OntoVAT, an ontology for knowledge extraction in VAT-related judgments*

Davide Liga[1][0000−0003−1124−0299], Alessia Fidelangeli[2], and Réka Markovich[1]

[1] University of Luxembourg, Esch-sur-Alzette, Luxembourg
{davide.liga,reka.markovich}@uni.lu
[2] Alma Mater Studiorum - University of Bologna, Bologna, Italy
alessia.fidelangeli2@unibo.it

Abstract. In this work, we introduce OntoVAT, a multilingual ontology designed for knowledge extraction in VAT-related legal judgments. To the best of our knowledge, this is the first comprehensive ontology in the field of VAT (Value-Added Tax). The main aims of this ontology are to capture the key concepts involved in the European VAT domain and to provide an extendible and reusable knowledge representation to facilitate the automated extraction or detection of VAT-related concepts in legal judgments. This ontology can also facilitate many other tasks of Artificial Intelligence and Law (AI&Law), e.g., legal knowledge extraction, keyword extraction, topic modeling, and semantic relations extraction. OntoVAT is created using OWL as the basic format of representation, with a SKOS lexicalization. We present here a first version of the ontological patterns and relations of the ontology, which we release in three different languages and which is the result of an ongoing effort between computer scientists and domain experts.

Keywords: Legal Knowledge Representation · Ontology · VAT · AI&Law.

1 Introduction

The field of Artificial Intelligence and Law (AI&Law) has seen a huge growth in recent years, with a range of applications being developed to assist legal professionals, improve access to justice, and facilitate the functioning of legal systems. One critical aspect in the development of AI&Law applications is the representation and management of knowledge, which is essential for ensuring that systems can operate effectively and deliver accurate results. Ontologies, which are formal representations of a specific domain's knowledge, can contribute to achieving this objective, and can have a crucial role in combination with

* This works has been supported by the Analytics for Decision of Legal Cases (ADELE), founded by the European Union's Justice Programme (grant agreement No. 101007420); Davide Liga was supported by the project INDIGO, which is financially supported by the NORFACE Joint Research Programme on Democratic Governance in a Turbulent Age and co-funded by AEI, AKA, DFG and FNR and the European Commission through Horizon 2020 under grant agreement No 822166

non-symbolic and sub-symbolic AI methods [10]. In fact, ontologies are crucial tools for the advancement of the field of AI&Law, since they provide a way to accurately represent complex symbolic knowledge in machine-readable format, while preserving the advantages coming from being modular and inter-operable components. In this work, we propose a first version of OntoVAT, an ontology designed for knowledge extraction from legal judgments related to Value Added Tax (VAT). The main aims of this ontology are to capture the key concepts involved in the European VAT domain and to provide an extendible and reusable knowledge representation for extracting VAT-related concepts for the analysis of judicial decisions or, more generally, for the analysis of judgments. These kinds of ontology can facilitate tasks such as the retrieval of keywords, topic modeling, the extraction of semantic relations, etc.

In the next sections, we will describe the few related works and our own contributions (see Section 2), the methodology we adopted (see Section 3), and the current structure of the ontology (see Section 4). Finally, in the last part of the work we will provide some suggestions for future developments in the field (see Section 5).

2 Related Works and Contributions

Ontologies are important tools in the field of AI&Law [11], and have been used in various contexts such as the modeling of privacy law [9] or the recent Artificial Intelligence Act [2]. Nonetheless, there are no attempts to build a comprehensive ontology related to Value-Added Tax (VAT). To the best of our knowledge, the only attempt to build an ontology in this field dates back to 20 years ago [7] [13], when Karremans et al. pursued to describe a few potential core ontological concepts related to VAT. However, their work was more dedicated to showing the obstacles related to the design of complex multilingual ontologies (where culture-specific or language-specific elements can create constraints or limitations during the design of the ontology) than aimed at creating a complete VAT ontology. The authors' proposal was limited to a few interesting conceptual suggestions for the development of a potential VAT ontology.

