DELIVERABLE

Project Acronym: **EUCases**
Grant Agreement number: **611760**
Project Title: **European and National Legislation and Case Law Linked in Open Data Stack**

**D1.1 Report on the state-of-the-art and user needs**

Revision: Final version

Authors:

Guido Boella  
Hristo Kostantinov

with

Luigi di Caro  
Llio Humphreys  
Matey Kolev  
Kornel Marko  
Robert Muthuri  
Petya Ossenova  
Livio Robaldo  
Kiril Simov  
Andrea Violato

UNITO  
APIS  
UNITO  
UNITO  
APIS  
VERBIS  
UNITO  
ICT-BAS  
ICT-BAS  
Nomotika
Status, abstract, keywords, statement of originality

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>final</td>
<td></td>
</tr>
</tbody>
</table>

Abstract (for dissemination)
The present report summarises the results of the analysis performed on the latest scientific and technological developments in the field of legal open data, legal XML, legal ontologies, legal semantic web standards in six EU countries and in Europe at large. Findings from a survey on user needs for pan-European legal information services are presented, and the state of play on the legal information market is explored.

The report concludes that the need for EU level legal information by legal practitioners goes beyond application of EU law. Moreover, existing information portals were found to have severe limitations, because of which they do not match all needs of users regarding access to EU wide legislation.

Keywords
Human language technologies, Ontologies, Classification, Summarization, Machine translation, Legal Open Data, Legal information portals, User needs

Statement of originality
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

The authors are grateful for the comments received from peer reviewer Dr Clemens Wass, Vienna University of Economics and Business.
Executive summary

This deliverable describes first of all the situation concerning Legal Open Data in Europe and of Legal Information Portals, with an assessment of the strengths and weaknesses of the legal portals in the countries addressed by the project.

Second, it reports the state-of-the-art in the scientific fields relevant for the project, in particular the situation of the resources available to the partners in Human Language Technologies.

Third, it outlines the main findings of the questionnaire survey conducted as part of the project, in which about 100 members of the target audience, i.e. potential users, have participated.

In summary, concerning the first point our investigation has shown that public legal resources in Europe in general, and in the six Member States at the core of the project (Austria, Bulgaria, France, Germany, Italy, U.K.) in particular, are becoming more and more open for use and re-use. Thus, no hindrances in obtaining legal data from the portals initially listed in the Description of Work have been identified.

At the same time, however, the research indicates that the quality of legal information available on public portals is not completely satisfactory, due to: lack of completeness, especially in case law; low level of legal added value; and restrictions on open data licensing.

The results of the questionnaire survey lend support to the assumption that the target group is strongly interested in accessing law at European level, because legal practitioners require access to legal information from other countries not only for matters which do directly concern domains regulated by EU law, but also for a range of other purposes.

Concerning the state-of-the-art in NLP technologies, the overview presented in the report shows that the tools available to the consortium represent the state-of-the-art and that they meet all the requirements of the project, from ontologies to machine translation, summarization and classification.
# Table of Contents

1 The need for accessing law and case law in Europe..................................................6
  1.1 Introduction.............................................................................................................6
  1.2 Social and economic issues ...................................................................................6
  1.3 Technical challenges..............................................................................................8
  1.4 The EUCases solution............................................................................................9

2 Legal Open Data in Europe.......................................................................................11
  2.1 Background and perspectives ..............................................................................11
    2.1.1 From promulgation to free access to law........................................................11
    2.1.2 Free access to law or legal open data...............................................................13
    2.1.3 Linking legal open data – the road to the future Legal Semantic Web.............16
    2.1.4 Conclusions .......................................................................................................18
  2.2 Public legal and open data portals: overview and assessment .........................19
    2.2.1 Scope and assessment criteria...........................................................................19
    2.2.2 Public legal and open data portals in the six project Member States...............20
    2.2.3 Public legal and open data portals of the EU.....................................................45
    2.2.4 Open Access journals.......................................................................................52
  2.3 Pan-European legal information services.............................................................53
    2.3.1 Non-commercial services ...............................................................................53
    2.3.2 Commercial services.......................................................................................56
  2.4 Concluding remarks..............................................................................................60

3 International standards in the legal information field.............................................64
  3.1 Legal XML............................................................................................................64
    3.1.1 Introduction.......................................................................................................64
    3.1.2 Legislative XML..............................................................................................64
    3.1.3 Existing standards............................................................................................65
    3.1.4 Conclusion.........................................................................................................67
  3.2 EU and national standards for identification of legal acts, metadata and citations: ELI, ECLI..............................................................67
    3.2.1 Motivations.......................................................................................................67
    3.2.2 ELI ..................................................................................................................68
    3.2.3 ECLI................................................................................................................69
  3.3 National standards................................................................................................69
    3.3.1 Italy..................................................................................................................69

4 State-of-the-art NLP technologies and legal ontologies used by project partners71
  4.1 NLP tools available to the consortium.................................................................71
    4.1.1 UNITO..........................................................................................................71
    4.1.2 AVERBIS.......................................................................................................72
    4.1.3 IICT-BAS.......................................................................................................75
    4.1.4 Conclusions ....................................................................................................77
4.2 Tools for integrating Human language technologies: Comparison between GATE and UIMA .................................................................77
4.3 Automatic text summarization......................................................................................................................................................79
  4.3.1 State-of-the-art in summarization ......................................................................................................................................79
  4.3.2 Available Libraries and Services ......................................................................................................................................79
4.4 Multi-label classification of legislative text into EuroVoc .................................................................81
  4.4.1 Introduction ...........................................................................................................................................................................81
  4.4.2 Background and Related Work ..........................................................................................................................................81
  4.4.3 Classification of legislative text ..........................................................................................................................................82
4.5 Machine Translation .................................................................84
  4.5.1 Machine Translation Systems ...........................................................................................................................................85
  4.5.2 Aligners ..................................................................................................................................................................................86
  4.5.3 Language Model Toolkits .....................................................................................................................................................87
4.6 Ontology annotation .............................................................................................88
  4.6.1 Introduction ...........................................................................................................................................................................88
  4.6.2 The structure of a domain ontology ....................................................................................................................................88
  4.6.3 Ontology-to-text relation ....................................................................................................................................................90
4.7 Legal ontologies and the European legal taxonomy Syllabus ...........................................95
  4.7.1 Introduction ...........................................................................................................................................................................95
  4.7.2 Background ............................................................................................................................................................................97
  4.7.3 Legal taxonomy Syllabus ....................................................................................................................................................106
  4.7.4 Implementation issues .........................................................................................................................................................112
  4.7.5 Comparison with state-of-the-art ....................................................................................................................................114
5 Identification of user needs for cross-border legal information services ..........124
  5.1 User questionnaire / online survey ..........................................................................................124
  5.2 Analysis of results from the survey ......................................................................................127
  5.3 Conclusions ...........................................................................................................................................................................129
6 APPENDIX I: Answers to the survey .................................................................130
7 APPENDIX II: Legal XML - example regulation ..............................................................136
8 APPENDIX III: Legal XML - example judgement ..............................................................142
1 The need for accessing law and case law in Europe

1.1 Introduction

Open data is gathering momentum with the increase of online publication of information by governments and institutions. In Europe the momentum is the result of legislative initiatives at EU and national level\(^1\). There is a new awareness of citizen demands for greater transparency, and a belief that Open Data, particularly reuse of data, has the potential to have a great impact on the economy and society. This is true also for Legal Open Data. The publicly available data now include more and more legal documents, from legislation to case law, and, as a result of the parallel growth of the open access initiative, even doctrinal work by legal scholars. While much Open Data is in the form of quantitative structured data, legal Open Data is often unstructured, textual and in technical language, which means they have to be managed with different methodologies. In the EU context, there is the additional complexity of the multilingual character of legislation, bringing in the need to break language barriers. Moreover, putting legal documents on an institutional portal is not enough to ensure a real accessibility of Law, due to the holistic character of norms which emerge from a network of documents and not from a single piece of legislation. These difficulties affect the accessibility of the law for judges, lawyers, scholars, enterprises, public administrations and citizens, and make it difficult for SMEs to exploit such legal Open Data to produce new services. These problems pose new research questions in the field of Human Language Technologies to deal with legal texts in the context of Linked Open Data.

