Hybrid AI to Enhance Legal Drafting with LEOS

Summary technical report

Monica Palmirani, Michele Corazza, Generoso Longo, and Salvatore Sapienza
Department of Legal Studies, ALMA-AI, University of Bologna
December, 2025

HyperModeLex
This project is funded by European Research Council
GA n° 101055185
Funded by the European Union
European Research Council

EUROPEAN COMMISSION

Directorate-General for DIGIT
Directorate A — Digital Transformation
Unit A.3 — HR & Legislative Solutions

Contact: Willy Van Puymbroeck

E-mail: Willy.Van-Puymbroeck@ext.ec.europa,eu

European Commission
B-1049 Brussels

EUROPEAN COMMISSION

Hybrid AI to Enhance Legal Drafting with LEOS

Summary technical report

Directorate-General for DIGIT

2025

Manuscript completed in December 2025

First edition

Luxembourg: Publications Office of the European Union, 2025

Print ISBN 978-9268-33290-0 doi:10.2799/4519264 NO-01-25-012-EN-C
PDF ISBN 978-9268-33289-4 doi:10.2799/1298773 NO-01-25-012-EN-N

Contents

Abbreviations

Acronym	Definition
AKN	Akoma Ntoso standard of OASIS
AKN4EU	Akoma Ntoso for European Legislation a profile of Akoma Ntoso standard
API	Application Protocol Interface
CSV	Comma Separated Values
DEFINA	Legal Definition Assistant
ERC	European Research Council
EUROVOC	EU's multilingual and multidisciplinary thesaurus
LEOS	Legislative Editing Open Source
LLM	Large Language Models
NLP	Natural Language Processing
PoC	Proof of Concept
RDF	Resource Description Framework
RECONTA	Recital Connector to Articles
REFERA	Reference Embedding Retrieval Assistant
SORTIS	Study On regulatory reporting standards
SQL	Structured Query Language
TREND	Template of Reporting Requirement Engine for Normative Drafting
UX	User Experience

Executive summary

The report presents the outcome of a study funded by the European Commission. The study sought to improve legislative drafting by developing AI-based microservices that assist lawyers and policy developers to uphold the rule of law, harmonise the language of legislation, retrieve relevant accurate legal references, enhance drafting clarity, and strengthen semantic connections within legal texts. LEOS, an open-source web editor developed by the European Commission, provides the platform for integrating these AI enhancements.

This document describes in detail four use cases in which hybrid AI combined with Akoma Ntoso XML (AKN) is applied to enhance LEOS. The work is done in cooperation with the European Commission's DG Informatics (Unit A3) and is supported by the ERC HyperModeLex - Hyperdimensional Modelling of the Legal System in Digital Society - and Erasmus+ Jean Monnet LEDS4XAIL - Legal Design and Data Science for Explicable AI in Legal Domain -projects.

The first use case, REFERA (Reference Embedding Retrieval Assistant), aims to reduce manual effort and citation errors, by retrieving the most relevant normative references even from incomplete input. The second use case, DEFINA (Legal Definition Assistant), identifies fitting legal definitions to ensure consistency and reuse of legal terminology. The third use case, RECONTA (Recital Connector to Articles), examines the correlation between recitals and legislative provisions, and suggests a way to record these relations in a machine-readable format through AKN-XML metadata. The fourth use case, TREND (Template of Reporting Requirement Engine for Normative Drafting), proposes, based on legal taxonomies and ontologies, templates for reporting obligations to simplify and harmonise the drafting of regulatory requirements.

Together, these use cases convincingly demonstrate how AI and semantic annotations in AKN-XML can be used to improve the quality, transparency, and searchability of legal texts. The hybrid AI methodology combines symbolic and statistical techniques, uses embeddings, retrieval-augmented generation, and large language models. The integration of the use cases in LEOS is through an easy-to-use interface with emphasis on 'explainability' of AI-generated output. Finally, AKN serialisation is used for the structured representation of legal documents, normative references, and lifecycle metadata such as entry into force and repeal, creating a robust foundation for interoperability and explainable AI in legislative drafting.

1. Introduction

The report presents the main outcomes of a study on 'The use of AI in legal drafting.' The study was funded by the European Commission[1] and executed in 2025 by the University of Bologna. Its main aims were to carry-out several proofs of concept of using hybrid AI, in combination with Akoma Ntoso XML (AKN-XML), in drafting EU legislation, and their integration in LEOS using micro-services. The study should moreover

- Address the full complexity of legal drafting.
- Contribute to building trust in the use of AI, and
- Illustrate the importance and the potential of the use of AI.

The study work has been conducted in cooperation with the DG-Informatics, European Commission Unit A3 - HR and Legislative Solutions. It is also part of the ERC "Hyperdimensional Modelling of the Legal System in Digital Society" (HyperModeLex) Grant agreement ID: 101055185[2] and of the Erasmus+ Jean Monnet "Legal Design and Data Science for Explicable AI in Legal Domain" (LEDS 4 XAIL) Grant No. 101085576[3].

1.1. Rationale

Legal drafting is a crucial task in the legislative process in any deliberative assembly. The goals of the study were to develop AI-based tools and micro-services:

(i) To support political decision-making.

(ii) To guarantee the rule of law.

(iii) To ensure adherence to the theory of law principles.

(iv) To assist the harmonisation of language used in specific parts of the legislation (e.g., definitions), thereby avoiding inconsistencies.

(v) To correctly retrieve relevant parts of the legislative corpora (e.g., normative references), thereby avoiding errors.

---

1 DIGIT C.1.004, Purchase order 2024-35337 Use of AI in Legal drafting – Ares (2024)8448887.

2 In this project the AI tools used for embedding AKN-XML structure were developed

3 In this project the user interface to explain the results of using the AI tools was developed

(vi) To facilitate the application of drafting rules, thereby improving quality and clarity.

(vii) To assist the formulation of legislative clauses with all the information in simple language (e.g., temporal parameters in the reporting requirement), and

(viii) To support the detection of the semantic connection between different parts of the legislative document (e.g., recitals connected with related articles).

In the last 15 years, many specialised editors have been developed [1],[2],[3],[4],[24], [25], [26] to achieve these goals. These make use of Natural Language Processing (NLP), Semantic Web approaches, symbolic AI based on rules [5], or XML based methods.

LEOS [6] is a most promising web editor for legal drafting. It has been developed by the European Commission to support its legal drafting, also with the aim of addressing the needs of the Member States. LEOS is an open-source web editor specific to legal drafting; its interface is developed using Angular. Its purpose is to be used in every step of law-making.

1.2. Scope

The aims of this study are (i) to design a framework architecture to enhance LEOS with add-ons that use AI, (ii) to improve the quality of the legal content, (iii) to help legal drafters in their daily work, and (iv) in general, to assist the law-making process.

The four add-ons provide the following features [7], [8]:

1. Retrieve the most relevant normative references considering the subject of the bill, using incomplete data (REFERA- REFerence Embedding Retrieval Assistant). The goal is to reduce manual error-prone work, and to avoid repetitions in legislative citations [9].
2. Retrieve the most relevant definition considering the subject of the bill, using a term (DEFINA- legal DEFINition Assistant). The goal is to maximise the reuse of similar legal concepts, limit the introduction of new terms, and avoid provisions that could lead to interpretation problems [10].
3. Correlate during legislative drafting recitals and normative text using terms identified as drivers. The goal is to track this correlation with AKN-XML in a machine-readable manner using metadata (RECONTA- REcital CONnector to Articles).
4. Suggest templates of reporting requirements according to a taxonomy defined in this report and the Reporting Requirement Metadata Vocabulary (RRMV)

ontology[4] and propose a semantic annotation in AKN4EU (TREND - Template of Reporting requirement Engine for Normative Drafting).

The work builds on the results of previous studies on 'Drafting in the era of digitisation and AI' [7], on the 'SORTIS' study on regulatory reporting standards[5] and on the work done on regulatory reporting requirements developed within the Proof of Concept (PoC) on 'Context Aware Legal Verification'[6]. The former developed the RRMV ontology, whilst the latter explored the integration in LEOS, combining AKN and Resource Description Framework (RDF), thereby demonstrating the feasibility of automatically capturing reporting requirements within LEOS.

All four use-cases illustrate the positive impact not only to attain better regulation, but also to increase transparency and searchability of legal knowledge using authoritative semantic annotation in AKN4EU (see Figure 1).

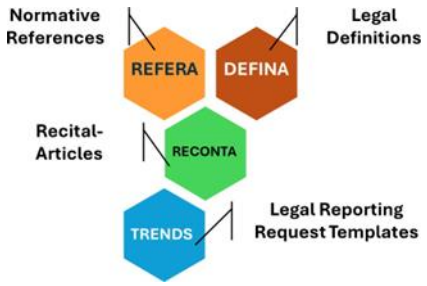


Figure 1: Four use cases using AI integrated in LEOS.

1.3. Methodology of hybrid AI

The methodology is based on hybrid AI [11], [12], and uses multiple techniques. No new text (e.g., using LLM or generative AI) is generated. We aim to suggest pertinent, contextual information. Suggestions are, based on extracting relevant legal knowledge from the legal corpora, using a similarity index according to the bill parameters [27].

---

4 See https://interoperable-europe.ec.europa.eu/collection/digital-ready-policymaking/solution/rrmv

5 See https://interoperable-europe.ec.europa.eu/collection/digital-ready-policymaking/news/streamlining-regulatory-reporting-sortis-project-results

6 See https://interoperable-europe.ec.europa.eu/collection/justice-law-and-security/solution/leos-open-source-software-editing-legislation/document/proof-concept-context-aware-legal-verification

We first use the EUROVOC classification and other contextual information provided by users during drafting (e.g., type of provision). Second, the approach considers the temporal validity of normative provisions, e.g., excluding those that are repealed, or suggesting the appropriate version of the consolidated text according to the view date given by the end-user. Third, we resolve the normative references to include in the model of indexing the text cited in a recursive way (only the first level), allowing to collect more information, especially when the description is imperfect in the text and consists only of normative citations toother provisions (e.g., “For the purposes of this Directive, the definitions laid down in Article 2 of Directive 2000/60/EC shall apply”). Fourth, context is important for providing relevant output of the suggestion. E.g., a definition depends on the bill. To illustrate there are many definitions of ‘accuracy;’ which definition to use depends on the document at hand. Fifth, we use embedding, Retrieval Augmented Generation (RAG), and Large Language Models (LLM) to support the end-user in drafting. Sixth, the user interface is a fundamental pillar to guarantee usability, transparency, and ‘explainability’ of the AI behaviour and output [13]. Finally, we use AKN [14] serialisation to structure legal documents, normative references, metadata of the lifecycle of the document, the date of entry into force/into operation/of repeal.

To learn more on the use of NLP, ‘law first’ and the new concept of Hybrid AI you can consult [15], [16],[17],[18], [19], [11],[20].

2. Use Cases REFERA and DEFINA

2.1. Introduction and motivation

2.1.1. REFERA

Normative references are widely used in legislative documents to cite other documents or provisions which are relevant within the context of the norm. The use of normative references, however, can lead to errors, if they break the relation between the text and the cited document. For this reason, we propose a system to suggest the appropriate references in case of an incorrect or partial text for a given reference. In particular, the system leverages temporal information, as well as metadata about the documents, to suggest references which (i) are valid, (ii) into force, (iii) which share the same number as the one provided in the query, and (iv) refer to documents that have a similar theme to the one specified by the user. As an example, if the user provides a partial reference of the form ``Regulation 406'', the system will return documents which are in force and share the document number 406. In this case, a result would be Regulation 406/2010. The user is then able to inspect the title and the document, and if relevant to incorporate it in the current draft. Furthermore, the tool can help in using the proper format for citations, since the format changed over the years. Regulations before 1968 appear as number/yy/EEC; those after 1968 are represented by number/yy and after 2009 they are written as yyyy/number.

