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

D4.2 Initial 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	Several colleagues from the pilot partners (DNV GL, Openlaws, CuatreCasas) as well as UPM and SWC have contributed to D4.2
REVIEWERS	María Navas Loro (UPM), Víctor Mireles Chavez (SWC)
VERSION | STATUS	V1.0 | Draft
NATURE	Other
DISSEMINATION LEVEL	Public
DOCUMENT DOI	10.5281/zenodo.1745324
DATE	30/11/2018 (M12)

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	01/06/2018	Julián Moreno Schneider (DFKI), Georg Rehm (DFKI)
0.2	First draft of structure	18/07/2018	
0.3	Final version of TOC	25/10/2018	
0.4	First version of introduction	29/10/2018	
0.5	Description of sections 2-5	30/10/2018	
0.6	First version of conclusions	31/10/2018	
0.7	Restructured document based on ToC reviews	06/11/2018	
0.8	Finalised first draft	09/11/2018	
0.9	Including reviews	16/11/2018	
1.0	Final version	30/11/2018	

LIST OF ACRONYMS

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

EXECUTIVE SUMMARY

This report provides an overview of the first description of the curation workflows associated with every use case (as defined in D4.1 [LynxD41]), specifically for the pilot use cases, structured in three phases:

1. Collection of all information related to the workflows.
2. Definition of the workflows needed to address every use case.
3. Conceptual reduction of the workflows based on their similarities.

This process resulted in eight workflows for the three business cases (pilots). The business cases are composed of four scenarios: *data protection*, *labour law*, *CE marking* and *geothermal energy*. From these scenarios, sets of use cases were extracted: three use cases for *data protection*, three use cases for *labour law*, two use cases for *CE marking* and two use cases for *geothermal energy*. The use cases have been defined so that they can guide the further development process by defining example users (persona approach). Each persona includes representative workflows provided by the Lynx platform.

The workflows are divided into two groups: (i) those that are commonly used in more than one use case; and (ii) those that are use case specific.

- The common workflows are:
  - LKG population
- The use case specific workflows are:
  - Contract Analysis
  - Contract Search
  - Labour Law Search
  - CE Marking Search
  - CE Marking Extended Search
  - Geothermal Search
  - Geothermal Extended Search

Apart from the workflows defined for the business cases, there is another use case, the General User Use Case, for which we will also define specific workflows to be tackled in future meetings.

Now that the initial definition of workflows has been completed, next steps include:

- Define the workflows associated with the **General User Use Case**.
- Define the workflows associated with the training of common services. For example, the Dictionary Access service will be used for the training of several services, such as EntEx.
- Finalize the definition of the workflows including the feedback obtain in the discussion held with all the members of the consortium in the plenary meeting in Vienna on November 14 and 15, 2018.
- Once all workflows are specified, the different implementation options will be explored.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

A document curation workflow is defined as the orderly execution of a series of natural language processing steps to perform a certain functionality related to the purpose of processing of documents under the umbrella of a certain task or use case. The definition of a workflow is informed by the input it receives, the output it generates and the functionality it is supposed to perform: annotate a document, add a document to a knowledge base, search for information, etc.

Lynx will offer compliance-related features and functionalities through common services and datasets included in the LKG (Legal Knowledge Graph). Therefore, the workflows must make use of these services, named Building Blocks (BB) in Work Package 3 (WP3) and datasets in order to implement the functionality required in each use case (and scenario).

In order to determine the document curation workflows required in each use case, we perform a systematic analysis of the available building blocks (performed in WP3) and match it with the required functionalities in every use case.

As described in D4.1 [LynxD41], the three pilots (business cases) are organized in four scenarios and, in turn, each of the scenarios is divided into several use cases. The initial definition of the workflows is done on a scenario level, because most of the functionalities are repeated in the different use cases, apart from some included modules.

The main steps accomplished for the initial definition of the workflows are:

- The first step of the (initial) workflow definition process was to define the different building blocks that are needed in every use case as defined in the workshops (described in D4.1 [LynxD41]).
- The second step of the (initial) workflow definition process was to define the order in which the building blocks have to be executed.
- The last step of the (initial) workflow definition process was to find shared components in the different workflows.

1.1 PURPOSE AND STRUCTURE OF THIS DOCUMENT

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

Section 2 describes the common parts or workflows needed in several use cases. Section 3 defines the different workflows required for every use case pilot (and scenario) defined in D4.1 [LynxD41]. Section 4 concludes the report.

2 GENERAL/COMMON WORKFLOWS

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

2.1 LKG POPULATION WORKFLOW

This workflow is needed in all the use cases, because in all of them documents (labour laws, GDPR, standards, etc.) have to be included in the LKG. This workflow processes documents in order to include them into the LKG. The different types or domains of documents are: GDPR, labour law and jurisprudence as well as standards, legislations and best practices related to the Geothermal domain.

The overview of the workflow overlapped with the architecture defined in WP1 (and the building blocks defined in WP3) is shown in Figure 9 in the appendix. In this figure, the sequential line shows the components used in the workflow, but it is not mandatory that the services are requested or executed sequentially. Therefore, the order of execution of the components is depicted in Figure 1.

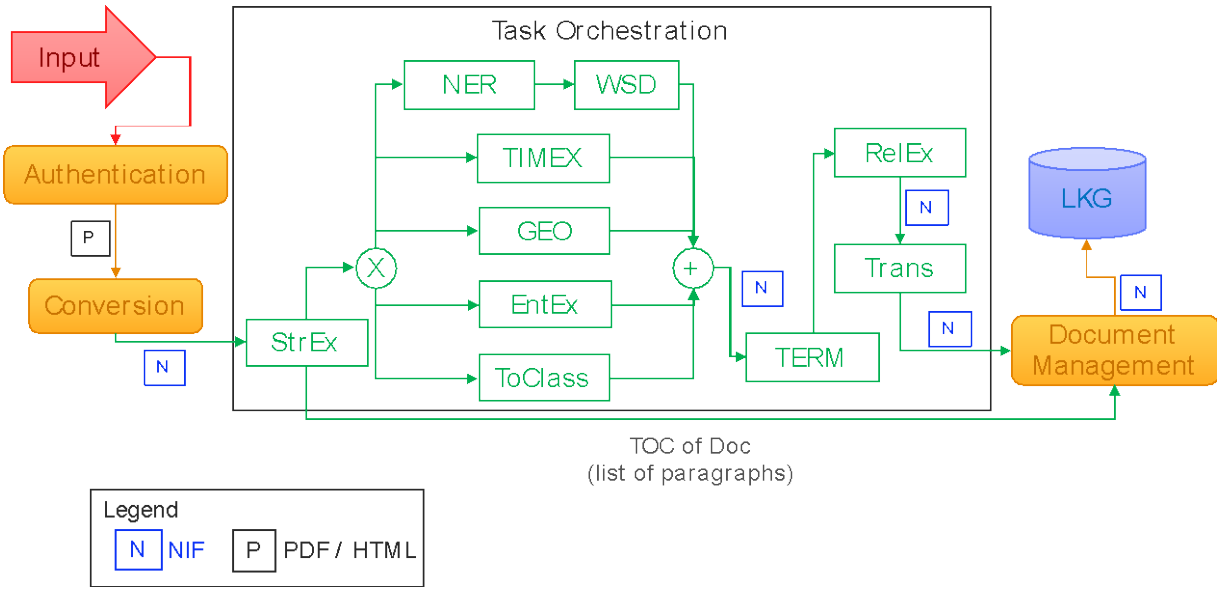

Figure 1. Execution ordering of the components used in the LKG Population Workflow.

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

2.1.1 Input

The input of the workflow will be a text document. The workflow will be able to process different formats of documents: PDF, HTML and XML.

2.1.2 Output

The output of this workflow is the generated RDF information, stored in the Legal Knowledge Graph (LKG).

2.1.3 Workflow Components (Building blocks)

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

- StrEx: analyses the structure of the document
- NER: annotates persons, locations and organizations
- TIMEX: analyses times, dates and deadlines

- GEO: analyses addresses (locations are already done in the NER)
- WSD: used to disambiguate analysed Named Entities
- EntEx: annotates specific terms (from thesaurus) in the document
- ToClass: detects the topic or main idea of a certain part of a document
- RelEx: analyses references between entities mentioned within a text and also with other documents in the LKG
- TERM: extracts specific terms from the document in order to include them in already existing or new terminologies
- TRANS: translates the document from and into different pairs of languages: ES, EN, DE and NL
- Document Manager: includes the document and the enriched information into the LKG apart from including it into the text search tool (Solr, Elasticsearch, etc)

2.1.4 Datasets

This workflow does not need any concrete datasets, apart from the LKG information in order to link the newly included document with the already existing documents.

