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

D5.6 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 (40 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); Jeroen Alberts (DNV GL)
REVIEWERS	Christian Sageder (OLS); Pascual Boil Ballesteros (Cuatrecasas)
VERSION | STATUS	v0.1 | Final
NATURE	Demonstrator
DISSEMINATION LEVEL	Public
DOCUMENT DOI	10.5281/zenodo.4298892
DATE	30/11/2020 (M36)

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	09/11/2020	Pieter Verhoeven (DNV GL)
02	Second version	23/11/2020	Pieter Verhoeven (DNV GL)
03	Final version	30/11/2020	Pieter Verhoeven (DNV GL)

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
MT:	Machine Translation
ET:	Entity linking
EC:	European Commission
EU:	European Union
SeSim:	Semantic Similarity
DCM:	Document Manager
ESG:	Environmental, Social, and Corporate Governance

EXECUTIVE SUMMARY

This deliverable constitutes the 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 are subject to accelerated or even a transformative future growth. During this transformation, friction in access to relevant regulatory information can cause delays (or require breakthroughs) which can be solved by the use of knowledge graph technology. 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 that show how Lynx services can interact with the end-user. Section 3 describes the pilot architecture and the current system in place. This section mainly talks about the front-end stack and the backend integration with the Lynx services. In this section the 'final' pilot is presented. Section 4 describes a reflection on the development in terms of 'learning and recommendation'.

TABLE OF CONTENTS

1 INTRODUCTION TO GEOTHERMAL ENERGY 6

2 USE CASE DESCRIPTION 8

2.1 Introduction 8

2.2 Objective of the Pilot 9

2.3 Envisioned Solution 11

2.4 Screen Design Prototypes of GTE Lynx Pilot System 17

2.5 Business case(s) modalities 19

3 PILOT ARCHITECTURE & SYSTEM 19

3.1 Introduction 19

3.2 Web server (user interface components) 19

3.3 Lynx services 20

3.3.1 Document Ingestion 20

3.3.2 End user interaction 21

3.3.3 Summary of Lynx services 21

3.4 Screenshots of the final pilot 23

3.4.1 Pilot landing page 23

3.4.2 Recommender upload page 24

3.4.3 Recommender result page 24

3.4.4 Search page 25

3.5 Promotion and Validation 26

4 REFLECTION AND RECOMMENDATIONS 27

5 References 30

5.1 Lynx Deliverables 30

LIST OF FIGURES

Figure 1 Overview on geothermal capacity - power & heating, EGEC Report 2018 (screenshot).....6
Figure 2 Geothermal energy overview (source: GeoVision report by the U.S. DOE)........................7
Figure 3 High level introduction of permits and consents required during the lifecycle of an Geothermal Energy operation..8
Figure 4 Mockup Start Screen GTE Pilot of Lynx...12
Figure 5 Mockup Start Screen of Recommender Feature of GTE Pilot of Lynx..........................13
Figure 6 Mockup Recommender result page of GTE Pilot of Lynx..14
Figure 7 Mockup Search result page of GTE Pilot of Lynx..15
Figure 8 Mockup Compliance / update feature result page of GTE Pilot of Lynx........................16
Figure 9 Design Prototype Start Screen GTE Pilot of Lynx...17
Figure 10 Design Prototype Recommender Upload Page of GTE Pilot of Lynx...........................18
Figure 11 Screen Design Prototype Recommender Results Page of Lynx GTE Pilot.....................18
Figure 12 GTE recommender Lynx services interaction...21
Figure 13 Design pilot Start Screen GTE Pilot of Lynx...23
Figure 14 Design Recommender Upload Page of GTE Pilot of Lynx..24
Figure 15 Design Recommender Results Page of Lynx GTE Pilot...25
Figure 16 Design Search Page of Lynx GTE Pilot..25
Figure 17 Agenda IGA Stay Engaged Campaign..26
Figure 18 Lynx "emerging technologies for standardisation" webinar for IGA..............................26

LIST OF TABLES

Table 1 'job to be done' key questions and use case focus..11
Table 2 Modalities and sources of exploitable value mechanisms...19

1 INTRODUCTION TO GEOTHERMAL ENERGY

Emerging digital technologies can contribute to the energy transition and the ambitions set in the Paris Agreement

Emerging alternative energy resources, including biomass, solar, wind, green 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. This is not only a challenge for the incumbent energy companies, but also a call for the whole supply chain to train their ‘innovation muscle’. Digital technologies, such as knowledge graphs, can contribute to the broader challenges of mitigating climate change. Therefore we aim to combine the Lynx consortium partners’ knowledge and expertise to address this challenge.

As a proxy to this challenge, we will focus on the domain of geothermal energy for which regulatory information is fragmented and for which most EU member countries will have ambitions to increase the use of geothermal energy in the energy mix. Figure 1 below pictures a map of the EU with the geothermal heat potential at 200 meters, which can be used for electricity and district heating[1].

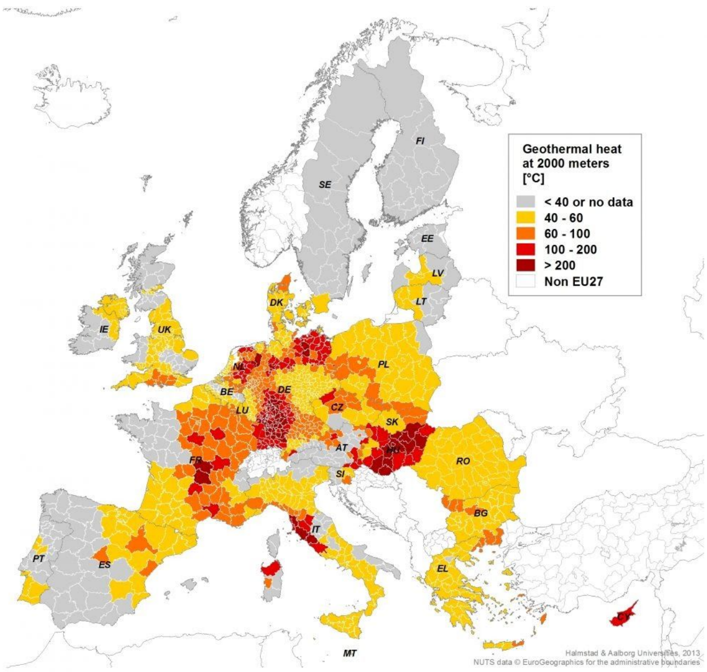

Figure 1 Overview of geothermal heat at 200 meters (˚C) in Europe

What is geothermal energy?

Geothermal energy[2] is heat generated in the sub-surface of the earth. Geothermal steam and fluids 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’ (see figure 2). Depending on the temperatures, this fluid can be used to produce clean electricity or as a baseload for municipal district or industry heating and cooling.

[1] Source: European Commission, Atlas of Geothermal Resources in Europe. https://www.euroheat.org/wp-content/uploads/2013/05/Heat-Roadmap-Europe-II-2013.pdf
[2] Geothermal energy on Wikipedia: https://en.wikipedia.org/wiki/Geothermal_energy

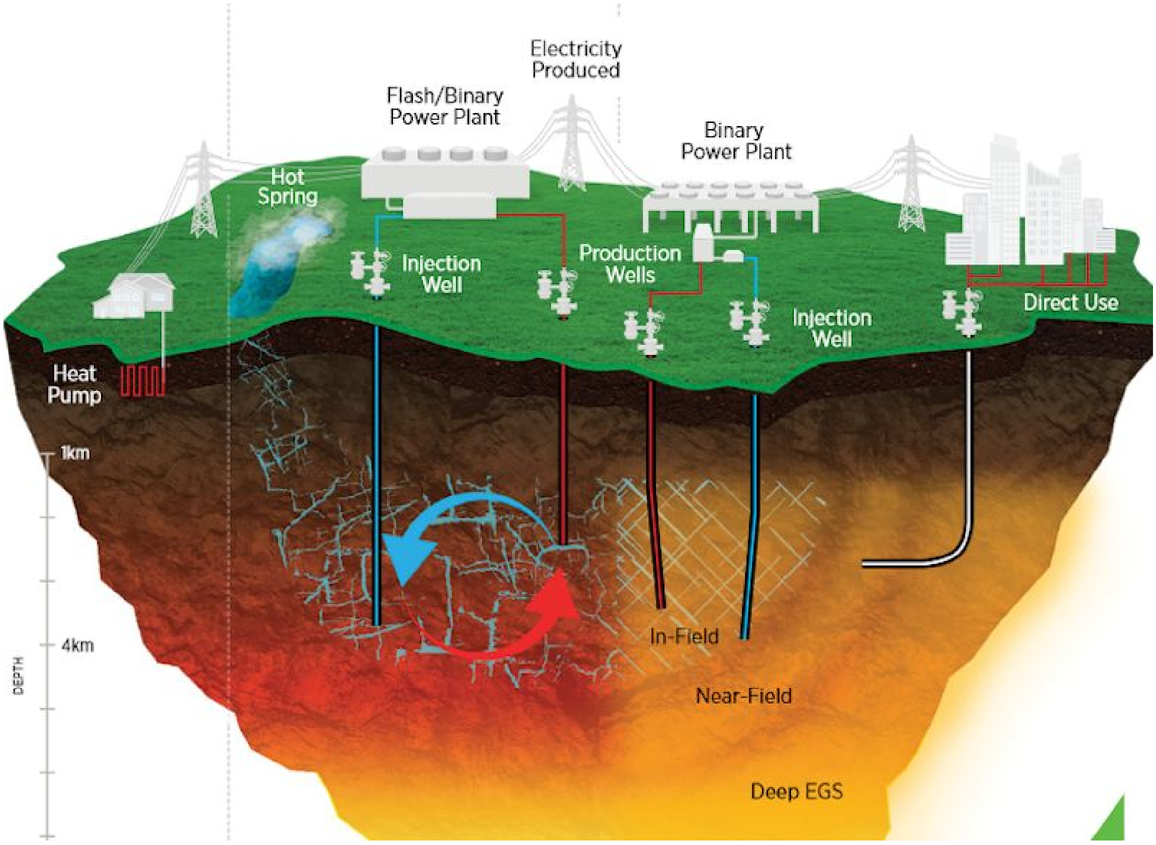

Figure 2 Geothermal energy overview (source: GeoVision report by the U.S. DOE)