2.1.2. DEFINA

Legal definitions are a sensitive part of the law as they define new legal concepts, introduce new terminologies, express equivalences between different definitions, and refer to exceptions in specific cases. In EU legislation, there is usually a clear article called “Definitions,” but sometimes technical definitions are given in the last part of the act or in the annexes. Additionally, definitions are sometimes organised in a long list of points and may be related to each other.

Definitions are composed of three main parts: definiens (term); definiendum (description); and legal concept (abstract class of concept). The use of the same term in multiple definitions is frequent, and a term may have a completely different meaning in different domains (e.g., the definition of pollution depends on the domain, like water, energy, industry, etc.). For this reason, the tool calculates the similarity of a given term (which can also be composed of multiple words) with the existing, valid, and updated (present in consolidated versions of documents) definitions in the legal corpus, using the similarity index as a criterion.

2.2. Dataset for REFERA and DEFINA

The dataset used consists of 10 years of European legislation (2010-2021), i.e., about 15.000 regulations and directives. It was provided by the European Publication Office in Formex 3.0 XML format. We converted all documents in AKN, and using an NLP approach we annotated the definitions and the normative references. All documents include normative references, and the dataset includes 899 documents with definitions. For definitions, we considered only explicit provisions typically titled “Definitions” or where a regular pattern can sufficiently reliably identify the relationship between a term (definiens) and description (definiendum). The definitions that include normative references are managed by navigating the link to include the complete information (e.g., ‘personal data’ means personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679).

2.3. Architecture of REFERA and DEFINA

The architecture of our system (see Figure 2) includes an eXist XML database, which includes all the AKN XML documents and allows the user to query its contents using XPath expressions. Furthermore, it comprises a relational SQL database which contains a correspondence between documents and their relevant EUROVOC terms, as these are found on the CELLAR knowledge graph. Each EUROVOC is also associated with its vector representation, which is used by the reference suggestion module.

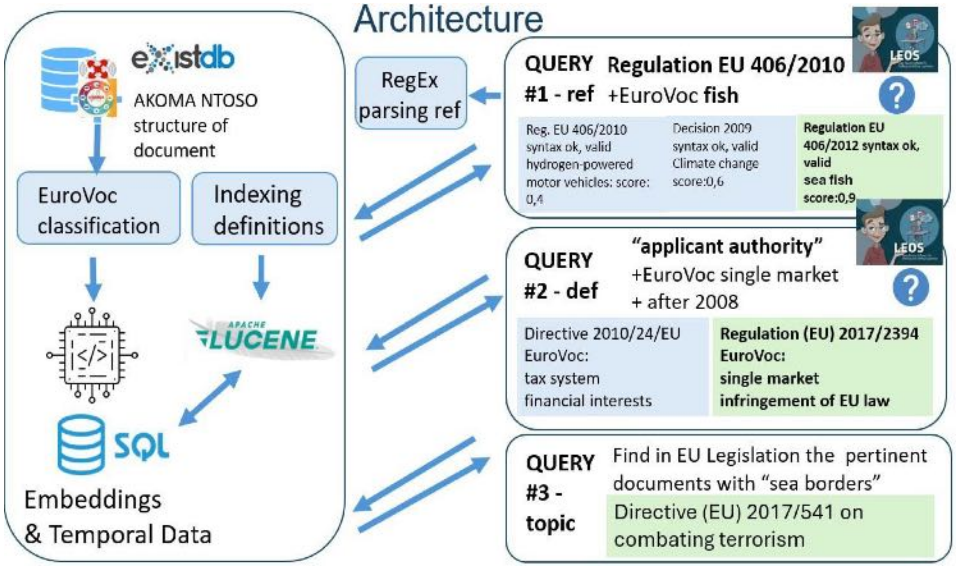


Figure 2: The architecture of the enhanced LEOS.

Finally, for the search by topic functionality, a list of vectors representing the title and each of the articles of each document is memorised as an embedding database. This permits the activation of the comparison between the topics selected by the user and the vectors of the title and articles using similarity methods.

2.3.1. References suggestion with REFERA

The input of the reference suggestion provided by the user is composed of a reference and EUROVOC terms. From this input the module for normative references extracts the document number and year. These values are compared with the database, and the system retrieves first the documents that have the same year and number as those provided by the user. If no document is found, a less stringent criterion is applied, where all documents that contain the numbers in the query are returned. A final set of references is selected using the Levenshtein distance [21], a metric used for the comparison of strings. It measures the number of additions, deletions and substitutions that separate two strings. For both document number and year, all documents that have a Levenshtein distance of at most 1 are included, meaning that they contain at most one substitution, deletion, or addition of characters (see the Figure 3).

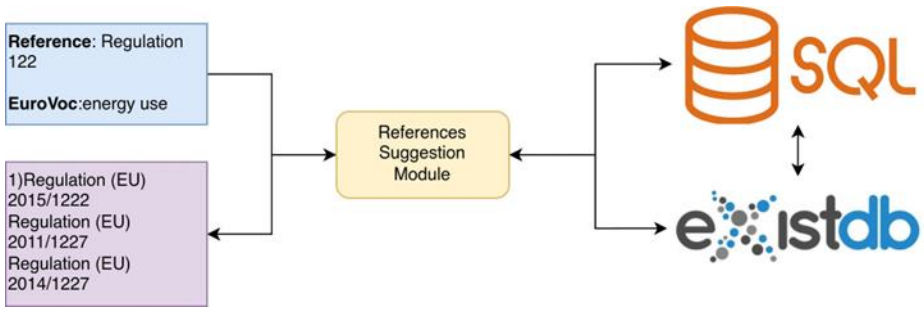


Figure 3: The reference suggestion module.

The refinement of all the selected documents, then, is derived from the EUROVOC terms the user provided. These EUROVOC terms are compared with the ones associated with each document included in the selected references. Each EUROVOC, both those which are provided by the user and those associated with documents, is represented as the sum of the fastText embeddings [22] of the words composing it. The vector representations of two lists of EUROVOC are then averaged, obtaining a vector representing the user inputs and vector representing each document in the database. Finally, the list of proposed references is sorted by decreasing cosine similarity between each document EUROVOC and the user-provided terms.

2.3.2. Definition suggestion with DEFINA

The definition suggestion module (see Figure 4) attempts to leverage information about prior existing definitions to suggest relevant definitions to the user. First, the system selects a set of relevant definitions from the user query. This is done using the Lucene information retrieval tool, which is integrated in the eXist database. This tool allows the user to query definitions and to provide a list of definitions sorted by relevance. Internally, the system uses an approach based on term frequency–inverse document frequency (Tf-idf), meaning that it retrieves definitions that have a nonzero overlap with the user query. Then, EUROVOC terms in the user query are compared with the ones in the selected definitions, to provide the user with the most relevant definitions in the desired context.

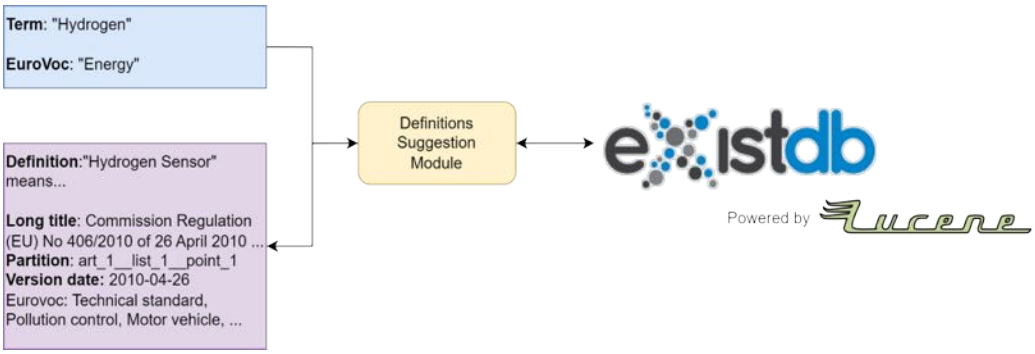


Figure 4: The definition suggestion module.

2.3.3. Results of REFERA and DEFINA

To assess the performance of the REFERA module, we selected several European legislative documents that were related to the following EUROVOC terms:

- Energy use.
- Energy efficiency; and
- Energy consumption.

Partial references for these documents were constructed by removing a part of the complete reference (e.g., Regulation (EU) 2017/1938 →Regulation 19/2017). Then, we performed three tests for each reference, i.e., by providing the model with one, two or three EUROVOC terms and the incomplete reference. The order of the terms is the order as presented above, and we used top 1,5 and 10 accuracy as metrics. The results of this evaluation, presented in Table 1, show that the model can suggest the correct reference most of the time within the top 5 suggestions, however only when provided with sufficient EUROVOC terms (at least 2).

Number of EUROVOC	Top 1 Accuracy	Top 5 Accuracy	Top 10 Accuracy
1	0.20	0.40	0.80
2	0.30	0.70	1.00
3	0.30	0.70	1.00

Table 1: Evaluation for the normative reference suggestions.

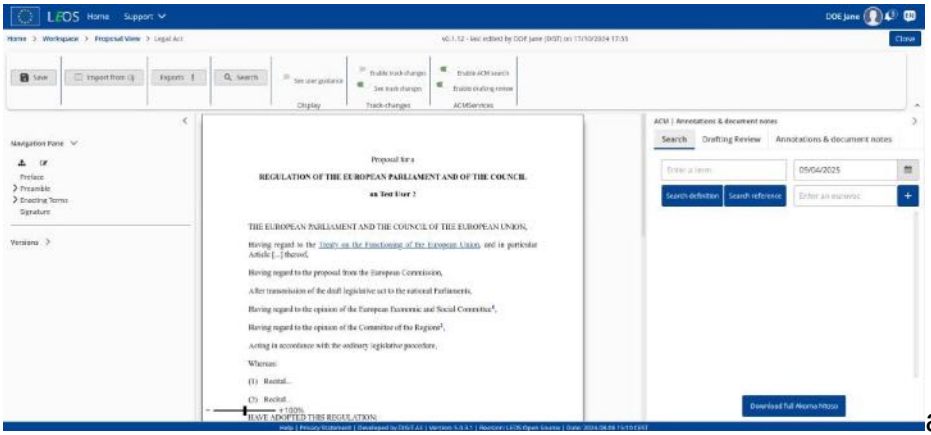
To evaluate the DEFINA module, a list of terms that are not defined in EU legislative documents, along with relevant EUROVOC terms was produced. Both the terms and the EUROVOC terms are used to retrieve relevant definitions. The definitions that are retrieved are assessed by legal domain experts (junior researchers) and marked as either “helpful” or “not helpful” for the given term. We then computed the top 5 accuracy, both including instances where the system does not return any definition and excluding those cases. The results, presented in Table 2, show that DEFINA can retrieve relevant definitions most of the time, especially if instances where the system returns a blank output are excluded.

Top 5 Accuracy with “blank” outputs	Top 5 Accuracy without “blank” outputs
0.62	0.72

Table 2: Top 5 accuracy values for definitions and document clustering.

2.3.4. Integration in LEOS of REFERA and DEFINA

We have integrated the micro-services in LEOS via an API. At interface level we added a panel in the right space of LEOS that shows boxes where to interact with the API. The end user enters the EUROVOC term, or the definition term or the partial normative citation to be investigated. The results are shown in Figure 5 and Figure 6 below:



an

Figure 5: REFERA and DEFINA integration in LEOS with user interface.

