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

D5.2 Intermediate demonstrator for pilot 2

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	Pieter Verhoeven (DNV GL)
CONTRIBUTORS	Martin Kaltenboeck (SWC); Artem Revenko (SWC); Martin Schauer (SCW); Eelco Kruizinga (DNV GL)
REVIEWERS	Christian Sageder (OLS); Pascual Boil Ballesteros (Cuatrecasas)
VERSION | STATUS	v1.0 | Final
NATURE	Demonstrator
DISSEMINATION LEVEL	Public
DOCUMENT DOI	10.5281/zenodo.3558450
DATE	30/11/2019 (M24)

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

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
01	First version	11/11/2019	Pieter Verhoeven (DNV GL);
02	Second version, after first review by Eelco Kruizinga (DNV GL)	18/11/2019	Pieter Verhoeven (DNV GL);
03	Second review Pascual Boil (CUATRECASAS) Stefanie Hegele (DFKI) Elena Montiel-Ponsoda (UPM) Christian Sageder (openlaws)	18/11/2019	Pieter Verhoeven (DNV GL);
04	Third review Pascual Boil (CUATRECASAS)	25/11/2019	Pieter Verhoeven (DNV GL);
05	Forth review	28/11/2019	Martin Kaltenboeck (SWC)
10	Final version	29/11/2019	Pieter Verhoeven (DNV GL);

DISCLAIMER

LIST OF TERMS AND ACRONYMS

Doublet: A combination of a production and an injection well

GTE: Geothermal Energy

LKG: Legal Knowledge Graph

SME: Small & Medium Enterprises / Subject Matter Experts

UI: User Interface

EXECUTIVE SUMMARY

This deliverable constitutes the intermediate results on the Lynx use case development for geothermal energy (GTE). Section 1 provides a brief introduction of the geothermal energy context which is used as a proxy for trends and future needs of multiple emerging alternative sustainable energy resources. The grand view that underpins this use case, is the applicability of emerging semantic technologies (knowledge graphs) for sustainable energy alternatives as they as subject to accelerated or even a transformative future growth for which friction in access to relevant regulatory data can cause delays (or require breakthroughs).

Section 2 describes the use case and the progress made in engaging with the industry to learn if problems / issues are recognized and how they can be subject to this use case. It further elaborates on plausible user journeys and mock-ups of screen design how Lynx services can interact with the end-user. Section 3 describes the pilot architecture and system current in place. This section mainly talks about the front-end stack and the envisioned backend and integration with the Lynx services. Section 4 described the outlook of use case specific activities planned for the rest of the Lynx project. The report is supplemented by appendix A) Envisioned business model canvas.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION TO GEOTHERMAL ENERGY

We think emerging digital technologies can contribute to the energy transition and the ambitions set in the Paris Agreement

Digital technologies, such as knowledge graphs, can contribute to the broader challenges of mitigating climate change. Emerging alternative energy resources, including biomass, solar, wind, hydrogen and geothermal energy, will probably face a period of experimental growth and learning, in order to develop industry standards and balanced regulation & policies comparable to mature industries. As a proxy to this challenge, we will focus on the domain of Geothermal Energy for which regulatory information is fragmented strongly these days and for which most EU member countries will have ambitions to use GTE. Figure 1 below pictures a map of the EU with the GTE energy potential for heating, cooling and /or electricity.

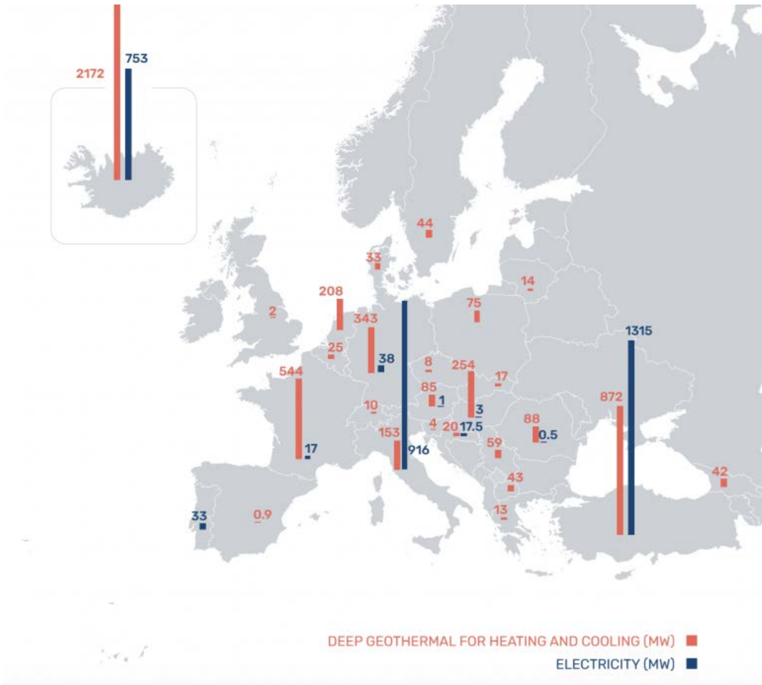

Figure 1 Overview on geothermal capacity - power & heating, EGEC Report 2018 (screenshot)

What is geothermal energy?

Geothermal energy[1] is heat generated in the sub-surface of the earth. A geothermal fluid and/or steam carry the geothermal energy to the earth's surface. Geothermal energy operators drill a production and an injection well (also known as a doublet) to a certain depth (between 100m - 4000m) to circulate fluid in order to produce 'heat'. Depending on the temperatures, this fluid can be used to produce clean electricity, or as baseload for municipal district or industry heating or cooling.

Challenges for the Geothermal Energy domain

Geothermal Energy (GTE) is a sustainable energy alternative with a significant potential for EU countries to achieve the goals set out in the Paris Agreement to reduce greenhouse gas emissions by at least 40 % by 2030 compared to 1990. Nevertheless, in order to fully reap the GTE potential against acceptable economics, breakthroughs are needed to bring cost, technical and market risks down. Emerging technologies in the domain of digitalisation can help the GTE industry accelerate

[1] Geothermal energy on Wikipedia: https://en.wikipedia.org/wiki/Geothermal_energy

its growth and become a professional & matured industry. For example, standardisation of technology & methodology captured in international recognised best practices and fit for purpose regulatory frameworks on a European and national scale. These challenges are recognised in recent industry reports like the *Strategic Research and Innovation Agenda, European Technology & Innovation Platform on Deep Geothermal, [link]*

2 USE CASE DESCRIPTION

2.1 INTRODUCTION

The GTE use case builds upon the following problem assumptions:

1) National stakeholders (see 2.2) facing regulatory risks, missing potential opportunities, are taking poor decisions due to compliance information being fragmented over multiple information sources. The first assumption is: "can value be generated by connecting regulatory information resources for geothermal energy?"