Challenges for the Geothermal Energy domain

Geothermal energy is a sustainable alternative energy with a significant potential for many EU countries to contribute to achieving 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 energy potential against acceptable economics, breakthroughs are needed to reduce cost, as well as risks related to financing, applied technology and market mechanisms. Emerging technologies in the domain of digitalisation can help to tackle the challenges related to the need to accelerate growth of the geothermal energy industry and become a professional & matured industry. For example, digitalisation has the potential to accelerate the development of 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].*

Government policies for advancing the energy transition

Governments play a crucial role in legislating and assuring compliance to mitigate safety and environmental risks in all industries and the Energy Industry in specific due to the transition it is in. With the expected growth in sustainable energy alternatives on the one side, continuous standardisation of technology to bring down costs and risks is expected on the other side. Most countries will individually or in coalition develop policies and laws[3]. Governments will seek balance in the use of subsidy schemes to accelerate growth and develop regulation or legislation to mitigate safety and environmental risks, in

[3] chapter 5.5 "POLICIES ADVANCING THE ENERGY TRANSITION" of DNV GL Energy Transition Outlook report © DNV GL AS. 2020. All rights reserved. url: www.eto.dnvgl.com

order to guide sustainable growth of both technologies and markets. To illustrate the regulatory mechanism, an example of the permitting policy for the Netherlands is included in the text box below:

***General introduction of the permitting process in the Netherland*[4]**
In the Netherlands, geothermal operators must comply with the same set of laws as other mining companies: the Dutch Mining Act. SodM is the regulator responsible for supervising geothermal activities which are classed as mining activities under the Dutch Mining Act (including the Mining Decree and the Mining Regulation). When a geothermal operator wants to develop a geothermal project, various permits and consents are required. A summary of the most important ones is given in the following paragraphs.
The first one is the exploration permit (opsporingsvergunning) issued by the Ministry of Economic Affairs and Climate Policy. The permit provides the operator with exclusive rights to explore for geothermal reservoirs in the licensed area. It covers the time frame from exploration (finding) of the reservoirs to drilling of the exploration wells. Typically, these exploration wells are subsequently also used as production wells. The permit sets out the framework and conditions for the geothermal project, the organization, perceived risks and mitigations.
To continue the process by drilling geothermal wells and building a geothermal installation, a location permit (omgevingsvergunning) is required. Then, if geothermal wells have been drilled successfully and a surface installation has been built, a production permit (winningsvergunning) is needed to produce heat. In addition, approval from the ministry is required for the production plan (winningsplan).
The various permits provide the framework within which geothermal heat can safely be produced with minimal damage to the environment.
During the various stages of the project, the Dutch Mining Act requires operators to meet a number of monitoring and reporting conditions concerning the geothermal site and wells. In addition to these, additional conditions with respect to monitoring, reporting, research and/or risk mitigation measures can be imposed per permit.

Figure 3 High level introduction of permits and consents required during the lifecycle of an geothermal energy operation

2 USE CASE DESCRIPTION

2.1 Introduction

Following the high level introduction of permits and consents required during the lifecycle of an geothermal energy operation (in the Netherlands) we assumed that access to the latest regulatory content, at the right time and for the right person, is a problem worth solving by exploring the feasibility of the Lynx services. In this chapter the use case is described by introducing the problem assumptions, the overall issues & challenges for the envisioned user segments as well as the mockups of the initial solution.

The viability it the LYNX use case for geothermal energy was done by testing 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 means a competitive disadvantage and limits international competition and, therefore, the potential benefits of economies of scale / standardisation. The second assumption is

[4] Source: Ensuring safe growth of the geothermal energy sector in the Netherlands by proactively addressing risks and hazards, EBN 2020, Cambridge University Press: 02 June 2020, https://doi.org/10.1017/njg.2020.3

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 member 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 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. Nowadays, two types of regulatory regimes coexist in Europe: countries that have specifically addressed geothermal energy clauses in the mining act and countries without any specific regulation for geothermal energy.

4) Along the tests with geothermal energy companies, we learned that regulatory concerns are also strongly related to other factors like safety and environmental risks. The geothermal energy pilot, initially built for permits, could probably also be used for safety and environmental risks related content. E.g. identifying concepts that can be related to known risks and best practices on mitigation measures.

The demonstrator for pilot 2 is mainly focused on the first problem assumption: geothermal energy regulatory information being fragmented. Although in the final stage of the project, we have pivoted to test other aspects as well (safety and environmental related content).

2.2 Objective of the Pilot

The objective of the pilot is to understand future barriers caused by the transformative mechanisms of the energy transition and to learn if semantic technologies can be used to take away friction in e.g. access to regulatory information. In this section we introduce potential issues and challenges to characterise the emerging 'problem to solve'. As well as an overview of stakeholders who are likely to have an interest in solving friction due to these forces.

Issues and challenges

Related to the problem assumption mentioned in the introductions, the following mechanisms are taken into account in the use case for Lynx:

- **Fragmented information** needed to assure regulatory compliance imposes a risk during decision making in early phase geothermal energy initiatives. High risk profiles lead to higher interest rates (risk mark-ups) and fewer bankable propositions. But also regulatory uncertainties like varying lead time to permit approval impose significant risks to early phase initiatives.
- **Growth:** geothermal energy is expected to play a pivotal role in the energy transition and the EU's ambition to reduce 40% CO_2 emission by 2030 and therefore needs to grow (e.g. today in the Netherlands, 17 doublets are operating. This should to grow to 175 doublets in 2030 and 700 in 2050)[5] and increase the need for intuitive access to regulatory information.
- **Maturing discipline:** Stakeholders acknowledge the geothermal energy 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.

[5]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 and effectively managed in order to improve bankability of geothermal energy initiatives.
- Mechanisms like **economies of scale, standardisation and knowledge sharing** between geothermal energy hotspots (EU) and governmental participation / risk taking, might help accelerate the required growth of the industry.
- 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 project life cycle[6]. On average it takes 7 years of exploration, design and construction before geothermal installations can be commissioned, and energy will be produced. During this period multiple stakeholders are involved in developing a bankable geothermal energy initiative within the boundaries of (inter)national regulations and policies. All these stakeholders need accurate regulatory information and might be a potential user of the Lynx platform as described in this report. The Lynx pilot for geothermal energy was built with the following potential stakeholders in mind:

- Regional Energy Coordinator
- Subject matter experts
- Project developers
- Drilling companies
- Operators
- End-users
- 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 pilot for geothermal energy:

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

[6] 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 comes first. During the engagements with geothermal energy practitioners we talked about what is keeping them busy and the type of friction they experience when looking for relevant regulatory information. The approach is inspired on value propositions design[7] 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 pilot, 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 if 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).	No
8	"How should risks be mitigated throughout the geothermal energy project life cycle, based on 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 the intentions of the regulator and what is the envisioned timeline? When will unconventional technology be allowed to be used?)	Yes

Table 1 'job to be done' key questions and use case focus

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

2.3 Envisioned Solution

The initial vision on solving the fragmented nature of regulatory information was to create a tool that would recommend relevant information sources, based on an input document from the user. For example, a document describing the specification of an envisioned geothermal energy operation. This chapter described the initial mock-up created at the start of this project and used to test with potential users of the product.