3 SCENARIO SPECIFIC WORKFLOWS

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

3.1 SCENARIO 1 (CONTRACTS) WORKFLOWS

This section describes the workflows for Scenario 1. This scenario corresponds to business case 1, "Compliance Assurance Services in Data Protection". Its objective is to enhance compliance with data protection obligations through automation, reducing costs, corporate risks and personal risks. 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 the Scenario 1 can be found in D4.2 [LynxD42].

3.1.1 Contract Analysis Workflow

This workflow processes a contract and extends it with some metadata and enriched information. The overview of the workflow, overlapped with the architecture defined in WP1 (and the building blocks defined in WP3), can be seen in Figure 10 in the appendix. This figure depicts the sequential line of components that are used in the workflow, but it is not mandatory that the services are requested sequentially. Therefore, the order of execution of the components is depicted in Figure 2.

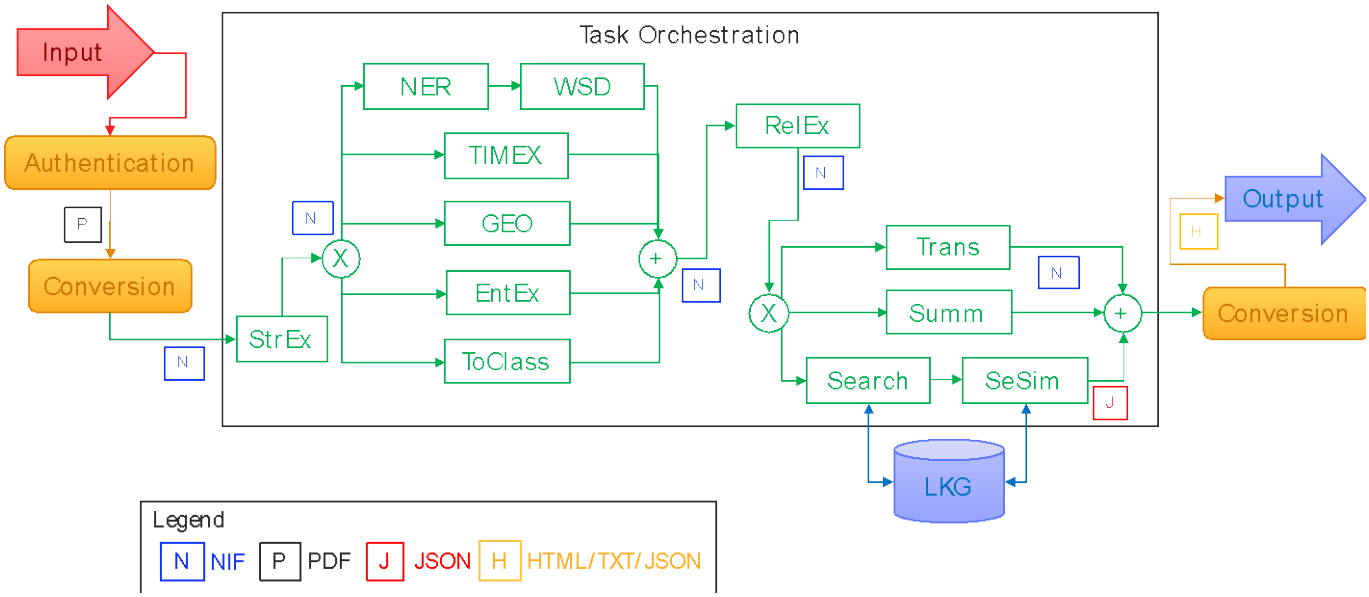

Figure 2. Specification of Scenario 1 Workflow regarding execution order of components.

3.1.1.1 Input

The input of the workflow is a contract, i.e., a PDF document containing a contract. The concrete format of the contracts' content is under discussion. We are currently experimenting with an XML-based annotation format that is based on the international DocBook standard [DocBook2010]. The idea is to transform, automatically, incoming PDF contracts into DocBook-annotated XML documents.

3.1.1.2 Output

The output of the workflow will be a PDF document containing the contract as well as some additional information that is added to the contract as metadata or as a kind of summary that can be included in the contract PDF file as its new first page:

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

3.1.1.3 Workflow Components (Building blocks)

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

- StrEx: analyses the structure of the document (including the title and possibly the signatures)
- NER: analyses persons (possibly parties) and places
- TIMEX: analyses times, dates and deadlines
- GEO: analyses addresses (locations are already done in the NER)
- WSD: used to disambiguate analysed Named Entities
- EntEx: annotates specific terms (from thesaurus) in the document
- ToClass: detects the topic or main idea of a certain part of a document
- RelEx: analyses references to the Legislation, e.g., GDPR
- TRANS: translates the document from and into different pairs of languages: ES, EN, DE, NL
- Search: this building block can be requested to find other relevant documents for the contract
- SeSim: performs a semantic similarity search that could help to filter the full text search
- Summ: generates a summary of the contract
- Privacy clauses: at the moment we are unable to specify which building block will perform this functionality. This information will be included in deliverable “D4.3 Final workflow definition”

3.1.1.4 Datasets

The datasets required by this workflow are:

- A set of contracts that can be used for training and structure analysis purposes, preferably with semantic annotations.
- The legislation and case law (e.g., GDPR) processed and included in the LKG.

3.1.2 Contracts Search Workflow

This workflow searches contracts and legal documents. There is an overlap with the architecture (see Figure 11 in the appendix). The order of execution of the components is depicted in Figure 3.

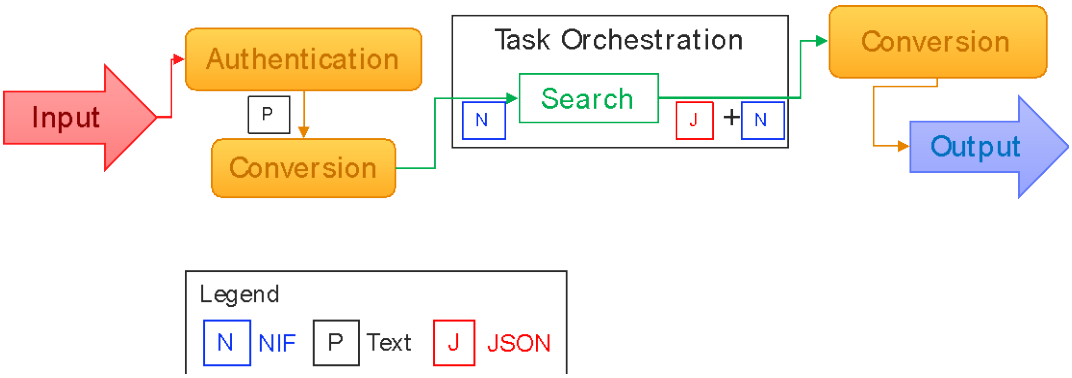

Figure 3. Specification of Contracts Search Workflow

3.1.2.1 Input

The input of the workflow will be a textual query.

3.1.2.2 Output

The output of the workflow will be a JSON document containing the list of contracts and laws that match the query.

3.1.2.3 Workflow Components (Building blocks)

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

- Search: processes the query and retrieves the related contracts and legal documents. It is a full text search based approach.

3.1.2.4 Datasets

The datasets that are required by this workflow are:

- The documents included in full text search index and in the LKG.

3.2 SCENARIO 2 (LABOUR LAW) WORKFLOWS

This section describes the workflows for the 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 two types of documents:

- EU and Member State Labour Law: labour legislations from the EU and 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 the Scenario 2 can be found in deliverable D4.2 [LynxD42].

3.2.1 Labour Law Search Workflow(s)

In this part we present two workflows together because their functionality only differs in the fact that the extended workflow summarizes the retrieved information into a text instead of returning the individual results (parts of legislation, etc). These workflows answer a set of questions related to the labour law domain. Their input are questions and the basic functionality is making a search of relevant documents. The first definition of the workflow(s), overlapped with the architecture defined in WP1 (and the building

blocks defined in WP3), can be seen in Figure 12 in the appendix. For making it more readable, Figure 4 only depicts the involved elements in the workflow(s) in the order that they are requested.

