



# LEGAL DOCUMENT MANAGEMENT: AN INTEGRATED APPROACH

Stevan Gostojić\*,  
Marko Marković

Faculty of Technical Sciences,  
University of Novi Sad,  
Novi Sad, Serbia

## Abstract:

In this paper, we introduce the main legal document management concepts. Furthermore, we review legal document types and formats, legal document life-cycle phases, and legal document management systems components implementing those phases. Several successful cases studies of legal document management are identified. Based on the review and the case studies, we identify several issues in legal document management and give recommendations on how to overcome those issues. The recommendations come down to using standardized document and metadata formats, publishing documents and metadata using open data recommendations if possible and using standards-abiding software.

## Keywords:

legal document management, legal document formats, legal document life-cycle, semantic documents, open access.

## 1. INTRODUCTION

According to [1], the most widely used document management standard, a document is information that is identified, structured, processed, controlled and interchanged as a unit. Therefore, a legal document can be defined as a document containing legal information (i.e. legal norms and legal facts). For more information about legal documents management, see [2].

In most contemporary legal systems, the government is divided into three branches: the legislative (i.e. the parliaments), the executive (i.e. the public administration) and the judicial (i.e. the courts of law). Therefore, legal documents can be broadly divided into legislative, executive and judicial documents (although other types such as contracts do exist) although other types of legal documents such as contracts also exist. In this paper, we focused on legislative documents as the most complex legal document type.

A document management system (DMS) is an information system that manages documents through their life-cycle by supporting document structural types, metadata assignment, version control, business process definition, execution and monitoring, document indexing, retrieval and browsing, user collaborations, security management and

## Correspondence:

Stevan Gostojić

## e-mail:

gostojic@uns.ac.rs



integration with other information systems. Documents are inputs and outputs of business processes consisting of a sequence of activities according to particular procedural rules.

In recent years, several legal document management systems, information systems that are specialized to manage legal document life-cycle, have been developed and came into operational use. Those systems are used by the government, legal offices and legal departments in large companies.

We review the application of ICT to legal document management in local self-governance, territorial autonomy and state bodies offering G2C (government to citizens), G2B (government to businesses) and G2G (government to government) services. The digitization of business process implementing those services can increase the effectiveness and the efficiency of business processes and saves time and money to citizens, business, and public officials participating in them.

The rest of this paper is structured as follows. In the next section, we review legal document formats (the static aspect of legal documents). Legal document life-cycle (the dynamic aspect of legal documents) is introduced in the third section. The fourth section explicates several successful case studies of legal document management. The paper is summarized and recommendations are given in the concluding section.

## 2. DOCUMENT FORMATS

There are numerous formats developed for representing legal documents. In [3] generations of legal document formats are recognized as a categorization based on characteristics of these formats.

The first generation consists of XML standards that are focused on text and structure of legal documents. This approach is similar to data presentation using database entities and typographic elements.

The second generation standards introduce a document model and descriptions of text, structure, and metadata. However, these layers of a legal document are loosely interconnected.

Document standards from the third generation are based on patterns and clearly differentiate layers of text, structure, metadata, and ontology.

Besides patterns, the fourth generation applies additional constraints using grammars (e.g. RELAX NG, Schematron, DSD) to ensure prescriptiveness in document model.

Legal semantic web [4] brings the idea of the semantic web application to the legal domain. This perspective distinguished several layers of legal semantic web technologies. The first layer enables the identification of legal resources. The second layer structures legal documents using XML. The third layer allows for creating assertions about legal documents. The fourth layer introduces legal ontologies connected with general ontologies. The fifth layer gives a formal representation of legal norms.

### *Legal Document Identifiers*

For identification of legal resources, there are several formats available.

CELEX numbers [5] are initially introduced by CELEX database for identification of legal documents in European Communities. CELEX number consists of four parts representing the sector, year of publication, document type, document number [6]. CELEX numbers lack support to embed elements of a national identification system.