2) International stakeholders (see 2.2) struggle with a lack of understanding of country specific regulatory frameworks (which is a competitive disadvantage) which limits international competition in turn limiting the potential benefits of economies of scale / standardisation. The second assumption is if we stimulate internationalisation by providing 'level playing field' access to relevant compliance information for different EU countries?

- Complex and incomplete regulations, fragmented among EU members states, and long and complex authorization processes (exploration / exploitation permits) have slowed down geothermal energy, while the energy transition outlook reports project the need for an accelerated growth towards 2030 in order to meet the Paris Agreement.
- In this context it is assumed that scaling and internationalisation will contribute to cost reduction and lower technical risk due to standardisation on a European level.

3) Regulators in EU member countries can contribute by further alignment of regulation and policies for which goals have been set. Today two types of regulatory regimes coexist in Europe: countries that have specifically addressed geothermal energy clauses in the mining act and countries lacking any specific regulation for GTE.

The intermediate demonstrator for pilot 2 is mainly focussing on the first problem assumption: GTE regulatory information being fragmented.

2.2 OBJECTIVE OF THE PILOT

Related to the previously mentioned problem assumption, the following factors are relevant to get a sense of the issues / challenges at stake:

- **Fragmented information** needed to assure regulatory compliance imposes a risk during decision making in early phase initiatives. High risk profile lead to higher interests (mark-ups) and fewer bankable propositions. But also, regulatory uncertainties like time to permit approval, imposes significant risks to early phase initiatives.
- **Growth:** GTE is expected to play a pivotal role in the energy transition and the EU's ambition to reduce 49% CO_2 emission by 2030 and therefore needs to grow (e.g. today in the Netherlands, 17 doublets are operated which is aimed to grow to 175 in 2030 and 700 in 2050)[2]
- **Maturing discipline:** Stakeholders acknowledge the GTE market is still in its infancy, and industry standards (technical, safety & environment) and regulation (different in each EU country) need to mature rapidly (resulting in short term uncertainties) in order to create a level playing field for the industry (and fair competition)

[2] Master Plan Geothermal Energy in the Netherlands (May 2018) https://geothermie.nl/images/bestanden/Masterplan_Aardwarmte_in_Nederland_ENG.pdf

- **Technical & Commercial risks** need to be understood better and mitigated in order to improve bankability of GTE initiatives.
- Mechanisms like **economies of scale, standardisation and knowledge sharing** between GTE hotspots (EU) and governmental participation / risk taking, might help accelerating this growth.
- The EU is encouraging SMEs to develop **cross border business** as international business generate healthier margins and lead to more sustainable markets with fair competition and more standardization.

User segmentation

Geothermal energy projects involve a wide range of stakeholders throughout the life cycle[3]. On average it takes 7 years of exploration, design and construction before operations can be commissioned, and energy is produced. During this period multiple stakeholders are involved in developing a bankable GTE initiative within boundaries of (inter)national regulations and policies and benefit from direct access information. The Lynx demonstrator for GTE is built with the following potential stakeholders in mind:

- Regional Energy Coordinator
- Subject matter experts
- GTE project developers
- Drilling companies
- Operators
- End-users
- GTE Service providers
- Engineering companies
- Regulators
- Industry associations
- Governments
- Banks

User roles or functions:

Within the hereinabove mentioned user segments, the following functions / roles are envisioned to benefit from the Lynx demonstrator for GTE.

- Project managers
- Technical Authorities
- Consultants
- Legal professionals
- Regulatory experts
- Health Safety & Environment (HSE) managers
- Liaisons with governmental organisations

[3] Research from the International Renewable Energy Agency (IRENA) identified 16 stakeholder categories being involved throughout the lifecycle of GTE project and facing regulatory risks [IRENA Project Navigator – Technical Concept Guidelines for Geothermal Power Projects]

Understanding pains & gains in *job to be done*

Before deep diving in the realm of knowledge graphs and semantics, understanding the problems of our envisioned end users came first. During the engagements with GTE practitioners we talked about what's keeping them busy and how easy / hard is to find and use regulatory information. The approach is inspired on value propositions design[4] wherein customer insights are described in:

- Job to be done: what is the customer trying to achieve? This could be tasks they're trying to complete, problems they're trying to solve, or wants or needs they're trying to satisfy.
- Pains: negative emotions, undesired costs or situations, and risks the customer (could) experience before, during, or after getting the job done.
- Gains: benefits the customer expects, desires or would be surprised by. This includes functional utility, social gains, positive emotions, cost savings, etc.

While building the demonstrator, the customer characteristics were used for steering the process of building functionality to relieve pain and result in a positive, desirable user experience. A selection of typical 'job to be done' questions are listed below. The pains and gains correspond with the problem assumptions introduced in 2.1: regulatory information being fragmented over various sources. If we can demonstrate the Lynx platform is able to address these pains & gains, it would prove the added value and the likelihood of a viable business case.

Key question: who is suffering most (risks / missing out opportunities) by lack of overview / insights?

No.	Key question	Use case focus
1	"What geothermal energy potential is located where?" (i.e. the properties of subsurface structures and is they are suited for a GTE exploitation).	no
2	"What geothermal energy potential can be exploited where?" (i.e. where the properties of the subsurface meet the basic conditions needed for specific uses).	no
3	"What are the possible consequences based on current insights?" (i.e. subsidence and seismic events, also with respect to subsurface structures such as faults).	no
4	"Who should be involved" (i.e. who is representing the industry, the regulator, which service providers are active?).	Yes
5	"What is the status of recent permit applications "(i.e. who applied for the permit and what was the outcome?").	Yes
6	"What well performance / seismic data is available for location x"(i.e. in the Netherlands operators are obligated to release seismic data after 5 years of production, to benefit new initiatives using similar technology / seismic conditions).	No
7	"Which state of the art technologies have proved to be feasible in the current regulatory space? (i.e. the use of advanced drilling technologies for deep Geothermal).	Yes

[4] Value Proposition Design https://www.strategyzer.com/books/value-proposition-design

8	"How should risks be mitigated throughout the GTE project life cycle, bases in current best practices? (i.e. technical qualification is mandatory in the pre-construction phase, what does regulation require, which services are available").	No
9	"How is the regulatory landscape expected to develop / change in the next 10 years? (i.e. what are intentions of the regulator and what is the envisioned timeline? When will unconventional technology be allowed to be used)?	Yes

2.3 ENVISIONED SOLUTION

Mock-up & User Journey

This section describes the envisioned user journey, taking the user segments / roles (section 2.2) as a starting point: typically, an engineer who is looking for relevant regulatory related information, based on an input document (PDF). This could be a PDF document describing a request for service on a geothermal energy project.

