



Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe

D4.3 Final version of Workflow definition

PROJECT ACRONYM	Lynx
PROJECT TITLE	Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe
GRANT AGREEMENT	H2020-780602
FUNDING SCHEME	ICT-14-2017 - Innovation Action (IA)
STARTING DATE (DURATION)	01/12/2017 (36 months)
PROJECT WEBSITE	http://lynx-project.eu
COORDINATOR	Elena Montiel-Ponsoda (UPM)
RESPONSIBLE AUTHORS	Julián Moreno Schneider (DFKI), Georg Rehm (DFKI)
CONTRIBUTORS	
REVIEWERS	Artem Revenko (SWC), Socorro Bernardos (UPM), María Navas-Loro (UPM), Víctor Rodríguez-Doncel (UPM)
VERSION STATUS	V1 Final
NATURE	Other
DISSEMINATION LEVEL	Public
DOCUMENT DOI	10.5281/zenodo.3235767
DATE	31/05/2018 (M18)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780602

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	First draft version	01/04/2019	Julián Moreno Schneider (DFKI), Georg Rehm (DFKI)
0.2	First draft of structure	18/04/2019	
0.3	Include content into common workflows	07/05/2019	
0.4	Include content into scenario workflows	07/05/2019	
0.5	Writing Introduction and conclusions	10/05/2019	
0.6	Writing Executive Summary	10/05/2019	
0.9	Including reviewers comments	24/05/2019	
1.0	Final remarks	29/05/2019	

LIST OF ACRONYMS

BB	Building Block
EntEx	Entity Extraction
DCM	Document Manager
GEO	Geolocation information extraction
LKG	Legal Knowledge Graph
NER	Named Entity Recognition
RelEx	Relation Extraction
RFP	Request for Proposal
SeSim	Semantic Similarity
StrEx	Document Structure Analysis
Summ	Summarization
TERM	Terminology extraction
TIMEX	Temporal Expression Analysis
ToClass	Topic Classification (Detection)
TRANS	Translation
WP	Work Package
WSD	Word Sense Disambiguation

EXECUTIVE SUMMARY

This report describes the final definition of the curation workflows associated with every business use case (as defined in D4.1 [LynxD41]). This process started with the initial definition provided in D4.2 [LynxD42], in which we outlined four scenarios: *data protection*, *labour law*, *CE marking* and *geothermal energy*. One of them has been put on hold (*CE marking*) and one of them has been adapted to cover a certain type of document (contracts). *Data protection* is now named *contract analysis*.

The specification of a workflow includes its input and output as well as the functionality it is supposed to perform: annotate or enrich a document, add a document to the knowledge base, search for information, etc. Workflows make use of the services (building blocks) to implement the required functionality. The content curation workflows for the different use cases that we prototypically implement in the project have been defined. We performed a systematic analysis of the microservices (developed in parallel) and matched them with the required functionalities for each use case.

- First, we determine the principal elements involved in each use case, i.e., the services, input and output.
- Second, we define the order in which the services have to be executed.
- Third, we identify the shared components in the different workflows.

Currently we have defined five different workflows, which are divided into two groups: (i) those that are commonly used in more than one use case; and (ii) those that are used case specifically.

- The common workflow is:
 - LKG population
- The use case specific workflows are:
 - Contract Analysis (OLS)
 - Labour Law Question Answering (CuatreCasas)
 - Geothermal Project Analysis (DNV GL)
 - Geothermal Project Extended Analysis (DNV GL)

Apart from the business cases, there is the General User/Public Portal use case, which is also considered for defining specific workflows.

TABLE OF CONTENTS

1	INTRODUCTION	5
1.1	PURPOSE OF THIS DOCUMENT	6
1.2	STRUCTURE OF THIS DOCUMENT	6
2	COMMON WORKFLOWS	7
2.1	LEGAL KNOWLEDGE GRAPH WORKFLOWS	7
2.1.1	Legal Knowledge Graph Population	7
3	SCENARIO SPECIFIC WORKFLOWS	9
3.1	SCENARIO 1 (CONTRACT ANALYSIS) WORKFLOWS	9
3.1.1	Contract Analysis Workflow	9
3.2	SCENARIO 2 (OIL&GAS – GEOTHERMAL ENERGY) WORKFLOWS	11
3.2.1	Geothermal Project Analysis Workflow	11
3.2.2	Geothermal Project Extended Analysis Workflow	13
3.3	SCENARIO 3 (LABOUR LAW ANALYSIS) WORKFLOWS	14
3.3.1	Labour Law Q&A Workflow(s)	14
3.4	SCENARIO 4 (GENERAL USER/PUBLIC PORTAL) WORKFLOW	15
3.4.1	Web Interface	15
3.4.2	Public Portal Search	16
3.4.3	Public Portal Document Navigation	16
4	CONCLUSIONS	17
	ANNEX 1. XML DEFINITION OF THE WORKFLOWS	18
	LKG POPULATION WORKFLOW	18

TABLE OF FIGURES

Figure 1. Graphical representation of LKG Population Workflow.....	7
Figure 2. Graphical representation of Contract Analysis workflow.	9
Figure 3. Example mockup of output summary (first page) for Contract Analysis Workflow	10
Figure 4. Graphical representation of the Geothermal Project Analysis Workflow.	12
Figure 5. Graphical representation of the Geothermal Project Extended Analysis Workflow.	13
Figure 6. Graphical representation of the Labour Law Q&A Workflow.	14
Figure 7. Mock-up design example for the web interface of the General User/Public Portal.....	16

TABLE OF TABLES

Table 1. List of defined workflows and their identifying acronyms in Lynx	6
Table 2. Possible answers of the LKG Population workflow.....	8
Table 3. Example of output JSON for the Geothermal Project Analysis workflow.	12
Table 4. Example of output JSON for the Geothermal Project Extended Analysis workflow.	13
Table 5. Example of input JSON for the Labour Law Search workflow.	15
Table 6. Relation between Workflows and Building Blocks.	17
Table 7. XML definition of the Legal Knowledge Graph Population workflow.	22

1 INTRODUCTION

The Lynx platform has been designed using a microservice architecture. A workflow manager is needed in order to orchestrate all the different services. Therefore, and to comply with the needs of every business case, several workflows must make use of these services, named Building Blocks (BB) in Work Package 3 (WP3), and datasets in order to implement the required functionality.

The specification of a workflow includes its input and output as well as the functionality it is supposed to perform: annotate or enrich a document, add a document to the knowledge base, search for information, etc. Workflows make use of the services (building blocks) to implement the required functionality. The content curation workflows for the different use cases that we prototypically implement in the project have been defined. We performed a systematic analysis of the microservices (developed in parallel) and matched them with the required functionalities for each use case.