European Legislation Identifier (ELI) [7] provides a mechanism for the identification of legislative documents. It is based on Uniform Resource Identifier (URI) and besides identification purposes, ELI offers a set of metadata and an ontology to facilitate information interchange between member states.

European Case-Law Identifier [8] provides a standard for identification of court decisions of EU member states. ECLI consists of five elements: ECLI keyword, country code, national court code, the year of decision and decision unique number. ECLI format is able to embed national coding of court decisions. In addition, ECLI provides a minimum set of metadata for decisions to facilitate searchability of case law.

URN:LEX [9] is proposed as a convention for identifying, naming, assigning, and managing persistent resources in the legal domain. It is applicable to different types of legal documents including legislation, case law, regulations, bills, etc. Extensibility of URN:LEX enables support for encapsulation of a national identification system.

### *Legal Document Metadata*

Metadata has an important role in the description of legal documents. Several standards listed below provide metadata sets and legal ontologies.



LKIF Core Ontology [10] is built for describing basic legal concepts. It consists of three layers: top level, intentional level, and legal level. The top-level defines mental, physical and abstract concepts, and occurrences. The intentional level defines concept and relations for describing the behavior of rational agents. The legal layer contains concept definitions for expressing normative statements. For applicability in a particular legal domain, besides LKIF Core ontology as a core ontology, a domain ontology is required.

Judicial Ontology Library (JudO) [11] is an ontology enabling the representation of knowledge for reasoning and argumentation in judgments. JudO consists of two modules: a core ontology and a domain ontology. JudO core ontology describes general legal concepts derived from LKIF-Core ontology. JudO domain ontology represents concepts and rules from the Italian Civil Code and Consumer Code and from related judgments. JudO ontology is a helpful resource for judgment annotation but has its limitations when modeling exceptions to some legal rule.

CEN Metalex provides a vocabulary of metadata for describing legal resources using RDF statements. Concepts are defined using OWL schema [12]. It distinguishes three basic classes: BibliographicThing, Event, and EventParticipant. CEN Metalex XML Schema allows integration of RDF statements and thereby secures that retrieved legal resources and corresponding metadata are up-to-date.

FOLaw is an ontology specifying functional dependencies between knowledge [13]. FOLaw recognizes six types of knowledge naming normative knowledge (typical legal knowledge), world knowledge (legal view on a real-world domain), responsibility knowledge (connection between norms and violation consequences), reactive knowledge (sanctions for a violating behavior), creative knowledge (creation of fictive legal entities), and meta-legal knowledge (resolves inconsistencies between legal rules).

LRI-Core ontology [14] connects common sense concepts with legal concepts. It consists of two layers: foundational ontology and legal core ontology. Foundational or upper ontology focuses on five categories of concepts: physical, mental, abstract, roles and occurrences. Core layer provides support for the development of legal domain ontologies.

To enable interoperability of legal information system in member states ELI standard gives a set of legislative metadata. The metadata set is divided into groups of metadata relevant for work level, expression level, and

manifestation level. Some of the work level metadata are an identifier, type of legislation, subject, authority, and publication date. Expression level metadata contains a language, title, description, etc. Metadata at the manifestation level are a URI of expression, link to file, and publisher.

ELI Ontology [15] defines a model of metadata to enable the description of legal resources and their sharing over the web. It is based on existing vocabularies e.g. Functional Requirements for Bibliographic Records (FRBR) model, Simple Knowledge Organisation System (SKOS) and Dublin Core [16].

ECLI standard gives a minimal set and an optional set of metadata that should be assigned to court decisions for their retrieval using search engines. The metadata is based on the Dublin Core Metadata Initiative. Some of the metadata from the minimal set are an identifier, creator, date, language, publisher, type, etc. Among optional metadata are a title, subject, abstract, description, references, etc.

### *Legal Document Content*

Many document formats are available for content representation of legal documents. Most of them are based on XML format. Some of them that belong to the third generation of legal document standards are described below.