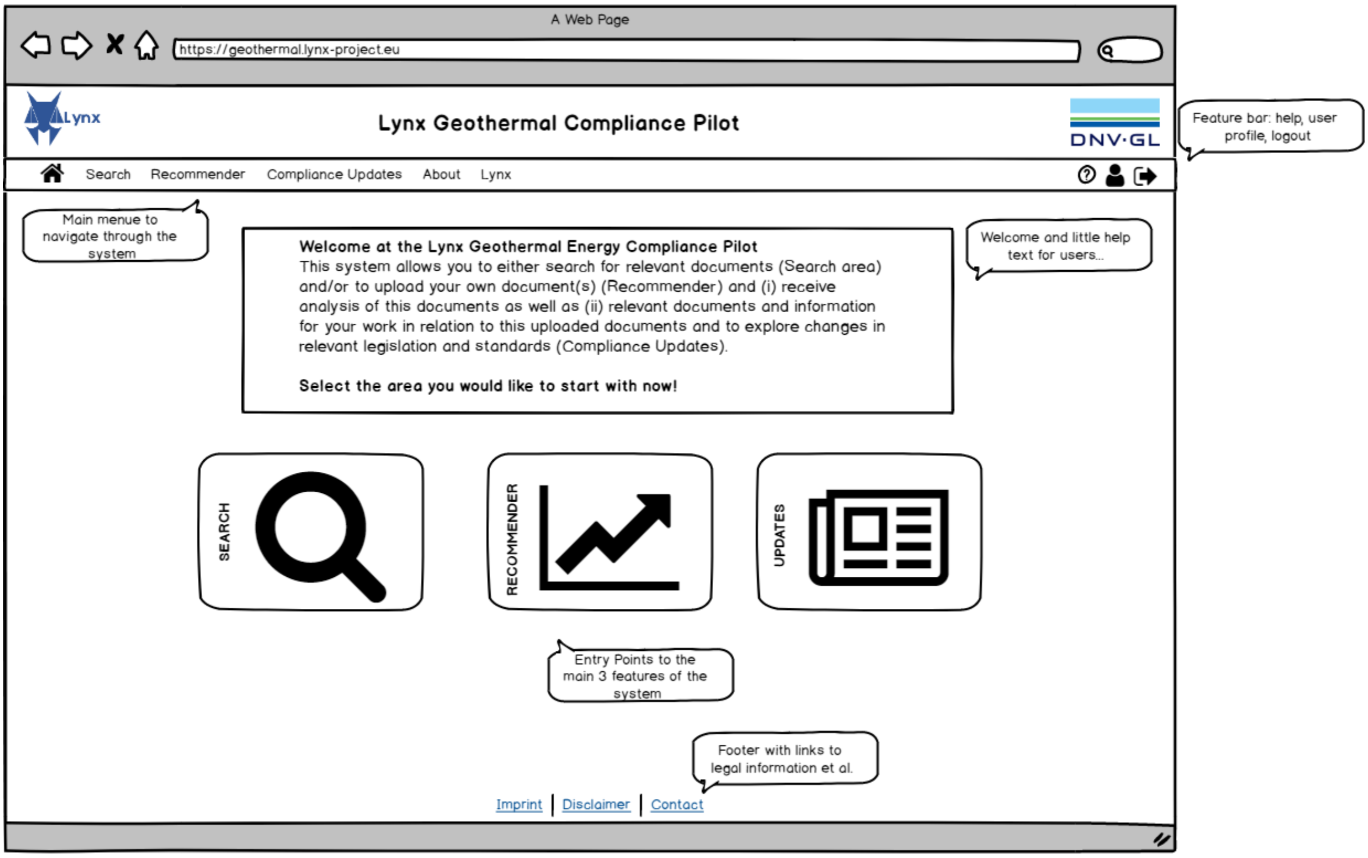

Figure 2 Mockup Start Screen GTE Pilot of Lynx

The Start screen (Figure 2) provides a first overview of the GTE Pilot system and introduces the different modalities to use the platform. Initially we focus on 1) the 'recommender' which allows the user to upload a PDF document to highlight recommendations on concepts relevant to GTE (mainly geographic location, regulation, GTE technological concepts),2) search, a semantic search engine across the known corpora using the GTE knowledge graph and 3) an 'update' feature to highlight latest changes relevant to GTE initiatives.

Feature: Recommender

Once the user selects the 'recommender' option, an upload screen is presented where the user can upload or select a file to the system (fig 2).

