Ontology of VAT Concepts

Grant Agreement No.:	101087342
Project Acronym:	POLINE
Project Title:	Principles of Law in National and European VAT
Website:	https://site.unibo.it/poline/en
Contributing WP:	WP2
Deliverable ID:	D2.1
Contractual delivery date:	31/10/2024
Actual delivery date:	05/11/2024
Dissemination level:	PU
Deliverable leader:	UNITO

This project is funded by the European Union's Justice Programme (2022)

1. Document History

Version	Date	Author	Partner	Description
1.0	15/10/2024	Rachele Mignone, Ivan Spada, Davide Audrito, Luigi Di Caro	UNITO	First draft
2.0	23/10/2024	Piera Santin, Alessia Fidelangeli	EUI, UNIBO	Contribution of Sections 6 and 7
3.0	28/10/2024	Rachele Mignone, Ivan Spada, Davide Audrito, Luigi Di Caro	UNITO	Second draft
4.0	29/10/2024	Rachele Mignone, Ivan Spada, Davide Audrito, Luigi Di Caro, Piera Santin, Alessia Fidelangeli	UNITO, UNIBO, EUI	General survey
5.0	30/10/2024	Rachele Mignone, Ivan Spada, Davide Audrito, Luigi Di Caro, Piera Santin, Alessia Fidelangeli	UNITO, UNIBO, EUI	Final survey

2. Contributors

Partner	Name	Role	Contribution
UNITO	Di Caro Luigi	Professor	Methodology, writing, reviewing
UNITO	Mignone Rachele	Research Fellow	Methodology, development, writing, reviewing
UNITO	Spada Ivan	Research Fellow	Methodology, development, writing, reviewing
UNITO	Audrito Davide	Ph.D. Fellow	Methodology, writing, reviewing
UNITO	Siragusa Giovanni	Researcher	Methodology, writing, reviewing
EUI	Santin Piera	Research assistant	Methodology, development, writing reviewing
UNIBO	Fidelangeli Alessia	Research fellow	Introduction, methodology, development, writing reviewing
UNIBO	Giuseppe Contissa	Associate Professor	Reviewing

3. Table of Contents

4. List of Acronyms

CJEU Court of Justice of the European Union

ECLI European Case Law Identifier

EU European Union

JPOL Judicial Principles Of Law

VAT Value Added Tax

5. Executive Summary

This deliverable details the development of WP2 of the POLINE project, focusing on the creation of a multilingual ontology for legal concepts related to Value Added Tax (VAT). Recognizing the limitations of existing resources like the EUR-Lex taxonomy, which relies primarily on legislative terminology, this project sought to create a more comprehensive tool that reflects the evolving interpretations of VAT law as established by the Court of Justice of the European Union (CJEU). The resulting ontology integrates the existing EUR-Lex structure with carefully selected keywords extracted from a corpus of CJEU judgments [3], providing a richer and more precise representation of key VAT concepts.

The construction of this ontology involved a rigorous, multi-stage process. First, a dataset of CJEU judgments focusing on the subdomains of "Exemptions" and "Taxable Amount" was compiled (for the choice of the sub-domains see para. 6). Keywords from these judgments were then extracted and meticulously evaluated by legal experts to ensure relevance and avoid redundancy. Selected keywords were then integrated into the EUR-Lex hierarchical structure, creating a six-level ontology that balances detail with practicality. This entire process was guided by a set of carefully defined rules to maintain consistency and coherence. Furthermore, recognizing the importance of accessibility across the EU, the ontology has been translated into multiple languages, with each translation validated by legal experts to ensure accuracy and consistency with official legal terminology.

This multilingual VAT ontology serves as the foundation for the Judicial Principles of Law (JPOL) extraction and retrieval tool. By leveraging the ontology's structured knowledge base, the tool will be able to identify and classify JPOLs with greater accuracy and efficiency, facilitating more targeted access to relevant legal information. This will ultimately empower legal professionals, tax authorities, businesses, and other stakeholders to navigate the complexities of VAT law with greater confidence and understanding.

6. Introduction

The main objective of the project is the development of a tool for the extraction of JPOLs and for their retrieval based on a common ontology of VAT concepts.

This tool addresses an inconsistency in accessing EU case law. While the European Union's Publications Office uses the EUR-Lex classification system for both EU law and case law accessible through the EUR-Lex platform, the Court of Justice of the European Union (CJEU) employs a different database and classification methodology (see para. 7).

Moreover, EUR-Lex is based on terms taken from VAT legislation. However, since the introduction of VAT in 1967, a significant number of decisions have been adopted. Over these sixty years, the case law of the CJEU has focused on specific topics that are not directly identifiable through the EUR-Lex taxonomy's classes [1]. This is particularly evident in two areas of VAT: exemptions for activities in the public interest and rules concerning the taxable amount. These areas are regulated by only a few provisions in the Directive, yet they address either crucial aspects of public life or the structure of the tax. This is one of the reasons why we decided to focus on these sub-topics.