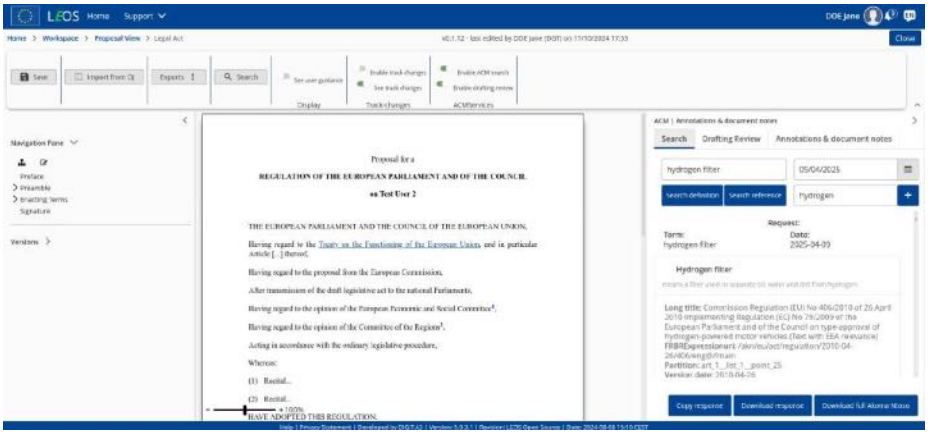


Figure 6: REFERA and DEFINA results in LEOS with user interface.

3. Use Case RECONTA

3.1. Introduction and motivation

Within European legislation, one of the key responsibilities of the drafter is to identify and establish the relationships between recitals and normative provisions (such as articles). This relation is fundamental to assist interpretation and enables a more comprehensive and more accurate understanding of the law. While recitals do not possess binding legal force, they constitute a crucial interpretative component, as they outline the rationale, principles, and objectives that underpin the legislative text. Frequently, they also reflect the content of soft law instruments (e.g., preliminary white papers) arising from negotiations among stakeholders, the positions of Member States, and the broader policy direction of the European Commission and other EU bodies. To support this interpretative process, we aim to automatically link recitals to the corresponding normative provisions in LEOS during the drafting phase and make these associations machine-readable and natively integrated within an AKN-XML structure. An additional objective is to establish correlations between recitals and enabling terms, that function as semantic anchors guiding the identification of relevant connections.

The goal of the RECONTA module is to produce a classifier that can determine which recitals are connected to a given article. For this reason, the approach taken is to use a recent transformer module, ModernBERT-base [23], and fine-tune it with a binary classification task, to determine whether a recital and an article are related or not. These models consider the semantics of words in the context in which they are found and are the basis of modern AI and NLP achievements. The specific model that we selected was chosen to allow for a longer context window (8k tokens) and because of its low computational cost.

3.2. Dataset of RECONTA

To train the module, we obtained JSON files available from the https://streamlex.eu website. We used thirteen regulations concerning he European Digital Strategy and Cybersecurity with the correspondence between recitals and articles (see Table 3). These were converted into a single CSV file where each row contains the number and text of an article, the number and text of a recital, and whether they are related. From this dataset, we constructed a training, a validation and a test set which were split to avoid any form of contamination between the training and validation/test sets. To achieve this, we selected two document types that are the least homogeneous compared to the training set. We used the Cyber Resilience documents (Digital Operational Resilience Act[7], Cyber Resilience Act[8]) as the validation set and Cyber Security documents (Cyber Security Act[9]) as a test set. The other documents were used to train the model. The total figures are as follows:

- Training set has 628 articles, 988 recitals.
- Test set has 115 articles, 235 recitals; and
- Validation set has 135 articles, 231 recitals.

Name of Legal Document	Number of Articles	Number of Recitals	Role
Artificial Intelligence Act	113	180	Training
Cyber Resilience Act	71	130	Validation
Cybersecurity Act	69	110	Test
Data Act	50	119	Training
Data GovernanceAct	38	63	Training
Digital MarketsAct	54	109	Training
Digital Operational Resilience Act	64	106	Validation
Digital ServicesAct	93	156	Training
Directive onMeasures for a High Common Level of Cybersecurity Across the Union	46	144	Test
European HealthData Space Regulation	105	115	Training
European MediaFreedom Act	29	78	Training

7 See https://eur-lex.europa.eu/eli/reg/2022/2554/oj/eng

8 See https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng

9 See https://eur-lex.europa.eu/eli/reg/2019/881/oj/eng

Name of Legal Document	Number of Articles	Number of Recitals	Role
General DataProtection Regulation	99	173	Training
Product LiabilityDirective	24	64	Training

Table 3: Table with acts examined.

3.3 Model of RECONTA

To train our binary classifier, we opted to use a concatenated representation of both the article and the recital. Formally, given an article A and a recital R:

$$A = (a1, \ldots, aj) \quad (3.3.1)$$

$$R = (r1, \ldots, rk) \quad (3.3.2)$$

Where a1, . . . , aj and r1, . . . , rk represent each token of the article and recital, respectively. We constructed the following representation for our inputs:

$$X = c \oplus A \oplus s \oplus R \oplus s \quad (3.3.3)$$

Where s denotes the [SEP] token, which is normally used to separate sentences in BERT or similar models, while c denotes the [CLS] token, normally used at the beginning of the input and when performing sequence classification. $\oplus$ is used to denote concatenation. Given a model M, the training is then performed using a cross-entropy loss as follows:

$$\mathcal{L}(X, Y) = -\log \frac{\exp(M(X)_{y_n})}{\sum_{i=1}^{C} \exp(M(X)_i)} y_c \quad (3.3.4)$$

Where $M(X)_i$ is the prediction obtained using model M on input X for the i-th class, and C is the total number of classes (two in our case). Y represents the gold standard labels obtained from https://streamlex.eu. A visualisation of the model behavior is shown in Figure 7.

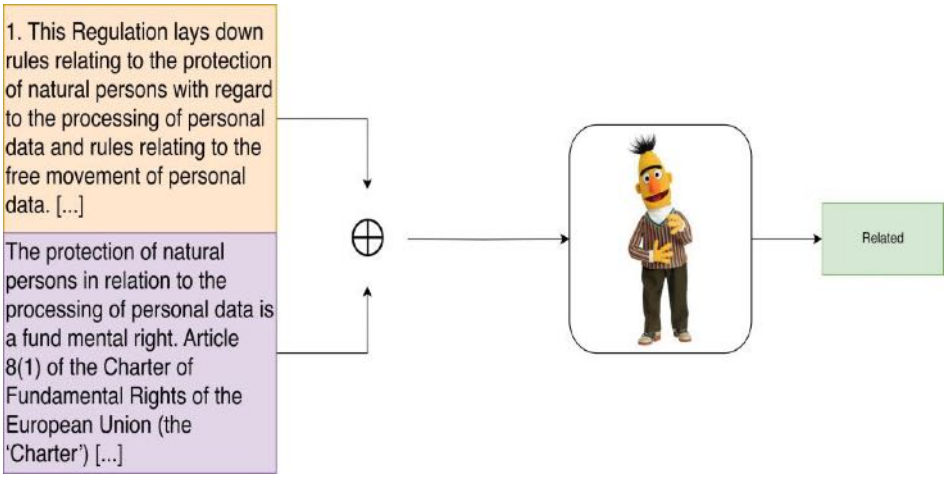


Figure 7: The RECONTA module.

3.3. Results of RECONTA

Since our dataset is heavily imbalanced (most of the pairs belong to the "nonrelated" class), we constructed a minibatch which is used during training to contain half samples annotated as "related," and the other half annotated as "non-related." This was achieved by selecting one "non-related" sample for each "related" sample at random.

The precision, recall and F1 scores[10] for the two classes, as well as for their macro-averaged values are shown in Table 4. Due to the high level of imbalance between the two classes, the performance on the minority ("related") class is markedly inferior to that on the majority class. We are currently training a larger model (ModernBERT-large) to improve the overall performance of the system.

Class	Precision	Recall	F1 Score
Non Related	0.99	0.99	0.99
Related	0.51	0.45	0.48
Macro Averaged	0.75	0.72	0.73

Table 4: Performance level on the test set.

3.4. Representation of the article-recitals graph

Using the CSV dataset produced in the previous step, we have used KNIME to calculate the similarity index between articles and recitals that have a value >0,70 and to represent them with a parallel graph (see the Figure 8). In this graph, each

[10] F1 score refers to harmonic mean of precision and recall (in classification metrics)

article relates to the relevant recital using a colored line. Many articles have an index of similarity very high (>0,90).

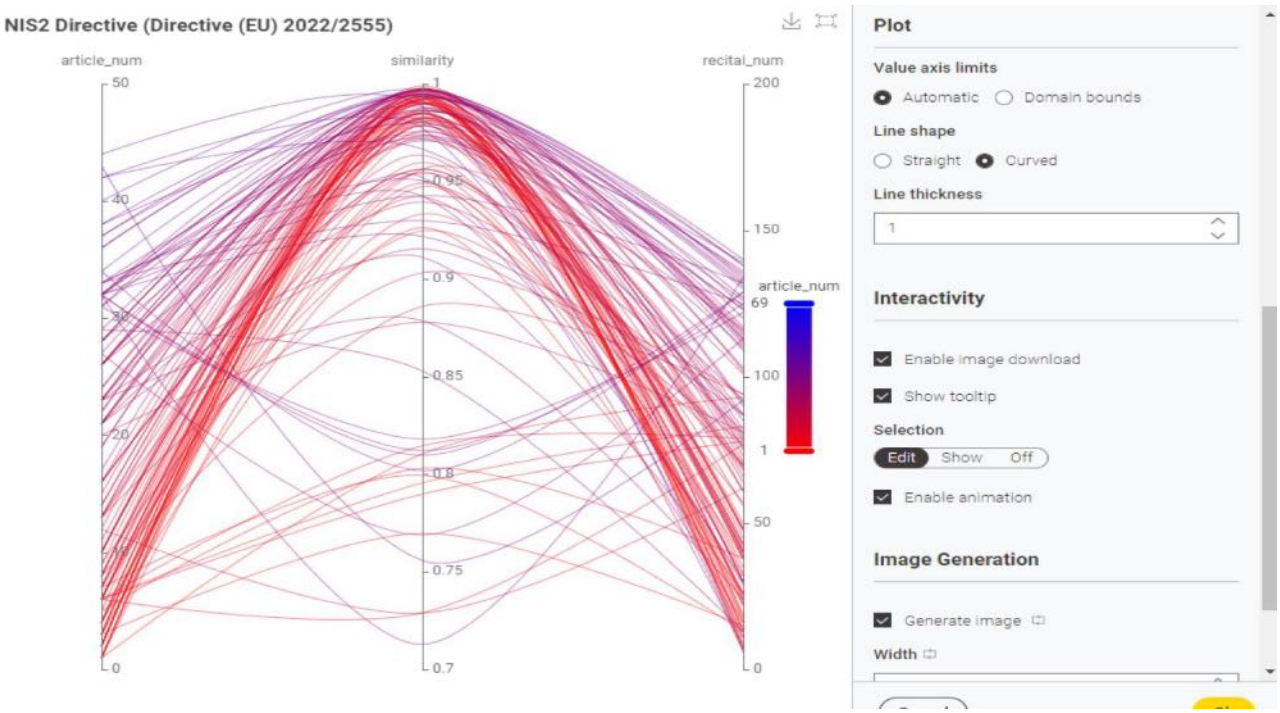


Figure 8: The RECONTA Parallel Graph Representation.

