Publications Office of the European Union

infeurope

Initiative for Modeling the Legal Analysis Methodology

Preliminary requirements specification for the Legal Analysis Methodology models

Deliverable WP 1.1

Eugeniu Costetchi

02 October 2019

Version 1.0

Disclaimer

The views expressed in this report are purely those of the Author(s) and may not, in any circumstances, be interpreted as stating an official position of the European Commission. The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission. All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscript s including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

Project acronym	LAM project
Project title	Initiative for Modeling the Legal Analysis Methodology
Document reference	Preliminary requirements specification for the Legal Analysis Methodology models
Author(s)	Eugeniu Costetchi
Editor(s)	Eugeniu Costetchi
Contractor	Infeurope S.A.
Framework contract	10688
Actual delivery date	02 October 2019
Delivery nature	Report (R)
Dissemination level	Public (PU)
Filename	wp1-1-ontology-requirements
Suggested readers	project partners, future users, legal practitioners, software architects

Abstract

This document aims to analyse and formulate the requirements of the EUR-Lex Legal Analysis team, focusing mainly on the data model and potential future applications. It also provides an approach for transposing the Legal Analysis Methodology (LAM) from a plain text document into structured data with semi-formal and formal underpinning.

Contents

1 Introduction 5
2 Context 5
3 Scope 7
4 Intended uses 8
4.1 Maintenance of the LAM description 8
4.2 Modeling and structuring LAM description 9
4.3 Consistency checking of the LAM description 9
4.4 Automatic validation and curation of the document descriptions 9
5 Modeling stack 10

1 Introduction

This document aims to analyse and formulate the requirements of the EUR-Lex Legal Analysis team, focusing mainly on the data model and potential future applications. It also provides an approach for transposing the Legal Analysis Methodology (LAM) from a plain text document into structured data with semi-formal and formal underpinning.

The benefit of having the LAM represented in a structured form is that it enables automation of multiple processes, such as document classification, metadata validation and 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 in the next section.

2 Context

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

EUR-Lex Legal Analysis Methodology methodology manual[1] 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.

[1] Publications Office, 2017, EUR-Lex legal analysis methodology, version 201706

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.

The process involving LAM starts with the legal documents (LD) 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 performs a further quality control check on the legal documents, XML notice and metadata, whilst also ensuring the metadata dissemination on the Eur-Lex website and other channels.

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 the LAM itself.

