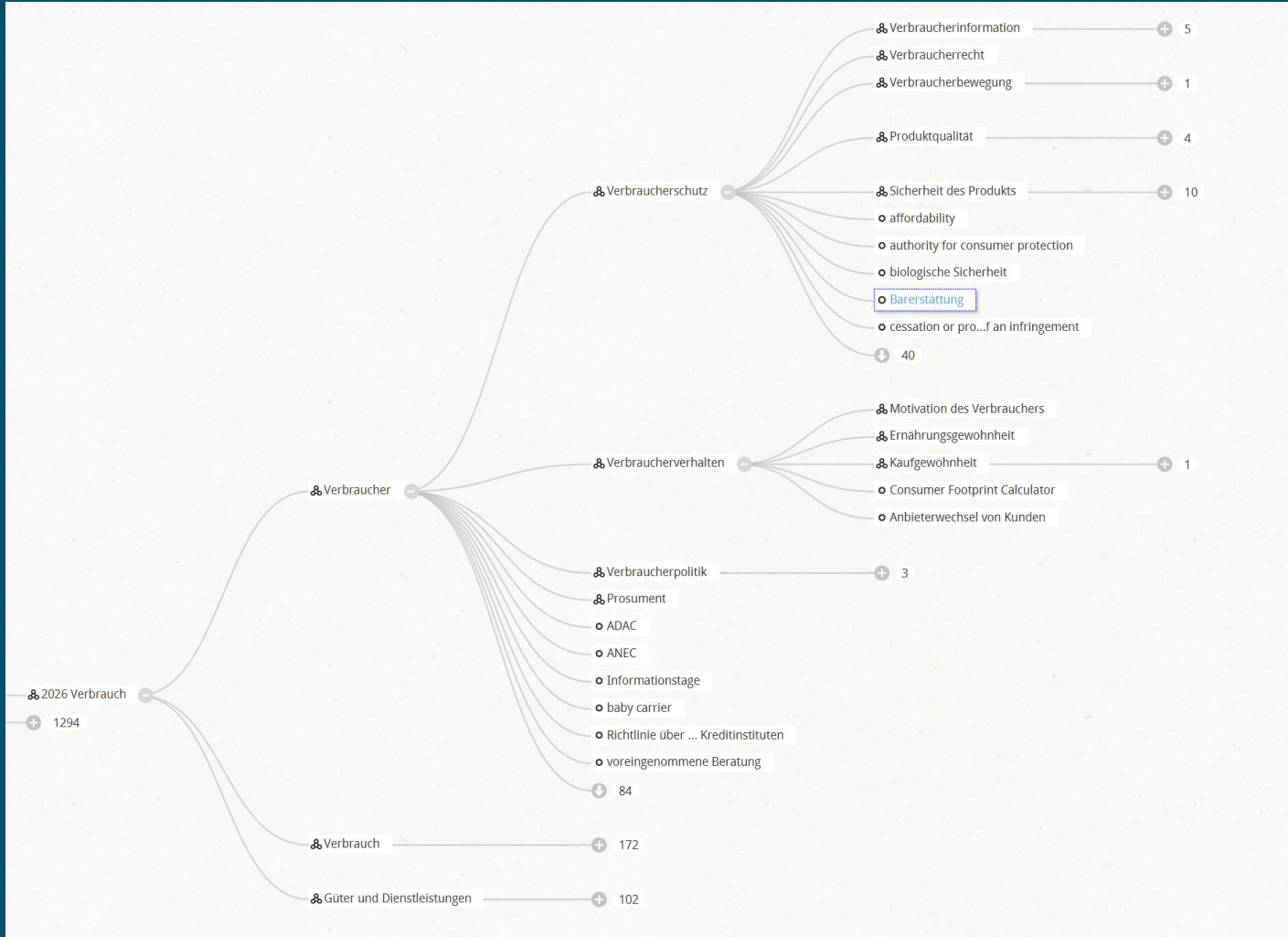
- The knowledge graph – as is Eurovoc – is based on relations between concepts, thus it is "language agnostic"
- Once deployed it can be used for any EU language

