Requirements Specification for EUR-Lex Legal Analysis Methodology Model

Eugeniu Costetchi

19/07/2019

Table of Contents

Chapter 1. Introduction

This document provides the preliminary requirements specification for developing an ontology for the EUR-Lex Legal Analysis Methodology.

1. Context

EUR-Lex legal analysis methodology (LAM) presents and describes the use of metadata elements that are relevant for the legal and documentary analysis of the EUR-Lex website's content.

The metadata elements employed in LAM are taken from the Common Data Model (CDM) of the CELLAR repository of the Publications Office.

CELLAR is an electronic database which contains the documents and their related metadata diffused on one of the websites of the Publications Office. The CDM is an ontology that describes the concepts and relationships (properties/elements) that can exist for the data stored in the CELLAR.

LAM documentation contains descriptions of classes of legal documents and a selection of metadata suitable for describing each document class. LAM aims at facilitating the understanding and the use of relevant CDM properties.

LAM gives some basic Definition for the metadata elements, determines their cardinality and lists the related properties. It also gives some methodological rules concerning the use of the elements in different contexts during the legal analysis. It also describes which kind of data has to be used when filling in the metadata elements. If a metadata element has to be completed with a value coming from a controlled vocabulary, it is indicated. If there is no indication, it means that the metadata element can be filled in with free text.

2. Purpose

This document aims at analysing and formulating the requirements of the LAM team regarding the data model mainly and potential future applications. It also provides an approach for transposing the LAM from a plain text documentation into structured data with semi-formal and formal underpinning.

The benefit of having the LAM represented in a structured form is enabling automation of multiple processes, such as document classification, metadata validation, metadata enrichment, which currently are performed manually by the OP staff or by external contractors. Such an automation can lead to significant reductions of cost and reduce the time needed for performing these processes.

This document does not intend to provide a detailed functional specification for every envisaged element but rather provide a general direction and describe the path for reaching different business and technical objectives. The main concern, at this stage is to describe how a LAM ontology can be created starting from the current state of affairs presented elsewhere in this document.

3. Document Conventions

Application profile - ...

LAM model …

LA team, LA sector - OP C2 003

Model …

Ontology ...

Chapter 2. Context of the project

1. LAM business context

This section briefly explains the general context and the processes using and impacted by LAM. This description reflects the situation to date and may serve as a basis for deriving improvements.

The process starts with the legal document (LD) being published in CELLAR. Then the legal analysis team receives an XML Notice and access to the HTML content. At this stage, the notice contains a minimal set of metadata which may or may not be correct.

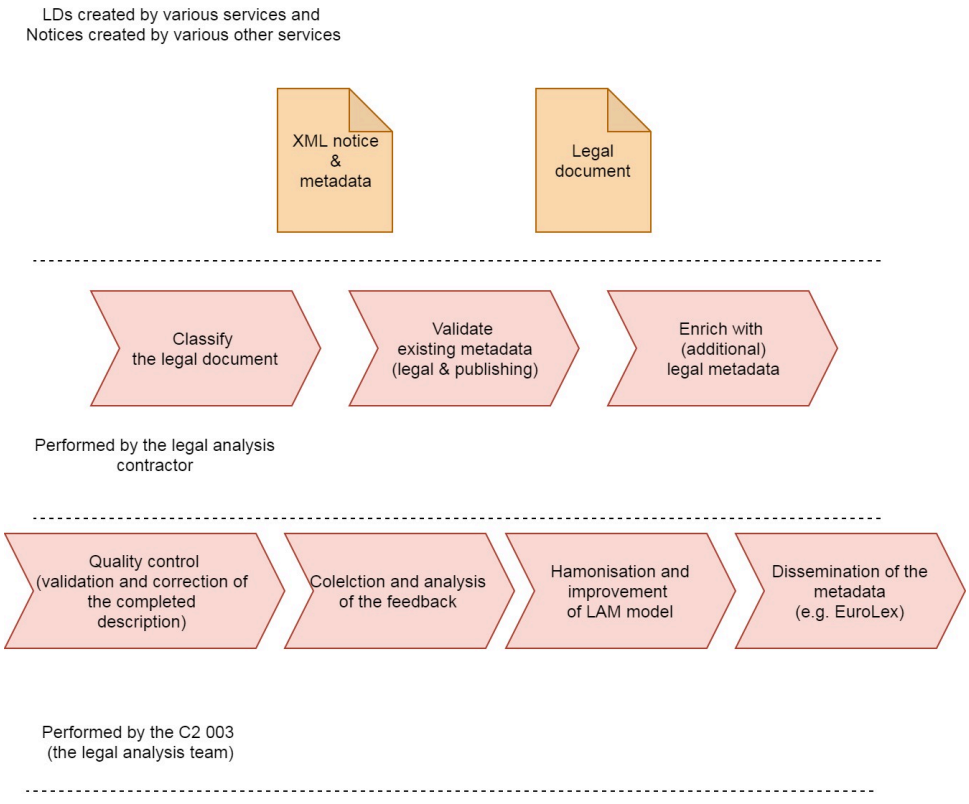
The goal is to (a) verify the correctness of the existing metadata, (b) classify the document according to LAM methodology (c) enrich the document with the corresponding legal metadata.