- First, we determine the principal elements involved in each use case, i.e., the services, input and output.
- Second, we define the order in which the services have to be executed.
- Third, we identify the shared components in the different workflows.

The final definition of the curation workflows is done on a scenario level and has been accomplished by the following steps:

- The initial definition of the workflows has been refined together with all the involved partners in every pilot use case.
- The General User/Public Portal workflows have been defined during the Lynx Coding Sprint taking place in Cercedilla (Madrid),¹ involving the whole consortium.
- A final refinement of the workflows has been carried out during the Barcelona plenary meeting² involving the whole consortium.

As already described in deliverable D4.2 [LynxD42], there are three pilots (business cases), which were divided in four scenarios and, in turn, each of the scenarios was divided into several use cases. Now the case is that one of the scenarios, Scenario 3a in deliverable 4.2 [LynxD42], has been frozen because the complexity of scenario 3b in deliverable D4.2 [LynxD42] (now renamed to Scenario 3) was enough for DNV-GL. Apart from keeping three scenarios, and defining a new scenario for the European citizen (General User/Public Portal), we have renumbered them in order to match the Description of Work. Now, the new numbering of the scenarios is:

- Scenario 1: Contract Analysis
- Scenario 2: Oil&Gas – Geothermal Energy
- Scenario 3: Labour Law
- Scenario 4: General User/Public Portal

Table 1 presents a list of the defined workflows together with an identifier Acronym.

¹ <http://lynx-project.eu/news>

² *ibid.*

Number	Name	Acronym	Scenario
1	Legal Knowledge Graph Population	LKG	1, 2, 3, 4
2	Contract Analysis	ConAn	1
3	Geothermal Project Analysis	GeoPA	2
4	Geothermal Project Extended Analysis	GeoPExA	2
5	Labour Law Question Answering	LabLaw	3
6	General User/Public Portal	GUPP	4

Table 1. List of defined workflows and their identifying acronyms in Lynx

1.1 PURPOSE OF THIS DOCUMENT

This report describes the final definitions of the workflows related to the different use case pilots. The document is based on D4.2 [LynxD42], which defines the initial version of the workflows, and aligned with D1.1 [LynxD11] and D4.1 [LynxD41], which define the requirements for the Lynx platform collected from the use case pilots.

1.2 STRUCTURE OF THIS DOCUMENT

Section 2 describes the workflows needed for pre-processing of information or training of services. Section 3 describes common workflows needed in several use cases. Section 4 defines the different workflows required for every use case pilot (and scenario) defined in D4.1 [LynxD41]. Section 5 concludes this deliverable.

2 COMMON WORKFLOWS

This section describes the workflows needed or used in several scenarios and use cases.

2.1 LEGAL KNOWLEDGE GRAPH WORKFLOWS

2.1.1 Legal Knowledge Graph Population

The inclusion of documents into the LKG is a feature required for all use cases (Contract Analysis, Labour Law, Geothermal Projects, General User). Therefore, the workflow that populates the LKG with documents is use case agnostic and a common workflow in the Lynx project. The workflow is responsible for processing documents, by annotating them with semantic information (through several enrichment services), and for including them into the LKG. Besides, this workflow is also responsible for including the document into the full text search module (indexing process).

The type of documents that are included in the LKG, i.e., type of documents that the LKG population workflow can manage, are different depending on the specific use case: GDPR (and other related documents) for the *Contracts Analysis Scenario*, labour law, collective bargaining agreements and court cases for the *Labour Law Scenario* and standards and best practice documents for the *Geothermal Scenario*. Figure 1 depicts the graphical representation of the LKG Population workflow. The different steps (building blocks) that are accessed and the order in which they are executed are outlined.

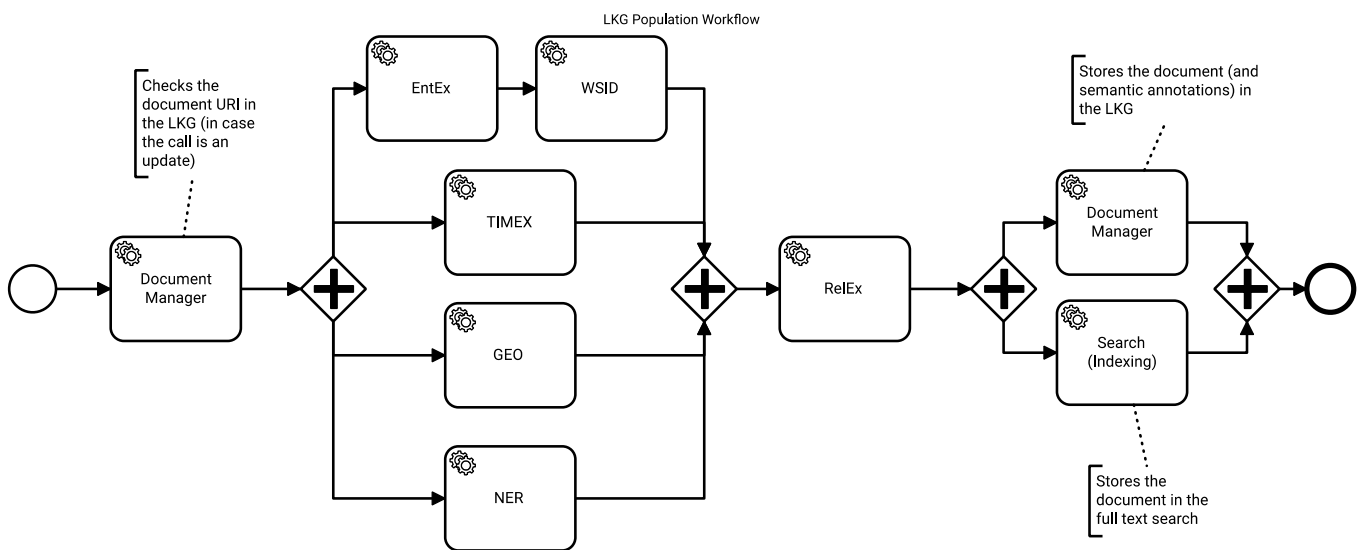


Figure 1. Graphical representation of LKG Population Workflow.

As can be seen in Figure 1, there are five services that can be requested simultaneously (in parallel) while the others cannot because they need enriched information that should be provided by the previous services. For example, RelEx needs information on the entities that have been identified in the document.

2.1.1.1 Input

The input of this workflow is a document. The format of the documents is independent from the workflow manager, because it accepts documents in NIF format. Therefore, the different formats of documents (XML, HTML, PDF, etc.) must be converted into NIF outside of this workflow. Depending on the implementation details, it could happen that this workflow must convert the documents as a first step of the workflow.