1.2 Social and economic issues

The accessibility of law, case law and doctrine at EU and national levels in all member states (MS) is essential for international business and harmonization of EU law\(^2\). When deciding a case which involves applying EU law, a national judge is acting as a guarantor of the European legal order itself\(^3\). His ruling becomes a decision not only on national laws but also on the application of EU legal provisions, and as such, (s)he needs access to the relevant case law of other MSs. Knowledge of case law is required not only by judges but also lawyers, companies subject to regulatory compliance, and even legislators as they seek to predict the impact of implementing new EU regulations looking at the other MSs’ choices. If the layman (particularly in Civil Law countries) usually thinks that the law is composed of regulations written by legislators, it is widely accepted in legal theory and practice that case law and doctrinal work are also highly influential and in some contexts binding. The interplay between regulations and case law is complex when looking at the European context. European treaties and regulations are directly applicable with immediate effect in the Member States while EU directives are also binding, but have to be transposed first into national legislation. National judges who apply national legislation deriving from the

\(^{\text{1}}\) E.g., in Italy, DLeg 24/01/2006 n. 36, “Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico”.
\(^{\text{3}}\) In its Resolution of 9 July 2008 (European Parliament resolution of 9 July 2008 on the role of the national judge in the European judicial system (2007/2027(INI)) the European Parliament determines the national judges as “the keystone of the European Union judicial system” who play “a central and indispensable role in the establishment of a single European legal order”.

Page 6 of 148

13/01/2014
transposition of EU law often wish to know how judges in other countries have interpreted transpositions of the same EU laws. Evidently, linking national case law to EU law has pan-European dimensions.

The need for cross-border multilingual access to national case law and legislation adopted in accordance with the requirements of EU law and to national case law on the application of Community law has long been recognised and repeatedly addressed by various initiatives of European institutions.

Developments can be observed in EU jurisprudence (see CILFIT Case\textsuperscript{4}):

\vspace{3pt}

4


the Court of Justice stresses that before deciding whether to refer a question to the Court for a preliminary ruling, the national judge must know how this question was dealt with by courts in other MSs. The growing importance of national judiciaries in the application of EU law was underlined in the resolution of the European Parliament of 9 July 2008 on the role of the national judge in the European judicial system\textsuperscript{5}. It determined that this “central and indispensable role in the establishment of a single European legal order” requires not only “knowledge of European law, but also mutual general knowledge of the legal systems of the other Member States.” The Resolution then emphasises the importance of the “availability of national databases on national court rulings concerning Community law” and expresses the consideration that “these databases should be as complete and user-friendly as possible”. Ensuring access to national case law was listed as one of the priorities in the European e-Justice action plan\textsuperscript{6}.

Despite many initiatives on the European level, there are still many problems that hamper the establishment of integrated services offering 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 is inadequate, as it is extremely difficult to search for relevant case law using only boolean queries or the references contained in the judgment, which most often refer only to the national level, neglecting the EU dimension. Selecting cases and linking them to the EU dimension is 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. The market is demanding a more sophisticated system allowing multilingual search and driven by knowledge of the relations between EU and national legislation is urgently needed. Despite the availability of data, the result is reminiscent of the old times of restricted access to a limited amount of case law resources, marked by full dominance and unbalanced competition of big legal publishers.

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 ELI\textsuperscript{7} and ECLI\textsuperscript{8}, elaborated by the Working Party on e-Law, on a voluntary basis. An European Legal Doctrine Identifier\textsuperscript{9} (ELDI) has been proposed to cover legal doctrine. The standards will probably be introduced only for newly published data.

\vspace{5pt}


7 Council conclusions inviting the introduction of the European Legislation Identifier (ELI), OJ C 325, 26.10.2012, p. 3-11.

8 According to Pt. 5 of the conclusions “Knowledge on the substance and application of European Union law cannot be solely acquired from EU legal sources, but also from national sources, in particular from national legislation implementing European Union law”.


Marc van Opijnen, The European Legal Semantic Web: Completed Building Blocks and Future Work, p. 10.
The commercial Caselex service\textsuperscript{10} remains the only attempt for commercial implementation of a multilingual system containing national case law of the Member states linked to EU law. The system was developed from 2005 to 2009 supported by two EU co-funded projects under the eContent and eTen programs. The Caselex database contained 4070 national court decisions from almost all the EU Member States in 11 subject areas of EU law. However, the service failed to achieve commercial sustainability and it ceased to be updated after EU funding came to an end.

The world of legal publishing is changing rapidly as a result of the Open Data movement. Most MSs are offering public access to portals with consolidated national legislation. Judgments which were previously available only to the case parties, their lawyers and occasionally the big legal information providers are now published freely on the courts’ internet sites. Some MSs (e.g., Bulgaria, France) have even introduced centralised interfaces for national case law. What was unimaginable until yesterday is today a fact – millions of court decisions are publicly available on the Web. Moreover, in line with the PSI Directive\textsuperscript{11}, the data are re-usable free of charge or on acceptable prices.

There is a broad market for this service, and potential customers of the service fit into two categories: legal professionals (judges and court staff, barristers and solicitors, in-house legal counsel, public prosecutors, policy makers, legal information managers, paralegals, legal researchers, academics and law students, lawyers, legislators, scholars, compliance managers, association of law practitioners) and legal information providers or publishers which aim at redistributing the Linked Open Data enriched by the platform, subject to suitable licensing terms. The emergence of events such as the Legal IT Business Show in London and LawTechCamp\textsuperscript{12} show the increasing market of IT in the legal domain, and at the same time, the very low technological level of current commercial solutions, mainly based on workflow and business administration rather than legal research.

1.3 Technical challenges

Despite the large amounts of legal Open Data accessible today, there are a number of obstacles that prevent or significantly hinder cross-border access to national legislation and case law with EU relevance:

- The amount of Open Data makes it difficult to find relevant and useful documents.
- Documents are often published as plain text without hyperlinks to the cited legal resources, preventing navigation and reasoning on the relationships among documents.
- Metadata (if any) are not in a standardised format, which hampers interoperability between the information systems of national and European institutions.
- National and EU websites are either poorly interconnected or not interconnected at all. They use different identification systems. On the national level, there is no tool for multilingual searches.
- Judicial websites in most countries do not display information on follow-up proceedings.
- Data are not published in computational formats like RDF or XML for Linked Open Data.

\textsuperscript{10} \url{http://www.caselex.com}, website probably discontinued


\textsuperscript{12} \url{http://www.legalitbizshow.com}, \url{http://www.lawtechcamp.com}
The various players in the field of legal publishing pursue their own interests, which in many cases hamper the development of open standards. Near monopoly in legal publishing sometimes discourages governments from publishing Open Data, which only benefits established market players.

Moreover, there are several challenges specific to applying Human Language Understanding tools to legal texts which prevent from directly reusing existing open source tools, besides their lack of expertise:

- A large amount of source data (millions summing legislation and case law) in heterogeneous format.
- The need to process a number of different languages (24 Official languages in EU member states). The hypertextual character of legal texts: (a) often content elements are replaced by references to other legislation (e.g., “the subjects defined by art.1 of law X”) which makes it difficult to use traditional information retrieval methodologies; (b) to identify the relevance of a judgement we must navigate from national law to European law and again to another national law.
- Legal language is very technical, and the meaning of terms is related to the context of the legislation defining it. In the EU Law context, even within the same language, the terminology used at the EU level can be different from that of the national level, even when implementing EU directives.
- Ontologies in the legal field are often developed in a top-down fashion, which result in definitions that are too abstract and generic. Moreover, it is also neglected that different legal systems use different ontologies, which are in turn different from the European level. Finally, they often forget the relation between terms and context. Thus, they received little attention from legal practitioners.
- The multilingual classification scheme promoted by the EU – Eurovoc – has thousands of labels for classification organized in a hierarchical way.