This absence of related works is probably due to the difficulty in reconstructing such a complex and articulated legal (and conceptual) domain. Indeed, the creation of an ontology in the field of VAT entails many critical issues: (1) while most VAT concepts are harmonized at the European level, others are regulated (or even mentioned) only at the national level; (2) the VAT regulation relies on the use of concepts belonging to other domains of law (such as civil law or commercial law) or common language concepts which are employed with a particular meaning in the field of VAT; (3) many VAT concepts are not defined by the VAT Directive or national legislation, but by the case law of the Court of Justice of the European Union (CJEU). Thus, on the one hand, the modeling of VAT concepts requires an analysis on multiple levels, considering: European legislation, case law, and national implementations. On the other hand, it requires an analysis of concepts belonging to several fields of law, as well as to

common language. We decided to build an ontology at an intermediate-low layer of abstraction while committing it to already existing upper ontologies. In this regard, there are already many other upper ontologies designed to represent higher levels of abstraction, including the Legal Knowledge Interchange Format (LKIF), an upper ontology designed for legal knowledge [6]).

Another important aspect behind the design of OntoVAT is that it has an applicative intended use. It has been designed with the purpose of capturing the concepts which might be crucial in the legal reasoning of VAT-related judgments and especially with decisions concerning taxable/exempt VAT transactions. Hence, we had to focus both on relatively abstract concepts such as "transaction" or "place", which were frequently mentioned in the above-mentioned decisions, as well as on more specific concepts beloning to the domain of VAT (such as the concepts of "exemption" or "supply of goods"), or to specific areas of knowledge (for example "vessels" or "human blood"). The above-mentioned challenges are related to the difficulty of building an ontology capable of being expressive and representing such a large number of layers of abstractions belonging to different conceptual areas. A further challenge was to ensure the consistency of the resulting model from a formal point of view. For this reason, we decided to create this ontology in OWL format, so as to provide the scientific community with a first formal ontology, on which to explore automated reasoning experiments. Here, we present this first version of OntoVAT as a multilingual ontology (implemented in English, Italian, and Bulgarian) which is both consistent from a formal point of view and tailored to a specific applicative goal, namely modeling the most crucial concepts in VAT-related legal judgments.

3 Methodology

For the creation of OntoVAT, we were inspired by [9], which adopted a methodology to minimise the difficulties for legal operators to define a legal ontology.

We followed a top-down approach applied on legal sources and made more robust by the partial reuse of pre-existing ontology patterns [5]. Our results are evaluated by using foundational ontologies (in particular LKIF [6], DOLCE [3] and DUL [1]), and we followed the principles in the OntoClean [4] method, according to which each ontological concept can be evaluated based on three meta-properties:

1. "identity" (making sure that a class uniquely identifiable)
2. "unity" (making sure that instances of a class form cohesive and meaningful wholes)
3. "rigidity" (whether a property is essential to the instances of a class or if it can change over time)

Our validation involved a strongly interdisciplinary group, mostly composed of computer scientists, lawyers, and philosophers, which allowed an integrated expertise coming from different disciplines.

We can summarise our approach in the following steps:

(i) a group of legal experts selected nearly 500 legal judgements related to the domain of VAT in Italian and Bulgarian;
(ii) the judgements were analyzed and the portions of text related with the judges' motivations were annotated;
(iii) Italian and Bulgarian legal experts analysed the most important concepts mentioned in the judgements, checking these concepts against their respective statutory backgrounds;
(iv) our technical team received the selected concepts and portions of text from the legal experts to map them into the ontology;
(v) for each element of the ontology our legal experts provided a range of linguistic variations/synonyms, a definition, the most common examples instantiating that concept, the most common related terms, and any relevant normative references related to the concept;
(vi) the gathered results were validated by the legal team that returned them to the technical team who implemented the new information in the ontology;
(vii) the steps from (iii) to (vi) were iterated several times to refine the ontology;

We are also in the process of implementing an algorithm which uses the OntoVAT to determine whether an ontological concept is relevant on judgements related to VAT, i.e. if a specific decision deals with one or more of the ontological concepts. This process can be summarised as follows:

1. legal experts were asked to select from OntoVAT the ontological concepts which are considered more relevant in the decisions of judges;
2. considering the concepts selected in the previous step, legal experts were asked to manually annotate nearly 70% of the judgements by including the information of whether each selected concept is relevant in each judgement by associating a binary value, where 0 means "non relevant" and 1 means "relevant" (the concept is considered relevant if the court's decision concerns that concept from the substantial point of view);
3. an algorithm designed by the technical team encodes the information contained in the ontology to predict whether or not a concept is relevant (comparing the results with the gold standard defined in the previous step);

We are currently in the process of completing step 2 and implementing step 3. Our preliminary results shows that by using OntoVAT we can catch the most important relevant concepts in the judicial decisions.

This methodology can be generalized and applied to different domains (and can be easily extended to other languages). For example, we employed the same approach for the development of another ontology, PaTrOnto, related to the domain of patents and trademarks [8]. The main difference between PaTrOnto and OntoVAT is related to the above-mentioned step (iii), since the statutory backgrounds for the field patents and trademarks is completely different, also in terms of harmonisation at the European level.

4 The design of OntoVAT

4.1 Core concepts

It is worth mentioning that the ongoing effort behind this work is the result of the cooperation between computer scientists and legal experts in the VAT domain. Regarding the design of OntoVAT, we proceeded by taking into account different sources of information. First of all, we considered the European VAT Directive, which is the main legal source at the European level. The Directive provides a harmonized and coherent perspective on the ontological concepts of the VAT domain and it is compatible with our target of creating a multilingual VAT ontology, as it is available in all the official languages of the European Union. Moreover, we also considered another source of information, namely the case law of the CJEU, which we found particular useful to find key concepts which were not defined by the Directive. Finally, we tried to model the key ontological concepts with an even more concrete source of information, namely the (above-mentioned) dataset of VAT-related judgments adopted by national courts. More specifically, we analyzed which concepts were particularly important in the legal reasoning of national judges, and how these concepts were employed by them. Thus, while the Directive was the fundamental starting point of the work, this was complemented by further research aimed at identifying the concepts that were actually relevant in real cases decided by national courts or by the CJEU. Therefore, one of the first challenges was to reconcile these two aspects (i.e., the more abstract normative dimension and the more concrete dimension of judicial cases).

Inspired by the first articles of our first source (i.e., the EU VAT Directive), we decided to put at the center of our ontology the concept of "Transaction", around which we added all the other concepts. The core ontological concepts are shown in Figure 1.

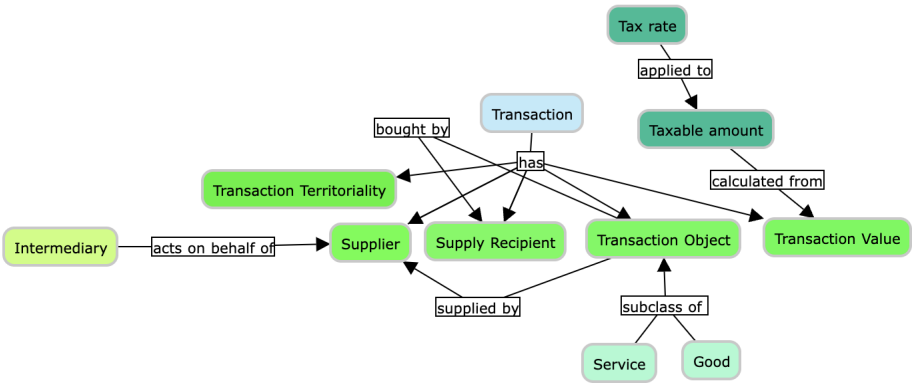

Fig. 1. The core elements of OntoVAT.

Thus, the idea is that any transaction which may (or may not) be subject to VAT will have some agents involved (a supplier, a recipient, and sometimes intermediaries), an exchanged object (generally, a service or a good), and an exchanged value from which the taxable amount is calculated. We also added the concept of territoriality, as it has consequences on the fact that the transaction is actually taxed. Starting from these core ontological concepts, we then further developed the ontology by extending their modeling. For example, a challenging step during the design of the ontology was related to the modeling of the *objective profiles* and the *subjective profiles* of the transaction, i.e., which people are subject to VAT according to the European VAT Directive (which people are taxable persons), and what kind of transactions and transactions objects (e.g., types of goods and services) are relevant for the judges to take their decisions. Furthermore, we included the concept of "Exemption" and "Right to deduction", modeling also the relation with the concept of "VAT Chargeable Event", since we realized that these concepts were very relevant in our dataset of national decisions.

4.2 OntoVAT details and lexicalisation

The ontology is currently composed of 129 concepts (i.e., OWL classes) and 36 properties (relationships between classes). A more exhaustive numerical description is reported in Table 1.

Element	Quantity
Number of classes	122
Number of properties	28
Number of datatype properties	8
Number of transitive properties	0
Number of disjoint class pairs	578
Number of subclass relations	101

Table 1. OntoVAT's statistics.

OntoVAT is a multilingual OWL ontology enriched with a SKOS lexicalisation and implemented in English, Italian and Bulgarian. This OWL+SKOS multilingual implementation has been implemented using VocBench 3 [12] and is a powerful approach to mitigate the issue of semantic non-uniformity in multilingualism, which has been pointed out in previous research [7]. Thanks to the use of SKOS, each ontological concept (i.e. each OWL class) is enriched with some specific properties which are incorporated in the SKOS data model, namely:

- skos:definition
- skos:scopeNote
- skos:altLabel

- skos:hiddenLabel
- skos:example

The addition of these properties to each ontological concept (in English, Italian and Bulgarian) facilitates the integration of crucial information within the ontology, making OntoVAT particularly expressive and powerful. In particular, **skos:definition** contains the definition of each single OWL class (i.e., the definition of each single concept). In **skos:scopeNote**, we added relevant specifications about the skos:definition field (whenever was necessary to further specify the interpretative angle of the chosen definition). Furthermore, scopeNotes also contain all relevant normative references (if any) describing the concept. We also added any relevant synonyms in the three different languages as **skos:altLabel** properties. In **skos:example**, we added some examples of the concept (which might look like further potential subclasses of the concept). Finally, the property **skos:hiddenLabel** is used to store terms in natural language which might signal the presence of the concept in the text (this can be useful for any application layers built on top of OntoVAT).

As mentioned before, we built OntoVAT using concepts taken from the European VAT Directive to grant a coherent and harmonic conceptual framework. Therefore all concepts are already designed to be appropriate for both Italy and Bulgaria. In fact, Italy and Bulgaria must grant the uniform application of European law.

In most cases, the semantic meaning of concepts is therefore harmonic between Italy and Bulgaria. In these cases, for each OWL class, a skos:definition is just provided in English and translated into Italian and Bulgarian with no adjustments. However, in few cases, definitions of concepts (i.e., their semantic meaning) vary at national level. In these situations, priority was given to national definitions, therefore the skos:definition in Bulgarian/Italian will not be just a translation from English, instead it will be a different definition (coherent with the national legislation). Moreover, whenever further specifications are needed to explain the scope of the concepts' meaning (at Bulgarian, Italian, and European level), we employed a skos:scopeNote property in Bulgarian/Italian/English.

Lastly, since national legislation may have alternative terms for referring to the Directive's concepts, we handled alternative terms as synonyms (skos:altLabel) in Italian/Bulgarian. For the time being we did not introduce any country-specific class, as our goal was to develop a common ontology which could be used by both Italian and Bulgarian judges. Moreover, the creation of a common ontology may be useful in developing a common conceptual framework that promotes the uniform application of EU law in a harmonised field. In the future, we will consider extending our ontology by adding specific classes based on concepts which are used by the legislator in national implementation. This might be useful for national judges, who might be more familiar with different country-specific concepts.

Hence, we handle the issue of multilinguality by specialising the skos properties skos:definitions, skos:scopeNotes and skos:altLabels whenever needed, with-

out affecting the coherence of the ontological concepts or their relations (Figure 2 shows an example of how multilinguality is handled for a specific concept/class).

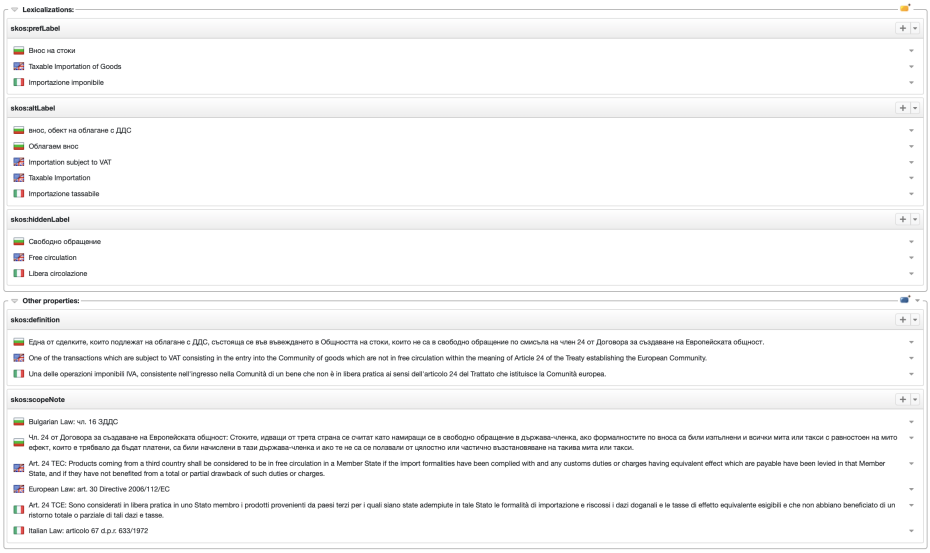

Fig. 2. An example of multilingual lexicalisation, related to the OWL class (i.e. the concept) "Taxable Importation of Goods".