2.1.1.2 Output

This workflow does not provide any information as output because its main functionality is populating (storing information in) the LKG. Therefore, the output is a message confirming the result of the document population. The output is a JSON message containing two fields: code and text. If the document has been properly populated, then the code will return “OK” and the text will be the identifier (ID or URI) of the document in the LKG. If an error occurs, the code and text will relate to the error, trying to be as descriptive as possible.

The different answers given by this workflow are described in Table 2.

Answer Code	Answer Description	Answer Text
OK	Success	Identifier of the document in the LKG
ERR_PRO	Semantic processing error: some processing step failed.	Textual description of the error. It would be as concrete as possible.
ERR_DUP	Duplication error: the document already exists in the LKG and the call was not an Update.	
ERR_LKG	LKG storage error: the storage of the semantic information together with the document in the LKG failed.	
ERR_SEA	Search indexing error: the inclusion of the document in the full text search index failed.	

Table 2. Possible answers of the LKG Population workflow.

2.1.1.3 Workflow Components (building blocks)

The building blocks that have been identified to be used by this workflow are:

- Document Manager (left part of Figure 1): adds the document to the LKG in order to obtain a Lynx Identifier. If the population action is an Update, then this step retrieves the actual Identifier of the document from the LKG
- NER: annotates persons and organizations as generic entities, and also a set of legal domain specific entities such as: person, judge, lawyer, country, city, street, landscape, public/social, state and economic institutions, institution, company, court and brand.
- WSID: induces and disambiguates analysed Named Entities
- TIMEX: annotates temporal expressions – times and dates
- GEO: annotates geographical entities
- EntEx: annotates specific terms (from thesaurus) in the document
- RelEx: analyses references between entities mentioned within a text and also with other documents in the LKG
- Document Manager (right part of Figure 1): includes the document and the enriched information into the LKG
- Search (Indexing): includes the document into the full text search building block

3 SCENARIO SPECIFIC WORKFLOWS

This section describes the specific workflows needed in every scenario and use case.

3.1 SCENARIO 1 (CONTRACT ANALYSIS) WORKFLOWS

This section describes the workflows for Scenario 1. This scenario corresponds to business case 1, “Contract Analysis”. Its objective is to process rental contracts in order to analyse its compliance with data protection. The prototype analyses two types of documents:

- Public regulatory data protection framework: data protection legislation and case law from the EU and Member States and public provisions and suggestions by authorities.
- Private data processing contracts: contracts between controllers/data subjects/processors, data processing policies of companies and general contracts which may include data processing clauses.

A complete description of Scenario 1 can be found in D4.2 [LynxD42].

3.1.1 Contract Analysis Workflow

This workflow processes a contract through some semantic services, which enrich it with general and use case specific semantic information. The contract also goes through other services (Summ and Trans) that generate new content to be included, and finally, the search functionalities (Search and SeSim) include links to relevant information (mainly documents) from the LKG. The datasets required by this workflow are: (i) a set of contracts that can be used for training purposes; and (ii) the legislation and case law (e.g., GDPR) processed and included in the LKG. A graphical representation of the Contract Analysis workflow is depicted in Figure 2.

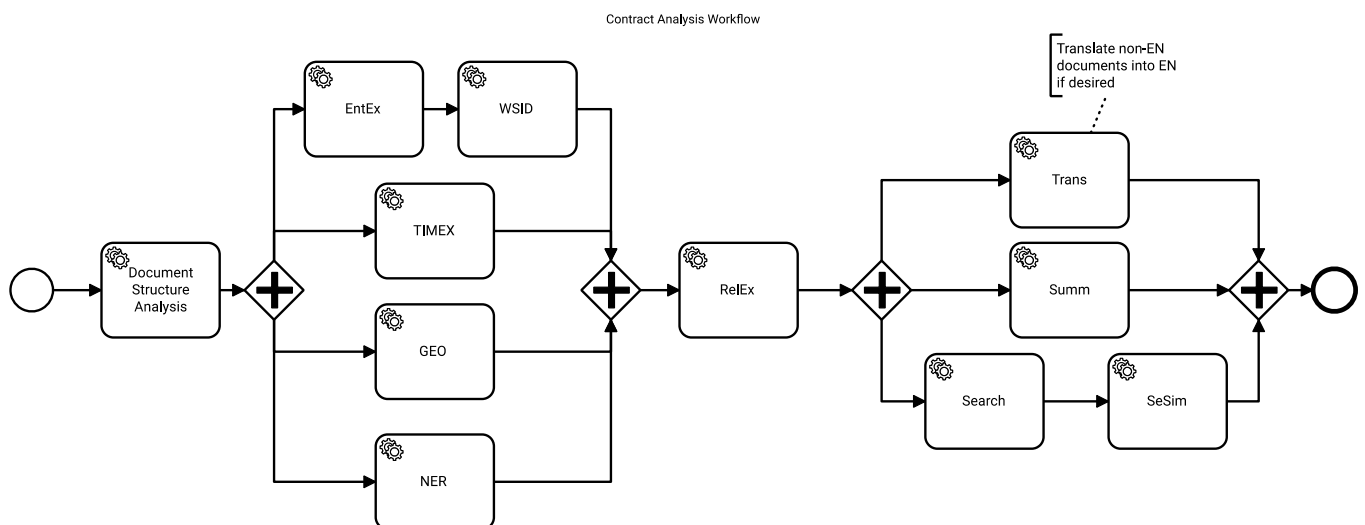


Figure 2. Graphical representation of Contract Analysis workflow.

3.1.1.1 Input


The input of the workflow is a contract, i.e., a PDF document containing a contract. We use an XML-based annotation format that is based on the international DocBook standard [DocBook2010] to annotate the example contracts. The idea is to transform, automatically, incoming PDF contracts into DocBook-annotated XML documents. The original document will be crawled and parsed to be available in NIF for the services.

3.1.1.2 Output

The output of the workflow will be a PDF document containing the contract as well as some additional information added to the contract as metadata (if possible) or as a summary included in the contract PDF file as its new first page including:

- Title
- Parties
- Places, addresses
- Times, dates
- Deadlines
- Applicable law
- Privacy clause
- Legislation References, e.g., to the GDPR
- Keywords
- Document structure
- Signatures

Figure 3 depicts a possible mock-up example of an output summary page for a contract. As can be seen, it contains the most relevant information retrieved and generated while processing the contract.



« TITLE »
« KEYWORDS »

Parties . . .
« PARTIES »
Entities (Places, Ad- dresses) . . .
Term1 1, 2, 3
Term2 1
Terms . . .
Term1 1, 2, 3
Term2 1
Dates and Deadlines . . .
« DEADLINES »
Applicable law (Legislation References) . . .
« APPLICABLELAW »
Privacy Clauses . . .
« PRIVACY »

Metadata
The metadata associated with this contract is:
Creation Date: Wed May 08 14:19:39 CEST 2019
Modification Date: Wed May 08 14:19:39 CEST 2019

« SIGNATURES »

Table of Content

Section 1 1

Summary

That is the summary of the document

Figure 3. Example mockup of output summary (first page) for Contract Analysis Workflow

3.1.1.3 Workflow Components (building blocks)

The building blocks that have been identified to be used by this workflow are:

- Document Manager (left part of Figure 2): includes the document into the LKG in order to obtain a Lynx Identifier. If the population action is an Update, then this step retrieves the actual Identifier of the document from the LKG
- NER: annotates persons (possibly parties) and organizations as generic entities, and also a set of legal domain specific entities such as: person, judge, lawyer, country, city, street, landscape, public/social, state and economic institutions, institution, company, court and brand
- WSID: induces and disambiguates analysed Named Entities and extracted entities
- TIMEX: annotates temporal expressions: times, dates and deadlines
- GEO: annotates geographical entities, specially fined-grained such as addresses
- EntEx: annotates specific terms (from thesaurus) in the document
- RelEx: analyses references between entities mentioned within a text and also with other documents in the LKG
- Trans (left part of the figure): translates the document into all language in the project: ES, EN, DE and NL (apart from the source language)
- Search+Sesim: retrieves similar and relevant documents for the contract
- Summ: generates a summary of the contract

3.2 SCENARIO 2 (OIL&GAS – GEOTHERMAL ENERGY) WORKFLOWS

This section describes the workflows for the Scenario 3 (Oil&Gas – Geothermal energy) in the Lynx project. This scenario focuses on compliance management support for geothermal energy projects and aims to obtain standards and regulations associated with certain terms in the field of geothermal energy, across the whole project life cycle (from inception to operation and decommissioning).

The idea is that a user can submit a Request for Proposal (RFP), feasibility study or other geothermal project description through the system and is then informed which standards, regulations and industry best practice must be taken into consideration to carry out the considered project in a compliant manner. The scenario corresponds to business case 2, “Compliance Assurance Services in Oil & Gas and Energy”. Its objective is to innovate both existing compliance related services offered by DNV GL as well as existing compliance management processes within DNV GL customers to achieve accelerated and more effective compliance. Within this scenario, the system identifies matches between two categories of documents:

1. RFPs, feasibility studies or other forms of geothermal project descriptions
2. Regulations, standards and industry best practice in the geothermal energy domain, as well as in adjacent domains such as the oil & gas sector

A complete description of the Scenario 3b can be found in D4.2 [LynxD42].

3.2.1 Geothermal Project Analysis Workflow

This workflow retrieves (parts of) documents related to the Geothermal Energy domain that are relevant for a technical description of a geothermal project. The overview of the workflow is shown in Figure 4.

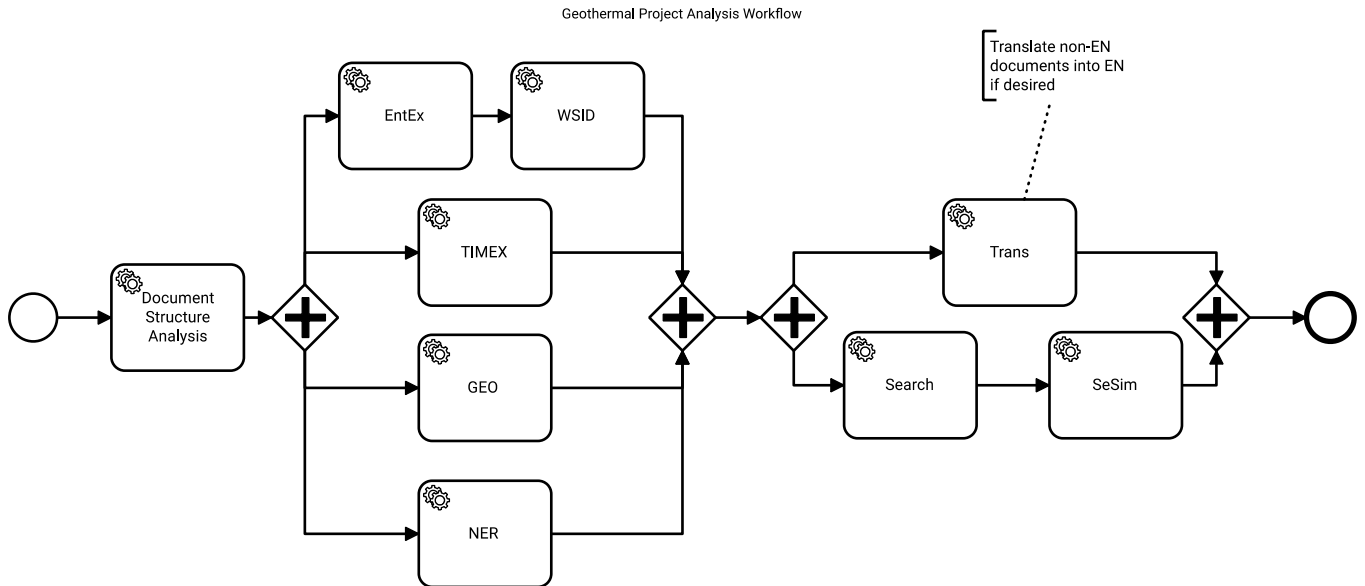


Figure 4. Graphical representation of the Geothermal Project Analysis Workflow.

Even though all documents in the LKG are available, the most interesting documents for this business use case are standards and best practice documents.

3.2.1.1 Input

The input of the workflow is a description of a geothermal project. This information is provided as a PDF/Word document.

3.2.1.2 Output

The output of the workflow is a JSON document containing regulations, standards and best practices (or parts of them) that are related for the geothermal project description. An example of such an output is shown in Table 3.

```

{
  "results":[
    {
      "result_id":"1",
      "text":" anticipada",
      "document_id":"lkg_document_328765",
      "target_lang":"en"
    },
    {
      "result_id":"2",
      "text":"minimum ",
      "document_id":"lkg_document_328765_section3",
      "target_lang":"en"
    }
  ]
}
    
```

Table 3. Example of output JSON for the Geothermal Project Analysis workflow.

3.2.1.3 Workflow Components (building blocks)

Although this is only a search workflow, in order to find the most relevant documents (standards, regulations and best practices) related to the project description, it is needed to analyse the document

and use the extracted semantic information to perform a better search. The building blocks that have been identified to be used by this workflow are:

- Document Manager (left part of Figure 4): adds the document to the LKG in order to obtain a Lynx Identifier. If the population action is an Update, then this step retrieves the actual Identifier of the document from the LKG
- NER: annotates persons and organizations as generic entities.
- WSID: induces and disambiguates analysed Named Entities
- TimEx: annotates temporal expressions – times and dates. Besides it can also identify numbers and values.
- GEO: annotates geographical entities
- EntEx: annotates specific terms (from thesaurus) related to the Energy and Geothermal domain in the document
- RelEx: analyses references between entities mentioned within a text and also with other documents in the LKG
- Trans: translates the documents into the destination language in case it is different from the document language (considering only the languages used in the project: EN, DE, ES, NL)
- Search: full text search to retrieve the most relevant documents (standards, best practices and regulations)
- SeSim: performs a semantic similarity search that could help to filter the full text search

3.2.2 Geothermal Project Extended Analysis Workflow

This workflow has exactly the same functionality as the previous workflow but includes a summary of the retrieved information in the results instead of just returning them (as parts of documents). The overview of the workflow (based on Geothermal Search Workflow) is depicted in Figure 5.

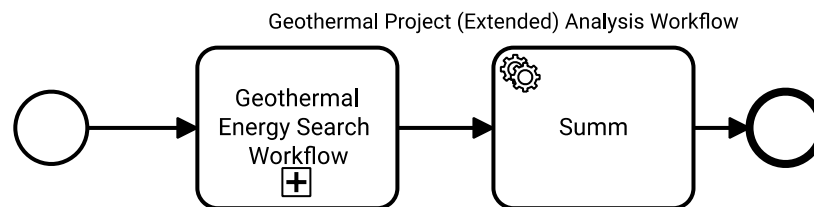


Figure 5. Graphical representation of the Geothermal Project Extended Analysis Workflow.

3.2.2.1 Input

The input of the workflow is a description of a geothermal project. This information is provided as a PDF/Word document.

3.2.2.2 Output

The output of the workflow is a JSON document containing regulations, standards and best practices (or parts of them) that are related to the geothermal project description, together with a summary of all the retrieved documents. An example of such an output is shown in Table 4.

```

{
  "results":[
    "Exactly the same as in Table 3"
  ],
  "summary":"This text contains a summary generated from all the resulting snippets or documents."
}
  
```

Table 4. Example of output JSON for the Geothermal Project Extended Analysis workflow.

3.2.2.3 Workflow Components (building blocks)

The building blocks that have been identified to be used by this workflow are:

- Geothermal Project Analysis Workflow
- SUMM: generates a summary of the retrieved (parts of the) documents

3.3 SCENARIO 3 (LABOUR LAW ANALYSIS) WORKFLOWS

This section describes the workflows for Scenario 2 (Labour Law). This scenario corresponds to business case 3, “Compliance Assurance Services in Labour Law”. Its objective is to provide access to aggregated and interlinked relevant legal information regarding law labour across multiple legal orders, jurisdictions, and languages. The prototype is meant to analyse three types of documents:

- EU and Member State Labour Law: labour legislations from the EU and Member States.
- Collective bargaining agreements from Member States.
- Labour law jurisprudence: jurisprudence related to labour law issues in the different jurisdictions that relate to the national or European labour laws.

A complete description of Scenario 2 can be found in deliverable D4.2 [LynxD42].

3.3.1 Labour Law Q&A Workflow(s)

This scenario encompasses two workflows that can be combined into only one description, because the second workflow can be considered an improvement (or an extension) of the first workflow. These workflows answer a set of queries (in specific jurisdictions) related to the labour law domain. The first workflow returns a set of answers (and supporting parts of documents) for every query. The second workflow uses these answers and supporting (parts of) documents to generate a piece of text (summary) that summarizes the information contained in all retrieved answers and supporting documents (results). Figure 6 depicts the involved elements in the workflow(s) in the order that they are requested.

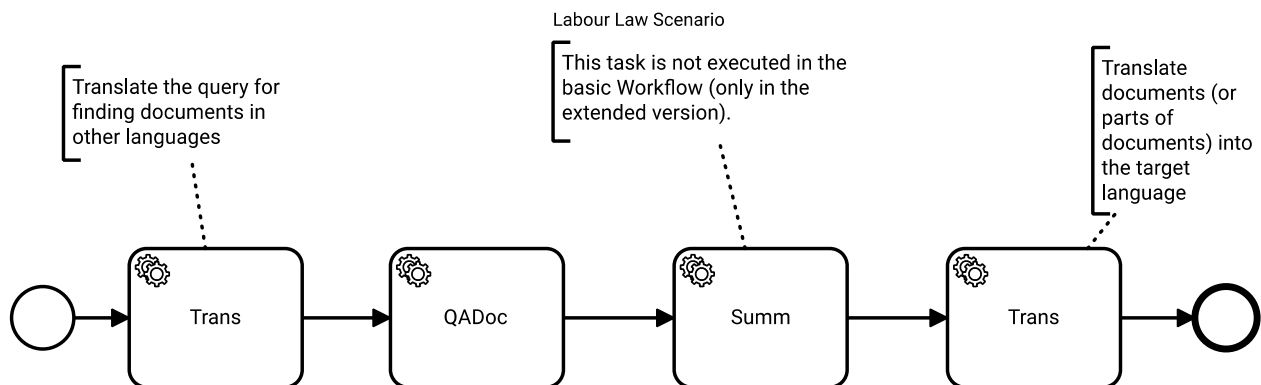


Figure 6. Graphical representation of the Labour Law Q&A Workflow.

3.3.1.1 Input

The input of the workflow is a set of queries together with some additional information for every question. This information is provided by the user as a JSON document. An example of such an input is shown in Table 5. This JSON description is not exhaustive here, but will be further refined at a later stage.

```

{
  "queries":[
    {
      "id":"q1",
    }
  ]
}
  
```

```

    "text": "edad mínima de jubilación anticipada",
    "source_lang": "es",
    "target_lang": "es",
    "additional_information": {
      "jurisdiction": "es",
      "sex": "female"
    }
  },
  {
    "id": "q2",
    "text": "minimum retirement age",
    "source_lang": "en",
    "target_lang": "en,de",
    "additional_information": {
      "jurisdiction": "eu",
      "sex": "female"
    }
  }
]

```

Table 5. Example of input JSON for the Labour Law Search workflow.

3.3.1.2 Output

The output of the workflow is a JSON document containing labour law documents (or parts) that are related to every question (and its additional information). In the case of the extended workflow (including the summarisation part) the answer also includes a summary for every query at the input.

3.3.1.3 Workflow Components (building blocks)

The building blocks that have been identified to be used by this workflow are:

- QADoc: is a building block that provides the whole functionality for this workflow, retrieving labour law documents (or parts) related to the input queries.
- SUMM: this module creates a summary of the information contained in the retrieved documents. This is only applicable in the Extended Use Case (UC2.3), not in the basic one (UC2.1 and UC2.2).

3.4 SCENARIO 4 (GENERAL USER/PUBLIC PORTAL) WORKFLOW

The Lynx platform is not only going to be used by the pilot use case partners, but it must also be suitable, interesting and usable for European citizens. Facing this goal, we have defined a new scenario in the project: the citizen's portal scenario (General user/public portal). Its main objective is to enable European citizens to access the information stored in the LKG, although they can access only the public part of the LKG.

Based on the necessary functionalities for the citizen's portal, it is not needed to define specific workflows, because these functionalities are provided directly by individual building blocks: the first functionality is provided by the Search building block, while the second functionality is provided by the SeSim building block. Next are described both functionalities with more detail.

3.4.1 Web Interface

The General User use case is going to be developed independently of the other business use cases, therefore a new web interface has to be design. In Figure 7, a proposed design (mock-up) of the interface can be seen.

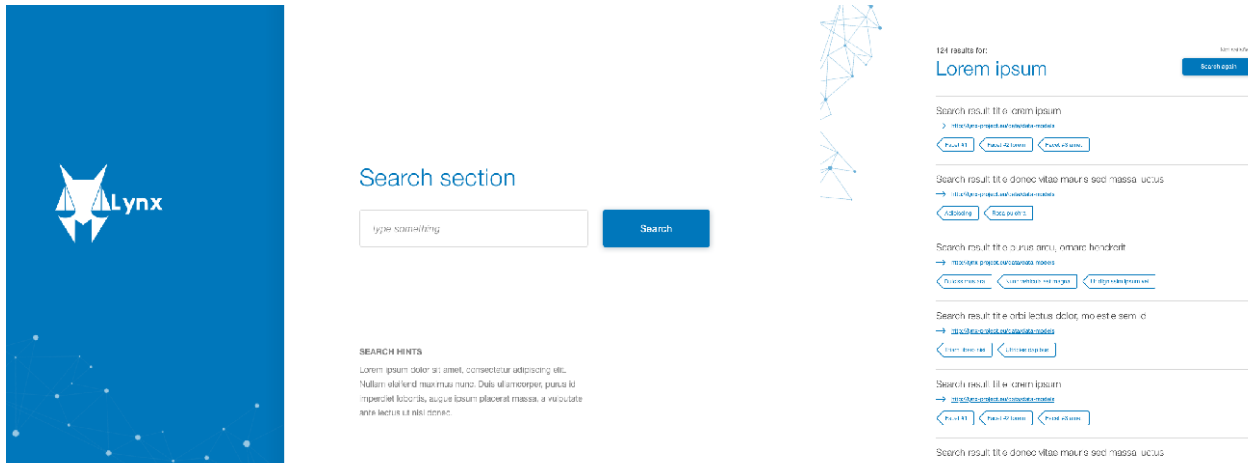


Figure 7. Mock-up design example for the web interface of the General User/Public Portal

3.4.2 Public Portal Search

This functionality allows (non-registered) users to access information in the legal domain through the web interface. Users can post queries (natural language) and will retrieve the information (documents or parts of documents) from the legal domain that are stored in the LKG. The private part of the LKG, standards and right-protected documents, are not going to be accessible through this functionality. The result of the search will be a kind of web search result list.

3.4.3 Public Portal Document Navigation

Once a result list is displayed to the user, they will be able to navigate through the LKG in order to find similar and relevant documents to the one they are interested in. This could be implemented as a pop-up menu when clicking on any of the results of the list in the web interface. This functionality is fulfilled by the SeSim module, that is able to find similarities between documents of the LKG.

4 CONCLUSIONS

This report provides the final description of the curation workflows associated with every use case (as defined in D4.1), specifically for the pilot use cases. These descriptions are based on deliverable D4.2 [LynxD42].

The initial definition of the workflows (deliverable D4.2 [LynxD42]) was created based on the information obtained from the business case partners in individual workshops. Based on several discussions, both with all partners and bilaterally, we have determined some conflictive points in the initial definition of the workflows, which have been solved leading to changes in the ordering of elements or the elements themselves that are included in several use cases workflows.

The most conflictive components were:

- Translation service: due to the multilingual nature of the Lynx project and the management of documents in different languages (EN, DE, ES, NL), it was difficult to determine in which position it would be best to establish the translation service, particularly in the LKG population workflow. The discussion arises when deciding if documents in other languages than EN must always be translated to English or must be processed in their original language (through the building blocks as far as they are available for the specific language).
- The Document Manager functionality was complicated to locate and to define. At the beginning, it was only regarded as a storage and retrieval module, but it also has the functionality of providing URIs for new documents and returning URIs of already existing documents. Therefore, it appears in more places than before.

In order to provide a better understanding of the relation between workflows and building blocks, we present the relations in Table 6.

	NER	TIMEX	GEO	EntEx	WSID	ReLex	QADoc	Trans	Summ	Sear	Sesim	DCM
LKG	X	X	X	X	X	X				X		X
ConAn	X	X	X	X	X	X		X	X	X	X	
LabLaw							X					
GeoPA	X	X	X	X	X	X		X		X	X	
GeoPEA	X	X	X	X	X	X		X	X	X	X	
GUPP										X	X	

Table 6. Relation between Workflows and Building Blocks.

Apart from the changes carried out on the workflows already included in the initial definition, we have also had some meetings and discussions with the whole consortium for defining the General User/Public Portal use case. This use case is first defined in this deliverable and could undergo modifications while the web interface is being implemented.

The following steps that must be carried out in the project would be the implementation of a curation workflow manager and the implementation of the workflows defined in this report. Based on the actual implementation, the definitions could be slightly modified.

ANNEX 1. XML DEFINITION OF THE WORKFLOWS

The XML definitions of the workflows defined in Lynx can be downloaded here: <http://lynx-project.eu/doc/api/>. Below, we present one example of XML workflow definition generated using Camunda Modeler (<https://camunda.com/de/products/modeler/>).

LKG POPULATION WORKFLOW

```

<?xml version="1.0" encoding="UTF-8"?>
<bpmn:definitions xmlns:bpmn="http://www.omg.org/spec/BPMN/20100524/MODEL"
xmlns:bpmndi="http://www.omg.org/spec/BPMN/20100524/DI"
xmlns:dc="http://www.omg.org/spec/DD/20100524/DC"
xmlns:camunda="http://camunda.org/schema/1.0/bpmn"
xmlns:di="http://www.omg.org/spec/DD/20100524/DI" id="Definitions_1d3v58z"
targetNamespace="http://bpmn.io/schema/bpmn" exporter="Camunda Modeler"
exporterVersion="3.0.1">
  <bpmn:process id="Process_0okrj44" isExecutable="true">
    <bpmn:startEvent id="StartEvent_1sxo1hd">
      <bpmn:outgoing>SequenceFlow_1cv04j0</bpmn:outgoing>
    </bpmn:startEvent>
    <bpmn:parallelGateway id="ParallelGateway_1x7c2cl">
      <bpmn:incoming>SequenceFlow_1cv04j0</bpmn:incoming>
      <bpmn:outgoing>SequenceFlow_0h7fjlb</bpmn:outgoing>
      <bpmn:outgoing>SequenceFlow_0pavv3j</bpmn:outgoing>
      <bpmn:outgoing>SequenceFlow_05b1igv</bpmn:outgoing>
      <bpmn:outgoing>SequenceFlow_0wh71kw</bpmn:outgoing>
    </bpmn:parallelGateway>
    <bpmn:parallelGateway id="ParallelGateway_1yyvq9u" camunda:asyncBefore="true">
      <bpmn:incoming>SequenceFlow_1eq5wl3</bpmn:incoming>
      <bpmn:incoming>SequenceFlow_0wx6x84</bpmn:incoming>
      <bpmn:incoming>SequenceFlow_1azpnjf</bpmn:incoming>
      <bpmn:incoming>SequenceFlow_0qgk02i</bpmn:incoming>
      <bpmn:outgoing>SequenceFlow_1y0oka0</bpmn:outgoing>
    </bpmn:parallelGateway>
    <bpmn:serviceTask id="ServiceTask_0m4p9b8" name="TimEx" camunda:type="external"
camunda:topic="TimEx">
      <bpmn:extensionElements>
        <camunda:inputOutput>
          <camunda:inputParameter name="language">en</camunda:inputParameter>
        </camunda:inputOutput>
      </bpmn:extensionElements>
      <bpmn:incoming>SequenceFlow_0h7fjlb</bpmn:incoming>
      <bpmn:outgoing>SequenceFlow_1eq5wl3</bpmn:outgoing>
    </bpmn:serviceTask>
    <bpmn:serviceTask id="ServiceTask_1jbdub2" name="GEO" camunda:type="external"
camunda:topic="geo">
      <bpmn:extensionElements>
        <camunda:inputOutput>
          <camunda:inputParameter name="models">ner-wikinerEn_PER</camunda:inputParameter>
          <camunda:inputParameter name="language">en</camunda:inputParameter>
          <camunda:inputParameter name="analysis">language</camunda:inputParameter>
          <camunda:inputParameter name="content">true</camunda:inputParameter>
          <camunda:inputParameter name="mode">spot</camunda:inputParameter>
        </camunda:inputOutput>
      </bpmn:extensionElements>
      <bpmn:incoming>SequenceFlow_0pavv3j</bpmn:incoming>
      <bpmn:outgoing>SequenceFlow_0wx6x84</bpmn:outgoing>
    </bpmn:serviceTask>
    <bpmn:sequenceFlow id="SequenceFlow_0h7fjlb" sourceRef="ParallelGateway_1x7c2cl"
targetRef="ServiceTask_0m4p9b8" />
    <bpmn:sequenceFlow id="SequenceFlow_0pavv3j" sourceRef="ParallelGateway_1x7c2cl"
targetRef="ServiceTask_1jbdub2" />
    <bpmn:sequenceFlow id="SequenceFlow_1eq5wl3" sourceRef="ServiceTask_0m4p9b8"
targetRef="ParallelGateway_1yyvq9u" />
    <bpmn:sequenceFlow id="SequenceFlow_0wx6x84" sourceRef="ServiceTask_1jbdub2"
targetRef="ParallelGateway_1yyvq9u" />
    <bpmn:serviceTask id="ServiceTask_098xws9" name="NER" camunda:type="external"
camunda:topic="geo">
      <bpmn:extensionElements>
        <camunda:inputOutput>
          <camunda:inputParameter name="models">ner-wikinerEn_PER</camunda:inputParameter>
    
```

```

        <camunda:inputParameter name="language">en</camunda:inputParameter>
        <camunda:inputParameter name="analysis">language</camunda:inputParameter>
        <camunda:inputParameter name="content">true</camunda:inputParameter>
        <camunda:inputParameter name="mode">spot</camunda:inputParameter>
    </camunda:inputOutput>
</bpmn:extensionElements>
<bpmn:incoming>SequenceFlow_05b1igv</bpmn:incoming>
<bpmn:outgoing>SequenceFlow_1aiovlx</bpmn:outgoing>
</bpmn:serviceTask>
<bpmn:sequenceFlow id="SequenceFlow_05b1igv" sourceRef="ParallelGateway_1x7c2c1"
targetRef="ServiceTask_098xws9" />
    <bpmn:serviceTask id="ServiceTask_0tskw6a" name="WSD" camunda:type="external"
camunda:topic="wsd">
        <bpmn:extensionElements>
            <camunda:inputOutput>
                <camunda:inputParameter name="models">ner-wikinerEn_PER</camunda:inputParameter>
                <camunda:inputParameter name="language">en</camunda:inputParameter>
                <camunda:inputParameter name="analysis">language</camunda:inputParameter>
                <camunda:inputParameter name="content">true</camunda:inputParameter>
                <camunda:inputParameter name="mode">spot</camunda:inputParameter>
            </camunda:inputOutput>
        </bpmn:extensionElements>
        <bpmn:incoming>SequenceFlow_1aiovlx</bpmn:incoming>
        <bpmn:outgoing>SequenceFlow_1azpnjf</bpmn:outgoing>
    </bpmn:serviceTask>
    <bpmn:sequenceFlow id="SequenceFlow_1aiovlx" sourceRef="ServiceTask_098xws9"
targetRef="ServiceTask_0tskw6a" />
    <bpmn:sequenceFlow id="SequenceFlow_1azpnjf" sourceRef="ServiceTask_0tskw6a"
targetRef="ParallelGateway_1yyvq9u" />
    <bpmn:serviceTask id="ServiceTask_1afzltv" name="EntEx" camunda:type="external"
camunda:topic="entex">
        <bpmn:extensionElements>
            <camunda:inputOutput>
                <camunda:inputParameter name="models">ner-wikinerEn_PER</camunda:inputParameter>
                <camunda:inputParameter name="language">en</camunda:inputParameter>
                <camunda:inputParameter name="analysis">language</camunda:inputParameter>
                <camunda:inputParameter name="content">true</camunda:inputParameter>
                <camunda:inputParameter name="mode">spot</camunda:inputParameter>
            </camunda:inputOutput>
        </bpmn:extensionElements>
        <bpmn:incoming>SequenceFlow_0wh71kw</bpmn:incoming>
        <bpmn:outgoing>SequenceFlow_0qgk02i</bpmn:outgoing>
    </bpmn:serviceTask>
    <bpmn:sequenceFlow id="SequenceFlow_0wh71kw" sourceRef="ParallelGateway_1x7c2c1"
targetRef="ServiceTask_1afzltv" />
    <bpmn:sequenceFlow id="SequenceFlow_0qgk02i" sourceRef="ServiceTask_1afzltv"
targetRef="ParallelGateway_1yyvq9u" />
    <bpmn:sequenceFlow id="SequenceFlow_1y0oka0" sourceRef="ParallelGateway_1yyvq9u"
targetRef="Task_1qc9nh1" />
    <bpmn:endEvent id="EndEvent_02bpwm6">
        <bpmn:incoming>SequenceFlow_0wwqc4z</bpmn:incoming>
    </bpmn:endEvent>
    <bpmn:sequenceFlow id="SequenceFlow_0wwqc4z" sourceRef="Task_1qc9nh1"
targetRef="EndEvent_02bpwm6" />
    <bpmn:sequenceFlow id="SequenceFlow_1cv04j0" sourceRef="StartEvent_1sxolhd"
targetRef="ParallelGateway_1x7c2c1" />
    <bpmn:serviceTask id="Task_1qc9nh1" name="RelEx" camunda:type="external"
camunda:topic="external">
        <bpmn:incoming>SequenceFlow_1y0oka0</bpmn:incoming>
        <bpmn:outgoing>SequenceFlow_0wwqc4z</bpmn:outgoing>
    </bpmn:serviceTask>
</bpmn:process>
<bpmndi:BPMNDiagram id="BPMNDiagram_1">
    
```

```

<bpmndi:BPMNPlane id="BPMNPlane_1" bpmnElement="Process_0okrj44">
  <bpmndi:BPMNShape id="StartEvent_1sxolhd_di" bpmnElement="StartEvent_1sxolhd">
    <dc:Bounds x="156" y="371" width="36" height="36" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNShape id="ParallelGateway_1x7c2c1_di"
bpmnElement="ParallelGateway_1x7c2c1">
    <dc:Bounds x="283" y="364" width="50" height="50" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNShape id="ParallelGateway_1yyvq9u_di"
bpmnElement="ParallelGateway_1yyvq9u">
    <dc:Bounds x="587" y="364" width="50" height="50" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNShape id="ServiceTask_0m4p9b8_di" bpmnElement="ServiceTask_0m4p9b8">
    <dc:Bounds x="406" y="219" width="100" height="80" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNShape id="ServiceTask_1jbdub2_di" bpmnElement="ServiceTask_1jbdub2">
    <dc:Bounds x="406" y="474" width="100" height="80" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNEdge id="SequenceFlow_0h7fjlb_di" bpmnElement="SequenceFlow_0h7fjlb">
    <di:waypoint x="308" y="364" />
    <di:waypoint x="308" y="259" />
    <di:waypoint x="406" y="259" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNEdge id="SequenceFlow_0pavv3j_di" bpmnElement="SequenceFlow_0pavv3j">
    <di:waypoint x="308" y="414" />
    <di:waypoint x="308" y="514" />
    <di:waypoint x="406" y="514" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNEdge id="SequenceFlow_1eq5wl3_di" bpmnElement="SequenceFlow_1eq5wl3">
    <di:waypoint x="506" y="259" />
    <di:waypoint x="612" y="259" />
    <di:waypoint x="612" y="364" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNEdge id="SequenceFlow_0wx6x84_di" bpmnElement="SequenceFlow_0wx6x84">
    <di:waypoint x="506" y="514" />
    <di:waypoint x="612" y="514" />
    <di:waypoint x="612" y="414" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNShape id="ServiceTask_098xws9_di" bpmnElement="ServiceTask_098xws9">
    <dc:Bounds x="338" y="81" width="100" height="80" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNEdge id="SequenceFlow_05b1igv_di" bpmnElement="SequenceFlow_05b1igv">
    <di:waypoint x="308" y="364" />
    <di:waypoint x="308" y="121" />
    <di:waypoint x="338" y="121" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNShape id="ServiceTask_0tskw6a_di" bpmnElement="ServiceTask_0tskw6a">
    <dc:Bounds x="484" y="81" width="100" height="80" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNEdge id="SequenceFlow_1aiovlx_di" bpmnElement="SequenceFlow_1aiovlx">
    <di:waypoint x="438" y="121" />
    <di:waypoint x="484" y="121" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNEdge id="SequenceFlow_1azpnjf_di" bpmnElement="SequenceFlow_1azpnjf">
    <di:waypoint x="584" y="121" />
    <di:waypoint x="612" y="121" />
    <di:waypoint x="612" y="364" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNShape id="ServiceTask_1afzltv_di" bpmnElement="ServiceTask_1afzltv">
    <dc:Bounds x="406" y="349" width="100" height="80" />
  </bpmndi:BPMNShape>
  <bpmndi:BPMNEdge id="SequenceFlow_0wh71kw_di" bpmnElement="SequenceFlow_0wh71kw">
    <di:waypoint x="333" y="389" />

```

```

        <di:waypoint x="406" y="389" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNEdge id="SequenceFlow_0qgk02i_di" bpmnElement="SequenceFlow_0qgk02i">
        <di:waypoint x="506" y="389" />
        <di:waypoint x="587" y="389" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNEdge id="SequenceFlow_1y0oka0_di" bpmnElement="SequenceFlow_1y0oka0">
        <di:waypoint x="637" y="389" />
        <di:waypoint x="714" y="389" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNShape id="EndEvent_02bpwm6_di" bpmnElement="EndEvent_02bpwm6">
        <dc:Bounds x="897" y="371" width="36" height="36" />
    </bpmndi:BPMNShape>
    <bpmndi:BPMNEdge id="SequenceFlow_0wwqc4z_di" bpmnElement="SequenceFlow_0wwqc4z">
        <di:waypoint x="814" y="389" />
        <di:waypoint x="897" y="389" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNEdge id="SequenceFlow_1cv04j0_di" bpmnElement="SequenceFlow_1cv04j0">
        <di:waypoint x="192" y="389" />
        <di:waypoint x="283" y="389" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNShape id="ServiceTask_18tsa4v_di" bpmnElement="Task_1qc9nh1">
        <dc:Bounds x="714" y="349" width="100" height="80" />
    </bpmndi:BPMNShape>
</bpmndi:BPMNPlane>
</bpmndi:BPMNDiagram>
</bpmn:definitions>
    
```

Table 7. XML definition of the Legal Knowledge Graph Population workflow.

REFERENCES

- [DocBook2010] Walsh, N., Hamilton, R. L. (2010). DocBook 5: The Definitive Guide. Sebastopol, CA: O'Reilly. ISBN: 978-0-596-80502-9
- [LynxD11] Jorge González-Conejero, Emma Teodoro, & Pompeu Casanovas. (2018). Lynx D1.1 Functional Requirements Analysis Report. Zenodo.
- [LynxD41] Julián Moreno-Schneider, & Georg Rehm. (2018). D4.1 Pilots Requirements Analysis Report. Zenodo.
- [LynxD42] Julián Moreno-Schneider, & Georg Rehm. (2018). D4.2 Intermediate version of Workflow definition. Zenodo.