D14 – Semantic graph of extracted data

Project funded by the European Union – Next GenerationEU through the Recovery and Resilience Plan of the Slovak Republic under project No. 09I05-03-V02-00049.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU
VAIA VÝSKUMNÁ A INOVAČNÁ AKTIVITA
VÝSKUMNÁ AGENTÚRA

INTRODUCTION

The project **"Automation of Legal Text Analysis Based on Machine Learning"** (hereinafter referred to as **"ALTAML"**) represents a key initiative aimed at integrating innovative approaches in the field of data processing and subsequent analysis, specifically within the legal domain, which also includes the law of information and communication technologies. The objective of this project is to develop and validate effective methods for the automated analysis of legal texts using machine learning techniques. This includes, in particular, the development of tools that facilitate the processing and analysis of large volumes of data in the form of various legal documents, as well as the extraction of relevant information (attributes) from such documents, including the identification of key terms, references to other legal acts, and other attributes.

The ALTAML project thus aims to contribute to a more efficient access to legal information and to the acceleration of legal processes, thereby ensuring a higher degree of legal certainty and improving the accessibility of legal texts, the results of their analysis, and relevant legal information for both professionals and the general public.

Within the KPB3 work package, a proposal for the semantic representation of Slovak court decisions was developed, compatible with the data representation of the Court of Justice of the European Union (CJEU) in the form of an RDF graph. The resulting dataset is an RDF graph containing both CJEU decisions and Slovak court decisions.

The RDF (Resource Description Framework) and OWL (Ontology Web Language) languages are used to represent information in the so-called Semantic Web, which focuses on the description of entities, their properties, and relationships, thus providing data with meaning/semantics that are moreover processable and usable automatically without human intervention, primarily based on the universal language of mathematical logic. The fundamental method for querying and navigating an RDF graph is querying using the SPARQL language.

1. Use of the Semantic Web in European Union institutions

For more than ten years, European institutions have been publishing legislation and case law as *a semantic web* (linked data). The benefits of this approach are mainly:

- **Global identifiers.** Decisions (*ECLI* identifier) and legal acts (*ELI* identifier) are given global URIs, which make them referable across systems.
- **Common RDF graph.** All data on judges, decisions, legal acts, etc. are stored in a single common RDF graph, where their relationships and connections are precisely defined using the *Common Data Model (CDM)* ontology.
- **One source of truth.** Web portals (*EUR-Lex*, *Curia*) and the SPARQL endpoint work with the same graph – no duplicate data conversion is necessary.
- **Querying and filtering.** It is possible to search the graph for decisions by date, type of proceedings, topics, cited legal acts and other criteria.
- **Multilingualism.** Document texts exist in many languages, URIs and relationships remain common; language can be switched using rdfs:label.

Slovak court decisions are currently scattered across various databases without a unified model. The dataset in this deliverable shows that the same data model currently used by the EU is sufficiently universal to represent Slovak case law – without the need for further conversion or specific formats, thus enabling the creation of a universal database of decisions.

The ontological layer contains these three main models:

- **Common Data Model (CDM** http://publications.europa.eu/ontology/cdm#**)** – ontology of the EU Publications Office. It defines classes such as ("cdm:case-law_work", "cdm:legal_resource", "cdm:Judge" or "cdm:court_national" and the relationships between them, such as ("cdm:case-law_delivered_by_judge", "cdm:work_cites_work" . . .). Each CDM instance appears only once in CELLAR, and the *EUR-Lex*, *Curia* and SPARQL endpoint portals all refer to the same URI.
- **SKOS + EuroVoc** (http://eurovoc.europa.eu/) – *Simple Knowledge Organisation System* (SKOS) is a W3C standard for publishing thesauri and hierarchical dictionaries. EuroVoc is just such a thesaurus – it contains approximately 7,000 multilingual concepts, ranging from general topics (such as "law") to detailed terms (such as "compensation"). Each decision or legal act can be assigned to one or more EuroVoc topics, so that thematic information remains separate from the CDM structure.
- **Dublin Core Terms** (**DcTerms** http://purl.org/dc/terms/) – a generic layer of the most common bibliographic attributes (e.g. "dcterms:title", "dcterms:language"). CDM does not duplicate these properties but directly adopts them (e.g. "cdm:work_date_creation" is subPropertyOf "dcterms:date"), thus it maintains interoperability with other open data that use the classic Dublin Core.