1.4 The EUCases solution

The EUCases project will develop a unique pan-European Law and Case Law Linking Platform (EUCases Linking Platform) which models the whole life cycle transforming multilingual legal Open Data in published Linked Open Data after a semantic and structural analysis. It will reuse the legal Open Data sets of the European (EUR-Lex) and national legislative and case law (high courts and appeal courts) portals of six selected Member States (BG, DE, FR, IT, UK, AT), as well as open access doctrinal work. The software stack we propose covers all steps in the publication process: we collect the data from institutional portals by means of web crawlers, we enrich them using Human language technologies and ontologies and we publish enriched documents and knowledge (metadata, references, ontologies, key terms, case summaries) as Linked Open Data in RDF to facilitate access, multilingual search and further reuse.

The role of this Deliverable 1.1. is to examine all the main elements which will be at the basis of the EUCases project, whose general picture is described by Figure 1, concerning:

---

13 EUR-Lex (EU), Legifrance (FR), legislation.gov.uk (UK), Supreme Court (UK), Bundesgerichtshof (DE), Gesetze im internet (DE), Rechtsinformationssystem (AT), Supreme Court of Cassation (BG), Supreme Administrative Court (BG), Consiglio di Stato (IT), DOAJ (Directory of Open Access Journals) and others.
1) The situation of the Legal Open Data in Europe and of Legal Information Portals, with an assessment of the strengths and weaknesses of the legal portals in the countries addressed by the project.

2) The State-of-the-art in the scientific fields relevant for the project:
   a. Legal XML standards for representing legal documents in a structured way and standards to express references to legal texts.
   b. Natural language processing tools which will be used to enrich the legal texts and improve the retrieval:
      i. Summarization, to extract an abstract from judgements.
      ii. Automated classification to associate legal documents with classifiers from the Eurovoc thesaurus of EU.
      iii. Machine translation to expand the queries from keywords in one language into keywords of another language.
      iv. Legal Ontologies: legal ontologies for EU law and tools for populating multilingual ontologies and linking the texts to the concepts.

3) Understanding the needs of the users, from lawyers to law scholars, from judges to compliance officers. The analysis of the need has been carried out through a questionnaire survey addressed to the categories of potential users listed above.

Figure 1: EUCases overall overview
2 Legal Open Data in Europe

2.1 Background and perspectives

2.1.1 From promulgation to free access to law

Making law publicly available (promulgation of law) and ensuring access to it has been considered for centuries as an immanent, inherent obligation of the state. From the Hammurabi's Code of Laws, the earliest-known example of a ruler proclaiming publicly to his people an entire body of laws, through the impressive developments of the Roman legal system between the Twelve Tables and the Justinian Code, to the Napoleon's Code civil des Français, the first European modern body of law, legal sources have been published on stone, clay tablets or paper to ensure enforceability of legislator's will.

Promulgation is necessary for the law to obtain its force. It is an irrevocable prerequisite, conditio sine qua non, for the application of the well-known presumption of legal knowledge formulated with the Latin maxim Ignorantia juris non excusat (ignorance of the law does not excuse). At the same time it is called upon to guarantee the authenticity of the source of law. Over the past few centuries promulgation of law in specialised printed state journals (official gazettes) has established itself as a worldwide practice. This practice was complemented by the publication of key decisions in official and unofficial case law reports series issued in most countries by the supreme courts and in some cases by private publishers. In fact, these publications reached only a relatively small elite of legal experts and clerks.

The development of the first computerised applications in the legal domain in the 1960s and 1970s made it possible to facilitate and broaden the access to legal resources. In many countries governments created and maintained centralised legal databases having the task to integrate in a single information retrieval system the full corpora of national legal instruments, enhanced later also with selected collections of case law. Due to the telecommunication and technology constraints of the time these systems remain inaccessible for the general public as well as for many legal professionals. Consequently, they were rapidly overcome by private legal information providers in the course of the 1970s and 1980s. With their regularly consolidated up-to-date and historic (point-in-time) versions of legal instruments, digitalisation of older, but still useful, legal and judicial acts (backlog production) and cross-links between legislation, case law and legal doctrine the private legal databases have elaborated the present-day standards for provision of legal information. However, the paid access to private legal databases was a major hindrance for broadening public access to legal information.

Nowadays, in the era of internet and e-government access to law is understood in much broader sense than it is rooted either in the dogmatic fiction of legal knowledge and the related obligation of state to promulgate the law or in the restricted, mostly paid access to electronic legal resources in the last decades of the previous century. Consequently, the public legal databases were given a new drive in the 1990s and afterwards. Law should be freely accessible – and for this reason often exempt from copyright protection – and more understandable; these are the two pillars of the modern 21st century concept.

Firstly, law should be accessible to all for free – for use and re-use. From a formal obligation of the state, publishing of law has turned into a public service for the benefit of citizens,
society and businesses. The free access to law paradigm is based on the three principles from Montreal:\(^\text{14}\)

- Public legal information is common heritage of humanity and maximising access to it promotes justice and the rule of law;
- Public legal information is digital common property and should be accessible to all on a non-profit basis and free of charge;
- Government bodies that create or control public legal information should provide access to it so that it can be published by independent organisations.

In most European countries the free access to law was additionally guaranteed by the Freedom of Information and Access to Information legislations adopted in the last decades of the 20th century and the first decade of the 21st century which ensured public access to government documents not published in the official gazettes. Directive 2003/98/EC of the European Parliament and the Council of 17 November 2003 on the re-use of public sector information\(^\text{15}\) (PSI Directive) paved the way for further commercial and non-commercial exploitation of the legal information produced by public bodies in all EU member states.

**Secondly**, law should be more understandable, i.e., clearer and more easily accessible either for legal professionals or lay persons. At European level this principle is an integral part of the EU better regulation policy as of its early origins. The principle of making law more understandable has two distinct dimensions:

- drafting good legislation (better law-making), and
- improving practical access to legislation by various editorial and technical presentation means (like consolidation, time-versioning, linking, classifying, authoring, etc.) offered by the public online legal databases.

Here, we will stress on the latter only. Consolidation is the most important precondition for making law understandable. There is no need to argue too long about its importance, if we consider for instance the case with the Bulgarian Labour Code, which has been amended 80 times after its promulgation in 1986. It is also obvious that not only a lay person, but an experienced lawyer will need a consolidated up-to-date version of the code, and if the concrete legal case is governed by provisions which are no longer in force, he or she will need in addition a consolidated historic version valid for the specific point in time.

The importance of consolidation of EU legal acts was already recognised with the conclusions of the Edinburgh European Council (11 and 12 December 1992). According to the conclusions, consolidation means the mechanical process of regrouping of the diverse fragments of legislation governing a given matter whereby the provisions of the basic act governing a particular matter, and all its amendments are brought together, without any examination or alteration of the text and without the recitals.\(^\text{16}\) The resulting consolidated text is for information only and has no legal status. In spite of the fact that consolidation does not have any legal effect, it is generally accepted today that it is almost impossible for the layman and the legal professional to access and use EU or national legislation if not published in a consolidated form.

---

\(^\text{14}\) As formulated in the Free Access to Law Declaration (see http://www.worldlii.org/worldlii/declaration/). This declaration was made by legal information institutes meeting at the 4th Law via Internet Conference in Montreal in October 2002 as a joint statement of their philosophy of access to law, and was subsequently amended at meetings in Sydney (2003), Paris (2004), Montreal (2007) and Ithaca (2012).


The task of making EU law more understandable has been declared as an integral part of the new smart regulation policy of the EU. In Section 2.4 “Making legislation clearer and more accessible” of the Commission communication “Smart Regulation in EU”\(^{17}\) the Commission confirms its commitment to further consolidate existing legislation and even “encourages Member States to consolidate national legislation which transposes EU legislation and to make it electronically available, including via the EUR-Lex portal”. The Office for Official Publications (OPOCE) is mandated since 1 September 1998 to carry out the consolidation of Community legislation under the control of the inter-institutional Working Group on Consolidation.