In addition, this decision serves a twofold purpose. First, these two domains contain numerous JPOLs, and access to them is crucial for the target groups identified in the “Impact and Ambition” section of the Proposal. Second, the EUR-Lex classification groups the decisions, and the JPOLs within these decisions, under a few macro-concepts. As a result, users cannot easily find judgements that address specific problems and sub-topics, they should also search for the JPOLs in it without a specific retrieval system, making research time-consuming and less effective.

To address these challenges, we decided to develop an ontology for classifying JPOLs. Creating this ontology was one of the first steps of the project, as it forms the foundation for both the retrieval system used on the platform and the classification of JPOLs.

The development of the ontology was driven by two guiding ideas. The first was to create an ontology based on both legislative concepts and concepts introduced by the CJEU to describe the content of its case law. Only such an ontology can effectively classify the JPOLs extracted from this case law. To this end, we built an ontology that incorporates both EUR-Lex concepts and the keywords found in each judgment of the CJEU. The second guiding principle was to rely on official concepts, i.e., concepts used in official European documents such as European legislation or case law.

This approach also allows for the creation of a multilingual ontology, where each version is an official one. The existence of such an integrated multilingual ontology aligns with Recital 5 of the recently published Interoperable Europe Act (Regulation 903/2024). According to this recital, “as the

language barrier is one of the obstacles to interoperability, to the reuse of solutions, and to the establishment of cross-border services, semantic interoperability is key to facilitating effective communication in diverse multilingual environments, including at regional and local levels". Hence, the existence of interoperable public tools is compliant with the objective of semantic interoperability, as well as legal interoperability.

This integrated multilingual ontology could be highly beneficial in the context of VAT, as the first VAT Directives were adopted before many countries joined the European Union. Many countries that implemented the Directive during the first years of the harmonisation process (during the Seventies) relied on national concepts and used national terminology, making semantic interoperability more difficult for both tax administrations and taxpayers, especially those operating in cross-border contexts [4]. Countries that joined the EU more recently, such as Bulgaria (2007), have often adhered more strictly to European terminology. The existence of a common ontology could help developing a shared terminology in the field of VAT. In addition, the integration of the ontology with JPOLs would improve the harmonisation of VAT.

Thus, this integrated multilingual ontology offers several advantages: it is likely to be accepted by European institutions as it relies on official documents; it organises (rather than creates) existing concepts; it promotes tax harmonization by serving as a common framework for both the EU and its member states; and it is compatible with national legal systems as it uses European terms that should be recognized within the legal systems that have implemented the Directive.

7. VAT Ontology

This section outlines the theoretical and computational foundations underlying the development of the VAT ontology and its linguistic versions (English, Swedish, Bulgarian, and Italian). It details the methodologies employed in constructing the ontology's core structure, derived from both legal texts and expert knowledge, and the subsequent process of translating this structure into multiple languages while maintaining semantic and legal precision.

7.1. Theoretical Basis

The case law of the CJEU can be accessed through two main official channels.
The first is the CJEU's website: curia.europa.eu. The retrieval system on the Curia website allows users to search by "subject matter". Under the available topics, there is one for "taxation," which includes the sub-topic "value added tax". However, there are no more specific sub-topics beyond "value added tax". This makes it impossible to search based solely on the subject matter, requiring users to add references to legislation, case law, or specific text. Since few years, it has also been possible to search via the so-called "systematic classification scheme" on the Curia website, which is the same system used by EUR-Lex.

The second channel is EUR-Lex, through the Directory of European Union case-law. EUR-Lex is the official online access to EU law and case law. It provides official and comprehensive access to EU legal documents, available in all 24 official EU languages, and is updated daily. EUR-Lex is run by the Publications Office of the European Union. Through EUR-Lex, users can access a wide range of EU legal documents, including treaties, legal acts from EU institutions, preparatory documents related to EU legislation, EU case-law, international agreements, EFTA documents, references to national transposition measures, and some national case-law related to EU law. EUR-Lex features a systematic classification of case law, based on the structure of the Treaty of Lisbon. This classification scheme was originally used by the Digest of case-law, a systematic collection of judgment and order summaries published in the European Court Reports and the European Court Reports – Staff Cases (ECR-SC). Each paragraph of a summary is assigned one or more classification codes corresponding to the relevant legal field.

For our purposes the terms falling under Class 4 - "Internal policy of the European Union" are the most relevant, especially those addressing "Tax provisions" (Class 4.10).