The next graph presents the same results using a sunburst graph (see the Figure 9). Clicking on an article, it is possible to see the recitals connected and the percentage of similarity.

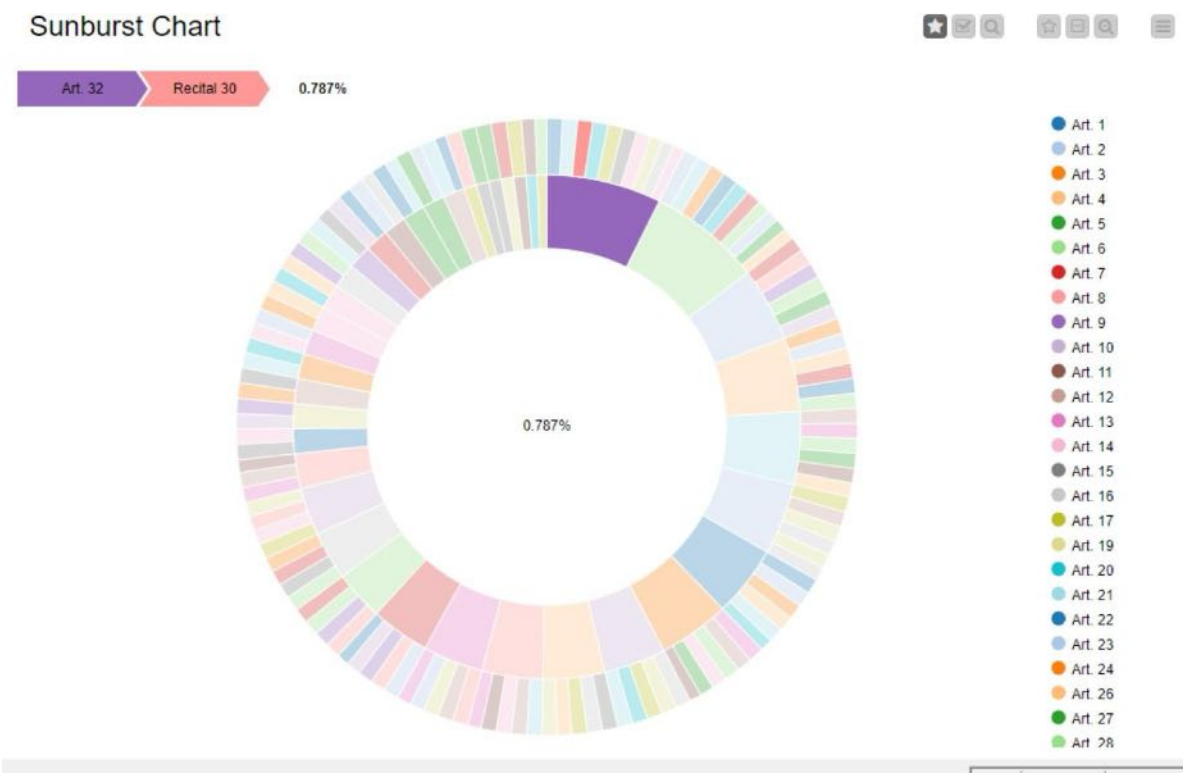


Figure 9: The RECONTA Sunburst Graph.

The scatter plot graph permits showing the correspondence between articles and recitals (see Figure 10). Each article in the X axis relates to the recitals in the Y axis.

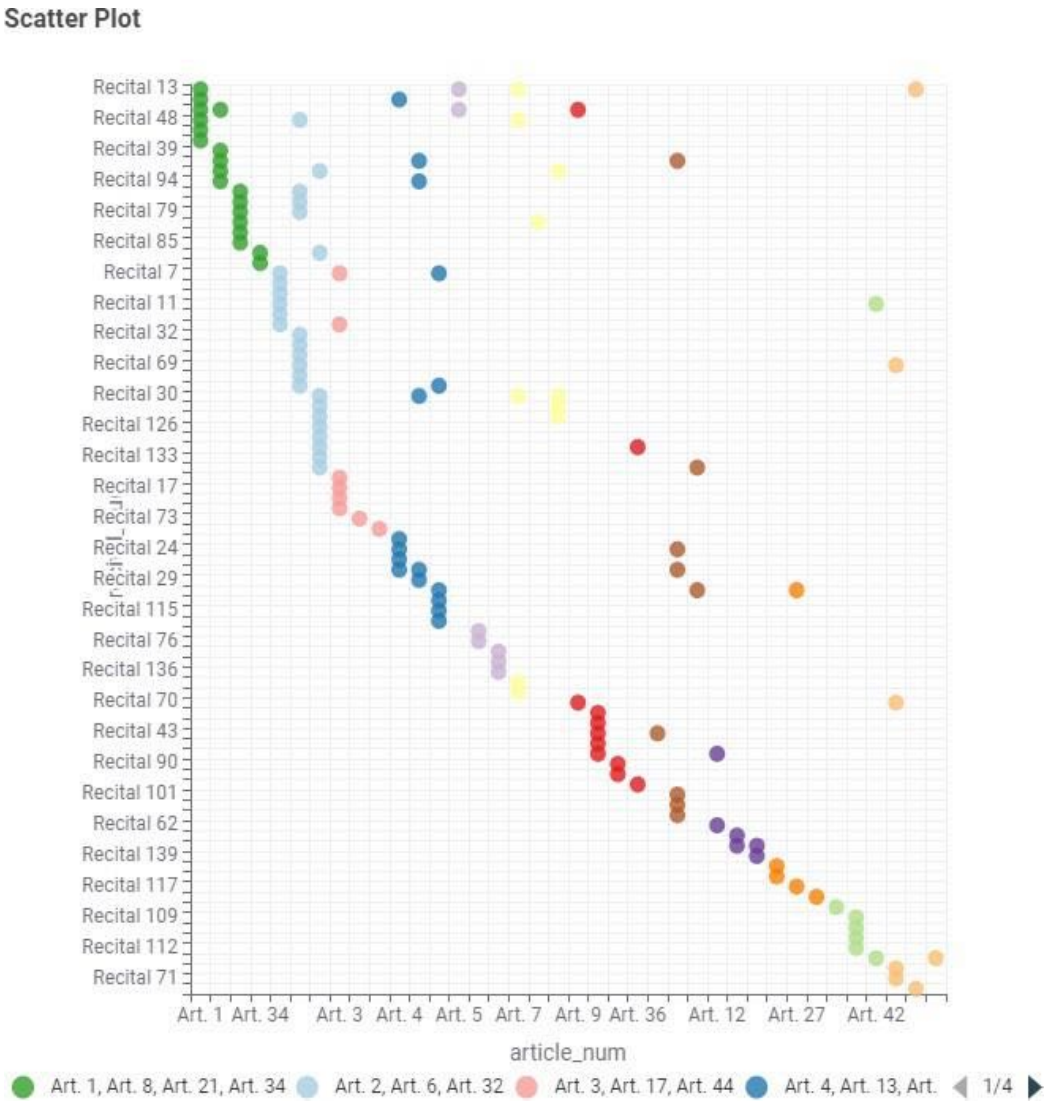


Figure 10: The RECONTA Scatter Plot.

This information could be provided to the legal drafter to confirm or disconfirm the connections and so to officially annotate within AKN the implicit references in an authoritative manner, using the following structure in the <meta> block:

```
<analysis source="">
    <otherReferences source="#MonicaPalmirani">
        <implicitReference for="#art_32" href="#recital_127"/>
    </otherReferences>
</analysis>
```

We use analysis because the correlation between articles and recitals is an output of legal interpretation and there can be many according to the source attribute. In this case the provenance of the legal interpretation is made by “MonicaPalmirani.” Different and multiple interpretations are admissible.

```
<otherReferences source="#FabioVitali">
  <implicitReference for="#art_32" href="#recital_133"/>
</otherReferences>
```

In this manner, it is possible to filter the fragments of annotation that are subjective and not objective using the provenance.

4. Use Case TRENDS

4.1. Introduction and motivation

Reporting Requirements (or Reporting Requests or Review Clauses - hereafter RR) are legislative provisions that command bodies to execute a task by an entity to another entity (e.g., from the European Commission to the European Parliament). A reporting requirement is not a legal obligation from the perspective of the theory of law [28] because there is no sanction and, consequently, no violation. They are not legal obligations in the classic meaning (addressed to the citizens, with a penalty) but meta-norms for managing and monitoring the law-making process within the legal system. In other words, it is an action addressed to the internal mechanism of the legal system.[11] Often, the RR requires an entity to prepare a report, plan, budget, opinion, or standard for another entity typically to monitor the implementation of the legislation.

The European Union legislation includes extensive reporting requirements involving a diverse range of entities (e.g., Member States, Agencies, …). Metadata plays a fundamental role in managing the legal knowledge of reporting requirement.[12] The volume of reporting requirements has grown substantially and is a main instrument to monitor the effectiveness of an act or the implementation of policies. However, derogations and exceptions [24], and their evolution and modifications over time may lead to very complex RRs making it difficult for legal experts to know exactly which reporting requirements apply and which ones are not relevant. Failure to manage this complexity has a negative impact e.g., on the implementation of the 'better regulation' system[13] and in general may lead to a less effective legislative European agenda[14]. To mitigate this risk, reporting requirements are continuously monitored and manually updated by amongst other agencies typically using SQL databases and NLP tools that scan EU legislation[15]. To arrive at a more structured approach, the use of AI tools

---

11 See the Case study analysis of regulatory reporting practices across the Euro- pean Commission, https://op.europa.eu/en/publication-detail/-/publication/ 5a5e5b13-e996-11e9-9c4e-01aa75ed71a1

12 See "The Importance of Metadata for Regulatory Reporting", Directorate-General for Informatics, D2 Interoperability, https://interoperable-europe.ec.europa.eu/sites/ default/files/document/2021-12/Issue%20-%20The%20Importance%20of%20Metadata.pdf

13 Better Regulation: Joining forces to make better laws, COM(2021) 219, 29 April 2021, European Commission See eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021DC0219

14 See the "Objectives of the Better Regulation agenda" at https://commission.europa. eu/law/law-making-process/planning-and-proposing-law/better-regulation_en

15 See for instance, https://commission.europa.eu/system/files/2023-10/Factsheet_ CWP_Burdens_10.pdf

combined with legal ontologies to detect and model the reporting requirements[16] was explored in the Study on Regulatory Reporting Standards (SORTIS) project[17] This resulted in the development of an ontology for Regulatory Reporting Metadata Vocabulary (RRMV).[18]

Here, we extended the use of RRMV and apply it:

- To identify existing reporting requirements during legal drafting.
- To simplify reporting requirements during legal drafting.
- To standardise some regular reporting requirements to improve interpretation and tracking over time; and
- To classify the reporting requirements to avoid duplications and reduce the number of reporting obligations.

4.2. Reporting Requirements

A reporting requirement usually has a pattern composed of following basic elements.

- Requirement is the fragment of the legislative source structure (e.g., article, paragraph) in a given version and language (e.g., version updated at 2025-04-10, in English) that produces at least one Action (e.g., present a report) in a given time (e.g., by 1 January 2026), where Agents (e.g., European Commission) are involved playing specific Roles (e.g., addresser).
- Action is an event with at least one Agent that is a participant in the action at a given time.
- Agent(s) act some action, addressed to other agents. Each agent plays a specific role (e.g., addressee, addresser) with a specific role type (e.g., reporter, legislator, etc.). The Agent also has a categorisation (e.g., institution, individual, company, group of individuals, etc.).
- TemporalEntitySpec is the specification of a temporal entity with non-zero extent or duration.