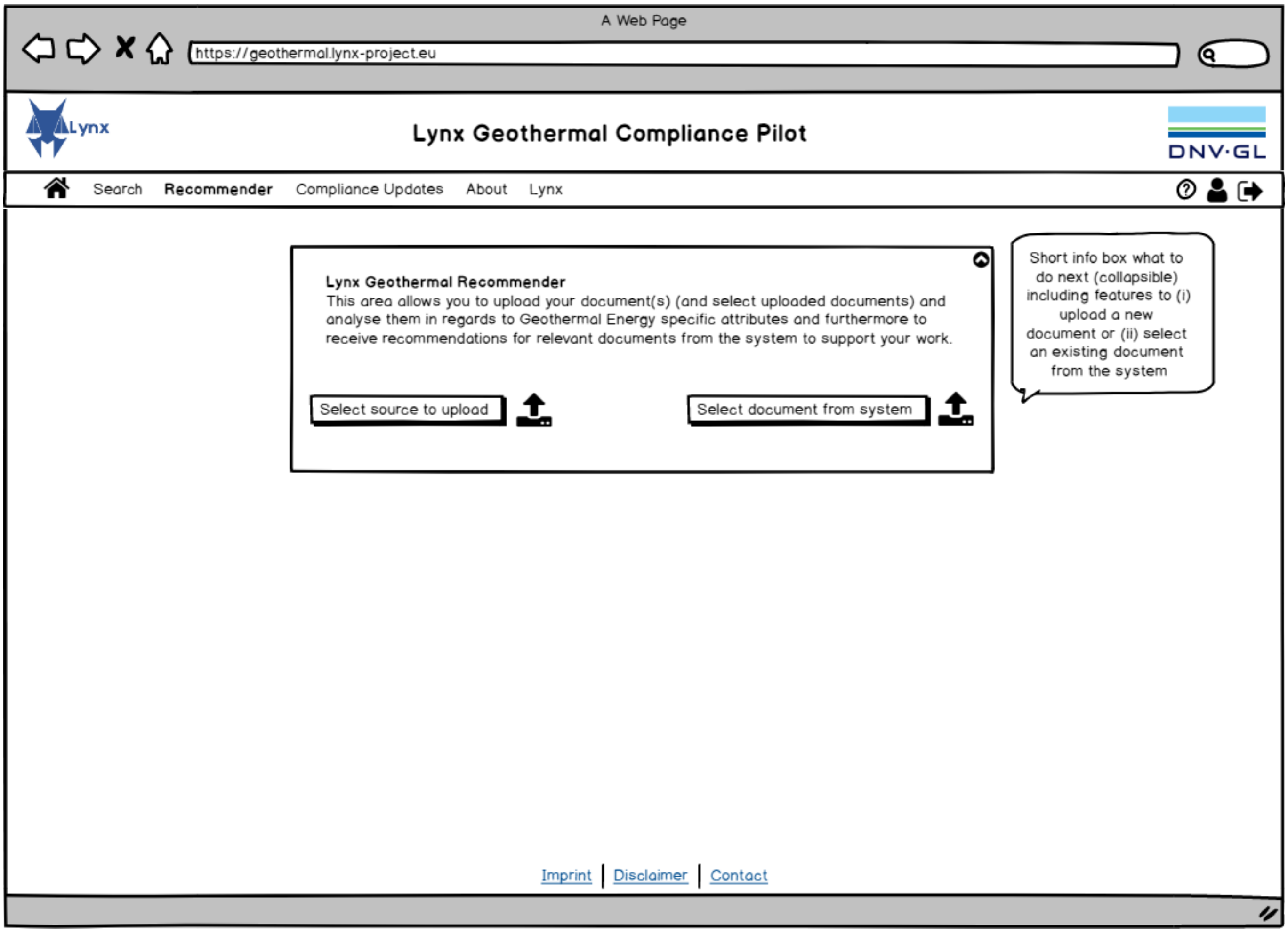

Figure 3 Mockup Start Screen of Recommender Feature of GTE Pilot of Lynx

After uploading the file, it will be processed which takes time (in order to allow asynchronous services to be completed) and needs to be indicated to the user. Figure 3 presents the uploaded document by the user (on the left column) with highlighted information categories found in the document. In this particular example, it found 'Drenthe', which is a city in the Netherlands. The highlight colour indicated it belongs to the category 'geographical location', for which the Recommender found documents relevant to that location. This is displayed in the content box on the right, which presents recommendations based on location (yellow), regulation & standards (blue), GTE concepts (red) or best practices / other featured content (grey).

The Legal Knowledge Graph, which contains both, domain specific as well as generic multilingual vocabularies, and taxonomies and documents etc. is used together with a set of Lynx services (for details on the used Lynx services please see section 3.3 Lynx Services in this document) – namely: SEAR - search service, SeSim - semantic similarity service, EntEx+WSID - entity extraction and disambiguation service, TRANS - machine translation, DCM - document manager and WM - workflow manager, makes it possible to suggest (more) relevant and more specific input (than for instance Google) to the end-user. The state of the technical backend is described in section 3.

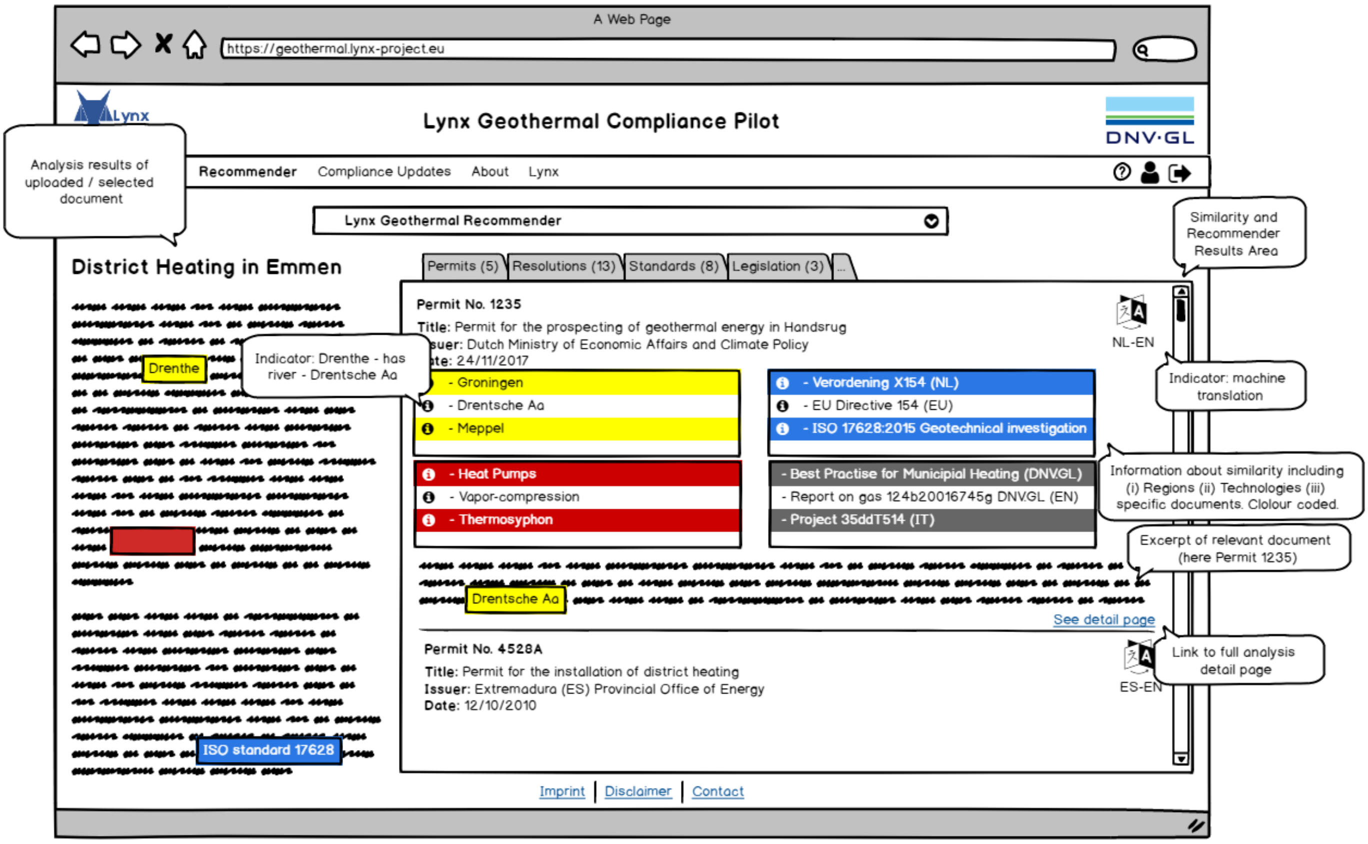

Fig. 3: Mockup Recommender result page of GTE Pilot of Lynx

Feature: Search

The second feature of the GTE Pilot System is semantic search: users can submit search keywords (or short phrases) and be presented categorized / faceted search results (see the MockUp in Figure 4 for better understanding) similar to the results of the recommender feature. Users are able to narrow down the search results using the ‘facets’ presented in the content types box at the right. The facets are provided dynamically by the Legal Knowledge Graph – from various interlinked vocabularies about GTE technologies, about geographical regions and/or document types etc.