Example:

4 Internal policy of the European Union
[...]
4.10 EEC/EC - Tax provisions * Tax provisions

4.10.00 General

4.10.01 Internal taxation

4.10.02 Harmonization of tax legislation

4.10.02.01 Value-added tax

4.10.02.01.00 General

4.10.02.01.01 Concept of value added tax and other similar taxes

4.10.02.01.02 Scope

4.10.02.01.03 Place of taxable transactions

4.10.02.01.04 Chargeable event, chargeability and collection of VAT

4.10.02.01.05 Taxable amount

4.10.02.01.06 Rates

4.10.02.01.07 Exemptions

4.10.02.01.08 Deductions and refunds

4.10.02.01.09 Obligations of taxable persons and certain non-taxable persons

4.10.02.01.10 Special schemes

4.10.02.01.11 National derogations

4.10.02.01.12 Transitional arrangements for the taxation of trade between Member States in the Sixth Directive

4.10.02.01.13 Administrative cooperation and combating tax fraud

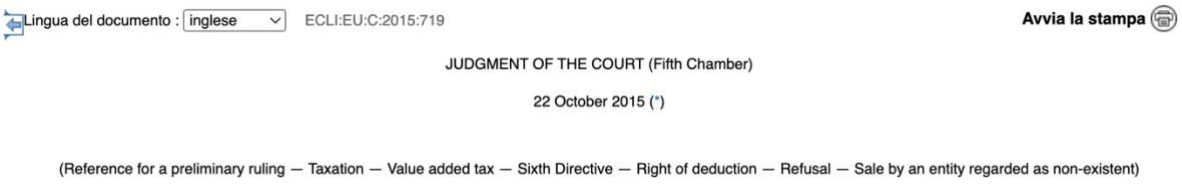
In addition to the Curia and EUR-Lex systems, many other taxonomies of case law exist, usually developed by private companies or national public bodies, such as courts. The former are typically accessible only to registered users and do not use an ontology or taxonomy that is openly available to the public. The latter often restrict access to members of the public administration concerned. For example, the Italian Court of Cassation allows the full access to its database of case law (CED), which includes a thesaurus for the classification of the case law, only to judges. Therefore, the taxonomies or ontologies used by these databases (if any) are not available as a starting point for

our research. Furthermore, these systems are often available only in the national language of the developing company or public body.

In the light of these observations, we considered the EUR-Lex classification scheme as the most appropriate starting point because it is:

- Public.
- Multilingual.
- Officially recognised and used by the EU institutions.
- Scalable.
- Complete (it covers any VAT topic).

Despite these positive features, the EUR-Lex taxonomy is not very detailed and does not include concepts developed by case law. For this reason, we chose to enrich the EUR-Lex taxonomy with some of the keywords used in CJEU decisions. The reason for using these keywords is that we wanted the additional concepts to be as official and multilingual as EUR-Lex ones. Each CJEU judgment includes a chain of keywords that describe the content of the decision, located just above the case number and below the date. These keywords are also visible in the short case description on EUR-Lex .


Lingua del documento : inglese ECLI:EU:C:2015:719 Avvia la stampa

JUDGMENT OF THE COURT (Fifth Chamber)

22 October 2015 (*)

(Reference for a preliminary ruling — Taxation — Value added tax — Sixth Directive — Right of deduction — Refusal — Sale by an entity regarded as non-existent)

The integration of the keywords only concerned the two subdomains chosen for the extraction of JPOLs (see para. 8). Since decisions concerning the taxable base or the exemptions may contain also JPOLs concerning other aspects of VAT (e.g., taxable person, chargeability, etc.), we maintained the entire EUR-Lex structure concerning VAT (code 4.10.02 of the Directory of European Union case-law, and subcodes). This way JPOLs not related to the selected subtopics could still be linked to some of the ontology labels.

In older decisions, the keywords chains are short, few, and quite generic. In more recent decisions, they form long chains, with the last keywords being very specific. This reflects the increasing importance of judicial concepts within VAT. Moreover, in each long chain, shorter chains can often be identified, corresponding to the preliminary questions asked by the national court. For the integration of these keywords into the EUR-Lex structure, we followed the following rules:

Rule No. 1 - Ontology Levels

The ontology should not go beyond the 6th level. We believe that an ontology with more than six levels would be too detailed and specific.

Rule No. 2 - Relationship Between Keywords and EUR-Lex Concepts
The keywords may correspond to an EUR-Lex classification concept. In such cases, they can be disregarded.

Example: CELEX 62017CJ0449

Reference for a preliminary ruling — Common system of value added tax (VAT) — Directive 2006/112/EC — Article 132(1)(i) and (j) — ***Exemption for certain activities in the public interest*** *— School or university education — Concept — Driving school tuition provided by a driving school.*

EUR-Lex taxonomy

4.10.02.01.07 Exemptions

4.10.02.01.07.00 General

4.10.02.01.07.01 **Exemptions for certain activities in the public interest**

Rule No. 3 - Integration of EUR-Lex Concepts with Keywords
If the keyword does not correspond to an EUR-Lex concept, it must be evaluated to determine if it should be included in the ontology. This evaluation was done manually by domain experts. Once a keyword is selected for inclusion, it must be assigned a level within the six-level ontology. The level assignment was also done manually.