---

16 See Streamlining regulatoryreporting https://interoperable-europe. ec.europa.eu/collection/better-legislation-smoother-implementationstreamlining-regulatory-reporting/

17 See the results available at https://interoperable-europe.ec. europa.eu/collection/better-legislation-smoother-implementation/news/ streamlining-regulatory-reporting-sortis-project-results, https://interoperable-europe.ec.europa.eu/collection/better-legislation-smoother-implementation/

18 See solution/rrmv

- Action is Executed and can change the Status of the Action planned (e.g., completed or suspended like during the COVID).
- Requirement has a Topic classification (e.g., EUROVOC) that semantically qualifies the legal context (e.g., Financial).

Here is an example from the Directive 2014/65/EU of the European Parliament and the Council of 15 May 2014 on markets in financial instruments versioned in 2019[19]:

> "Art. 51 6. Commission... ESMA shall submit those draft regulatory technical standards to the Commission by 3 July 2015."

ESMA is the Agent with the Role of addresser that must function as an Action (submit) a Results (draft regulatory technical standards) delivered to another Agent (Commission) in the Role of addressee, at a given time (by 3 July 2015). See Figure 11.

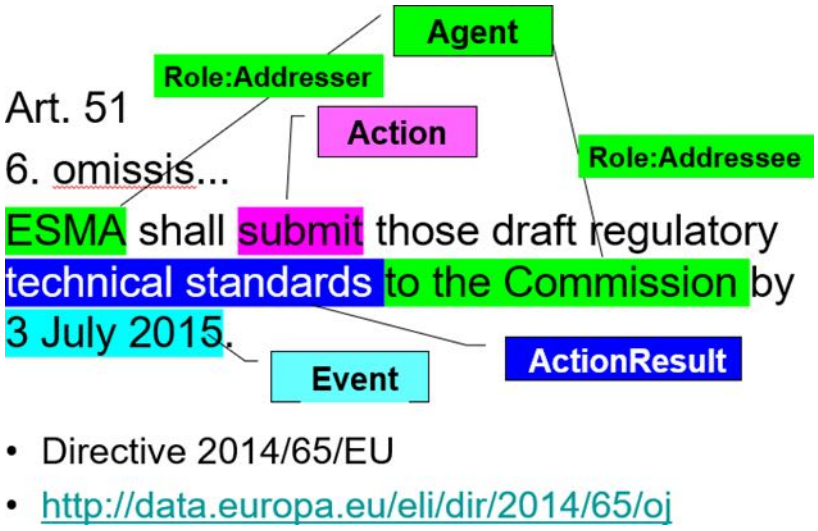


Figure 11: Art. 51 example.

A more complex case is to monitor the RR over time especially in view of the cascade implications on the whole legal system. E.g., in Directive 2014/65/EU[20] consolidated on 31 December 2019, paragraph number 9 was inserted in Article 33, creating a new Reporting Requirements:

> "Art. 33 Commission... 9. The Commission shall set up an expert stake-holder group by 1 July 2020 to monitor the functioning and success of SME growth markets. By 1 July 2021, the expert stakeholder group shall publish a report on its conclusions."

19 http://data.europa.eu/eli/dir/2014/65/2019-12-31

20 http://data.europa.eu/eli/dir/2014/65/2025-01-17

In this provision, we find two RRs connected to each other by a relationship of consequentiality. RR1 and RR2 see in Figure 12 where it is evident the connection with the two RRs.

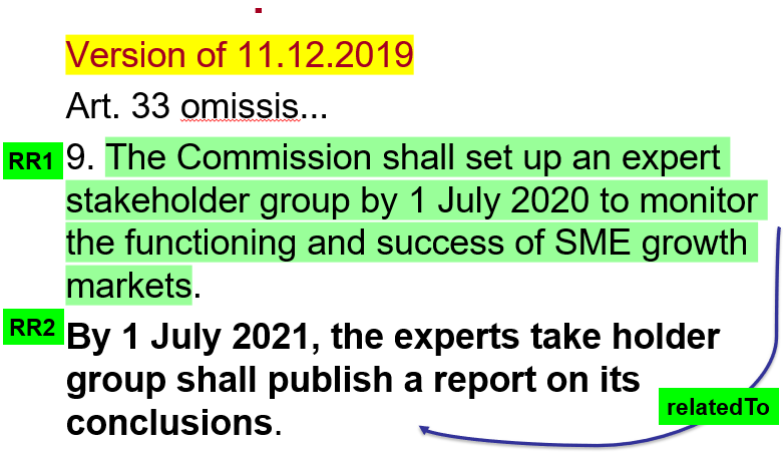


Figure 12: Art 33 example.

4.3. Methodology

To identify the patterns, we use a hybrid methodology considering the following:

(i) linguistic analysis of RRs.

(ii) legal analysis of the role of RRs in the legal drafting context.

(iii) semantic analysis of RRs.

We already argued that RRs are not a strong obligation type under the legal theory; however, we can classify RRs following the classification of public law:

- Organisational Norms: These norms establish the framework within which public power can be exercised. They are foundational to the principle of legality, ensuring that public bodies have explicitly delegated authority to act. It is also a pillar of the division of powers.
- Procedural Norms: These norms establish" how-to" of administrative actions, ensuring fairness, neutrality, and accountability.

In this context RRs are meta-norms, with the purpose to establish organisational and procedural norms that have side-effects within the European normative system (e.g., provide a report) or in the dialogue with the legal systems of the Member States (e.g., collect information from the other countries). For this reason, it is fundamental to distinguish the norms that are commands addressed to private sectors, citizens, society (external to the legal system process), and the norms that have the given purpose (teleological norms) to manage the normative system as the set of norms and procedures to maintain over the time the legal system.

4.4. Extracting information on reporting requirements using RRMV

To extract RRs from our dataset, we have used a Hybrid AI tool developed by the University of Bologna, using embedding database, open LLM, binary classification first and extracting entities later (see Figure 13) The paper was presented in ICAIL2025[21].

The architecture is designed to first take the AKN legislative text produced by LEOS and to analyse each article and each paragraph to binary classify the partition depending on whether includes a RR. Second, using the RRMV ontology, we extract the entities. Third, these entities are combined in RDF triples. Finally, we include inside of the AKN XML the serialisation of the corresponding tagging of the entities instead of the text (e.g., organisation, event, date, role) and in the metadata the RDF complete fragment. As a result, the legislative text is neutrally annotated using an agnostic approach, and the metadata provides an interpretation of this annotation using RRMV. The separation between metadata and content representation is one pillar in the AKN methodology to avoid a crystallisation of the legal interpretation within the content of the official legislative text. Multiple RDF variants are possible to guarantee a plurality of interpretations, applications, or purposes.

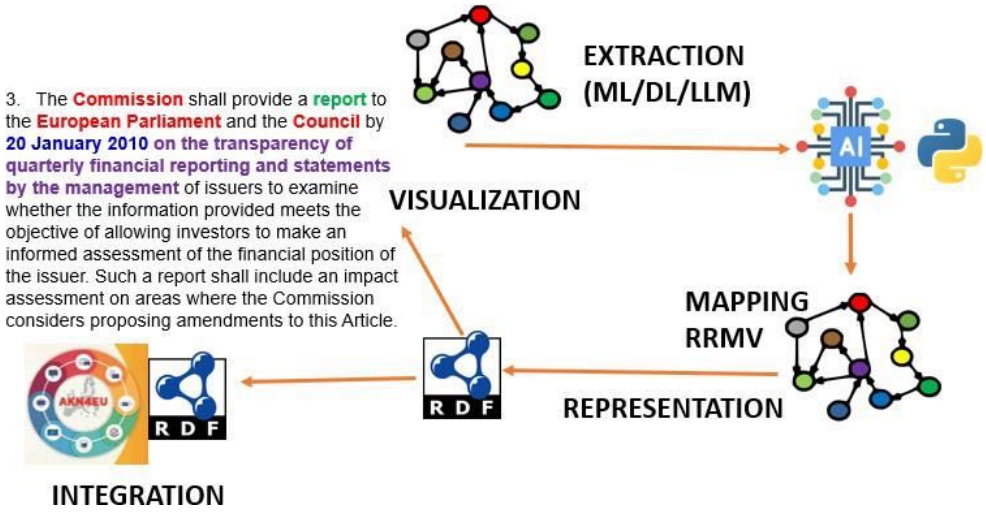


Figure 13: Architecture of RR Extraction.

21 [3]https://sites.northwestern.edu/icail2025/accepted-papers/

4.5. Linguistic and legal analysis of reporting requirements

We started to process the 891 reporting requirements of the datasets using the cloud map (see Figure 14). We can notice thematic words (e.g., food), names of agency (e.g., ESMA), the name of the type of document (e.g., Regulation), and the name of institutions (e.g., Commission).

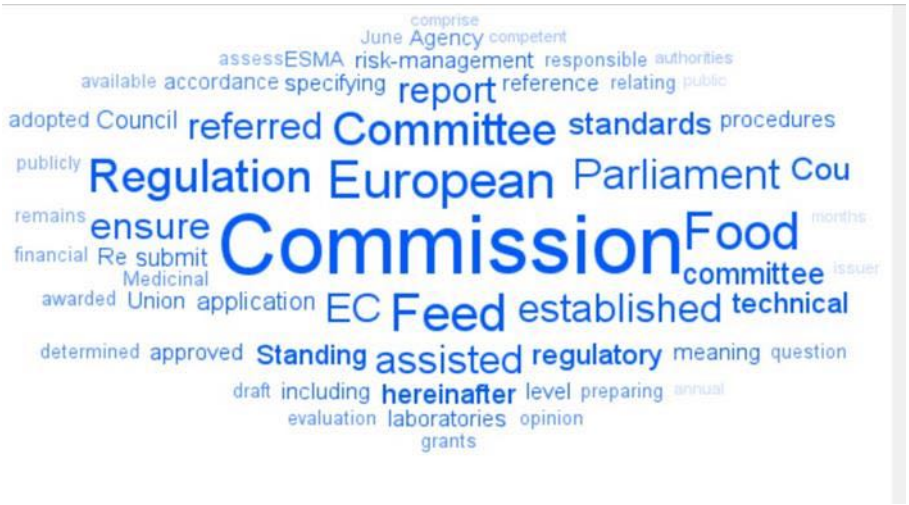


Figure 14: Cloud words.

Using the same set of reporting requirements, we extracted 259 verbal predicates with the Stanford NLP entity library and calculated their frequency. We also filtered those verbs > than 100 items, selecting a list of 25 major verbs: We notice in the Table 5 verbs' statistics (see also Figure 15) that 'submit' is the most frequent, followed by 'develop'.

Frequency	Verb
56	manage
62	consult
63	draw
64	approve
68	promote
70	review
82	specify follow
94	relate
101	carry
56	manage
62	consult

Frequency	Verb
63	draw
56	manage
62	consult
63	draw
64	approve
68	promote
70	review
82	specify follow
94	relate
101	carry
128	take
134	register
70	review
82	specify follow
94	relate
101	carry
128	take
134	register
136	regard
152	ensure
162	publish
180	adopt
196	implement
207	assess
286	determine
316	establish
369	provide
413	notify
499	refer
668	specify
749	develop
1488	submit

Table 5: Frequency of the most used verbs.

This analysis permits to define groups of verbs according to the linguistic categories (Table 6). To do this, we used Gemini to identify these linguistic categories.

Linguistic Category	Verb
Action/Movement	draw, carry, take, follow
Creation/Development	develop, establish
Information/Communication	consult, notify, publish, refer, register, relate, specify, submit
Judgment/Evaluation	approve, assess, determine, regard review
Implementation/Support	adopt, ensure, implement, manage, promote, provide

Table 6: Linguistic Categories.