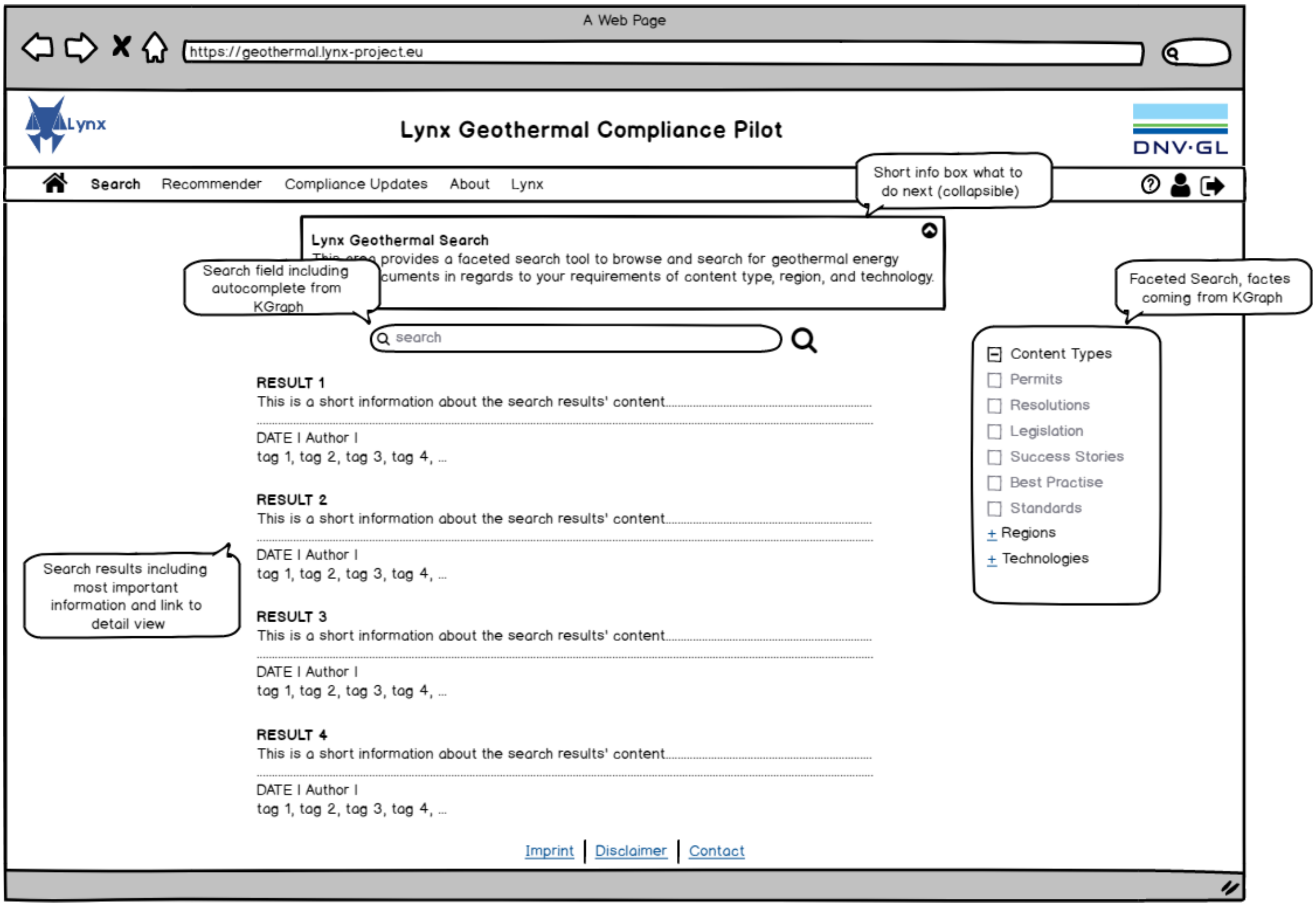

Figure 4 Mockup Search result page of GTE Pilot of Lynx

Feature: GTE updates / notifications

The third feature is an ‘aggregator’ informing the user on most relevant updated regarding their preferences of search keyword (see Figure 5). As most GTE practitioners have acknowledged the dynamics of the market, it would benefit if they could somehow better track topics of interest based on their preferences (e.g. location or specific technology or specific regulation). Early notifications, e.g. on permits of policy development, can be used to act upon or mitigate risks.

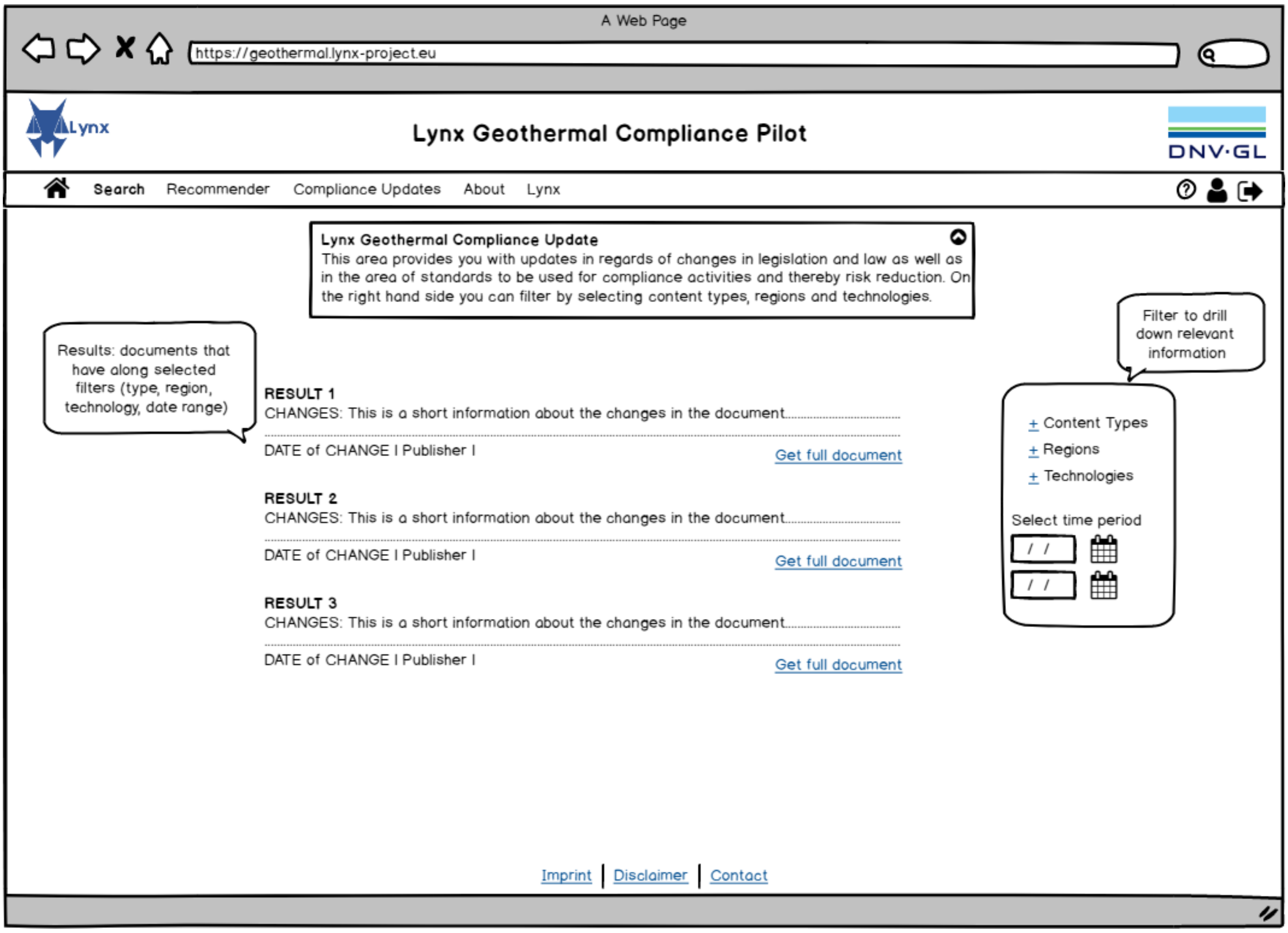

Figure 5 Mockup Compliance / update feature result page of GTE Pilot of Lynx

2.3.1 SCREEN DESIGN PROTOTYPES OF GTE LYNX PILOT SYSTEM

Based on the specification and mock-ups (section 2.3.1), the screen design has been developed for the areas of (i) start page (Figure 6), (ii) the upload function (Figure 7) and (iii) the recommender (Figure 8). The implementation of these areas has started and an alpha version is already in place (see Section 3).

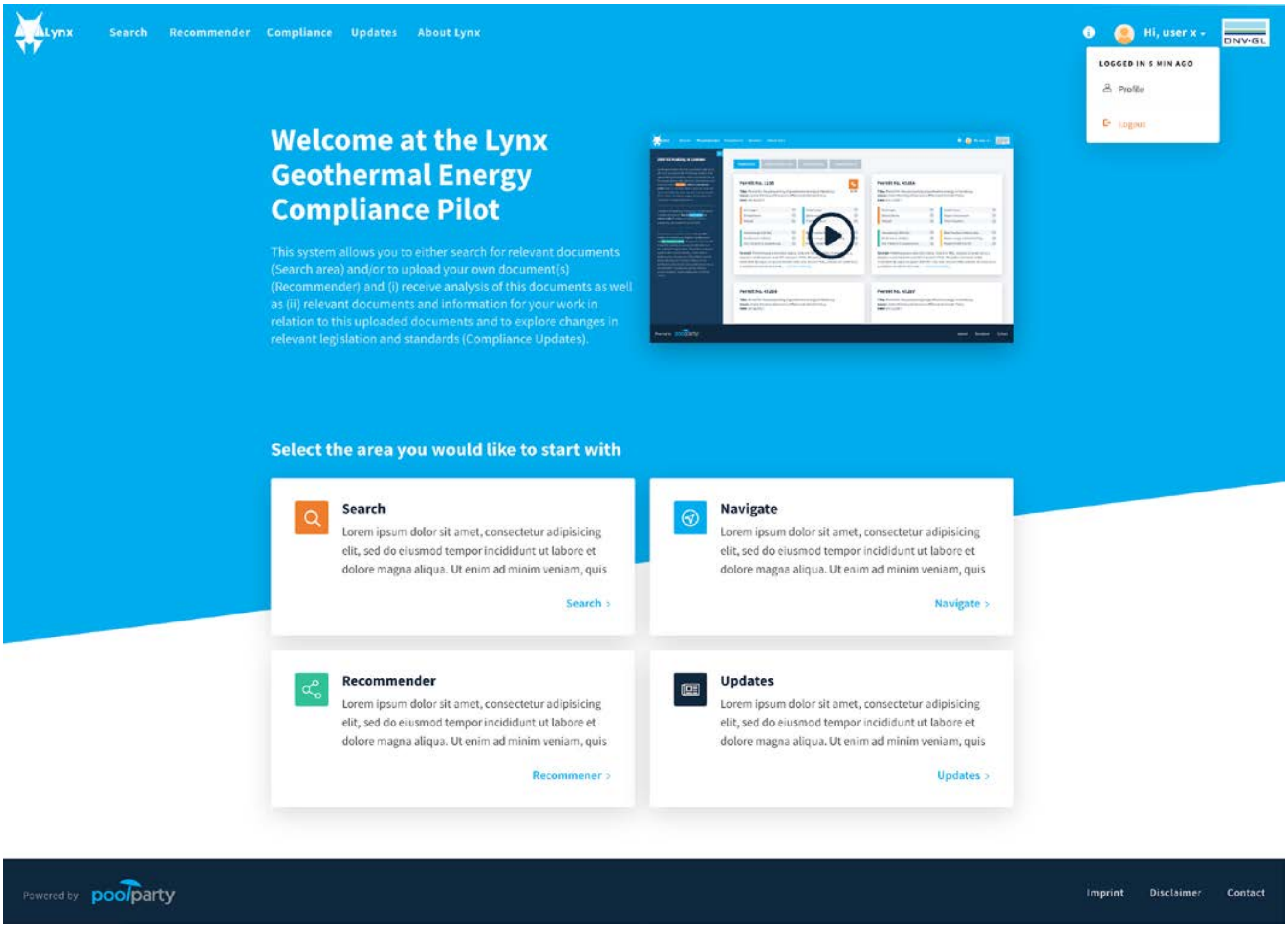

Figure 6 Design Prototype Start Screen GTE Pilot of Lynx

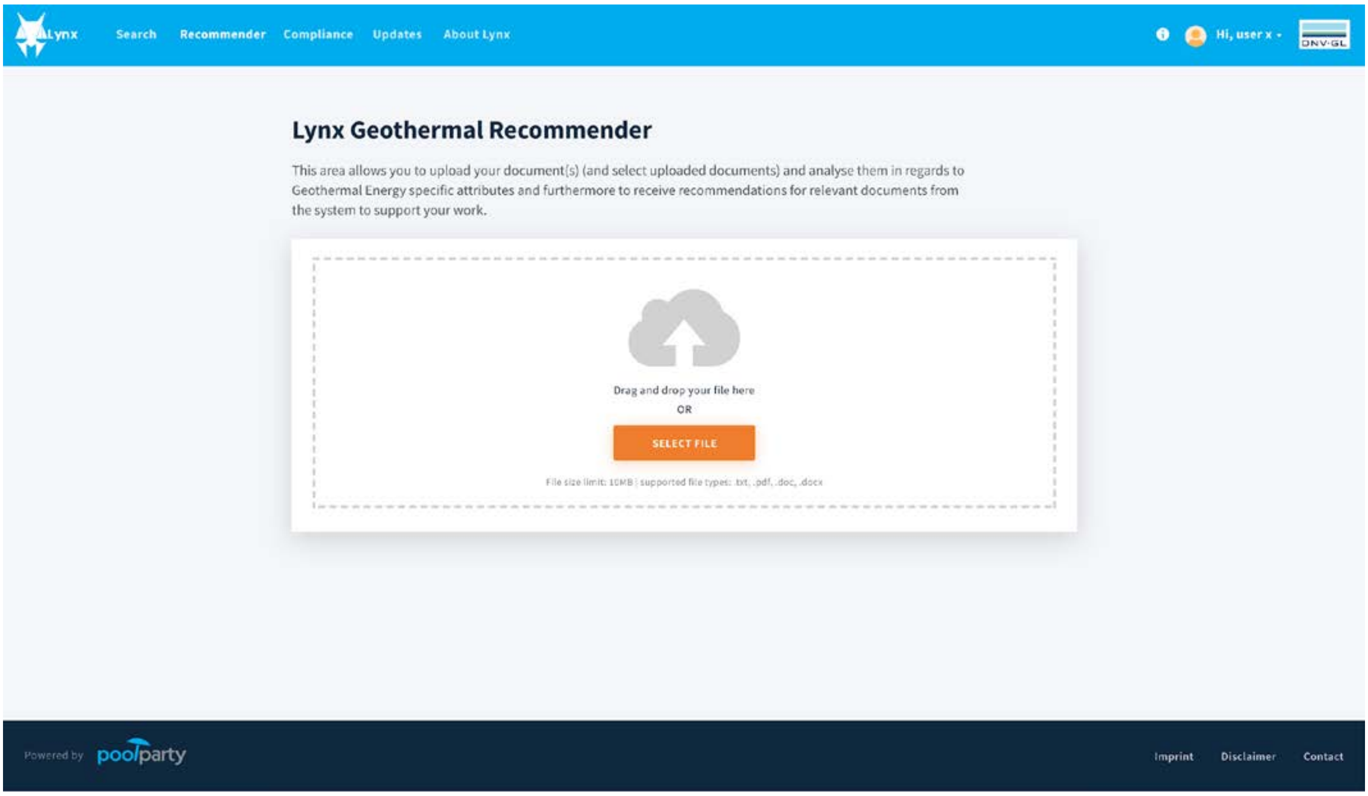

Figure 7 Design Prototype Recommender Upload Page of GTE Pilot of Lynx

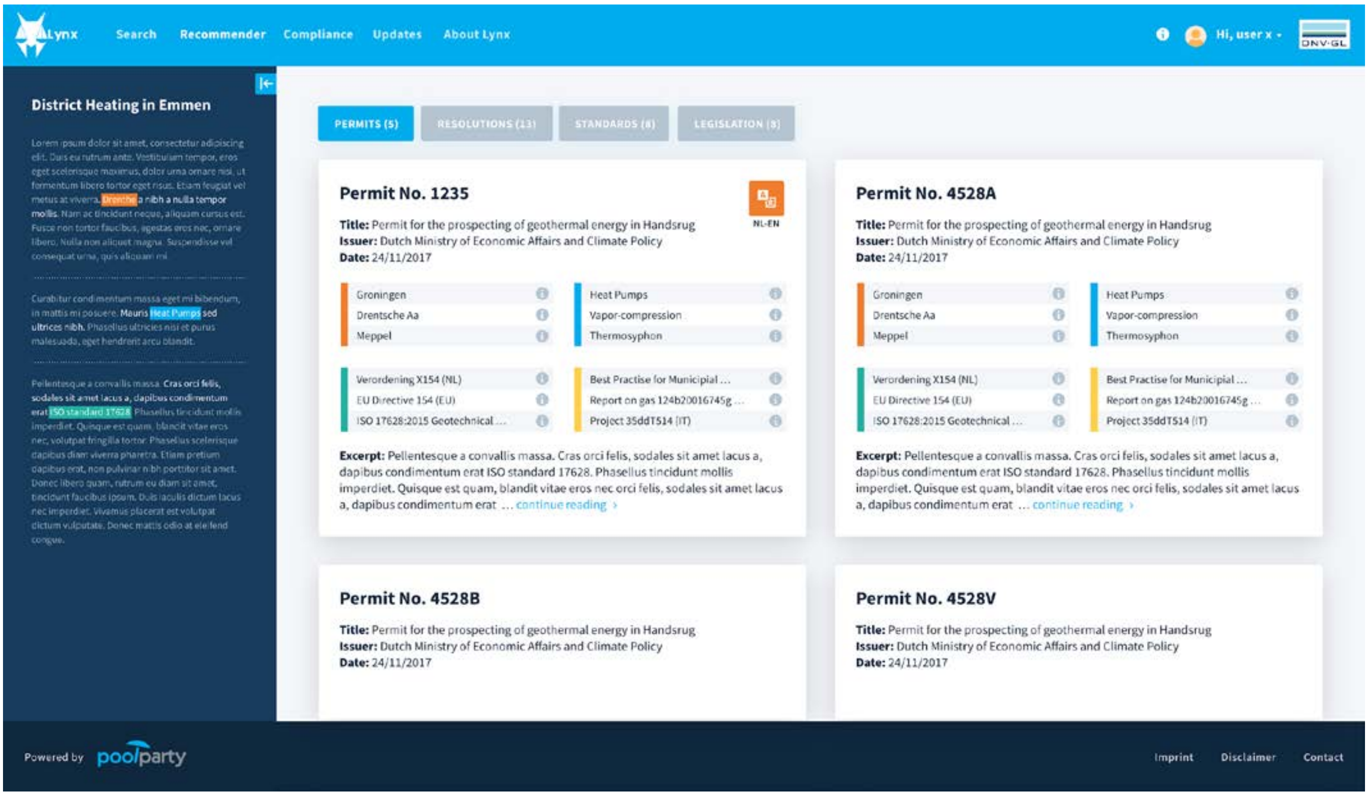

Figure 8 Screen Design Prototype Recommender Results Page of Lynx GTE Pilot

2.4 BUSINESS CASE(S)

Modalities

If the solution proves to be able to add value to the problem assumptions of the GTE use case it is assumed wider adaptation for other sustainable energy alternatives is possible. Nevertheless, for the Lynx project we will focus on the domain of GTE. For the business case we look for evidence that value can be generated, for various stakeholder groups, which exceeds the total operating cost of a Lynx 'instance' for GTE. The following modalities and sources of exploitable value mechanisms are considered (Table 1.

Modality	Value exploitation mechanism
Internal decision support system The target audience would be consultants or engineers who are normally involved in due diligence, technical advisory, inspection or verification services for various stakeholders in the GTE value chain	Used internal: improve efficiency of reducing time to find relevant information and be able to deliver more quality / value to the client. Used external: subscription for customers as a self-service platform or licensed to Engineering / consultancy firms.
External information aggregator: with the current fragmented GTE information landscape, there is room for an 'umbrella' platform aggregating the various information sources.	Potential for 2-sided platform model with the purpose to consummate matches among users and facilitate the exchange of information (and later perhaps services?) and thereby enabling value creation for all participants. The size of the community served can be exploited in various manners and not only by promoting content and services, but also catalise alignment of policies between individual EU countries, which is high on the agenda of the EC (refer to source).