Rule No. 4 - Elimination of Terms
We decided to eliminate generic terms (e.g., "reference for a preliminary ruling," "taxation") or overly specific terms (e.g., "Service used by an insurance company to review the accuracy of a diagnosis of serious illness and provide the best possible care abroad"), as well as ignorable terms such as synonyms (e.g., "principle of fiscal neutrality," since it is a synonym of "principle of neutrality" which is already included).

For the hierarchical structure of the ontology, we adhered to the EUR-Lex hierarchical structure. We copied this structure into an Excel file and worked on this structure. In one case, we reversed the third and fourth levels, moving "determination of the taxable amount" (4th level in EUR-Lex) to the third level and shifting "supply of goods or services," "intra-community acquisitions of goods", and "importation of goods" (3rd level in EUR-Lex) to the fourth level.

We then enriched this structure with the relevant keywords. The keywords were first automatically extracted by computer scientists from the decisions in the dataset, whose composition will be

explained in detail in Deliverable D2.2 (First release of the annotated corpus + Draft guidelines). Afterward, VAT experts manually selected the relevant keywords and assigned them to appropriate levels, based on both the structure of keyword chains in CJEU case law and their domain knowledge.

7.2. Legal Ontologies

Legal ontologies provide a robust framework for representing and analysing complex legal knowledge by formally defining legal concepts and their interrelationships [5, 6].

This structured approach enables sophisticated analyses, including uncovering hidden patterns and similarities within legal systems [7]. When dealing with JPOLs as this approach not only enhances the precision of legal analysis but also promotes cross-lingual legal interoperability and harmonization, preventing redundancy and promoting integration [8, 9], ultimately fostering a more comprehensive understanding of legal principles in a globalized legal landscape.

7.3. Dataset

The dataset at the foundation of the VAT ontology comprises two primary components. The first consists of two tabular datasets containing sentences extracted from EUR-Lex and classified as either "exemptions" or "taxable amount." The second component is a hierarchical tree encoding the subsumption relationships defined by the EUR-Lex platform for its labels. This tree has been manually augmented by legal experts with keywords extracted from the sentences to provide a comprehensive overview of the topic hierarchy.

7.3.1. Judgment Datasets

Two datasets were extracted from EUR-Lex: `CJEU_Taxable_Amount_EURlex` (taxable_dataset) containing 50 judgments and `CJEU_exemption_public_interest_EURlex` (exemption_dataset) containing 51 judgments.

The two datasets contain judgment metadata as stored on the online archive, downloaded in their English translation and presented in a tabular structure comprised of the following columns:

- **Title:** the title of the judgment
- **Parts:** the parties involved
- **Object:** brief thematic description
- **Keywords:** list of keywords characterizing the case
- **Case number**
- **CELEX number:** EUR-Lex identifier
- **ECLI identifier:** European Case Law Identifier
- **Subject matter:** general document classification

- **Subject matter, 1st level** general document classification
- **Date of document:** date of issuing
- **Author:** issuing authority and judge

Example:

- **Title:** Judgment of the Court (Fourth Chamber) of 21 September 2017.
- **Parts:** Minister Finansów v Aviva Towarzystwo Ubezpieczeń na Życie S.A. w Warszawie.
- **Object:** Request for a preliminary ruling from the Naczelny Sąd Administracyjny.
- **Keywords:** Reference for a preliminary ruling — Taxation — Value added tax — Directive 2006/112/EC — Article 132(1)(f) — Exemptions for certain activities in the public interest — Exemption for the supply of services by independent groups of persons for their members — Applicability to insurance.
- **Case number:** Case C-605/15.
- **CELEX number:** 62015CJ0605
- **ECLI identifier:** ECLI:EU:C:2017:718
- **Subject matter:** Value added tax, Taxation
- **Subject matter, 1st level:** Value added tax, Taxation
- **Date of document:** 2017-09-21
- **Author:** Court of Justice, Kokott

Data Processing

The tabular datasets were initially merged into a single Pandas DataFrame. A new attribute, "Dataset," was added to each row, categorizing them according to the source dataset ("taxable" or "exemption").

Subsequent this initial step, additional processing was performed to refine the documents' content.

1. **Keyword Processing:** most entries in the merged dataset contains a "Keyword" column wherein keywords were grouped within a single string and delimited by variations of the hyphen character. Processing these keyword strings was challenging due to the inconsistent delimiters and the occasional presence of hyphens within keywords themselves (e.g., "self-employed," "non-profit"). This processing resulted in a new "**Split Keyword**" column which includes a list of individual keyword strings. The processed dataset was then stored in TSV format for reuse without further processing.

