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

D3.5 Initial Platform Prototype

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	Filippo Maganza (ALP)
CONTRIBUTORS	Julian Moreno-Schneider (DFKI), Sotiris Karampatakis (SWC)
REVIEWERS	Artem Revenko (SWC), María Navas Loro (UPM), Víctor Rodríguez Doncel (UPM)
VERSION | STATUS	V1.0 |Final
NATURE	Other
DISSEMINATION LEVEL	Public
DOCUMENT DOI	10.5281/zenodo.3558245
DATE	29/11/2019 (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)
0.1	ToC and introduction	02/11/2019	Filippo Maganza
0.5	Microservices-platform integration	08/11/2019	Julian Moreno Schnider, Filippo Maganza
0.8	JSON-LD	08/11/2019	Sotiris Karampatakis
1.0	Integration with pilots and conclusion	08/11/2019	Filippo Maganza

ACRONYMS LIST

HTTP	Hypertext Transfer Protocol
LKG	Legal Knowledge Graph
NIF	NLP Interchange Format
NLP	Natural Language Processing
RDF	Resource Description Framework
REST	Representational State Transfer
TCP	Transmission Control Protocol
URL	Uniform Resource Locator
WP	Work Package

EXECUTIVE SUMMARY

This Lynx deliverable describes the work carried out during T3.5 *Services/platform integration* and focuses on two different problems that we are currently dealing with:

- The integration of the microservices developed in WP3 with the basic platform implemented in WP1.
- The integration of the Lynx platform with the use case applications.

With regards to the first point, two important internal milestones have been achieved:

- The format for representing documents and annotations has been defined, this is the NLP Interchange Format (NIF). Using this format, we are now able to integrate enriched documents in the LKG.
- Most of the enrichment microservices are now able to process NIF, hence they can be easily plugged in and out of the workflows designed in WP4.

As for the second point, this document tackles the problem of long-running REST services, and suggests several strategies that can be adopted in such situations to develop a fault-tolerant system. These strategies will be used for the realisation of both the Contract Analysis and the Geothermal Project Analysis pilot use cases.

The Lynx platform prototype is currently deployed on a PaaS Openshift instance and is remotely accessible through the following URL: https://alp-api-88-staging-int.cloud.itandtel.at. Moreover, the services that compose the platform prototype can now be tested, their OpenAPI descriptions are available at this link: http://lynx-project.eu/doc/api/.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The requirements gathered in T1.1 *Industry requirements elicitation*, T1.2 *Technical requirements specification* and T4.1 *Pilot Use Cases and Requirements Study* served as input of the Lynx platform design process started in *T1.3 Technical architecture design and development*. The main milestone achieved was the definition of the Lynx architecture, which is based on the microservice architectural pattern and the REST architectural style.

The software architecture designed in T1.3 served as a blueprint for the implementation of the Lynx basic platform which is composed of four main logical components with well-defined responsibilities:

- Identity management and authorization service: which is responsible for the authorization of Lynx clients and users; moreover, it provides a way of managing their identities and roles.
- API Manager: which provides a single point of access to the Lynx platform by routing incoming requests to the right services; moreover, it is also responsible of authenticating the clients.
- Workflow Manager: which is responsible for the microservice orchestration for the execution of workflows. The detailed description of the inner architecture of this component is part of D4.4.
- Document Manager: is where the LKG is stored and maintained. The architecture of this component is explained in D1.4 *Setup and implementation of the basic platform* and in D4.4 *Initial implementation and report of Data and Content Curation Services*.

T3.5 *Services/platform integration* started in November 2018 and is still ongoing. Its main objectives are:

- The integration of the Lynx microservices developed in WP3 with the basic platform developed in WP1.
- The integration of the Lynx platform with the client applications.

This deliverable (D3.5, Initial platform prototype) consists of two parts:

- The platform prototype, which is currently deployed on a PaaS instance and accessible in a URL pointing to the Lynx API gateway[1].
- This document, where the work carried out in T3.5 is presented.

The functionality of the prototype can now be tested, the complete OpenAPI description is available at http://lynx-project.eu/doc/api/. Being a first prototype, some of its functionalities are subject to change.

1.1 PURPOSE AND STRUCTURE OF THIS DOCUMENT

This document mainly describes the work performed during T3.5. The first chapter, *Microservices-platform integration*, describes the general strategies adopted for the integration of Lynx microservices and the representation adopted for Lynx documents. The second chapter, *Integration with client applications*, explains for each pilot use case how we designed the interaction between the Lynx platform and its client applications. The third chapter, Citizen portal, describes the design of the Lynx citizen portal.

[1] https://alp-api-88-dev-int.cloud.itandtel.at

The last chapter, *Conclusion and future work,* closes this document and describes the next steps for the realization of the Lynx platform.

2 MICROSERVICES-PLATFORM INTEGRATION

In order to support the compliance-related workflow, the services delivered in WP3 should be integrated together with the Lynx Workflow Manager. Moreover, since the main goal of the Lynx project is the creation of the Legal Knowledge Graph, the documents and the annotations produced by the services should have an RDF representation. These two problems are strongly related to each other and, in order to simplify the integration with the platform, it has been established that all the enrichment services should support the same representation, whichi will be based on NLP Interchange Format 2.1 (NIF).

The first section of this chapter presents JSON-LD, the format that we are using to integrate the Lynx Natural Language Processing (NLP) services with other services and applications that are not able to process RDF. The second section describes NIF, its purpose and usefulness for Lynx, and some examples of its usage.

2.1 JSON-LD

Currently JSON (Wikipedia, 2019).is a dominant data transmission language across the Web, perhaps the most preferable across web developers (Cagle, 2017). They are familiar with JSON, and also with tools and frameworks that they use to create and manipulate data in JSON format. On the other hand, there is a steep learning curve for web developers to learn how to use RDF on their applications. Indexers, visualization frameworks, message brokers and other commonly used tools for web application development are usually based on JSON. But can the two worlds merge for the benefit of both?

The answer seems to be JSON-LD (W3C). JSON-LD is a W3C recommendation released in January 2014 and it is a method of serializing RDF using JSON. This allows data to be serialized in a way that is similar to traditional JSON RDF/JSON (W3C, 2013) format existed already, but there was still a gap between the design of the format and the requirements of a web developer. JSON-LD bridges this gap by using the very simple but powerful concept of context.

