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

D4.1 Pilots requirements analysis report

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 D4.1
REVIEWERS	Socorro Bernardos (UPM), María Navas Loro (UPM), Jorge González Conejero (UAB)
VERSION | STATUS	V1.0 | Final
NATURE	Report
DISSEMINATION LEVEL	Public
DOCUMENT DOI	10.5281/zenodo.1256830
DATE	31/05/2018 (M6)

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

VERSION	MODIFICATIONS	DATE	AUTHORS
0.1	First draft version	16/04/2018	Julian Moreno-Schneider (DKFI), Georg Rehm (DFKI)
0.2	Second draft version	30/04/2018	
0.3	Third draft version	09/05/2018	
0.4	Include conclusions and surveys in Annex 1	25/05/2018	
0.5	Include Annex 2	28/05/2018	
1.0	Final version	31/05/2018	

ACRONYMS LIST

CE:	Conformité Européenne
DoA:	Description of Action
DS:	Dataset
EU:	European Union
GDPR:	General Data Protection Regulation
GUI:	Graphical User Interface
LKG:	Legal Knowledge Graph
SC:	Scenario
UC:	Use Case

EXECUTIVE SUMMARY

This report provides an overview of the process and methodology applied in Lynx for gathering the requirements of the pilot use case partners, structured in three phases.

1. Preparation of a survey to collect, from the pilot use cases partners, their primary needs.
2. Organisation of face-to-face workshops with the pilot use case partners to collect, in a systematic and detailed way, the needs in their respective use cases.
3. Elicitation process to further specify and prioritise the requirements obtained from the face-to-face workshops to formulate concrete sets of use case specific requirements.

This three-phase-process resulted in a complex set of 42 requirements 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 in such a way that they can guide the further development process by defining example users (persona approach). Each of the personas includes representative workflows provided by the Lynx platform.

The detailed list of prioritised requirements for the three Lynx pilot business cases is shown in and 18. The requirements are divided into two groups: (i) those that focus on the functionalities of the different pilots (scenarios); (ii) those that focus on the functionality of the Lynx platform.

- Selected example features of individual pilots (scenarios) are:
  - User management
  - Upload of files and queries
  - In-document navigation
  - Simple payment feature (optional)
- Selected example features of the Lynx platform are:
  - Access to relevant documents in each scenario (GDPR, labour law, standards etc.)
  - Offer services through APIs so that they can be integrated in existing platforms
  - Allow storage of private information in the Legal Knowledge Graph
  - Manage information (and translations) in EN, ES, DE and NL
  - Accept and generate different file formats: PDF, XML, HTML, plain text

Apart from the prioritised list of business case requirements (related to functionalities of the platform), it is equally important to determine the requirements related to datasets. Each business case (and scenario and use case) requires different datasets in order to perform its functionality. The main datasets that are needed in the business cases are:

- GDPR
- Labour law
- Standards

Now that the pilots requirements analysis has been completed, the next steps include:

- Preparation of concrete software development specifications based on the requirements
- Discuss and agree upon the division of the tasks, i.e., to map services onto requirements
- Discuss and agree upon an (optional) process to update the requirements.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Driven by the legal knowledge graph, Lynx will offer compliance-related solutions based on AI-driven language technologies applied in three pilot use cases. The pilot use cases cover three different areas in our target domains. Therefore, each pilot use case requires its individual set of features and functionalities, especially with regard to the common services provided by the Lynx platform, the included datasets and document curation workflows.

In order to collect the information needed to fully develop the prototypes we perform a systematic requirements gathering process together with the three use case partners. Capturing the requirements for software systems is a challenge, especially with regard to communication: “Those who want the new software (either to use or to sell) must communicate with those who build the new software” [Nielsen1993].

There are several sources for requirements: end users, pilot use cases (knowledge workers, IT experts, legal domain experts) and other stakeholders. The main requirement input considered in this report are the pilot use cases, while the other sources are covered in Task 1.1 (Deliverable D1.1). Table 1 provides an overview of the three pilots (business cases) of the project.

Scenario	Topic	Business Case (according to the Description of Action)	Lynx Partner
1	Data protection	BC1: Compliance Assurance Services in Data Protection	Openlaws
2	Labour law	BC3: Compliance Assurance Services in Labour Law	Cuatrecasas
3a	CE marking	BC2: Compliance Assurance Services in Oil & Gas and Energy	DNV GL
3b	Energy	BC2: Compliance Assurance Services in Oil & Gas and Energy	DNV GL

Table 1. Overview of the main pilot and use case scenarios

These three pilots (business cases) are organised in four scenarios. In turn, each of the scenarios is divided into several use cases. Pilots (business cases) refer to general domains. The scenarios are application domains and, hence, more concrete, focused and specific than the pilots. The use cases (defined in each scenario) are concrete example uses of the respective pilot by the pilot partners in that scenario.

In addition to the pilots, scenarios and use cases, the Lynx platform is providing the necessary backend functionality, i.e., the pilots communicate with the platform. The architecture is shown in Figure 1.

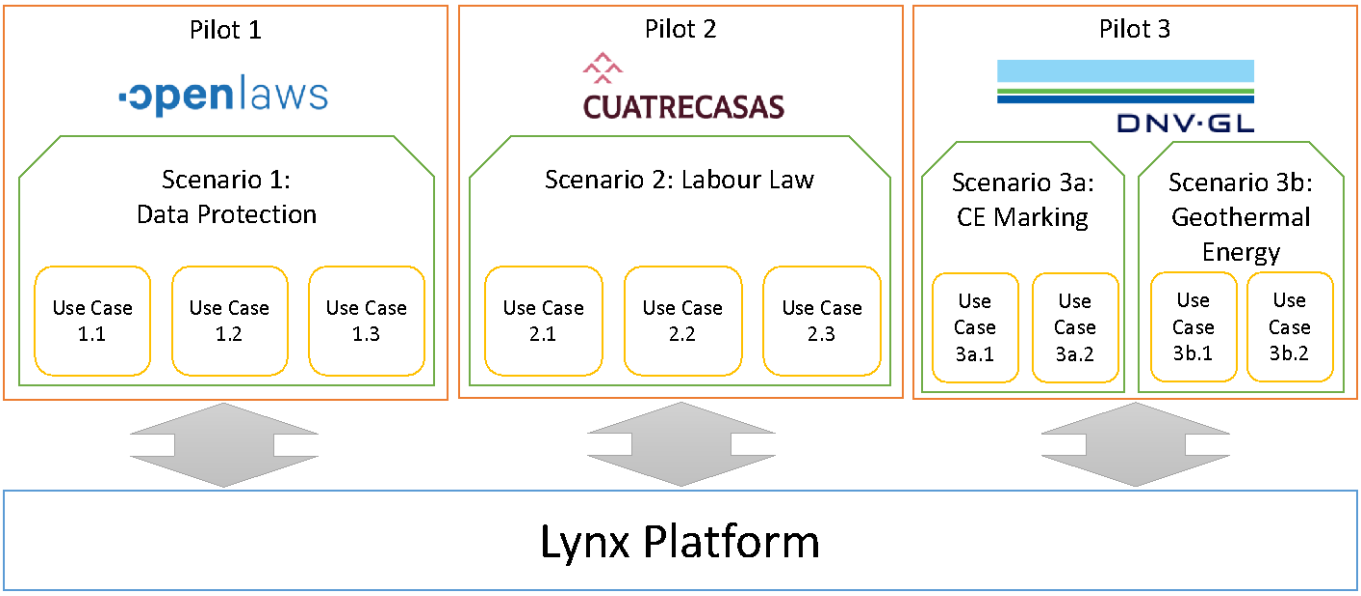


Figure 1. Pilots, scenarios and use cases

1.1 PURPOSE AND STRUCTURE OF THIS DOCUMENT

This report gathers the requirements for the use case pilots, especially with regard to their target user groups, potential customers, experts and key stakeholders. The document is aligned with Deliverable D1.1, which defines general requirements for the Lynx platform collected from general stakeholders.

It also tries to describe the methodology applied in Lynx for gathering the requirements between the pilot use cases and the platform, as well as the results from this methodology, i.e., elicited requirements. Although the requirements of the end users have to be taken into account, in this step we are not going to include the information extracted from the potential users or stakeholders because this part is covered by Deliverable D1.1.

Section 2 summarizes the requirements gathering process (methodology and overview). Section 3 describes the results of the requirements gathering surveys. Similarly, Section 4 provides the detailed results of the requirements gathering face-to-face workshops. Section 5 presents the final requirements. Section 6 concludes the report.

2 REQUIREMENT GATHERING PROCESS

One of the first steps towards the design and definition of a software project is a systematic list of requirements. Collecting these requirements is by no means a standardised process – there are many different available techniques that can be applied. Some examples are: document analysis (evaluating the documentation of a present system), feasibility study (studying similar existing systems), interview (with one more future users), observation (studying users in their workplace), prototyping (gathering preliminary requirements to build an initial prototype), surveys/questionnaires (gathering information from a small or large number of users), brainstorming (identifying all possible solutions to problems) and requirements workshop (more organised and structured than a simple brainstorming session). A complete description of requirement gathering techniques can be found, among others, in [Fricker2015].

Based on all these existing and commonly used techniques (and also taking into account the typical constraints of a research project) we opted for a combined (hybrid) approach that consists of the following three steps:

1. The first step is defining the **requirements gathering survey**, its sections and questions, the main objective of which is to collect a first set of requirements and needs that the use case partners have, i.e., trying to get an initial idea of what they expect from the Lynx platform.
2. The second step comprises **face-to-face workshops** with the pilot partners to collect more specific information on the three use cases based on the results obtained from the surveys.
3. The last step is the **elicitation process**, where the results obtained from the workshops are filtered, analysed and translated into requirements for the implementation of the pilot use cases.

This process was completed in five months (see Figure 1). The process was divided into several tasks with the participation of different entities: tasks shown in blue were taken care of by DFKI, while tasks shown in yellow also involved the pilot use case partners and tasks shown in red involved the whole Lynx consortium; deadlines are shown in green.

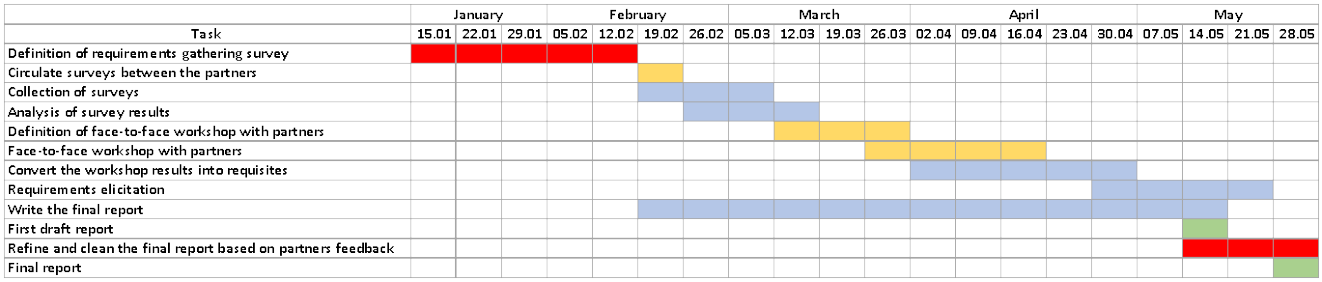


Figure 2. The requirements gathering process

3 PHASE 1: SURVEY

The first step of the requirements gathering process is a survey, the goal of which is to help us define the general needs of the use case pilots. There are several questions that have to be answered to identify the needs the pilot partners have with regard to their use cases, among others:

- What is the current usage of automatic tools that are used for legal information processing?
- Which type of documents are going to be processed in the business case?
- Which processing steps are expected/needed in the scenarios?
- What is the expected output in the use cases?

3.1 REQUIREMENTS GATHERING PROCESS

We prepared a requirements gathering survey, which is divided into several parts, each one focusing on gathering specific information related to the respective pilot use case.

3.1.1 Survey Participants

The first part is intended to collect basic information about the team involved in filling in the survey. This is to make sure that all roles involved in the development and usage of the system for this specific pilot use case are adequately covered (IT experts, knowledge workers, legal experts, etc.). The requested information was the number of colleagues involved in filling in the survey, and their role in the company.

3.1.2 Non-functional Requirements

This section is meant to compile all information related to the non-functional requirements, that is, information related to the high level characteristics of the pilot. Table 2 shows the questions of this section, as well as their respective objective.

Question	Objective
Please describe, as specifically as possible, your use case (or use cases): what kind of functionality or processing capabilities do you want to realise or achieve with the help of the Lynx platform?	Understand how and for which purpose the pilot use cases will make use of the Lynx platform.
What kind of devices do you work with predominantly? (Desktops/laptops, touch-interface devices, speech interfaces etc.)	List the corresponding devices the Lynx system should support.
Do you plan to integrate the Lynx platform into existing in-house systems and graphical user interfaces (GUIs)?	Determine if the Lynx system will be integrated into existing platforms or systems.
Please specify the system into which you want to integrate Lynx? Please provide screenshots or screencasts of the system.	Define the systems and user interfaces currently used by the pilot use case partners.
Do you currently use a stand-alone application with a GUI or web-based GUI?	Determine the type of user interfaces the pilot use case partners currently use.
How are you planning to use the services developed in Lynx? (REST API calls, Web services, Web browser, Mobile phone/tablet applications, Other)	Define how the Lynx system (and its services) will be used.

Would your preference be to develop a new (web-based) GUI to connect to the Lynx services or would you prefer some other way?	Determine if it is necessary to design a new graphical user interface.

Table 2. Questions of the non-functional requirements section of the survey

3.1.3 Automatic Processing Tools

This section compiles the information related to the legal information processing tools currently used by the pilot partners (see Table 3).

Question	Objective
How do you analyse or process legal documents in your company? (For example, with the help of human experts, fully automatically, semi-automatically etc.? Please be as specific and descriptive as possible.)	Determine the processing steps currently used by the pilot use case partner for processing legal information.
Do you use automatic solutions and tools for analysing and processing legal documents in your company? If yes, which ones?	Determine if the steps from the previous question are performed through automatic tools, and which ones are used.
What kind of documents from the legal domain (or your use case domain) do you work with (official law texts, letters, case law, EU regulations and directives, client specifications etc.)?	Determine the type of legal documents that are currently processed.
If you already use software for processing legal documents, please provide screenshots or screencasts of your software/GUIs.	Provide examples of currently used interfaces and determine their functionalities.
In terms of use cases and workflows, please specify all (or a representative set of) typical workflows that you use in-house (e.g., types of documents, analyses, approaches of producing new content, etc.).	Determine the processing workflows that are currently in place. This step includes manual and automatic steps.

Table 3. Questions of the 'automatic processing tools' section of the survey

3.1.4 Users

This section compiles the information related to the users that will use the pilots (see Table 4).

Question	Objective
What types of users are going to use Lynx services (e.g., JavaScript developers, lawyers, knowledge workers, customers, etc.)?	Determine the different type of users the pilot use case is intended for.
Do you need a multi-user solution?	Determine if a multi-user solution is needed.
Do you need authentication (login/password)?	Determine if authentication is needed.
Do you need access control lists with different roles and different permissions?	Determine if user roles and different types of permissions are needed.

Table 4. Questions of the 'users' section of the survey.

3.1.5 Datasets

This section compiles information related to the datasets necessary for the operation of the respective pilot. The idea is to describe the collections of information that will be included by the pilot partners and those needed by end users (see Table 5).

Question	Objective
What kind of reference materials or reference data sets do you use on a regular basis?	Determine the currently used datasets.
Which online data sets or reference materials would help you in your daily work?	Determine the datasets that pilot partners are interested in.
File Formats: Which are the formats of files that you want to process with Lynx? Do you want the same file format in the request you send to Lynx as well as in the responses you get back from Lynx?	Determine the needed/wanted file formats to be made available as input/output formats.

Table 5. Questions of the 'datasets' section of the survey

3.1.6 Functional Requirements

This section compiles the information related to the functionalities that pilot partners expect, would like to have or need. The questions seek to obtain specific information on each of the services available (a priori) in the Lynx system. Table 6 shows the respective questions. The "objective" column was removed because the objective of each question is to identify the need (if exists) that the pilot partner has for using a concrete common service.

Question
Do you need a tool that can **identify and highlight named entities (persons, locations, organizations, etc.) in legal documents**? For example, this could result in a colour-based highlighting of person, location, organisation names in documents or the filtering of document collections based on the names contained in them.
Do you need a tool that can **identify and highlight time expressions and normalize them**? Such a function could enable a timeline view of a large document collection, for example, of a series of letters or correspondence.
Do you need a tool that can **identify and highlight geographical information related to locations in legal documents**? For example, the output of such a function could be an interactive map containing all documents or content of the documents.
Do you need a tool that can **identify and highlight events (or other types of important keywords) in legal documents**? For example, the output of such a function could be a list of events (words, phrases, expressions, etc.) that require some kind of action or reaction from the reader.
Do you need a tool that can **identify relations between entities (some judge is related to a criminal because they are involved in a court case) in legal documents**? For example, the output of such a function could result in capabilities for searching documents containing relations through certain entities.
Do you need a tool that can **identify specific domain terminology (legal terms, oil & gas related terms, etc.) in legal documents**?