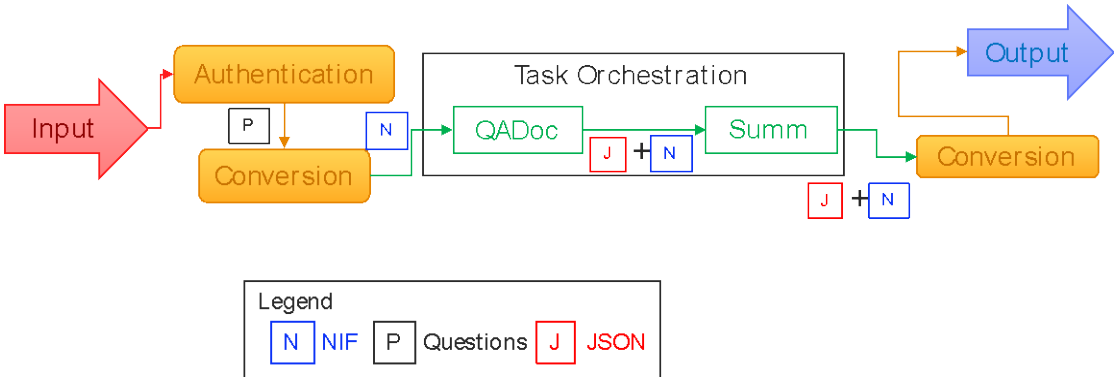

Figure 4. Detailed description of the Labour Law Search Workflow.

3.2.1.1 Input

The input of the workflow is a set of questions (or sentences) together with some additional information for every question. This information is provided as a JSON document.

3.2.1.2 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).

3.2.1.3 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).

3.2.1.4 Datasets

The datasets that are required by this workflow are:

- A set of examples of queries and additional information that can be used for training purposes.
- Labour laws of the respective jurisdictions (Spain, Europe, Austrian and maybe Dutch) included in the LKG and accessible through the QADoc.
- Labour law jurisprudence of the respective jurisdictions (not complete but as examples from: Spain, Europe, Austria) included in the LKG and accessible through the QADoc.

3.3 SCENARIO 3A (CE MARKING) WORKFLOWS

This section describes the workflows for the Scenario 3a (CE Marking). This scenario is loosely based and corresponds to business case 2, “Compliance Assurance Services in Oil & Gas and Energy”. Its objective is to explore how existing compliance-related services offered by DNV GL and existing compliance regimes within DNV GL customers could benefit from the Lynx platform. This scenario is focused on certification of CE marking. The prototype analyses two types of documents:

- Technical design: a technical design of a piece of machinery that has to be reviewed.

- Standards and regulations used in the CE Marking and certification processes to determine if a piece of machinery is suitable for being certified and gets the CE Marking.

This scenario is focused on the analysis of technical designs of products, particularly machines for which a CE marking is sought. This mark shows that a product complies with all the regulations necessary to be sold in the EU. In order to achieve this marking, it is necessary to ensure that the technical definition of a product complies with all the regulations and standards necessary in the area of application.

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

3.3.1 CE Marking Search Workflow

This workflow retrieves documents (or parts of documents) related to the CE Marking process that are relevant to the technical description of a machinery piece. The overview of the workflow, overlapped with the architecture defined in WP1 (and the building blocks defined in WP3), can be seen in Figure 13 in the appendix. In this figure is depicted the sequential line of components that are used in the workflow, but it is not mandatory that the services are requested sequentially. Therefore, the order of execution of the components is depicted in Figure 5.

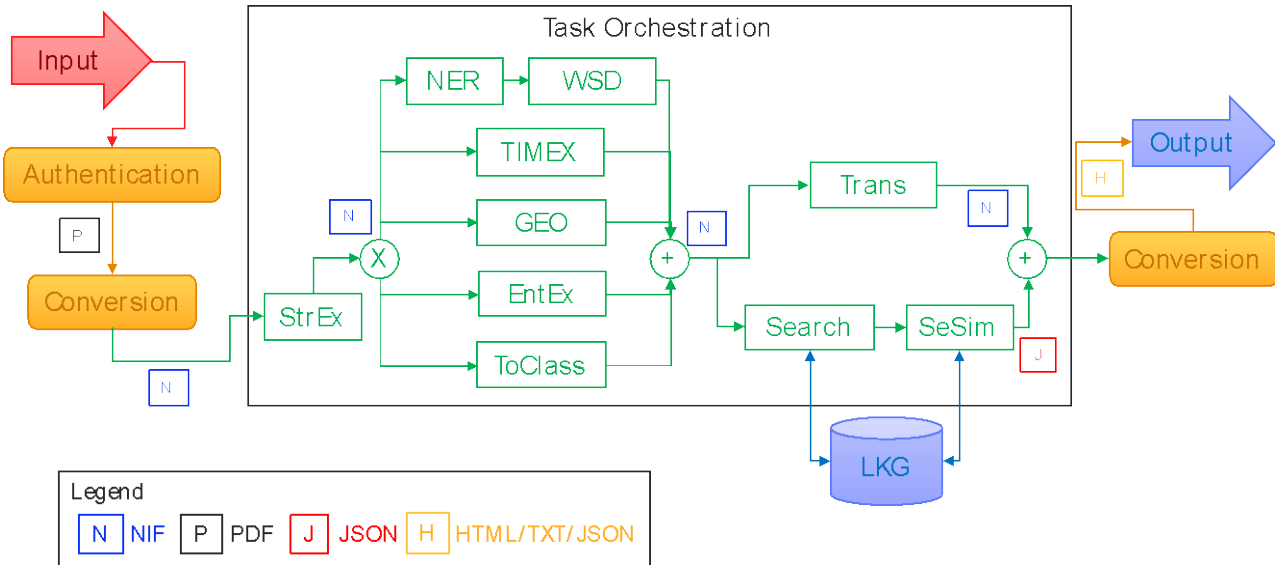

Figure 5. Execution ordering of the components used in the CE Marking Search Workflow.

As can be seen in the figure above, there are six (two of them sequentially) services that can be requested simultaneously (in parallel) while the others cannot because they need some information that should be enriched by the previous services.

3.3.1.1 Input

The input of the workflow is a technical description of a machinery piece. This information is provided as a PDF/Word document.

3.3.1.2 Output

The output of the workflow is a JSON document containing regulations, standards and best practices (or parts of them) that are related to the CE Marking process (and the technical description) of the machinery piece.

3.3.1.3 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 machinery piece technical 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:

- NER: analyses persons, locations and organizations
- TIMEX: analyses times, dates, deadlines, numbers and values
- GEO: analyses addresses (locations are already done in the NER)
- WSD: used to disambiguate analysed Named Entities
- EntEx: annotates specific terms (from thesaurus) in the document
- ToClass: detects the topic or main idea of a certain part of a document
- StrEx: analyses the structure of the document (including the title and possibly the signatures)
- TRANS: translates the document from and into different pairs of languages: ES, EN, DE and 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

3.3.1.4 Datasets

The datasets that are required by this workflow are:

- A set of technical descriptions of machinery pieces for training purposes, preferably annotated with semantic information.
- The information stored in the LKG (both public and private part) together with the information stored in the full-text search index.

3.3.2 CE Marking Extended Search Workflow

This workflow retrieves documents (or parts of documents) related to the Geothermal Energy that are relevant to a technical description of a geothermal project. The overview of the workflow, overlapped with the architecture defined in WP1 (and the building blocks defined in WP3), can be seen in Figure 14. In this figure is depicted the sequential line of components that are used in the workflow, but it is not mandatory that the services are requested sequentially. Therefore, the order of execution of the components is depicted in Figure 6.

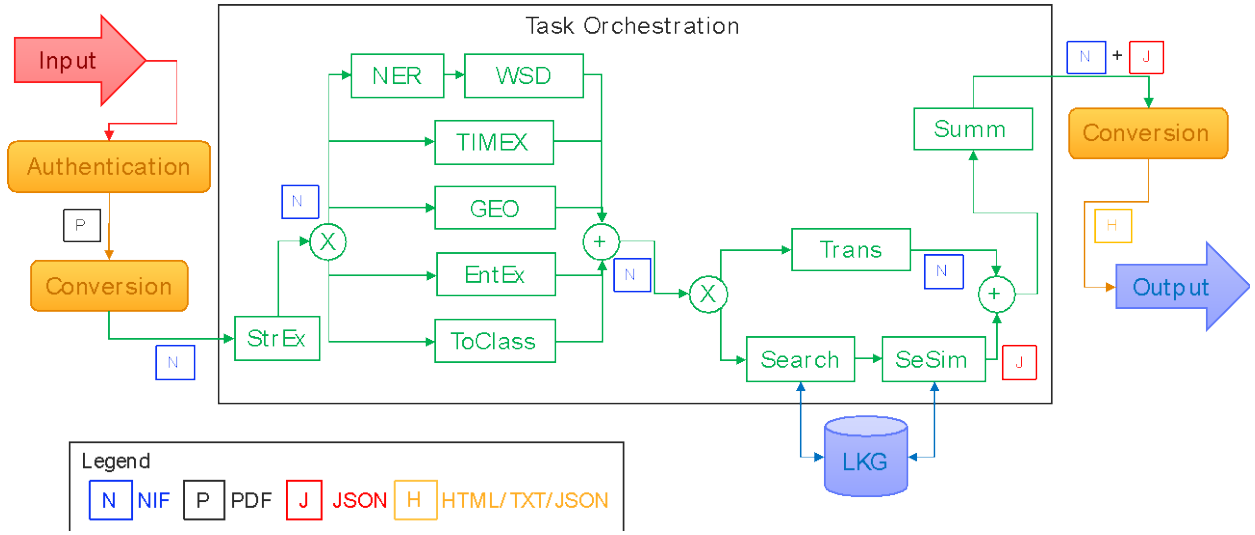

Figure 6. Execution ordering of the components used in the LKG Population Workflow.