Eurovoc + IATE: Observations

Many Eurovoc nodes are not used at all by IATE

Whereas many IATE concepts are just tagged with often generic Eurovoc nodes
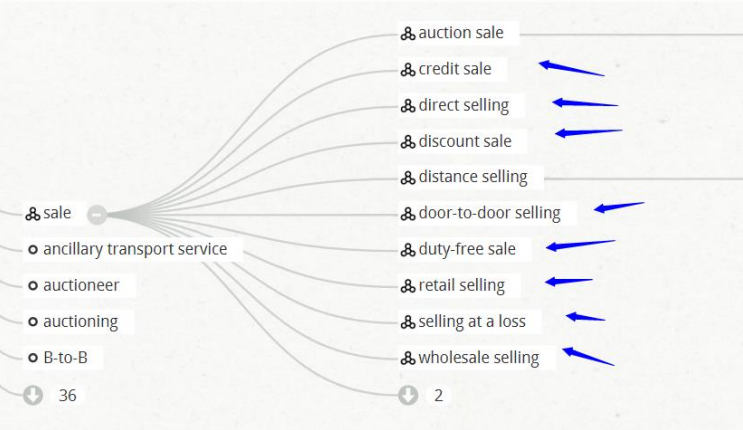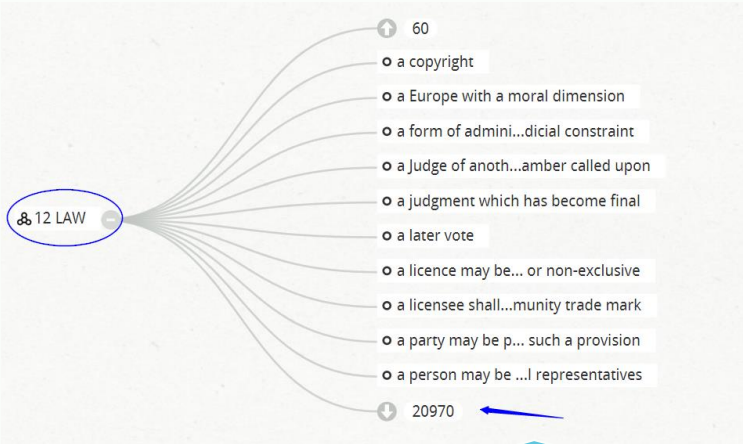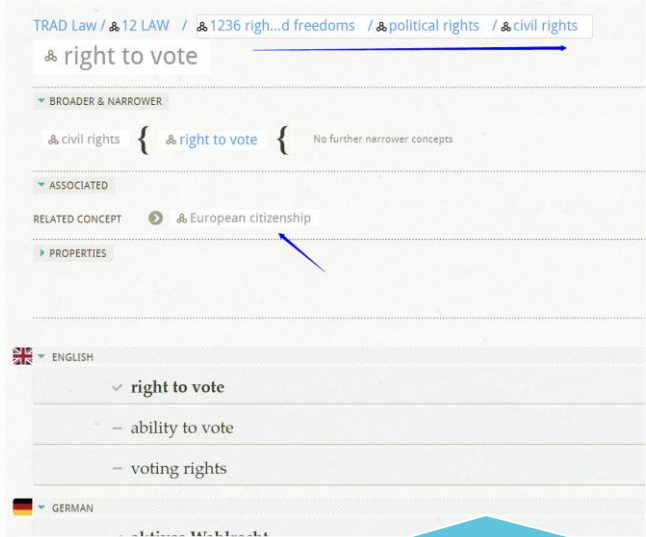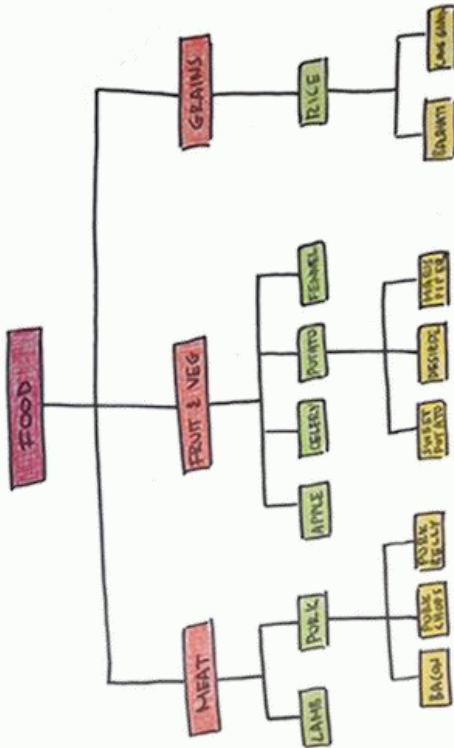
IATE concepts hooked in Eurovoc may immediately benefit from Eurovoc's associative relations