The document classification, for instance, is performed by considering the structure of the document title (i.e. presence/absence of keywords), structure of the CELEX number (if present), author and other metadata. For example, the title which contains string "communication of the commission" required that the author is the European Commission (EC) and is classified as a certain LAM document class.

Enrichment of the corresponding legal metadata is manually performed under the guidance of the LAM documentation.

The diagram below depicts operations performed using the published Legal Document and the XML notice & metadata. Document classification, validation of publishing metadata and enrichment with legal metadata are performed by an external contractor. The legal analysis team further performs a quality control of the legal document and XML notice and metadata and ensure the dissemination of the metadata dissemination on EuroLex website and other channels.

Figure 2-1. Digital assets and operations related to legal analysis methodology



In addition, the legal analysis team is the LAM owner and main editor. This role implies collecting feedback from various stakeholders and partners; initiating projects to harmonise and improve the quality of LAM itself.

2. LAM stakeholders

This section provides an overview on the main stakeholders of the legal analysis methodology.

- Legal analysis sector of the European Publications Office
  - LAM (model) owner
  - LAM (model) editor and maintainer/manager
- Legal analysis contractor
  - instance data creator = LAM model user
  - feedback provider (of the LAM model suitability as they deal with the practical cases and concrete document instances. Therefore they need some sort of user friendly interface to LAM to report mistakes, inconsistencies, contradictions, incomplete or ambiguous descriptions etc.)
- GIL-GM forum
  - discussion group of LAM model design decisions (open questions in the methodology, creation of a new document type, etc.)
- Publication agents and contractors
  - instance data creator (only for the documents they are mandated to create CLEX code) = LAM model user
  - responsible for publishing OJ, therefore they must follow / check the metadata constraints are respected

Chapter 3. Towards LAM model

The modelling of LAM should be approached in two steps.