The context is simply a map between the actual properties defined in a JSON schema and the according properties in an ontology or vocabulary. Thus, a web developer can transform data in JSON into RDF just by defining the context. The context is then either included within the same JSON file, referred to as an external link, or even specified on transmission time by specifying the URL of the context in the header of the request.

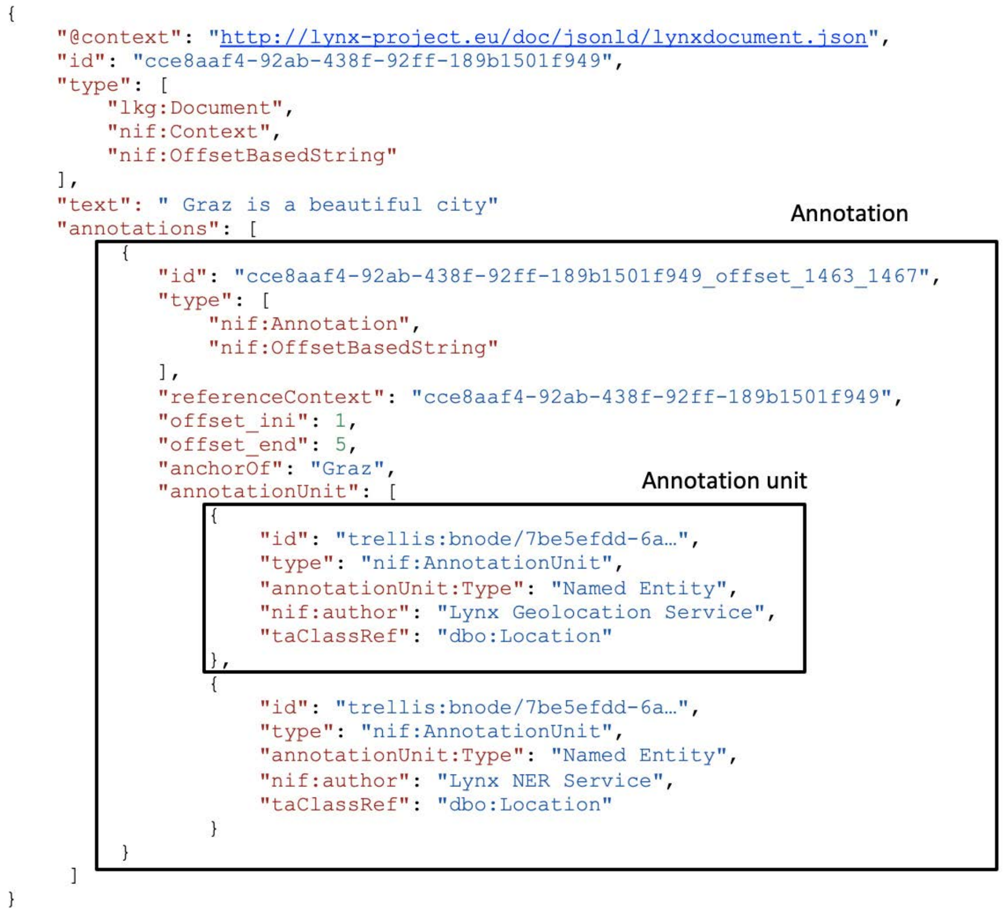
Using semantic web applications and traditional web applications at the same time is also a requirement of the Lynx platform. On the one hand, both indexers and frontend applications consume data in JSON. On the other hand, annotation services need to process the same information to produce annotations on the documents, and usually these services consume data in RDF format, specifically using the NIF vocabulary. One option is to use both formats, synchronize and provide mappings between them, a procedure that has to be very precise to be consistent. The other option is to use JSON-LD.

In the frame of the Lynx Project we decided to follow the second option, as to be able to store the document on the DCM and at the same time use the same file to index it into a Solr indexer. We developed a context based on the definition of the LynxDocument class as described in http://lynx-project.eu/data2/data-models. The current state of the Lynx specific schema can be found in *ANNEX 1: Context of the Lynx Json-ld*.

2.2 NLP INTERCHANGE FORMAT

The main goal of the Lynx project is the generation of the Legal Knowledge Graph. To make it possible, in T3.5 we designed an RDF representation for documents and annotations. The representation we are currently using is based on NIF 2.1, which comprises a collection of specifications and ontologies that are useful to model documents and annotations using RDF. NIF is designed to achieve interoperability between Natural Language Processing (NLP) services (Sebastian Hellmann), especially for enriching documents with annotations. The reference NIF ontology we are using for the project is available at https://github.com/NLP2RDF/ontologies/blob/7143b534d5f70f65c2d915680dfa56508f174214/nif-core/nif-core.ttl.

Figure 1 shows a NIF document serialized in JSON-LD. It describes the adopted RDF representation, as well as the naming conventions that we use to identify meaningful parts of the RDF graph that represents the enriched document.


Annotation
Annotation unit

```
{
    "@context": "http://lynx-project.eu/doc/jsonld/lynxdocument.json",
    "id": "cce8aaf4-92ab-438f-92ff-189b1501f949",
    "type": [
        "lkg:Document",
        "nif:Context",
        "nif:OffsetBasedString"
    ],
    "text": " Graz is a beautiful city"
    "annotations": [
        {
            "id": "cce8aaf4-92ab-438f-92ff-189b1501f949_offset_1463_1467",
            "type": [
                "nif:Annotation",
                "nif:OffsetBasedString"
            ],
            "referenceContext": "cce8aaf4-92ab-438f-92ff-189b1501f949",
            "offset_ini": 1,
            "offset_end": 5,
            "anchorOf": "Graz",
            "annotationUnit": [
                {
                    "id": "trellis:bnode/7be5efdd-6a…",
                    "type": "nif:AnnotationUnit",
                    "annotationUnit:Type": "Named Entity",
                    "nif:author": "Lynx Geolocation Service",
                    "taClassRef": "dbo:Location"
                },
                {
                    "id": "trellis:bnode/7be5efdd-6a…",
                    "type": "nif:AnnotationUnit",
                    "annotationUnit:Type": "Named Entity",
                    "nif:author": "Lynx NER Service",
                    "taClassRef": "dbo:Location"
                }
        }
    ]
}
```

Figure 1 A Lynx document represented in NIF and serialized in JSON-LD

Although the Lynx platform is using the basic NIF features to represent the documents, the already existing annotations in NIF are not enough for representing all enriching information in documents for the legal domain. Therefore, some additional annotations were defined and used.

