

A Pragmatic Approach to Semantic Annotation for Search of Legal Texts - An Experiment on GDPR

Adeline NAZARENKO ^{a,1}, François LÉVY ^a and Adam WYNER ^b

^aLIPN, University Sorbonne Paris Nord, France

^bDepartment of Computer Science, Swansea University, United Kingdom

Abstract.

Tools must be developed to help draft, consult, and explore textual legal sources. Between statistical information retrieval and the formalization of textual rules for automated legal reasoning, we defend a more pragmatic third way that enriches legal texts with a coarse-grained, interpretation-neutral, semantic annotation layer. The aim is that legal texts can be enriched on a large scale at a reasonable cost, paving the way for new search capabilities that will facilitate mining of legal sources. This new approach is illustrated on a proof-of-concept experiment that consisted in semantically annotating a significant part of the French version of the GDPR. The paper presents the design methodology of the annotation language, a first version of a *Core Legal Annotation Language* (CLAL), together with its formalization in XML, the gold standard resulting from the annotation of GDPR, and examples of user questions that can be better answered by semantic than by plain text search. This experimentation demonstrates the potential of the proposed approach and provides a basis for further development. All resources developed for that GDPR experiment are language independent and are publicly available.

Keywords. Text annotation, Semantic search, Annotation methodology, Semantic markup language, Law

1. Introduction

One of the main early objectives of AI and Law [1] has been to analyse legislation and regulations to allow for querying and facilitate reasoning. The [European General Data Protection Regulation \(GDPR\)](#)², imposes data protection by design and by default rules on all companies and organisations that collect and use personal data. This requires all organisations in all EU member states to evaluate their compliance with the GDPR, which is why it is so important to provide tools assisting legal analysis.

Taking the reasoning approach, researchers have designed ontologies adapted to privacy issues [2,3] combined an ontology with a deontic logic [4], formalized the rules of GDPR provisions [5], and encoded rules in a machine-readable form [6], amongst others. Despite these efforts, automating reasoning for compliance evaluation seems un-

¹Corresponding Author: A. Nazarenko, LIPN, France; E-mail: adeline.nazarenko@lipn.univ-paris13.fr.

²Adopted in 2016 and entered into force the 25th May 2018.

likely in the near term or on a significant scale. The *knowledge bottleneck* in the process of translating from Natural Language to formal and exploitable logical representation introduces several difficulties: the process of interpretation resists formalization because of the ambiguity and vagueness of legal texts along with the plurality of interpretation; operationalization – from few articles to a whole law and a set of laws – is a challenge in itself; and legal practitioners have to be convinced that they can use the new services.

Even if such issues remain of great research interest, we consider with [7] that information retrieval represents a promising alternative approach, more practicable on a large scale and more flexible to address the diversity of drafters’ and legal professionals’ uses (e.g. retrieving, clustering, comparing and contextualizing provisions). While public services³ are adapted for the browsing of multiple legal texts, they do not offer advanced semantic search functionalities, which could facilitate the daily work of legal professionals as well as contribute to the impact of digital and AI methods in the long term. Semantic annotation and search can usefully identify the named entities, other relevant textual segments, and relationships amongst them.

To navigate between the difficult analysis of statutory rules and needs of current legal practice, we propose a coarse-grained and interpretation-neutral approach to annotating legal texts with semantic information, enabling semantically-based information retrieval capabilities. Since the annotations apply to textual passages, leave out finer details, and constitute a simple language, people familiar with legal sources can annotate the text; the annotators do not need to be professional lawyers or logicians. Even if any annotation involves some interpretation, annotation in CLAL should remain ”neutral” or consensual, in the sense that annotators must agree significantly (high inter-annotator agreement) as to the annotation of textual passages, leaving aside what does not meet with consensus.

This paper reports on a proof-of-concept experiment on the GDPR. It provides an annotation language, methodology, and annotated text. We demonstrate semantic search by answering user queries, such as *What are the obligations of a data controller?*, *What are the rights of the data subject?*, *What are the possible fines and sanctions issued in response to violations by a data controller?* *Who supervises a data controller?*⁴. This shows the practicability of the language, method, and corpus for semantic search.