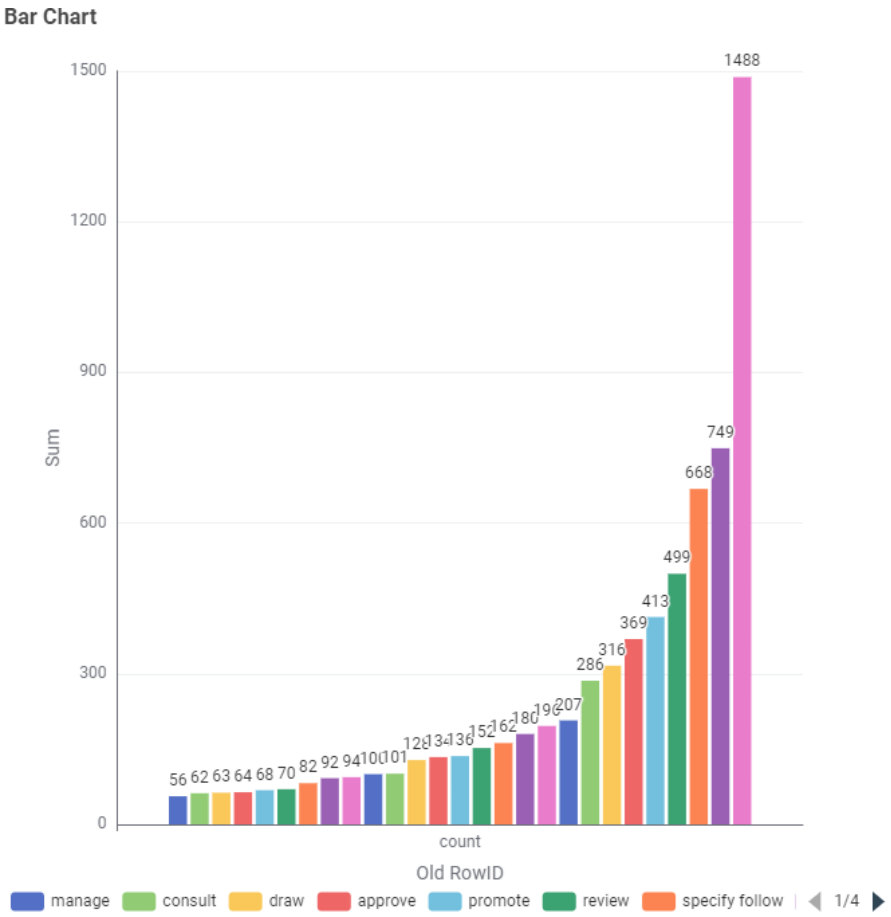


Figure 15: Verbs statistics.

After this linguistic analysis, we did a legal analysis to understand the role of these verbs in the light of the legal drafting techniques dimension. We have also extracted the most frequent legal sentence to analyse the functions of the RR (see Figure 16):

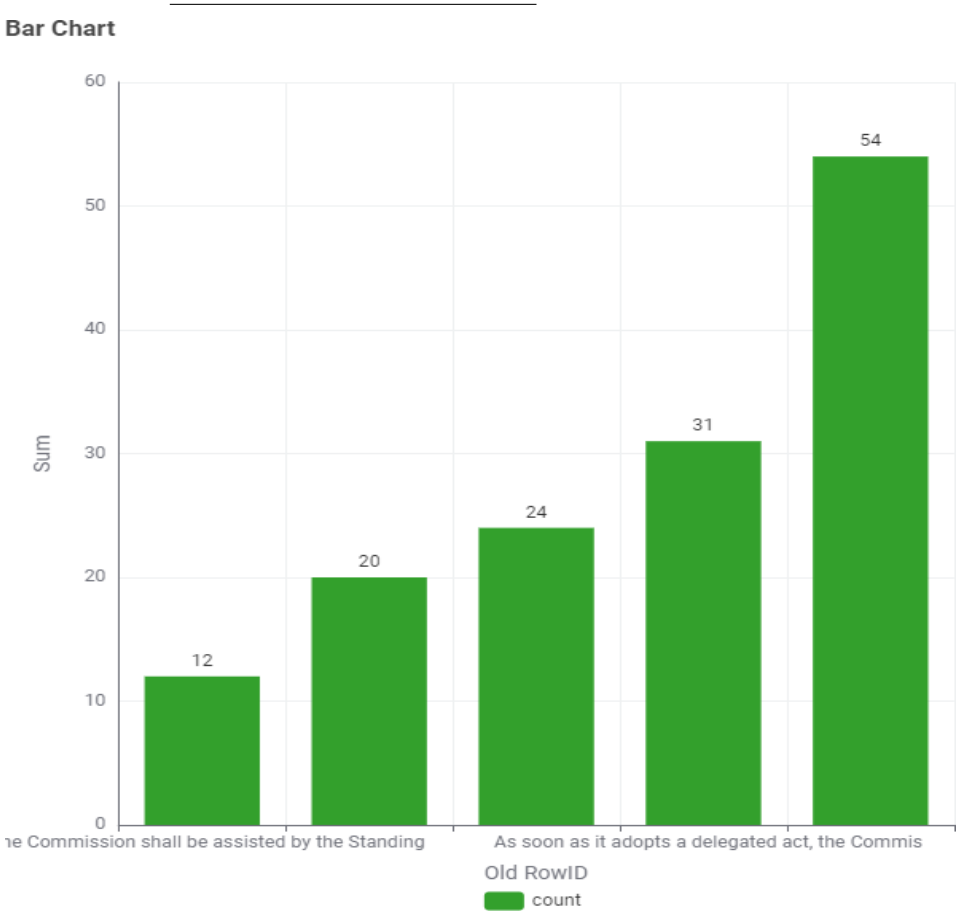


Figure 16: Sentences' statistics.

The analysis of the most frequent sentences is an important input to create the semantic taxonomy presented in the following paragraph 4.6.

After this step, we created patterns and calculated the frequency to understand which template to suggest in LEOS (see Figure 17 and Table 7).

Pattern	Frequency
<agent> shall <verb> <list>	12
<object> shall be <verb> <agent> <deadline>	17
<agent> shall develop draft implementing technical standards	22
The Commission shall be assisted by <agent>	22
<agent> shall <verb> <deadline>	23
<agent> shall, <deadline>, <verb>	24
confused	27
In order to <verb>, <agent> shall, in close cooperation with <agent>, <verb>	29

Pattern	Frequency
By <date>, <agent> shall <verb>	83
<agent> shall develop draft regulatory technical standards	85
<agent> shall <verb>	166

Table 7: Patterns and frequency.

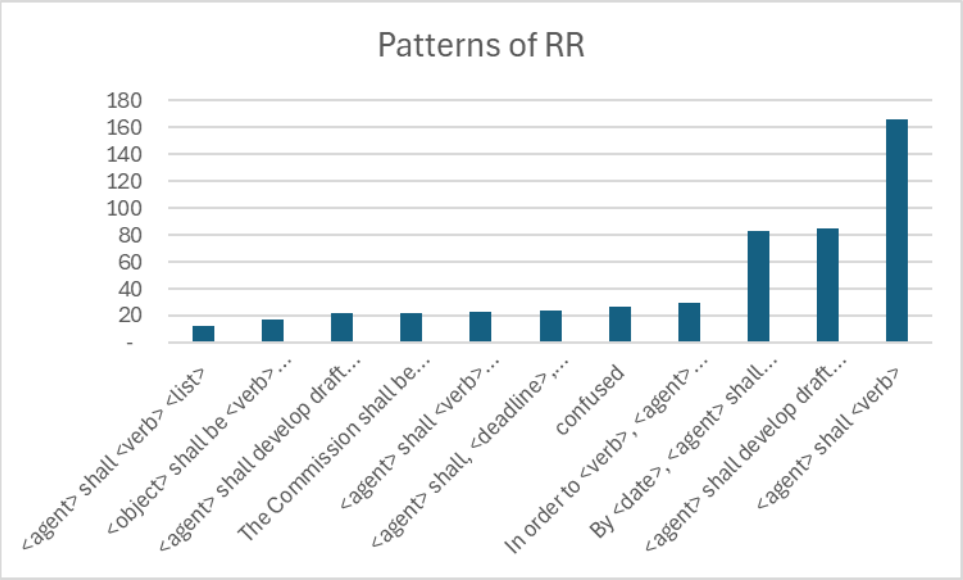


Figure 17: Patterns of sentences of RRs.

4.6. Semantic taxonomy of reporting requirements

Based on the above analysis, we identified the following semantic taxonomy of RRs that reflects the main legal functional categories of RRs inside of European legislation:

- EU Implementation: reporting requirements for implementing EU legislation.
- Member State Compliance: reporting requirement that involve Member States to comply with the EU legislation.
- Evaluation/Monitoring/Assessment: reporting requirements that aim to monitor, determine, evaluate, or assess.
- Plan/Strategy/Coordination: reporting requirements that aim to provide plan, strategy, coordination between institutions.

- Transparency/Accountability/Notification: reporting requirements that aim to notify, inform, provide opinion, communicate or record information.
- Budgetary/Financial Analysis: reporting requirements that define actions concerning budgetary or financial outcomes; and
- Solution/Achievement/Accomplishment: reporting requirements that aim to design, implement, document a solution or to provide achievement or accomplishment.

4.7. Templates of reporting requirements

We notice that there are three most frequent patterns as follows:

- By <date>, <agent> shall <verb>
- <agent> shall develop draft regulatory technical standards
- <agent> shall <verb>

The first pattern is the most complete with <date>, <agent>, <action>. For this reason, we have implemented this pattern in LEOS as proof-of-concept.

However, we found different variants of the same pattern:

- <agent> shall <verb> to <agent> <objective> of <topic> by <date>
- By <date>, <agent> shall <verb> <object> to <agent> on <topic>
- By <date>, and subsequently every <duration> <unit> thereafter, <agent> shall <verb> a <object> to <agent> on <topic>

Moreover, as the frequency of the RRs has the European Commission as the addresser, and European Parliament and the Council as addressees, we formulated in the LEOS integration a wording to help drafters to follow a clear pattern of provision.

4.8. Integration in LEOS

The integration in LEOS aims to demonstrate the use of reporting requirements templates, and subsequently to make the information contained in the reporting requirement available in AKN4EU and in RDF using RRMV.

We implemented the functionality to qualify the article as a reporting requirement in LEOS. At first, using the above taxonomy, we can identify the main components of the reporting requirement and the nature of the legal text being drafted (see Figure 18). Second, we offer a draft text for three common templates:

Template 1: The Commission shall [verb] to the European Parliament and to the Council a [object] of [topic or title] by [dd month year]

Template 2: By [dd month year], the Commission shall [verb] a [object] to the European Parliament and to the Council on [topic or title]

Template 3: By [dd month year], and subsequently every [duration] [unit] thereafter, the Commission shall [verb] a [object] to the European Parliament and to the Council on [topic or title].

The following Figure 19 and Figure 20 illustrate the implementation of these templates and the corresponding RDF automatically generated using AI. We first select the partition that we intend to draft as a Regulatory Rule (RR), as well as the specific type of RR. At this stage of the legal drafting process, no textual content is yet available for analysis. Based on these initial parameters, the system proposes a template in the LEOS interface, offering three variants derived from the previous data analysis: (i) a simple sentence, (ii) a sentence starting with the deadline, and (iii) a sentence starting with the deadline and including the duration.