Every Lynx NLP service should be able to process and return NIF documents, and also to enrich documents with specific information. For example, the Named Entity recognition service (NER) will include annotations for entities, the translation service (TRANS) will include a translated text annotation or the word sense induction, and the disambiguation service (WSID) will include additional information for already annotated term (entity-like) annotations. The service-specific annotations that we are using in the scope of Lynx project are described in Table 1.

In order to validate NIF and Lynx documents, we are defining several SHACL (Shapes Constraint Language) shapes. A SHACL shape can be viewed as a description of an RDF graph that does satisfy a certain set of conditions (W3C). Having defined a shape for Lynx documents, the integration of the services will be much easier, since we will know exactly whether the NIF output of a service is valid or not. Currently, Lynx shapes are still not ready, their design and implementation are some of the next steps of T3.5.

Annotation Property	Description	Microservice
itsrdf:taClassRef	Determines the type of the named entity: • Person: http://dbpedia.org/ontology/Person • Location: http://dbpedia.org/ontology/Location • Organization: http://dbpedia.org/ontology/Organization • Temporal entity: http://www.w3.org/2006/time#TemporalEntity	NER, GEO, EntEx, TimEx
itsrdf:taItemRef	Determines (several) external link(s) referring to the same entity. For example, a value could be *“https://www.wikidata.org/wiki/Q90”* for *“Paris”*.	NER, GEO, EntEx
itsrdf:target	Used by the Machine Translation service to include the translation of a text (document).	Trans
nif:normalizedDate	Used by the Temporal Expression Analysis service to include a normalized value for temporal expressions.	TimEx
lynxnif:summary	Used by the Summarization service to include a summary in the NIF document.	Summ
lynxnif:object lynxnif:subject	Object and subject of a relation between entities - should be nif:phrases	RelEx
lynxnif:relation	URI of the identified relation between two entities annotated in the relation extraction service	RelEx

Table 1 Description of the annotation properties used to represent Lynx documents

3 INTEGRATION WITH CLIENT APPLICATIONS

This section describes how the Lynx platform should be integrated with the Openlaws, Cuatrecasas and DNV GL applications to build the pilot use cases. It focuses on the specification of the integration strategies used to bring together the Lynx platform and its client applications.

3.1 CONTRACT ANALYSIS AND GEOTHERMAL PROJECT ANALYSIS

The Contract Analysis and the Geothermal Project Analysis workflows were specified in D4.3. They both require converting a JSON document to RDF, enriching it and searching for relevant documents. This process typically requires a long time (even more than 5 minutes for long documents), therefore an integration based on a single synchronous HTTP POST request presents several problems:

- The client cannot query for the status of the process.
- The client thread is blocked until the results are not returned.
- The TCP connection used by HTTP must be kept alive even when it is not being used for transferring data, hence the time in which the application can be affected by network faults is substantially higher than what is the actual time needed to transfer the data between the client and the server.

In such cases, there are different standard patterns that can be used to design a more robust integration based on REST (RESTalk Patterns). Considering the pilot and the technical requirements, we have decided to offer two possibilities:

1. Webhook (see Figure 2):
    1. The Lynx consortium and the client agree on a callback URL.
    2. The client sends a request to Lynx, if the request is malformed, the server responds with an error, otherwise it responds with a 202 Accepted code and schedules the processing of the client request.
    3. When the result is ready, Lynx sends it back by using the callback URL provided by the client in the initial request.
2. Polling (see Figure 3):
    1. The client sends a request to Lynx, which decides whether to accept or reject it.
    2. If the request has been accepted, the client can start to periodically ask the server whether the result is ready or not.
    3. When clients know that the result is ready, they can download it from the server.

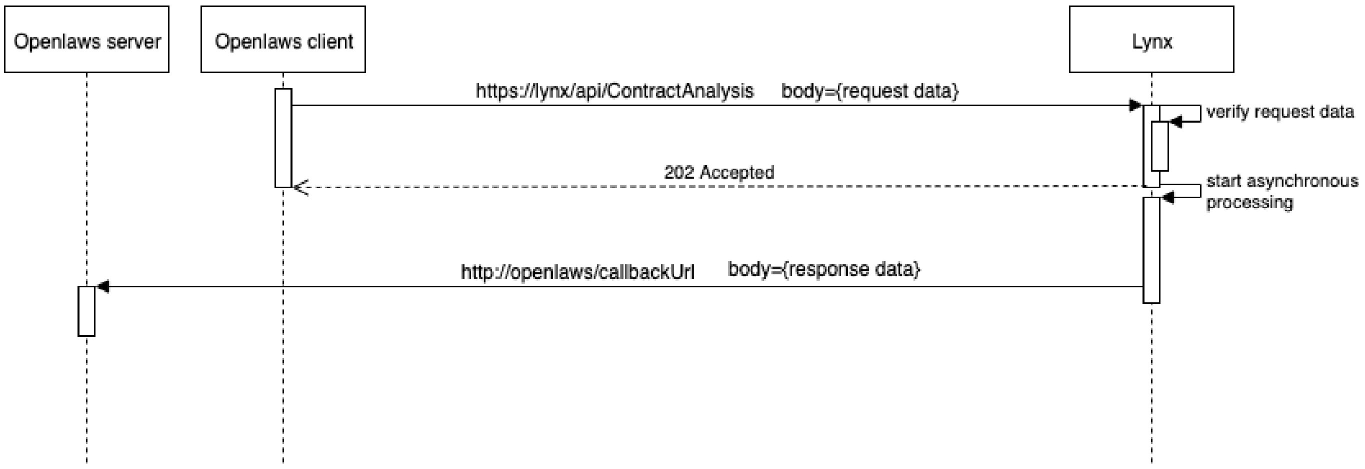

Openlaws server
Openlaws client
Lynx
https://lynx/api/ContractAnalysis body={request data}
verify request data
202 Accepted
start asynchronous processing
http://openlaws/callbackUrl body={response data}