Next to the impressive consolidation project OPOCE is continuously developing and improving the EUR-Lex web portal\(^{18}\) as a complex legal information system with various practical functionalities facilitating the access to EU law like links between legal acts (incl. links between EU legal instruments and case law of the Court of Justice), references to external legal resources (summaries of EU legislation, national execution measures, national case law), browsing of legal documents via several classifiers (subject matter, directory of EU legislation in force, Eurovoc thesaurus, digest of case-law classification scheme), etc. Such complex legal information services in the public domain have been developed also in many EU member states (United Kingdom, France, Belgium, Italy, Austria, Netherlands, Finland, a. o.), thus proving the new role of the state not only for publishing, but also for helping citizens and businesses to understand and apply the law.

### 2.1.2 Free access to law or legal open data

In 2003 the PSI Directive set the framework for re-use of information produced by public bodies in Europe. Public legal data was and is a huge and important domain of it. However, the intensive investment for digitalisation of older legislation and case law, production of consolidated, point-in-time versions of legal acts, cross-linking, classification, metadata extraction, etc. has prevented the governments in many EU member states from allowing real and unimpeded re-use of public legal data. In some countries public databases were not re-usable or re-use was allowed for non-commercial purposes only, in other the further exploitation was subject to high or even prohibitive fees.\(^{19}\)

Such hindrances existed also in other public data domains. Therefore non-governmental organisations and commercial players pushed national and EU institutions to overcome the barriers limiting the re-use of public sector information in favour of openness, transparency and economic growth. Similar to other “open” initiatives such as “open source”, “open content” and “open hardware”, the “open data” movement began to gain momentum over the past few years.\(^{20}\) It promoted the idea that certain data, especially data produced by government bodies with public money, should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control\(^{21}\).


\(^{20}\) One of the emblematic forms of this movement became the legal hackathons. A non-exhaustive list of legal hackathons held in 2012 and 2013 is available on: [http://legalinformatics.wordpress.com/2012/04/13/legal-hacking-upcoming-events/](http://legalinformatics.wordpress.com/2012/04/13/legal-hacking-upcoming-events/).

\(^{21}\) See the Open Data Definition on: [http://opendefinition.org/](http://opendefinition.org/).
2009 was the turning point for open data. In the USA President Obama issued the Open Government Directive\(^ {22} \) instructing all federal agencies to actively open their data to the public, and in UK Sir Tim Berners-Lee, the British computer scientist and (co-)inventor of the World Wide Web, convinced Prime Minister Gordon Brown of the possibilities of opening up public sector data.\(^ {23} \)

In December 2011 the European Commission introduced a package of measures to overcome existing barriers across the EU (the so called “open data package”).\(^ {24} \) It included a proposal for revision of the PSI Directive, financing instruments in support of open data and creation of open data portals and coordination and experience sharing action across the Member States. Referring to the Vickery study\(^ {25} \) and the McKinsey report\(^ {26} \) the Commission concluded in its communication “Open data – an engine for innovation, growth and transparent governance”\(^ {27} \) that public data exploitation holds enormous potential for the EU economy and consumer welfare, job creation and market growing, and appealed to the Member States not to wait for the new package to become law, to release their public data today so that they can benefit immediately from the revenues and jobs such policy will create. The “Open Data Directive”\(^ {28} \) amending the PSI Directive was adopted in 2013. It has to be implemented by the Member States till 18 July 2015.

With the inclusion of open data in today’s agenda of the society governments in Europe are facing a new dilemma. How will the investments in creating valuable public legal portals ensuring free access to law for citizens and businesses will pay off if the well-structured legal data will be released for re-use to everyone free of charge or against a small fee covering the marginal costs? Is it reasonable for governments to add further value to legal information or should better publish the legal databases as legal open data “as they are” and leave the production of value-added products and services to private sector players to avoid being a competitor?

Actually legal open data does not contradict or substitute the free access to law, but rather complements and extends it.\(^ {29} \) As before, public bodies are not obliged to produce information for re-use where and when this is not an integral part of their public tasks. Developing value-added public legal databases is a manifestation of the modern concept about governments’ role for ensuring free access to law (see previous section) and not an obligation under the Open Data Directive. Each EU member state is free to decide to what extent it could invest in the enrichment and improvement of the public legal portals it offers to citizens and businesses. However, having spent public money for such activities for the

---


\(^ {25} \) Review of recent studies on PSI re-use and related market developments, G. Vickery, August 2011 ([http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/final_version_study_psi.docx](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/final_version_study_psi.docx)). The Vickery study has indicated that the overall economic gains from further opening up public sector information by allowing easy access are around €40 billion a year for the EU27 whereas the total direct and indirect economic gains from PSI applications and use across the whole EU27 economy could be estimated in the order of €140 billion annually.


\(^ {29} \) In his presentation “Free access to law” and “Open Data” - Similarities and Differences” ([http://www.austlii.edu.au/austlii/seminars/2013/5_slides.pdf](http://www.austlii.edu.au/austlii/seminars/2013/5_slides.pdf)) at the research seminar organised by the Australian Legal Information Institute in Sydney on 15.11.2013 prof. Daniel Poulin argues that free access to law can be inspired by and benefit from open data business models to expand the provided services and to achieve sustainability. In spite of the fact that his considerations refer to free access to law ensured by the non-government legal information institutes all over the world, his conclusions are also valid for the public sector.
benefit of society governments should restrain from collecting unreasonably high fees for delivering the already paid legal resources as bulk data. Thus legal open data is not an alternative, but an extension of the free access to law.

The modern concept about the desirable obligations of governments in respect of ensuring free access to law and opening legal data for re-use was initially formulated at the “Expert” meeting on Global Co-operation on the Provision of Online Legal Information (October 2008), called by the Permanent Bureau of The Hague Conference on Private International Law in October 2008. The experts agreed almost unanimous on the first draft of 18 principles that states should adopt (are encouraged to implement). In February 2012 a formal international conference in Brussels involving state parties was convened jointly by The Hague Conference on Private International Law and the European Commission to consider the next steps in this process. The joint conference unanimously endorsed a set of conclusions and recommendations, which gave implied endorsement to the approach taken by the Expert meeting in 2008, and annexed the principles it set out, retitled as “Guiding Principles to be Considered in Developing a Future Instrument”. According to Prof. Greenleaf, Prof. Mowbray and Philip Chung, the 18 principles set out in Annexure 3 to the “Recommendation and Conclusions” of the Conference can be summarised in the following six essential elements:

1. **Ensuring free access** – States are to ensure that their “main” legal materials are “available for free access in electronic form by any persons”. The materials to be so made available include legislation, court and tribunal decisions and international decisions, but States are also encouraged to make available for free access historical materials, preparatory documents, amended legislation (i.e. consolidations), repealed legislation and explanatory materials.

2. **Assisting re-publication** – States are encouraged to allow and facilitate others reproducing and re-using their legal materials (for free electronic access), and to remove any impediments to such publication. This clearly implies that there should be no licence fees for republication, because that would be an impediment to free access republication.

3. **Integrity and authoritativeness** – States are encouraged to “make available authoritative versions of their legal materials in electronic form”, and to do whatever they can to ensure those “authoritative legal materials can be reproduced or re-used by others with clear indications of their origins and integrity (authenticity)”. They are also encouraged to remove obstacles to the admissibility of these materials in their courts. If the principles of integrity and authoritativeness will be adopted by the States, this will have major implications for the future of free access to law:

   (i) Courts would have to provide the most authoritative versions of their cases to all republishers, thus abandoning the situation where a commercial publisher is given a monopoly over disseminating the “authorised” version in return for some editorial work to make the case report complete. The result would be that there would be no exceptions to “Court-issued authorised versions”.