Do you need a tool that can **recognize citations, references and relations between legal documents**? For example, the output of such a function could be an interactive graph display showing the relations between all the documents of a court case or piece of legislation.
Do you need a tool that **can disambiguate the sense of a term determining if it is referring to labour law (as an example) or any other domain in legal documents**? For example, the output of such a function could be used for better determining concrete topics the document is talking about.
Do you need a tool that can **translate legal documents to other languages (if yes, which languages and language pairs?)**?
Do you need a tool that can **summarise documents or sets of documents in the legal domain**?
Do you need a tool that can **search through collections of legal documents**?
Do you need a tool that can **recommend other legal documents related to a certain task**?
Do you need a tool that can **alert you about changes in existing legal documents or the appearance of new legal documents**?
Do you need a tool that can **determine the main topic of a legal document or part of a document (paragraphs, etc.)**? For example, the output of such a function could help in searching documents for certain legislations, such as Oil&Gas or labour law.
Do you need a tool that can **determine the main type of a legal document (e.g., letter, law, contract, technical report, case report etc.)?** For example, the output of such a function could help further process and visualise a large and heterogeneous set of documents.
Do you want to combine several automatic processing steps? For example: When you get a document, the first thing you do is to translate (if it is in a language other than English), then you read it to learn which people are mentioned (locations and time expressions are also important but first are people). After that you focus on the references of other laws and finally you try to identify arguments and events.

Table 6. Questions of the 'functional requirements' section of the survey

3.1.7 Additional Requirements

The last section of the survey only had one question ('Please write down any additional requirements you may have that are not covered by the questions above.') providing the pilot use case partners with a place to share everything that they deemed important and that was not covered by any other section.

3.2 RESULTS AND ANALYSIS

Each Lynx pilot is associated with one use case partner. Each main use case partner was asked to fill in the survey on their respective pilot. We received three completed surveys (see Annex 1). Since the surveys were only the first step of the requirements gathering process, the result was not a list of requirements but a first indicative collection of the pilot partners' needs and ideas.

The first analysis of the surveys we carried out was to determine all similarities (and, therefore, differences). For this we used a spreadsheet, in which we included the results of each survey (Figure 3).

2			o Optional: Other prominent keywords, "dangerous" keywords, potentially checked against some knowledge graph ("profiling", "automated decisions", "health data", "liability", "contractual penalty"...) o Optional: Structure the document based on headings, possibly detect annexes * Save meta-information (in a PDF as custom meta-data or in a specific lynx metadata file/files, incl. links to GDPR) * Wizard and summaries o Based on the information collected, we would offer a summary and some wizards/guide throughs with recommendations about what to do in order to become GDPR compliant * Secondary target: Notifications from privacy RSS feeds * Optional: Translations of the documents * Optional: Register of processing activities * Optional: Show parties (data flows?) on a map One could even think of a plugin for ERP Systems (e.g. SAP) and/or Sharepoint/Alfresco	requirements for a given product, e.g.: does it need a CE marking at all? Which are the applicable EU directives (Machine Directive, Pressure Equipment Directive, Atex directive, etc.)? How are the directive received in a specific country? Which are the applicable standard related to the above directives? Does the product need a third party certificate? The above questions are particularly problematic if the importer is not himself an expert of the product. The "lynx portal" would have certainly be a great tool for us. The above problem can also be an issue for manufacturers who wish to sell their products in the EU (e.g. from China and the USA): which standards should they use during the design? Is there a correspondence among standards from different bodies (e.g. EN standards vs ISO standards vs ASME standards)? And so on. Also in this case, the "lynx portal" would support us as consultant	
3	2 What kind of devices do you work with predominantly? (Desktops/laptops, touch-interface devices, speech interfaces etc.)	Laptops. But we normally work on RWD (responsive) mode to be able to work on several kind of devices.	* Main Device - Desktop browsers * Further Devices will be mobile / handheld devices * Openlaws is a responsive web application and can be used from any device, certain functionalities are not available on mobile devices because of usability concerns	Laptops	Laptops and Desktop computer Web browser
4	3 Do you plan to integrate the Lynx platform into existing in-house systems and graphical user interfaces (GUIs)?	Not for this use case	YES	YES	Trend is to integrate it in existing tools.
5	4 Please specify the system into which you want to integrate Lynx? Please provide screenshots or screencasts of the system.	We have an internal framework (SOAP integration platform) that covers almost 90% of key information in our systems (SAP ERP, SAP CRM, ...) and every (99%) of our system (standard or internally developed) have public/private API to interact with	We will integrate it to our existing openlaws platform (https://openlaws.com) Openlaws provide already now following high level features: * Legal Monitoring: The customer will be actively informed of changes in the legal situation (Austria and EU). * A combined search for Austria and the EU legislation and case law. * Collaboration: Keep an overview of standards together. * Underline and comment functionalities (can be shared within a group). * Contract and Policy Management: Organizing contracts and internal policies, keeping track of deadlines and sharing in the team. * Management of contractors: keep contract partners in the overview and automatically request company excerpts from the contract partner. * Public folders	Microsoft Azure Cloud Ecosystem	
6	5 Do you currently use a stand-alone application with a GUI or web-based GUI?		Web-based gui, based on AngularJS which communicate with the backend through a REST API	NO	
7	6 How are you planning to use the services developed in Lynx?	REST API calls and Web services	REST API calls	Web browser	REST API calls
8	7 Would your preference be to develop a new (web-based) GUI to connect to the Lynx services or would you prefer some other way?	I will not provide an API (initially) I will provide an end-user application (internal and/or external). As a timeline ... I imagine first an internal solution for my lawyers and after that, and depending on accuracy and final value of the outputs, we could do the next step, enriching or simplifying the final solution for my customers.	We plan to integrate the Lynx service via our existing backend and provide our own developed GUI for the features.	Yes, preferable within our ecosystem: the marketplace on www.veracity.com. However, there may be budgetary constraints to have LYNX adopt a DNV GL look-and-feel, so having some sort of transition period in which we could use LYNX 'as is' would be likely.	

Figure 3. Spreadsheet comparing survey outcomes

The similarities are as follows:

- Every pilot partner uses automatic technologies to process, to a certain degree, legal information, although every pilot use case partner uses different technologies.
- All pilots plan to access the Lynx platform through REST APIs.
- Every pilot will develop its own system, platform or ecosystem, into which Lynx will be integrated, most likely through a self-developed web-based graphical user interface.
- Controlling user access is a must. The platform must define user roles and permissions to control access, i.e., there is pilot-specific private information, information only accessible through payments (e.g., standards), or public information, among others. Every type of information must only be accessible to the users that have the rights for it.
- A few datasets are needed in more than one pilot use case: GDPR law, standards.
- Regarding functional requirements, there are not many shared characteristics as of now, but most of them are nice to have (with different priorities) in each pilot use case.

The most salient differences between the three surveys are:

- Each pilot use case expects different high level functionalities from the Lynx system.
- Each pilot partner applies different methods and technologies to process legal information.
- Each pilot use case needs specific datasets. These datasets can contain both public and/or private information.
- The priority towards Lynx's common services differs among the three pilot partners.

Although the forms have provided us with important and very helpful information towards defining the requirements, the information gathered was still vague, which is why we carried out a total of four face-to-face workshops to make them more concrete (see section 3.2).

4 PHASE 2: FACE-TO-FACE WORKSHOPS

The second step of the requirements gathering process includes face-to-face workshops with each of the pilot partners. These workshops serve two purposes: to get a better understanding of the use case in question, given that the descriptions in the Description of Action (DoA) are not too detailed and also to clarify and prioritise the results obtained from the surveys.

4.1 REQUIREMENTS GATHERING PROCESS

The workshops were planned based on the results obtained from the surveys. We were especially considering two different formats:

1. If the implementation of a new or redesigned graphical user interface is a must (through the survey), the respective workshop will be planned as a design thinking workshop where the main focus will be put on the generation of GUI mockups of the new interface. Thus, the output of the workshop will not only be a list of requirements, but also a set of designs of the new interface, while the requirements would be less concrete and would have to be refined in yet another step.
2. If the result of the survey reveals that the implementation of a new graphical user interface is *not* necessary but an integration is planned (i.e., an integration of the Lynx services into the system currently used by the pilot partner), then only one workshop is needed. This workshop will focus on the study and analysis of the currently used technologies and how the users interact with them, as well as how the future workflows can be integrated into their current working processes. This format would provide more concrete results close to the final requirements.

Based on the results obtained in the surveys, none of the pilot partners need a new graphical interface. Therefore, all face-to-face workshops were planned based on the second format mentioned above, to gather more insights how the pilot partners process in-situ documents and how they currently use the semantic processing tools. In each workshop we made an effort to develop graphical representations of architectures and workflows (see below).

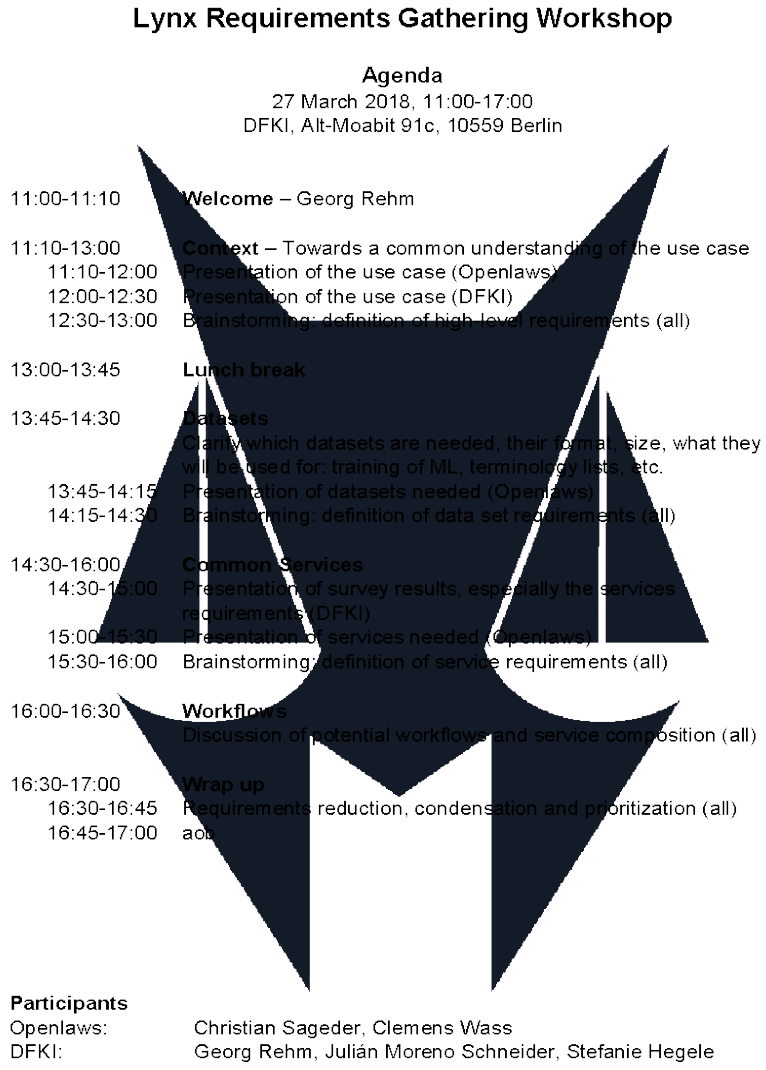
We organised a total of four workshops; two workshops were held with partner DNV GL. The basic information (date, place and participants) of each face-to-face workshop is described in Annex 2.

4.2 RESULTS AND ANALYSIS

The following sections describe the results obtained from each workshop.

4.2.1 Scenario 1: Openlaws – Data Protection

This was our first workshop. Figure 4 shows the agenda. The main points to discuss in the workshop were the general idea of the use case, the services needed to carry out the desired functionality, the necessary datasets and the interconnection between the different parties to achieve with functionality.



Lynx Requirements Gathering Workshop

Agenda
27 March 2018, 11:00-17:00
DFKI, Alt-Moabit 91c, 10559 Berlin

11:00-11:10 **Welcome** – Georg Rehm

11:10-13:00 **Context** – Towards a common understanding of the use case
- 11:10-12:00 Presentation of the use case (Openlaws)
- 12:00-12:30 Presentation of the use case (DFKI)
- 12:30-13:00 Brainstorming: definition of high-level requirements (all)

13:00-13:45 **Lunch break**

13:45-14:30 **Datasets**
Clarify which datasets are needed, their format, size, what they will be used for: training of ML, terminology lists, etc.
- 13:45-14:15 Presentation of datasets needed (Openlaws)
- 14:15-14:30 Brainstorming: definition of data set requirements (all)

14:30-16:00 **Common Services**
- 14:30-15:00 Presentation of survey results, especially the services requirements (DFKI)
- 15:00-15:30 Presentation of services needed (Openlaws)
- 15:30-16:00 Brainstorming: definition of service requirements (all)

16:00-16:30 **Workflows**
Discussion of potential workflows and service composition (all)

16:30-17:00 **Wrap up**
- 16:30-16:45 Requirements reduction, condensation and prioritization (all)
- 16:45-17:00 aob

Participants
Openlaws: Christian Sageder, Clemens Wass
DFKI: Georg Rehm, Julián Moreno Schneider, Stefanie Hegele

Figure 4. Agenda of the face-to-face workshop between DFKI and Openlaws

The workshop was held as planned, all topics and questions were discussed in detail. Notes were taken in a shared document. The main aspects were collectively developed on a whiteboard (see Figure 5).

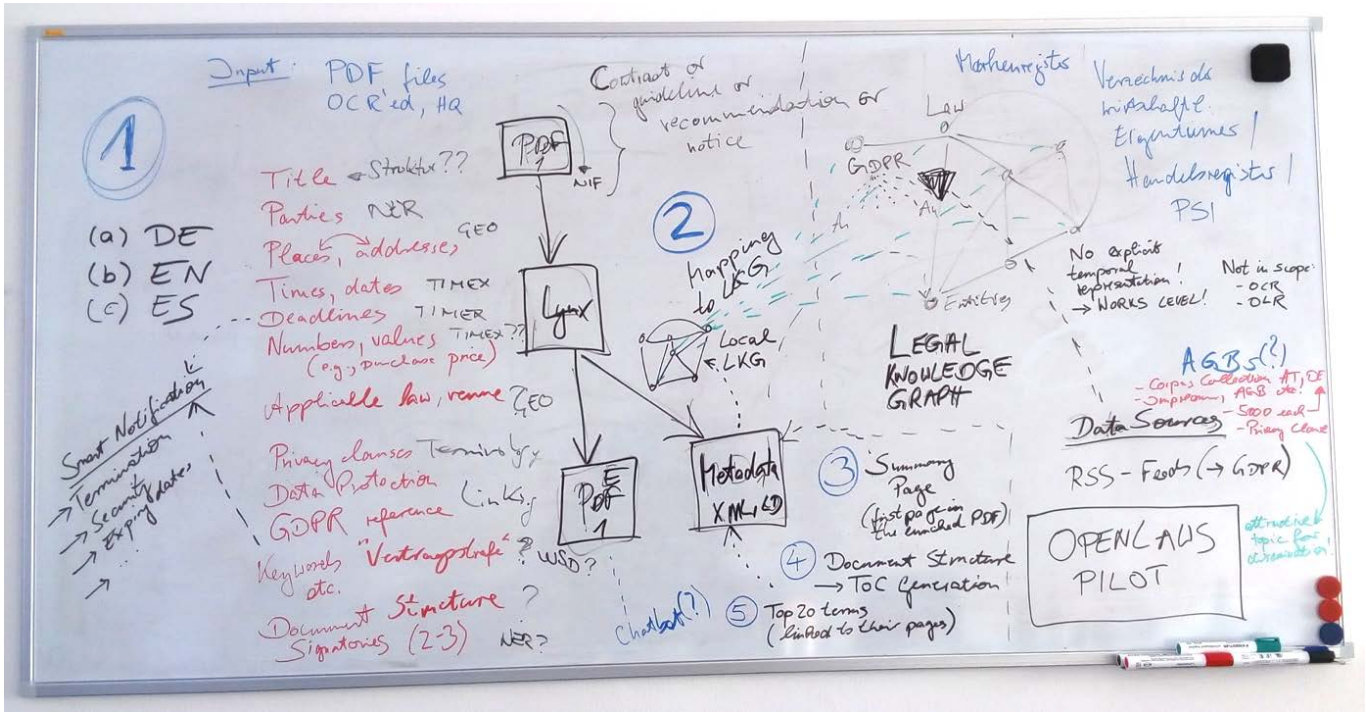


Figure 5. Graphical representation of the first pilot use case (contract analysis)

We also prepared a digital representation of this diagram (Figure 6).