As can be seen in the figure above, there are six services that can be requested simultaneously (in parallel) while the others cannot because they need some information that should be enriched by the previous services.

3.3.2.1 Input

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

3.3.2.2 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.

3.3.2.3 Workflow Components (Building blocks)

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

- NER: analyses persons, locations and organizations
- TIMEX: analyses times, dates, deadlines, numbers and values
- GEO: analyses addresses (locations are already done in the NER)
- WSD: used to disambiguate analysed Named Entities
- EntEx: annotates specific terms related to CE Marking in the document
- ToClass: detects the topic or main idea of a certain part of a document
- StrEx: analyses the structure of the document (including the title and possibly the signatures)
- TRANS: translates the document from and into different pairs of languages: ES, EN, DE, 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
- Summ: generates a summary of the resulting information related to the CE Marking process

3.3.2.4 Datasets

The datasets that are required by this workflow are:

- A set of technical descriptions of machinery pieces for training purposes, preferably annotated with semantic information.
- The information stored in the LKG (both public and private part) together with the information stored in the full-text search index.

3.4 SCENARIO 3B (OIL&GAS – GEOTHERMAL ENERGY) WORKFLOWS

This section describes the workflows for the Scenario 3a (Oil&Gas – Geothermal energy) in the Lynx project. This scenario is focused 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 RFP, feasibility study or other geothermal project description to the system and then is informed which standards, regulations and industry best practice must be taken into consideration to carry out the considered project in a compliant manner. This 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, 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].

3.4.1 Geothermal Search Workflow

This workflow retrieves documents (or parts of documents) related to the Geothermal Energy that are relevant to a technical description of a geothermal project. The overview of the workflow, overlapped with the architecture defined in WP1 (and the building blocks defined in WP3), can be seen in Figure 15. In this figure is depicted the sequential line of components that are used in the workflow, but it is not mandatory that the services are requested sequentially. Therefore, the order of execution of the components is depicted in Figure 7.

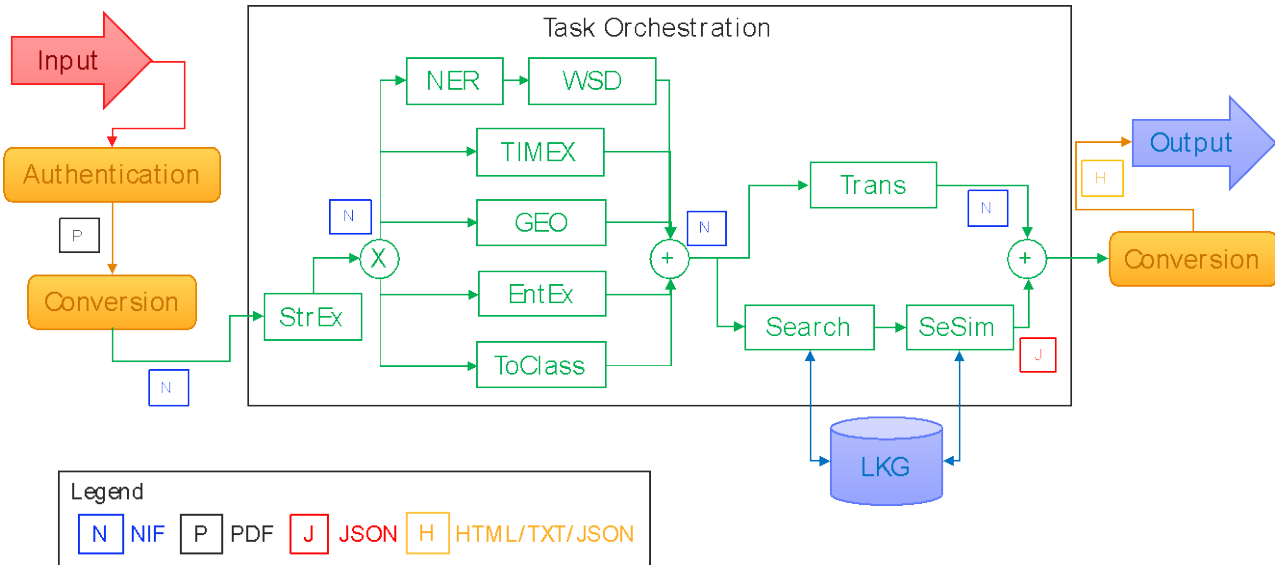

Figure 7. Execution ordering of the involved building blocks.

As can be seen in the figure above, there are six services that can be requested simultaneously (in parallel) while the others cannot because they need some information that should be enriched by the previous services.

3.4.1.1 Input

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

3.4.1.2 Output

The output of the workflow is a JSON document containing regulations, standards and best practices (or parts of them) that are related to the geothermal project description.

3.4.1.3 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:

- NER: analyses persons, locations and organizations

- TIMEX: analyses times, dates, deadlines, numbers and values
- GEO: analyses addresses (locations are already done in the NER)
- WSD: used to disambiguate analysed Named Entities
- EntEx: annotates specific terms (from thesaurus) related to Energy and Geothermal domain in the document
- ToClass: detects the topic or main idea of a certain part of a document
- IX: analyses the structure of the document (including the title and possibly the signatures)
- TRANS: translates the document from and into different pairs of languages: ES, EN, DE, 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

3.4.1.4 Datasets

The datasets that are required by this workflow are:

- A set of project descriptions for training purposes, preferably annotated with semantic information.
- The information stored in the LKG (both public and private part) together with the information stored in the full-text search index.

3.4.2 Geothermal Extended Search Workflow

This workflow retrieves documents (or parts of documents) related to the Geothermal Energy that are relevant to a technical description of a geothermal project. The overview of the workflow, overlapped with the architecture defined in WP1 (and the building blocks defined in WP3), can be seen in Figure 16. In this figure is depicted the sequential line of components that are used in the workflow, but it is not mandatory that the services are requested sequentially. Therefore, the order of execution of the components is depicted in Figure 8.

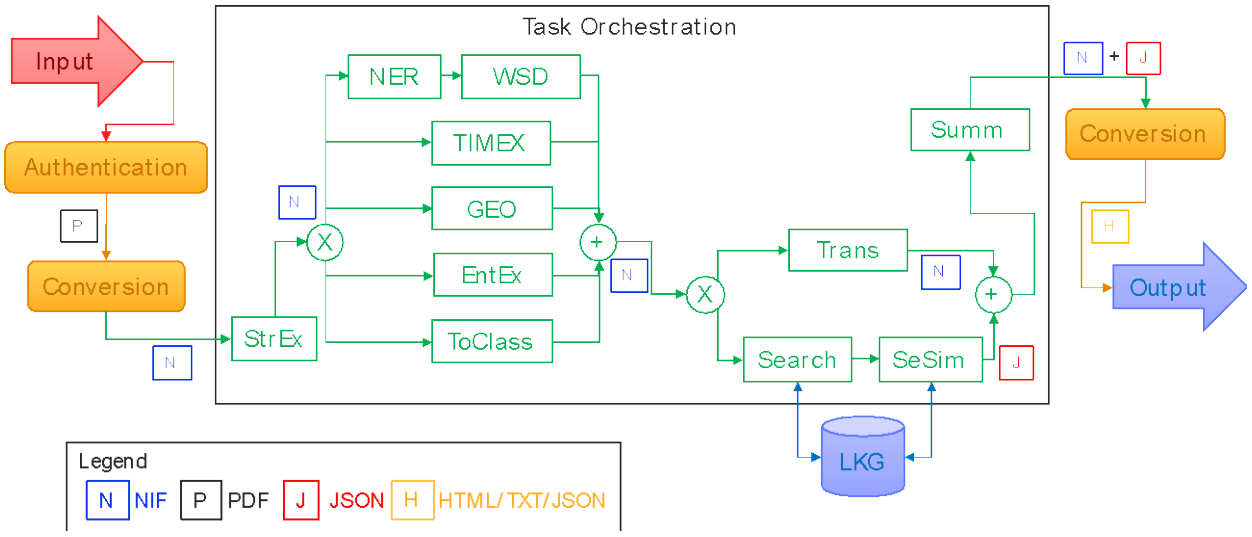

Figure 8. Execution ordering of the components used in Geothermal Extended Search Workflow.