   (ii) Legislatures would have to declare that the online versions of their legislation, and delegated legislation, were as authoritative as the print versions (as some already have);

---


(iii) Courts, legislatures etc. would have to provide authenticated versions of their outputs (e.g. digitally signed copies) to all republishers, in a way that the republishers could distribute copies with the authenticity intact, so that they did not retain a monopoly over “authenticated” versions.

(iv) Courts would have to regard as admissible these republished versions, accepting them as both authoritative and authentic.

(4) **Preservation** – States are encouraged to preserve their historical legal materials.

(5) **Citations** – States are encouraged to adopt neutral methods of citation (medium and provider-neutral, and internationally consistent). Impediments to admissibility are unlikely to be overcome unless Courts adopt neutral citations.

(6) **Open formats and metadata** – States are encouraged to use open formats for legal materials, and provide metadata with them; and to cooperate in developing metadata standards. Open formats are vital, because if official bodies only provide data in formats such as PDF text (or, far worse, PDF image) then the process of republication, if it is to provide any value-adding, becomes unnecessarily difficult and expensive. When official bodies have already created metadata, they should provide it with the source data to republishers.

### 2.1.3 Linking legal open data – the road to the future

#### Legal Semantic Web

Law is a complex system of interrelated legal norms. It is not possible to understand law without tracking the nets of relationships and references between various legal acts. Legislation and case law are densely interconnected. The same is valid also for all other sources of law. Thus law is linked by definition.

Despite the fact that public legal data in Europe is more or less open, they are published as linked data in rare cases only (e.g. the UK legislation portal provides the legislation data in XML/RDF format[^33]). Even if we deliberately deviate from the strictly technological definition of linked data[^34], and think about linking legal data as publishing them cross-linked by using simple HTML-hyperlinks, we can observe that most of the national legal portals provide their data as plain text without inline- or metadata links between documents. Legislation and case law often are published on separate national portals that are not interlinked. Even when all the data is gathered in one portal (e.g. Legifrance[^35]) no links between judicial acts and cited legal acts and their provisions are established. There are many reasons (which could vary from country to country) for the widespread practice to publish on-line official legal documents without links. There are no doubts, however, that the most important from these is the problem with the complexity of legal citation techniques which hampers the automatic


[^34]: Wikipedia defines Linked Data as “a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP, RDF and URLs, but...... it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried.” See also the definition of the World Wide Web Consortium (W3C) on [http://www.w3.org/standards/semanticweb/data](http://www.w3.org/standards/semanticweb/data) and the note on Linked Data of the inventor of Internet Tim Berners-Lee: [http://www.w3.org/DesignIssues/LinkedData.HTML](http://www.w3.org/DesignIssues/LinkedData.HTML).

detection of references and the lack of standards for establishing of legal links. Legal references are formulated in natural language, often following rather vague rules which is especially valid for case law. The diversity of citation techniques is complemented with the lack of uniform identifiers. Subsequently, lawyers and judges often cite vendor specific publications of legal instruments and court decisions, and all this additionally hinders the development of citation indexes and the hyperlinking of legal documents.

At EU level, the situation is completely different. From its early origins the paid CELEX database, which became freely available on the public EUR-Lex portal as of 2004, provided links (stored outside the documents' text as metadata) between all types of legal, administrative and judicial acts – treaties, international agreements, secondary and complementary legislation, preparatory acts, case law of the Court of Justice in Luxembourg, etc. In the recent decade the European Commission and the Publications Office of the European Union were making significant efforts to link EU law to national legislative and judicial acts of the Member States on the application of its provisions. Sector 7 of EUR-Lex database offers lists of references to the so called National Execution Measures, i.e., to legal instruments of the Member States transposing EU directives into national law. However, these lists are available as plain text only without any (hyper-) links to the referred national acts. The same is valid also for the national case law on the application of EU law collected in Sector 8 of EUR-Lex. The collection includes metadata of more than 27,000 court decisions of national courts with links to the cited EU provisions and case law, but without offering either the full text of the judicial acts or (at least) links to their texts in the national case law databases. References to national legislation in the decisions' metadata are provided as plain text without links to national case law portals.

Because of its decentralised architecture, the N-Lex portal launched in April 2006 as a common gateway to national law could not solve the problem with the missing links between EUR-Lex, Sector 7 and national legal portals of the Member States. If a user would like to consult the national execution measures (MNE is the French abbreviation used in EUR-Lex) adopted by a Member State in transposition of a directive he or she should first open the list with MNE, copy one by one the titles of the respective legal instruments and then open the N-Lex portal and paste the titles in the search box for the country of question. As a result (if any) he or she will receive a variety of formats, consolidated or not consolidated legal texts with or without metadata. The same decentralised methodology has been chosen for the architecture of the future ECLI portal for national case law which is expected to be launched in 2014 as integral part of the European e-Justice portal which means that the linking problem will remain unresolved unless a different and, most importantly, uniform approach will be applied.

Such uniform approach was proposed by the EU Council with its invitation to the Member States to introduce common standards for online publications of legislation and case law –

36 Recently, several OASIS members with expertise in the legal field drafted a proposal for establishing a Technical Committee to develop a free, open mark-up standard for citations to legal materials. See http://XML.coverpages.org/legal-citation-markup-problem-statement-draft.html.


the so called ELI\textsuperscript{41} and ECLI\textsuperscript{42} standards. Both standards were elaborated by the EU Council Working Group on e-Law and e-Justice with a view to establishing the common area of freedom, security and justice by enabling European and national legislation and case law to be referenced in a harmonised and stable way. This will allow the provision of simple and fast online access to legal data at regional, national and EU level for public authorities, professional users, academics and citizens.

The idea behind ELI and ECLI standards is to introduce 1) unique identifiers for legislative and judicial acts in order to unequivocally identify, link and access national and European legislation and case law online, and 2) a minimum set of uniform and structured metadata to ensure a common format for legal data with the aim to facilitate the efficient search, exchange of legal information and interoperability between information systems of national and EU institutions. ELI is fully compatible with the W3C linked data standards. It uses HTTP URIs to uniquely identify pieces of legislation, and promotes RDFa serialisation of the metadata elements in compliance with the W3C Recommendation ‘RDFa in XHTML: Syntax and Processing’.

According to Marc van Opijnen, senior adviser on legal informatics at the Netherlands Council for the Judiciary and member of the EU Council Working Group on e-Law and e-Justice, ELI and ECLI standards pave the way for the future European legal semantic web. “While the legal semantic web is that specific part of the semantic web that is about legal data, – so Opijnen – the European legal semantic web refers to the interconnected and interoperable legal semantic webs of all Member States and that of the European Union. [ … ] With full respect of the competence of each Member State to develop its own building blocks, the development and use of European standards and services will facilitate the functioning and interoperability of all these national legal semantic webs.”\textsuperscript{43}

ELI and ECLI standards have to be implemented on a voluntary basis from the Member States. ELI is introduced so far in Luxembourg only (see http://eli.legilux.public.lu/eli), whereas ECLI is already applied fully or partially in Slovenia, France, Czech Republic, Netherlands and Malta.

### 2.1.4 Conclusions

In conclusion we could summarise that in general EU Member States have adopted the modern concept of making law more accessible and more understandable instead of simply promulgating it. In most countries, open data is already part of the public agenda. Where this is not the case, the forthcoming transposition of the “Open Data Directive” will soon change the situation. However, at this stage the developments towards linked legal open data as part of the European legal semantic web are still in their initial phase.


\textsuperscript{42} Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law, OJ C 127, 29.4.2011, p. 1–7 (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52011XG0429(01):EN:NOT).

2.2 Public legal and open data portals: overview and assessment

2.2.1 Scope and assessment criteria

After the general overview of the latest developments towards legal open data in Europe, in this section we explore in more detail the public legal and open data portals established in the six project EU member states (Austria, Bulgaria, France, Germany, Italy and United Kingdom) and in the EU itself. Portals of other EU member states and of European countries which are not Union members are falling outside the scope of this study. The same is also valid for portals developed and maintained by private bodies and non-governmental organisations, i.e. only official legal data sources are described below. Both official sites ensuring public access to legislation and such publishing case law are considered relevant.