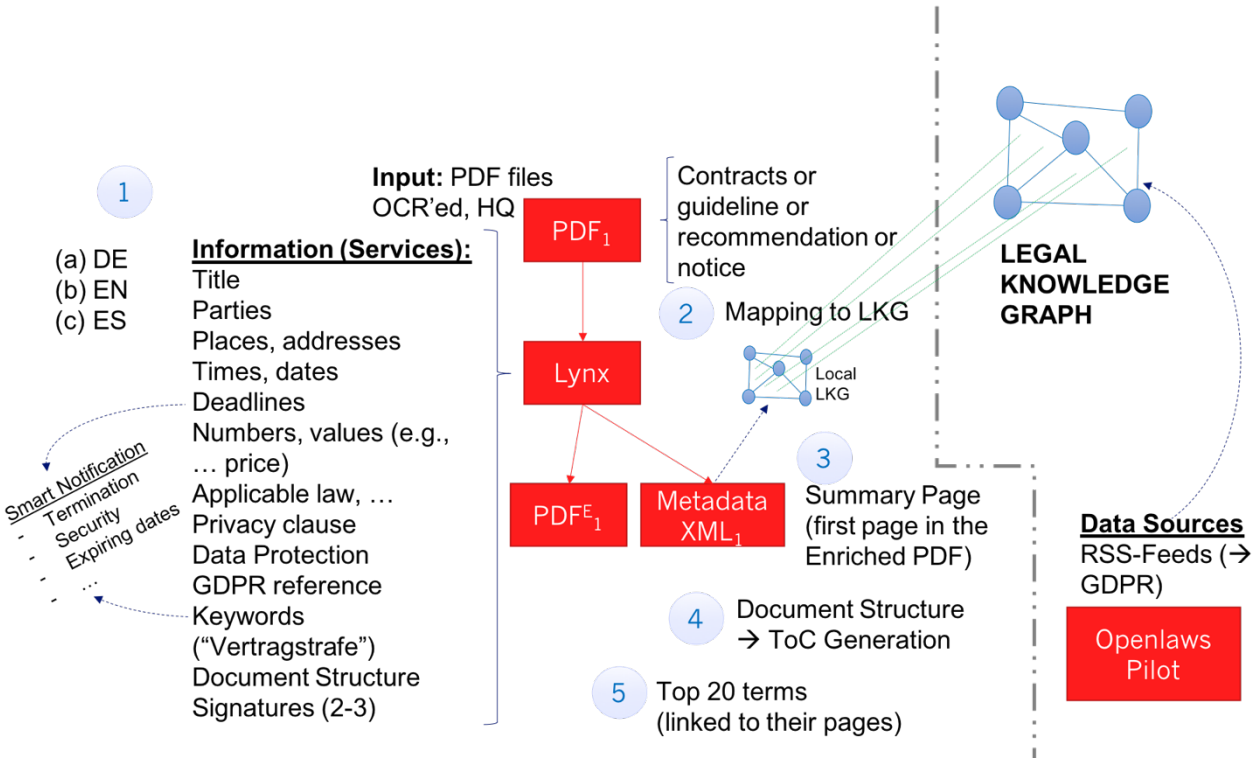


Figure 6. Graphical representation of the first pilot use case (contract analysis)

This pilot scenario corresponds to business case 1, "Compliance Assurance Services in Data Protection". The objective of this business case 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.

4.2.1.1 Use Case 1.1 (UC1.1): Free information search

Description

Alice is the controller of datasets with personal data and wants to know about specific details of a GDPR article and its application in practice. She starts Pilot1, activates a form-based search and retrieves related articles in the GDPR, relevant judgements in a jurisdiction of choice and other doctrine documents.

Alice is interested in European, Austrian, German, Spanish and perhaps Italian legislation, but she can only speak English. She would also like to get enriched documents, optionally.

Users

User	Type of user	What it does
Alice	Pilot1 free user	1. Logs in to the Pilot1 platform. 2. Makes a form-based query. 3. Gets a list of pointers to related documents. 4. Documents are optionally annotated with links and highlighted key information.

Table 7. Users descriptions of the Use Case 1.1

4.2.1.2 Use Case 1.2 (UC1.2): Premium information search

Description

Bob is the controller of datasets with personal data who pays for up-to-date information. He acts like Alice, but he gets access to curated, up-to-date information, maintained by Openlaws.

Bob likes receiving updates on relevant information.

Bob is interested in European, Austrian, German, Spanish and perhaps Italian legislations. Bob only speaks English.

Users

User	Type of user	What it does
Bob	Pilot1 premium user	1. Logs in to the Pilot1 platform (after payment). 2. Makes a form-based query. 3. Gets a list of updated documents (or pointers to them).
Openlaws	Maintains the Pilot1 premium features	Gets paid by Bob. Constantly adds and updates information.

Table 8. Users descriptions of the Use Case 1.2

4.2.1.3 Use Case 1.3 (UC1.3): Contract analysis

Description

Carol is a Data Protection Officer who reports to the CEO of her company, ensuring that the organisation applies the laws protecting individuals' personal data. She has access to dozens of contracts, and she needs to make fast inquiries on the contracts:

- Which datasets are managed by the company?
- In which contracts is party X involved?
- Which contracts are terminating next month?

Also, she would like to have better access to specific information in the contract.

Carol opens the Pilot1b website and uploads the contracts. She interactively converts some of the PDF contracts into structured information. Contracts must be represented as full-text PDF files from which the text can be extracted, i.e., not as images or scans. Important information to be annotated are:

1. Title
2. Involved parties
3. Places and addresses
4. Dates, deadlines, numbers, values (such as prices)
5. Applicable laws
6. Privacy clauses
7. Data processing / handling clauses
8. GDPR references
9. Keywords (domain-specific terminology)
10. Document structure
11. Signatures

The annotated contract will contain links to other documents in the Legal Knowledge Graph; a summary will be offered as well. Point number 5 (in the Figure 6) is the inclusion of the most relevant terms contained in the contract as well as links to the pages in which these terms appear in the manner of a glossary. Carol handles documents in English, German and Spanish, although she can only read English. The system performs the proper localisation in the output documents.

Users

User	Type of user	What it does
Carol	Pilot1b user	1. Logs in to the Pilot1b platform (possibly after payment). 2. Uploads contracts (possibly in an application installed locally). 3. Uses a wizard-based web application, in which she verifies and corrects the extracted information. 4. Downloads PDF-XML contracts. 5. Makes queries from a standard form.
Openlaws	Maintains the Pilot1 web site	Sends notifications to Carol about important events.

Table 9. Users descriptions of the Use Case 1.3

4.2.2 Scenario 2: CuatreCasas – Labour Law

This was the second workshop. The agenda is not reproduced in this deliverable because it mirrors the one of the first workshop (see above).

First, we discussed the overall scenario of the pilot. 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.

The business case begins with a company making an inquiry of labour nature in several jurisdictions where they want to open a new branch or acquire a company. In terms of input, the company will make a series of queries to answer in addition to the jurisdictions where the status of labour legislation should be consulted, as well as the number of employees that the company has or expects to have in these jurisdictions. The number of employees that exist in each jurisdiction is important because many laws and regulations depend on the number of people affected (employees).

- A system that is capable of answering a series of queries related to different labour laws is needed.
- At the input the system must receive the questions (textual consultations) in addition to additional information made up of the jurisdictions (countries) where you want to consult the legislation and the number of employees that the company that has the consultation has (or will have).
- The output of the system is foreseen in at least two formats:
    - The first output format is a basic format where the system returns, for each query and jurisdiction, the legal documents where the information relative to the query is located.
    - The second output format is optional, more advanced and based on the processing of the relative information retrieved for each query. In this case, a text composed of the most relevant information extracted from the documents is returned, processed for legibility by personnel not specialised in legislation.

Considering the multilingual nature of documents from different jurisdictions, automatic translation plays an important role. It is essential for the use case that the system can recognize the language of the input queries, translate them into the needed languages (depending on the jurisprudence where labour legislation has to be requested) and finally translates the answers into the same language in which the queries were provided.

Once the pilot use case was clarified, we summarised the information in a tentative architecture on the whiteboard (Figure 7, Figure 8).

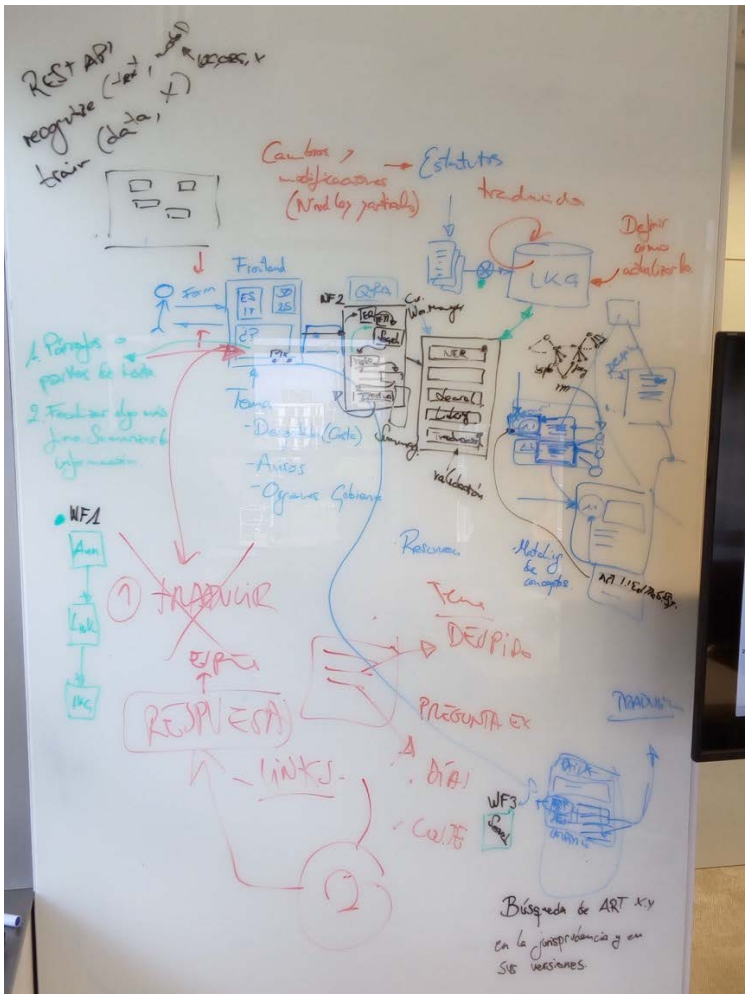


Figure 7. Graphical representation of the second pilot use case (labour law)

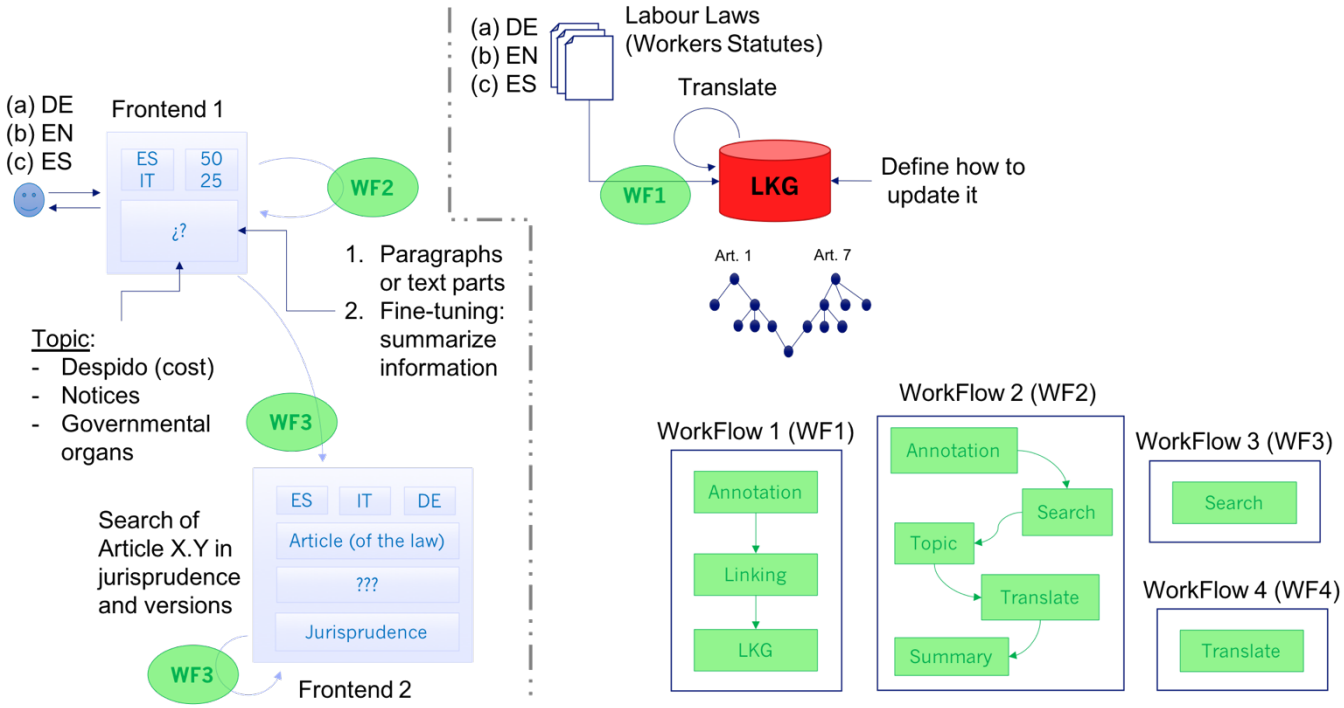


Figure 8. Graphical representation of the second pilot use case (labour law)

4.2.2.1 Use Case 2.1 (UC2.1): Expansion request

Description

Noah is a lawyer in CuatreCasas specialised in labour law. Company1 is a company that contacts Noah in order to know specific details of a labour legislation and its application in practice in different jurisdictions because they want to expand into a new market. Company1 provides Noah with a list of queries, a list of jurisdictions and some relevant related information, such as number of employees in every jurisdiction, if the company has a representative entity, etc. Noah retrieves related articles in the jurisdictions of choice (EU and Member States labour laws) and processes the information in order to provide a comprehensive report to Company1.

In order to retrieve the needed information, Noah opens Pilot2 and puts in the information provided by Company1 (queries, list of jurisdictions, i.e., country names, and related information). He retrieves related articles or their parts (sections or paragraphs) and related jurisprudence in the jurisdictions of choice (EU and Member States labour laws).

The request (queries) performed by Company1 can be expressed in any of the languages covered by the platform: German, Spanish, English or Dutch. Therefore, the retrieved documents have to be in the same language as the input queries.

Users

User	Type of user	What it does
Company1	Pilot2 enterprise	Requests CuatreCasas for information about specific jurisdictions where they want to expand its business.
Noah	CuatreCasas lawyer	1. Access the Pilot2 platform. 2. Uploads the queries, the jurisdictions list and the related information 3. Obtains parts of legislation articles in the desired jurisdictions and related jurisprudence

Table 10. Users descriptions of the Use Case 2.1

4.2.2.2 Use Case 2.2 (UC2.2): Acquisition request

Description

Company2 wants to acquire Company3, that has presence in Spain, Germany, Italy and Austria. Company2 needs to know specific details of the labour legislation and its application in practice in the jurisdictions where Company3 is present.

Company3 provides CuatreCasas (where William is a labour lawyer) with a list of queries, a list of jurisdictions and some relevant related information, such as number of employees in every jurisdiction of Company3, if Company3 has representative entities and in which jurisdictions, etc. William retrieves related articles in the jurisdictions of choice (EU and Member States labour laws) and related jurisprudence and processes the information in order to provide a comprehensive report to Company2.

In order to retrieve the needed information, William opens the Pilot2 website and puts in the information provided by the company (list of queries, list of jurisdictions and related information). He retrieves related articles in the jurisdictions of choice (EU and Member States labour laws) and related jurisprudence.

The request (queries) performed by Company2 can be expressed in any of the languages covered by the platform: German, Spanish, English or Dutch, although they are normally done in English. Therefore, the retrieved documents have to be in the same language as the input queries.

Users

User	Type of user	What it does
Company2	Pilot2 enterprise	Requests CuatreCasas for information about specific jurisdictions where they want to expand its business.
William (CuatreCasas)	CuatreCasas lawyer	1. Access the Pilot2 platform. 2. Uploads the queries, the jurisdictions list and the related information. 3. Obtains parts of legislation articles in the desired jurisdictions and related jurisprudence.