MetaLex [17] is a legal document interchange format. MetaLex proposes a document model, a metadata set and an ontology for expressing metadata. It is also called the lowest common denominator for other legal document standards. MetaLex becomes the CEN MetaLex standard, proposed by CEN standardization committee [18]. CEN MetaLex is built on experience collected from implementations of Norme In Rete and Akoma Ntoso standards.

Akoma Ntoso [19] project managed by United Nations Department of Economic and Social Affairs (UN/DESA) [20] offers a standard for representing parliamentary, legislative and judiciary documents in XML format. Akoma Ntoso XML schema provides building blocks for structural and semantic markup of legal documents. Besides several supported types of legal documents, Akoma Ntoso XML schema provides a generic type of document enabling representation of custom document types.

LegalDocML format [21] proposed by OASIS LegalDocML Technical Committee [22] provides an XML



format called LegalDocumentXML. It defines a legal document format for parliamentary, legislative and judicial documents. It is based on Akoma Ntoso standard adopting Akoma Ntoso Version 1.0 specification along with Akoma Ntoso naming conventions.

United States Legislative Markup (USLM) [23] is an XML schema designed to support the representation of legislative documents in the US. USLM is modeled to maximize compatibility with Akoma Ntoso standard. Therefore, USLM shares many similarities with Akoma Ntoso Schema, particularly in the naming of elements and attributes. USLM provides two document models: abstract and concrete models. The abstract model is a general model composed of basic tags, while the concrete model is a derivation of abstract model introducing specific terminology.

LegalRuleML is standard proposed by OASIS LegalRuleML Technical Committee [24] as a rule interchange language for legal domain. It is based on RuleML language [25] extending its capabilities to support features of legal rules. Extension of RuleML is performed using two modules, one for legal metadata and second is for legal operators i.e. deontic operators and operators for managing behaviors (violation and reparation).

### 3. DOCUMENT LIFE-CYCLE

Document life-cycle is “a period from the conceptual idea to the logical and physical deletion of a document” [1]. A document goes through several life-cycle phases: initiation, preparation, establishment, use, revision, withdrawal, and deletion. During the initiation phase, a document is identified and classified according to some classification scheme. The document content is created during the preparation phase. In the establishment phase, the document content is verified and the document is available for active use. During the revision phase, the document content is changed. This implies that the document has to pass through the preparation, establishment and use phases again. When a document is no longer actively used, it is withdrawn from the system. Usually, it remains in the system in the archived state. Finally, depending on the document type and the system rules, the document is completely (physically) removed from the system.

We established relationships between the activities executed on legislative documents (i.e. enactment, promulgation, publication, coming into force and repealing) and legal document life-cycle phases (i.e.

initiation, preparation, establishment, use, revision, withdrawal, and deletion). Those relationships were established in order to ease interoperability between components and systems.

The legislation is initialized and prepared by its sponsors (or their representatives) in the form of a bill. From that point onward, the content of legislation can be changed with amendments only.

The establishment life-cycle phase corresponds to enactment, promulgation, publication and coming into force of legislation. As a consequence of the enactment phase, the content of the legislation is created in the prescribed form, by the prescribed subject in accordance with the prescribed procedure. The publication of legislation is an activity that results with the content of legislation being communicated to the subjects it applies to.

The legislation is used from the moment it comes into force to the moment it is repealed (or even after it has been repealed in some cases). Coming into force is an activity that includes legal norms contained in the legislation into a legal system and the repealing is an opposite activity that excludes legal norms from a legal system. In the use phases, lawyers and citizens retrieve and browse legislation and apply legal norms contained in it to concrete legal cases.

The withdrawal phase corresponds to the repealing.

The deletion phase is usually not relevant to legislation because legislation is usually not deleted, but is archived for future use after it has been repealed.

The revision phase corresponds to amending. The amending procedure is similar to enactment procedure.