Next to the traditional legal data portals (more or less open for re-use, i.e. being more or less “open data portals”), we examine the newly established “genuine” open data portals (mainly with respect to the presence or absence of published legal datasets and the formats in which they are published).

In order to ensure a uniform approach and means for measurement of quality and added value one and the same set of assessment criteria are applied. Building on the modern “public access to law” concept discussed in the previous section and the latest standardisation efforts of the EU Council in conformity with the linked open data and semantic web principles, we use the following assessment criteria for legal data portals:

1. **Completeness** – Whether the published legal data are complete and fully / almost fully cover the needs of the users or the latter still need to search for other non-official legal information resources in order to have access to the full corpus of legislation and the great mass of the case law available in digital format.

2. **Legal added value** – Whether the legal data is published “as is” or they are enriched in order to be more understandable (e.g. consolidated and point-in-time versions of legislative acts, case summaries, annotations, editorial notes, etc.), whether legislation and case law are classified by meaningful classification criteria (e.g. subject matter or hierarchical legal thesaurus), whether the portal offers satisfactory search facilities, whether most important sources of law are translated into foreign languages, etc.

3. **Legal links** – Whether and to what extent the legal data are provided with hyperlinks to the referred internal (within the portal) or external legal resources (e.g. in-line or metadata links to the cited legal acts and/or provisions, links between legislation and case law collections or between national and EU legal resources), whether legal data are published in conformity with the ELI/ECLI and/or linked data standards, etc.

4. **Data structure and formats** – Whether legal data is well structured and marked-up, in which formats legal documents are available for export/download, are these formats open or proprietary, etc.

5. **Open data and licensing policy** – To what extent published legal data are freely available for use and re-use, whether they are offered for bulk download in machine-readable format, etc.

For each of the above five assessment criteria the evaluated public legal portals are given one of the following three scores: *fair*, *good* and *very good*. Open data portals are being
explored only as far as they offer or not access to public legal data without being evaluated according to the above criteria.

2.2.2 Public legal and open data portals in the six project Member States

Austria

Austria together with France is among those EU Member States that apply a centralised approach for ensuring access to public legal data. Legislation, case law and preparatory legal documents are published on a ‘one-stop-shop’ portal maintained by the Federal Government.

Rechtsinformationssystem (http://www.ris.bka.gv.at)

Rechtsinformationssystem (RIS) (in English: Legal Information System) is an on-line free accessible legal database, which is operated by the Austrian Federal Chancellery. The portal provides a comprehensive collection of Austrian federal and state legislation, federal and state law gazettes and case law of the federal and state courts. Launched initially in 1983 as internal system accessible only to the public administration, RIS was made publicly available on the Internet free of charge after the Federal Chancellery decided to open it in June 1997.

Completeness

If not the most, RIS is definitely one of the most comprehensive national legal collections in Europe. The database covers:

- **Austrian federal legislation** – the full body of Austrian federal legal instruments in force in their applicable up-to-date version incorporating all amendments (i.e. consolidated version) with the opportunity to access previous (point-in-time) versions for many of them as well as all officially published in the Federal law gazette initial versions of these instruments. The collection of the Federal law gazette covers all issues of the paper editions from 01.05.1945 to 31.12.2003 in machine-readable format and all authentic electronic issues published on-line as of 01.01.2004. From the RIS portal there are also links to the ALEX portal of the Austrian National library (http://alex.onb.ac.at) providing on-line access to the scanned images of the Law Gazette of the Habsburg Empire since 1849 as well as archive law collections from the age of the Empress Maria Theresa and afterwards.

- **Austrian state legislation** – the applicable consolidated versions and with some of the historic (point-in-time) versions of the legal instruments adopted by the governments of the Austrian States (Provinces) as well as all non-authentic issues of the State Law Gazettes of the following Austrian States: Burgenland, Carinthia, Upper Austria, Salzburg, Styria, Tyrol, Vorarlberg. Since 2014 RIS provides access to the authentic electronic issues of the State Law Gazettes of the Provinces Carinthia, Styria, Tyrol and Vienna.

- **Austrian municipal law** – selected municipal law collection of some of the municipalities in the following Austrian Provinces: Carinthia, Lower Austria, Salzburg, Styria, Vienna.

- **Austrian federal and state case law** – decisions of the Constitutional Court (since 1980), Administrative Court (since 1990), selected decisions of the Supreme Court, Regional Appeal Courts, Regional and District Courts, and other courts in civil and criminal matters.
• **By-laws of Austrian ministries** – selected decrees and instruction edicts of Austrian federal ministries.
• **Bills** – collection draft bills of Austrian ministries and government bills.
• **Austrian laws in English** – translated in English consolidated versions of more than 80 selected Austrian federal laws.

**Score:** Very good.

**Legal Added Value**

RIS portal offers many value-added features like consolidation of legislation, classification of federal law by subject matter, detailed document metadata, authorised annotations of selected court decisions, editorial notes, and translation of selected federal laws in English:

• RIS provides access to a comprehensive collection of consolidated versions of all Austrian federal legislative acts in force with the option to access previous (historic) versions of most of them. For each document the user can choose a date and receive the full text of the act to this specific point in time.

• Legal documents are provided with a detailed set of metadata, incl. links to modifying/modified acts and their publication in the Federal Law Gazette.

• The Austrian legislation is classified by subject matter according to the hierarchical 3-level taxonomy “Index of Federal Law” (“Index des Bundesrechts”).

• Where needed, editorial notes are added to metadata, for instance to inform users about specific conditions related to the entry into force of the document.

• Selected court decisions are provided with expert annotations of the most important legal rule(s) deriving from the reasoning part of the judgment (the so called “Rechtssatz”) and relevant key provisions are added to document metadata.

• The consolidated up-to-date versions of more than 80 most important Austrian federal laws are provided in English translation.

**Score:** Very Good.

**Legal links**

RIS portal provides a number a hypertext links between its numerous legal collections. Most of them are established as metadata links, but there also some in-line links within document texts as well as links to external legal resources like the EUR-Lex portal and the web site of the Austrian parliament:

• **Metadata links** – these are typical for the metadata field presenting the history of modifications. Links are established:
  o to the initially published PDF- or electronic version in the Federal Law Gazette of the act and to the PDF- or electronic versions of all subsequent modifying (amending) acts,
  o to the various stages of the pre-parliamentary and parliamentary procedures for these acts (links to the web site of the Austrian parliament where the government bills, reports of the relevant parliamentary commissions, debates in the parliament, etc. are published),
  o to the text of the directives in EUR-Lex portal when the initial or modifying legal act transposes provisions of the directives,
  o to the up-to-date consolidated version of the act and to a form for choosing a date for obtaining the historical version in force to the that specific point in time.
• **In-line links** – these are established within the texts of the legal documents in exceptional cases only. Such links could refer to:
  o the issues of the Federal Law Gazette where these are explicitly cited together with the reference to the legal act,
  o annotations of cited court decisions.

However, RIS does not provide in-line hyperlinks in cases where one legal provision refers to a provision of another legal act or when legal provisions are cited in case law. The latter is partially compensated by the option to search for case law by entering the number of a concrete legal provision and the abbreviation of the respective legal act in the “Norm” (Key provision) field of the case law search menu.

*Score:* Good.

**Data structure and formats**

RIS is characterised by a well-established data structure. The database architecture follows the classical models of the most legal information systems developed in the last decades of the 20th century where legal documents are split into articles (legal paragraphs) which form the basic logical information unit. For each legal paragraph RIS maintains a set of metadata. When executing search queries, users receive a list of legal paragraphs containing the search terms instead of a list of legal acts. The same is also valid when users search for a concrete act by searching its title or abbreviation: the result is a list of all paragraphs and articles of the act. When opening a legal paragraph from the list, users can see the text and the metadata related to this paragraph. In addition there is an option to receive the full text of the legal act in its consolidated up-to-date version or to choose a date and to receive the consolidated historic version valid to this specific point in time.