Table 11. Users descriptions of the Use Case 2.2

4.2.2.3 Use Case 2.3 (UC2.3): Private lawyer search

Description

Liam is a law enthusiast who wants to know about specific details of a labour legislation and its application in practice in different jurisdictions. He opens Pilot2b and puts in a list of queries, a list of jurisdictions and some relevant complementary information, such as number of employees in every jurisdiction, if there is a representative entity or member, etc. He expects to get a concrete answer for every query for every jurisdiction apart from parts of legislation articles and jurisprudence.

The request (queries) performed by Liam can be expressed in any of the languages covered by the platform: German, Spanish, English or Dutch. Therefore, the retrieved documents have to be in the same language as the input queries.

Users

User	Type of user	What it does
Liam	Pilot2 free user	1. Logs in to the Pilot2 platform. 2. Makes a request providing queries, jurisdictions and related information. 3. Gets a summary of the most relevant information.

Table 12. Users descriptions of the Use Case 2.3

4.2.3 Scenario 3a: DNV GL – CE Marking

This was our third workshop. The theme had changed from Oil & Gas & Energy to certificates and CE marking. This pilot 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 analyzes two types of documents (Figure 8, Figure 10):

- 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.

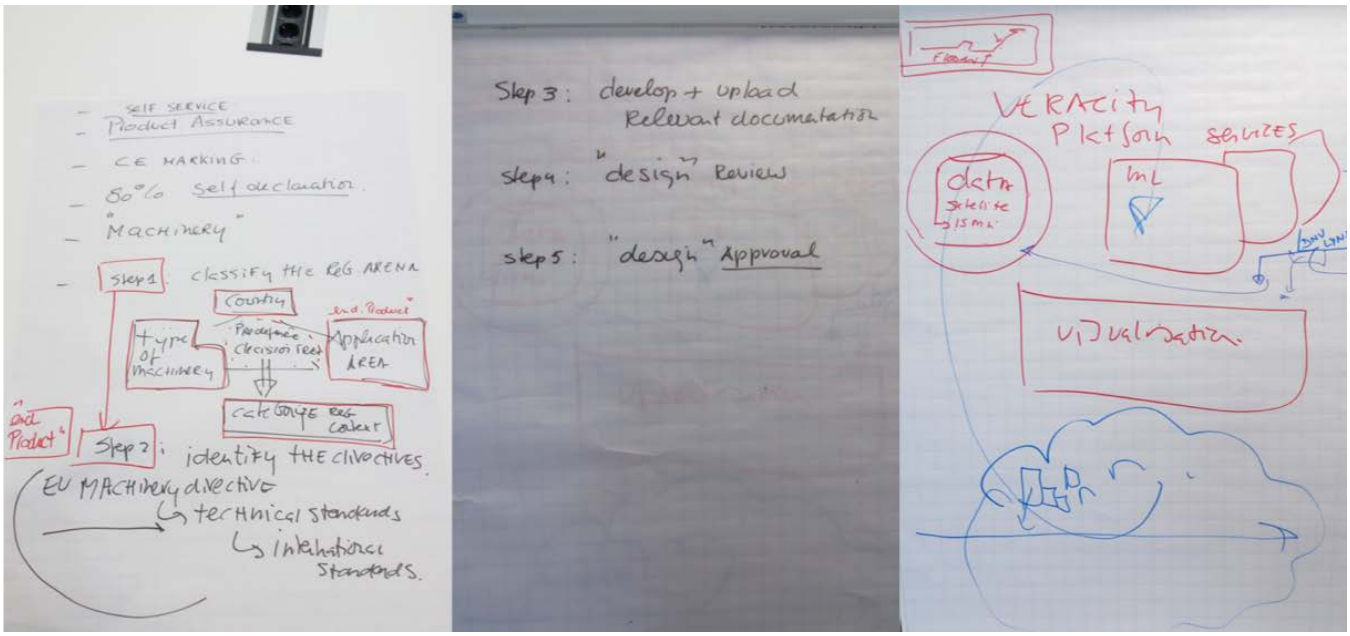


Figure 9. Graphical representation of the third pilot use case (CE Marking)

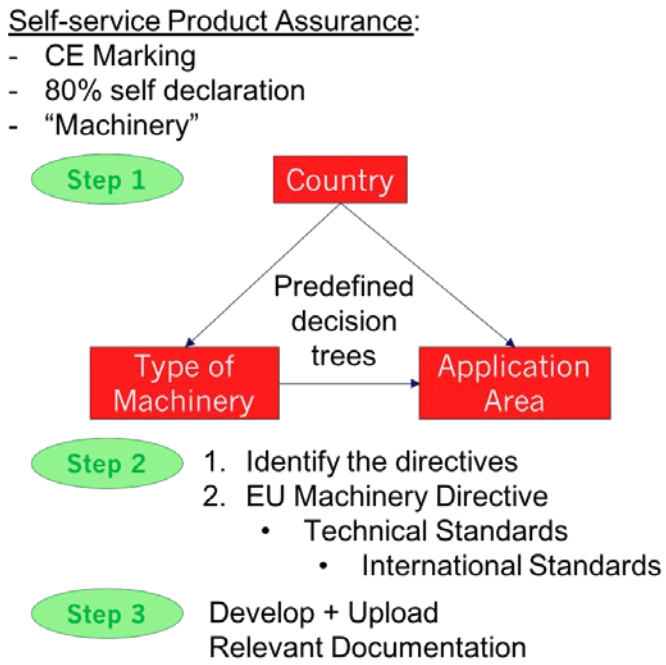


Figure 10. Graphical representation of the third pilot use case (CE Marking)

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.

As can be seen in Figure 10, this scenario is composed of three steps.

1. The first step is to classify the regulatory arena. This includes identifying the area of application of the equipment, the country in which the customer wants to sell the product and the type of

machinery. This process is based on a set of predefined decision trees, which define the regulatory context. The decision trees have been created internally in DNV GL and are part of their internal knowledge.

2. The second step is the identification of the directives that must be applied within the regulatory context. Some possible documents are standards, regulations and best practice documents. Within all available documents, the user must be offered those parts that are relevant to the product or machine in question and that are needed for the certification.
3. The third step focuses on developing the necessary documentation to carry out the product certification. When making a certification there is a certain set of documents that must be prepared and uploaded for approval. The purpose of this step is to help the user to complete the necessary documentation in a manner appropriate to the previously encountered standards.

In this scenario several different users are considered: the first user is an DNV GL worker who must check if the technical definition of the product conforms to the standards, so he would only need steps 1 and 2; a second user would be a final customer who would like to know if his design is adequate to request a certification therefore he would need steps 1 to 3.

4.2.3.1 Use Case 3a.1 (UC3a.1): Technical Definition Analysis

Description

Company5 wants to check whether and how a machine, that they want to sell in the EU can be (self)-declared for CE marking. Company5 is interested in selling products in Austria, Germany, Spain and perhaps the Netherlands.

They contact DNV GL. James has to assess which directives and standards are relevant. It is also necessary to ensure that the technical definition of the piece of machinery complies with all the regulations and standards necessary in the country where it is going to be sold. It is also crucial that the appropriate documents to be provided are compliant.

DNV GL provides Company5 with a web-based portal (Pilot3a enterprise) that ensures that the relevant data about the piece of machinery is provided in an easy way.

In order to perform the regulatory check, James runs the Pilot3a website based on the technical description of the piece of machinery provided by the client. He retrieves relevant standards and regulations in the EU and Member States. He also identifies which documents are required to self-declare the machinery equipment.

Then, James will check manually if the data of the technical description complies with the retrieved documents generating a report for Company5 with comments that guide Company5 to comply with the relevant standards and regulations.

Users

User	Type of user	What it does
Company5	Pilot3a enterprise	1. Requests DNV GL for checking the technical description of a machinery piece. 2. Provides relevant information about the type of machinery, application area and designated countries.
James	DNV GL worker	1. Access the Pilot3a-a platform. 2. Uploads the technical description of the machinery piece.

		3. Obtains directives and standards that are relevant for the certification of the concrete machinery piece. 4. Checks manually the technical description against the retrieved directives and standards. 5. Creates a report with comments for Company5.

Table 13. Users descriptions of the Use Case 3a.1

4.2.3.2 Use Case 3a.2 (UC3a.2): Certification

Description

Company6 wants to check whether and how a machine, that they want to sell in the EU can be (self)-declared for CE marking. Company6 is interested in selling products in Austria, Germany, Spain and perhaps the Netherlands.

In order to get this CE Marking, they access the Pilot3a website and introduce the technical description of the piece of machinery, countries and application. The Lynx platform has to assess which directives and standards are relevant. It is also necessary to ensure that the technical definition of the machine complies with all the regulations and standards necessary in the country where it is going to be sold. It is also crucial that the appropriate documents are being provided to be compliant.

Pilot3a website ensures that the relevant data about the machinery is provided in an easy way and that it performs the regulatory check, retrieving relevant standards and regulations in the EU and Member States, identifying which documents are required to self-declare the machine and generating a report for Company6 with comments that guide Company6 to comply with the relevant standards and regulations.

Users

User	Type of user	What it does
Company6	Manufacturer enterprise	1. Access the Pilot3a-b platform. 2. Provides relevant information about the type of machinery, application area and designated countries. 3. Obtains standards and directives, together with a report about the standards and directives that have to be considered. Besides, the system also provides a list of documents that have to be prepared for the self-declaration.

Table 14. Users descriptions of the Use Case 3a.2

4.2.4 Scenario 3b: DNV GL – Energy

This was our fourth face-to-face workshop. 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.

The result of the workshop can be seen in Figure 11 and Figure 12.

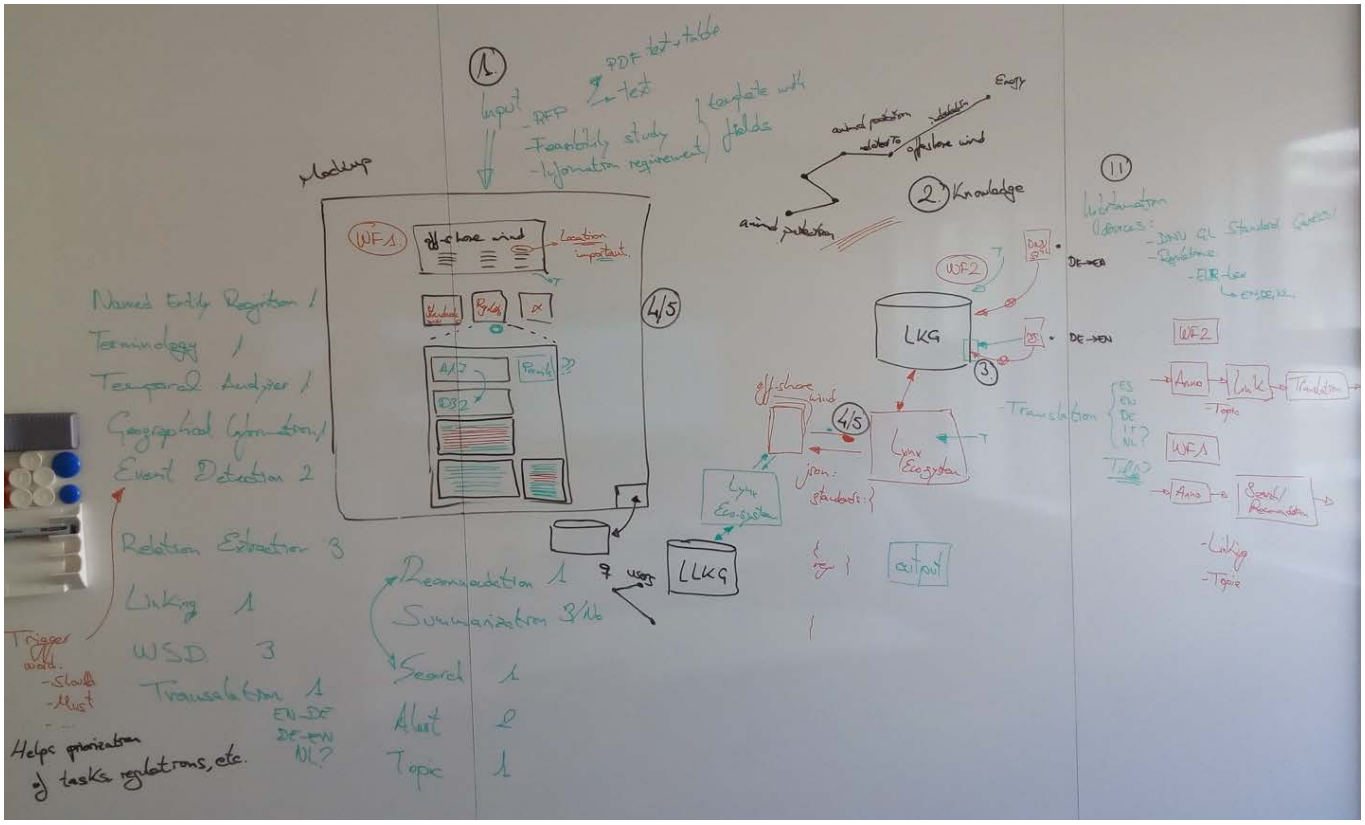


Figure 11. Graphical representation of the third pilot use case (Geothermal Energy)

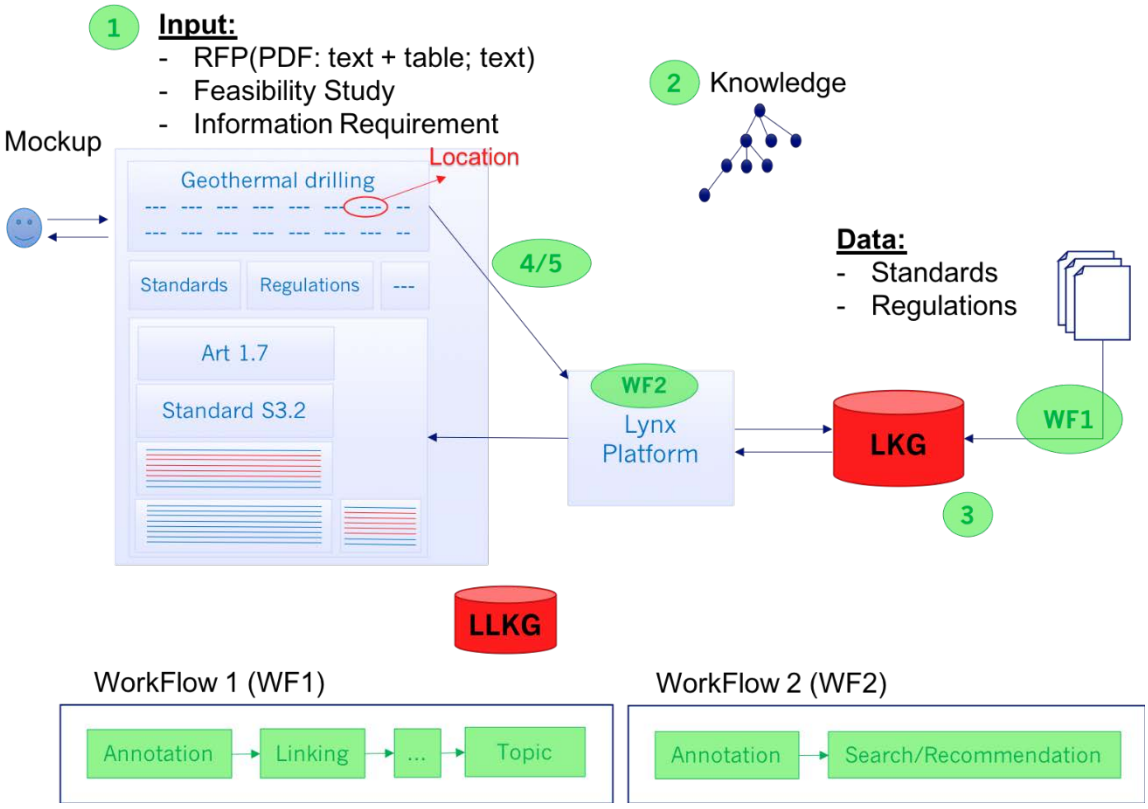


Figure 12. Graphical representation of the third pilot use case (Geothermal Energy)

As can be seen in the figure above, the scenario is composed of five steps:

1. Processing the input. As the input of the platform the user will provide a document describing the project (such as a RFP or a feasibility study (PDF), always considering that the quality of the input must be sufficient to extract the text.
2. Identification of knowledge, which is generated during the realisation of the project and will allow the identification of specific terms related to the domain in addition to existing relationships between domains, i.e., if the term drilling in the oil & gas domain is equivalent to the term drilling in the geothermal energy domain.
3. Ingestion of documents into the Legal Knowledge Graph. During this step, the set of documents and web resources (both public and private standards and international and national regulations) from which information is to be extracted, will be processed through a workflow and subsequently added to the LKG along with all the semantic information annotated in them.
4. Processing the information obtained in the RFP or feasibility study to look for applicable compliance content. In this case, another workflow is applied in which the entry document is processed and compliance content is retrieved that is applicable based on the annotated semantic information.
5. The result of the process is represented to the end user in the user interface, with categorisation of results according to their nature (e.g., directives, national law, industry standards, industry best practice). Furthermore, faceted drill down of results is possible (e.g., by country, by publication date/validity date, language). A report download option is available.