According to [26], the main components of a legislative document management systems are an editor, a converter, a name resolver, a validator, and post-editing software tools.

An editor is a fundamental tool for the generation of legislation in a standardized XML format used as an interface to activate, control and verify the automatic conversion tool, as a tool to manually mark-up a document, and as a tool for direct insertion of both text and markup.

A converter works by examining the typographical and textual regularities of a document in the legacy format and converting that document to a standardized XML format.

A validator checks if a document complies with the requirements.

Post-editing tools enable indexing, retrieval, and browsing of legislation, business process management, transformation to other formats, etc.



Therefore, editors, converters, and validators are used to implement initiation, preparation, establishment, and revision life-cycle phases and post-editing tools are used to implement use, withdrawal, and deletion life-cycle phases. Several legal document management system components, that entered into operational use, are presented in the next section.

## 4. CASE STUDIES

Most legal document management systems do not use standardized legal document formats or components. Thus, proprietary formats and software prevail. In this section, we review several (components of) legal document management systems that can be considered case studies of standards-abiding integrated document management approach.

According to ELI Implementation Overview [27], 12 countries currently adopted ELI standard. However, in many other countries, there are portals introducing ELI standard. For example, the official portal for publishing legislation in Serbia PIS [28] proposed the identification of legislative documents using ELI.

ECLI identifier for case law is implemented in the Court of Justice of the European Union, the European Court of Human Rights, the Boards of Appeal of the European Patent Office and 12 Member States [29]. Furthermore, ECLI implementation in seven other EU countries is in progress.

Language Independent Markup Editor called LIME [30] is a web-based document editor enabling markup of nonstructured documents. Using appropriate plugin different document formats are supported. Currently, there are entry points for creating documents in Akoma Ntoso format and LegalRuleML format. LIME is an editor and a validator according to the classification given in the previous section.

Indigo Platform [31] is a document management system for legislative documents. Indigo is developed as a web application providing an environment for managing legislation to end users. Additionally, Indigo implements REST API enabling developers to use legislative documents in their projects. Akoma Ntoso XML format has been chosen for representation of document content and metadata. Indigo manages indentation for subsections and automatically generates the document cover page and table of contents. Still, only the Akoma Ntoso act document type is supported. Indigo Platform is an editor, a converter, and a validator.

In 2013, the UK started the LDAPP project for development of legislative drafting, amending and publishing tools [32]. The project is based on open standards and open source technologies and it has entered in beta phase. Akoma Ntoso has been chosen as XML markup language. Existing API for accessing UK legislation is extended to support Akoma Ntoso format. Government bodies participating in this project are the two Houses of the UK Parliament (the House of Lords and the House of Commons), the Scottish Parliament, the Office of the Parliamentary Counsel, the Scottish Government's Parliamentary Counsel Office, and The National Archives. According to the description given in [32], LDAPP implements all document life-cycle phases.

MetaLex Document Server (MDS) [33] provides access to Dutch regulation in CEN MetaLex format. The document metadata is available as RDF triples in accordance with MetaLex ontology. Document identifiers are based on URI patterns used by UK legislation portal. Initially, a document set from legacy regulations database is transformed from proprietary document format to CEN MetaLex XML format and RDF format. For each document, MDS offers three manifestations, XML, RDF and HTML. MDS is a post-editing tool.

LexML project in Brasil was started to provide a platform for collecting and enabling access to legal documents from executive, legislative and judiciary branch [34]. Portal was launched in 2009 and it provides documents produced by 15 government bodies. Identification of legal documents is based on URN:LEX standard and documents are represented using XML format similar to Akoma Ntoso Schema. LexML is a post-editing tool.

Authoring Tool for Amendments (AT4AM) is a web-based authoring tool for amendments [35]. It enables users to create and table amendments and is used by the European Parliament for amending legislation. On the basis of AT4AM another open source tool called AT4AM for All has been developed. Both tools rely on Akoma Ntoso XML schema as a document format. As an amendment management system, AT4AM implements all phases of legal document life-cycle.