Figure 2 Asynchronous request processing using a webhook

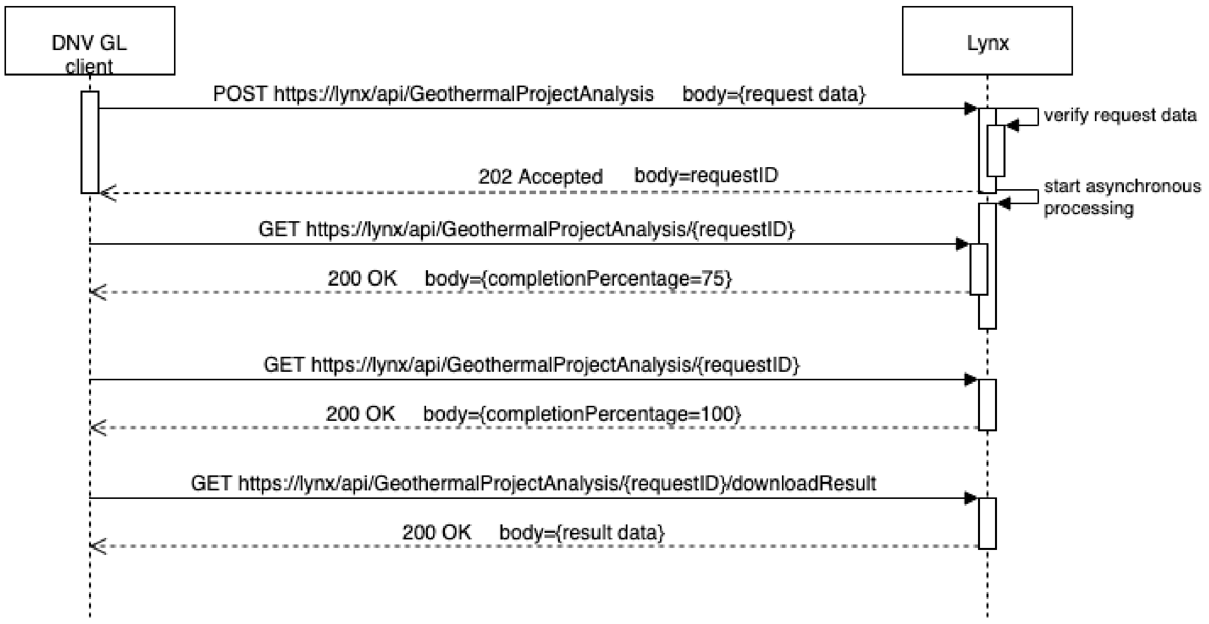

DNV GL client
Lynx
POST https://lynx/api/GeothermalProjectAnalysis body={request data}
verify request data
202 Accepted body=requestID
start asynchronous processing
GET https://lynx/api/GeothermalProjectAnalysis/{requestID}
200 OK body={completionPercentage=75}
GET https://lynx/api/GeothermalProjectAnalysis/{requestID}
200 OK body={completionPercentage=100}
GET https://lynx/api/GeothermalProjectAnalysis/{requestID}/downloadResult
200 OK body={result data}

Figure 3 Asynchronous request processing using short polling

3.2 QUESTION ANSWERING AND FREE-TEXT SEARCH

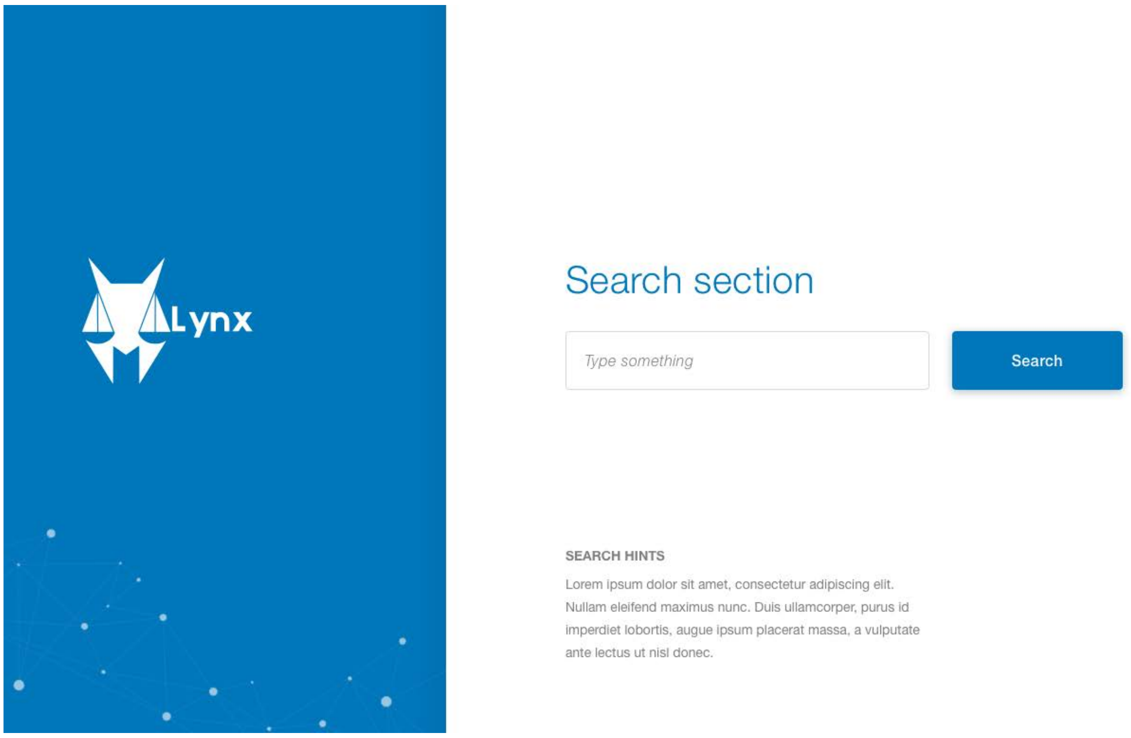
For question answering and free-text search, asynchronous request processing is not needed because Lynx is able to respond in a much shorter time. Therefore, in these two cases, the integration is simply based on a single HTTP POST request.