2. **Long Keyword Evaluation:** Keywords with a length of five characters or more (n=5, as determined by domain experts) were extracted from both datasets for manual evaluation. This extraction yielded 134 keywords from the `taxable_dataset` and 121 from the `exemption_dataset`. Domain experts reviewed these extracted keywords, retaining only those deemed relevant to the research task and removing the others from the respective document keyword lists.
3. **Substring Identification for JPOL Keywords:** Keywords containing the substring "principle," indicative of relevance to JPOL, were extracted. This process identified 13 keywords from the `taxable_dataset` and 3 from the `exemption_dataset`. These keywords were also presented to domain experts for further analysis.

7.3.2. Dependency Tree

The second dataset (Figure 1) comprises a hierarchical representation of the EUR-Lex label taxonomy. Domain experts manually augmented this tree by integrating the keywords extracted from the judgment (described in the previous section) together with the corresponding articles from Directive 112/2006/EC. These keywords were positioned at various levels within the hierarchy based on their conceptual relationship to the existing labels.

	2 level	3 level	4 level	5 level	6 level
VAT					
	Taxable amount				
		Determination of taxable amount (artt. 91, 92)	*Supply of goods or services (artt. da 73 a 77, 81, 82)*	Concept of 'supplies which give rise to successive payments'	
				Fees laid down by law (art. 78)	
			Price reduction (art. 79, 87)	Sales promotion scheme	
				Discount card	
			Intra-Community acquisition of goods		
			Importation of goods (art. 86, 88, 89)		
			Concept of 'subsidy directly linked to the price'		
			Consideration in the form of goods and services		

Figure 1: Hierarchical label tree

Data Processing

The initial tabular representation of the EUR-Lex taxonomy was created in an Excel table and then augmented by integrating the corresponding numerical codes for each textual label. This table was then transformed into a set of pairwise entries, each representing a subclass relationship within the

hierarchy. Each pair consists of a child node and its parent node, effectively capturing the hierarchical structure of the taxonomy.

7.4. Linguistic Versions

To ensure cross-lingual applicability, the hierarchical taxonomy and its associated keywords were translated into Italian, Swedish, and Bulgarian.

This process involved several steps for each keyword in the original English dataset. First, identification of the English keywords in each judgement. Next, leverage of the document's CELEX number and the use of web scraping techniques for the retrieval from EUR-Lex of each judgment in the target language (Italian, Swedish, or Bulgarian).

Afterward, the keywords present within the target language document were extracted and individually processed. We assumed that the keywords' order across different language versions of the same judgment is consistent. According to our assumption, a keyword in the target language occupies the same position as the English keyword within the English keywords chain from which we started building the English version of the ontology.

> *Example:*
>
> *Directive 2006/112/EC — Article 14(2)(b) — **Supply of goods** — Motor vehicles — Finance lease with an option to purchase*
>
> *Direttiva 2006/112/CE – Articolo 14, paragrafo 2, lettera b) – **Cessione di beni** – Autoveicoli – Contratto di leasing finanziario con opzione d'acquisto*
>
> *Direktiv 2006/112/EG – Artikel 14.2 b – **Leverans av varor** – Motorfordon – Avtal om finansiell leasing med köpoption*
>
> *Директива 2006/112/ЕО — Член 14, параграф 2, буква б) — **Доставка на стоки** — Моторни превозни средства — Договор за лизинг с опция за изкупуване*

This automated translation process was then subjected to a manual review by legal experts fluent in the respective target languages. This expert review ensured the accuracy and appropriateness of the translations, both linguistically and within the specific legal context.

7.5. Validation Guidelines

This section details the validation guidelines provided to the legal experts involved in evaluating the accuracy and relevance of the generated multilingual VAT ontology. These guidelines outline the criteria used to assess the quality of the translated keywords, the appropriateness of their placement within the hierarchical structure, and the overall coherence of the ontology across different languages. A clear understanding of these guidelines is crucial for interpreting the expert feedback and appreciating the rigor of the validation process.

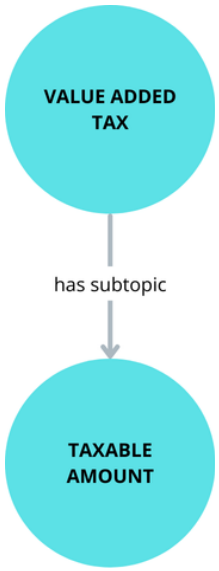
7.5.1. General Guidelines

This part of the document describes the structure and content of an ontology designed to represent the VAT domain. An ontology, in simple terms, is a way to organize knowledge about a specific subject. Think of it as a structured vocabulary that defines the key concepts and how they relate to each other. In our case, this ontology will help us understand and navigate the complex articulations of VAT law.