We carefully assigned a definition to each concept by giving priority to definitions coming from the domain-specific legislative sources, whenever the concept exists in that domain. If the concept is not mentioned neither in the national nor in the European legislative sources, we searched for a definition in the case law of the Court of Justice of the European Union (CJEU). If the concept is not defined neither in the legislation nor in the case law of the CJEU, as it frequently happens for "factual concepts", it is defined following a simple description based on legal encyclopedias or dictionaries. In this way, we made sure that the definition of each concept coherently anchored to the legals sources.

4.3 Commitment and scope

Figure 3 shows a simplified conceptual map that gives a clearer understanding of the formal structure of the ontology, showing most ontological classes and properties which can be found in the OWL ontology[3]. In this map, one can see the previously mentioned core elements having the class "Transaction" as

[3] Relations such as "has" connecting to a target concept are represented in OWL as "hasTargetConcept", while relations such as "can be" are translated in OWL as datatype properties with a boolean value.

central concept, as we previously described in Figure 1. To make the picture more readable, the classes “Supplier”, “Good” and “Service” have been duplicated and expanded at the bottom of the map and some classes have been omitted. Please see Figure 4 for the complete hierarchies of classes and properties in OntoVAT.

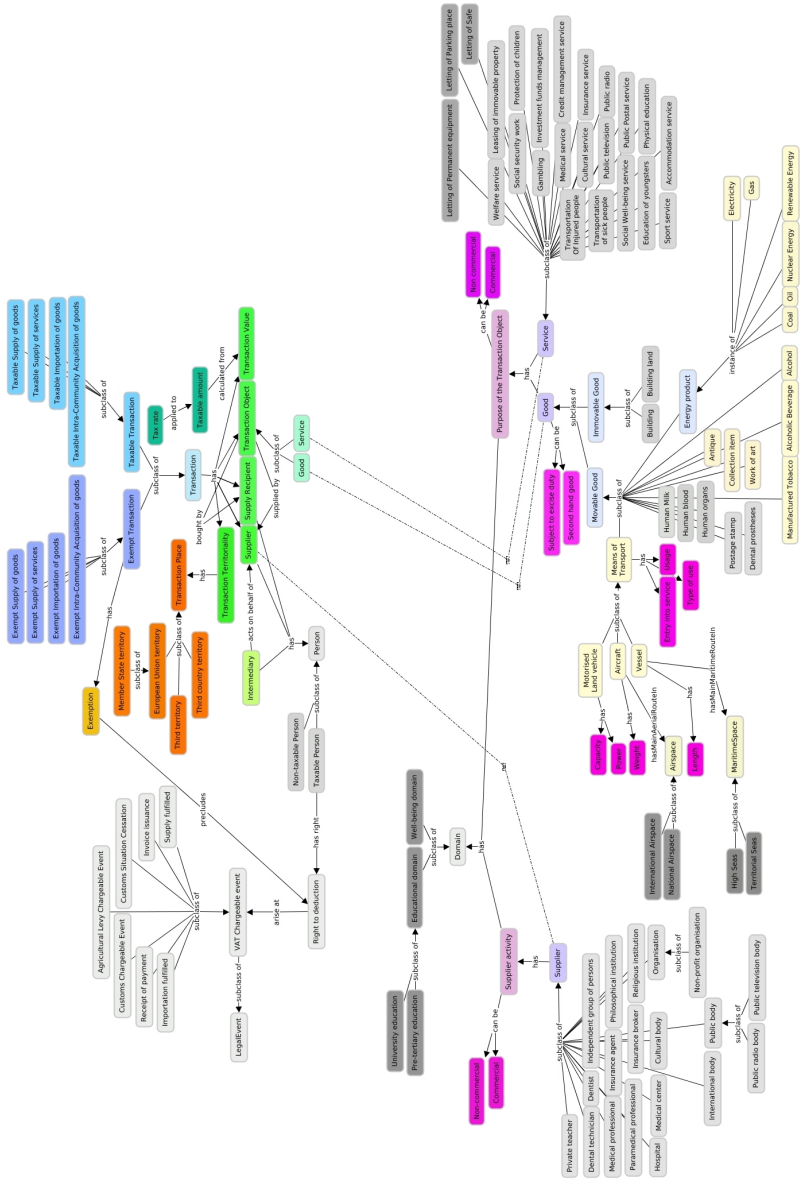

Fig. 3. Simplified map of the main concepts and relations in OntoVAT.

.

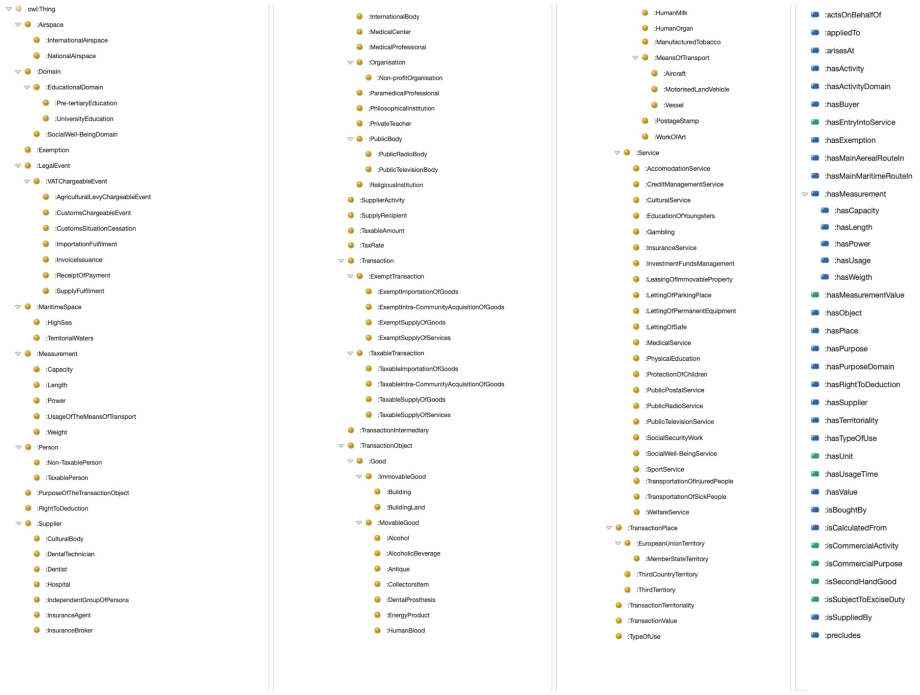

Fig. 4. All hierarchies of classes and properties.

.

To grant ontological robustness across the conceptual framework, most classes in OntoVAT are designed to be disjointed. The only class we decided not to disjoin are **VAT Chargeable Event**, **Domain**, and **Supplier**.

As can be seen in Figure 5, we did not disjoin the subclasses of "VAT Chargeable Event" to allow an instance of VAT chargeable event to belong to multiple types of chargeable event. Regarding the "Domain" concept, we preferred to allow an instance of domain to belong to multiple classes because the supplier's activity might sometimes involve an overlap of multiple domains, and because a domain might sometimes be defined as an intersection of multiple sub-domains. For the same reason, we also wanted to allow potential overlaps in the subclasses of the concept "Supplier".

These choices of allowing the overlap in the above mentioned cases (i.e., VAT chargeable events, domains and suppliers) might be made clearer with an example: an individual of the class "Dentist" could also be, in principle, an individual of the class "Private Teacher". Similarly, we decided that it was safer to leave potential overlapping among the sub-classes of "VAT chargeable event" as well as among the sub-classes of "Domain".