4.2.4.1 Use Case 3b.1 (UC3b.1): Geothermal Search

Description

Company8 is interested in developing geothermal projects in Austria, Germany, Spain and the Netherlands, each representing new business regions to the company. In order to ensure that they are compliant as they develop and deliver their projects, they contact DNV GL, where Thiago is the responsible person.

Due to the fact that, at the moment, not all aspects of Geothermal Energy projects are standardised, Thiago also has to look for relevant standards and best practices from other, adjacent industries, including Oil & Gas, that could apply to the geothermal project as well. Also, Thiago requires that he can identify requirements written in languages that are foreign to him. Thiago is Spanish-speaking and especially Dutch content is difficult for him. English is no problem though, so he'd rather have Dutch relevant content being translated to English.

In order to retrieve the required information, Thiago opens the Pilot3b website and submits Company8's RFP, feasibility study or other form of project description. He retrieves standards and best practices related to Geothermal Energy and other applicable fields, grouped and facetted per category. Thiago is interested to understand in which phases of the geothermal project, which compliance actions need to be undertaken, so that he can help his client with advising them about e.g. the timing of permits to be acquired.

Then, Thiago will check manually which compliance content (e.g. standards and best practices, web sources) apply to the project. Thiago generates, from the system, a report for Company8 using selections from the aforementioned results. He does some post-processing to the report before submission to Company8.

Users

User	Type of user	What it does
Company8	Pilot3b-a enterprise	Requests DNV GL for information about compliance requirements for geothermal projects in a certain region.
Thiago	DNV GL worker	1. Access the Pilot3b platform. 2. Uploads the RFP (or feasibility study) of a geothermal project. 3. Obtains compliance content relevant for the concrete project. This content is grouped into categories and can be explored using facets. 4. Checks manually how the content apply to the project and recommends how to become compliant with it. 5. Generates report based on selected relevant items.

Table 15. Users descriptions of the Use Case 3b.1

4.2.4.2 Use Case 3b.2 (UC3b.2): Geothermal project Analysis

Description

Emily is a geothermal project developer for Company9, who intends to develop a geothermal energy project in a region in the European Union. In order to do that, she has to assure that their project complies with regulations, standards and industry best practice. She has no previous experience in this process and relies on the Pilot3b website to assist her. Emily logs in to the Pilot3b website with her paid account. She remembers that at first, Company9 only had free accounts to access the Pilot3b website, but with those they could only get summary information on compliance content and the breadth of coverage across regions and types of projects was also limited. She then successfully made a case for getting paid accounts so that Company9 could comprehensively screen compliance requirements across regions and types of projects.

She then submits the RFP (feasibility study or other project description) of the project to the Pilot3b website. She retrieves a summary containing the different regulations, standards and industry best practice that have to be considered for this project in the region at hand.

Users

User	Type of user	What it does
Emily	Geothermal project developer	1. Logs in the Pilot3b platform (possibly with paid account) 2. Uploads the RFP (or feasibility study) 3. Obtains compliance content relevant for the concrete project. This content is grouped into categories and can be explored using facets. 4. Downloads a report summarising the compliance requirements (with links to the sources) that apply to the project.

Table 16. Users descriptions of the Use Case 3b.2

5 PHASE 3: REQUIREMENTS ELICITATION

The goal of the elicitation process is to prepare, based on the results of the surveys and face-to-face workshops, a concrete list of requirements. A requirement needs to meet several criteria to be considered a "good requirement" [Zielczynski2007]. Good requirements should have the following characteristics:

- Unambiguous
- Testable (verifiable)
- Clear (concise, terse, simple, precise)
- Correct
- Understandable
- Feasible (realistic, possible)
- Independent
- Atomic
- Necessary
- Implementation-free (abstract)

In addition, sets or groups of requirements should be:

- Consistent
- Nonredundant
- Complete

5.1 GATHERING PROCESS

This process took as input the requirements obtained in the second phase (workshops). The requirements and lists of requirements should have the characteristics defined above. In order to achieve this goal, all partners involved in WP4 participated in this task.

The first step was to present, to the partners, the results made in and after the workshops, both the detailed descriptions of the scenarios and the extracted requirements. During this discussion we discovered that the descriptions of the scenarios have been too general to be able to extract 'concrete' and 'atomic' requirements. This is why we also developed use cases for each scenario. The preparation of the list of concrete requirements was made a lot easier through the use cases.

We discussed the pre-final list as a basis, a discussion was held to filter and refine the list of final requirements. The final result is the list of requirements shown in Section 5.2.

5.2 RESULTS AND ANALYSIS – FINAL REQUIREMENTS

This section presents the final list of requirements regarding the pilot use cases (Table 18) and regarding the Lynx platform (Table 19). We also present dataset-related requirements (Table 20). The lists will be further refined and specified during the further development of the pilots. The priority of the requirements is indicated on four levels, from A (highest priority) to D (lowest priority).

Pilot Use Case Requirements

ID	Requirement	Applies to Pilot				Use Case	Priority
		1	2	3a	3b		
P-R1	User signup, login, logout and account management	X	X	X	X	All	A
P-R2	Support payments and manage premium users	X	X	X	X	UC1.2, UC2.3, UC3a.2, UC3b.3	C/D
P-R3	Manage PDF documents (upload, describe, list, delete)	X				UC1.3	A
P-R4	Provide and process documents	X	X	X	X	UC1.3, UC2.1, UC2.2, UC2.3, UC3a.1, UC3b.1	A
P-R4-1	Process private documents (especially contracts)	X				UC1.3	A
P-R4-2	Process labour law and jurisprudence (European Union and Member States)		X			UC2.1, UC2.2, UC2.3	A
P-R4-3	Process standards and directives related to pieces of machinery			X		UC3a.1	A
P-R4-4	Process standards and regulations related to geothermal energy				X	UC3b.1	A
P-R5	Upload queries and additional information (company information, etc.)		X			UC2.1, UC2.2	A
P-R6	Navigate documents using hyperlinks		X	X	X	UC2.1, UC2.2, UC3a.1, UC3b.1, UC3b.2	A
P-R6-1	Navigate legislation and jurisprudence		X			UC2.1, UC2.2	A
P-R6-2	Navigate standards and directives			X		UC3a.1	A
P-R6-3	Navigate standards and regulations to geothermal and other relevant domains such as Oil & Gas				X	UC3b.1, UC3b.2	A
P-R7	Provide relevant metadata, especially for technical descriptions (type of machinery, countries, application) of pieces of machinery			X		UC3a.1, UC3a.2	A
P-R8	Manage RFPs (upload, modify, delete)				X	UC3b.1, UC3b.2	A

Table 17. Pilot Use Case Requirements

Lynx Platform Requirements

ID	Requirement	Use Case	Priority
L-R1	Access to GDPR related legislation, case law and doctrine (or links to) in the LKG	UC1.1, UC1.2	A
L-R2	Summarize documents (e.g., a contract)	UC1.3	C
L-R3	Translate metadata elements (e.g., of a contract)	UC1.3	A
L-R4	Extract the metadata from a contract	UC1.3	A
L-R5	Identify GDPR-related clauses in a contract	UC1.3	B
L-R6	Generate PDF/XML documents	UC1.3	A
L-R7	Provide and process documents (including import)	UC2.1, UC2.2, UC2.3, UC3a.1, UC3a.2, UC3b.1, UC3b.2	A/B
L-R7-1	Process European Union and Member States labour legislation and its changes within the LKG	UC2.1, UC2.2, UC2.3	A
L-R7-2	Process European Union and Member States jurisprudence related to labour legislation within the LKG (private access)	UC2.1, UC2.2, UC2.3	B
L-R7-3	Process public standards and regulations within the LKG	UC3a.1, UC3a.2	A
L-R7-4	Process public standards and regulations within the LKG (regarding geothermal energy and domains such as Oil & Gas)	UC3b.1, UC3b.2	A
L-R7-5	Process private documents (standards, etc.) within the LKG	UC3a.x, UC3b.x	A
L-R8	Faceted document search and retrieval	UC2.1, UC2.2, UC2.3, UC3b.x	A
L-R8-1	Labour law articles and jurisprudence	UC2.1, UC2.2, UC2.3	A
L-R8-2	Standards and regulations (geothermal and related domains)	UC3b.x	A
L-R9	Retrieve legislation and jurisprudence mentioned in a document	UC2.1, UC2.2	A
L-R10	Provide concrete answers extracted from labour law articles and jurisprudence	UC2.3	D
L-R11	Crosslingual search (for the Lynx languages ES, EN, DE, NL)	UC2.1, UC2.2, UC2.3	A
L-R12	Translation services for documents and content	UC2.1, UC2.2, UC2.3, UC3b.x	A/C
L-R12-1	Translate services from and to Spanish, English, German	UC2.1, UC2.2, UC2.3	A
L-R12-2	Translation services from and to Dutch	UC3b.x	C
L-R13	Provide information to track the use of resources of a user	UC2.1, UC2.2, UC2.3	C
L-R14	Summarise technical information to determine relevant information from standards and directives for certification	UC3a.2	C
L-R15	Determine documents to be prepared for certification	UC3a.2	D
L-R16	Accept PDF documents as input	UC1.x, UC3a.x, UC3b.x	A
L-R17	Accept plain text queries and plain text documents as input	UC2.1, UC2.2, UC2.3	A
L-R18	Provide access to the platform through REST APIs	All	A

Table 18. Lynx Platform Requirements

Datasets

ID	Feature	Description	Priority	Scenario
DS1	GDPR	GDPR law	A	SC1
DS2	Protection Law	National Data Protection Law	B	SC1
DS3	Data Protection Jurisprudence	Data Protection Jurisprudence	C	SC1
DS4	Contracts	A set of contracts	A	SC1
DS5	Labour Law	Labour law from covered jurisdictions: Spain, Germany, Austria and Europe	A	SC2
DS6	Labour law related jurisprudence	A set of labour related jurisprudence where information regarding the labour law queries	B	SC2
DS7	Predefined Questions	A set of queries that are normally requested in the labour law domain	A	SC2
DS8	Machinery Standards	A collection of available standards (public and private) regarding Machinery	A	SC3a
DS9	Geothermal Energy Standards	A collection of available standards (public and private) regarding Geothermal Energy, and any other related topic such as Oil&Gas that could be applicable	A	SC3b
DS10	Regulations	A collection of regulations that are applicable for the geothermal energy domain	C	SC3a, SC3b

Table 19. Datasets requirements in the Lynx pilot use cases

6 CONCLUSIONS

In this report we describe the Lynx requirements gathering process for the pilot use cases. For this purpose we applied a hybrid approach that consists of a requirements gathering survey (Phase 1), together with face-to-face workshops with the pilot use case partners (Phase 2) and a final requirements elicitation (Phase 3).

Phase 1: The survey was designed to gather concrete information about the three pilot use cases. The first part collects general information about the intended use of our technologies in the respective use case. The second part is designed to learn more details about the workflows the pilot partners currently use, in addition to determining if and which automatic processes are used. The survey includes questions to learn more about the intended users of the system and questions regarding the functional requirements, which are more detailed questions about how to use each service offered by the Lynx platform. The result of the surveys is an initial and general collection of the requirements of the pilot use cases.

Phase 2: The second step of the requirements gathering process were a total of four face-to-face workshops, one for each scenario: data protection, labour law, CE marking and geothermal energy. The workshops were developed with the pilot use case partners to get a better understanding of their concrete needs related to the scenarios. In each workshop we defined a tentative architecture of the use case, together with a list of needed functionalities (which was later converted into requirements).

Phase 3: The last step of the requirements process, the requirements elicitation, resulted in a final list of requirements for the Lynx pilot prototypes and the Lynx platform.

Now that the pilots requirements analysis has been completed, the next steps include, among others:

- Preparation of concrete software development specifications based on the requirements
- Discuss and agree upon the division of the tasks, i.e., to map services onto requirements
- Discuss and agree upon an (optional) process to update the requirements.

This document will be presented at an upcoming technical Lynx meeting in Vienna on 13/14 June 2018. At this meeting we will also discuss and agree upon the next steps, some of which are mentioned above.

ANNEX 1 – SURVEYS

This annex shows the surveys collected from the use case partners.

OpenLaws

1 About the Survey Participants

As mentioned above, please fill in the survey together with more than one person, and possibly by people holding different roles (e.g., analyst, legal expert, technical expert, product manager, account manager, etc).

NOTE: We do *not* need any personal information (name etc.) but just the number of participants who contributed to filling in the survey and their respective job titles or roles.

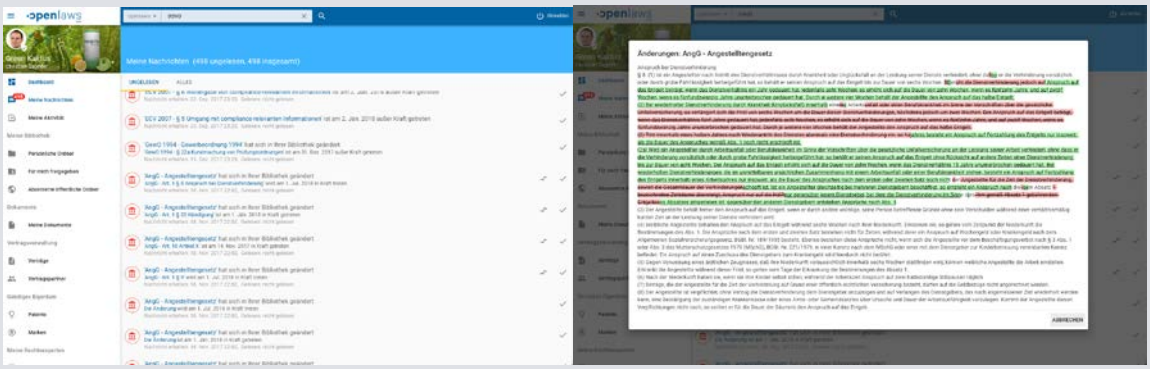
- Name of the use case partner: Openlaws
- Number of participants involved in filling in the survey: 2
- Job titles or roles:
  - Legal expert: 1
  - Analyst / Software architect: 1

2 Non-Functional Requirements

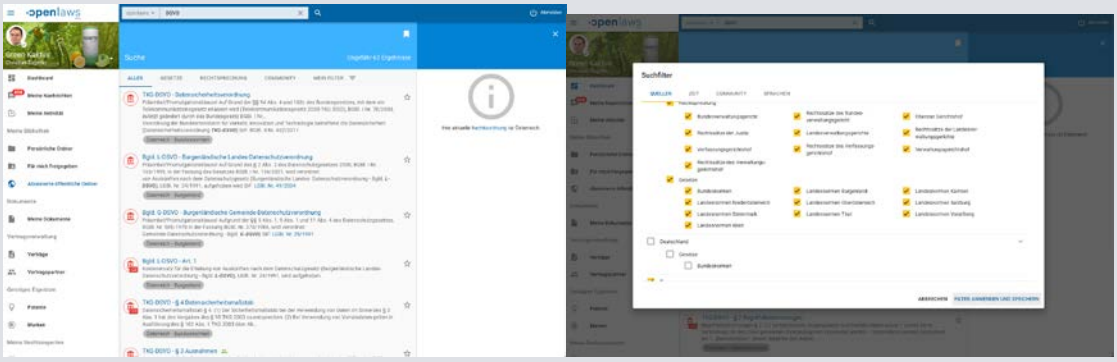
1. Please describe, as specifically as possible, your use case (or use cases): what kind of functionality or processing capabilities do you want to realise or achieve with the help of the Lynx platform?
   - (Batch) load of private documents/contracts (this service should be run locally in order to mirage concerns of companies)
   - Extract information from documents (in practice: from PDF scans, OCR should have been applied already before, not a LYNX requirement):
     - Title of the document
     - Parties (typically at the beginning of a document)
     - Places, addresses (relevant to check if there is a potential connection to outside the EU)
     - Recognize privacy clauses
     - Recognize data processing activities
     - Times & dates, deadlines
     - Numbers/values
     - Recognize references to GDPR
     - Applicable law and venue (typically at the end of a document)
     - Optional: Other prominent keywords, „dangerous" keywords, potentially checked against some knowledge graph („profiling", "automated decisions", „health data", "liability", "contractual penalty"...)
     - Optional: Structure the document based on headings, possibly detect annexes
   - Save meta-information (in a PDF as custom meta-data or in a specific lynx metadata file/files, incl. links to GDPR)
   - Wizard and summaries
     - Based on the information collected, we would offer a summary and some wizards/guide throughs with recommendations about what to do in order to become GDPR compliant
   - Secondary target: Notifications from privacy RSS feeds
   - Optional: Translations of the documents
   - Optional: Register of processing activities
   - Optional: Show parties (data flows?) on a map