The article is organized as follows. Sections 2-4 present the core of the work: the approach taken to defining annotation guidelines, the proposed vocabulary and the study on GDPR. Section 5 describes previous work on the semantic analysis and annotation of legal texts. Section 6 concludes and outlines the perspectives of this work.

2. GDPR annotation methodology

For this work, we followed the methodological recommendations inherited from the major annotation campaigns of the 1990s and 2000s [9].

Design of the annotation language Ideally, the annotation should identify elementary provisions so that they can be easily found and browsed by users. The *provision fragment* (hereafter *fragment*) is the main element of the annotation language: its text span

³Such as Legifrance (www.legifrance.gouv.fr) or UK legislation (www.legislation.gov.uk).

⁴For demonstration purpose, we started with a small test set of user questions, some of which were derived from the competency questions of [8].

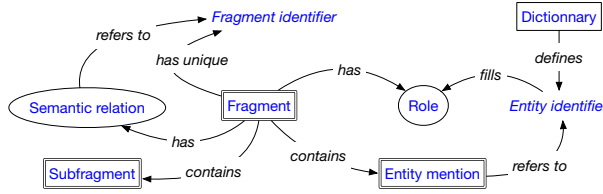


Figure 1. Relations between the CLAL components (in blue) encoded as XML elements with or without text content (double and simple square boxes), XML attributes (round boxes) and identifiers (no box).

corresponds to that of a sentence and it is a practical approximation of an elementary provision. Different types of fragments have been introduced, so that all sentences can be categorized, and the regulation be entirely annotated.

We also defined: some entity categories to account for the actors and concepts that play key roles in provisions; some relations to model the regulation as a semantic graph of entities and provision fragments; and a sub-fragment for encoding exceptions expressed not as independent sentences or fragments but as clauses or prepositional phrases.

Annotation tuning The tuning of the vocabulary and annotation instructions was done together with the annotation of the GDPR in an incremental and iterative way by two annotators working in parallel. We first selected the concepts on which there was a broad consensus for the description of elementary provisions, e.g. obligation and permission, and added more as necessary. The discussions on disagreements or misunderstanding led to adding, merging, splitting or redefining categories, until the whole process stabilized when no further language revision was needed after four months. We also had to give up some elements that could not be annotated in a consistent way, like conditional sub-fragments or actions. Approximately, one third of the GDPR was annotated by then, but the language may still evolve according to the needs dictated by the remaining text, for example to add a sanction type.

Writing the annotation instructions Annotation guidelines were written at the end of the annotation tuning phase. They include a presentation of the annotation vocabulary, the semantics associated with each element, a description of its syntax, some examples, and recommendations to facilitate the arbitration of difficult cases.

Evaluation of the annotation The quality of an annotation is first measured by its homogeneity and stability, using inter- and intra-annotator agreements. Comparing the annotations of parallel versions of the text (e.g. English/French) would also be interesting.

3. CLAL Semantic annotation vocabulary

This section presents and discusses the main components of the XML CLAL vocabulary, focusing on the most original or less obvious aspects. Figure 1 and Table 1 give an overview of the language, its vocabulary components, and how they relate to each other in the annotation.

3.1. Fragment categories

The language has two types of provision fragments, deontic and non-deontic ones.

Table 1. CLAL vocabulary: main elements and attributes

XML elements				XML attributes	
Entity mentions	Fragments		Sub-fragments	Roles	bearer target obj
	Deontic	Non-deontic			
CONCEPT	OBLIGATION	DEFINITION	EXCEPT	Semantic relations	rel except
PERSON	PROHIBITION	LEGAL_PRECISION			
LEGAL_ENTITY	PERMISSION	QUALITY_ATTRIBUTION			
	POWER RIGHT	EXCEPTION		Other	type

The deontic categories correspond to the traditional, familiar deontic concepts for which CLAL provides 5 types of XML fragment elements: `OBLIGATION`, `PROHIBITION`, `PERMISSION`, `RIGHT`, and `POWER`. The `POWER` case deserves an explanation. The GDPR gives the power to some institutions to specify the impact of rules according to various contextual parameters, as in Art 45 §3,⁵ which delegates to the Commission the definition of an "adequate level of protection". The empowered decisions have statutory consequences, which differs from permissions.⁶ A `POWER` is further specified as `ruling` or `execution` depending if it concerns a rule definition or an actual case.