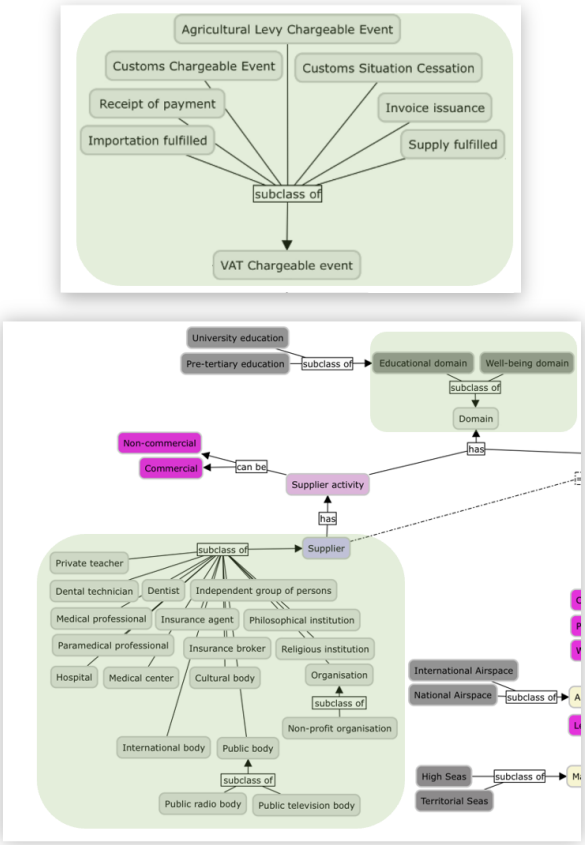

Fig. 5. The portions of OntoVAT which allow internal overlaps (i.e. where individuals can belong to multiple classes) are the subclasses of "VAT Chargeable Event" (image in top), "Domain" and "Supplier" (both depicted in the image at the bottom).

.

4.4 Alignment with upper ontologies

To make OntoVAT more robust and interoperable, we are exploring alignments to other well-known legal upper ontologies, in particular LKIF (Legal Knowledge Interchange Format) [6]. We also align to the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) and to the DOLCE+DnS Ultralite (DUL) ontology [1].

We list the alignment of our classes in Table 2, while Figure 6 shows the alignments and commitment to the upper ontologies as described in Table 2.

OntoVAT class	Aligned with class	In	Comment
Airspace	place:Place	LKIF	
Domain	expression:Qualification	LKIF	
Exemption	norm:Norm	LKIF	An “exemption” is the result of interactions between norms, which are meant to assess if an exemption occurs.
Legal Event	top:Spatio Temporal Occurrence	LKIF	
Maritime Space	place:Place	LKIF	
Measurement	dul:Unit Of Measure	DUL	
Purpose Of The Transaction Object	expression:Qualification	LKIF	
Person	legal-action:Legal Person	LKIF	
Right To Deduction	norm:Right	LKIF	
Supplier	legal-role:Legal Role	LKIF	
Supplier Activity	expression:Qualification	LKIF	
Supply Recipient	legal-role:Legal Role	LKIF	
Taxable Amount	dul:Amount	DUL	
Tax Rate	dul:Amount	DUL	
Transaction	action:Trade	LKIF	action:Trade seems more appropriate than action:Transaction
Transaction Intermediary	legal-role:Legal Role	LKIF	
Transaction Object	dolce:Substantial	DOLCE	Regarding its subclasses, “Good” aligns to “dolce:Agentive physical Object”; “Service” aligns to “dolce:SocialObject”
Transaction Place	place:Place	LKIF	
Transaction Territoriality	norm:Norm	LKIF	The concept of “territoriality” is the result of interactions between norms, which are meant to assess a given geographical space (i.e. the “Transaction Place”).
Transaction Value	dul:Amount	DUL	
TypeOfUse	expression:Qualification	LKIF	

Table 2. Alignment and interoperability with upper ontologies.

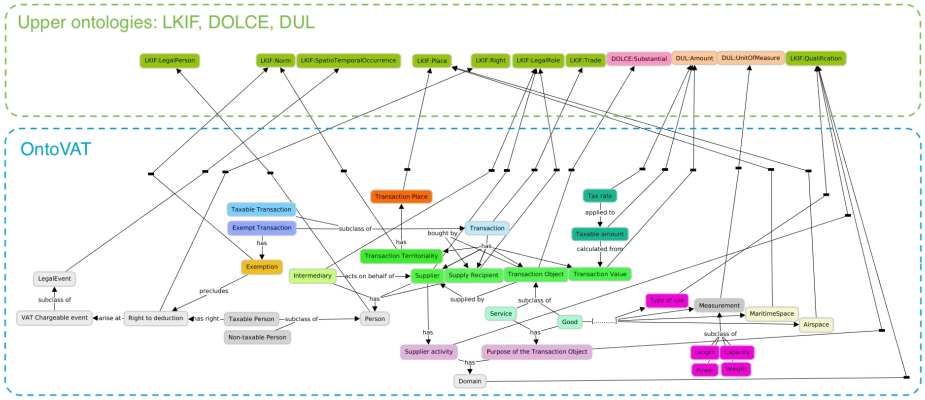

Fig. 6. OntoVAT alignments with upper ontologies.