As can be seen in the figure above, there are six services that can be requested simultaneously (in parallel) while the others cannot because they need some information that should be enriched by the previous services.

3.4.2.1 Input

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

3.4.2.2 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.

3.4.2.3 Workflow Components (Building blocks)

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

- NER: analyses persons, locations and organizations
- TIMEX: analyses times, dates, deadlines, numbers and values
- GEO: analyses addresses (locations are already done in the NER)
- WSD: used to disambiguate analysed Named Entities
- EntEx: annotates specific terms (from thesaurus) related to Energy and Geothermal domain in the document
- ToClass: detects the topic or main idea of a certain part of a document
- StrEx: analyses the structure of the document (including the title and possibly the signatures)
- TRANS: translates the document from and into different pairs of languages: ES, EN, DE, 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
- SUMM: generates a summary of the retrieved documents

3.4.2.4 Datasets

The datasets that are required by this workflow are:

- A set of project descriptions for training purposes, preferably annotated with semantic information.
- The information stored in the LKG (both public and private part) together with the information stored in the full-text search index.

4 CONCLUSIONS

This report provides a first description of the curation workflows associated with every use case (as defined in D4.1), specifically for the pilot use cases. For this purpose, we followed an approach based on three steps:

1. Collect all the information related to the workflows
2. Define the workflows needed in every use case
3. Filter the workflows based on their similarities

The first step of the workflow definition process is collecting all the information related to the workflows: information about the use cases (from D4.1), the requirements of the Lynx platform and architecture (D1.1) and the definition of the building blocks (WP3). All this information was put together and use in the second step.

The second step of the (initial) workflow definition process is preparing a description of the different workflows needed for every use case in the Lynx project. Every workflow definition is composed of four main parts: the input, the output, the datasets that are required in the workflow and the components that compose it.

The last step of the (initial) workflow definition focuses on finding overlapping or similar workflows (or parts of workflows) in the use cases. After that, the final list of workflows is divided into two main blocks: common workflows and scenario specific workflows.

The following steps that must be carried out in the project would be:

- Define the workflows associated with the **General User Use Case.**
- Define the workflows associated with the training of common services. For example, the Dictionary Access service will be used for the training of several services, such as EntEx.
- Finalize the definition of the workflows including the feedback obtain in the discussion held with all the members of the consortium in the plenary meeting in Vienna on November 14 and 15, 2018.
- Once all workflows are specified, the different implementation options will be explored.

ANNEX 1 – WORKFLOWS

This section includes the figures describing all the workflows defined in the document overlapped with the architecture defined in WP1.

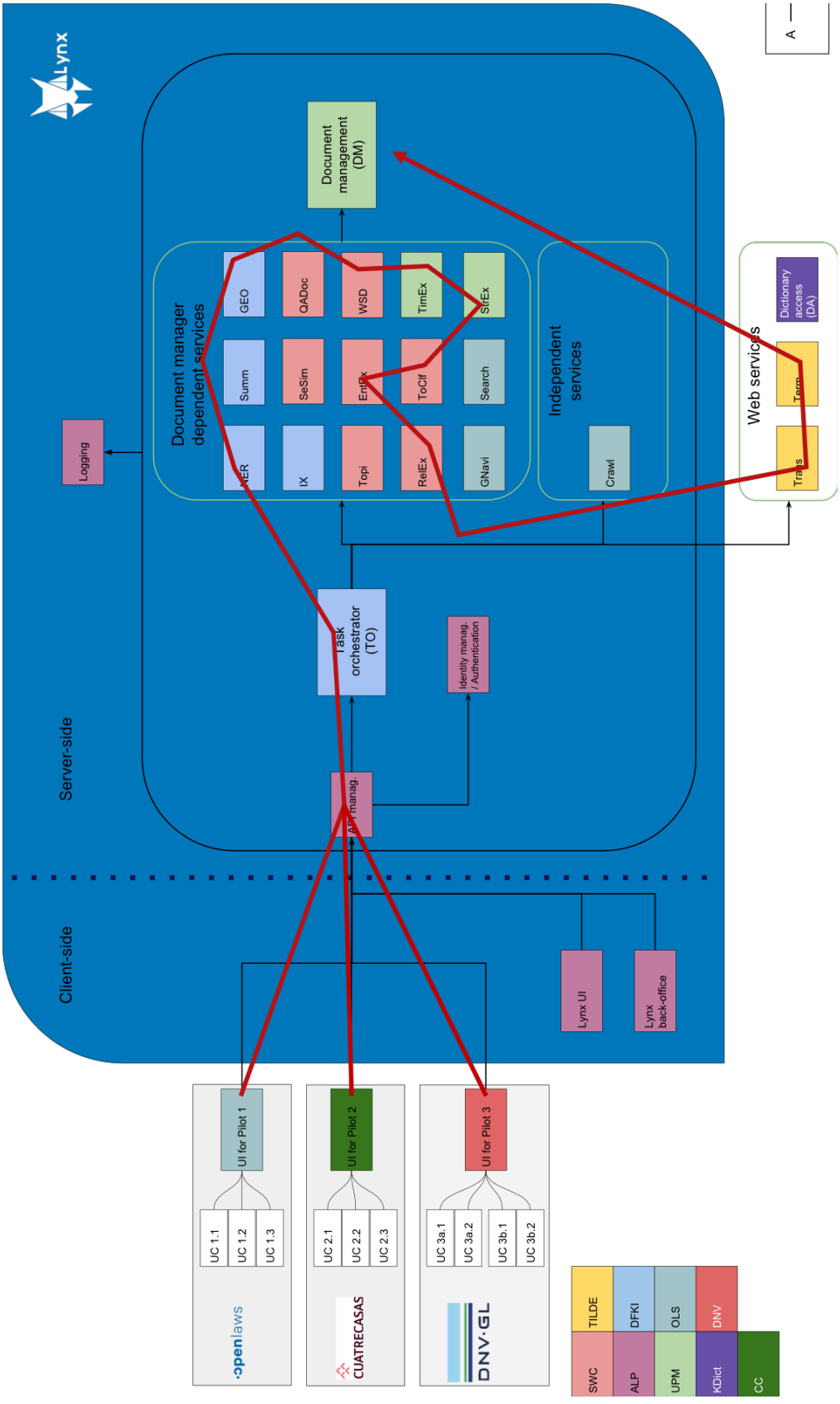

Figure 9. Specification of the LKG Population Workflow in the Lynx architecture.