4 CITIZENS PORTAL

The Lynx project aims to provide European citizens a better access to legal and regulatory information from multiple jurisdictions. The Lynx citizen portal is the web application that we are developing to achieve this objective. The application is structured in three screens:

- Splash screen (shown in Figure 4). It features the Lynx logo and a form for free-text queries. Its main purpose is presenting to the user the Lynx logo and some information regarding the Lynx search functionalities. When a search is performed the *document search screen* is loaded.
- Document search screen (depicted in Figure 5). It features a form for free-text queries, search filters and search results with excerpt and tags. In this screen the users can visualize the results of their search grouped into pages and select the document they are interested in. Once a document is selected, the *document view screen* is loaded.
- Document view screen (see two mockups of this screen in Figure 6 and Figure 7). This screen has three main components: The document corpus component, the left sidebar and the right sidebar. The document corpus component displays the text of the document and its annotations, the latter have different colour depending of their provenance (ex. NER or TimEx). The right sidebar contains the metadata of the document. The left sidebar features the controls to toggle annotations' rendering, changing the view mode or the language. Moreover, in the top part of this screen, there is a free text query form (it is missing in the mockups). Once the user enters input a query, the *document search screen* is loaded.

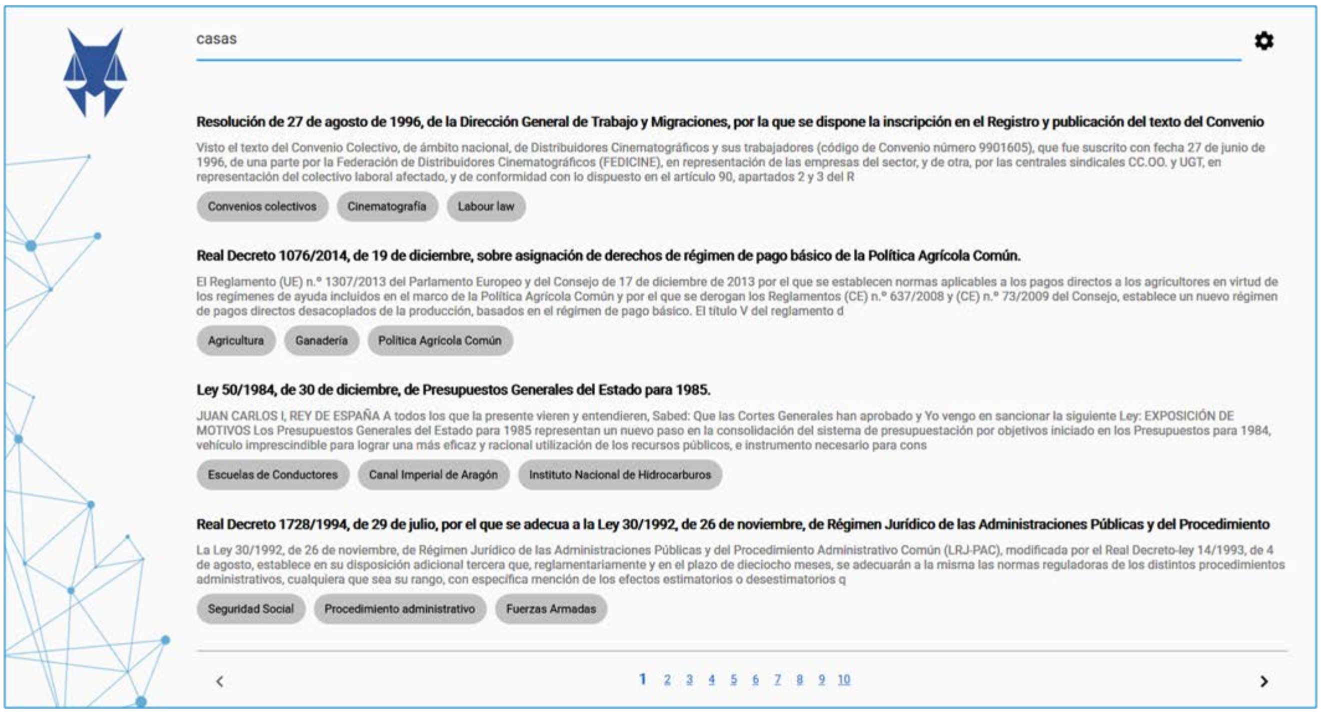
The citizen portal design started in November 2018 while its implementation started in May 2019. The application is not currently available for testing at the moment.



SEARCH HINTS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam eleifend maximus nunc. Duis ullamcorper, purus id imperdiet lobortis, augue ipsum placerat massa, a vulputate ante lectus ut nisl donec.

Figure 4 Mockup of the citizen portal's "splash screen"


casas

Resolución de 27 de agosto de 1996, de la Dirección General de Trabajo y Migraciones, por la que se dispone la inscripción en el Registro y publicación del texto del Convenio

Visto el texto del Convenio Colectivo, de ámbito nacional, de Distribuidores Cinematográficos y sus trabajadores (código de Convenio número 9901605), que fue suscrito con fecha 27 de junio de 1996, de una parte por la Federación de Distribuidores Cinematográficos (FEDICINE), en representación de las empresas del sector, y de otra, por las centrales sindicales CC.OO. y UGT, en representación del colectivo laboral afectado, y de conformidad con lo dispuesto en el artículo 90, apartados 2 y 3 del R

Convenios colectivos | Cinematografía | Labour law

Real Decreto 1076/2014, de 19 de diciembre, sobre asignación de derechos de régimen de pago básico de la Política Agrícola Común.

El Reglamento (UE) n.º 1307/2013 del Parlamento Europeo y del Consejo de 17 de diciembre de 2013 por el que se establecen normas aplicables a los pagos directos a los agricultores en virtud de los regímenes de ayuda incluidos en el marco de la Política Agrícola Común y por el que se derogan los Reglamentos (CE) n.º 637/2008 y (CE) n.º 73/2009 del Consejo, establece un nuevo régimen de pagos directos desacoplados de la producción, basados en el régimen de pago básico. El título V del reglamento d

Agricultura | Ganadería | Política Agrícola Común

Ley 50/1984, de 30 de diciembre, de Presupuestos Generales del Estado para 1985.

JUAN CARLOS I, REY DE ESPAÑA A todos los que la presente vieren y entendieren, Sabed: Que las Cortes Generales han aprobado y Yo vengo en sancionar la siguiente Ley: EXPOSICIÓN DE MOTIVOS Los Presupuestos Generales del Estado para 1985 representan un nuevo paso en la consolidación del sistema de presupuestación por objetivos iniciado en los Presupuestos para 1984, vehículo imprescindible para lograr una más eficaz y racional utilización de los recursos públicos, e instrumento necesario para cons

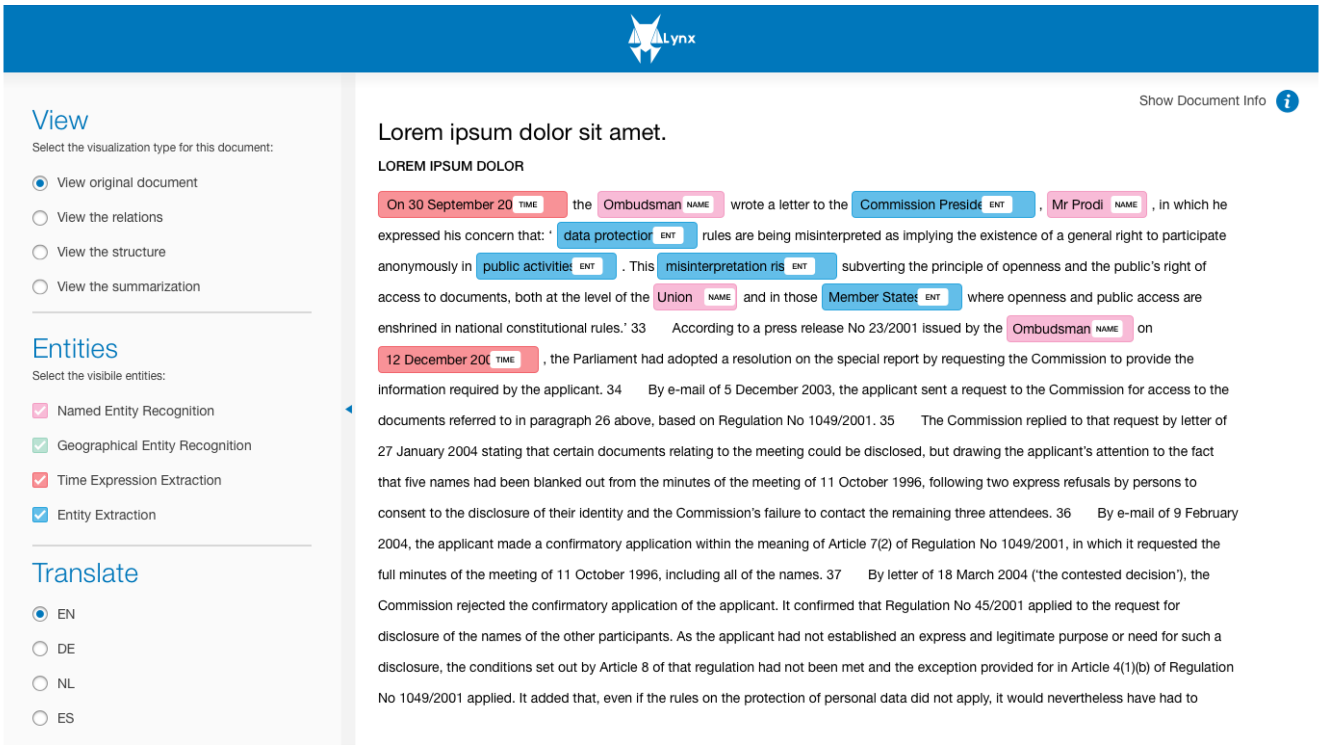
Escuelas de Conductores | Canal Imperial de Aragón | Instituto Nacional de Hidrocarburos

Real Decreto 1728/1994, de 29 de julio, por el que se adecua a la Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento

La Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común (LRJ-PAC), modificada por el Real Decreto-ley 14/1993, de 4 de agosto, establece en su disposición adicional tercera que, reglamentariamente y en el plazo de dieciocho meses, se adecuarán a la misma las normas reguladoras de los distintos procedimientos administrativos, cualquiera que sea su rango, con específica mención de los efectos estimatorios o desestimatorios q

Seguridad Social | Procedimiento administrativo | Fuerzas Armadas

Figure 5 A screenshot of the "document search screen" of the citizen portal first prototype



View

Select the visualization type for this document:

- [✓ View original document]
- View the relations
- View the structure
- View the summarization

Entities

Select the visible entities:

- [✓ Named Entity Recognition]
- [✓ Geographical Entity Recognition]
- [✓ Time Expression Extraction]
- [✓ Entity Extraction]

Translate

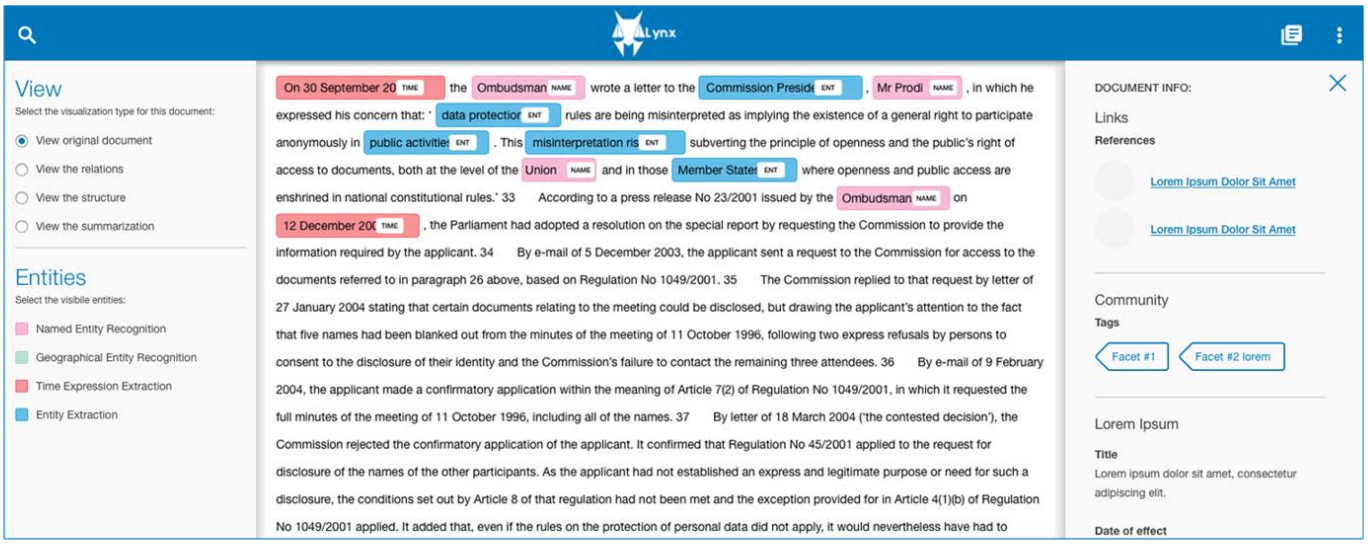
- [✓ EN]
- DE
- NL
- ES

Lorem ipsum dolor sit amet.

LOREM IPSUM DOLOR

On 30 September 20 TIME the Ombudsman NAME wrote a letter to the Commission Preside ENT , Mr Prodi NAME , in which he expressed his concern that: ' data protectior ENT rules are being misinterpreted as implying the existence of a general right to participate anonymously in public activitie ENT . This misinterpretation ris ENT subverting the principle of openness and the public's right of access to documents, both at the level of the Union NAME and in those Member States ENT where openness and public access are enshrined in national constitutional rules.' 33 According to a press release No 23/2001 issued by the Ombudsman NAME on 12 December 20(TIME , the Parliament had adopted a resolution on the special report by requesting the Commission to provide the information required by the applicant. 34 By e-mail of 5 December 2003, the applicant sent a request to the Commission for access to the documents referred to in paragraph 26 above, based on Regulation No 1049/2001. 35 The Commission replied to that request by letter of 27 January 2004 stating that certain documents relating to the meeting could be disclosed, but drawing the applicant's attention to the fact that five names had been blanked out from the minutes of the meeting of 11 October 1996, following two express refusals by persons to consent to the disclosure of their identity and the Commission's failure to contact the remaining three attendees. 36 By e-mail of 9 February 2004, the applicant made a confirmatory application within the meaning of Article 7(2) of Regulation No 1049/2001, in which it requested the full minutes of the meeting of 11 October 1996, including all of the names. 37 By letter of 18 March 2004 ('the contested decision'), the Commission rejected the confirmatory application of the applicant. It confirmed that Regulation No 45/2001 applied to the request for disclosure of the names of the other participants. As the applicant had not established an express and legitimate purpose or need for such a disclosure, the conditions set out by Article 8 of that regulation had not been met and the exception provided for in Article 4(1)(b) of Regulation No 1049/2001 applied. It added that, even if the rules on the protection of personal data did not apply, it would nevertheless have had to

Figure 6 Mockup of the citizen portal's "document view screen" with the left sidebar activated



View

Select the visualization type for this document:

- [✓ View original document]
- View the relations
- View the structure
- View the summarization

Entities

Select the visibile entities:

- 🟪 Named Entity Recognition
- 🟩 Geographical Entity Recognition
- 🟥 Time Expression Extraction
- 🟦 Entity Extraction

On 30 September 20 TIME the Ombudsman NAME wrote a letter to the Commission Preside ENT , Mr Prodi NAME , in which he expressed his concern that: ' data protectior ENT rules are being misinterpreted as implying the existence of a general right to participate anonymously in public activitie: ENT . This misinterpretation ris ENT subverting the principle of openness and the public's right of access to documents, both at the level of the Union NAME and in those Member States ENT where openness and public access are enshrined in national constitutional rules.' 33 According to a press release No 23/2001 issued by the Ombudsman NAME on 12 December 20(TIME , the Parliament had adopted a resolution on the special report by requesting the Commission to provide the information required by the applicant. 34 By e-mail of 5 December 2003, the applicant sent a request to the Commission for access to the documents referred to in paragraph 26 above, based on Regulation No 1049/2001. 35 The Commission replied to that request by letter of 27 January 2004 stating that certain documents relating to the meeting could be disclosed, but drawing the applicant's attention to the fact that five names had been blanked out from the minutes of the meeting of 11 October 1996, following two express refusals by persons to consent to the disclosure of their identity and the Commission's failure to contact the remaining three attendees. 36 By e-mail of 9 February 2004, the applicant made a confirmatory application within the meaning of Article 7(2) of Regulation No 1049/2001, in which it requested the full minutes of the meeting of 11 October 1996, including all of the names. 37 By letter of 18 March 2004 ('the contested decision'), the Commission rejected the confirmatory application of the applicant. It confirmed that Regulation No 45/2001 applied to the request for disclosure of the names of the other participants. As the applicant had not established an express and legitimate purpose or need for such a disclosure, the conditions set out by Article 8 of that regulation had not been met and the exception provided for in Article 4(1)(b) of Regulation No 1049/2001 applied. It added that, even if the rules on the protection of personal data did not apply, it would nevertheless have had to

DOCUMENT INFO:

Links

References

Lorem Ipsum Dolor Sit Amet

Lorem Ipsum Dolor Sit Amet

Community

Tags

Facet #1 | Facet #2 lorem

Lorem Ipsum

Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Date of effect

Figure 7 Mockup of the citizen portal's "document view screen" with both the left and right sidebar activated

5 CONCLUSIONS AND FUTURE WORK

In the first part of this document we have explained how we represent Lynx documents and how the Lynx microservices are integrated together using NIF and JSON-LD. In the second part, we have described how the Lynx platform will interact with the client applications to realize the pilot use cases. The design of the integration strategies and their implementation constitute the main work carried out during T3.5.

Task 3.5 has started in November 2018 and is expected to finish in August 2020. Now, most of the Lynx enrichment microservices are able to process NIF, therefore, they can be easily plugged in and out of the workflows. The population workflow has been implemented and can be tested -the complete guidelines are presented in D4.4.

The next steps to be carried out during T3.5 are:

- Improvement of the Java library that we are using for the serialization, manipulation and validation of Lynx documents and annotations.
- Integration of Search, SeSim and RelEx with the Workflow Manager.
- Definition of SHACL shapes for validating NIF and Lynx documents.

Other envisioned steps important for the implementation of the Lynx platform are:

- Conversion of documents from PDF to JSON.
- Population of the LKG.
- Realization of the Contract Analysis and the Geothermal Project Analysis workflows.
- Development of the Lynx citizen portal.

6 REFERENCES

Cagle, K. (2017, 09 17). *JSON-LD rewrites the Semantic Web.* Retrieved from linkedin: https://www.linkedin.com/pulse/json-ld-rewrites-semantic-web-kurt-cagle/

RESTalk Patterns. (n.d.). *Long Running Operation with Polling.* Retrieved from http://restalk-patterns.org/long-running-operation-polling.html

Sebastian Hellmann, M. B. (n.d.). Retrieved from https://nif.readthedocs.io/en/latest/#

RDF 1.1 JSON Alternate Serialization (RDF/JSON). W3C (2013, 11 07). Retrieved from https://www.w3.org/TR/rdf-json/

JSON-LD 1.1. W3C (n.d.). Retrieved from https://json-ld.org/spec/latest/json-ld/

Shapes Constraint Language (SHACL). W3C (n.d.). Retrieved from https://www.w3.org/TR/shacl/

Wikipedia. (2019, 07 22). *JSON-LD.* Retrieved from Wikipedia: https://en.wikipedia.org/wiki/JSON-LD

ANNEX 1: CONTEXT OF THE LYNX JSON-LD

An up-to-date version of the context can be found online[2].

```
{
 "@context" : {
   "@base": "http://lkg.lynx-project.eu/res/",
   "nif": "http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#",
   "itsrdf": "http://www.w3.org/2005/11/its/rdf#",
   "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
   "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
   "owl" : "http://www.w3.org/2002/07/owl#",
   "skos": "http://www.w3.org/2004/02/skos/core#",
   "dct": "http://purl.org/dc/terms/",
   "dbo": "http://dbpedia.org/ontology/",
   "dbr": "http://dbpedia.org/resource/",
   "lkg": "http://lkg.lynx-project.eu/def/",
   "xsd": "http://www.w3.org/2001/XMLSchema#",
   "eli": "http://data.europa.eu/eli/ontology#",
   "foaf": "http://xmlns.com/foaf/0.1/",
   "id" : "@id",
   "type": "@type",
   "sameAs": "owl:sameAs",
   "text": "nif:isString",
   "translations" : {
     "@id" : "itsrdf:target",
     "@container" : "@language"
   },
   "Concept": "skos:Concept",
   "ConceptScheme": "skos:ConceptScheme",
   "prefLabel": "skos:prefLabel",
   "altLabel": "skos:altLabel",
   "notation": "skos:notation",
   "definition": "skos:definition",
   "broader": "skos:broader",
   "narrower": "skos:narrower",
   "inScheme": "skos:inScheme",
   "hasTopConcept": "skos:hasTopConcept",
   "topConceptOf": "skos:topConceptOf",
   "parts": {
     "@id": "eli:has_part",
     "@container": "@set"
   },
   "offset_ini": {
     "@id": "nif:beginIndex",
     "@type": "xsd:integer"
   },
   "offset_end": {
     "@id": "nif:endIndex",
     "@type": "xsd:integer"
   },
   "parent": {
     "@id": "lkg:parent",
     "@type": "@id"
   },
   "annotations": {
     "@reverse": "nif:referenceContext"
   },
   "metadata": {
     "@id": "lkg:metadata"
   },
   "subject": {
     "@id": "dct:subject",
     "@container": "@set"
   },
   "comment": {
```

[2] http://lynx-project.eu/doc/jsonld/lynxdocument.json

```
    "@id": "rdfs:comment"
  },
  "source": {
    "@id": "dct:source"
  },
  "title": {
    "@id": "dct:title"
  },
  "nick": {
    "@id": "foaf:nick",
    "@container": "@set"
  },
  "first_date_entry_in_force": {
    "@id": "eli:first_date_entry_in_force",
    "@container": "@set"
  },
  "version_date": {
    "@id": "eli:version_date",
    "@container": "@set"
  },
  "version": {
    "@id": "eli:version",
    "@container": "@set"
  },
  "hasAuthority": {
    "@id": "lkg:hasAuthority",
    "@container": "@set"
  },
  "jurisdiction": {
    "@id": "eli:jurisdiction",
    "@container": "@set"
  },
  "language": {
    "@id": "dct:language",
    "@container": "@set"
  },
  "type_document": {
    "@id": "eli:type_document",
    "@container": "@set"
  },
  "links": {
    "@id": "rdfs:seeAlso",
    "@type": "@id",
    "@container": "@set"
  },
  "uri": {
    "@id": "dct:uri",
    "@type": "@id"
  },
  "taClassRef": {
    "@id": "itsrdf:taClassRef",
    "@type": "@id"
  },
  "taIdentRef": {
    "@id": "itsrdf:taIdentRef",
    "@type": "@id"
  },
  "referenceContext": {
    "@id": "nif:referenceContext",
    "@type": "@id"
  },
  "taConfidence": {
    "@id": "itsrdf:taConfidence",
    "@type": "xsd:decimal"
  },
  "taClassConf" : {
    "@id" : "nif:taClassConf",
    "@type": "xsd:decimal"
  },
  "anchorOf": {
        "@id": "nif:anchorOf"
      },
      "annotationUnit": {
        "@id" : "nif:annotationUnit",
        "@container": "@set"
      }
    },
    "@type": "nif:Context",
    "has_parts":{
      "@type" :"nif:Section"
    },
    "annotations" : {
      "@type" : "nif:String",
      "annotationUnit": {
        "@type" : "nif:AnnotationUnit",
        "@embed" : "@always"
      }
    },
    "metadata":{
    }
  }
```