A lot of attention was paid to fragments having no obvious deontic semantic value. Definitions and exceptions are usually easy to identify, but 48 out of 178 fragments appeared to pertain neither to a clear deontic category, nor to definitions or exceptions. After discussion and different tests, we included in the vocabulary two additional categories which are sub-typed with the help of an attribute. `QUALITY_ATTRIBUTION` labels fragments specifying that an entity has a given quality entailing legal consequences⁷. `LEGAL_PRECISION` labels fragments that complement other fragments. For instance, the fragment "*The information shall be provided in writing*" (Art. 12) complements a previous `OBLIGATION`.⁸ In this case, the precision has a procedural type (`type="procedure"`) but there are also text specifications (`type="text_specification"` for provisions constraining the content of a text, such as contracts or statutory decisions) and underspecified precisions (`type="default"`).

3.2. Entities

The annotation points out the entities that play key roles in the regulation. In the annotation perspective, the entity which is referred to is and must be distinguished from its mentions, the occurrences of (multi-)words that refer to it. We identified three main types of entities in the GDPR: concepts, persons, and legal entities, the latter two being actors.

3.3. Roles and semantic relationships

To relate the identified elements, the language allows for entity to fragment relations (roles) and fragment to fragment relations (semantic relationships), which are all en-

⁵The Commission may decide, . . . , that a third country, . . . ensures an adequate level of protection

⁶Transfers to the third country is allowed without further control.

⁷The fragment "*Any controller [. . .] shall be liable for the damage caused by processing which infringes this Regulation*" (Art 82 §2) is encoded as a `QUALITY_ATTRIBUTION` of a `responsibility` type.

⁸The controller shall take appropriate measures to provide any information [. . .] to the data subject [. . .].

coded as XML fragment attributes. The goal of annotation is not to provide a full semantic graph of the regulation provisions and entities but to encode the relations that are indisputable, useful for answering user queries, and otherwise inaccessible to users.

An entity to fragment relation denotes the role played by the entity in the provision. For instance, the `obj` attribute relates the concept to its `DEFINITION` fragment, the `bearer` attribute indicates the actor primarily concerned by a provision (e.g., the one who is granted a permission or to whom is attributed a quality), and the `target` attribute specifies the actor who incurs an obligation with respect to the right of a person .

The annotation language provides two types of relationship between fragments. 1) The generic dependency one (`rel`) indicates that one fragment specifies the meaning or must be interpreted in the light of another. E.g., it is used to relate a `LEGAL_PRECISION` to the fragment it complements. 2) The exception relationship is more specific.

Because of their importance for legal reasoning, special attention is given to exceptions. There are several ways to annotate an exception relation, allowing the annotation to follow the structure of the text, but all of them indicate that a piece of text *B* introduces an exception to the rule expressed in a fragment *A*. If *B* is a fragment, it has an `except` attribute with the identifier of *A* as value. *B* can be any type of fragment and it is annotated as `EXCEPTION` if it has no other identified semantic value.⁹ Besides that, a specific sub-fragment element `EXCEPT` is introduced to mark-up exceptions when the rule and its exception are part of the same sentence.

4. GDPR experiment

The proposed pragmatic approach supporting semantic search has been implemented and tested on the text of the GDPR. This included the formalization of the language itself in XML (annotation vocabulary and rules), the actual annotation of a large part of the French version of the regulation, and the evaluation of the overall approach. The XML CLAL description and the annotated corpus are openly available together with guidelines for the annotation of other versions of GDPR or other regulations.¹⁰ The Oxygen XML Editor¹¹ and an *ad hoc* semantic search tool were used for the development phase.

4.1. Implementation of the annotation language

The annotation language includes the vocabulary and the set of constraints that regulate the use of annotations. It is formalized as an XML schema which follows two design principles: independence of annotation layers and maximum control of annotations.