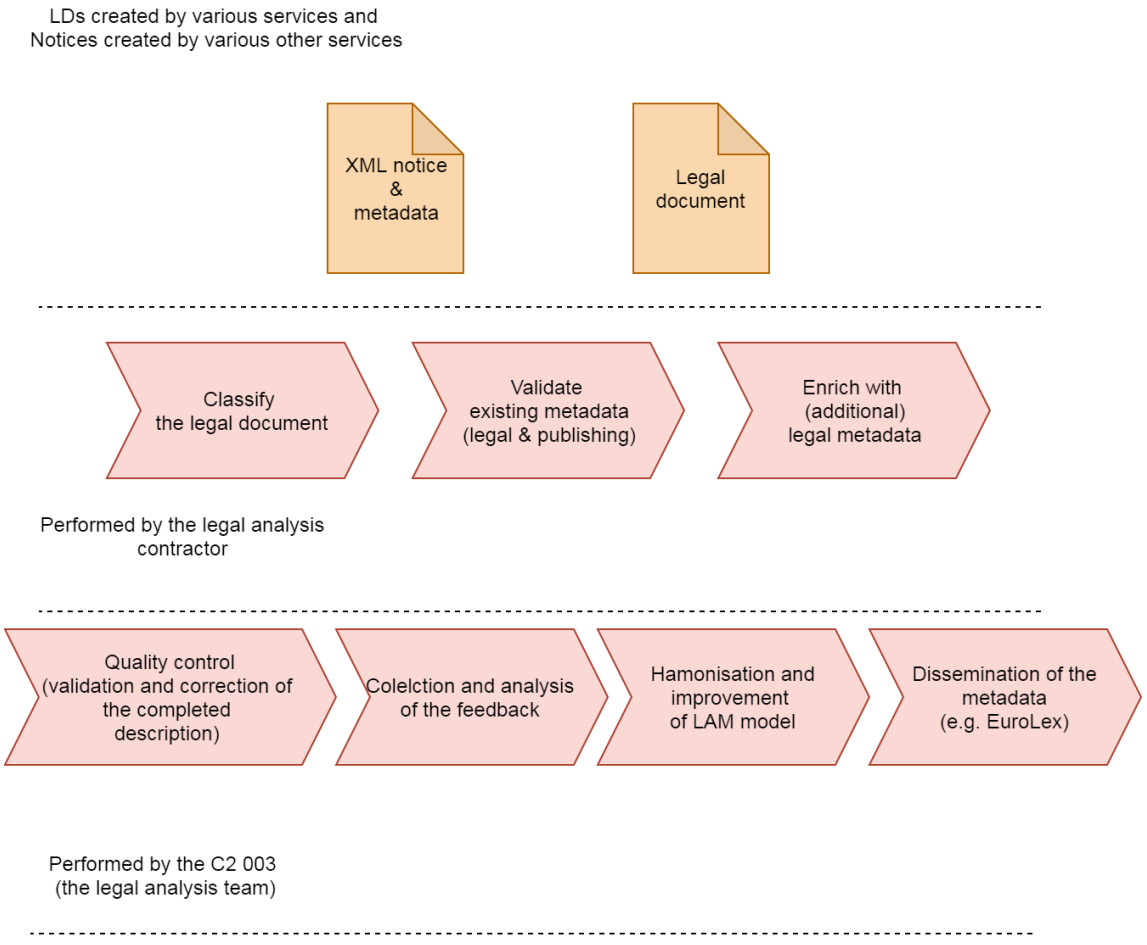

Figure 1: Digital assets and operations related to legal analysis methodology

3 Scope

The LAM model should cover similar degree of detail as in the current documentation. It should cover the following types of information.

- description of legal document classes
  - human readable definitions and descriptions
  - correspondences to other classifications (CELEX/ELI, CDM, Resource-

    Type etc.) as (a) mappings and as (b) code grammars
  - publication metadata
- description of document metadata properties
  - human readable definitions and descriptions
  - correspondences to formal properties defined in existing or new LAM model
- description of the metadata annotations (property annotations)
- description of LD classification rules (fields hinting at how to automatically/manually classify LDs)
- description of class shapes, i.e. constraints on how the class is instantiated
  - e.g. specific property value constraints
  - cardinality constraints,
  - range class constraints, etc.
- descriptions of LD classes and LD property groupings (table of contents in the LAM word document and templates in the LAM webApp description)

4 Intended uses

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

4.1 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 Legal Analysis Methodology document, 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, for both experts and lay people, as the main point of access to the LAM for consultation purposes.

4.2 Modeling and structuring LAM description

Currently LAM is described in an informal manner, therefore it is not possible to automate or implement any automated processes relying on it. To enable automation 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.

The modeling need is situated at two levels of abstraction or perhaps even two meta-levels. First, the LAM documentation covers descriptions of classes of legal documents, so there is a need to formalise these descriptions, lets call it Legal Document definition model. Second, there is a need to formalise how the descriptions of classes should be structured; let's call it the LAM meta-model.

4.3 Consistency checking of the LAM description

It is not easy to verify whether the document class is consistent with other classes and properties described in the model. The model should enable performing tests to determine if the data has any internal conflicts. The exact type of conflicts is not determined at this point and they may refer to redundancy, cycles, or contradiction detection.

4.4 Automatic validation and curation of the document descriptions

The legal document descriptions instantiating the LAM model should conform to model instantiation rules and fit specific data shapes. The model should enable automatic verification and validation of instance data.

In a similar manner, it should be possible to derive, for invalid documents, what parts need to be modified in order to correct the description.

5 Modeling stack

The approach to creating LAM ontology shall be performed in two steps. First the LAM model shall be expressed in a semi formal representation based on Simple Knowledge Organization System (SKOS) model[2] in order to enable the domain experts to maintain the content. Nonetheless, this content must already be sufficiently precise to support, in a second step, a transformation into a formal KR language such as Web Ontology Language (OWL)[3]. SKOS is a common data model for sharing and linking knowledge organization systems via the Web. The SKOS data model provides a standard, low-cost migration path for porting existing knowledge organization systems to the Semantic Web. SKOS also provides a lightweight, intuitive language for developing and sharing new knowledge organisation systems.

Depicted in Figure 2 is the process and the main intermediary steps to systematise the LAM knowledge down to a formal ontology. The diagram represents the flowchart for creating the LAM project deliverables, where each shape signifies as follows. Parallelograms (brown) represent static assets such as input and output data and models. The rectangles with an extra set of bars (blue) represent automatic processes executed by scripts, while arrow shaped rectangles (pink) represent manually executed processes. The assets which represent models (i.e. instantiable resources) are marked in bold.