1. Challenges

unstructured data

need to have an intuitive and easy to use authoring for LAM model

2. Technologies

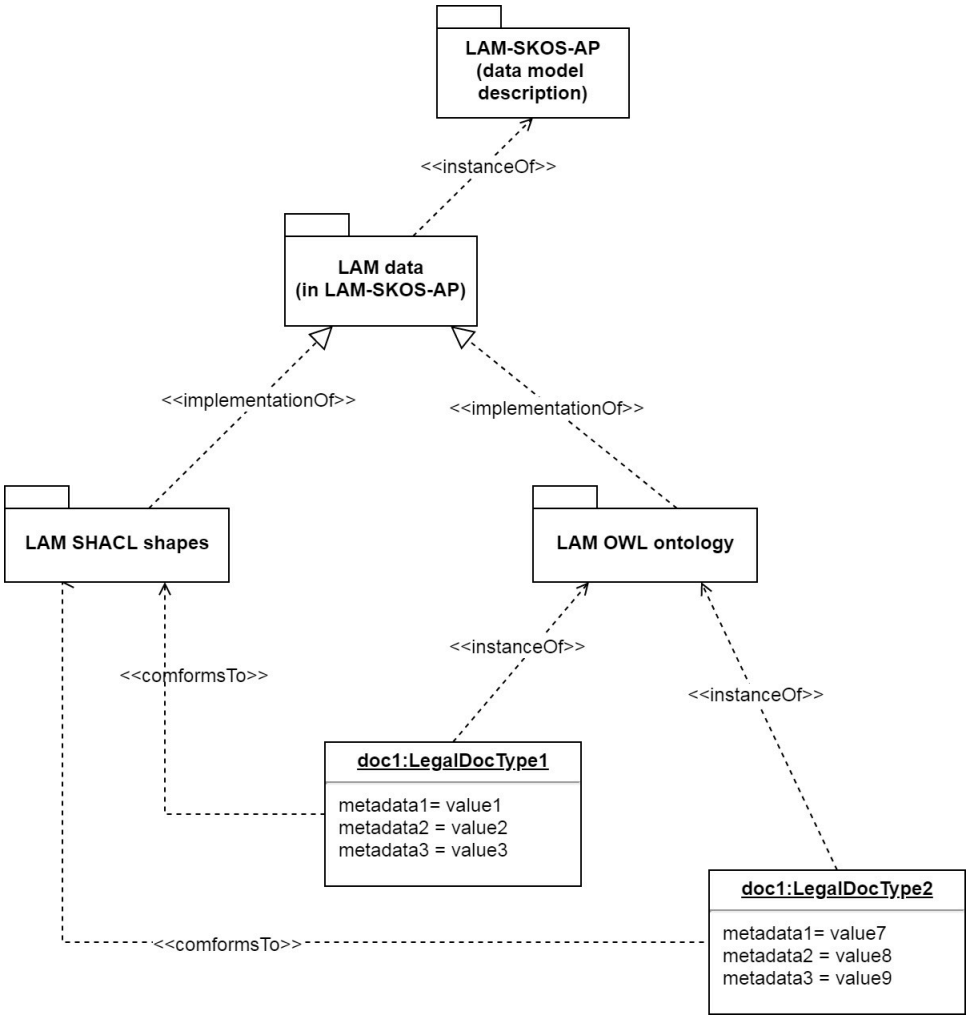
The Semantic Web technological stack is recommended for this project.

3. Modelling stack

LAM ontology should describe classes of documents and (application) profiles for each of the classes.

Since the LAM model is insatiable then the model for describing LAM model is a meta-description for the legal document descriptions.

Figure 3-1. LAM modeling stack



Chapter 4. Use cases

There are several use cases motivating the current modelling exercise. This section briefly briefly describes them.

Maintenance of the LAM description

Currently LAM is maintained as a set of Word and Excel documents, this makes the editing cumbersome and most importantly error prone. Making references to formal properties in CDM ontology, manually tracking rules and dependencies between LAM elements becomes increasingly difficult as it becomes larger. Moreover, sharing the documents, collecting input and the interaction with stakeholders and partners becomes difficult.

There is a need for providing an interactive documentation of Legal Analysis Methodology, which should include a complete description of document classes, properties, metadata and constraints; and enable an easy navigation based on the dependencies between them. This interactive documentation should also enable collecting feedback, corrections and suggestions for improvement on any part of the methodology. It should serve as the main point of access for the LAM for consultation purposes for both experts and lay people.

Modelling and structuring LAM description

Because currently LAM is described in purely informal manner it is not possible to automate or implement any automated processes relying on it. To enable automatisation of any sort the domain model must be created first and formalised in a machine readable format. The model provides a vocabulary for describing LAM concepts and the structural connections between them

Consistency checking of the LAM description

It is not easy to verify whether the text document

Curation of existing/legacy document descriptions

Assist/guide creation of the document descriptions

Automatic validation of the document descriptions

Index