Table 1 Modalities and sources of exploitable value mechanisms

Business Model Canvas

During the transition between WP5 and WP6, the feasibility of the Lynx use case for GTE has been verified by means of a business model canvas exercise. Main purpose is to validate the minimum 'problem-solution' combination is proofing the potential value in order to create healthy business model. The initial outcome of the business model is included in appendix 1.

3 PILOT ARCHITECTURE & SYSTEM

3.1 INTRODUCTION

This section describes the architecture and services as well as the technologies used to implement the Lynx GTE Pilot System. Finally, the section provides information about the current status of the Pilot System at the time of writing this deliverable. Outlook and time plan for the Pilot System can be found in the following section.

3.2 WEB SERVER (USER INTERFACE COMPONENTS)

In order to keep the authentication to the API secure and to overcome browser-specific limitations a sole client-side implementation is not sufficient for the UI component. Therefore, the architecture of the GTE Pilot System user interface can be separated into a client-side frontend and a separate backend application.

The backend runs on Node.js, which is an event-driven JavaScript runtime, and makes use of the Node-js-framework Express in order to simplify basic tasks. Node.js applications can be run easily and are supported by most container application platforms like OpenShift out of the box.

The frontend is based on the JavaScript library and web-framework React, which encapsulates functionality into components. Redux is being used to handle the state of those components in a more scalable way in case of any future additions to the application.

All code of the frontend is written in TypeScript 3.6, which is being compiled to ECMA Script 5. This allows a type-strict and cleaner development approach.

The HTML-templates are built on a theme named Stisla, which itself is based on Bootstrap 4. Additionally, Sass is being used as a CSS-preprocessor in order to support additional functionality like variables in custom stylesheets.

The whole application does not rely on having a database in place, which is caused by the fact that the application currently only separates between public and private use of the APIs. In case the application gets extended to a multi-user environment the use of MongoDB has been contemplated in order to have a full MERN-stack (MongoDB + Express + React + Node) in place.

3.3 LYNX SERVICES

The Lynx services used in this pilot are as follows

- **SEAR – cross-lingual search service**. The search services enablers the full text search over the documents as well as search for specific annotations and/or facets.
- **SeSim - semantic similarity service**. The service computes a score of similarity between documents. The score is based on the underlying LKG, the thesaural part of it, and can only take parts of the LKG into account, therefore implementing aspect-based similarity, for example, only showing documents with relevant locations and/or technologies, etc.
- **EntEx+WSID - entity extraction and disambiguation service**. This service produces the necessary annotations of the documents enriching the documents with entities from the LKG. These entities are afterwards used by SeSim service to compute the similarity score.
- **TRANS - machine translation**. The service provides the translations of the documents. Though the EntEx+WSID and the SeSim services are multilingual, the user might benefit

from being able to translate the documents. Moreover, the service might be used to integrate even more languages.

- **DCM - document manager**. The document manager stores the documents and their annotations.
- **WM - workflow manager**. The service is used to implement the necessary workflows, for example, EntEx+WSID followed by SeSim.

Whereby SeSim, EntEx and TRANS services have been trained and customised with geothermal energy (GTE) specific data and documents as well as models (in the form of taxonomies and terminologies) to ensure the services work well for the Pilot System specific requirements in regards of languages and translation, and the recommender service. In the course of the tests of the GTE Pilot System – see section 4 Outlook & Timeplan – the Pilot System will be tested by experts and consultants (the main target group) and the feedback by these stakeholders will be directly used for continuous improvements.

3.4 Status of Alpha System in place

The Alpha System = 1st release has been setup over the last months. The development system is established at SWC infrastructure to ensure and follow a fast and easy as well as agile development process. The system will be deployed at Lynx infrastructure later. This means that Lynx Services and software components are available for development, for example SWCs PoolParty Semantic Suite where Knowledge Graph management takes place but also other services as described in section 3.3 in this document.

The backend has been deployed, means the pilot is communicating and working with the Lynx Services that have been specified and orchestrated for this Pilot. The pilot system so far is working with a subset of the overall data sources and documents and the Legal Knowledge Graph as specified in the requirements specification. These sources and the pilot specific Knowledge Graph will be further docked onto the system / will be further developed over time.

The frontend has been designed and setup / implemented (please see section 3.2 above for technical details) and has been deployed at SWC infrastructure at the moment. Please see screenshots and description of interfaces and features provided in section 2.3.1. Mock-ups and User Journey.

This Alpha System or 1st Release of the GTE Pilot of Lynx will be presented to different stakeholders and thereby further and continuously developed over the coming weeks based on the feedback of these stakeholders (that represent the end users / target groups of the Pilot System). For more information on the time plan please see section: 4 Outlook and time plan.

4 OUTLOOK & TIMEPLAN

The following time plan has been developed for the Pilot implementation and continuous review and refinement.

Alpha Pilot (Proof of Concept, First Release) Version in place: November / December 2019

Internal Validation campaign (Jan-Feb 2020)

1) Finalize demonstrator, including a visualization of the envisioned system architecture;
2) Internal DNV GL by means of one or more webinar(s);
3) Client visit(s) to present and discuss the pilot with geothermal energy experts;
4) Presentation to European Geothermal Energy Council (e.g. online meeting, https://www.egec.org/);

Promotion & Engagement & Improvements (Feb - Oct2020)

5) Continue engagement with the Geothermal energy industry to test interest & to raise appetite;
6) Private preview for advocates with the purpose to 1) complement the ontology / knowledge graph and 2) complement the identified information sources;
7) Evaluate 1) and 2) and define what (minimum) improvements are needed for second release (Beta version, see below);
8) Continuous improvements and expansions taking place based on continuous feedback by several stakeholders (feature set, Knowledge Graph, data and documents);
9) Release of Beta Version (2nd release) in place by end of early Q2/2020 (04/05-2020);
10) 3rd and final Release in place by 10/2020 (remark: several minor / small releases in between);
11) Support Exploitation Plan and Activities with regards to the pilot system - as demonstrator for Lynx as well as a standalone 'vertical solution;

Evaluation, Engagement & Uptake (Sept2020 - November/ December 2020)

12) Continuous evaluation of the pilot system by means of continuous feedback by stakeholders
13) Continuous improvements and expansion of the Pilot System regarding the available feature set and data & documents. Continuous expansion of the GTE Knowledge Graph.
14) Identify and specify which improvements can be done within the Lynx project timeline and what needs to be done afterwards to bring the pilot to final production.
15) Continuously work on the uptake and sustainability plan (aligned with exploitation activities) AND bring the Pilot System to a stage that it is independent from Lynx project.

APPENDIX 1 ENVISIONED BUSINESS MODEL CANVAS

During the WP6 exploitation activities, the value proposition of the Lynx use case for GTE was translated in a business model canvas below. The colour coding indicates elements of the use case that are domain specific (blue) or more generic for the lynx platform and its services.