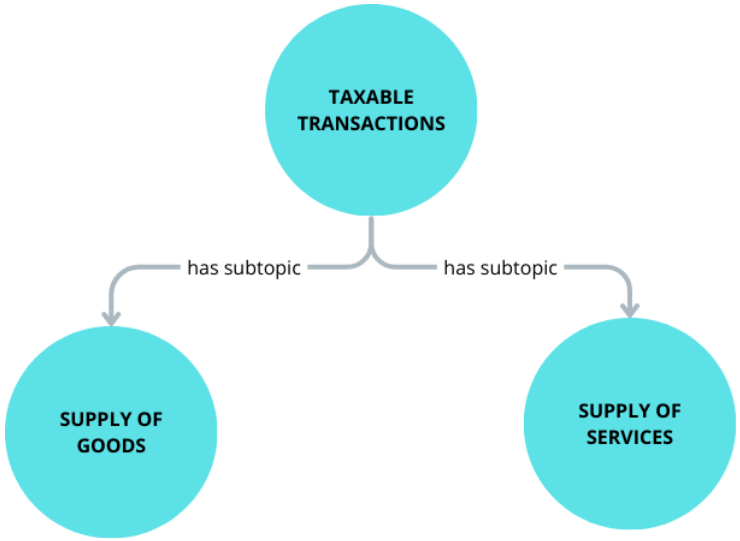
An ontology example [2], developed for the ADELE project, can be found at https://adele-tool.eu/it/1/1/Ontology/View
the aim for this phase of the project is to develop something similar, in the context of VAT regulations.

Our ontology will take the form of a hierarchical tree structure. Imagine a family tree, but instead of family members, we have legal concepts related to VAT. Each concept is represented by a "node" in the tree. These nodes are derived from two main sources: the official EUR-Lex taxonomy used to categorize VAT-related judgments, and keywords extracted from actual legal judgments in our project's dataset.

The connections between these nodes are called "edges." All the edges in our ontology are of the same type: "has subtopic." This implies that if a node A is connected to a node B with a "has subtopic" edge, then node B is a specification of node A. For example, "VAT" might be a parent node, and "taxable amount" might be a child node connected by the "has subtopic" edge, indicating that "taxable amount" is a specific aspect of the broader concept of "VAT." This hierarchical structure allows us to organize the concepts from general to specific, providing a clear and logical framework for understanding VAT law.



We use the concept of “Taxable Transaction” as an example to explain how the multilevel hierarchical ontology works. “Taxable Transaction” is a parent node. "Supply of Goods" and "supply of services" are nodes connected to "Taxable Transactions" by the "has subtopic" edge. This means that the concept of “Taxable Transactions” contains both the concepts of "Supply of Goods" and "Supply of Services". In the hierarchy, these two concepts rely on a more specific level than “Taxable Transaction”, since they differ from each other, but both are actually qualifiable as “Taxable Transaction”. This hierarchical organization continues down the tree, with each level representing a greater level of detail.



The purpose of this ontology is to create a comprehensive and accurate representation of the VAT legal domain. To ensure its accuracy, we require expert annotation.

7.5.2. Annotation Request

We provided to the legal experts a Google sheets file containing 6 different sheets:

- **tree** which contains the English version of the final tree, where the topics were previously manually aligned by the legal experts
    - short preview:

	2 level	3 level	4 level	5 level	6 level
VAT					
	Taxable amount				
		Determination of taxable amount (artt. 91, 92)	*Supply of goods or services (artt. da 73 a 77, 81, 82)*	Concept of 'supplies which give rise to successive payments'	
				Fees laid down by law (art. 78)	
			Price reduction (art. 79, 87)	Sales promotion scheme	
				Discount card	
			Intra-Community acquisition of goods		
			Importation of goods (art. 86, 88, 89)		
			Concept of 'subsidy directly linked to the price'		
			Consideration in the form of goods and services		

translation table which contains the official translation of every English topic into Italian, Swedish and Bulgarian. Some blank cells are left where the translation of a topic was not present on EUR-Lex; they have been manually imported from the corresponding linguistic version in Curia. This table was used to translate the English terminology into national languages.

- **relations_en** contains the triplets of the form "node-edge-node". Each triplet represents a connection in the ontology. For example, a triplet might look like this: "VAT-has subtopic-taxable amount". This file was created starting from the "tree" table.