The result is a model: **CDM** describes the structure and links between documents, **SKOS/EuroVoc** provides thematic classification, and **DcTerms** provides basic bibliographic data. Above this ontological layer are the metadata part (CELLAR) and the document part (REST API), which together form the complete EU Linked Data ecosystem.

The *CELLAR* repository contains instances of CDM model classes (specifically judgments, laws, judges, etc.). Instances are linked by properties, e.g. "cdm:work_cites_work" (the decision refers to another document) or "cdm:judge_delivered_case-law" (the judge is the author of the case law). Only brief information such as source URIs, names, dates, identifiers and links are stored in the RDF graph. The full texts of documents (e.g. the judgments or laws themselves) are not included in the graph. They are stored separately in a separate document layer.

In this project output, we focus on extending the metadata layer (CDM model) with Slovak court decisions.

2. CDM data model

In this subchapter, we provide an overview of all elements of the Common Data Model (CDM) related to court decisions that were analyzed and used in the extension of the ontology for Slovak court decisions. This is followed by a detailed description of the meaning of a given class or property in the context of CDM and legal jurisprudence – what it represents, what subjects and objects it links, and what implications arise from its use.

2.1. Prefixes used in the EU RDF graph

The most important prefixes used in defining ontologies based on RDF and OWL standards in the EU are:

- **cdm:** – prefix for the official Common Data Model ontology published by the EU Publications Office: http://publications.europa.eu/ontology/cdm#.
- **foaf:** – *Friend of a Friend* ontology (http://xmlns.com/foaf/0.1/) designed to represent people, organizations and their relationships – often serves as a bridge to external identities (e.g. via foaf:Agent).
- **owl:** – prefix *for Web Ontology Language* (http://www.w3.org/2002/07/owl#).
- **xsd:** – prefix for data types defined in *XML Schema Definition* (http://www.w3.org/2001/XMLSchema#), for example "xsd:string", "xsd:date" or "xsd:boolean".

2.2. Important classes used in the EU RDF graph

- **owl:Thing** is the most general object in OWL ontologies – it represents the root node under which all classes fall. It is the "top concept"; all classes are **owl:Thing**.
- **cdm:Work**
  - URI http://publications.europa.eu/ontology/cdm#Work
  - Inheritance: $cdm:Work \subseteq owl:Thing$

    - Basic CDM class for *a work* – abstract intellectual or legal content independent of language or form of publication. All court decisions are a subclass **of cdm:Work**
- **cdm:case-law_work**
    - URI http://publications.europa.eu/ontology/cdm#case-law_work
    - Inheritance: $cdm: Case - law_work \subseteq cdm: Work \subseteq owl: Thing$
    - Specifies *a court decision* (judgment) as a work. It retains all the properties **of cdm:Work**, but allows the addition of specialized links (judge, senate, country of origin).
- **cdm:Judge**
    - URI http://publications.europa.eu/ontology/cdm#Judge
    - Inheritance: $cdm: Judge \subseteq cdm: Agent \subseteq owl: Thing$
    - Represents an individual judge – a natural person working in the judiciary. Acts as an agent when issuing a decision; links to the decision via the properties cdm:case-law_delivered_by_judge or cdm:judge_delivers_case-law.
- **cdm:court_national**
    - URI http://publications.europa.eu/ontology/cdm#court_national
    - Inheritance: $cdm: Court_national \subseteq cdm: Agent \subseteq owl: Thing$
    - The court that decides. It is linked to decisions via the property cdm:case-law_delivered_by_court_national or via the property cdm:court_national_delivers_case-law.
- **cdm:court_type**
    - URI http://publications.europa.eu/ontology/cdm#court-type
    - Inheritance: $cdm: Court_type \subseteq cdm: Concept \subseteq owl: Thing$
    - Type of court that issued the decision
- **cdm:Country**
    - URI http://publications.europa.eu/ontology/cdm#Country
    - Inheritance: $cdm: Country \subseteq cdm: Agent \subseteq owl: Thing$
    - State as a legal entity.

2.3. Significant properties used in the EU RDF graph

- **cdm:work_cites_work**
    - **Type:** object property
    - **Inheritance:** cdm:work_cites_work ⊆ cdm:cites ⊆ owl:topObjectProperty
    - **Domain:** cdm:Work
    - **Value domain:** cdm:Work
    - **Description:** Expresses *the reference* of one work to another (e.g. a decision refers to another decision or legislative act). Directional, non-inverse, allows modelling of a network of references.
- **cdm:case-law_ecli**
    - **Type:** data property
    - **Inheritance:** cdm:case-law_ecli ⊆ cdm:ecli ⊆ owl:topDataProperty
    - **Domain:** cdm:case-law_work

    - **Value domain:** xsd:string
    - **Description:** European Case Law Identifier (ECLI). Each decision can have exactly one unique ECLI string.
- **cdm:case-law_delivered_by_judge**
    - **Type:** object property
    - **Inheritance:** cdm:case-law_delivered_by_judge ⊆ cdm:delivered_by ⊆ owl:topObjectProperty
    - **Domain:** cdm:case-law_work
    - **Value domain:** cdm:Judge
    - **Description:** Links the decision to the judge who issued it
- **cdm:judge_delivers_case-law**
    - **Type:** object property
    - **Inheritance:** cdm:judge_delivers_case-law ⊆ cdm:delivers⊆ owl:topObjectProperty
    - **Domain:** cdm:Judge
    - **Value domain:** cdm:case-law_work
    - **Description:** Inverse view – which decisions were issued by a specific judge.
- **cdm:case-law_delivered_by_court_national**
    - **Type:** object property
    - **Inheritance:** cdm:case-law_delivered_by_court_national ⊆ cdm:delivered_by ⊆ owl:topObjectProperty
    - **Domain:** cdm:case-law_work
    - **Value domain:** cdm:court_national
    - **Description:** Specifies that the decision was issued by a specific court.
- **cdm:court_national_delivers_case-law**
    - **Type:** object property
    - **Inheritance:** cdm:court_national_delivers_case-law ⊆ cdm:delivers ⊆ owl:topObjectProperty
    - **Domain:** cdm:court_national
    - **Value domain:** cdm:case-law_work
    - **Description:** Inversion of the previous property – from the perspective of the senate.
- **cdm:case-law_originates_in_country**
    - **Type:** object property
    - **Inheritance:** cdm:case-law_originates_in_country ⊆ cdm:originates_in⊆ owl:topObjectProperty
    - **Domain:** cdm:case-law_work
    - **Value domain:** cdm:Country
    - **Description:** Specifies the country in which the decision originated (jurisdiction of origin).
- **cdm:country_initiates_case-law**
    - **Type:** object property
    - **Inheritance:** cdm:country_initiates_case-law ⊆ cdm:initiates ⊆ owl:topObjectProperty
    - **Domain:** cdm:Country

- **Value domain:** cdm:case-law_work
- **Description:** Inversion of the previous relationship – from the perspective of the state that initiated the case.

- **cdm:work_date_creation**
  - **Type:** data property
  - **Inheritance:** cdm:work_date_creation ⊆ cdm:date_document ⊆ cdm:date ⊆ owl:topDataProperty
  - **Definition domain:** cdm:Work
  - **Value domain:** xsd:date /xsd:dateTime
  - **Description:** Date of creation of the work; for decisions, this is the date of issuance (announcement) of the judgment.

3. Integration of Slovak court decisions into the EU RDF graph

The metadata of court decisions made available via the Ministry of Justice API contains several metadata that have a non-empty intersection with the metadata on court decisions in the EU RDF graph. On the other hand, there are also attributes that are part of the metadata of Slovak court decisions but are not present in the CDM. Our ambition was to represent overlapping attributes in a uniform and semantically correct manner so that both European and Slovak decisions could be processed with in a uniform manner.

The *Common Data Model* (CDM) ontology for court decisions focuses primarily on the case law of the Court of Justice of the European Union. Its key class is cdm:case-law_work, there are no court decisions which are not case law.

Although our dataset of Slovak court decisions also contains case law, this information is not explicitly stored. Related *legal sentences* indicating that a decision is a case law are often published separately as unstructured text documents (typically in .doc/.docx format) on the websites of individual courts, under the title Collections of Opinions and Decisions.

The Ministry publishes not only case law, but also *ordinary* judgments and resolutions of the first or second instance courts. Lower-level decisions do not appear in the EU RDF graph because these decisions have properties that are not recognised by CDM, such as information about the district and region to which the court belongs, etc.

In order to maintain semantic consistency and at the same time be able to model these specifics, we are introducing a new superclass upjs_ontology:court_decision

cdm:case-law_work ⊑ upjs_ontology:court_decision ⊑ cdm:Work ⊑ owl:Thing ,

As a result, *every* European (including Slovak) court ruling becomes a valid upjs_ontology:court_decision, but ordinary judgments are not also instances of cdm:case-law_work. This preserves the integrity of the original CDM while allowing us to model our own attributes without conflicting with the basic model. Court of Justice of the European

Union judgments that are of the type cdm:case-law_work are therefore automatically displayed in our dataset as upjs_ontology:court_decision.

3.1. Ontological principles and inheritance

- **Upward extension** – we create superordinate (more general) classes/properties, thereby not interfering with existing CDM definitions.
- **Use of subClassOf and subPropertyOf** – new resources are classified into the existing hierarchy without disrupting it.
- **Consistent domain and range** – each subtype respects the constraints of its supertype.

3.2. Classification of new classes and properties used for the representation of Slovak court decisions

When designing the extension, some classes and properties from the original CDM model were taken over without change and used directly in the data. These include in particular:

- cdm:work_cites_work – to express citations between decisions,
- cdm:work_date_creation – for the date of creation of the decision,
- cdm:Region – for geographical classification,
- cdm:Judge – to represent judges,
- cdm:court_national – to designate national courts.

The remaining entities were designed as new ones or were derived from existing CDM elements, respecting the structure of the original model and ensuring the possibility of inference and interconnectivity with European data.

Prefixes used:

- **upjs_ontology: = https://court.ics.upjs.sk/ontology/** – for classes and properties,
- **upjs_resource: = https://court.ics.upjs.sk/resource/** – for individual instances

3.2.1. New classes

- **upjs_ontology:court_decision**
  - **URI:** https://court.ics.upjs.sk/ontology/court_decision
  - **Inheritance:** cdm:case-law_work ⊆ upjs_ontology:court_decision ⊆ cdm:Work ⊆ owl:Thing
  - **Description:** Covers all decisions in general, including Slovak courts, and case law remains a direct instance of cdm:case-law_work.
- **upjs_ontology:decision_nature**
  - **URI:** https://court.ics.upjs.sk/ontology/decision_nature
  - **Inheritance:** upjs_ontology:decision_nature ⊆ skos:Concept ⊆ owl:Thing

    - **Description:** Term for the attribute **nature of the decision**, e.g. "first instance not subject to appeal".
- **upjs_ontology:decision_form**
    - **URI:** https://court.ics.upjs.sk/ontology/decision_form
    - **Inheritance:** upjs_ontology:decision_form ⊆ skos:Concept ⊆ owl:Thing
    - **Description:** Formal type of document (attribute decision_form), e.g. "judgment" or "resolution".
- **upjs_ontology:legal_area**
    - **URI:** https://court.ics.upjs.sk/ontology/legal_area
    - **Inheritance:** upjs_ontology:legal_area ⊆ skos:Concept ⊆ owl:Thing
    - **Description:** Main legal area (attribute oblast_pravnej_upravy); linkable to EuroVoc, e.g. "civil law".
- **upjs_ontology:legal_sub_area**
    - **URI:** https://court.ics.upjs.sk/ontology/legal_sub_area
    - **Inheritance:** upjs_ontology:legal_sub_area ⊆ skos:Concept ⊆ owl:Thing
    - **Description:** A finer division of the legal area (attribute subarea_of_legal_regulation), e.g. "enforcement and execution of decisions".