Why a Knowledge Graph for Terminologists and Linguists

- ⚥ Effective way to add structure to terminology collections

- ⚥ Improve data quality
  - ⚥ avoid duplicates and overlapping concepts
  - ⚥ associative relations

- ⚥ Easier data maintenance: similar "things" are close to each other

- ⚥ Formalize multilingual knowledge, make it machine-digestible

- ⚥ Boost performance of AI algorithms, priming them with structured data

Agenda: Merging Terminology and Knowledge

Method: Eurovoc + IATE in One

Illustration of the Results

Use Cases

Discussion and Outlook

Agenda: Merging Terminology and Knowledge

Method: Eurovoc + IATE in One

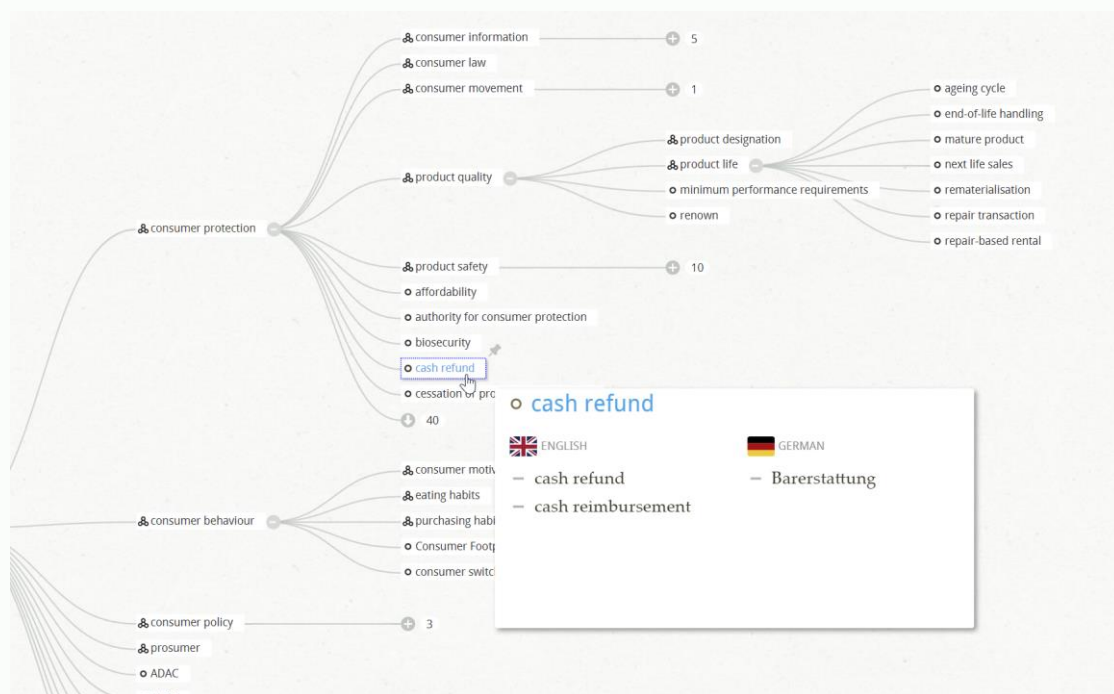Illustration of the Results

Use Cases

Discussion and Outlook