   One could even think of a plugin for ERP Systems (e.g. SAP) and/or Sharepoint/Alfresco
2. What kind of devices do you work with predominantly? (Desktops/laptops, touch-interface devices, speech interfaces etc.)
   - Main Device – Desktop browsers
   - Further Devices will be mobile / handheld devices
   - Openlaws is a responsive web application and can be used from any device, certain functionalities are not available on mobile devices because of usability concerns
3. Do you plan to integrate the Lynx platform into existing in-house systems and graphical user interfaces (GUIs)?
   - Yes (x)
   - No

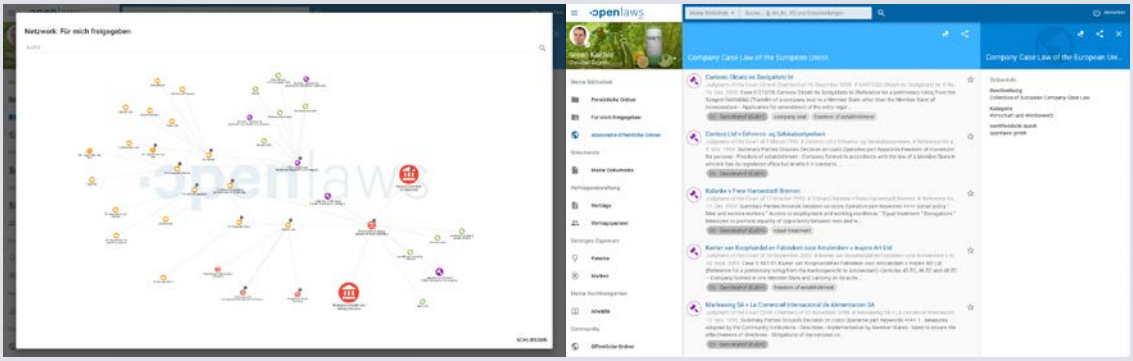
   If the answer to question 3 is "Yes", please also reply to 4 and 5:
4. Please specify the system into which you want to integrate Lynx? Please provide screenshots or screencasts of the system.
   We will integrate it to our existing openlaws platform (https://openlaws.com)
   Openlaws provide already now following high level features:
   - Legal Monitoring: The customer will be actively informed of changes in the legal situation (Austria and EU).



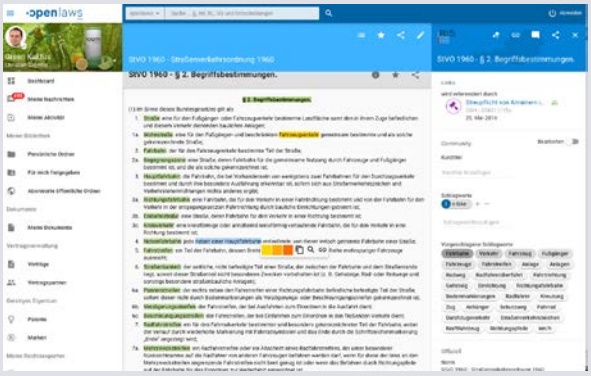
- A combined search for Austria and the EU legislation and case law.

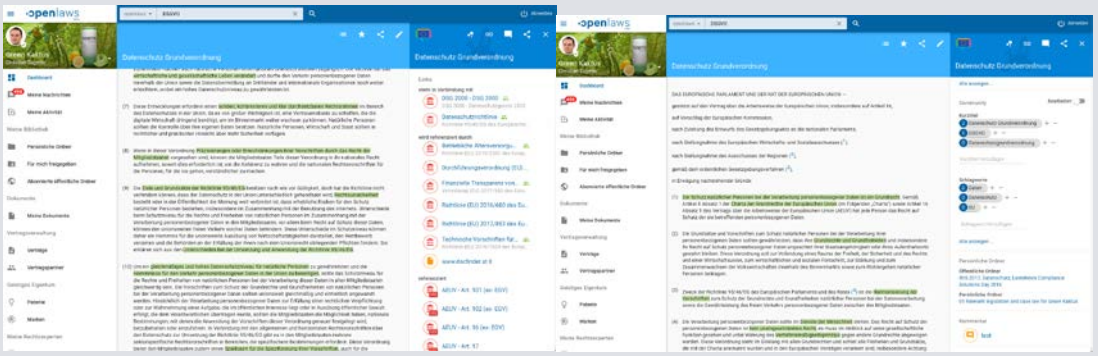


- Collaboration: Keep an overview of standards together.

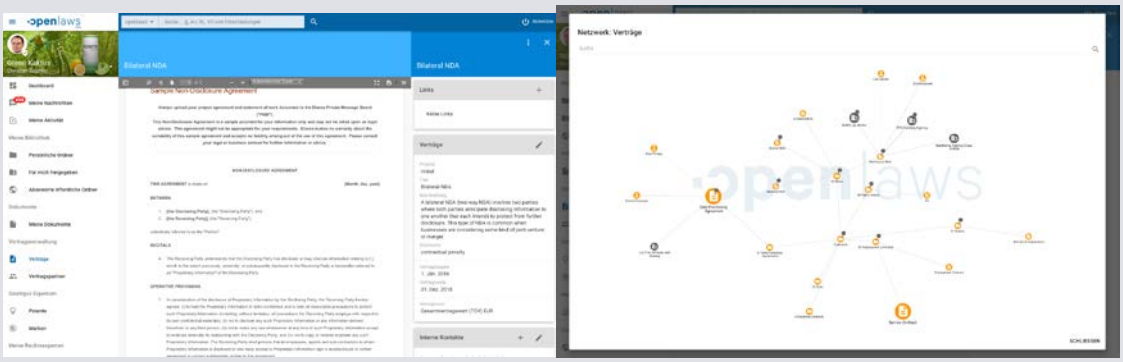


-
- Underline and comment functionalities (can be shared within a group).

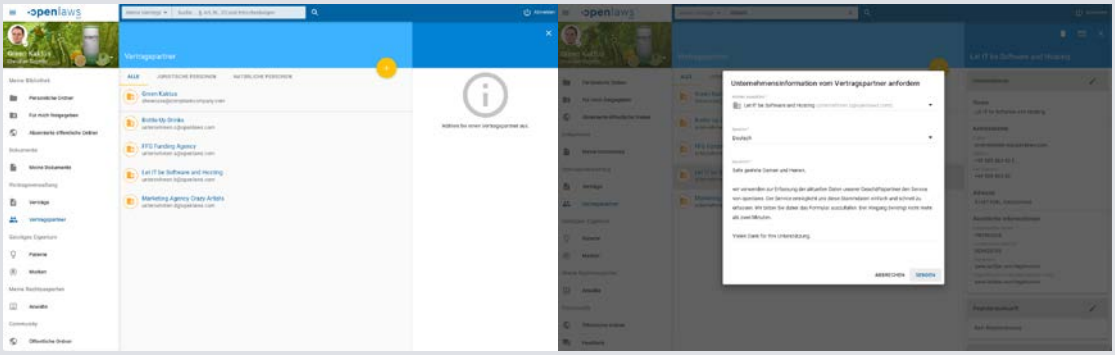


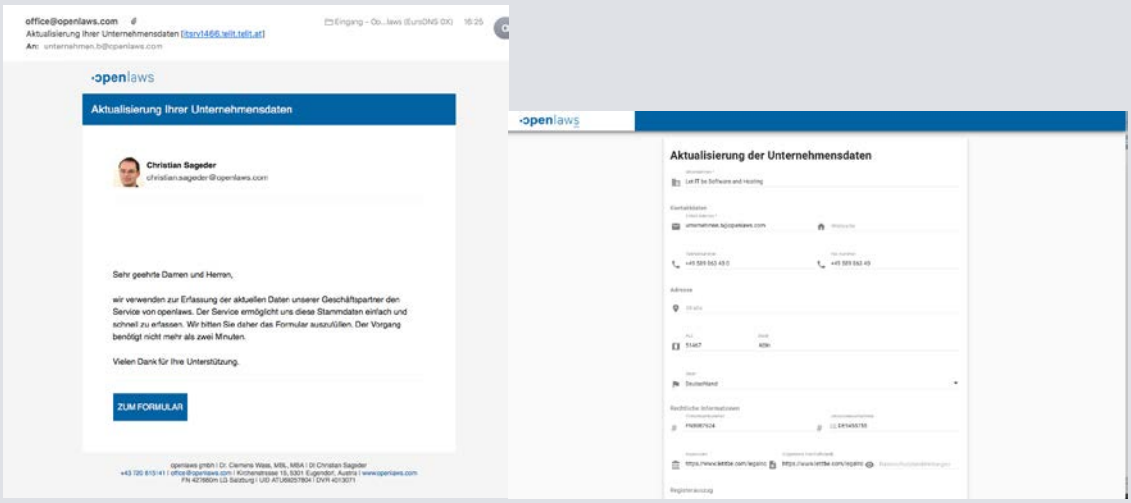


- Contract and Policy Management: Organizing contracts and internal policies, keeping track of deadlines and sharing in the team.

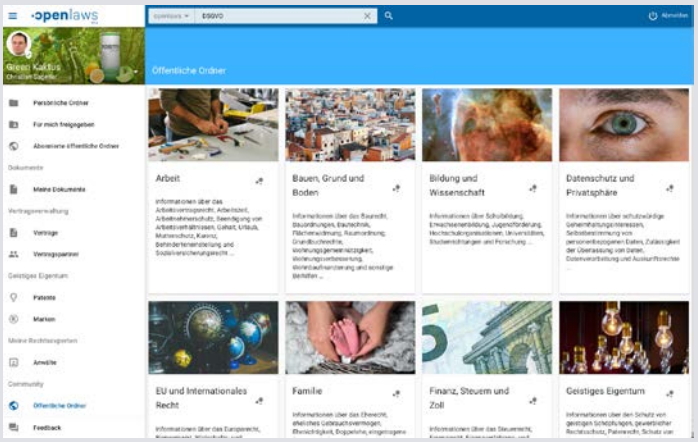


- Management of contractors: keep contract partners in the overview and automatically request company excerpts from the contract partner.





- Public folders



5. Do you currently use a stand-alone application with a GUI or web-based GUI?
   Web-based gui, based on AngularJS which communicate with the backend through a REST API
   If the answer to question 3 is "No", please also reply to 6 and 7:
6. How are you planning to use the services developed in Lynx?
   - REST API calls (x)
   - Web services
   - Web browser
   - Mobile phone/tablet applications
   - Other:
7. Would your preference be to develop a new (web-based) GUI to connect to the Lynx services or would you prefer some other way?
   We plan to integrate the Lynx service via our existing backend and provide our own developed GUI for the features.

3 Actual usage of Semantic Capabilities and Automatic Tools

8. How do you analyse or process legal documents in your company? (For example, with the help of human experts, fully automatically, semi-automatically etc.? Please be as specific and descriptive as possible.)
   - Processing of documents is fully automatically
   - Austrian / EUR-Lex legislation and case law is imported on a daily base from the official web services. During the import process these documents are split into fragments. Based on the meta data these documents are linked to other fragments / documents.
   - Austrian documents are also analyzed (Name Recognition) with the Poolparty (Semantic Web Company)
   - All documents are added to our full text search index. In addition, the search index includes also meta information, e.g. document number, case numbers, ...
   - In addition individuals
     - can proof the name recognitions (Tagging)
     - add additional links between documents
     - comment documents
     - underline documents
9. Do you use automatic solutions and tools for analysing and processing legal documents in your company? If yes, which ones?
   Yes,
   - Own developed tool
   - Poolparty (Semantic Web Company)
10. What kind of documents from the legal domain (or your use case domain) do you work with (official law texts, letters, case law, EU regulations and directives, client specifications etc.)?
    Legislation / Case law,
    - Austrian legislation and case law
    - EurLex regulations, directives, case law
    - German federal legislation, case law (beta)
    - Dutch legislation, case law (alpha)
    - Italian legislation (proof of concept)

    Client documents
    - All kind of documents, which are available as pdf
11. If you already use software for processing legal documents, please provide screenshots or screencasts of your software/GUIs.
12. In terms of use cases and workflows, please specify all (or a representative set of) typical workflows that you use in-house (e.g., types of documents, types of analysis, types of processing, types or approaches of producing new content, etc.).
    Import legislation / case law
    1. Import legislation / case law from official source
       - Text & meta data
    2. Split documents into article / paragraphs
    3. Build version graph

4. Link articles
5. Auto tag articles
6. In case law we search also for references to legislation
7. Add documents / meta data to search engine (Elastic Search)

New Contract

- Adding new contract
- Adding contract partner and person in charge
- Provide additional Meta Data
- Upload contract

4 Users and Profiles

13. What types of users are going to use Lynx services (e.g., JavaScript developers, lawyers, knowledge workers, customers, etc.)?
    - (Java) Developers, as GUI will be provided by openlaws itself
14. Do you need a multi-user solution?
    - o (X) Yes, please specify: it MUST be guaranteed that the data which is processed cannot be access by any other user. There must be at least a separation by each party using the REST interfaces from Lynx
    - o No
15. Do you need authentication (login/password)?
    - o (X) Yes, please specify: it MUST be guaranteed that the data which is process cannot be accessed by any other user. For accessing public information, e.g. legislation an authentication is not necessary.
    - o No
16. Do you need access control lists with different roles and different permissions?
    - o Yes, please specify _____
    - o (X) No, will be handled by the openlaws application itself

5 Data Sets

17. What kind of reference materials or reference data sets do you use on a regular basis?
18. Which online data sets or reference materials would help you in your daily work?

Dataset (e.g. DBPedia)	Domain (e.g. labour law, oil & gas)	Language of the dataset	How do you want to use it? Why this and not another?

19. File Formats: Which are the formats of files that you want to process with Lynx? Do you want the same file format in the request you send to Lynx as well as in the responses you get back from Lynx?