All the information is then processed by an API, which extracts the relevant entities and serialises them in AKN4EU format. The entities are inserted both into the text of the provision and into the metadata block of the AKN structure (RDF of RRMV). This approach ensures that the RDF remains separate from the annotations of the original text, allowing other end-users (e.g., institutions) to add additional interpretations without duplicating the text.

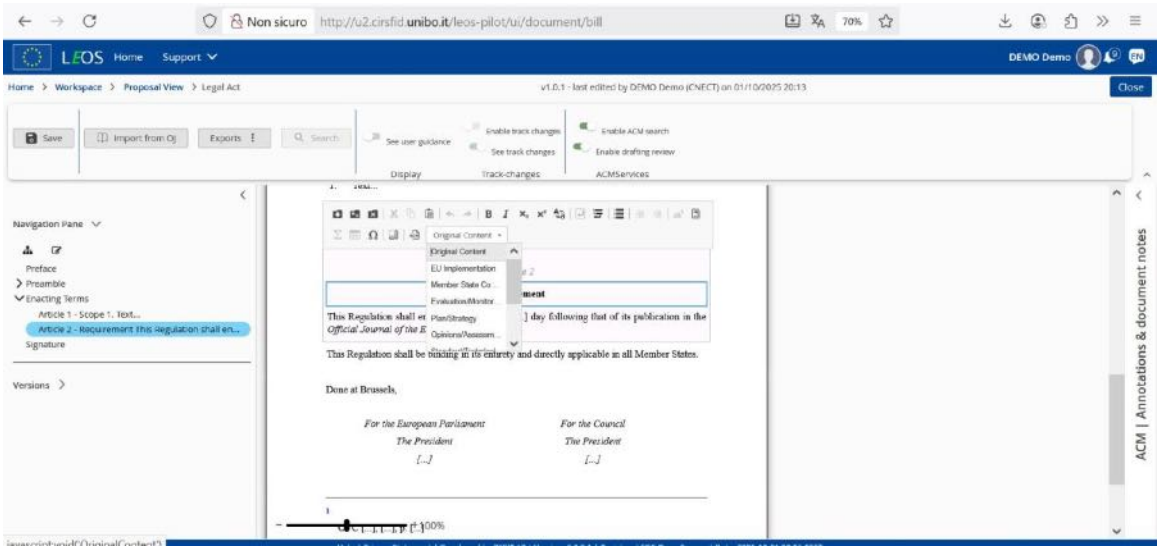


Figure 18: LEOS – selection of the type of provision.

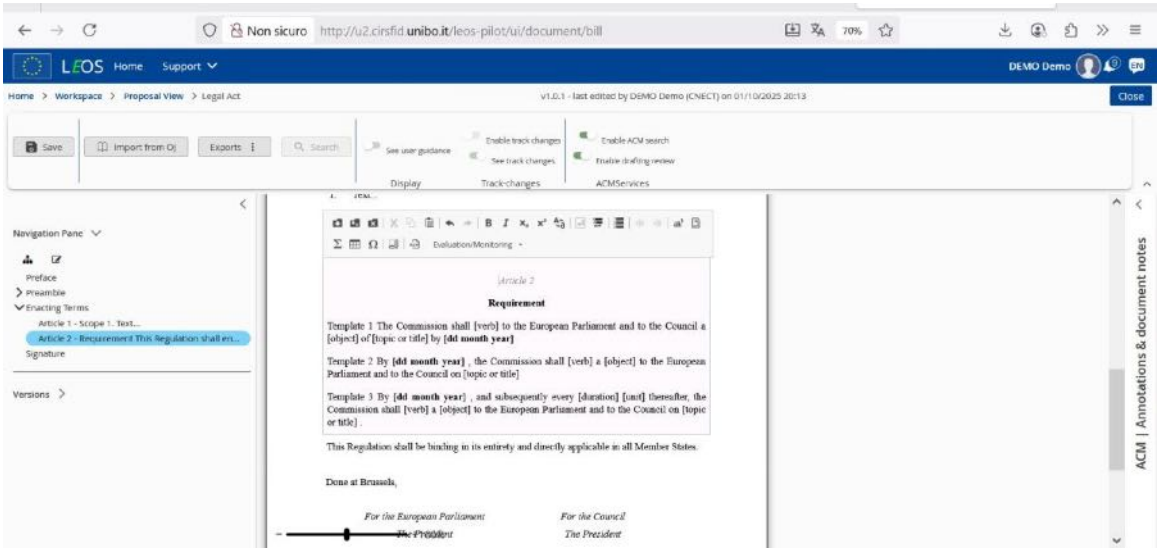


Figure 19: LEOS – template generation based on the type of provision.

In Figure 19 we can see the automatic creation of the template in the text.

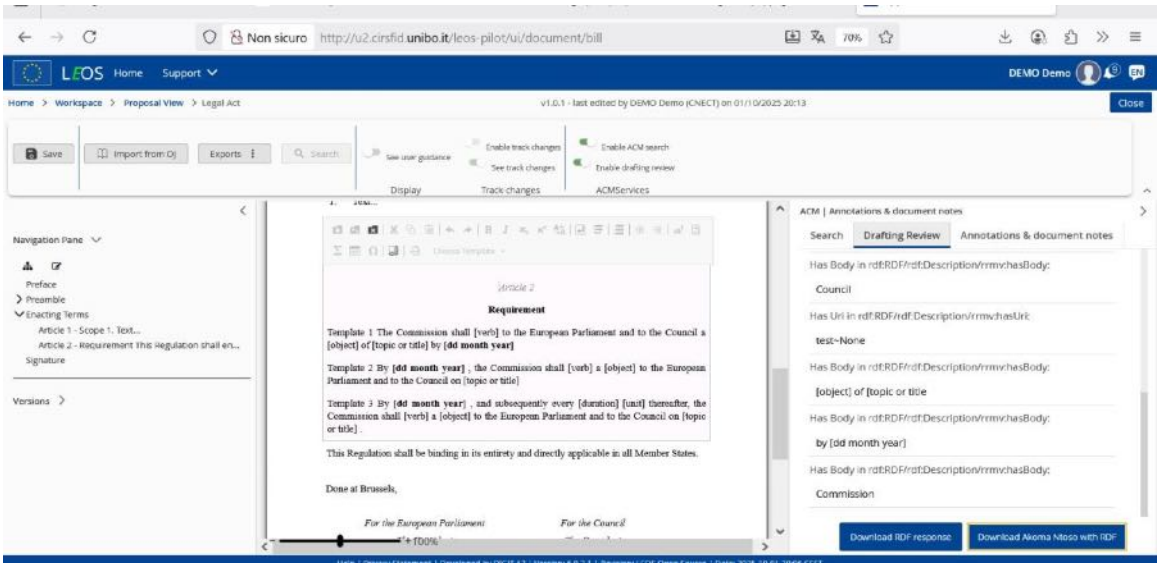


Figure 20: LEOS – analysis by the AI tool and extraction of the entities for the formalization in RDF.

In Figure 20 we can see the API elaboration of the text and the extractions of the semantic entities, according with the RRMV ontology. The two buttons in the box permit to download the RDF or the completely annotated AKN-XML file.

4.9. Serialisation in AKN4EU

In this session, we intend to show the output of the serialisation of AKN used in LEOS template.

First, hereafter we see how the three variants of template are proposed in the AKN content of the article. You see in the <content> Template 1, 2, 3 proposed as three separated <p>. Inside the template already offers the tagged content <organization>, <concept>, <date>:

```
<article leos:listIdAttr="_akn_article_nUUlqO" refersTo="~_ART_REQ"
 xml:id="__akn_article_tu0re9" class="current">
 <num leos:editable="false" xml:id="__akn_article_tu0re9_1IGSeM"> Article 7 </num>
 <heading xml:id="__akn_article_tu0re9_ntJv1J"> Requirement </heading>
 <paragraph>
  <content>
   <p> Template 1 The <organization refersto="#EUCommission"> Commission </organization>
    shall <concept> [verb] </concept> to the <organization refersto="#EUParliament">
     European Parliament </organization> and to the <organization refersto="#EUCoucil">
     Council </organization> a [object] of <concept> [topic or title] </concept> by <date
     date="9999-01-01"> [dd month year] </date>
   </p>
  </content>
 </paragraph>
 <paragraph>
  <content>
   <p> Template 2 By <date date="9999-01-01"> [dd month year] </date> , the <organization
     refersto="#EUCommission"> Commission </organization> shall <concept> [verb]
    </concept> a [object] to the <organization refersto="#EUParliament"> European
     Parliament </organization> and to the <organization refersto="#EUCoucil"> Council
    </organization> on <concept> [topic or title] </concept>
   </p>
  </content>
 </paragraph>
 <paragraph>
  <content>
   <p> Template 3 By <date date="9999-01-01"> [dd month year] </date> , and subsequently
    every [duration] [unit] thereafter, the <organization refersto="#EUCommission">
     Commission </organization> shall <concept> [verb] </concept> a [object] to the
     <organization refersto="#EUParliament"> European Parliament </organization> and to
    the <organization refersto="#EUCoucil"> Council </organization> on <concept> [topic or
     title] </concept> . </p>
  </content>
 </paragraph>
</article>
```

In the <analysis> metadata block we find the RDF serialisation of RRMV inside of the <otherAnalaysis> sub-block. This permits to other institutions to add further <otherAnalysis> sub-block with the specification of the provenance using @source.

In this case the source is cirsfid. The provenance of the semantic annotation is fundamental to understanding the 'authoritativeness' of the RDF assertions:

```
<analysis source="#cirsfid">
 <otherAnalysis source="#cirsfid">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:rrmv="http://data.europa.eu/2qy/rrmv#">
   <rdf:Description rdf:about="https://system.org/example#Requirement1">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#Requirement"/>
    <rrmv:isPartOf rdf:resource="http://publications.europa.eu/resource/celex/test#test"/>
    <rrmv:hasUri rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"> test~None
</rrmv:hasUri>
    <rrmv:produces rdf:resource="https://system.org/example#Requirement1Action1"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="N287a3aad04a74a618bb0413b82aaed2c">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#Annotation"/>
    <rrmv:hasBody> Template 1 The Commission shall [verb] to the European Parliament and
     to the Council a [object] of [topic or title] by [dd month year] </rrmv:hasBody>
    <rrmv:hasTarget rdf:resource="https://system.org/example#Requirement1"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Requirement1Action1AgentRole3">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#AgentRole"/>
    <rrmv:withRole rdf:resource="http://data.europa.eu/2qy/rrmv#addressee"/>
    <rrmv:forAgent rdf:resource="https://system.org/example#Agent3"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="N3085eda5595c4b81b541d3356dac2891">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#Annotation"/>
    <rrmv:hasBody> by [dd month year] </rrmv:hasBody>
    <rrmv:hasTarget
     rdf:resource="https://system.org/example#Requirement1Action1InstantSpec1"/>
   </rdf:Description>
   <rdf:Description rdf:about="http://data.europa.eu/2qy/rrmv#Role">
    <rrmv:hasType rdf:resource="http://data.europa.eu/2qy/rrmv#RoleType"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="Nee03865adaea468e9f243ea0c2ca56ce">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#Annotation"/>
    <rrmv:hasBody> Commission </rrmv:hasBody>
    <rrmv:hasTarget rdf:resource="https://system.org/example#Agent1"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Requirement1Action1AgentRole2">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#AgentRole"/>
    <rrmv:withRole rdf:resource="http://data.europa.eu/2qy/rrmv#addressee"/>
    <rrmv:forAgent rdf:resource="https://system.org/example#Agent2"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="N09a29996467c4737a5c10245d6e769ec">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#Annotation"/>
    <rrmv:hasBody> European Parliament </rrmv:hasBody>
    <rrmv:hasTarget rdf:resource="https://system.org/example#Agent2"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Requirement1Action1InstantSpec1">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#InstantSpec"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Agent2">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#Agent"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="Ne3864e4f89b046a1b5f75a90d075caf6">
```

```
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#Annotation"/>
    <rrmv:hasBody> [object] of [topic or title </rrmv:hasBody>
    <rrmv:hasTarget
      rdf:resource="https://system.org/example#Requirement1Action1ActionResult1"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Requirement1Action1">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#Action"/>
    <rrmv:atTime rdf:resource="https://system.org/example#Requirement1Action1InstantSpec1"/>
    <rrmv:hasAgentRole
      rdf:resource="https://system.org/example#Requirement1Action1AgentRole1"/>
    <rrmv:hasAgentRole
      rdf:resource="https://system.org/example#Requirement1Action1AgentRole2"/>
    <rrmv:hasAgentRole
      rdf:resource="https://system.org/example#Requirement1Action1AgentRole3"/>
   </rdf:Description>
   <rdf:Description rdf:about="http://data.europa.eu/2qy/rrmv#bearer">
    <rrmv:subClassOf rdf:resource="http://data.europa.eu/2qy/rrmv#RoleType"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Requirement1Action1AgentRole1">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#AgentRole"/>
    <rrmv:withRole rdf:resource="http://data.europa.eu/2qy/rrmv#addresser"/>
    <rrmv:forAgent rdf:resource="https://system.org/example#Agent1"/>
   </rdf:Description>
   <rdf:Description rdf:nodeID="N2499d898eee54a6095bc047c56e7d0e2">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#Annotation"/>
    <rrmv:hasBody> Council </rrmv:hasBody>
    <rrmv:hasTarget rdf:resource="https://system.org/example#Agent3"/>
   </rdf:Description>
   <rdf:Description rdf:about="http://data.europa.eu/2qy/rrmv#addressee">
    <rdfs:subClassOf rdf:resource="http://data.europa.eu/2qy/rrmv#RoleType"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Agent1">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#Agent"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Agent3">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#Agent"/>
   </rdf:Description>
   <rdf:Description rdf:about="https://system.org/example#Requirement1Action1ActionResult1">
    <rdf:type rdf:resource="http://data.europa.eu/2qy/rrmv#ActionResult"/>
   </rdf:Description>
   <rdf:Description rdf:about="http://data.europa.eu/2qy/rrmv#addresser">
    <rdfs:subClassOf rdf:resource="http://data.europa.eu/2qy/rrmv#RoleType"/>
   </rdf:Description>
  </rdf:RDF>
 </otherAnalysis>
</analysis>
```

5. Conclusions

We elaborated and demonstrated:

(i) The architecture based on micro-services using RESTful approach to integrate LEOS web editor with advanced AI functionalities.

(ii) The power of Hybrid AI that permits to mitigate risks of AI in legal drafting domain using Semantic Web technologies (e.g., RRMV ontology and RDF databases), RAG (e.g., AKN XML repository in eXsist), embeddings, and the use of LLMs; and

(iii) The value of a robust legal-informatics methodology involving an interdisciplinary group of researchers where AI is used not only to provide a solution, but also to discover hidden information (e.g., linguistic patterns) that is essential to orient AI applications.

Additionally, the work shows how to implement use-cases, where legal, linguistic, semantic, and technical analysis need to be combined using a solid teleological approach to produce tangible benefits in the legislative domain relevant for democracy in our societies.

Acknowledgements

This work is funded by the European Commission within the DG-Informatics, Unit DG DIGIT A3 - HR and Legislative Solutions. The ERC HyperModeLex also supports the work. Grant agreement ID: 101055185 and by the Erasmus+ Jean Monnet Legal Design and Data Science for Explicable AI in Legal Domain (LEDS 4 XAIL) Grant No. 101085576.

We would like to thank Aurora Brega, who programmed in her thesis the User interface module in LEOS, and Emanuele Di Sante, who developed in his thesis part of the module REFINA and the serialization of the AKN-dataset of EU Legislation. Fabio Vitali of the University of Bologna revised the work of these two theses. We would like to thank Fernando Nubla Durango, Head of Sector of DIGIT.A3.005 - Decision making solutions, Willy Van Puymbroeck, and Cristina Stanciulescu, who contributed to defining the scope of the project, assessing the results, disseminating the output, and creating a community of stakeholders around these use-cases.

6. References

[1] Palmirani Monica RB. An XML Editor for Legal Information Management.In: Traunm¨uller R, editor. Electronic Government, Second International Conference, EGOV 2003, Prague, Czech Republic, September 1-5, 2003. vol. 16049 of Lecture Notes in Computer Science. Springer; 2003. p. 421-9.

[2] VV A. legisPro; 2025. Available from: https://joinup.ec.europa.eu/collection/justice-law-and-security/solution/leos-open-source-software-editing-legislation.

[3] Agnoloni T., E Francesconi, P. Spinosa. xmLegesEditor: an open source visual XML editor for supporting legal national standards. In: Proceedings of the V legislative XML workshop. European Press Academic Publishing; 2007. p. 239-51.

[4] Vergottini G. legisPro Standards; 2025. Available from: https:// xcential.com/legispro/standards/.

[5] Vergottini G. Comparing Legislative Drafting Software Solutions; 2025. Available from: https://xcential.com/resources/legislative-drafting-and-publishing-software-solutions.

[6] VV A. Leos Manual; 2025. Available from: https://joinup.ec.europa.eu/collection/justice-law-and-security/solution/leos-open-source-software-editing-legislation/releases.

[7] Palmirani M, et al. Drafting legislation in the era of AI and digitisation;2025. Available from: https://joinup.ec.europa.eu/sites/default/files/document/2022-06/Drafting%20legislation%20in%20the%20era%20of%20AI%20and%20digitisation%20%E2%80%93%20study.pdf.

[8] Corazza M, Zilli L, Palmirani M. Topic Similarity of Heterogeneous Legal Sources Supporting the Legislative Process. In: Dell'Orletta F, Lenci A, Montemagni S, Sprugnoli R, editors. Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024). Pisa, Italy: CEUR Workshop Proceedings; 2024. p. 244-50. Available from: https://aclanthology.org/2024.clicit-1.29/.

[9] Corazza M, Sapienza S, Palmirani M, Longo G, Zilli L, Sante ED. Hybrid AI for Supporting the European Drafting Legislation. In: Nakano YI, Suzumura T, editors. New Frontiers in Artificial Intelligence – JSAI International Symposium on Artificial Intelligence, JSAI-isAI 2025, Osaka, Japan, May 26-27, 2025, Proceedings. vol. 15692 of Lecture Notes in Computer Science. Springer; 2025. p. 80-95. Available from: https: //doi.org/10.1007/978-981-96-7071-0_6.

[10] Asif M, Palmirani M. Mapping EU Legislative Definitions on Sustainable Development Goals Using AI. In: Ko A, Buccafurri F, Kotsis G, Tjoa AM, Khalil I, editors. Electronic Government and the Information Systems Perspective - 14th International Conference, EGOVIS 2025, Bangkok, Thailand, August 25-27, 2025, Proceedings. vol. 16049 of Lecture Notes in

Computer Science. Springer; 2025. p. 3-17. Available from: https://doi.org/10.1007/978-3-032-02225-7_1.
[11] Palmirani M, Sovrano F, Liga D, Sapienza S, Vitali F. Hybrid AI framework for legal analysis of the EU legislation corrigenda. In: Legal knowledge and information systems. IOS Press; 2021. p. 68-75.
[12] Salvatore Sapienza MP Kevin Ashley. A Hybrid Artificial Intelligence Methodology for Legal Analysis. BioLaw. 2024;3/2024.
[13] Sapienza S, Palmirani M. Algorithmic Knowability: A Unified Approach to Explanations in the AI Act. In: Guidotti R, Schmid U, Longo L, editors. Explainable Artificial Intelligence. Cham: Springer Nature Switzerland; 2026. p. 185-209.
[14] Palmirani M, Vitali F. Akoma-Ntoso for Legal Documents. In: Sartor G, Palmirani M, Francesconi E, Biasiotti MA, editors. Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management. Dordrecht: Springer Netherlands; 2011. p. 75-100.
[15] Agnoloni T, Marchetti C, Battistoni R, Briotti G. Clustering Similar Amendments at the Italian Senate. In: Fiˇser D, Eskevich M, Lenardiˇc J, de Jong F, editors. Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association; 2022. p. 39-46. Available from:https://aclanthology.org/2022.parlaclarin-1.7/.
[16] Agnoloni T, Bartolini R, Frontini F, Montemagni S, Marchetti C, Quochi V, et al. Making Italian Parliamentary Records Machine-Actionable: the Construction of the ParlaMint-IT corpus. In: Fišer D, Eskevich M, Lenardič J, De Jong F (editors). Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association; 2022. p. 117-24. Available from: https://aclanthology.org/2022.parlaclarin-1.17/.
[17] Lesmo L, Mazzei A, Palmirani M, Radicioni D. TULSI: an NLP system for extracting legal modificatory provisions. Artificial Intelligence and Law. 2013 05;21.
[18] Asif M, Palmirani M. Legal Definition Annotation in EU Legislation Using Symbolic AI. In: K¨o A, Kotsis G, Tjoa AM, Khalil I, editors. Electronic Government and the Information Systems Perspective. Cham: Springer Nature Switzerland; 2024. p. 34-9.
[19] Anim J, Robaldo L, Wyner AZ. A SHACL-Based Approach for Enhancing Automated Compliance Checking with RDF Data. Information. 2024;15(12). Available from: https://www.mdpi.com/2078-2489/15/12/ 759.
[20] Kebede MG, Sileno G, Van Engers T. A Critical Reflection on ODRL. In: AI Approaches to the Complexity of Legal Systems XI-XII: AICOL International Workshops 2018 and 2020: AICOL-XI@JURIX 2018, AICOLXII@ JURIX 2020, XAILA@JURIX 2020, Revised Selected Papers. Berlin, Heidelberg: Springer-Verlag; 2020. p. 48–61. Available from: https: //doi.org/10.1007/978-3-030-89811-3_4.
[21] Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Proceedings of the Soviet physics doklady. 1966.

[22] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics. 2017;5:135-46. Available from: https://aclanthology.org/Q17-1010/
[23] Warner B et al. Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference; 2024. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics, Vienna, Austria, 2025. p. 2526-47.
[24] Palmirani M, Liga D. Derogations Analysis of European Legislation Through Hybrid AI Approach. In: Electronic Government and the Information Systems Perspective: 11th International Conference, EGOVIS 2022, Vienna, Austria, August 22–24, 2022, Proceedings. Berlin, Heidelberg: Springer-Verlag; 2022. p. 123–137. Available from: https://doi.org/10.1007/978-3-031-12673-4_9.
[25] Laws.africa, Open Law Technology; 2025. Available from: https://laws.africa/open-law-technology
[26] The Parliamentary Counsel Office of New Zealand, Legislative data and technology; 2025. Available from: https://www.pco.govt.nz/about-us/legislative-data-and-technology/
[27] Bakker, R.M., Schoevers, A.J., van Drie, R.A.N. et al. Semantic role extraction in law texts: a comparative analysis of language models for legal information extraction; 2025. Artificial Intelligence and Law (2025). Available from: https://doi.org/10.1007/s10506-025-09437-x
[28] Kelsen, H., Pure Theory of Law; 1967. Berkeley, CA: University of California Press.

7. List of figures

8. List of tables

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696,
– via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

NO-01-25-012-EN-C

Publications Office
of the European Union

ISBN 978-92-68-33290-0