3.2.2. New properties

- **upjs_ontology:ecli**
    - **URI:** https://court.ics.upjs.sk/ontology/ecli
    - **Type:** data property
    - **Inheritance:** cdm:case-law_ecli ⊆ upjs_ontology:ecli ⊆ cdm:ecli ⊆ owl:topDataProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** xsd:string
    - **Description:** ECLI identifier (attribute ecli), derived from the original **cdm:ecli**, which has the domain "cdm:case-law_work", but we needed the domain to be "upjs_ontology:court_decision"
- **upjs_ontology:docket_label**
    - **URI:** https://court.ics.upjs.sk/ontology/docket_label
    - **Type:** data property
    - **Inheritance:** upjs_ontology:docket_label ⊆ owl:topDataProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** xsd:string
    - **Description:** File reference number (attribute spisova_znacka).
- **upjs_ontology:docket_id_number**
    - **URI:** https://court.ics.upjs.sk/ontology/docket_id_number
    - **Type:** data property
    - **Inheritance:** upjs_ontology:docket_id_number ⊆ owl:topDataProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** xsd:string
    - **Description:** File identification number (attribute identifikacne_cislo_spisu).

- **upjs_ontology:decision_delivered_by_judge**
  - **URI:** https://court.ics.upjs.sk/ontology/decision_delivered_by_judge
  - **Type:** object property
  - **Inheritance:** cdm:case-law_delivered_by_judge ⊆ upjs_ontology:decision_delivered_by_judge ⊆ cdm:delivered_by ⊆ owl:topObjectProperty
  - **Domain:** upjs_ontology:court_decision
  - **Value domain:** cdm:Judge
  - **Description:** Links the decision to the judge (attribute sudca_guid, sudca_meno), originated from the original case law **cdm:case-law_delivered_by_judge**.
- **upjs_ontology:judge_delivers_decision**
  - **URI:** https://court.ics.upjs.sk/ontology/judge_delivers_decision
  - **Type:** object property
  - **Inheritance:** cdm:judge_delivers_case-law ⊆ upjs_ontology:judge_delivers_decision ⊆ cdm:delivers ⊆ owl:topObjectProperty
  - **Domain:** cdm:Judge
  - **Value domain:** upjs_ontology:court_decision
  - **Description:** Inverse direction (which decisions the judge issued), derived from the original case law **cdm:judge_delivers_case-law**
- **upjs_ontology:decision_delivered_by_court_national**
  - **URI:** https://court.ics.upjs.sk/ontology/decision_delivered_by_court_national
  - **Type:** object property
  - **Inheritance:** cdm:case-law_delivered_by_court_national ⊆ upjs_ontology:decision_delivered_by_court_national ⊆ cdm:delivered_by ⊆ owl:topObjectProperty
  - **Domain:** upjs_ontology:court_decision
  - **Value domain:** cdm:court_national
  - **Description:** The court issues a decision, originating from the original case law **cdm:case-law_delivered_by_court_national**
- **upjs_ontology:court_national_delivers_decision**
  - **URI:** https://court.ics.upjs.sk/ontology/court_national_delivers_decision
  - **Type:** object property
  - **Inheritance:** cdm:court_national_delivers_case-law ⊆ upjs_ontology:court_national_delivers_decision ⊆ cdm:delivers ⊆ owl:topObjectProperty
  - **Domain:** cdm:court_national
  - **Value domain:** upjs_ontology:court_decision
  - **Description:** Inverse direction to the previous property (attributes sud_guid, sud_nazov), originated from **the** original **case law cdm:court_national_delivers_case-law**
- **upjs_ontology:decision_originates_in_country**
  - **URI:** https://court.ics.upjs.sk/ontology/decision_originates_in_country

