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

D2.7 Catalogue of relevant legal and regulatory datasets

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
01	ToC and review of related literature	29/01/2020	Pompeu Casanovas (UAB)
03	1st version of the table	04/02/2020	Socorro Bernardos (UPM), Elena Montiel-Ponsoda (UPM)
04	2nd version of the table	13/02/2020	Socorro Bernardos (UPM), Patricia Martín-Chozas (UPM)
05	Review of content for some jurisdictions	25/02/2020	Artūrs Vasiļevskis (TILDE), Christian Sageder (openlaws)
06	General review	26/02/2020	Christian Sageder (openlaws), Pieter Verhoeven (DNV.GL)
1.0	Final review	28/02/2020	Elena Montiel-Ponsoda

ACRONYMS LIST

ELI: European Legislation Identifier

ECLI: European Case Law Identifier

LKG: Legal Knowledge Graph

EXECUTIVE SUMMARY

In this document we summarise the work carried out in WP2 to offer an overview of the publication status of legislation and case law in Europe, with regard to the implementation of ELI/ECLI, accessing criteria, representation formats and accessibility options.

The reviewed datasets are now part of the C-KAN Lynx Data Portal and have been used as starting point for the creation of the Lynx Legal Knowledge Graph.

TABLE OF CONTENTS

1 INTRODUCTION

This catalogue is the result of an exhaustive search for available legal and regulatory resources on the web from different jurisdictions across the European Union. We have specifically analysed the legal information sources provided by the 27 countries that conform the European Union (as of February 2020), plus the European Union itself and the United Kingdom.

For this analysis, we have considered several criteria of interest for the Lynx project, namely, type of documents available (law and/or case law), level of implementation of the ELI (European Resource Identifier) standard, formats in which the legislation is published, language of the data sources, or accessibility options (access interface, possibility of a bulk download).

This task is part of the WP2 that focuses in the data management, this is the collection and generation of data sources required for the Legal Knowledge Graph. Works related to legal language resources and corpora have already been reported in previous deliverables (D2.2, D2.3, D2.4 and D2.5).

2 LEGAL DATASETS IN THE EUROPEAN UNION

2.1 BACKGROUND AND MOTIVATION

This review of legal data sources in Europe takes as starting point several traditional sources of legal information published by various institutions of the European Union, namely:

1. *Access to Legislation in Europe*[1] report (from 2009)
2. *General Official Journals EU Forum*[2] (2004-2019) of the European Forum of Official Gazettes
3. General Information on All Official Gazettes[3] of the European Forum of Official Gazettes
4. European e-Justice Portal[4]

The latter aims to be a "one-stop shop in the area of justice" to facilitate access to information around justice systems and justice itself throughout the EU.

All mentioned sources provide highly valuable information, and the present report by no mean aims at replacing those, but complementing the information available in them with some updated links, and information about the availability of data sources as open data and the means of programmatically access those data sources.

Additionally, we provide information on the level of implementation of ELI for law and ECLI for case law. As reported in the EurLex web page[5], ELI is an initiative of all countries and institutions of the European Union adopted in 2012 to harmonise the publication of legislation in Europe, and to contribute to the exchange and reuse of that information, as well as to a better access by experts and citizens in general.

The adoption of ELI by the European Union and all Member States is foreseen as a complex long term process that consists of 3 main stages or pillars, as summarized below:

- Pillar 1: identification of legal resources (documents) by unique web identifiers (URIs) readable by both people and computer systems. URIs follow an agreed template, which captures the singularities of each jurisdictions.
- Pillar 2: all legal documents are to be described by a set of common metadata of basic information about the legal resource that expands the information in the URI (for instance, title, date, publisher, or language). The metadata are defined in the so-called ELI ontology
- Pillar 3: documents are to be published according to Semantic Web standards that allow for their structuring and the establishment of relations (links) between their parts and also with other documents. In this final phase, legal resources will be exposed according to these formats to enhance exchange and interoperability. Consequently, in Annex 1 we propose an updated table of legal datasets with links to the corresponding web portals publishing legal information in Europe. All data sources have been documented in the C-KAN based Lynx Data Portal (), and any additional updates until the end of the project will be done there.

For more details on this, please visit the EU vocabularies web page (`https://op.europa.eu/en/web/eu-vocabularies/eli`), and related links.

---

[1] ttps://op.europa.eu/documents/2895081/3069147/accesstolegislationineuropeguide2009.pdf/1ef71daf-8441-17b9-d433-7b0e03524cad

[2] https://op.europa.eu/en/web/forum_official_gazettes/home

[3] https://op.europa.eu/en/web/forum_official_gazettes/european-union

[4] https://e-justice.europa.eu/

[5] https://eur-lex.europa.eu/eli-register/about.html

As for the ECLI initiative, it also proposes a template for a unique identifier that consists of 5 elements. For more information, please visit the Eur-Lex web page (`https://eur-lex.europa.eu/content/help/faq/ecli.html`).