Product landing page

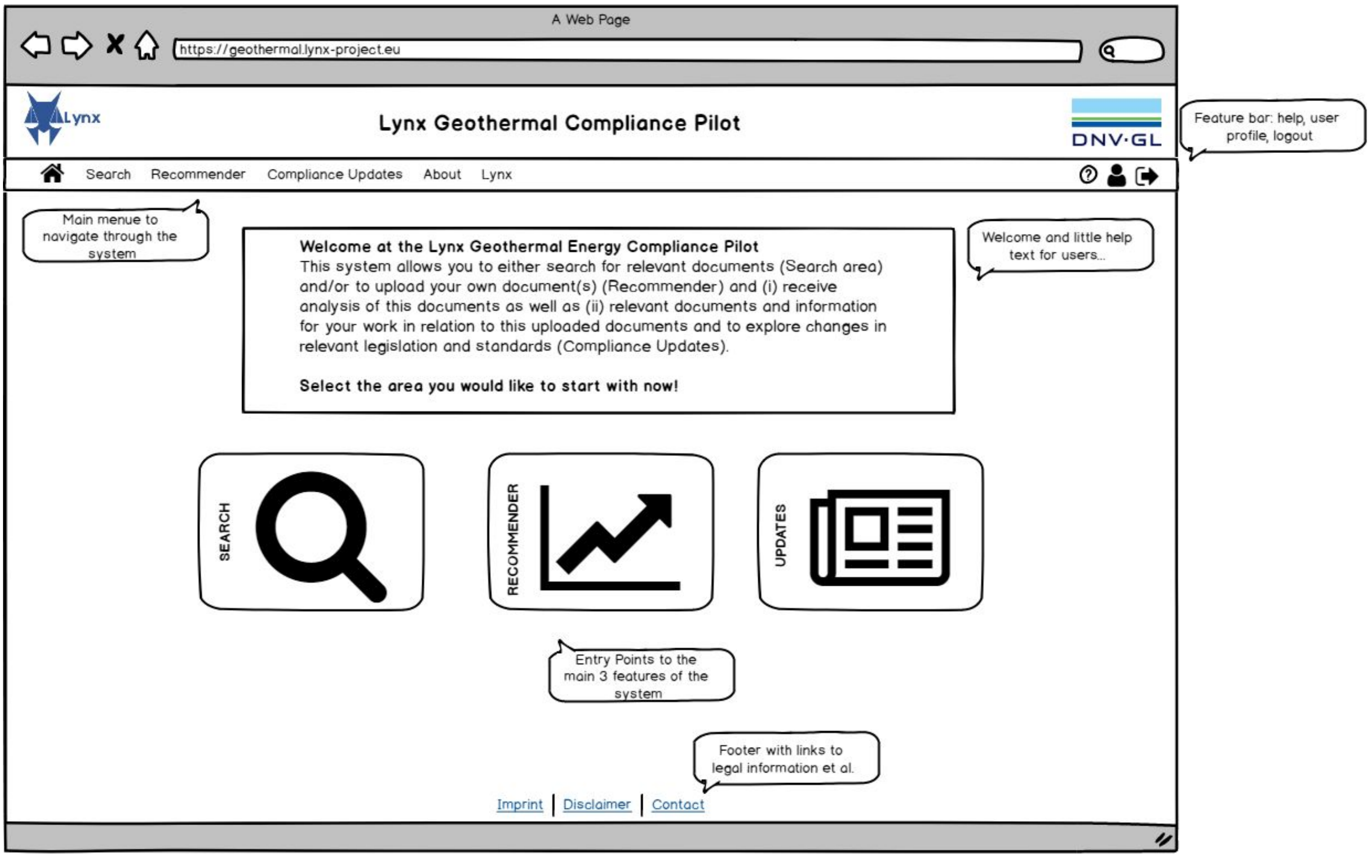

Figure 4 Mockup Start Screen GTE Pilot of Lynx

The Start Screen (Figure 4) provides a first overview of the geothermal energy 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 geothermal energy (mainly geographic location, regulation, geothermal energy technological concepts), 2) search, a semantic search engine across the known corpora using the geothermal energy knowledge graph and 3) an 'update' feature to highlight latest changes relevant to geothermal energy initiatives.

Feature: Recommender

Once the user selects the 'recommender' option, an upload screen is presented where the user can select or upload a file to the system (fig 5).

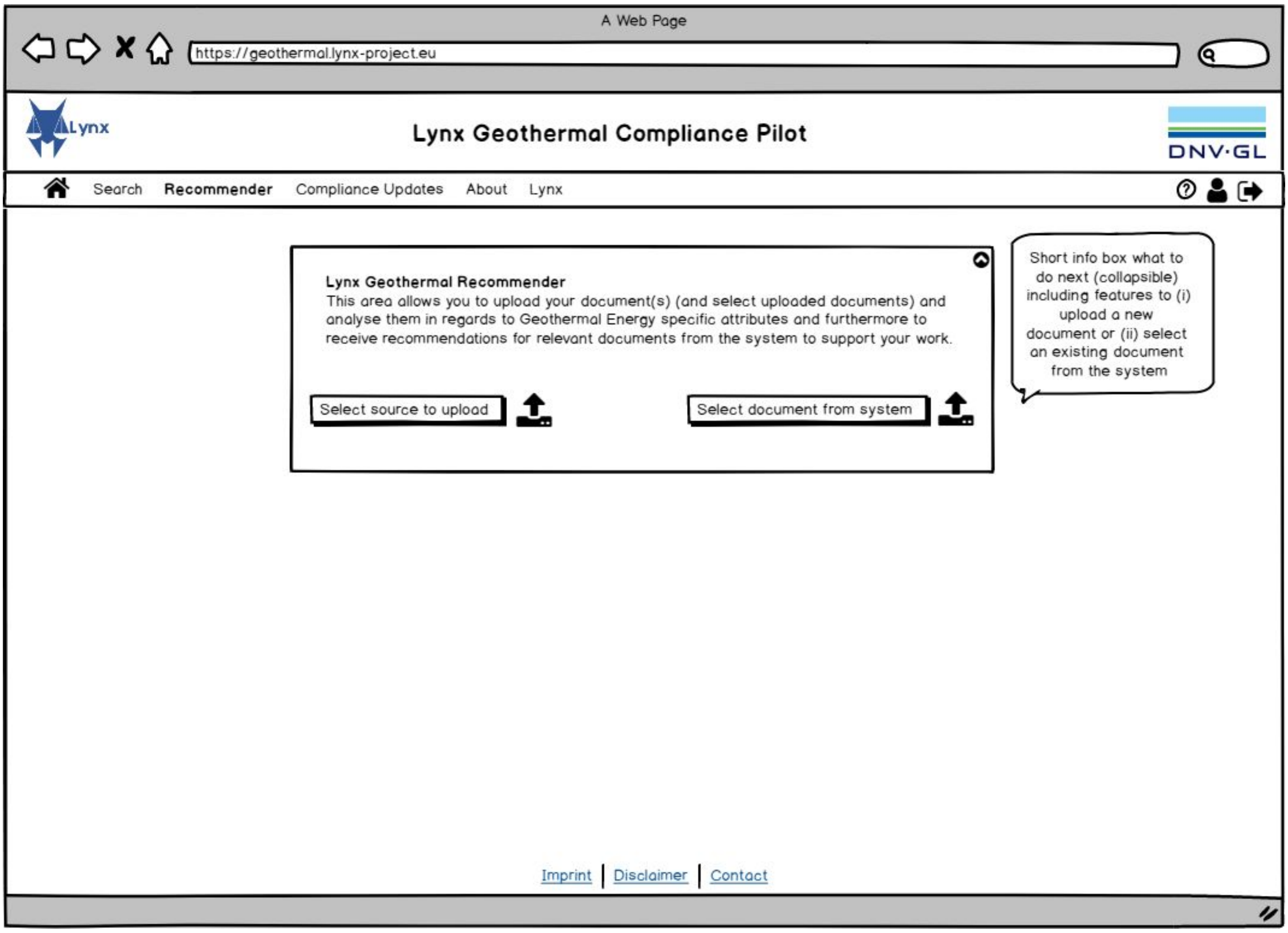

Figure 5 Mockup Start Screen of Recommender Feature of GTE Pilot of Lynx

After uploading the file, this will be processed, which takes time in order to allow asynchronous services to be completed. This information needs to be indicated to the user. Figure 6 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 province in the Netherlands. The highlight colour indicates that 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), geothermal energy concepts (red) and best practices / other featured content (grey).

The Legal Knowledge Graph contains domain specific and generic multilingual vocabularies, taxonomies, documents etc. The state of the technical backend is described in section 3.

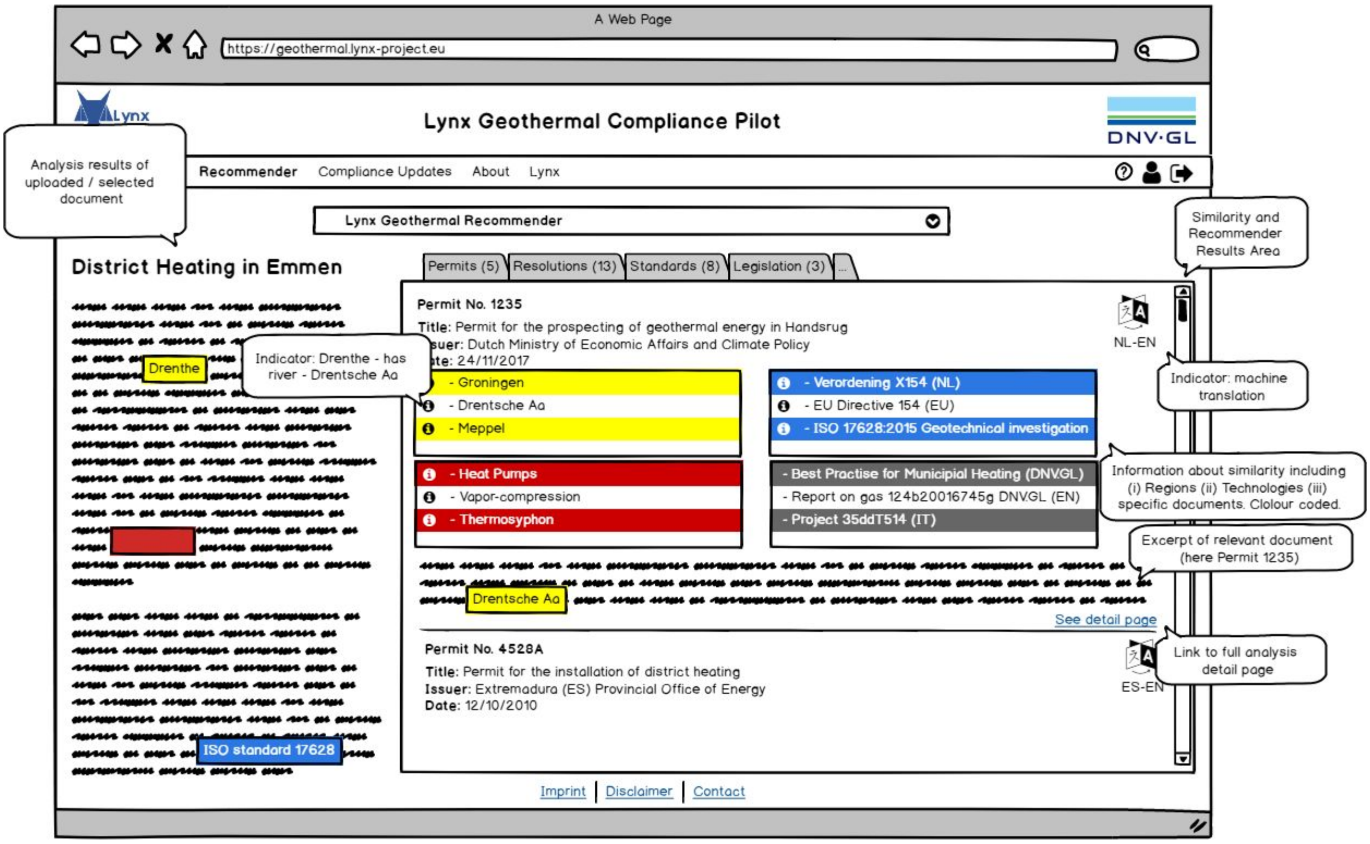

Figure 6: Mockup Recommender result page of GTE Pilot of Lynx

Feature: Search

The second feature of the geothermal energy Pilot System is semantic search: users can submit search keywords (or short phrases) and retrieve categorized / faceted search results (see the Mockup in Figure 7 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 geothermal energy technologies, about geographical regions and/or document types etc.

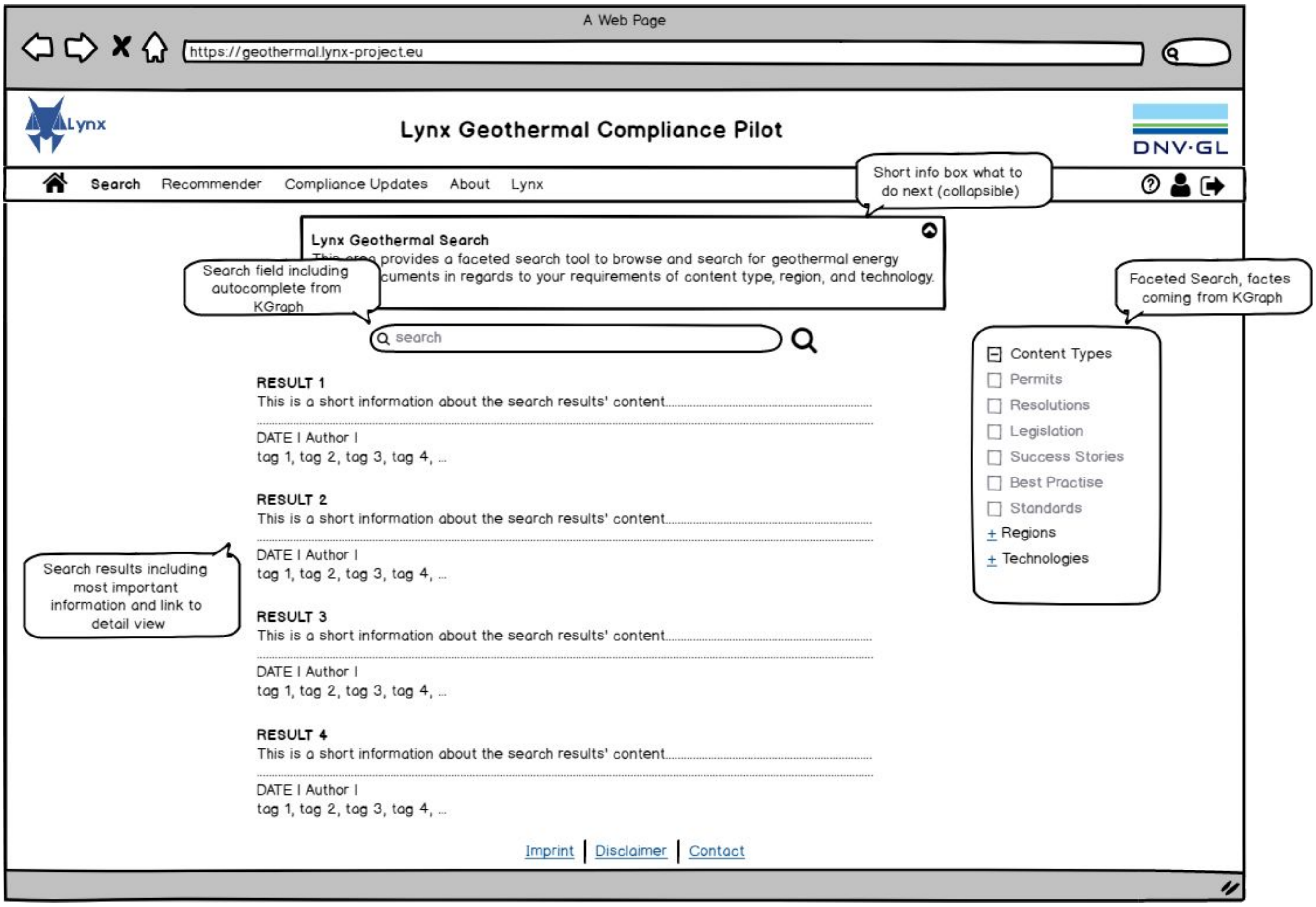

Figure 7 Mockup Search result page of GTE Pilot of Lynx

Feature: GTE updates / notifications

The third feature is an ‘aggregator’ informing the user on most relevant updates regarding their preferences of search keywords (see Figure 8). As most geothermal energy 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 on permits of policy development, for instance, can be used to act upon and mitigate risks.

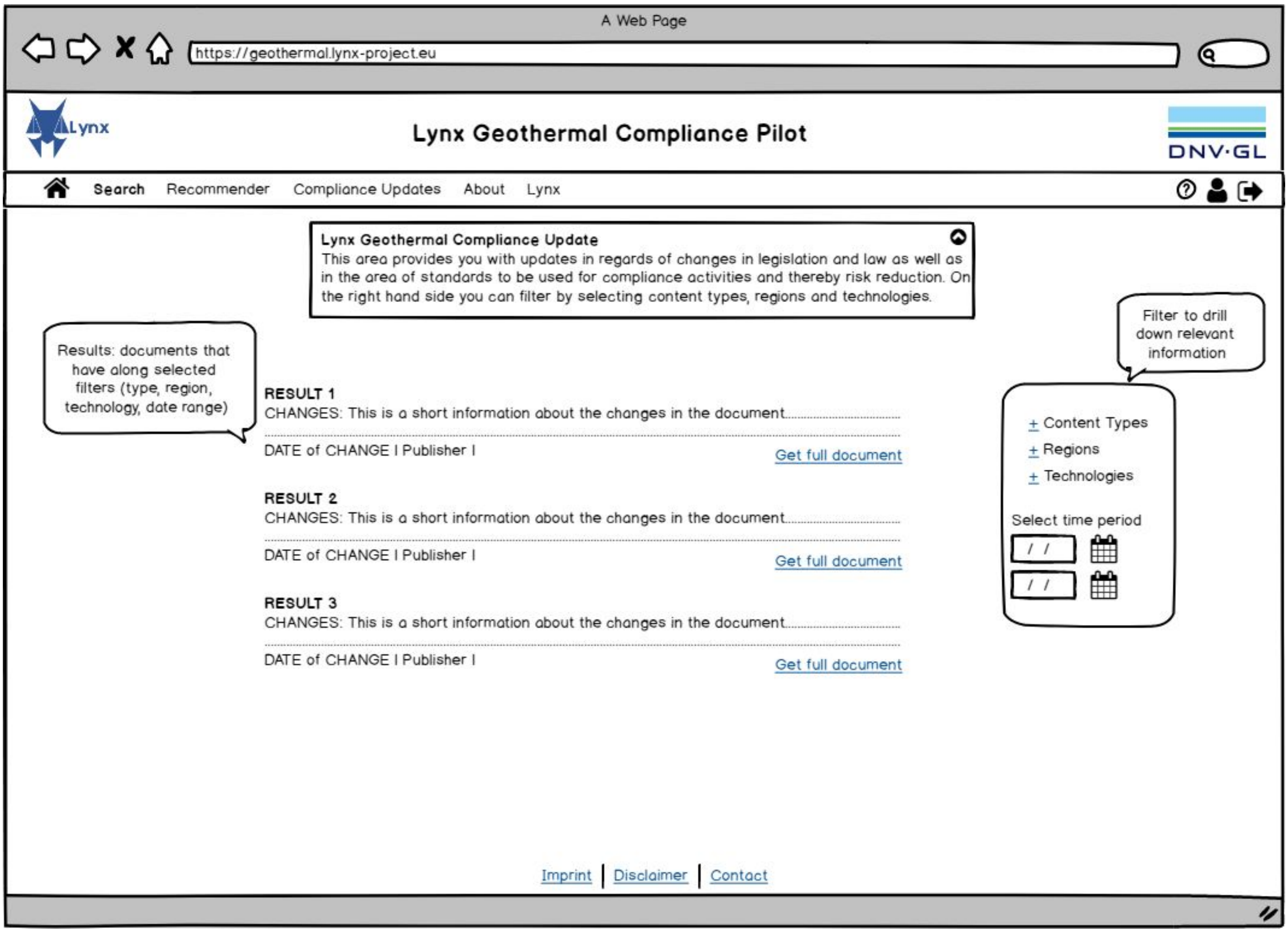

Figure 8 Mockup Compliance / update feature result page of GTE Pilot of Lynx

2.4 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 9), (ii) the upload function (Figure 10) and (iii) the recommender (Figure 11). The implementation of these areas has started as an alpha version is already in place (see Section 3).

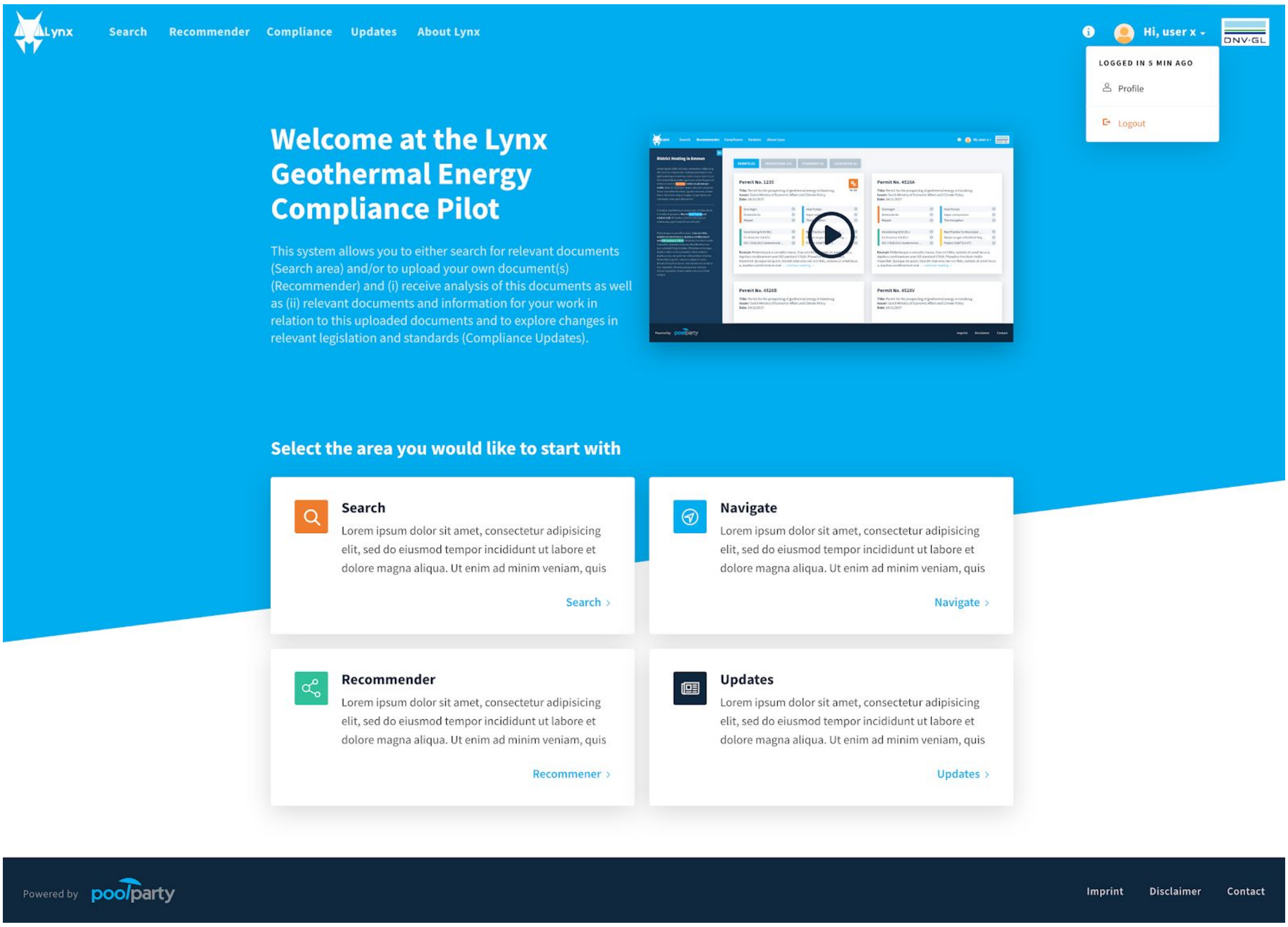

Figure 9 Design Prototype Start Screen GTE Pilot of Lynx

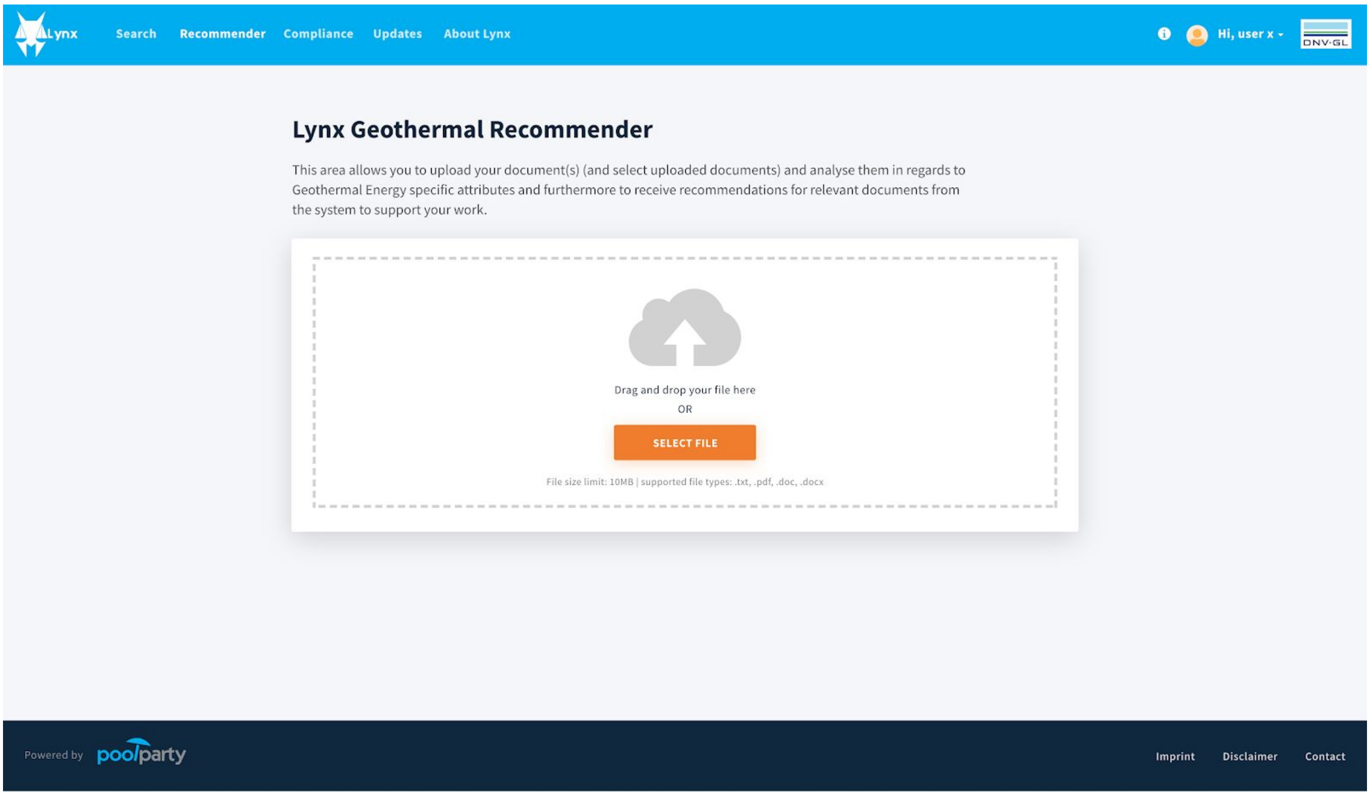

Figure 10 Design Prototype Recommender Upload Page of GTE Pilot of Lynx

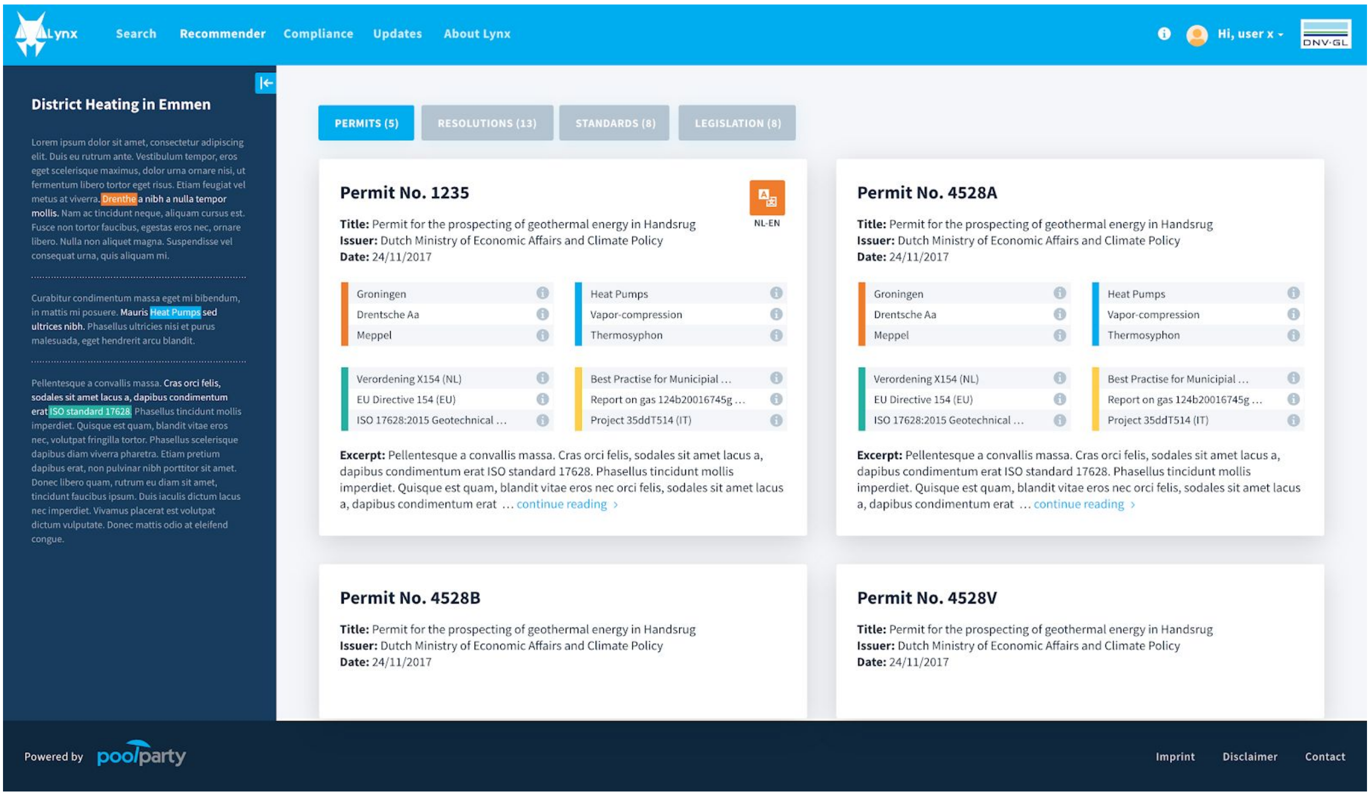

Figure 11 Screen Design Prototype Recommender Results Page of Lynx GTE Pilot

2.5 Business case(s) modalities

If the solution proves to be able to add value to the problem assumptions of the geothermal energy 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 geothermal energy. 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 geothermal energy. The following modalities and sources of exploitable value mechanisms are considered (Table 2).

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 geothermal energy value chain	Internally used: improves efficiency by reducing the time to find relevant information and be able to deliver more quality / value to the client. Externally used: subscription for customers as a self-service platform or licensed to Engineering / consultancy firms.
External information aggregator: with the current fragmented geothermal energy information landscape, there is room for an 'umbrella' platform aggregating the various information sources.	Potential for 2-sided platform model with the purpose of consummating matches among users and facilitating 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 catalysing the alignment of policies between individual EU countries, which is a priority on the agenda of the EC[8].

Table 2 Modalities and sources of exploitable value mechanisms

3 PILOT ARCHITECTURE & SYSTEM

3.1 Introduction

This section describes the architecture and services as well as the technologies used to implement the Lynx geothermal energy pilot system.

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 geothermal energy pilot system user interface can be separated into a client-side frontend and a backend application.

The backend runs on Node.js[9], which is an event-driven JavaScript runtime, and makes use of the Node-js-framework Express[10] in order to simplify basic tasks. Node.js applications can be run easily and are supported by most container application platforms like OpenShift[11] out of the box.

[8] Source: Regulation on the governance of the energy union and climate action https://ec.europa.eu/energy/topics/energy-strategy/energy-union_en#regulation-on-the-governance-of-the-energy-union-and-climate-action

[9] https://nodejs.org/en/
[10] https://expressjs.com/
[11] https://www.openshift.com/

The frontend is based on the JavaScript library and web-framework React, which encapsulates functionality into components. Redux[12] 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[13]. Additionally, Sass[14] 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, currently, the application only distinguishes between public and private use of the APIs. In case the application gets extended to a multi-user environment the use of MongoDB[15] has been contemplated in order to have a full MERN-stack (MongoDB + Express + React + Node) in place.

3.3 Lynx services

3.3.1 Document Ingestion

The document ingestion makes use of the workflow manager (WM) and in particular the population workflow that is described in D3.10, D4.4. The workflow is highly customizable, for the geothermal energy pilot the essential services in the workflow are Machine Translation (MT, described in D3.7) and Entity Linking (EL, described in D3.6). The MT service is used to output the document in the language of the user. The EL service makes use of the thesauri prepared by the terminology extraction process in order to extract the terms in the text.

[12] https://redux.js.org/
[13] https://getbootstrap.com/docs/4.0/getting-started/introduction/
[14] https://sass-lang.com/
[15] https://www.mongodb.com/

3.3.2 End user interaction

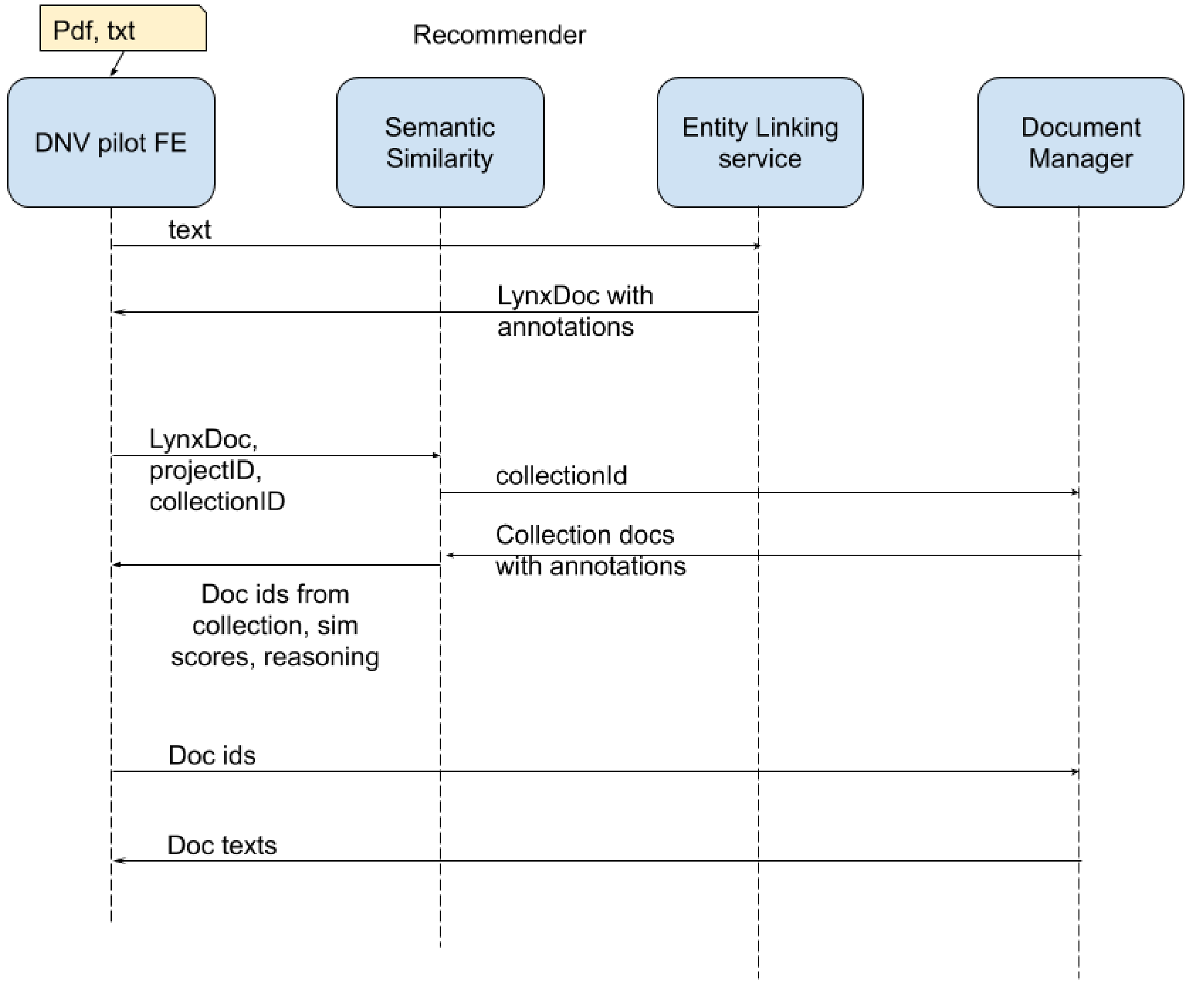

Figure 12 GTE recommender Lynx services interaction

Recommender. In the interaction with the end-user the first service that is used is the frontend application. The user uploads text input (.txt or .pdf) format that is sent to EL for annotating. The returned document with annotations is then sent to SeSim service (described in D 3.9). The SeSim service fetches the document from DCM and compares the input document with the fetched documents. Finally, the frontend fetches some additional information, including text of the documents, from DCM and displays these documents to the user.

Search. For the search functionality the frontend application forwards the user request to the Sear service (D3.9).

3.3.3 Summary of Lynx services

- **SEAR - cross-lingual search service**. The search service enables 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.
- **TermEx - terminology extraction.** The results of this service are used in SeSim to get the similarity between concepts.

- **EL - entity linking**. 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.
- **MT - machine translation**. The service provides the translations of the documents. Though the EL 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. Most documents used in this pilot are public documents, and the user input is not stored in the platform. However, for the sake of covering a broader range of use cases we foresee the option to create private document collections with restricted access.
- **WM - workflow manager**. The service is used to execute document ingestion.

Whereby SeSim, EntEx and TRANS services have been trained and customised with geothermal energy specific data and documents as well as models (in the form of taxonomies and terminologies) to ensure the good performance of the services for the Pilot System specific requirements with regard to languages and translations, and the recommender service.

3.4 Screenshots of the final pilot

Based on the mockups and screen design (section 2.3.1), the final pilot's UI is presented in this section covering the main pages: (i) start page (Figure 13), (ii) the upload function (Figure 14) and (iii) the recommender (Figure 15).

3.4.1 Pilot landing page

The pilot landing page (figure 13) introduces the services to the user. In contrast to the mockups, the "Lorem Ipsum" text is replaced with narrative to convey the use and benefit of the pilot to the user. The pilot will mainly focus on the recommender service.

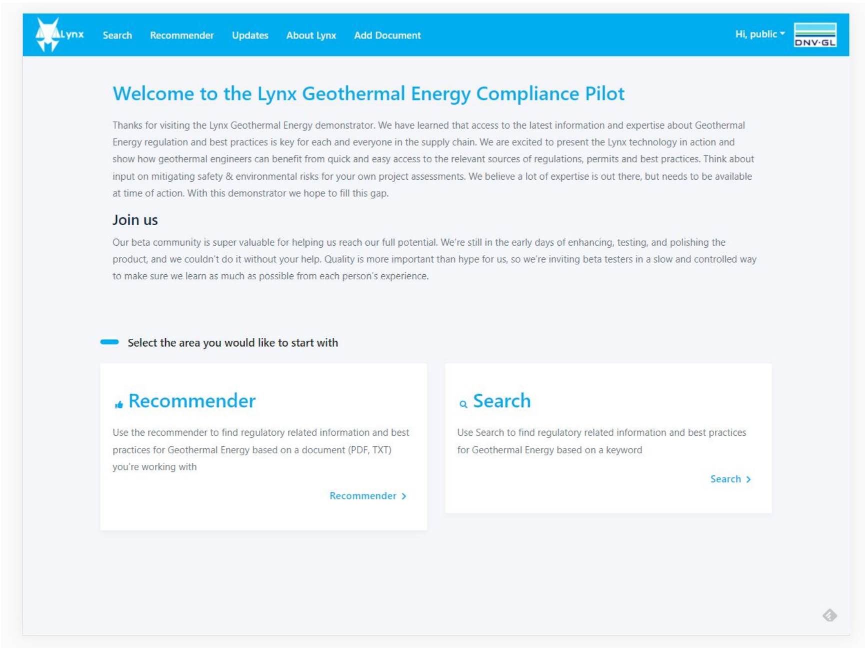

Figure 13 Design pilot Start Screen GTE Pilot of Lynx

3.4.2 Recommender upload page

The recommender upload page (figure 14) facilitates ‘drag and drop’ or uploading a file (in *.pdf of *.txt format) to the recommender. Users need to select the language of the uploaded document. Once uploaded, a ‘spinning wheel’ indicates the file is being processed. Shortly after this, the recommender result page will be displayed.

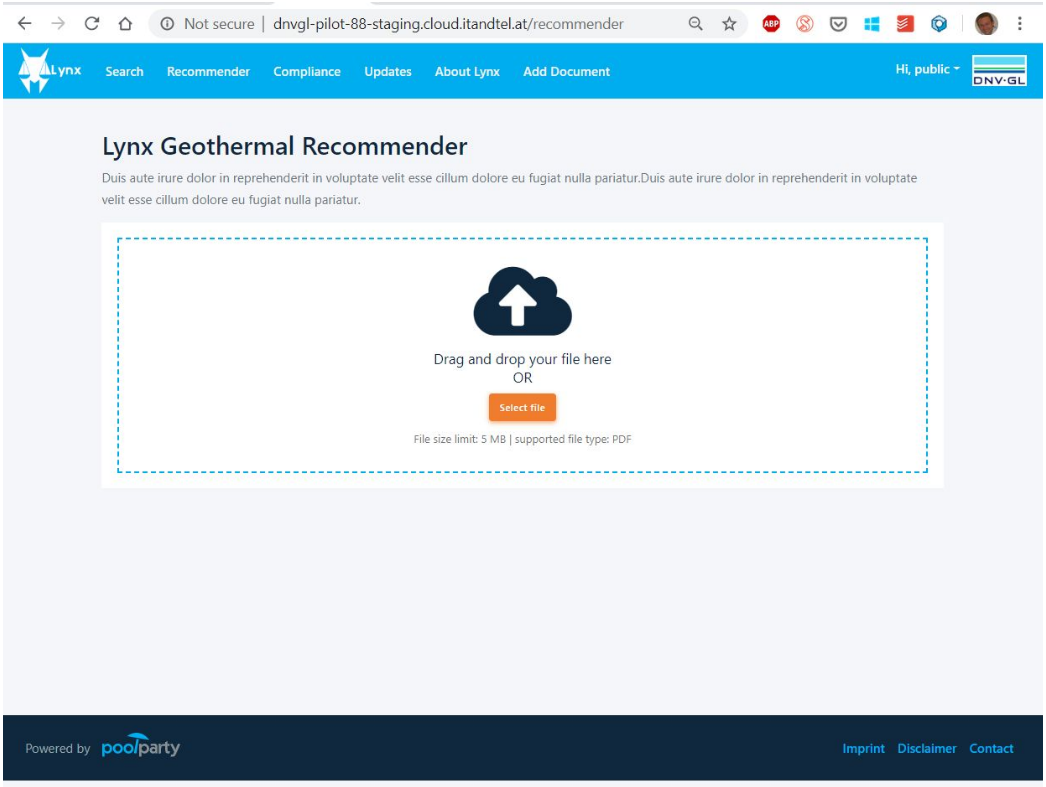

Figure 14 Design Recommender Upload Page of GTE Pilot of Lynx

3.4.3 Recommender result page

The recommender result page (figure 15) finally lists the recommended source based on the input document. This page consist of the following features:

A) A display box presenting the content from the uploaded document by the user with all the concepts (or words) known in the LKG, highlighted. The highlighting color indicated the category the concept belongs to (see the infobox “Color Legend ?”) at the right.

B) Recommender results are clustered in themes which can be selected by the user in the horizontal menu (Legislation, Best Practice, Permit, Standards and other.)

C) The recommender results are presented in ‘boxes’ including the following elements:

   a) Title of the document (in original language). It is envisioned that the title will link to the source link of the recommended document.
   b) Icon to indicate which languages were translated. The current screenshot shows that the original language, Dutch, has been translated into English.
   c) Detailed metadata about the document issuer and publication date.
   d) An infobox with an overview of terms similar to the document ingested by the user. By clicking on the info icon, the similarity score is presented (This is foreseen as a temporal feature to test the accuracy of the knowledge graph. For the envisioned end user it might not be relevant.)
   e) The last element in the recommender result box is a summary of the document for which the starting paragraphs of the document is translated by the Machine Translation service.

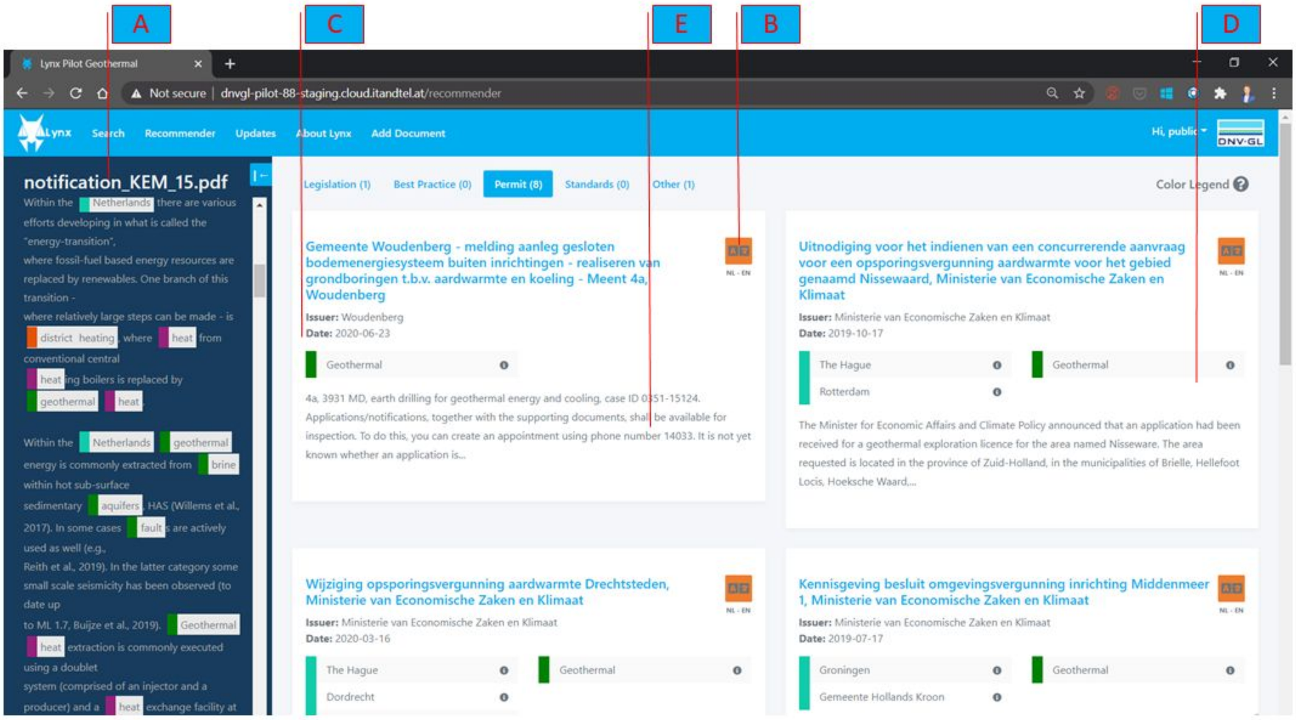

Figure 15 Design Recommender Results Page of Lynx GTE Pilot

3.4.4 Search page

The current release of the pilot also includes a search functionality, which presents relevant documents from the document manager based on a search keyword. As presented in Figure 16, the search result interface shows metadata and a short summarization of the document. The search result title (blue) includes a hyperlink which guides the user to the original document source (via hyperlink).

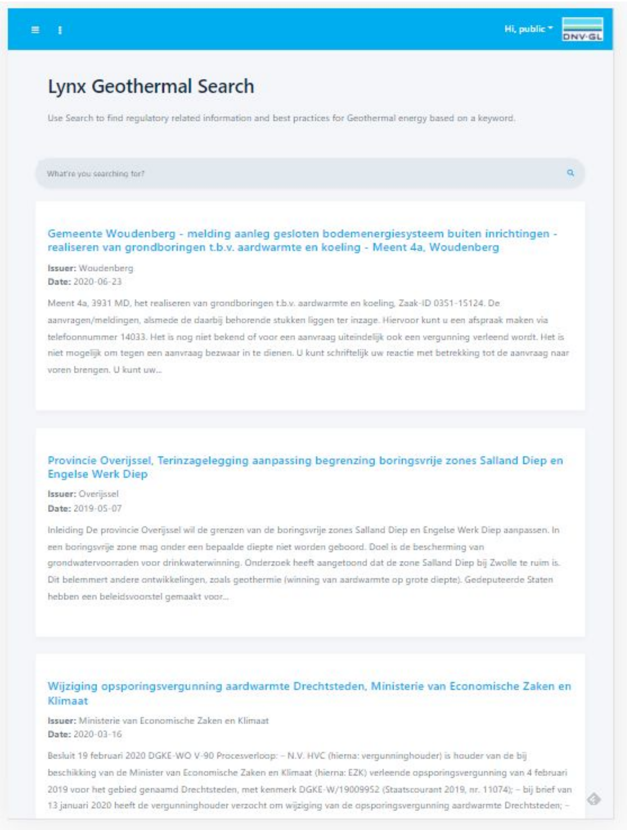

Figure 16 Design Search Page of Lynx GTE Pilot

3.5 Promotion and Validation

With the alpha version of the pilot in place, a number of initiatives took place to promote and validate the solution and encourage geothermal energy experts to think about the potential value of these emerging technologies for their industry.

The most memorable event was our collaboration with IGA (International Geothermal energy Association) which was triggered by an open tweet from the Head of IT, calling for digital innovation for the geothermal energy sector. After pitching the Lynx use case for geothermal energy, IGA was thrilled as they strongly believe bringing in new perspectives and non-traditional competencies (compared to drilling, seismic and petroleum engineers) has huge potential to accelerate innovation. With the support of IGA we developed a storyline with the title 'emerging technologies for standardisation' which was included in their 'stay engaged' campaign (figure 17), an 3-day online event (due to the corona pandemic) targeting geothermal professionals globally. The recording of the webinar (figure 18) is available on YouTube [link].

Figure 17 Agenda IGA Stay Engaged Campaign

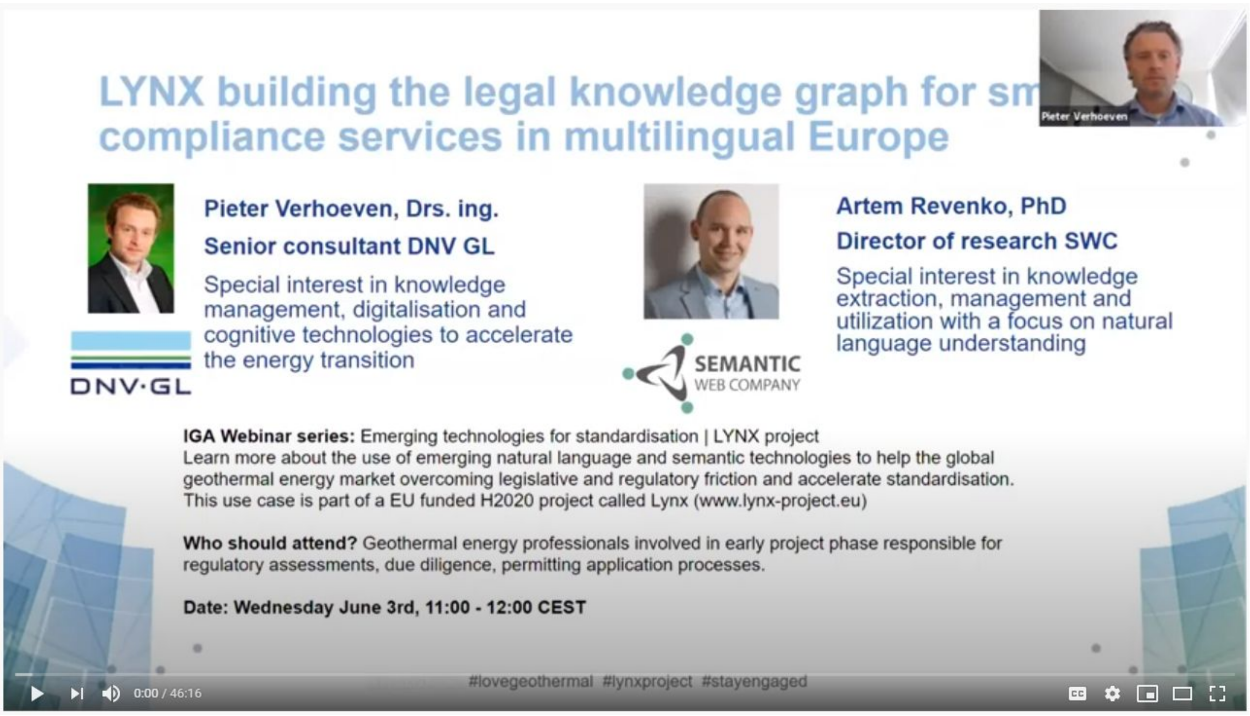

Figure 18 Lynx "emerging technologies for standardisation" webinar for IGA

As feedback on the collaboration with IGA the following testimonials where received:

- *"The Lynx project touches upon an important and emerging topic demonstrated through a very high engagement rate of our audience. The way how the geothermal sector deals with data and information will be key for actual geothermal development on the ground."* Gregor Rumberg, Technical Director International Geothermal Association
- *"Lynx is a great project, since it addresses the need for standardisation in geothermal and we are looking forward to our collaboration and to continue showcasing the added value for the sector."* Dr. Marit Brommer, Executive Director International Geothermal Association

4 REFLECTION AND RECOMMENDATIONS

During the test phase of the geothermal energy use case, several engagement activities were organised with stakeholders across the value chain to get feedback on the current features of the pilot. Based on these learning, we have looked for improvements for the current pilot as well as recommendations for future development. As mentioned in the executive summary, the geothermal energy use case was a proxy to learn how to use and understand the Lynx services and its use in a transformative context like the energy transition. In this chapter the reflections and recommendations are described.

Reflection 1: what makes recommended documents relevant?

This reflection was a key finding during end-user interaction: it was recognized the recommender is able to highlight certain concepts and translate into categorised recommended results. Although the recommender results were presented with the 'fingerprint' parameters (matching concepts in upload document and similarity scores), these metrics did not necessarily increase the end users' sense of relevancy. These fingerprint metrics are more parameters to assess the technical performance of the LKG and underpinning Lynx services.

Again, the question about relevancy appeared to be complicated. Pilot users suggest to use indicators like 1) whether or not it is in the law or enforced by regulation; 2) whether or not it is published by a trusted source; 3) whether or not it is peer reviewed by authoritative professionals; 4) whether or not it was published recently. We note that all these criteria are external to Lynx, yet need to be addressed.

A resolution of this point might come from handling the external sources differently, for example, adding source trust level of such. This should be taken into account during the early identification and assessing of external data sources: are there any external data sources which are recognized as a 'relevancy' indicator for the envisioned users?

Reflection 2: amount of documents to recommend

Users easily compare the recommender with their own experience (and habits) using search engines on the Internet (e.g google.com). Typically a search attempt starts by submitting one or a couple of search keywords and in most cases only the first search result page is used to look for anything relevant. If these results are not satisfying, more or other keywords are submitted, rather than clicking to the next search result pages.

With a 'search' user experience expectation in mind, e.g. expecting a narrow search result, the recommender user interface overwhelmed some users due to the (potential) large amount of categorised search results. This was not seen as useful, as the amount of search results could potentially grow exponentially and indicators to narrow down for relevancy (see reflection 1) were still limited.

Nevertheless, users do recognise the value of the recommender mechanisms to get a board picture needed during e.g. a kick off phase.

So far we have mostly focused on recall, i.e. finding all the relevant documents + noise. Tackling precision, i.e. all the returned results are relevant (no matter how many we missed) might be seen as an identified challenge and could be the next step - likely outside of Lynx

Reflection 3: how big is big data?

Compared to the other two Lynx use cases [D5.5 and D5.7], the amount of relevant corpuses readily available at the start of the project differed: in these use cases they took EU and country Law and thousands of contract-related documents, even semi structured, as a starting point, which directly kick started the value of the LKG. Given the early phase of geothermal energy, getting documents was complicated in most cases, let alone gathering large quantities of these documents. We've seen a large variety of modalities in EU countries, and regionally, the whereabouts of this sector is available in public sources. In Germany, for example, regional regulation relevant to geothermal energy is available, but not publicly online. Presumably working in an early phase domain might limit the power of knowledge graphs, as concepts are consolidated and the variety of practices is still highly unstructured.

As a recommendation, it would make sense to do a high level technical due diligence to test the feasibility of the knowledge graph technology given the maturity and availability of publicly accessible corpusses or relevant taxonomies or concept schemes.

Reflection 4: Who's problem should we solve first? Who is currently underserved?

The geothermal energy pilot was built with an envisioned end user in mind; starting with engineers involved in permitting processes where fragmented regulatory information resulting in friction covering a complete picture of the regulatory implications. Not only the envisioned efficiency gains, but also limiting opportunity costs are envisioned benefits of this use case. Although the problems regarding the permitting process were recognized by the industry, it also generated valuable insights in the complexity of the permitting process itself. One may conclude that, due to this complexity, it is yet not possible to isolate a specific problem, for which the LKG may be a solution which generates enough value (at scale) to make a healthy business case.

As a result we started to explore other use cases in the domain of geothermal energy, e.g. use cases that benefit from information regarding best practices related to safety and environmental risks. A platform to support geothermal energy operators to understand all (known) risks though the life cycle of a geothermal energy project, enriched with recommended standard, best practices or relevant research publications on measures to mitigate known risks. This would also be of interest for the state supervisor of Mines who is tasked to carry out measures to ensure certain standards to assure Safety and environmental performance in the industry. The Lynx geothermal energy pilot can be a platform for state supervisors to improve access to relevant information and expertise for a complete supply chain.

As a recommendation we would also suggest exploring financial analysts at banks or funds as well. Those who look into financing geothermal energy projects and need to understand the risk profile of an envisioned project as part of a due diligence process. Given the growing interest in 'green' or 'sustainable' loans or loans with a proven ESG footprint, this might be a segment currently underserved with relevant knowledge and expertise.

Reflection 5: Before pivoting, assess feasibility

We knew upfront that any of the wide array of sustainable energy alternatives could be subject to a use case for the Lynx services. With the learning we expect to increase the success rate of the Lynx use cases as we have better insights in which cases the technology adds value. So in case a new use case presents itself or we explore for other use cases ourselves; we would take the following aspects into consideration:

- Subject matter experts from both domains need mutual understanding of the problem - solution space. In the geothermal energy use case, for example, ample time was spent to introduce knowledge engineers and programmers in the field of geothermal energy, and the other way around.
- The characterization of the envisioned end-user needs to include the level of experience of knowledge such a person. Typically a solution will not target users who are new to the domain or who have years of experience.
- The ideal user is somewhere in the middle, and should be present in the market at scale, in order to suffice the investment required for a digital solution based on the Lynx
- Disadvantages of 'emerging' domains is the lack of standardisation, lack of 'common' vocabularies which can limit the power of the LKG and increases the effort needed to gather a good corpus of documents. Initially this can be quite unstructured of nature, and needs to contain information elements or terms that can be used for reasoning or analyzing a problem the end user might have. Again, also in an emerging domain there might be lack of consensus what the most pressing problem is and who would benefit if it is solved. So extensive user needs & behaviours research is recommended.

5 References

5.1 Lynx Deliverables

References to public Lynx deliverables are indicated with a 'D', e.g. 'D 3.6' throughout this document. Below the current public deliverables are listed as available on the Lynx website

D3.10 Intermediate extraction and linking services
D3.7 Translation Services
D3.6 Extraction and Linking Services
D3.9 Information Retrieval and Recommender Services
D4.4 Initial implementation and report of Data and Content Curation Services
D5.5 demonstrator for pilot 1
D5.7 demonstrator for pilot 3