    - **Type:** object property
    - **Inheritance:** cdm:case-law_originates_in_country ⊆ upjs_ontology:decision_originates_in_country ⊆ cdm:originates_in ⊆ owl:topObjectProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** cdm:Country
    - **Description:** Country of origin of the decision, in Slovak decisions always "SVK", originated from the original case law **cdm:case-law_originates_in_country**
- **upjs_ontology:country_initiates_decision**
    - **URI:** https://court.ics.upjs.sk/ontology/country_initiates_decision
    - **Type:** object property
    - **Inheritance:** cdm:country_initiates_case-law ⊆ upjs_ontology:country_initiates_decision ⊆ cdm:initiates ⊆ owl:topObjectProperty
    - **Domain:** cdm:Country
    - **Value domain:** upjs_ontology:court_decision
    - **Description:** Inversion of the previous relationship, derived from the original case law **cdm:country_initiates_case-law**
- **upjs_ontology:court_located_in_district**
    - **URI:** https://court.ics.upjs.sk/ontology/court_located_in_district
    - **Type:** object property
    - **Inheritance:** upjs_ontology:court_located_in_district ⊆ cdm:located_in ⊆ owl:topObjectProperty
    - **Domain:** cdm:court_national
    - **Value domain:** cdm:Region
    - **Description:** Links a court to a district (attribute sud_okres).
- **upjs_ontology:district_under_jurisdiction_of_region**
    - **URI:** https://court.ics.upjs.sk/ontology/district_under_jurisdiction_of_region
    - **Type:** object property
    - **Inheritance:** upjs_ontology:district_under_jurisdiction_of_region ⊆ owl:topObjectProperty
    - **Domain:** cdm:Region (*district*)
    - **Value domain:** cdm:Region (*region*)
    - **Description:** Hierarchy district → region (derived from the attribute sud_kraj).
- **upjs_ontology:has_court_type**
    - **URI:** https://court.ics.upjs.sk/ontology/has_court_type
    - **Type:** object property
    - **Inheritance:** upjs_ontology:has_court_type ⊆ cdm:has_type ⊆ owl:topObjectProperty
    - **Domain:** cdm:court_national
    - **Value domain:** cdm:court-type
    - **Description:** Type of court (*district, regional, etc.*) – attribute sud_typ.
- **upjs_ontology:decision_has_nature**

    - **URI:** https://court.ics.upjs.sk/ontology/decision_has_nature
    - **Type:** object property
    - **Inheritance:** upjs_ontology:decision_has_nature ⊆ owl:topObjectProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** upjs_ontology:decision_nature
    - **Description:** Assigns the nature of the decision (attribute povaha_rozhodnutia).
- **upjs_ontology:decision_has_form**
    - **URI:** https://court.ics.upjs.sk/ontology/decision_form
    - **Type:** object property
    - **Inheritance:** upjs_ontology:decision_has_form ⊆ owl:topObjectProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** upjs_ontology:decision_form
    - **Description:** Specifies the form of the decision (attribute forma_rozhodnutia).
- **upjs_ontology:decision_has_legal_area**
    - **URI:** https://court.ics.upjs.sk/ontology/decision_has_legal_area
    - **Type:** object property
    - **Inheritance:** upjs_ontology:decision_has_legal_area ⊆ owl:topObjectProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** upjs_ontology:legal_area
    - **Description:** Main legal area (attribute oblast_pravnej_upravy).
- **upjs_ontology:decision_has_legal_sub_area**
    - **URI:** https://court.ics.upjs.sk/ontology/decision_has_legal_sub_area
    - **Type:** object property
    - **Inheritance:** upjs_ontology:decision_has_legal_sub_area ⊆ owl:topObjectProperty
    - **Domain:** upjs_ontology:court_decision
    - **Value domain:** upjs_ontology:legal_sub_area
    - **Description:** Sub-area according to the attribute podoblast_pravnej_upravy.

4. The resulting dataset

The final unified graph of European and Slovak decisions (containing metadata about decisions without the text content of the decisions) in the structure described above in the form of an RDF graph with RDFS interference can be found in the model.rdf file.