## 5. CONCLUSIONS

In this paper, we reviewed legal document types and formats, legal document life-cycle phases, and software components that support legal document management and presented several case studies of legal document management systems that we identified as examples of good practice.



Based on the review of legal document formats, legal document management system components and case studies, we propose several recommendations that can advance legal document management.

Firstly, we recommend the adoption of a standardized legal document format. That format should support document identification, metadata assignment, and representation of both the content and the form of documents. It should be based on W3C recommendations and specialization of those recommendations in the legal domain.

Secondly, we recommend publishing legal documents according to open linked data recommendations if privacy and confidentiality considerations do not dictate otherwise. Attention should be paid to the technical and legal openness of published documents and metadata. Documents and metadata should be linked to other documents and metadata to promote the network effect.

The third recommendation is that each public body should be allowed to introduce standards-abiding (components of) legal document management systems on their own. However, one public body should offer hosting, maintenance and support for the generic implementation of a legal document management system to public bodies without the resources necessary for an independent operation of the system.

In the future, we plan to investigate how similar principles can be applied to executive and legislative documents.

As proposed in [36], the semantic document approach can be applied to the legal domain. The semantic representation of legal documents is the first and necessary step of the approach, continuing with the semantic modeling of metadata (the implementation of legislative metadata layer) and ending with the semantic enrichment of legislative document workflow. This approach can facilitate easier implementation of domain-specific services.

## REFERENCES

- [1] International Organization for Standardization (ISO), ISO IEC 82045-1: Document Management – Part 1: Principles and Methods, ISO, Geneva: International Organization for Standardization (ISO), 2001.
- [2] S. Gostojić, Stvaranje i primena pravnih akata tehnikama semantičkog veća, Novi Sad: Fakultet tehničkih nauka, 2017.
- [3] M. Palmirani and F. Vitali, “Akoma-Ntoso for Legal Documents,” *Legislative XML for the Semantic Web*, pp. 75–100, 2011.
- [4] G. Sartor, “Legislative Information and the Web,” *Legislative XML for the Semantic Web*, pp. 11–20, 2011.
- [5] M. V. Opijnen, *The European Legal Semantic Web: Completed Building Blocks and Future Work* on European Legal Access Conference, November 22, 2012.
- [6] European Union. EUR-Lex, *Frequently Asked Questions*, 2018. [Online] Available: <http://eur-lex.europa.eu/content/help/faq/intro.html> [Accessed March 2019]
- [7] Council of the European Union. *Council conclusions inviting the introduction of the European Legislation Identifier (ELI)*. CELEX:52012XG1026(01), 2012.
- [8] The Council of the European Union, “Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law”, *Official Journal C 127 of 29.4.2011*, [Online]. Available: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:127:0001:0007:EN:PDF> [Accessed March, 2019]
- [9] P. Spinosa, E. Francesconi, and C. Lupo, “A Uniform Resource Name (URN) Namespace for Sources of Law (LEX)”, [Online]. Available: <https://tools.ietf.org/html/draft-spinosa-urn-lex-13>. [Accessed March, 2019]
- [10] R. Hoekstra, J. Breuker, M. Di Bello, and A. Boer, “The LKIF Core Ontology of Basic Legal Concepts”, *LOAIT*, 321, 43-63., 2019
- [11] M. Ceci and A. Gangemi, “An OWL ontology library representing judicial interpretations,” *Semantic Web*, vol. 7, no. 3, pp. 229–253, 2016.
- [12] CEN, *CEN MetaLex OWL Ontology*, (2019). [Online]. Available: <http://www.metalex.eu/metalex-cen.owl> [Accessed March, 2019]
- [13] T. V. Engers, A. Boer, J. Breuker, A. Valente, and R. Winkels, “Ontologies in the Legal Domain,” *Digital Government Integrated Series In Information Systems*, pp. 233–261, 2008.
- [14] J. Breuker, R. Hoekstra, *Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law*, in: Proceedings of EKAW Workshop on Core ontologies, CEUR, 2004.
- [15] [Online]. Available: <https://publications.europa.eu/en/web/eu-vocabularies/eli> [Accessed March, 2019]
- [16] ELI Task Force, *ELI a technical implementation guide*. Publications Office, 2018
- [17] A. Boer and R.G.F. Winkels, “CEN MetaLex: Facilitating interchange in e-Government”, *Government and Electronic Participation: Joint Proceedings of Ongoing Research and Projects of IFIP EGOV and ePart 2011*, pp. 240-248, 2011.