The process starts from unstructured descriptions in the form of Word documents which are organised into tabular form and saved as Excel files. In accordance with a set of conventions recommended in Deliverable WP1.4[4], this step is performed by the LAM team, who are mainly legal analysis domain experts. This results in a semi-structured LAM representation based on the SKOS model. This is a bootstrapping step and is performed only once. Next, this data asset is transformed through a script into a knowledge base (KB), which will further be maintained using VocBench3

[2]Bechhofer, S., & Miles, A. (2009). Simple Knowledge Organization System Reference. `https://www.w3.org/TR/skos-reference/`

[3]Parsia, B., Motik B., & Patel-Schneider P. (2012). OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. `https://www.w3.org/TR/owl2-syntax/`

[4]Costetchi E., 2019, The structure of Excel workbook for bootstrapping the Legal Analysis Methodology descriptions

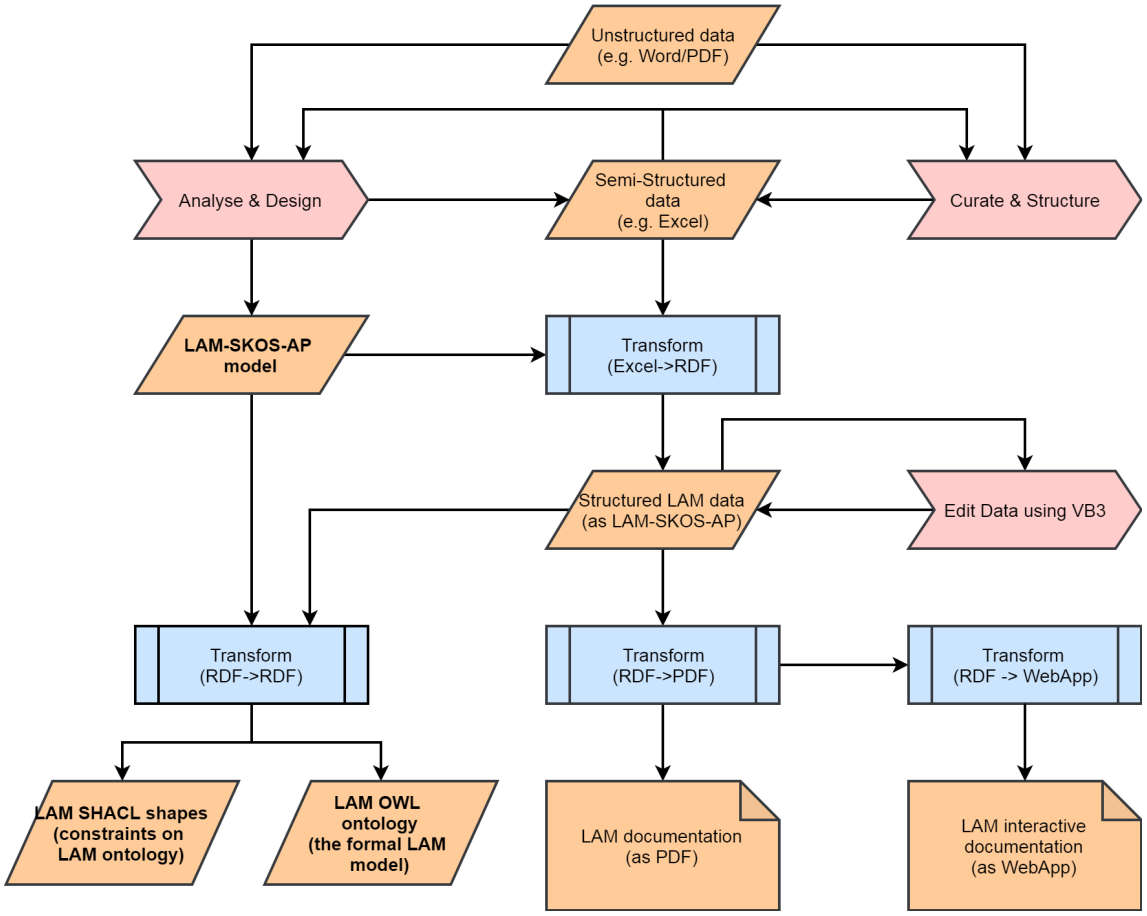

Figure 2: Approach to create LAM project deliverables

tool[5]. VocBench3 is a web-based, multilingual, collaborative development platform for managing OWL ontologies, SKOS(XL) thesauri and generic RDF datasets.

As the SKOS model is fairly simple it does not cover the needs of LAM project; therefore a SKOS extension should guide the structuring of LAM data in VocBench3. This extension should take the form of a SKOS based application profile (AP), which we here call LAM-SKOS-AP. An Application Profile (AP) is a specification that re-

[5]Stellato, A., Fiorelli, M., Turbati, A., Lorenzetti, T., Van Gemert, W.,Dechandon, D., Laaboudi-Spoiden, C., Gerencser, A., Waniart, A.,Costetchi, E., and Keizer, J. (forthcoming). VocBench 3: a Collaborative Semantic Web Editor for Ontologies, Thesauri and Lexicons.Semantic Web journal. link