Since two different annotation layers – pre-existing structural and CLAL semantic annotations – enrich the same text, it is essential to preserve their independence. Each one has its own namespace and XML schema. An integration layer is added to articulate the structure and the semantics, so that CLAL annotations can be combined with various structural schemas and the integrity of the initial document be preserved if the semantic annotations are to be removed. Technically, a semantically annotated text is associated with 3 (sets of) schemas corresponding to the three mentioned layers.

⁹E.g. In that case, Article 43 does not apply.

¹⁰www.lipn.univ-paris13.fr/~fl/CLAL.

¹¹www.oxygenxml.com

```

<xs:element name="PERMISSION">
  <xs:complexType >
    <xs:complexContent>
      <xs:extension base="leg:fragment_base_type">
        <xs:attribute name="IDENTIFIER" use="required" type="leg:fragment_identifier"/>
        <xs:attribute name="rel" type="leg:fragment_identifier_list"/>
        <xs:attribute name="except" type="leg:fragment_identifier_list"/>
        <xs:attribute name="bearer" use="required" type="xs:IDREFS"/>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>

```

Figure 2. Definition of the `PERMISSION` element in the semantic schema. The type `fragment_base_type` specifies the elements allowed in the content of a fragment (namely entities and sub-fragments). `IDENTIFIER` and `bearer` are required attributes. Most attributes may have lists of identifiers as values.

```

<leg:PERMISSION IDENTIFIER="012.001.003" except="012.001.002" bearer="p_CONT">When
requested by the <leg:PERSON ref="p_DS">data subject</leg:PERSON>, the information
may be provided orally, provided that the identity of the <leg:PERSON
ref="p_DS">data subject</leg:PERSON> is proven by other means.</leg:PERMISSION>

```

Figure 3. Example of a `PERMISSION` fragment annotation: the identifier `012.001.003` indicates the fragment localisation (Art 12, §1); the `bearer` is a person identified as `p_CONT`; the optional attribute `except` indicates that this permission is an exception to the previous fragment (`012.001.002`). The `bearer` is not mentioned in the fragment text content, which only includes two mentions of another person `p_DS` whose role is not specified.

The CLAL XML schema has been defined so as to control the annotation, thus to guide and ease the annotators' work: at a given point in the text, only the elements/attributes authorized for insertion should be accessible in the annotation tool and the mandatory ones should be indicated. This also allows a strict validation of the annotation syntax. Among others, there are constraints for the verification of the existence and uniqueness of identifiers, and the differentiation of identifier types. Fragment signatures indicate which attributes (roles and semantic relations) are allowed and required.

Figures 2 and 3 show how the element `PERMISSION` is defined in the CLAL schema and used for annotating a text fragment.

4.2. GDPR annotation

31 of the 99 articles of the French version of the GDRP have been annotated in CLAL by two annotators, following the methodology described in section 2. The annotated corpus is made available as a gold standard. Figure 4 gives an example of combined structural and semantic annotation. Table 2 gives an overview of the resulting annotation.

The stability of the annotation from one annotator to another and from one period to another for a given annotator is a good indicator of its quality. To measure the quality of the CLAL annotation of the GDPR, we performed an additional experiment consisting in having the two trained CLAL annotators who initially developed the annotation guidelines annotate a new part of the RGPD after a break of several months after the design of the annotation language and the annotation of the gold standard. Concretely, they annotated the same 8 additional articles on their own, and we measured the inter-annotator agreement by comparing their two annotations. They agreed on 90% of the fragments

```

</ARTICLE>
<ARTICLE IDENTIFIER="041">
<TI.ART>Article 41</TI.ART><STI.ART>Monitoring of approved codes of conduct</STI.ART>
  <PARAG IDENTIFIER="041.001"><NO.PARAG>1.</NO.PARAG>
    <ALINEA><leg:POWER IDENTIFIER="041.001.001" type="execution" bearer="le_SA">
      Without prejudice to the tasks and powers of the competent <leg:LEGAL_ENTITY
      ref="le_SA">supervisory authority</leg:LEGAL_ENTITY> ..., the monitoring of
      compliance with a code of conduct...may be carried out by a <leg:LEGAL_ENTITY
      ref="le_BOD">body</leg:LEGAL_ENTITY> which has an appropriate level of
      expertise ... and is accredited for that purpose by the competent
      <leg:LEGAL_ENTITY ref="le_SA">supervisory authority</leg:LEGAL_ENTITY>
    </leg:POWER></ALINEA>
  </PARAG>
  ...
  <PARAG IDENTIFIER="041.006"><NO.PARAG>6.</NO.PARAG>
    <ALINEA><leg:EXCEPTION IDENTIFIER="041.006.001" except="041">This Article
      shall not apply to processing carried out by public authorities and bodies.
    </leg:EXCEPTION></ALINEA>
  </PARAG>
</ARTICLE>