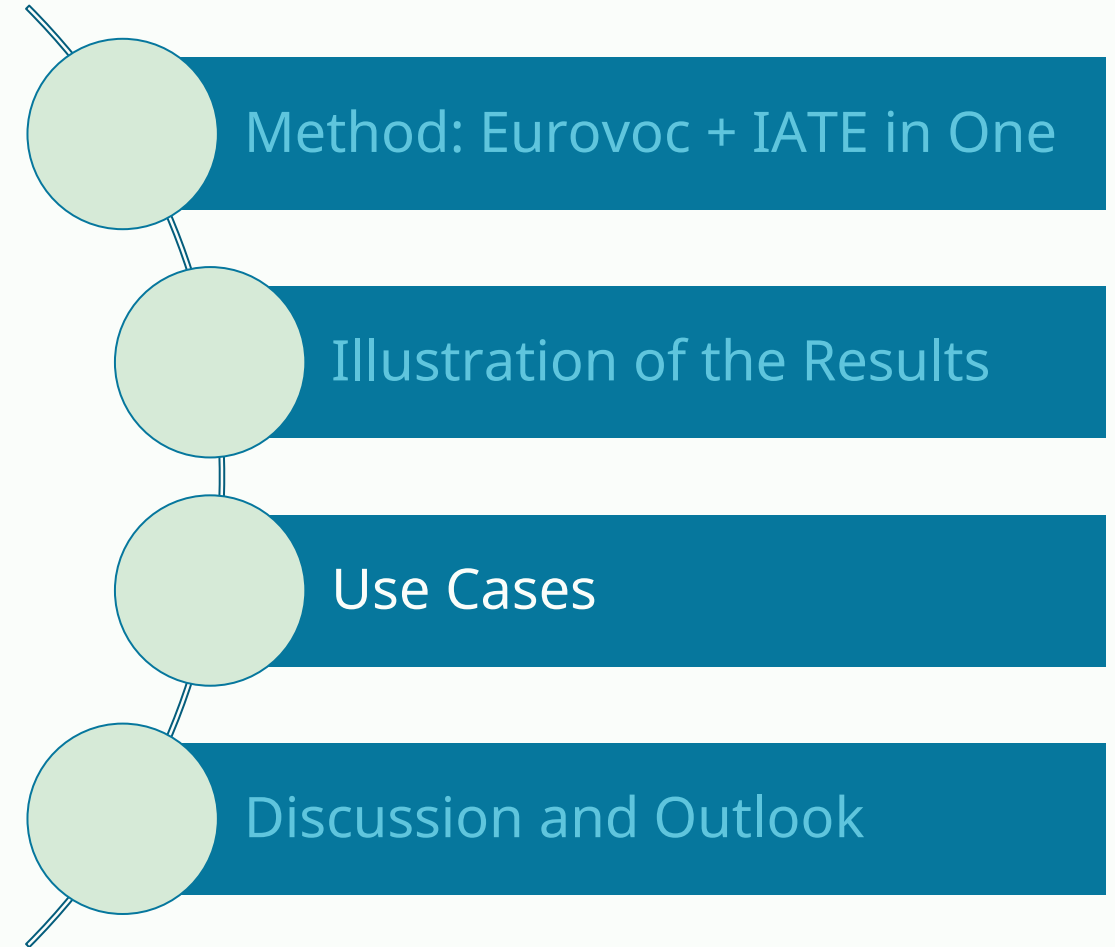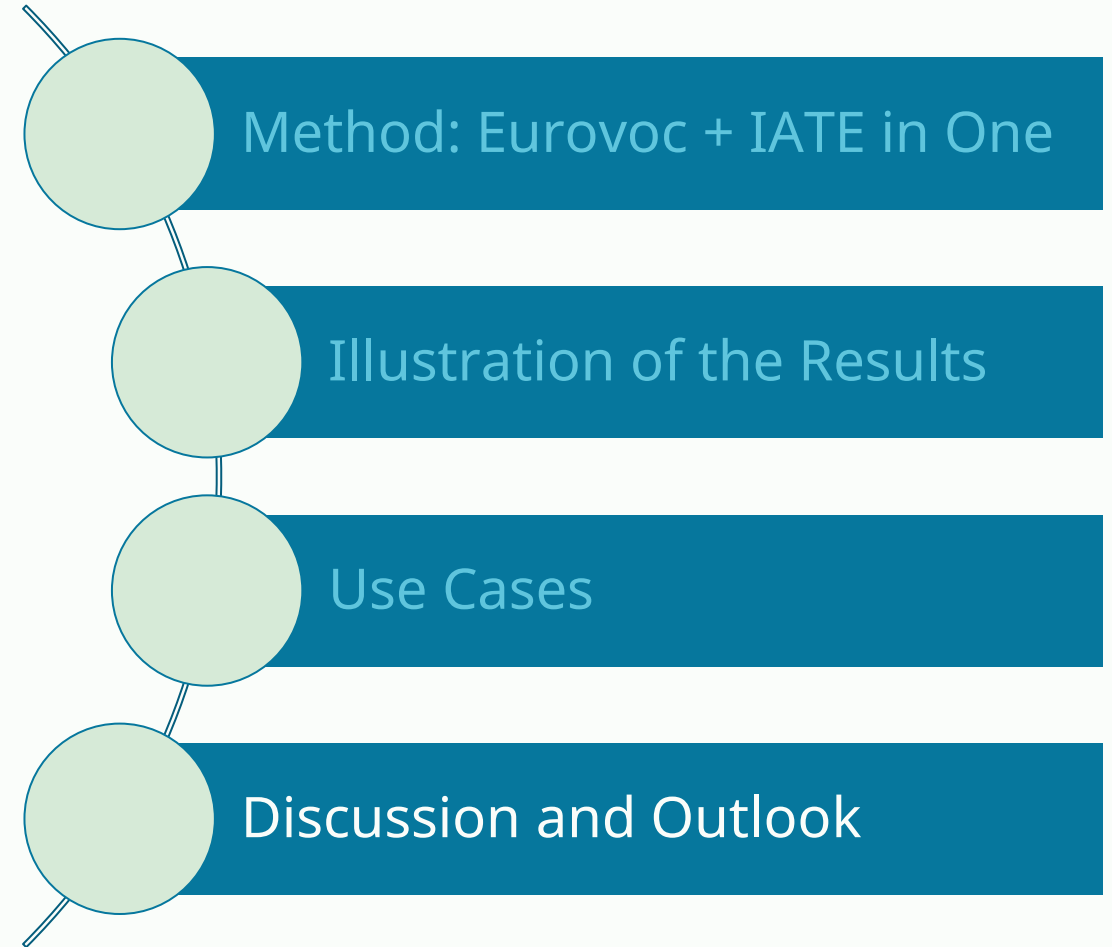
Key Messages: Merging Terminology with Knowledge

- Benefits for data maintainers and consumers
- Eurovoc adds semantics to IATE, IATE is like a "linguistic amplifier" for Eurovoc
- Linguistically rich knowledge graphs enable M-RAG, multilingual indexing, interoperability, cross-language search ...

Thanks a lot!