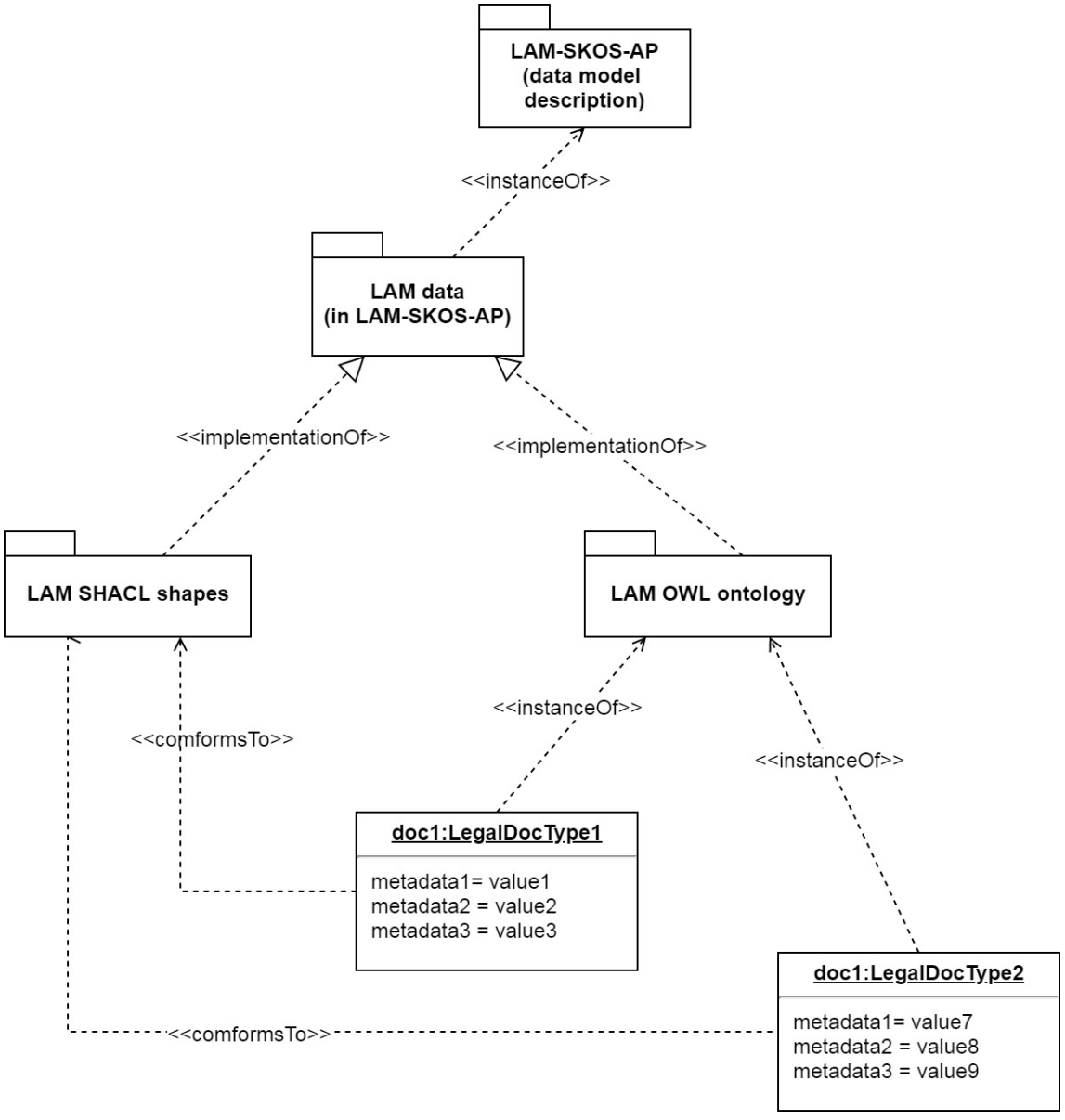

Figure 3: LAM modeling stack

uses terms from one or more base standards, adding more specificity by identifying mandatory, recommended and optional elements to be used for a particular application, as well as recommendations for controlled vocabularies to be used. The data

expressed using LAM-SKOS-AP constitute a easy to manage proxy for the final model. This means that it can be further transformed through a script into a formal LAM ontology (expressed in OWL2 language) and into a set of formal data shapes (expressed in SHACL language[6]) .

In Figure 3 are depicted the above mentioned assets and the relations between them. At the top most is the LAM-SKOS-AP which is the application profile for structuring the informal LAM knowledge base. The actual LAM data instantiate of the LAM-SOKS-AP; while the LAM OWL ontology and LAM SHACL shapes are implementations of LAM data (expressed in LAM-SKKPS-AP form). Two legal document instances are provided as examples to highlight the conformance relation to the data shapes underlining the relevance of the SHACL shapes.

[6]Knublauch H. & Kontokostas D., (2017). Shapes Constraint Language (SHACL). `https://www.w3.org/TR/shacl/`