Input file format	Output file format	Comments, challenges, additional information
PDF	PDF + XML	Prio 1
XHTML	XHTML	Prio 1
XML	XML	Challenge: Validation of the document Main advantage (XML/XLST)
JSON	JSON	
MS Office Formats		Prio 2
Open Office Formats		Prio 3

6 Common Services

20. Do you need a tool that can **identify and highlight named entities (persons, locations, organizations, etc.) in legal documents**? For example, this could result in a colour-based highlighting of person, location, organisation names in documents or the filtering of document collections based on the names contained in them.
    - o Yes, please specify in which context
    - o (X) Yes, but
      We only would need identifying the name entities. As we would like to process mainly contracts we would like to identify within this contract organizations, persons, locations, ... to automatically suggest this information to the user

- o No, but _____
- o No

21. Do you need a tool that can **identify and highlight time expressions and normalize them**? Such a function could enable a timeline view of a large document collection, for example, of a series of letters or correspondence.
    - o (X) Yes, please specify in which context:
      We would like to identify the date when contracts are signed, how long they are valid, identify termination periods.
    - o Yes, but _____
    - o No, but _____
    - o No
22. Do you need a tool that can **identify and highlight geographical information related to locations in legal documents**? For example, the output of such a function could be an interactive map containing all documents or content of the documents.
    - o (X) Yes, please specify in which context
      We would like to identify the applicable law (typically a country), venue (typically a city), and the addresses of the involved parties
    - o Yes, but _____
    - o No, but _____
    - o No
23. Do you need a tool that can **identify and highlight events (or other types of important keywords) in legal documents**? For example, the output of such a function could be a list of events (words, phrases, expressions, etc.) that require some kind of action or reaction from the reader.
    - o (X) Yes, please specify in which context
      We want to add this information to create tasks for individuals who then have have to take care and/or to follow up
    - o Yes, but _____
    - o No, but _____
    - o No
24. Do you need a tool that can **identify relations between entities (some judge is related to a criminal because they are involved in a court case) in legal documents**? For example, the output of such a function could result in capabilities for searching documents containing relations through certain entities.
    - o (x) Yes, please specify in which context
      We would like to add relations between a company and managers/employees of that company; ideally different roles for people working in a company are supported (e.g. managing director, data protection officer, etc.)
    - o Yes, but _____
    - o No, but _____
    - o No
25. Do you need a tool that can **identify specific domain terminology (legal terms, oil & gas related terms, etc.) in legal documents**?
    - o (X) Yes, please specify in which context
      We want to identify, data processing clauses and privacy clauses
    - o Yes, but _____
    - o No, but _____
    - o No
26. Do you need a tool that can **recognize citations, references and relations between legal documents**? For example, the output of such a function could be an interactive graph display showing the relations between all the documents of a court case or piece of legislation.
    - o Yes, please specify in which context_____
    - o (X) Yes, but we need only the recognition of such references. Main concerns are, that this can only be to public documents and not to all other documents within a company.
      Visualization is already provided by openlaws
    - o No, but _____
    - o No
27. Do you need a tool that **can disambiguate the sense of a term determining if it is referring to labour law (as an example) or any other domain in legal documents**? For example, the output of such a function could be used for better determining concrete topics the document is talking about.
    - o (X) Yes, please specify in which context_____
    - o Yes, but _____
    - o No, but _____
    - o No
28. Do you need a tool that can **translate legal documents to other languages (if yes, which languages and language pairs?)**?
    - o (X) Yes, please specify in which context
      Translation to English is sufficient,
    - o Yes, but _____
    - o No, but _____
    - o No
29. Do you need a tool that can **summarise documents or sets of documents in the legal domain**?
    - o (X) Yes, please specify in which context

A summary of a contract would be helpful; this would then be the default description in the contract management tool for users.
- o Yes, but _____
- o No, but _____
- o No

30. Do you need a tool that can **search through collections of legal documents**?
    - o Yes, please specify in which context_____
    - o Yes, but _____
    - o No, but _____
    - o (X) No, this functionality is already provided by openlaws.com
31. Do you need a tool that can **recommend other legal documents related to a certain task**?
    - o (X) Yes, please specify in which context_____
    - o (X) Yes, but not necessary the highest priority. It is not clear what is meant with certain tasks. If it is based on question yes.
    - o No, but _____
    - o No
32. Do you need a tool that can **alert you about changes in existing legal documents or the appearance of new legal documents**?
    - o Yes, please specify in which context_____
    - o (X) Yes, but only the appearance of new legal documents, the alert on changes of existing document is already provided by the service of openlaws.
    - o No, but _____
    - o No
33. Do you need a tool that can **determine the main topic of a legal document or part of a document (paragraphs, etc.)**? For example, the output of such a function could help in searching documents for certain legislations, such as Oil&Gas or labour law.
    - o Yes, please specify in which context_____
    - o (X) Yes, but for public legislation, case law we prefer to use the public topic provided by the data provider, classification of private documents is interesting, mainly if paragraphs in contracts can be identified what kind they are. In addition they could be mapped to public legislation and again used for alerts.
      Such functionality would be in particular useful in combination with RSS feeds. A user could subscribe to several legal RSS feeds. These feeds could then be clustered per legal topic. In our case, all privacy related RSS feeds could be identified, possibly going even into more detail (RSS feeds about "data protection impact assessments", about "data protection officers", about "processing registers", and so on. These could then be matched to the relevant article within the GDPR.
    - o No, but _____
    - o No
34. Do you need a tool that can **determine the main type of a legal document (e.g., letter, law, contract, technical report, case report etc.)?** For example, the output of such a function could help further process and visualise a large and heterogeneous set of documents.
    - o Yes, please specify in which context_____
    - o Yes, but _____
    - o No, but _____
    - o (X) No
35. Do you want to combine several automatic processing steps?
    For example: When you get a document, the first thing you do is to translate (if it is in a language other than English), then you read it to learn which people are mentioned (locations and time expressions are also important but first are people). After that you focus on the references of other laws and finally you try to identify arguments and events.
    Yes, but the pipeline can also performed by openlaws. It does not necessarily have to be a feature of Lynx

7 Additional Requirements

36. Please write down any additional requirements you may have that are not covered by the questions above.
    - o As our main business case is the processing of private contracts, a key requirement for our customers is security and confidentiality. Most of our customers would prefer a solution which can be used ON PREMISE and not a cloud service. For this reason, a solution where the common services (NER, annotiation, ...) can be on premise and use data from a public service would be a preferred option.
    - o We do not want that the Lynx platform become a direct competitor to the openlaws platform, existing services have to be considered by the participants.

CuatreCasas

1 About the Survey Participants

As mentioned above, please fill in the survey together with more than one person, and possibly by people holding different roles (e.g., analyst, legal expert, technical expert, product manager, account manager, etc).

NOTE: We do *not* need any personal information (name etc.) but just the number of participants who contributed to filling in the survey and their respective job titles or roles.

- Name of the use case partner: Cuatrecasas
- Number of participants involved in filling in the survey: 1
- Job titles or roles:
  - "Software developer": 1 (Applications Director)

2 Non-Functional Requirements

1. Please describe, as specifically as possible, your use case (or use cases): what kind of functionality or processing capabilities do you want to realise or achieve with the help of the Lynx platform?

In general, I expect LYNX to be a domain specific platform, with powerfull AI and semantics functionalities and legal resources (up-to-date) on European languages and jurisdictions, that could be "trained", customized and finally used for Cuatrecasas needs, currently in the Labour law pilot but also in future ones.
I expect from LYNX, accurate translation services (legal context), provide access to official Laws (Labour Law in this specific use case), translated in multiple languages (again, accurate translation) and provide other related legal content (based on articles of law) like jurisprudence, recent legal changes, ...
I expect also be able to train the system to be able to recognize and key legal concepts and also be able to answer some "simple" legal questions (Q&A functionality)

2. What kind of devices do you work with predominantly? (Desktops/laptops, touch-interface devices, speech interfaces etc.)

Laptops. But we normally work on RWD (responsive) mode to be able to work on several kind of devices.
Desktops/laptops.

3. Do you plan to integrate the Lynx platform into existing in-house systems and graphical user interfaces (GUIs)?

Not easy to answer for me …
We have an internal framework (SOAP integration platform) that covers almost 90% of key information in our systems (SAP ERP, SAP CRM, …) and every (99%) of our system (standard or internally developed) have public/private API to interact with

- Yes
- No [X] Not for this use case

If the answer to question 3 is "Yes", please also reply to 4 and 5:

4. Please specify the system into which you want to integrate Lynx? Please provide screenshots or screencasts of the system.

We have an internal framework (SOAP integration platform) that covers almost 90% of key information in our systems (SAP ERP, SAP CRM, …) and every (99%) of our system (standard or internally developed) have public/private API to interact with

5. Do you currently use a stand-alone application with a GUI or web-based GUI?
   If the answer to question 3 is "No", please also reply to 6 and 7:
6. How are you planning to use the services developed in Lynx?
   - REST API calls [X]
   - Web services [X]
   - Web browser
   - Mobile phone/tablet applications
   - Other:
7. Would your preference be to develop a new (web-based) GUI to connect to the Lynx services or would you prefer some other way?

I will not provide an API (initially) I will provide an end-user application (internal and/or external). As a timeline … I imagine first an internal solution for my lawyers and after that, and depending on accuracy and final value of the outputs, we could do the next step, enriching or simplifying the final solution for my customers.

3 Actual usage of Semantic Capabilities and Automatic Tools

8. How do you analyse or process legal documents in your company? (For example, with the help of human experts, fully automatically, semi-automatically etc.? Please be as specific and descriptive as possible.)

YES. We, as a lawyers, continually read and review legal documents, and we generate legal content (contracts, ...)
We dedicate a lot of human time to that task. We have an specific group of people dedicated to analyse legal information, legal changes, experiences and best practices and also create models /standards and LEGAL OPINION for specific complex legal questions.

9. Do you use automatic solutions and tools for analysing and processing legal documents in your company? If yes, which ones?

Several on specific use cases. But the most generalistic and broad solution we are using is EXPERT SYSTEM COGITO DISCOVER as a specially trained platform for Cuatrecasas documents.
But also, other sector specific solutions like Kira for Document/Contract Review (mostly) and Due Diligence operations.
EXPERT SYSTEM COGITO DISCOVER

- Self-development
  - We do our own developments enhancing the standard base technology and building specific applications that use standard technology (we are thinking on same approach to build Labor law use case in Lynx project)
- Commercialized business solution provided by which company? [X]
  - EXPERT SYSTEM

Difficult to explain … main use case is document enrichment and classification. We have a massive process that sends documents from our internal document management repository (internal, clients and knowledge content) to the Expert System (pre-trained and pre-customized) via Web Services/API and we receive a lot of extra-information (entity extraction, and internal classification) that we use to enrich the internal repository for better auto-classification and improve related analytics (experts/whoknowswhat …)

10. What kind of documents from the legal domain (or your use case domain) do you work with (official law texts, letters, case law, EU regulations and directives, client specifications etc.)?

Answered later in the Data Sets section.

11. If you already use software for processing legal documents, please provide screenshots or screencasts of your software/GUIs.
12. In terms of use cases and workflows, please specify all (or a representative set of) typical workflows that you use in-house (e.g., types of documents, types of analysis, types of processing, types or approaches of producing new content, etc.).

4 Users and Profiles

13. What types of users are going to use Lynx services (e.g., JavaScript developers, lawyers, knowledge workers, customers, etc.)?

Directly, an also regarding our specific current labor law pilot. I think that for that … the only direct user of the LYNX services will be **IT Developers**.
But if I think on the final usage through an specific application integrated with lynx, then:

- **Labour lawyers**
- Labour knowledge workers (labour lawyers as a skills)
- **Final customers** (could be lawyers or non-lawyers [HR])

14. Do you need a multi-user solution?
    - Yes [X]
    - No
15. Do you need authentication (login/password)?
    - Yes [X]
    - No
16. Do you need access control lists with different roles and different permissions?
    - Yes [X]
    - No

My "front-end" solution will require authentication, and I have 2 different users with different economic business case (internal Cuatrecasas lawyers and expernal users from Cuatrecasas currently clients/prospects).
I will not provide an API (initially) I will provide an end-user application (internal and/or external). As a timeline … I imagine first an internal solution for my lawyers and after that, and depending on accuracy and final value of the outputs, we could do the next step, enriching or simplifying the final solution for my customers.

User Type	Description
Internal Cuatrecasas Users	
External Clients	The second model could require some kind of monetize accountability based on legal question and the different language, jurisdictions and maybe even the level of complexity and value on the results (answer from the system).

Availability

If we are thinking in real usage (not piloting, not R&D, …), our internal applications always offer high availability.
When you say 1 hour of un-availability … you mean 1h/day? 1h/week? 1h/moth? …
1h/week ["working hours" non-weekends] could be accepted (but our difficult to explain because this is something only accepted in non-production environments or absolutely additional not relevant secondary information systems).
More than 1h/week is un-acceptable
If we think on an external service to our customers … then we should reach > 99% availability. (well at the end is a question of price, but … by default high standard)

5 Data Sets

17. What kind of reference materials or reference data sets do you use on a regular basis?

This use case will not use Cuatrecasas internal or any other client/case documentation, the corpus and knowledge base is only based on regulatory legal information from several countries/jurisdiction

- **Labour Law (official) on each country, by article**
  - Spain: Estatuto de los Trabajadores
    - BOE (public, …),
    - …
  - Italy: Statuto dei Lavoratori
    - http://www.altalex.com/documents/codici-altalex/2014/10/30/statuto-dei-lavoratori ...
  - Germany
  - Austria
  - Europe (CEE)
- **Related jurisprudence, legal sentences, … on each country**
  - Spain:

- CENDOJ y la web del Poder Judicial http://www.poderjudicial.es/search/indexAN.jsp
  - Italy
  - Germany
  - Austria
  - Europe (CEE)

18. Which online data sets or reference materials would help you in your daily work?

Dataset (e.g. DBPedia)	Domain (e.g. labour law, oil & gas)	Language of the dataset	How do you want to use it? Why this and not another?
Estatuto de los Trabajadores	Labour law	Spanish	
Statuto dei Lavoratori	Labour law	Italian	
	Labour law	German	
	Labour law	European Languages	
CENDOJ y la web del Poder Judicial	Related Jurisprudence	Spanish	

19. File Formats: Which are the formats of files that you want to process with Lynx? Do you want the same file format in the request you send to Lynx as well as in the responses you get back from Lynx?
From Lynx as a platform, the different sources (labour country law, and jurisprudence) they can have different formats. Normally local law will be founded on xml or pdf. And jurisprudence can be published in several format (metadata/xml, csv, … and also in pdf format)

Input file format	Output file format	Comments, challenges, additional information
PDF		
XML		
CSV		

6 Common Services

20. Do you need a tool that can **identify and highlight named entities (persons, locations, organizations, etc.) in legal documents**? For example, this could result in a colour-based highlighting of person, location, organisation names in documents or the filtering of document collections based on the names contained in them.
    - Yes, please specify in which context_____
    - [X] Yes, but
      - We have 3 main types of legal documents in our concrete labour law use case:
        - Labour law articles (by country/jurisdiction)
          - Normally is no relevant entities in there (as a maximum organization [courts/source that creates and publish them, and the dates of vality]
          - Here we have to be able to return {jurisdiction, law, article} related to some related concepts in it
        - Concrete question (based on legal labour law)
          - Here could be named / specified some locations. Maybe something
        - Jurisprudence / Case law
          - Here we have to be able to return {jurisdiction, law, article} related to some related concepts in it
          - And here could be interesting to identify some named-entities like court, judge, parties, organizations (companies), lawyers, legal specialty, sector/industry, …
    - No, but _____
    - No
21. Do you need a tool that can **identify and highlight time expressions and normalize them**? Such a function could enable a timeline view of a large document collection, for example, of a series of letters or correspondence.
    - Yes [X]
    - Yes, but _____
    - No, but _____
    - No

22. Do you need a tool that can **identify and highlight geographical information related to locations in legal documents**? For example, the output of such a function could be an interactive map containing all documents or content of the documents.
    - o Yes, please specify in which context_____
    - o [X] Yes, but … to think about, but not mandatory/not relevant
    - o [X] No, but _____
    - o No
23. Do you need a tool that can **identify and highlight events (or other types of important keywords) in legal documents**? For example, the output of such a function could be a list of events (words, phrases, expressions, etc.) that require some kind of action or reaction from the reader.
    - o Yes, please specify in which context_____
    - o Yes, but _____
    - o No, but _____
    - o No

    What do you mean? What is an event for you??
24. Do you need a tool that can **identify relations between entities (some judge is related to a criminal because they are involved in a court case) in legal documents**? For example, the output of such a function could result in capabilities for searching documents containing relations through certain entities.
    - o Yes [X], in general YES. In our specific use case, key relations will be between articles of different country-laws, between sentences/jurisprudence …
    - o Yes, but _____
    - o No, but _____
    - o No
25. Do you need a tool that can **identify specific domain terminology (legal terms, oil & gas related terms, etc.) in legal documents**?
    - o Yes [X]
    - o Yes, but _____
    - o No, but _____
    - o No
26. Do you need a tool that can **recognize citations, references and relations between legal documents**? For example, the output of such a function could be an interactive graph display showing the relations between all the documents of a court case or piece of legislation.
    - o Yes, please specify in which context_____
    - o [X] Yes, but … I don't have a clear idea about that, but, for sure in court sentences will appear citations to articles of law, and other sentences, that I want to use and suggest/co-relate
    - o No, but _____
    - o No
27. Do you need a tool that **can disambiguate the sense of a term determining if it is referring to labour law (as an example) or any other domain in legal documents**? For example, the output of such a function could be used for better determining concrete topics the document is talking about.
    - o Yes, please specify in which context_____
    - o [X] Yes, but … that's very interesting as a general solution, but in my specific use case all my information domain is about labour law and then is not necessary to disambiguate
    - o No, but _____
    - o No
28. Do you need a tool that can **translate legal documents to other languages (if yes, which languages and language pairs?)**?
    - o Yes [X], mandatory and KEY functionality in my use case. Absolutely key to be able to provide high reliable translation of local law in several languages
    - o Yes, but _____
    - o No, but _____
    - o No
29. Do you need a tool that can **summarise documents or sets of documents in the legal domain**?
    - o Yes, please specify in which context_____
    - o [X] Yes, this an important feature/functionality for Cuatrecasas (we have an internal group of people dedicating lots of hour to summarize knowledge documents and legal changes). I don't know if is really mandatory for our specific use case with Labour law
    - o No, but _____
    - o No
30. Do you need a tool that can **search through collections of legal documents**?
    - o Yes, please specify in which context_____
    - o Yes, but _____
    - o [X] No, but … I don't know if this is relevant in my Labour law use case, my feeling is not.
    - o No
31. Do you need a tool that can **recommend other legal documents related to a certain task**?
    - o [X] Yes, interesting in general, and specifically recommending jurisprudence or even similar answer to similar questions
    - o Yes, but _____
    - o No, but _____

- o No

32. Do you need a tool that can **alert you about changes in existing legal documents or the appearance of new legal documents**?
    - o Yes, please specify in which context_____
    - o [X] Yes, very interesting in general, but not for this particular use case on labor law Q&A
    - o No, but _____
    - o No
33. Do you need a tool that can **determine the main topic of a legal document or part of a document (paragraphs, etc.)**? For example, the output of such a function could help in searching documents for certain legislations, such as Oil&Gas or labour law.
    - o [X] Yes, KEY functionality in my use case. If we have all the labor country-law very well clasified by topic by paraghaps (and into articles), then … our pilot should be easy to develop and a real success, if not … so complicated
    - o Yes, but _____
    - o No, but _____
    - o No
34. Do you need a tool that can **determine the main type of a legal document (e.g., letter, law, contract, technical report, case report etc.)?** For example, the output of such a function could help further process and visualise a large and heterogeneous set of documents.
    - o Yes, please specify in which context_____
    - o Yes, but _____
    - o No, but _____
    - o No
35. Do you want to combine several automatic processing steps?
    For example: When you get a document, the first thing you do is to translate (if it is in a language other than English), then you read it to learn which people are mentioned (locations and time expressions are also important but first are people). After that you focus on the references of other laws and finally you try to identify arguments and events.