.

5 Conclusion

In this work, we presented the first version of OntoVAT, the first formal ontology in the legal domain of VAT. The ontology has been created in cooperation with domain experts and computer scientists, and is designed to capture key VAT-related concepts in judicial decisions. The ontology is designed in OWL and is enriched with a SKOS lexicalisation in three different languages (English, Italian, and Bulgarian).

Regarding the applicative level, we are currently using this ontology to support a Natural Language Processing (NLP) pipeline, designed to extract the relevance of VAT-related concepts in our dataset of annotated legal judgments. We are also using OntoVAT to facilitate automated legal knowledge extraction from VAT-related legal documents and to build a navigation tool, through which one can find relevant judgments depending on the selected ontological concepts, through the use of semantic similarity measures.

Combining OntoVAT with an NLP pipeline is only one of the potential applications of this ontology. In the future we plan to explore other kind of targets related to legal knowledge extraction, in combination with Machine Learning.

References

1. Borgo, S., Masolo, C.: Foundational choices in dolce. In: Handbook on ontologies, pp. 361–381. Springer (2009)
2. Dimou, A., et al.: Airo: An ontology for representing ai risks based on the proposed eu ai act and iso risk management standards. In: Towards a Knowledge-Aware AI: SEMANTiCS 2022—Proceedings of the 18th International Conference on Semantic Systems, 13-15 September 2022, Vienna, Austria. vol. 55, p. 51. IOS Press (2022)
3. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with dolce. In: Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web: 13th International Conference, EKAW 2002 Sigüenza, Spain, October 1–4, 2002 Proceedings 13. pp. 166–181. Springer (2002)
4. Guarino, N., Welty, C.A.: An overview of ontoclean. Handbook on ontologies pp. 201–220 (2009)
5. Hitzler, P., Gangemi, A., Janowicz, K.: Ontology engineering with ontology design patterns: foundations and applications, vol. 25. IOS Press (2016)
6. Hoekstra, R., Breuker, J., Di Bello, M., Boer, A., et al.: The lkif core ontology of basic legal concepts. LOAIT **321**, 43–63 (2007)
7. Kerremans, K., Temmerman, R., Tummers, J.: Representing multilingual and culture-specific knowledge in a vat regulatory ontology: Support from the termontography method. In: On The Move to Meaningful Internet Systems 2003: OTM 2003 Workshops: OTM Confederated International Workshops, HCI-SWWA, IPW, JTRES, WORM, WMS, and WRSM 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings. pp. 662–674. Springer (2003)
8. Liga, D., Amitrano, D., Markovich, R.: Patronto, an ontology for patents and trademarks. In: New Frontiers in Artificial Intelligence: JSAI-isAI 2023 Workshops, AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers. Springer (2024)

9. Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L.: Pronto: Privacy ontology for legal compliance. In: Proc. 18th Eur. Conf. Digital Government (ECDG). pp. 142–151 (2018)
10. Rodríguez-Doncel, V., Palmirani, M., Araszkiewicz, M., Casanovas, P., Pagallo, U., Sartor, G.: Introduction: A hybrid regulatory framework and technical architecture for a human-centered and explainable ai. In: AI Approaches to the Complexity of Legal Systems XI-XII, pp. 1–11. Springer (2020)
11. Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M.: Approaches to legal ontologies: Theories, domains, methodologies. law. Governance and Technology series. Springer (2011)
12. Stellato, A., Fiorelli, M., Turbati, A., Lorenzetti, T., Van Gemert, W., Dechandon, D., Laaboudi-Spoiden, C., Gerencsér, A., Waniart, A., Costetchi, E., et al.: Vocbench 3: A collaborative semantic web editor for ontologies, thesauri and lexicons. Semantic Web **11**(5), 855–881 (2020)
13. Temmerman, R., Kerremans, K.: Termontography: Ontology building and the sociocognitive approach to terminology description. Proceedings of CIL17 **7**, 1 (2003)