Each legal paragraph and each legal act can be opened and exported / downloaded in HTML, PDF and RTF format. The data is offered also in XML format for re-use (see below).

*Score:* Very Good.

**Open data and licensing policy**

RIS is one of the best examples in Europe for legal open data. The data contents of the portal are opened for re-use with the only requirement to quote RIS as source of the data. Data re-users may acquire the data in one of the following ways:

• **Free of charge via online link queries** – RIS offers an application interface that allows making various search queries via URL links. Documents can be referenced individually by document number or in groups by defining search criteria. A detailed documentation how such link queries can be established is published on RIS portal.\(^\text{44}\)

• **Free of charge via the Government Data portal** – RIS is presented in the newly established open data portal of Austria with one of its databases, namely “Federal Law consolidated” (“Bundesrecht konsolidiert”), and provides a well-documented Web service for the data download (see “Austrian Government Data Portal” below).

• **Via the paid Document-service of the Federal Chancellery** – the Austrian Federal Chancellery offers a paid document service that allows the legal data re-users to order a single bulk data download and/or daily or weekly updates of the so called RIS applications (“RIS Anwendungen”). The RIS applications cover the different types of legal data governed by the portal. Currently the portal comprises six applications: Federal Law, State Law, Municipal Law, Case Law, Decrees and Austrian Laws in English. Download of data is done through FTP-server. According to the new price list

\(^{44}\) See [http://www.ris.bka.gv.at/RisInfo/LinksaufDokumenteimRISsetzen.pdf](http://www.ris.bka.gv.at/RisInfo/LinksaufDokumenteimRISsetzen.pdf).
for 2014 the data purchase for a single data download amounts € 110 per application. The costs for the annual subscription for daily or weekly updates of the data are € 1000 per application.

Considering the RIS open data policy we may conclude that the Austrian Federal Government offers the legal data on very favourable terms. The paid service is cost-effective, especially if compared with similar paid services in other EU Member States (like Legifrance), and in the same time the re-users may choose the option to download the data free of charge by using the provided RIS API or the RIS Web service on the Austrian Government Data portal.

Score: Very Good.

**Austrian Government Data Portal (http://www.data.gv.at)**

The Open Government Data Portal of Austria was presented on 20 April 2012. Currently it comprises more than 1000 datasets published by various federal, state and municipal public bodies. The data are distributed under the Creative Commons Attribution License (CC BY 3.0). In the portal the legal data are presented by the RIS application “Federal Law consolidated” and the “Salzburg City Law, 1966” (the Statutes of Salzburg from 1966).

The RIS application is accessible via a web service offering three methods for machine requests to the URL `http://data.bka.gv.at/ris/OGDSерvice.asmx`: `version` (returns the actual version number), `request` (search for documents) and `getDocument` (retrieve a single document). The Government Data Portal provides a detailed guide describing the web service.\(^{45}\)

**Bulgaria**

Bulgaria is one of the few countries in Europe in which public access to legislation is still limited to the promulgation of legislative acts in the official gazette – the State Gazette, extended only with the provision of a simple search interface. However, the access to case law was significantly improved in the recent years and now almost all acts of the 186 courts in Bulgaria are accessible online.

**State Gazette Portal (http://dv.parliament.bg)**

The State Gazette (the official journal for the Republic of Bulgaria) is issued by the National Assembly. The technical process of issuing of the gazette is prepared and carried out by an editorial board under the supervision of the President of the National Assembly.

After a substantial reconstruction of the pretty ordinary previous web page of the State Gazette, the State Gazette Portal was launched in the end of June 2008. The portal is governed by the same editorial board which is responsible for the issuing of the State Gazette. As of 1 July the portal offers free access to a PDF copy of the legally binding paper edition of the State Gazette. In the same time users can use a relatively simple search interface to query the documents published in the State Gazette, which are available also in HTML-format.

---

**Completeness**

The State Gazette Portal offers public access to the initial publications of the Bulgarian legal acts only. Consolidated up-to-date and historic versions are not provided. Next to the legal instruments published in the official section of the State Gazette, the portal contains also all individual administrative acts of ministers and heads of central and local government bodies, notifications and announcements of ministries, other departments, higher schools and research institutes, municipalities, summonses issued by courts and other notifications published in the unofficial section of the State Gazette.

After the launch of the portal in 2008, in parallel with the ongoing publication of new regulations and other documents, the contents of the portal are gradually expanded through the digitalisation of old issues of the State Gazette published before 2008. Currently users of the site can find Bulgarian legal instruments (as well as the documents in the unofficial section) published after 01.01.2001. However, this is not sufficient because there are still many legal acts in force which have been published after the Second World War and before the year 2001 (e.g. Law on obligations and contracts from 1950, Law on succession from 1949, Criminal Code from 1968, Labour Code from 1986, a. o.). This fact combined with the lack of consolidated version of legal instruments determines the relatively modest usage of the portal by the legal professionals in comparison with the services of private legal information providers.

*Score: Fair.*

**Legal Added Value**

As already mentioned above, the State Gazette Portal does not provide consolidated Bulgarian legislation which reduces significantly its usage by citizens and legal professionals. This need is partially compensated by private suppliers or non-governmental organisations which maintain free available portals containing various information resources, incl. consolidated texts of legal acts (e.g. lex.bg, econ.bg, a. o.). However, the public need for establishment of an official source of actual legal information remains unsatisfied. Bulgarian government made a serious mistake in 2009 when it closed the small editorial office of the loose-leaf paper edition “Normative acts”, part of the administration of the Council of Ministers, instead of restructuring its activity and directing it to electronic publications. This small administrative unit of three civil servants was the only official institution having the experience and knowledge to consolidate legal texts. Thus Bulgaria remains one of the few EU Member States not having a public portal with consolidated national legislation.

*Score: Fair.*

**Legal links**

Documents in State Gazette Portal are published without any hyperlinks. There are neither links between legal acts within the portal nor links to external legal resources.

*Score: Fair.*

**Data structure and formats**

The State Gazette Portal publishes the legal data in PDF and HTML formats. The issues of the State Gazette are available as searchable PDF files representing an exact copy of the paper edition. All acts published in the issues of the State Gazette are uploaded also in HTML format. The database behind the portal has a quite simple architecture. Only a limited set of metadata is maintained for each document allowing the user to search documents by title, number and/or date of the issue of the State Gazette, section and rubric of the State Gazette, document type, author of the document and city of the author. Searches by user search terms are also possible. However, when opening a document from the result list, users have to call the search function of the browser to find the matches.

*Score: Fair.*
Open data and licensing policy

There is no formal statement on the State Gazette Portal giving any notion of the conditions for data re-use. No interfaces or web services for documents download are available as well. However, in Bulgaria it is generally accepted, that public legal data is free for re-use and no database rights can be claimed even if the data is offered with high added value. The lack of consolidated legal texts makes the data not so attractive for re-use. One of the few data re-users is the Spanish company vLex. The State Gazette Portal is also searchable via the N-Lex search interface.

The editorial board which is responsible for the issuing of the State Gazette provides a very specific paid data delivery service targeting the needs of the private legal information providers. It sends per e-mail the final electronic version of each issue of the State Gazette (which just has been sent to the printing house) to the service subscribers in the day of printing, i.e. the day before the actual date of the printed issue. This service allows the private legal information providers to consolidate the texts of the legal acts with the new issue of the State Gazette and to offer them to the customers of their legal databases on the next day simultaneously with publication of the amending acts in the State Gazette. The fee for the above described data delivery service amount 7000 Leva (approx. € 3500), a fee which is not fully consistent with current understanding for marginal costs and open data.

Score: Good.

Case Law Portal of the Supreme Judicial Council (http://legalacts.justice.bg)

The Case Law Portal was developed on initiative of the Supreme Judicial Council (SJC), the supreme administrative body of the Bulgarian judiciary. It was launched in 2009 with the idea to provide a centralised interface for publishing and accessing the decisions of all Bulgarian courts without those of the Supreme Administrative Court, the Supreme Court of Cassation and of the Constitutional Court. The portal uses an API which allows for interaction and data interchange with the case management systems of the courts in order to receive from them the decisions texts and metadata in XML format.