- [18] European Committee for Standardization, [Online]. Available: <https://www.cen.eu/> [Accessed March, 2019]
- [19] Akoma Ntoso, [Online]. Available: <http://www.akomantoso.org/> [Accessed March, 2019]
- [20] United Nations Department of Economic and Social Affairs, [Online]. Available: <https://www.un.org/development/desa/en/> [Accessed March, 2019]
- [21] OASIS, *Akoma Ntoso Version 1.0. Part 1: XML Vocabulary*, 2019. [Online]. Available: <http://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0-part1-vocabulary.html> [Accessed March, 2019]
- [22] OASIS, *LegalDocumentML TC*, [Online]. Available: <https://www.oasis-open.org/committees/legaldocml> [Accessed March, 2019]
- [23] United States Legislative Markup: User Guide for the USLM Schema, [Online]. Available: <http://uscode.house.gov/download/resources/USLM-User-Guide.pdf> [Accessed March, 2019]
- [24] OASIS LegalRuleML TC, [Online]. Available: <https://www.oasis-open.org/committees/legalruleml/> [Accessed March, 2019]
- [25] [Online]. Available: <http://www.ruleml.org> [Accessed March, 2019]
- [26] F. Vitali, "A Standard-Based Approach for the Management of Legislative Documents," in *Legislative XML for the Semantic Web. Law, Governance and Technology Series*, vol. 4, G. Sartor, M. Palmirani, E. Francesconi and M. Biasiotti, Eds. Dordrecht: Springer, 2011.
- [27] ELI Implementation Overview, [Online]. Available: <https://eur-lex.europa.eu/eli-register/about.html> [Accessed March, 2019]
- [28] Pravno Informacioni Sistem, [Online]. Available: <http://www.pravno-informacioni-sistem.rs> [Accessed March, 2019]
- [29] ECLI Current Implementation, [Online]. Available: <http://bo-ecli.eu/ecli/current-implementation>. [Accessed March, 2019]
- [30] LIME, [Online]. Available: <http://lime.cirsfid.unibo.it/> [Accessed March, 2019]
- [31] Indigo Platform, [Online]. Available: <https://indigo.readthedocs.io/en/latest/index.html> [Accessed March, 2019]
- [32] The National Archives, [Online]. Available: <http://www.legislation.gov.uk/projects/drafting-tool> [Accessed March, 2019]
- [33] R. Hoekstra, "The MetaLex Document Server," *The Semantic Web – ISWC 2011 Lecture Notes in Computer Science*, pp. 128–143, 2011.
- [34] All in one place: Brazilian Legal and Legislative Information Portal, [Online]. Available: <http://projeto.lexml.gov.br/documentacao/resumo-em-ingles>. [Accessed March, 2019]
- [35] K. Malanga, J. Mehat, I. Ganchev, J. Wandeto, and C. Shikali, "Evaluation of Open Source Software with QualiPSO OMM: a case for Bungeni and AT4AM for All," presented at the Free and Open Source Software Conference FOSSC-15, 2015, pp. 41–46.
- [36] G. Sladić, I. Cverdelj-Fogaraši, S. Gostojić, G. Savić, M. Segedinac, and M. Zarić, "Multilayer document model for semantic document management services," *Journal of Documentation*, vol. 73. pp. 803-824, 2017