2.2 TABLE VARIABLES

In the following we explain the rationale behind the features of the datasets analysed in the table:

Area

The state-of-the-art done in the context of the Lynx project and reported in this deliverable focuses on the legislation at EU and Member-State levels. Some of the portals mentioned in the table include links to sub-national level resources, especially for countries with federal or autonomous structure, but that is not reflected in the table.

Web site

URIs to the resources publishing legal information.

Content

The focus of this work has been on law and case law resources, the most relevant sources of legal information for the Lynx project. Several Member States only provide case law from Supreme/Superior Courts (SC in the table).

We also account for the level of implementation of ELI or ECLI, since it can be very helpful when searching for documents.

License Type

Most of the datasets do not explicitly adhere to a particular type of license. In those cases, the table reflects if the access is free (F) or restricted (R). Sometimes, a note in brackets specifies if that access only applies to some data/documents, or if registration is needed before accessing.

Format

The documents can be downloaded in different formats (PDF, XML, HTML, etc.). In many cases, there is an HTML version, but it cannot be directly downloaded. This is reflected by the expression (online).

Language

The two-letter code in ISO 636-1 is used to capture the available languages for each dataset

Access Interface

The table reflects if an API exists and/or a bulk download of documents is allowed and/or an SPARQL endpoint is available. When this is the case and the access it to be found on sites different from the one in the first column, the links are included.

Each of the resources included in the table are also available through the Lynx Data Portal[6].

---

[6] http://data.lynx-project.eu/dataset

3 CONCLUSIONS

Legal data in Europe is published in heterogeneous formats, including non-machine readabl andproprietary formats. This fact has two main consequences:

- o The information is barely updated, and a great part of the datasets are outdated.
- o It is more difficult (or even impossible) to harvest the data from such resources.

In the same line, not all the resources are available for bulk download, hindering data retrieval.

- - The selection of metadata to describe the resources is barely unified; this hinders the harvesting and processing of legal information.
- - We found that in some countries, the information is only published in their native language, which makes difficult the identification and retrieval of data.

Due to the reasons enumerated above, we have given priority to the data sources that belong to the jurisdictions involved in the Lynx business cases for their ingestion in the current version of LKG. However, the analysis performed in the context of WP2 and reported here was a necessary exercise for the planning of the next steps in the LKG enrichment.

ANNEX 1 – TABLE OF RELEVANT DATASETS

Area	Web site	Content		License type	Format				Access interface	Language (ISO 639-1)
		Law	Case law		PDF	XML	HTML	Other		
European Union	https://eur-lex.europa.eu/	x (ELI)	x (ECLI)	F	x		x			EU languages
	https://data.europa.eu/euodp/es/data/dataset/official-journals-of-the-european-union-in-english	x (ELI)		F		Formex format	x	xsl	API	EN (EU languages)
	https://data.europa.eu/euodp/es/data/dataset/eu-case-law		x (ECLI)	F	x		x		API	EU languages
België/Belgique/ Belgium	https://belgiumlex.be/	x	x	F	x		x (law, online)			NL and/or FR
Bulgaria	https://dv.parliament.bg/DVWeb/index.faces	x		F	x		x (online)			BG
	http://www.vks.bg/		x (SC)	F	x					BG
	http://www.sac.government.bg/		x (SC) (ECLI)	F (registration)						BG
Česká republika/Czech Republic	http://www.nsoud.cz/Judikatura/		x	F			x	doc		CS
	https://www.mvcr.cz/web-legislativa.aspx	x		F	x					CS
Danmark/Denmark	https://www.retsinformation.dk/	x (ELI)		F			x (online)			DA