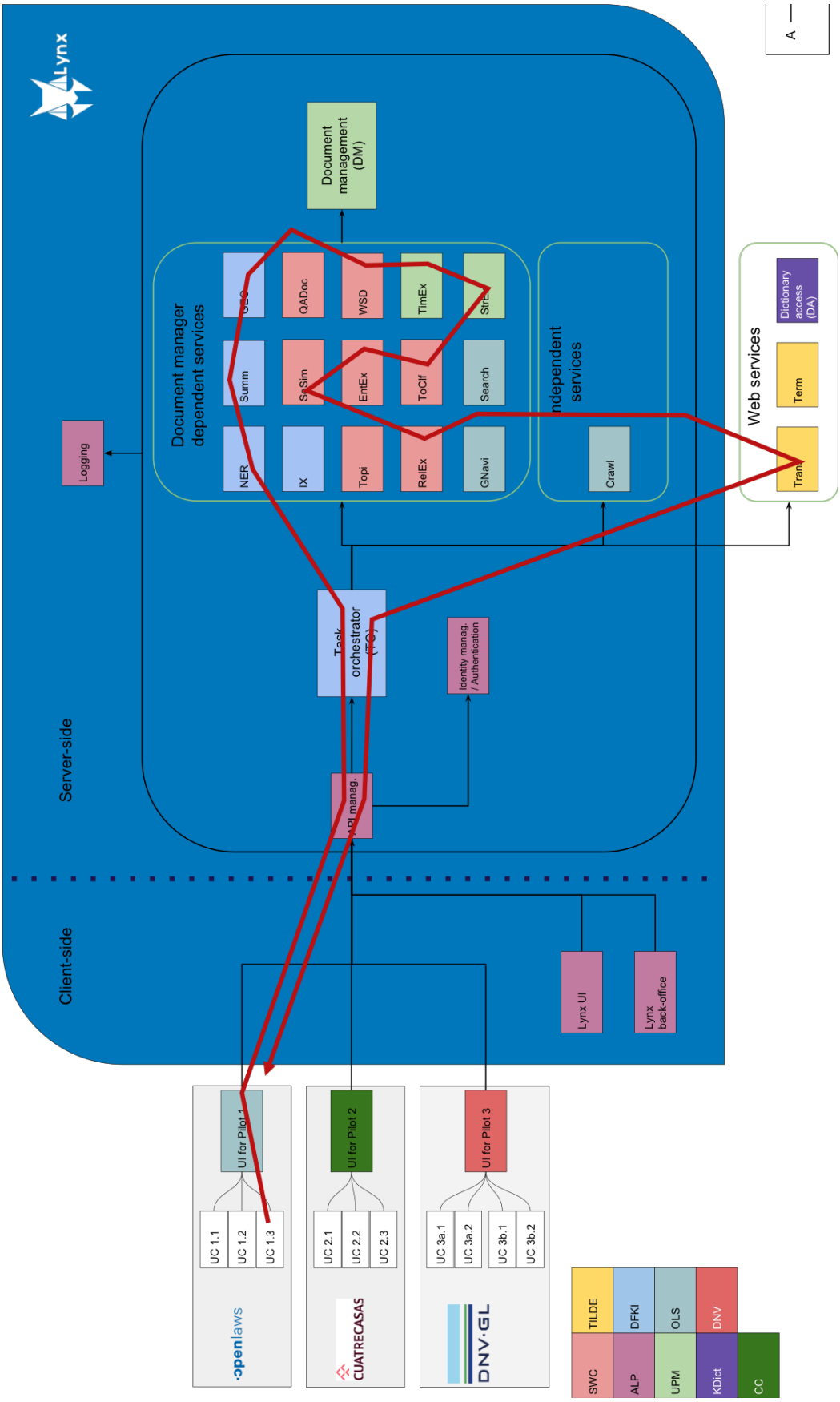

Figure 10. Specification of Contracts Analysis Workflow in the Lynx architecture

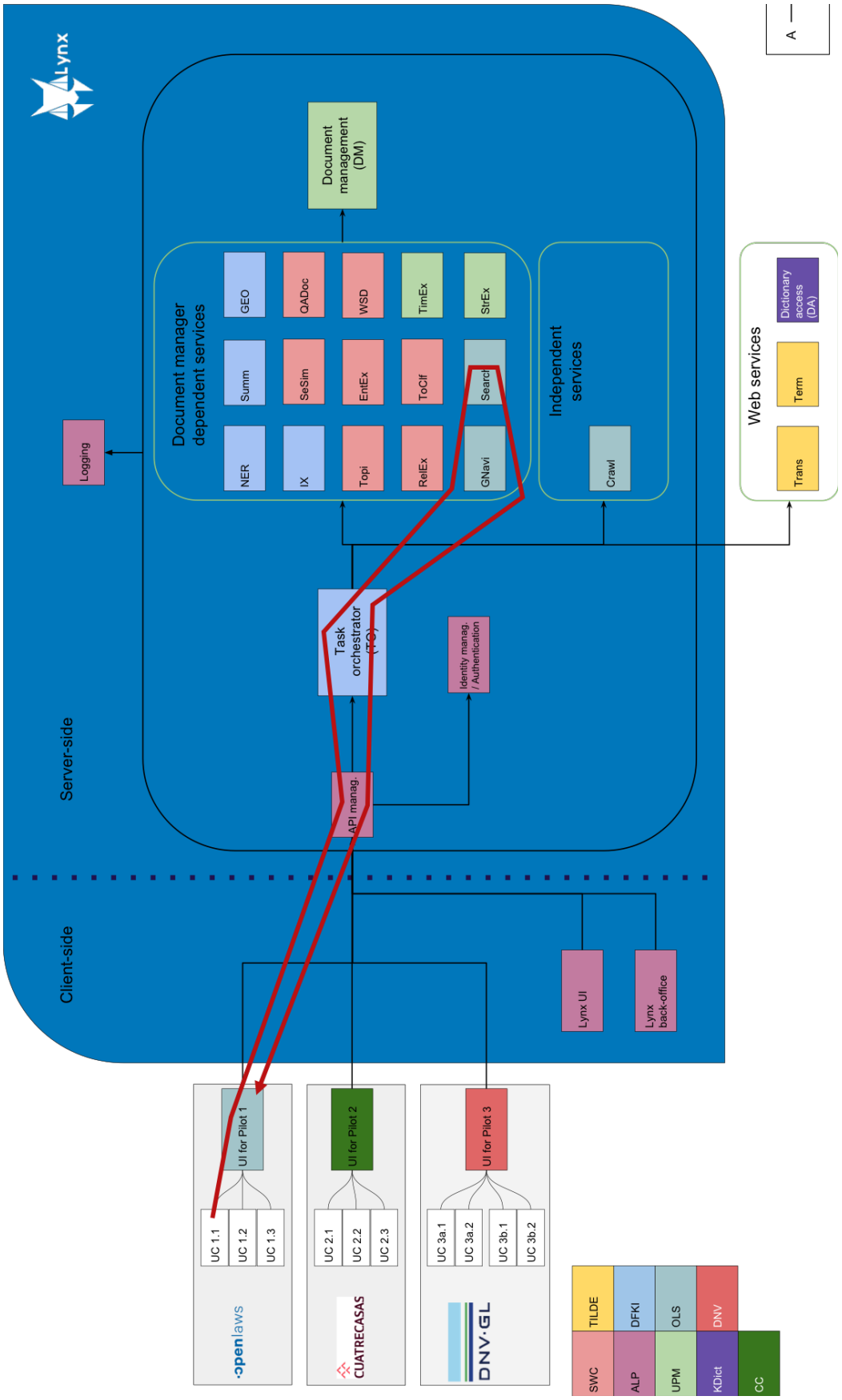

Figure 11. Specification of Contracts Search Workflow in the Lynx architecture

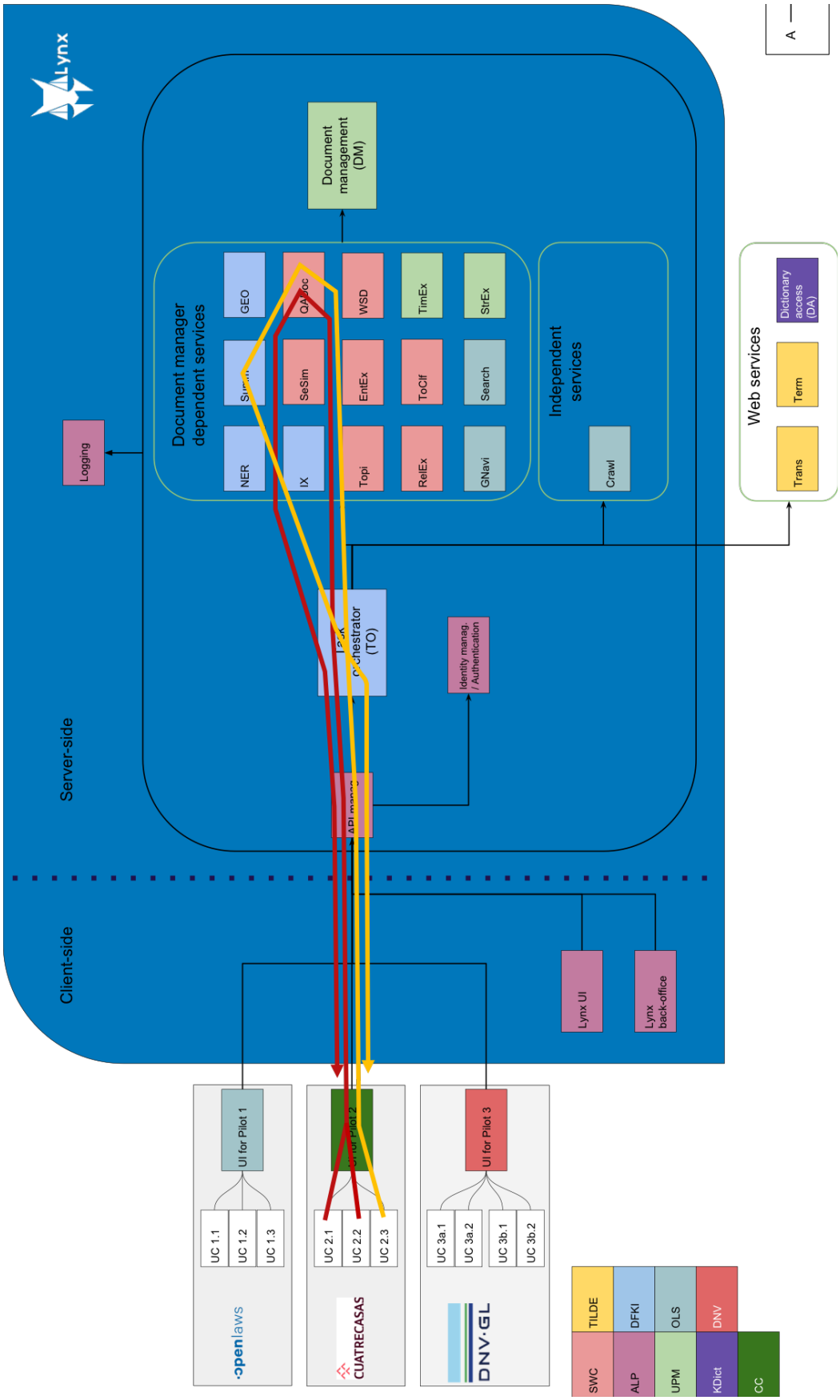