```

Figure 4. Example of annotation: Article 41 of the GDPR includes 6 different paragraphs, the first one being a `POWER` and the last one introducing an exception to the article itself. The structural and semantic layers (annotations resp. with `no` and `leg:` prefix) are intertwined.

Table 2. Distribution of the most frequent types of the XML elements and attributes in the annotated part of the GDPR. Mind that this distribution is probably not representative of the full regulation annotation.

Fragments (178)		Elements (709)				Attributes (890)	
		Entities mentions (515)		Sub-fragments (14)			
OBLIGATION	60	PERSON	290	EXCEPT	14	ref	515
LEGAL_PRECISION	31	LEGAL_ENTITY	217			bearer	136
POWER	23	CONCEPT	8			rel	34
QUALITY_ATTRIBUTION	17					except	18

(segmentation and typing), 94% of the roles and 60% of the semantic relations.¹² The good scores for fragments and roles give credibility to the proposed approach based on interpretation-neutrality and consensus. Unsurprisingly, the agreement is lower on `rel` attributes for which there is more annotation flexibility: the guidelines will probably have to be further specified. Annotators also reported that annotating these 8 articles took them 3 and 5 hours respectively, giving an average of 0.5 hour per article and approximately 50 hours for the entire GDPR.¹³ These first figures show that large portions of text can be annotated quickly, without the need for legal experts.¹⁴ This suggests that the proposed approach opens the way to a large-scale semantic search for legal texts.

¹²There is too few sub-fragments to give a reliable agreement measure for this category of elements. The figures for the annotation of entities are more difficult to interpret, due to the freedom left by the guidelines.

¹³This is certainly a very high estimate as the annotation task could be alleviated by using more user-friendly annotation tools and may be partly automated.

¹⁴The annotators should simply be familiar with the reading of legal documents.

4.3. Semantic search on GDPR

To show the benefit of semantic annotation for search, we designed a small experimental search engine based on an SQL-like querying language combining semantic and plain-text criteria and we tested our set of test questions.¹⁵ We illustrate the semantic search on 3 of these user questions in an informal way.

Q1. What are the rights of the data subject? The translation of the question in a semantic query is straightforward (*Which fragments are annotated as RIGHT which bearer role is filled by the identifier of the data subject?*¹⁶). It returns 6 fragments.¹⁷ On the other side, a search for sentences that contain the strings "right" and "data subject" provides the same fragments plus 17 additional ones which are noisy answers, except for one. The semantic search therefore appears to be more precise than full text search, which is an advantage for legal practitioners who have to browse large quantities of legal sources. It is always possible to broaden the search with a full text search to ensure one is not missing any information, at the cost of an important additional effort to analyse the results.

Q2. What are the obligations of a data controller? The question seems to translate directly into a semantic query: *Which fragments are annotated as OBLIGATION with the bearer role filled by the identifier of the data controller?* However, rights also express obligations in some cases, which leads to a second query: *Which fragments are annotated as RIGHT with a target role filled by the identifier of the data controller?*¹⁸ This double query returns 26 obligations and 6 rights. In comparison, a plain text search (*Which sentences contain both "data controller" and "obligation—obligatory" keywords?*) gives 13 fragments among which only one right and one obligation are relevant. This is due to the many ways to express obligations in the text and the absence of constraint on the controller's role in plain text search.

Q3. What are the obligations of the controller in case of data breach? This question shows how semantic categories and strings can be combined in the same query and how fragment relationships, especially for exceptions, can be exploited. *Q3* is similar to *Q2* with an additional condition (*in case of data breach*) that does not translate into a semantic restriction because the term "data breach" is not marked-up as a concept. However, *Q3* can be translated into a hybrid query combining semantic criteria and keywords (*Which fragments are annotated as OBLIGATION with the bearer role filled by the data controller identifier and containing the string "data breach"?*). This query directly returns a single fragment of Art. 34 §1¹⁹ but this fragment happens to be related to two exceptions that are relevant for answering *Q3*. The first exception appears in the close context of the OBLIGATION statement (in the same article, two paragraphs apart) but the second one (in Article 23) would be difficult to spot for the user if it were not explicitly marked in the annotation.

¹⁵Note that semantic criteria can only be matched with the annotated part of the GDPR.

¹⁶SELECT fragments WHERE name ~ "leg:RIGHT", bearer ~ "p_DS". The formalization of the two other queries is omitted due to space limitations.

¹⁷Such as the right to obtain from the controller, without undue delay the rectification of inaccurate personal data concerning him or her.

¹⁸The rights of third parties towards the controller are as many obligations imposed on the latter.

¹⁹When the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay.

5. Previous works

It has long been understood that *enriching legal sources with metadata* is essential to make them more accessible. XML vocabularies [10] have been designed to account for document information, the structure of the legal sources, and cross-references (e.g. case law, citations, modification). Much effort has also been devoted to the semantic enrichment of legal documents to facilitate legal reasoning. This metadata ranges from tags associated to key textual elements (e.g. actors, dates) to rule annotations [11], possibly associating formal descriptions to textual fragments and allowing for associating multiple interpretations. Nevertheless, bridging between the natural language of the legal sources and the logical or rule formalizations remains an open challenge despite efforts to define a methodology for deriving formal rules from texts, relying on controlled or semi-formal languages, or providing interfaces to support human formalization [12,13].

The modeling of the semantic information has also been the subject of numerous works, especially for the design of legal ontologies, core legal ontologies [14] or domain ontologies, some of which focusing on privacy and the specific issues of the GDPR (e.g. ODRL [2], PrivOnto [15], ProNTO [3]). Another approach aims to develop machine-readable languages for formally representing legal rules, such as LegalRuleML [16], but with few examples from source texts.

A sound methodology has been developed over time for manual annotation of corpora [9]. The issues have mainly concerned the annotation format, the tagset choice, the underlying theory, the training of annotators, the complexity/cost of an annotation task, and the quality of the annotation (inter- and intra-annotator agreement).

It is well known that legal information retrieval has to meet specific requirements, such as the size and interdependence of legal sources or the needs for retrieval completeness [17]. These requirements favor semantic or hybrid approaches, which the present work shows the potential of.

6. Conclusion and future work

This paper advocates a new approach of legal text mining relying on semantic technologies, which represents an alternative and middle ground to the traditional statistical-based information retrieval methods – which are applicable on a large scale and have a good recall but low precision – and those aiming at formalizing the content of rules – which are quite ambitious but difficult to implement due to the plurality of interpretations and the complexity of translating natural language rules into logic.

Considering that the text (be it legislation, case law, decisions, contracts...) is the reference for any legal work, we propose to keep the text at the center of attention while enriching it with a semantic annotation layer, thus enabling access using semantic search services. Our approach of regulatory texts is based on a coarse-grained, interpretation-neutral annotation that nevertheless semantically enriches the text. The approach is illustrated by a proof of concept experiment of annotating the French version of the GDPR and searching over the annotations. The paper presents the methodology followed for the design of the annotation language, the language itself as well as its implementation in XML, the annotated GDPR that might serve as a gold standard for training automatic annotation tools, and finally queries that illustrate the benefit of annotation for semantic

search. The resources provided also include an annotation guide that allows the experiment to be extended to other texts. In the end, this GDPR experiment shows the added value of annotation for legal text mining and exploration.

Further work is required. To evaluate the robustness of the proposed language and annotation approach, the experiment must be extended to new annotators, to other versions of the GDPR, and to other legal texts. While it may be necessary to revise the annotation language, the gold standard, and the annotation guide, the balance between annotation granularity, cost, and reliability will have to be maintained. It would be interesting to use the manually annotated corpora as training data in a machine learning study. Finally, with regard to users, we must refine the requirements analysis and provide easy-to-use tools for querying the GDPR and annotating new texts.