Country	URL									Language
Deutschland/Germany	https://www.bundesanzeiger.de/	x		F	x		x (online)			DE
	https://www.juris.de/jportal/index.jsp	x	x	F (some)	x		x			DE
Eesti/Estonia	https://www.riigiteataja.ee/	x	x (ECLI)	F	x	x (law)	x (law, online)	txt (law)		ET EN (some laws)
Éire/Ireland	http://www.irisoifigiuil.ie/	x		F	x					EN, GA
	https://beta.courts.ie/		x (SC)	F	x					EN
Elláda/Greece	http://www.et.gr/index.php/	x		F	x					EL
	http://www.areiospagos.gr/		x (SC)	F			x (online)	txt		EL
	http://legislation.di.uoa.gr/	x (ELI)		F			x (online)	rdf, json	Endpoint	EL
España/Spain	https://www.boe.es/	x (ELI)		F	x	x			API (information at https://www.boe.es/datosabiertos/)	ES (CA,GL,EU)
	http://www.poderjudicial.es/search/indexAN.jsp		x (ECLI)	F	x					ES
France	https://www.legifrance.gouv.fr/Bases-de-donnees/	x (ELI)	x (ECLI)	F	x			rtf		FR EN (some texts)
	https://www.dila.premier-ministre.gouv.fr/repertoire-des-informations-publiques/les-donnees-juridiques	x (ELI)	x	LO V2.0		x			Bulk (https://echanges.dila.gouv.fr/OPENDATA/<DB name>))	FR

Country	URL									Lang
Hrvatska/Croatia	https://pravosudje.gov.hr/pristup-informacijama-6341/zakoni-i-ostali-propisi/zakoni-i-propisi-6354/6354	x (ELI)		F			x (online)			HR
	https://sudskapraksa.csp.vsrh.hr/home		x (ECLI)	F	x			txt		HR
Italia/Italy	https://www.normattiva.it/	x		F		x	x			IT
	https://www.gazzettaufficiale.it/	x		F	x					IT
	http://pst.giustizia.it/PST/		x	R						IT
	http://www.italgiure.giustizia.it/		x	F (some)						IT
Kýpros/Cyprus	http://www.cylaw.org/	x	x	F	x (law)		x (case law)			EL
Latvija/Latvia	https://www.vestnesis.lv/	x		F	x		x (online)			LV
	https://likumi.lv/	x		F			x (online)			LV, EN
Lietuva/Lithuania	https://www.e-tar.lt/portal/lt/index	x		F	x		x (online)	docx, odt	?	LT
Luxembourg	http://legilux.public.lu/	x (ELI)		F	x	x	x	rdf	Endpoint, Bulk (http://legilux.public.lu/editorial/casemates)	FR
	https://justice.public.lu/fr/jurisprudence/jurisprudence-judoc.html		x (ECLI)	F	x					FR

Country	URL									
Magyarország/Hungary	http://www.njt.hu/	x		F		x (online)				HU, EN (some laws)
	https://birosag.hu/birosagi-hatarozatok-gyujtemenye		x (ECLI)	F				rtf		HU
Malta	http://www.justiceservices.gov.mt/LegalServices.aspx?pageid=72	x		F	x					EN
	https://ecourts.gov.mt/onlineservices/Judgements		x (ECLI)	F	x					EN or MT
Nederland/Netherlands	https://www.officielebekendmakingen.nl/	x		F	x	x	x	odt	API (information at https://www.koopoverheid.nl/documenten/instructies/2018/03/24/handleiding-open-data-webservice-van-overheid.nl---sru)	NL
	https://uitspraken.rechtspraak.nl/		x (ECLI)	F						NL
Österreich/Austria	https://www.ris.bka.gv.at/	x (ELI)	x (ECLI)	F	x		x	rtf	API (https://data.bka.gv.at/ris/api/v2.5/)	DE
Polska/Poland	http://isip.sejm.gov.pl/	x		F	x (some)					PL
	http://orzeczenia.nsa.gov.pl/cbo/query		x (ECLI)	F				rtf		PL

Country	URL									Language
Portugal	https://dre.pt/	x (ELI)		F	x		x (online)			PT EN (summaries)
	http://www.dgsi.pt/		x	F			x			PT
România/Romania	http://www.clr.ro/RO_Update/LexUpdate.aspx	x		F			x (records)			RO
	http://www.scj.ro/736/Cautare-jurisprudenta		x (ECLI)	F	x		x (online)			RO
Slovenija/Slovenia	http://www.pisrs.si/Pis.web/	x	x	F	x		x (online)			SL
Slovensko/Slovakia	https://www.slov-lex.sk/vyhladavanie-pravnych-predpisov	x		F	x		x (online)			SK
	https://obcan.justice.sk/infosud/-/infosud/zoznam/rozhodnutie		x (ECLI)	F	x					SK
Suomi/Finland	http://www.finlex.fi/	x	x (ECLI)	F	x (law)		x (online)			FI, sámi SV, EN (some texts)
Sverige/Sweden	https://lagrummet.se/	x	x (ECLI)	F	x (case law)		x (online)			SV
United Kingdom	http://www.legislation.gov.uk/	x		F	x		x (online)			EN, CY
	https://www.bailii.org/		x	F	x		x	rtf		EN
	https://www.supremecourt.uk/decided-cases/		x (SC)	F	x					EN