Figure 12. Specification of Labour Law Search (Standard/Extended) Workflows in the Lynx architecture

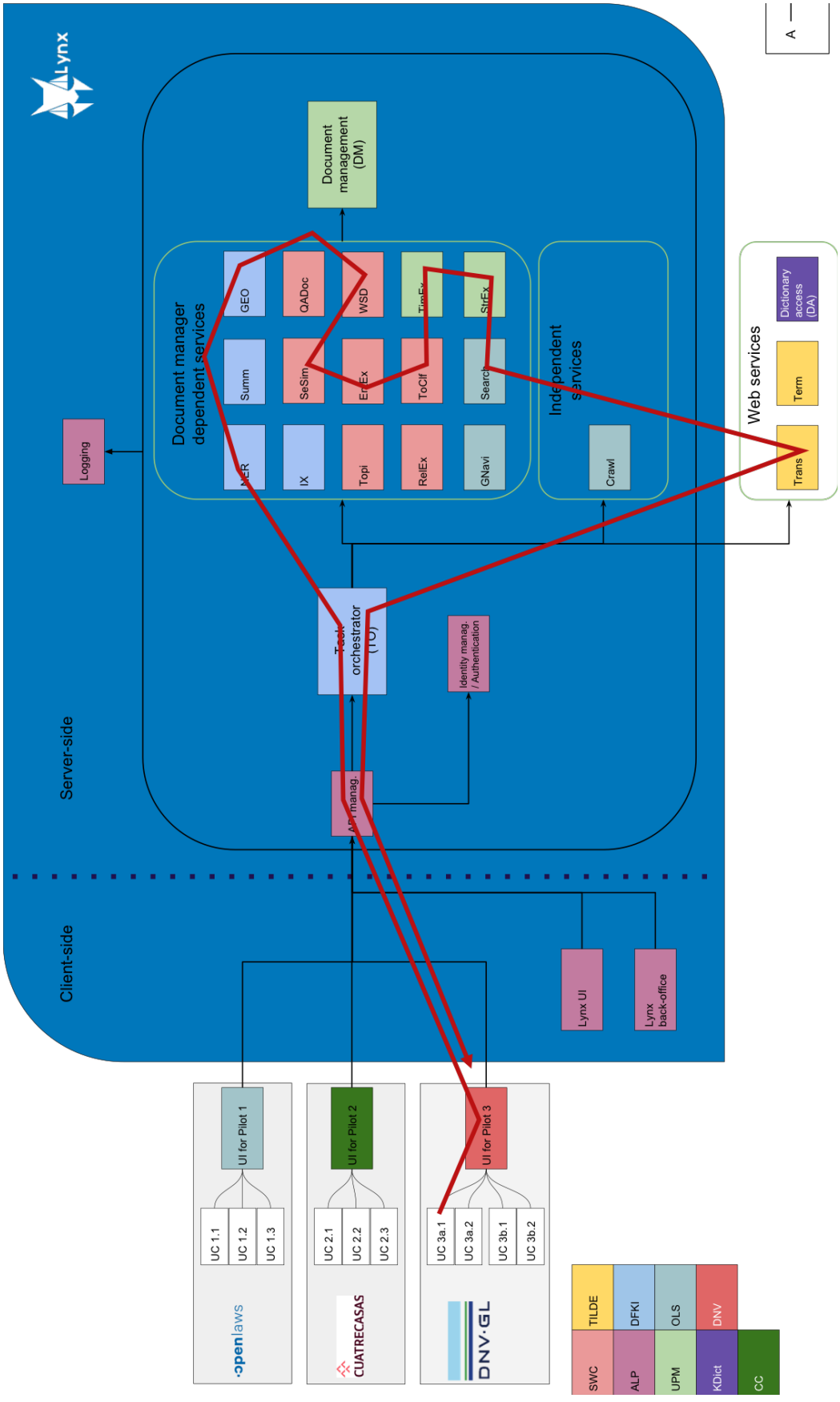

Figure 13. Specification of CE Marking Search Workflow in the Lynx architecture

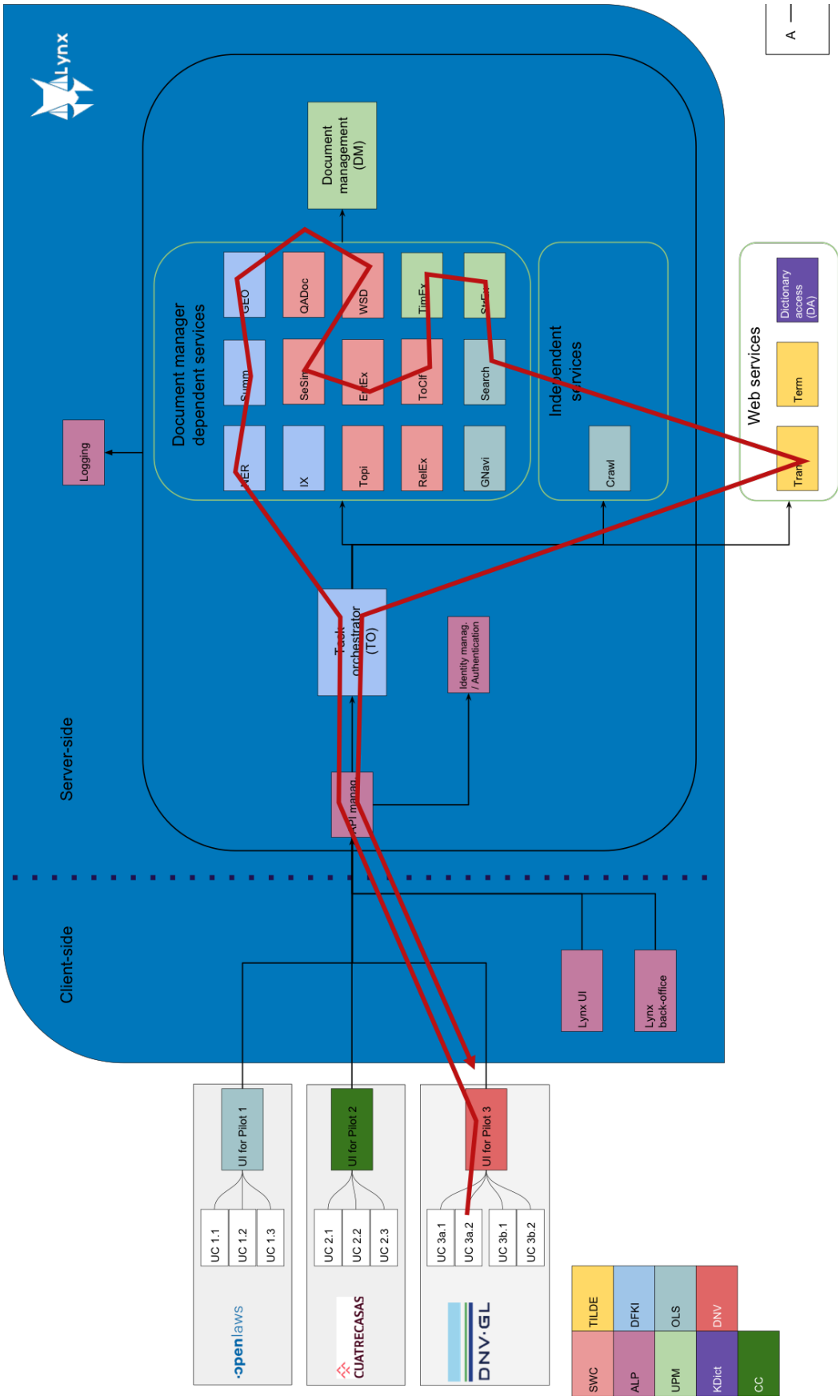

Figure 14. Specification of CE Marking Extended Search Workflow in the Lynx architecture