- I don't know, I imagine that yes but … no clear idea about that

7 Additional Requirements

36. Please write down any additional requirements you may have that are not covered by the questions above.

INTRODUCTION / General description of the Cuatrecasas Labour Law use case:

- The aim of this use case is be able answer several legal questions about the labour law, in several countries-jurisdictions.
- Cuatrecasas as a law firm, with international coverage. Frequently needs to analyze several aspects of its customer international operations: Normally M&A or simply natural growing (new buildings, new offices, …).
- We receive some labour law questionaries to answer, fulfilling them with the different countries-jurisdictions considerations.
- The system will provide complementary information related to the different legal questions (articles of the law): legal precedents, case law and other jurisprudence.

My use case could need some kind of Machine Learning or cognitive services, and Q&A functionalities to be able to arrive to a more specified human kind of answer to an specific question, that enrich to an "easy" document parts of related-to-topic text extraction

DNV GL

1 About the Survey Participants

As mentioned above, please fill in the survey together with more than one person, and possibly by people holding different roles (e.g., analyst, legal expert, technical expert, product manager, account manager, etc).

NOTE: We do *not* need any personal information (name etc.) but just the number of participants who contributed to filling in the survey and their respective job titles or roles.

- Name of the use case partner: {DNV GL}
- Number of participants involved in filling in the survey: <2>
- Job titles or roles:
  - Legal expert or lawyer: <0>
  - Software developer: <0>
  - Analyst: <0>
  - Business developer: <1>
  - Consultant: <1>
  - Key account managers: <0>
  - …

2 Non-Functional Requirements

1. Please describe, as specifically as possible, your use case (or use cases): what kind of functionality or processing capabilities do you want to realise or achieve with the help of the Lynx platform?
   Use case 1: Renewable Energy Projects
   DNV GL is supporting early phase renewable energy projects to assess feasibility and viability of the initiative (e.g. technical, financial due diligence). This could be a wind turbine project, geothermal or tidal etc. Each phase includes several regulatory challenges laid out in Directives, Laws, regulation, country specific regulation (planned or implemented) and down to standards and best practices. Getting a complete picture of these requirements for a specific country can be a tedious job for which Lynx could offer useful functionality
   Use case 2: CE Marking
   Many products must be CE marked for being commercialized in the EU. The CE marking is normally performed by manufacturers themselves if they have a legal representative in the EU. Otherwise, it must be carried out by the importer. However, it can be complicated for importers to understand which are the requirements for a given product, e.g.: does it need a CE marking at all? Which are the applicable EU directives (Machine Directive, Pressure Equipment Directive, Atex directive, etc.)? How are the directive received in a specific country? Which are the applicable standard related to the above directives? Does the product need a third party certificate?
   - The above questions are particularly problematic if the importer is not himself an expert of the product. The "lynx portal" would have certainly be a great tool for us.
   - The above problem can also be an issue for manufacturers who wish to sell their products in the EU (e.g. from China and the USA): which standards should they use during the design? Is there a correspondence among standards from different bodies (e.g. EN standards vs ISO standards vs ASME standards)? And so on. Also in this case, the "lynx portal" would support us as consultant
2. What kind of devices do you work with predominantly? (Desktops/laptops, touch-interface devices, speech interfaces etc.)
   Laptops
3. Do you plan to integrate the Lynx platform into existing in-house systems and graphical user interfaces (GUIs)?
   - Yes
   - ~~No~~

   If the answer to question 3 is "Yes", please also reply to 4 and 5:
4. Please specify the system into which you want to integrate Lynx? Please provide screenshots or screencasts of the system.
   Microsoft Azure Cloud Ecosystem
5. Do you currently use a stand-alone application with a GUI or web-based GUI?
   No

   If the answer to question 3 is "No", please also reply to 6 and 7:
6. How are you planning to use the services developed in Lynx?
   - REST API calls
   - Web services
   - Web browser
   - Mobile phone/tablet applications
   - Other:
7. Would your preference be to develop a new (web-based) GUI to connect to the Lynx services or would you prefer some other way?
   Yes, preferable within our ecosystem: the marketplace on www.veracity.com. However, there may be budgetary constraints to have LYNX adopt a DNV GL look-and-feel, so having some sort of transition period in which we could use LYNX 'as is' would be likely.

3 Actual usage of Semantic Capabilities and Automatic Tools

8. How do you analyse or process legal documents in your company? (For example, with the help of human experts, fully automatically, semi-automatically etc.? Please be as specific and descriptive as possible.)
   human experts based on prior experience, web searches.

9. Do you use automatic solutions and tools for analysing and processing legal documents in your company? If yes, which ones?
   No, not that we are aware of. There are some LYNX-alternatives (at least that is what our internal stakeholders voiced, but this is may be skewed due to little knowledge of what LYNX can actually deliver).

10. What kind of documents from the legal domain (or your use case domain) do you work with (official law texts, letters, case law, EU regulations and directives, client specifications etc.)?
    EU Directives, Legislation, Technical standards, Industry Standards, Recommended Practices

11. If you already use software for processing legal documents, please provide screenshots or screencasts of your software/GUIs.
    Not available at the moment

12. In terms of use cases and workflows, please specify all (or a representative set of) typical workflows that you use in-house (e.g., types of documents, types of analysis, types of processing, types or approaches of producing new content, etc.).
    Deskresearch
    Document analyses, comparison
    Human Judgement
    Peer review (colleagues or in norm committees)

4 Users and Profiles

13. What types of users are going to use Lynx services (e.g., JavaScript developers, lawyers, knowledge workers, customers, etc.)?
    Knowledge workers and probably customers

14. Do you need a multi-user solution? (Not sure if we know what that is. But I can imagine we distinguish internal user segments based on the business area and likely the same for the corresponding customers.

    o Yes, please specify _____
    o No

15. Do you need authentication (login/password)?

    ~~o~~ Yes, please specify in case of a customer portal, use authentication should be done via the marketplace on veracity.com. Internal authentication can be done via single sign on (Active Directory)
    ~~o No~~

16. Do you need access control lists with different roles and different permissions?

    o Yes, please specify TBC_____
    ~~o No~~

User Type	Description

5 Data Sets

17. What kind of reference materials or reference data sets do you use on a regular basis?
    Online portals like:
    https://ec.europa.eu/energy/en/topics/renewable-energy/renewable-energy-directive

- http://www.irena.org/geothermal

- http://www.nlog.nl/wetgeving

  http://www.iea.org/policiesandmeasures/

18. Which online data sets or reference materials would help you in your daily work?

Dataset (e.g. DBPedia)	Domain (e.g. labour law, oil & gas)	Language of the dataset	How do you want to use it? Why this and not another?

19. File Formats: Which are the formats of files that you want to process with Lynx? Do you want the same file format in the request you send to Lynx as well as in the responses you get back from Lynx?

Input file format	Output file format	Comments, challenges, additional information
HTML	HTML	
PDF	WORD	
XML	XML	

6 Common Services

20. Do you need a tool that can **identify and highlight named entities (persons, locations, organizations, etc.) in legal documents**? For example, this could result in a colour-based highlighting of person, location, organisation names in documents or the filtering of document collections based on the names contained in them.
    - o Yes, please specify in which context specific renewable energy sources e.g. heat network (district heating), energy storage, etc needs to be identified from the texts. Also, it could be useful if LYNX could extract from client provided documentation which component parts/technologies/hazards are in scope, and then identify relevant regulations ('matching')
    - o _____
    - o Yes, but _____
    - o No, but _____
    - o No

21. Do you need a tool that can **identify and highlight time expressions and normalize them**? Such a function could enable a timeline view of a large document collection, for example, of a series of letters or correspondence.

    - o Yes, please specify in which context : may be it is useful to have a facility that can map out on a timeline the validities of regulations and which document supersedes what document_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

22. Do you need a tool that can **identify and highlight geographical information related to locations in legal documents**? For example, the output of such a function could be an interactive map containing all documents or content of the documents.
    - o Yes, please specify in which context: this could be useful to see what geographical scope is covered by a regulation, or vice versa, given a scope, which regulations are relevant_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

23. Do you need a tool that can **identify and highlight events (or other types of important keywords) in legal documents**? For example, the output of such a function could be a list of events (words, phrases, expressions, etc.) that require some kind of action or reaction from the reader.

- o Yes, please specify in which context: this may be useful if the tool could highlight trigger words or phrases that indicate what are requirements ('musts'), what are recommendations ('suggestions'), what are practices ('how to's) in guidance documents, directives, policies, laws_____
- ~~o Yes, but _____~~
- ~~o No, but _____~~
- ~~o No~~

24. Do you need a tool that can **identify relations between entities (some judge is related to a criminal because they are involved in a court case) in legal documents**? For example, the output of such a function could result in capabilities for searching documents containing relations through certain entities.
    - o Yes, please specify in which context: this could be interesting e.g. to see what similar products have got which CE markings from which manufacturers?_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

25. Do you need a tool that can **identify specific domain terminology (legal terms, oil & gas related terms, etc.) in legal documents**?
    - o Yes, please specify in which context: yes, this would be useful to highllight various technology components and relate them to the requirements posed upon them across various documents_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

26. Do you need a tool that can **recognize citations, references and relations between legal documents**? For example, the output of such a function could be an interactive graph display showing the relations between all the documents of a court case or piece of legislation.
    - o Yes, please specify in which context: possibly: we have done some experimentation ourselves with creating graphs that depicts how various rules and instructions relate to each other in a company management system. Need to explore further in use case workshop(s)._____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

27. Do you need a tool that **can disambiguate the sense of a term determining if it is referring to labour law (as an example) or any other domain in legal documents**? For example, the output of such a function could be used for better determining concrete topics the document is talking about.
    - o Yes, please specify in which context: may be it is not disambiguation, but more categorisation/classification: if the tool knows that a certain piece of technology belongs to a certain class, then it could highlight: 'hey I know this is a piece of technology and I suspect it is an instance of this broader class of technologies, and for that class, I actually know that these are hazards associated with them, and for those hazards I know that you have to put in these barriers (as specified in guidance documents xyz'. To be explored in use case workshop(s)_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

28. Do you need a tool that can **translate legal documents to other languages (if yes, which languages and language pairs?)**?
    - o Yes, please specify in which context: this is not known to use, to be explored in use case workshop(s)_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

29. Do you need a tool that can **summarise documents or sets of documents in the legal domain**?
    - o Yes, please specify in which context: not really sure about this requirement, I don't think summarisation is necessary_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

30. Do you need a tool that can **search through collections of legal documents**?
    - o Yes, please specify in which context: we guess so, as the documentation sources are scattered _____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

31. Do you need a tool that can **recommend other legal documents related to a certain task**?

- o Yes, please specify in which context: absolutely, I think it is the whole point of our use cases that the system suggests where to look_____
- ~~o Yes, but _____~~
- ~~o No, but _____~~
- ~~o No~~

32. Do you need a tool that can **alert you about changes in existing legal documents or the appearance of new legal documents**?
    - o Yes, please specify in which context__definitely, keeping track of what rules hold and when they change/have changed is essential in compliance advisory service _____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

33. Do you need a tool that can **determine the main topic of a legal document or part of a document (paragraphs, etc.)**? For example, the output of such a function could help in searching documents for certain legislations, such as Oil&Gas or labour law.
    - o Yes, please specify in which context__we think it would be useful if the tool could identify documents (or parts of those) that are in scope of the compliance question at hand. This would probably mean that it acts as a suggestion engine, and say: look I believe this document is relevant for you, as, according to my knowledge graph, I think it is about the piece of technology that you are exploring. _____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

34. Do you need a tool that can **determine the main type of a legal document (e.g., letter, law, contract, technical report, case report etc.)?** For example, the output of such a function could help further process and visualise a large and heterogeneous set of documents.
    - o Yes, please specify in which context: not sure, it is probably important to have some understanding of the role of a directive, a standard, a guidance note, a good practice, etc. So, indeed, some typology of content and their legal ' strength' would be required_____
    - ~~o Yes, but _____~~
    - ~~o No, but _____~~
    - ~~o No~~

35. Do you want to combine several automatic processing steps?
    For example: When you get a document, the first thing you do is to translate (if it is in a language other than English), then you read it to learn which people are mentioned (locations and time expressions are also important but first are people). After that you focus on the references of other laws and finally you try to identify arguments and events.

    Not clear at this stage

7 Additional Requirements

36. Please write down any additional requirements you may have that are not covered by the questions above.

ANNEX 2 – FACE-2-FACE WORKSHOPS

During this stage of the project we have developed four requirement gathering workshops, one with each use case partner of the project (two workshops in the case of DNV GL). This section describes the basic information regarding the workshops.

Pilot 1: Openlaws – Contracts Analysis (27 March 2018)

The workshop took place the 27th March 2018 in the office of DFKI in Berlin, Germany. The participants in the workshops were:

- Christian Sageder (OpenLaws)
- Clemens Wass (OpenLaws)
- Julián Moreno Schneider (DFKI)
- Georg Rehm (DFKI)
- Stefanie Hegele (DFKI)

Pilot 2: CuatreCasas – Labour Law (11 April 2018)

The workshop took place the 11th April 2018 in the office of CuatreCasas in Barcelona, Spain. The participants in the workshops were:

- Pascual Boil (CuatreCasas)
- Iria Estévez López (CuatreCasas)
- Elsa Gómez Díaz (CuatreCasas)
- Jennifer Bel (CuatreCasas)
- Lara Vivas (CuatreCasas) [Partially]
- Julián Moreno Schneider (DFKI)

Pilot 3a: DNV GL – CE Marking (13 April 2018)

The workshop took place the 13th April 2018 in the office of DNV GL in Rotterdam, The Netherlands. The participants in the workshops were:

- Rob van der Spek (DNV GL)
- Hans Groothuis (DNV GL)
- Mike Norman (DNV GL)
- Ramkumar Palanivelu (DNV GL)
- Víctor Rodríguez Doncel (UPM)
- Julián Moreno Schneider (DFKI)

Pilot 3b: DNV GL – Energy (20 April 2018)

The workshop took place the 20th April 2018 in the office of DNV GL in Arnhem, The Netherlands. The participants in the workshops were:

- Eelco Kruizinga (DNV GL)
- Bart in ‘t Groen (DNV GL)
- Koen Broess (DNV GL)
- Maroeska Boots (DNV GL)
- Julián Moreno Schneider (DFKI)

REFERENCES

[Cockburn2001] Alistair Cockburn. Writing Effective Use Cases. Addison-Wesley, 2001.

[Cohn2004] Mike Cohn. User Stories Applied: For Agile Software Development. Addison-Wesley, 2004.

[Fricker2015] Fricker, S. A., Grau, R., and Zwingli, A., (2015). Re- quirements Engineering: Best Practice, pages 25–46. Springer International Publishing, Cham.

[Hammond2002] Judy Hammond, Tom Gross, and Janet Wesson, editors. Usability: Gaining a Competitive Edge, volume 226 of IFIP Conference Proceedings. Kluwer, 2002.

[IEEE1998] IEEE Computer Society. Software Engineering Standards Committee and IEEE-SA Standards Board. IEEE recommended practice for software requirements specifications. IEEE Std. 830-1998, 1998.

[McEwen2013] Scott McEwen. Requirements: An introduction. http://www.ibm.com/developerworks/rational/library/4166.html, April 2004. Accessed: April 30, 2018.

[Nielsen1993] Jacob Nielsen. Usability Engineering. Academic Press, 1993.

[Wake2013] Bill Wake. INVEST in good stories, and SMART tasks. http://xp123.com/articles/invest-in-good-stories-and-smart-tasks/, August 2003. Accessed Jan 15, 2013.

[Zielczynski2007] Peter Zielczynski. 2007. Requirements Management Using Ibm® Rational® Requisitepro® (First ed.). IBM Press.