Completeness

The portal provides access to more than 1.000.000 decisions of the district, regional and administrative courts, courts of appeal and military courts in Bulgaria rendered after 2008. Probably due to technical difficulties part of these courts do not publish their case law on the portal of the SJC. Unfortunately, among them are Sofia District Court, Sofia City Court and Sofia Court of Appeal whose jurisprudence is particularly prized by the legal professionals. The access to the case law of these three courts is possible via their own web sites. Nevertheless, the Case Law Portal of the SJC offers one of the most comprehensive collections of national case law in Europe.

Score: Very Good.

Legal Added Value

The texts of the court decisions are published “as is” without any attempt to add value in order to improve the access and searches or to support the better understanding of case law. The acts are equipped with the minimum set of metadata provided by the case management systems of the courts. In some cases decisions are published with corrupted, incomplete or even blank texts. No classification headings are available. The only possible distinction is by

46 http://us.vlex.com
47 The Bulgarian N-Lex connector was developed by EUCases partner APIS in 2012.
type of the proceedings: civil, criminal or administrative. Case summaries or any editorial notes are not provided. In addition, the search queries are executed very slowly which hampers significantly the normal use of the portal.

*Score:* Fair.

**Legal links**

Decisions in the Case Law Portal of the SJC are published without any hyperlinks. Exceptionally, links are established to connect cases at different stages of the court proceedings.

*Score:* Fair.

**Data structure and formats**

The Case Law Portal of the SJC publishes the court decisions in PDF or HTML formats depending on the format chosen by administration of the various courts. For each case there is a set of metadata and a list of the acts rendered by the court in this case. The metadata allows the users to search by court, case number, decision number, decision date, decision type, decision status. Search by terms is not allowed.

*Score:* Fair.

**Open data and licensing policy**

As in the case with the State Gazette Portal, the Case Law Portal of the SJC does not provide any meaningful information about the conditions for data re-use. No interfaces or web services for documents download are available as well. The traditional for web sites of Bulgarian public institutions copyright symbol in the bottom corner is not supported by a copyright notice. Unlike legislative acts, court decisions are not listed within the public acts not being considered as objects of copyright. However, in the legal doctrine has long been accepted that judicial decisions are not objects of copyright. This fact is enough for all legal information providers in Bulgaria to develop own applications for download of the case law published on the portal of the SJC.

*Score:* Fair.

**Case Law Portals of the Supreme Courts and the Constitutional Court**

Here we will make a brief common overview of the case law portals of the Supreme Court of Cassation (http://www.vks.bg/vks_p10.htm), the Supreme Administrative Court (http://www.sac.government.bg) and the Constitutional Court of the Republic of Bulgaria (http://www.constcourt.bg).

**Completeness**

With 15-16 cases in average per year the Constitutional Court definitely possesses the most complete own case law collection. Actually, all around 360 rendered acts after its foundation in 1991 are published on the web site of the court.

Almost complete is also the case law collection of the Supreme Administrative Court (SAC). The case law portal of the SAC was launched in 2002 as the first public source of case law in Bulgaria. All decisions rendered from the court after 01.01.1999 are accessible via the search interface of SAC portal. In fact, the court database does not offer case law from the first two years (1997 and 1998) after the SAC was restored as the supreme administrative jurisdiction in Bulgaria on 1st December 1996. A few decisions of the court from these first two years are available in the databases of the private legal information providers.

Not so complete, however, is the case law collection of the Supreme Court of Cassation. It comprises decisions rendered after 2008 when the newly reconstructed web site of the court
was launched. Selected decisions with authorised summaries from previous years can be found in printed journals like the Case Law Reports of the court (their exact name is: “Bulletin of the Supreme Court of Cassation”) and the legal journals issued by universities and research organisation or private entities as well as in the paid databases of the private legal information providers. These collections go back in time till the 1950s and are still used by legal professionals.

**Score:** Very Good.

**Legal Added Value**

The web site of the Constitutional Court provides to its users a simple search form for finding cases and some classificatory criteria for browsing the case law – by Article of the Constitution being interpreted, by legal basis (i.e., by legal acts which provisions are examined for unconstitutionality), by year and by judge rapporteur.

A more complex search interface provides the portal of the Supreme Court of Cassation. In section "Decisions" the site offers four different types of search forms by various search criteria and two lists with most important decisions – a list with the interpretative decisions (which are considered as source of law) and a list with the decisions on cases of exceptional public interest. Two of the search forms perform searches based on metadata of the cases and decisions (number and year of the case, incoming number, date of the decisions, decision type, court chamber, etc.), and one of them allows also to combine metadata criteria with keyword searches. The other two search forms offer the user the opportunity to browse the case law by key words and phrases like "rental contract", "public companies", etc. To this end selected decisions are annotated by the judge assistants to the court.

The portal of the Supreme Administrative Court offers only a form to search decisions by metadata criteria related to the cases or the court acts. One advantage of the site is the possibility to check the status of the pending cases and to access scanned images of the appealed administrative act and of the appeal. However, the previous functionality to browse the case law by key provisions of the legal acts cited was cancelled after the reconstruction of the portal in the autumn of 2013.

**Score:** Good.

**Legal links**

Only the site of the Constitutional Court offers links between the texts of legal acts which provisions have been examined for unconstitutionality and the texts of the respective decisions. To this end on the web site are published the full texts of the Bulgarian legislative acts being cited in the decisions of the court.

As mentioned above, the previous version of the case law portal of the Supreme Administrative Court (SAC) offered a functionality to browse the case law by key provisions of the legal acts cited, but this option is not available anymore after the reconstruction of the portal. SAC provides also information about all questions for preliminary rulings referred by the court and by all lower instance administrative courts in Bulgaria and the subsequent rulings of the Court of Justice of the EU (CoJ). However, this information is not provided by links to EUR-Lex portal or the web site of the CoJ, but only by publishing a list with the file in DOC-format of the orders and decisions of the national courts and the CoJ.

**Score:** Fair.

**Data structure and formats**

The decisions are published in HTML format as a rule. On the web site of the Supreme Administrative Court some of documents (Interpretative decisions and references for preliminary rulings) are published in MS Word DOC format. All three web sites provide also metadata of the cases, thus allowing the user to search cases by multiple criteria typical for judicial case management systems.
Score: Fair.

**Open data and licensing policy**

None of the three case law portals provides any re-use facilities for bulk data download. In addition the web sites of the Constitutional Court and of the Supreme Court of Cassation do not have copyright or legal notices. The Supreme Administrative Court has published on its web site a detailed legal notice prohibiting any kind of commercial re-use without prior permission. However, as in the case with the portal of the Supreme Judicial Council, it is generally accepted that court decisions are not objects of copyright, and therefore the published case law is re-used by all Bulgarian legal publishers.

Score: Fair.

**Government Open Data Portal** ([http://opendata.egov.bg](http://opendata.egov.bg))

The only reason to include this “portal” of the Council of Ministers in the present document is the fact that its existence is the best example for the lack of real public open data in Bulgaria. The site all in all provides access to three public datasets: documents of the prime minister created between 2009 and 2013, decrees of the Council of Ministers between 1990 and 1994 and protocols of the Ministry of Defence after 2010. The datasets are accessible in RDF format under the Open Data Commons Attribution License.

**France**

Similar to Austria public access to legislation and case law in France are ensured by a single centralised governmental service – the portal Legifrance.

**Legifrance ([http://www.legifrance.gouv.fr](http://www.legifrance.gouv.fr))**

Legifrance is an on-line free accessible public legal database which is operated by the Directorate of Legal and Administrative Information (DILA – Direction de l'information légale et administrative). The website is under the editorial responsibility of the General Secretariat of the French Government. The portal provides access to the consolidated texts of the French legislation, Official Journal (Journal Officiel), case law of the French courts and other valuable legal information resources. