

## **Lifting Legal Open Data to Linked Open Data Overcoming Barriers of Access and Language in European Law and Case Law**

**Guido Boella**, University of Turin, Italy, boella@di.unito.it  
**Luigi Di Caro**, University of Turin, Italy, dicaro@di.unito.it  
**Michele Graziadei**, University of Turin, Italy, michele.graziadei@unito.it  
**Loredana Cupi**, University of Turin, Italy, loredana.cupi@unito.it  
**Carlo Emilio Salaroglio**, University of Turin, Italy, salaroglio@hotmail.com  
**Llio Humphreys**, University of Luxembourg, lliobh@gmail.com  
**Hristo Konstantinov**, APIS EOOD, Bulgaria, ico@apis.bg  
**Kornel Marko**, Averbis GmbH, Germany, kornel.marko@averbis.com  
**Livio Robaldo**, University of Turin, Italy, robaldo@di.unito.it  
**Claudio Ruffini**, Nomotika Srl, Italy, claudio.ruffini@nomotika.it  
**Kiril Simov**, ICT-BAS, Bulgaria, kivs@bultreebank.org  
**Andrea Violato**, Nomotika Srl, Italy, andrea.violato@nomotika.it  
**Veli Stroetmann**, empirica GmbH, Germany, veli.stroetmann@empirica.com

This paper covers a wide range of issues addressed by EUCases FP7 project, whose aim was to develop a unique pan-European Linking Platform (EUCases Linking Platform) and services linking law and case law of Member States (MSs) to European Union legal system, thus providing access and search facilities at the European level and transforming multilingual legal open data into linked open data after both semantic and structural analysis in order to improve their usability and retrieval.

It has reused the legal Open Data sets of the European (EUR-Lex) and national legislative and case law (high courts and appeal courts) portals of 6 selected MSs (AT, BG, DE, FR, IT, UK), as well as open access doctrinal work.

Open data creates a social and economic value, and this is true also for Legal Open Data, but while much Open Data is in the form of structured data, Legal Open Data is unstructured, textual, multilingual and in a technical language, which means they have to be managed with different methodologies.

A number of initiatives have been introduced in recent years for ensuring cross-border access to national legislation and case law on the application of EU law, with financial support and/or direct involvement of various EU institutions.

In April 2006, the EU Publications Office launched the N-Lex portal as a common gateway to official legal databases of the MSs. To guarantee a common system for the identification, citation, metadata annotation and publication of national case law, the Council of Ministers has invited the MSs to introduce the standards European Legislation Identifier (ELI) and European Case Law Identifier (ECLI), elaborated by the Working Party on e-Law, on a voluntary basis. An European Legal Doctrine Identifier (ELDI) has been proposed to cover legal doctrine.

Despite the aforementioned initiatives and other attempts on the European level, many problems continued to exist hampering the establishment of integrated services in the offering of simultaneous cross-border and cross-linked access to EU and national legal resources.

Access to raw data, legislation and case law in the different countries was in fact inadequate and selecting cases and linking them to the EU dimension was a slow and costly process requiring legal expertise, thus it is commercially not viable nor sustainable by PAs without an increase of productivity in the document lifecycle by help of technologies.

EUCases software stack has addressed the problem of lifting Legal Open Data to Linked Open Data responding to market demand of a more sophisticated system and allowing multilingual search driven by knowledge of the relations between EU and national legislation. It covers all steps in the publication process: collecting the data from institutional portals by means of web crawlers; enriching them using Human Language Understanding technologies and ontologies; publishing enriched documents and knowledge (metadata, references, ontologies, key terms, case summaries) as Linked Open Data in RDF in order to facilitate access, multilingual search and further reuse.

Furthermore, besides publishing the Linked Open Data, two end-user applications have been developed within the project to prove the viability of the proposed platform.

First, the Consumer-Cases online service provides access to a multilingual collection of tens of thousands of national court decisions of the aforementioned MSs cross-linked with relevant EU and national legislation and case law as well as other Open Data legal resources (articles, commentaries, bills, administrative guidelines).

Second, as an add-in application that works in the most common text editors and browsers, the EULinksChecker (connected via the internet to the EUCases Linking Platform) assists legal professionals in their research by identifying and establishing connections between EU legislation, EU case law, and national cases and is automatically able to connect terminology in the documents to definitions in the legal ontology, helping to clarify the meaning of the document, and making its terminology more uniform.

On the other hand, the technologies used have been first of all the open legislative XML standard Akoma Ntoso and the above-mentioned ELI and ECLI standards to structure the documents by suitable parsers.

The conversion of the downloaded raw legal open data available in various text formats into a homogeneous XML format specifically designed for structural and semantic representation of legal documents has been a crucial stage of the EUCases project implementation.

Partners in the EUCases consortium have chosen the Akoma Ntoso legal XML standard for modelling in a harmonized structure the various legislative and judicial documents downloaded from the open data portals of the European Union and the six project EU Member States.

Important reasons for this choice are related with the fact that Akoma Ntoso is increasingly adopted by parliaments, among which the European Parliament and the Italian Senate.

In addition to this European Commission and the EU Council aims at adopting Akoma Ntoso for document management in order to favour the interoperability with EU Parliament.

The XML schema (XSD) of Akoma Ntoso is able to mark-up different type of documents but for the purposes of EUCases project three main types of documents of Akoma Ntoso XML schema have been used:  Act, for legislative documents; Judgment, for case law documents; Doc, for open access articles.

Akoma Ntoso documents includes a metadata section that reports the references occurring in the documents as well as other conceptual information such as key terms or the classification of the documents in accordance with the well-known EuroVoc Thesaurus produced by the European Union.

In the legal domain, knowledge engineers have to specify the category to which each text belongs, selecting key terms within a domain-specific ontology and EuroVoc has represented a wide-coverage and faceted thesaurus able to manage documents and activities.

In detail, it is a large multidisciplinary, multilingual, and hierarchical thesaurus of around 7,000 categories covering the activities of the EU.

The problem of classify these multi-label text collections has been solved using Support Vector Machines algorithm (SVM) which, like others, only works, with other mono-label text, thus a preprocessing step is needed in order make use of the SMV advantages.

EUCases Linking Platform has incorporated Legal Taxonomy Syllabus (LTS), a specialist multilevel multilingual ontology, and it has extended the LTS framework to include an ontology of prescriptions.

The LTS ontology of terms is used to explain the meaning of terms of art to users. The ontology of prescriptions is used to explain norms and all their components dealing with the specificities of EU legal ontologies: it distinguishes between the EU level ontology and the different national ones, assigning terms in different languages and relevant documents to each concept.

It allowed us to express correlations between terms and concepts in the different systems, thus also providing a source of knowledge for fine-grained translations.

It is clear from this brief description that EUCases has faced several challenges, such as managing a large amount of source data and metadata in heterogeneous formats and in a number of different languages, dealing with the fact that legal language is very technical and ontologies are often created in a top-down fashion, and so on.

Notwithstanding technical obstacles and challenges, the project has allowed the SMEs – APIS, Nomotika (both operating in the market of legal information provision), Averbis, IICT-BAS, and University of Torino (other academic partners) – to transform technologies from the academic community into innovative high tech products.

#### **ACKNOWLEDGMENT**

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 611760 (project name: EUCases).