A	B	C	D
Start		End	
value added tax	has subtopic	taxable amount	
value added tax	has subtopic	taxable persons	
value added tax	has subtopic	taxable transactions	
value added tax	has subtopic	place of taxable transactions	
value added tax	has subtopic	exemptions	
value added tax	has subtopic	obligations of taxable persons and certain non-taxable persons	
value added tax	has subtopic	national derogations	
value added tax	has subtopic	deductions and refunds	
value added tax	has subtopic	rates	
value added tax	has subtopic	tax evasion	
value added tax	has subtopic	administrative cooperation and combating tax fraud	
value added tax	has subtopic	chargeable event, chargeability and collection of vat	
value added tax	has subtopic	principle of fiscal neutrality	
value added tax	has subtopic	principle of effectiveness	
value added tax	has subtopic	direct effect	
value added tax	has subtopic	principle of proportionality	
value added tax	has subtopic	principle that national law must be interpreted in conformity with eu law	
value added tax	has subtopic	discretion of the member states	
value added tax	has subtopic	transitional arrangements for the taxation of trade between member states in the sixth directive	
value added tax	has subtopic	principle of equal treatment	
value added tax	has subtopic	validity and interpretation of the directive	
value added tax	has subtopic	abuse of rights	

- **relations_sv:** translation of "relations_en" into Swedish
- **relations_it:** translation of "relations_en" into Italian
- **relations_bg:** translation of "relations_en" into Bulgarian

We are asking annotators to perform the following tasks:

- **Validation of Translations:** Some of the concepts in the ontology have been translated from their original language into the national language of the project. Please verify the accuracy of these translations. If a translation is incorrect or unclear, please provide the correct translation.
- **Identification and Correction of English Terms:** While most concepts are in the national language, a few might still be in English. Please identify these remaining English terms and provide their correct translation in the national language. Ensure consistency by correcting the English term in every instance where it appears in the dataset. It's crucial to replace all occurrences of the English term with the same, correct translation to maintain uniformity within the ontology.

8. Ontology Creation

8.1. Neo4J

We used Neo4j for building the VAT ontology due to its proven effectiveness in managing complex interconnected data, a capability widely recognized in the field [1]. This graph database technology is particularly well-suited to represent the intricate web of relationships between legal concepts and legislative articles inherent to the VAT domain.

A key strength of Neo4j lies in its powerful visualization tools, which enable users to readily grasp the complex network of concepts and their interconnections within the ontology. This visual clarity is of paramount importance in legal domains, where a clear understanding of the interplay between different concepts is essential. Moreover, Neo4j's sophisticated querying mechanisms empower users to perform intricate searches within the ontology, facilitating the exploration of relationships between various VAT-related concepts.

The scalability offered by Neo4j is another compelling factor. As the ontology grows and evolves with the incorporation of new concepts and relationships, Neo4j maintains consistent performance, ensuring its long-term viability. This characteristic is particularly relevant in the context of ongoing legal research, where the data volume is expected to increase over time. Furthermore, Neo4j's robust transaction management capabilities guarantee the integrity and consistency of the data, a critical requirement in legal applications where accuracy is paramount. Lastly, the ability to export the graph database in established ontological formats like RDF, RDFS, and TTL promotes interoperability and simplifies integration with other knowledge representation systems.

8.2. Ontology Implementation

The creation of the multilingual VAT ontology involved processing the linguistic relationship files for each target language. For every entry in each language file, a *has_subtopic* relationship was defined, resulting in a structured representation consisting of the source concept, corresponding articles and EUR-Lex labels (used as node attributes), the *has_subtopic* relationship itself, and the target concept with its associated attributes. These structured representations were exported as CSV files. Finally, these CSV files were imported into Neo4j, employing a constraint on individual triples to prevent the creation of duplicate relationships, ensuring data integrity within the graph database.

Each node within the VAT ontology possesses specific attributes that provide contextual information and facilitate navigation and querying. These attributes are:

- **Node_name** holds the textual label representing the concept associated with the node. This label provides a human-readable name for the concept, facilitating understanding and interpretation within the ontology.

- **Eurlex_code** stores the official EUR-Lex code associated with the concept represented by the node. This code serves as a unique identifier and allows for direct linking to the corresponding EUR-Lex entry. For nodes representing keywords extracted from legal texts, this attribute may be empty if no direct EUR-Lex code is applicable.
- **Articles**:contains a list of Directive 112/2006/EC articles relevant to the concept represented by the node. These articles provide legal grounding and context for the concept, allowing users to understand its practical application within the VAT framework.
- **Node_type** indicates the source of the concept represented by the node determined by the presence or lack of a EUR-Lex code associated to the node. It can take one of two values:
    - **EUR-Lex** indicates that the node represents a concept derived directly from the EUR-Lex taxonomy.
    - **Keyword** indicates that the node represents a keyword extracted from the analysed legal texts. This distinction allows users to differentiate between concepts originating from the official EUR-Lex structure and those derived from the specific legal judgments under consideration.

VAT Ontology Schema

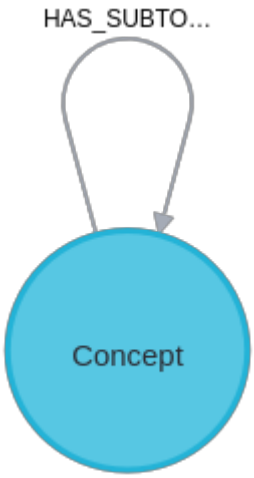


Figure 2

VAT Ontology Sample

Figure 3

8.3. Future Work

The current VAT ontology provides a solid background for future research and development. A key area for future work is the integration of automatically extracted JPOLs with the concepts present in the ontology. This integration will enable a deeper analysis of the interplay between VAT concepts and relevant JPOLs, providing a richer understanding of the legal landscape. This enhanced ontology

will facilitate more nuanced queries and analyses, allowing researchers to explore the influence of JPOLs on VAT legislation and case-law.

9. Conclusions

This deliverable marks a significant achievement in the POLINE project, establishing a robust and multilingual ontology for VAT legal concepts.
This complex framework relies on the integration of the official EUR-Lex taxonomy with keywords painstakingly extracted from a corpus of CJEU judgments, represents a substantial advancement in the field. It provides a significantly more nuanced and comprehensive resource compared to existing tools, accurately reflecting the fundamental role played by the Court of Justice of the European Union in developing the VAT field.

The development of this ontology involved a rigorous and multi-faceted process. Legal experts meticulously curated and validated each stage, from the initial selection of relevant keywords to their hierarchical integration within the existing taxonomy. This careful approach ensures the ontology's accuracy, relevance, and practical applicability across different legal systems and jurisdictions. Furthermore, the commitment to multilingualism, with translations already available in several key languages, underscores the project's dedication to accessibility and broad utility across the EU.

This structured knowledge base will serve as the cornerstone of the JPOL extraction and retrieval tool. By leveraging the ontology's rich and interconnected network of legal concepts, the tool will enable users to identify and classify relevant JPOLs with significantly greater efficiency and precision. This enhanced access to crucial legal information will empower legal professionals, tax authorities, businesses, and other stakeholders to navigate the complexities of VAT law with greater confidence and understanding. The successful completion of this ontology therefore represents a major milestone in the POLINE project, paving the way for the development of a truly impactful tool with the potential to transform the application of VAT law within the EU.

Looking ahead, the project team envisions further enhancements and expansions to the foundational ontology. Future work will prioritize incorporating additional VAT subdomains beyond the current focus on "Exemptions for Activities in the Public Interest" and "Taxable Amount," providing a more holistic and comprehensive representation of the entire legal field. Furthermore, recognizing the dynamic nature of VAT law, the team will implement mechanisms for continuous updating, ensuring the ontology remains current with the latest legislative changes and CJEU rulings. Expanding language support to encompass all official EU languages remains a key objective, further enhancing accessibility and promoting wider adoption across the Union. Crucially, the seamless integration of the ontology with the JPOL extraction tool will be a central focus, enabling the tool to fully leverage the ontology's structured knowledge for optimal performance. Finally, the project team is committed to incorporating user feedback throughout the development process, ensuring the ontology's practical utility and responsiveness to the needs of its diverse user base. By pursuing these future directions, the POLINE project aims to deliver a dynamic, comprehensive, and user-

centric resource that serves as an invaluable tool for navigating the intricacies of VAT law within the European Union.

10. References

[1] G. Sartor, P. Santin, D. Audrito, E. Sulis, L. Di Caro, Automated extraction and representation of citation network: a cjeu case-study, in: International Conference on Conceptual Modeling, Springer, 2022, pp. 102–111.

[2] A. Fidelangeli, D. Liga, R. Markovich, Using Ontological Knowledge and Large Language Model Vector Similarities to Extract Relevant Concepts in VAT-Related Legal Judgments, Cham : Springer, 2024.

[3] L. N. Brown, F. G. Jacobs, The Court of Justice of the European Communities, Sweet & Maxwell, 1989.

[4] M. Aujean, Harmonization of VAT in the EU: Back to the Future, in EC tax rev. 2012, pp.134 ff.

[5] G. Ajani, G. Boella, Luigi Di Caro, L. Robaldo, Llio Humphreys, S. Praduroux, P. Rossi and Andrea Violato. "The European Taxonomy Syllabus: A multi-lingual, multi-level ontology framework to untangle the web of European legal terminology." *Appl. Ontology*, 11 (2017): 325-375.

[6] P. Visser and Trevor J. M. Bench-Capon. "A Comparison of Four Ontologies for the Design of Legal Knowledge Systems." Artificial Intelligence and Law, 6 (1998): 27-57.

[7] Harith Alani, S. Dasmahapatra, K. O'Hara and N. Shadbolt. "Identifying Communities of Practice through Ontology Network Analysis." IEEE Intell. Syst., 18 (2003): 18-25.

[8] C. Rodrigues, F. Freitas, E. Barreiros, R. Azevedo and Adauto Trigueiro de Almeida Filho. "Legal ontologies over time: A systematic mapping study." Expert Syst. Appl., 130 (2019): 12-30.

[9] P. Visser and Trevor J. M. Bench-Capon. "A Comparison of Four Ontologies for the Design of Legal Knowledge Systems." Artificial Intelligence and Law, 6 (1998): 27-57.