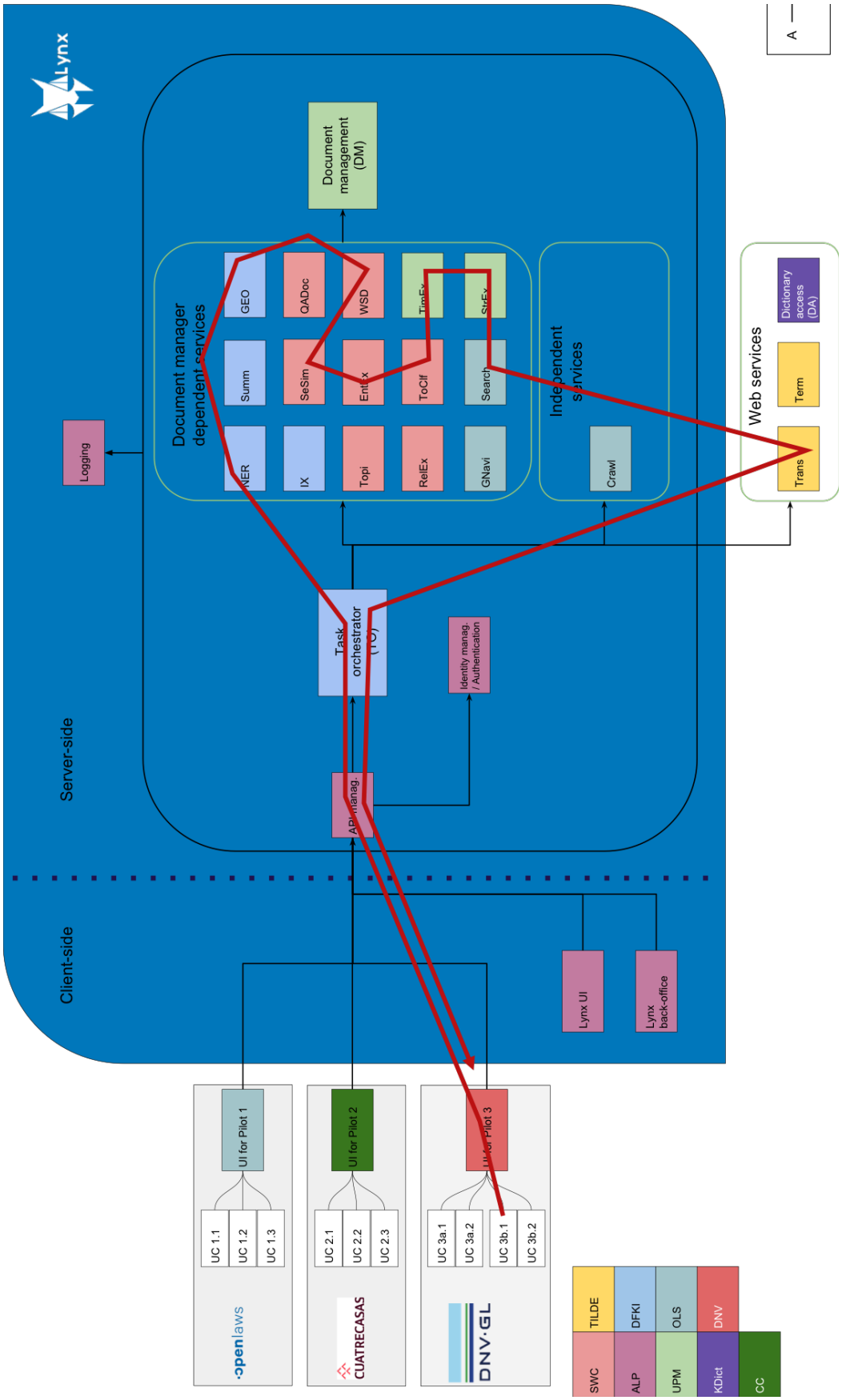

Figure 15. Specification of Geothermal Search Workflow in the Lynx architecture

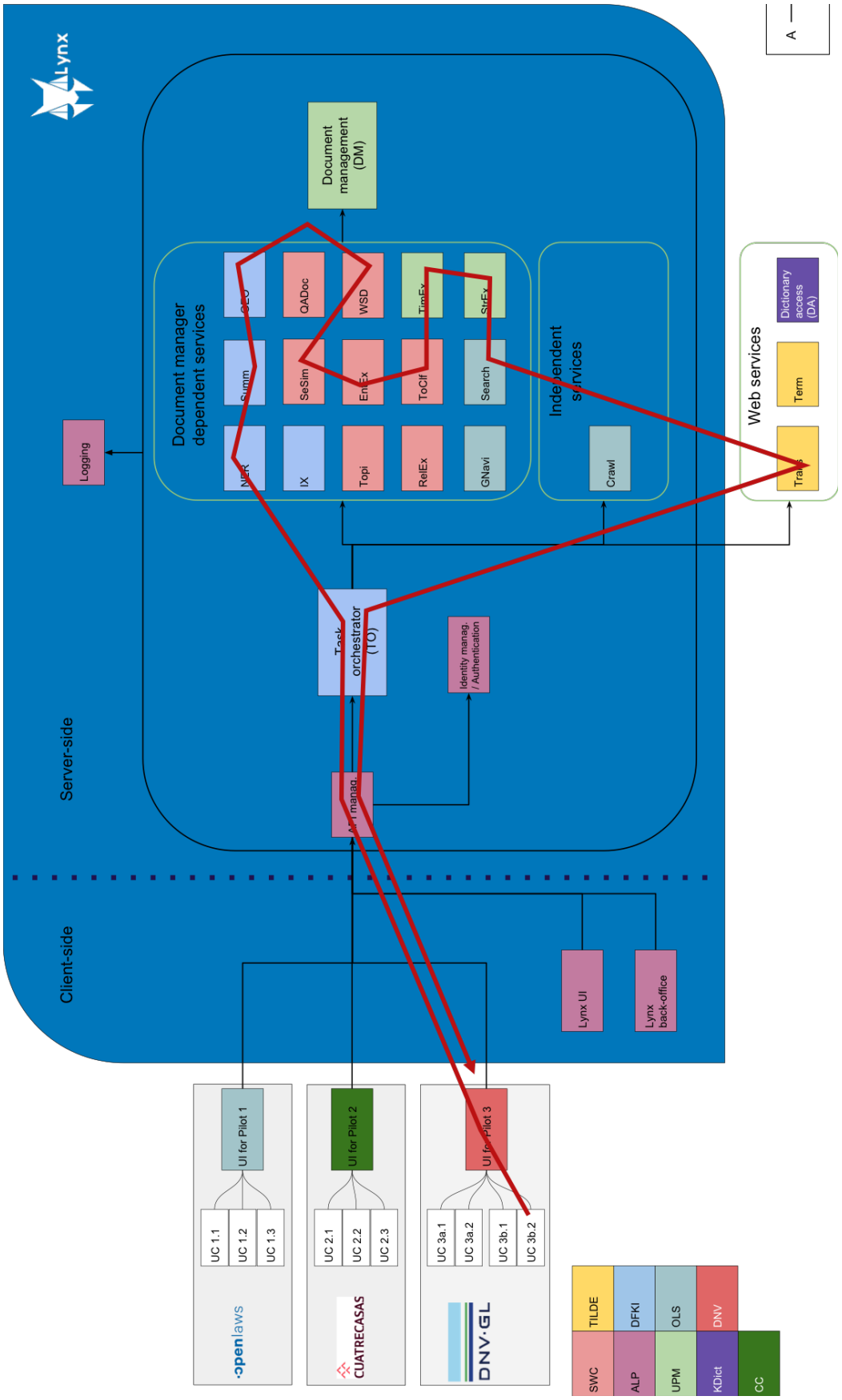

Figure 16. Specification of Geothermal Extended Search Workflow in the Lynx architecture

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.