References

- [1] Stamper R. LEGOL: Modelling Legal Rules by Computer. *Computer Science and Law*. 1980:45–71.
- [2] Group WOC. Open Digital Rights Language: Vocabulary & Expression 2.2. W3C Recommendation; 2018. Available from: www.w3.org/TR/odrl-vocab/.
- [3] Palmirani M, Martoni M, Rossi A, Bartolini C, Robaldo L. PrOnto: Privacy Ontology for Legal Reasoning. In: K6 A, Francesconi E, editors. *Electronic Government and the Information Systems Perspective*. Cham: Springer International Publishing; 2018. p. 139–152.
- [4] Palmirani M, Governatori G. Modelling Legal Knowledge for GDPR Compliance Checking. In: Palmirani M, editor. *Proc. of the 31st JURIX*. vol. 313 of *Frontiers in A.I. and Applications*. IOS Press; 2018. p. 101–110.
- [5] Libal T. A Meta-level Annotation Language for Legal Texts and Argumentation. In: Dastani M, Dong H, van der Torre L, editors. *Logic and Argumentation. CLAR 2020*. Cham: Springer; 2020. p. 131–150.
- [6] Robaldo L, Bartolini C, Lenzini G. The DAPRECO Knowledge Base: Representing the GDPR in LegalRuleML. In: Calzolari N, et al., editors. *Proc. of the 12th LREC*. Marseille: ELRA; 2020. p. 5688–5697.
- [7] Maxwell KT, Schafer B. Concept and Context in Legal Information Retrieval. In: *Oric. of the 21st JURIX*. IOS Press; 2008. p. 63–72.
- [8] Bartolini C, Muthuri R. Reconciling Data Protection Rights and Obligations: An Ontology of the Forthcoming EU Regulation. In: *Language and Semantic Technology for Legal Domain*; 2015. p. 8.
- [9] Fort K. *Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects*. Wiley-ISTE; 2016.
- [10] Barabucci G, Cervone L, Palmirani M, Peroni S, Vitali F. Multi-layer Markup and Ontological Structures in Akoma Ntoso. In: Casanovas P, Pagallo U, Sartor G, Ajani G, editors. *AI Approaches to the Complexity of Legal Systems*. Berlin, Heidelberg: Springer; 2010. p. 133–149.
- [11] Wyner AZ, Peters W. On Rule Extraction from Regulations. In: Atkinson K, editor. *Proc. of the 24th JURIX*. vol. 235 of *Frontiers in A.I. and Applications*. Vienna: IOS Press; 2011. p. 113–122.
- [12] Nazarenko A, Lévy F, Wyner A. Towards a Methodology for Formalizing Legal Texts in LegalRuleML. In: Bex F, Villata S, editors. *Proc. of the 29th JURIX*. vol. 294 of *Frontiers in A.I. and Applications*. IOS Press; 2016. p. 149–154.
- [13] Libal T, Steen A. NAI: The Normative Reasoner. In: *Proc. of the 7th ICAIL*. New York: ACM; 2019. p. 262–263.
- [14] Hoekstra R, Breuker J, Di Bello M, Boer A. The LKIF Core ontology of basic legal concepts. *International Journal of High Performance Computing Applications (IJHPCA)*. 2007 01:43–63.
- [15] Oltramari A, Piraviperumal D, Schaub F, Wilson S, Cherivirala S, Norton TB, et al. PrivOnto: A semantic framework for the analysis of privacy policies. *Semantic Web*. 2018;9:185–203.
- [16] Athan T, Governatori G, Palmirani M, Paschke A, Wyner AZ. LegalRuleML: Design Principles and Foundations. In: Faber W, Paschke A, editors. *Reasoning Web. Web Logic Rules - 11th International Summer School, Tutorial Lectures*. Berlin: Springer; 2015. p. 151–188.
- [17] Van Opijnen M, Santos C. On the Concept of Relevance in Legal Information Retrieval. *Artif Intell Law*. 2017 Mar;25(1):65–87.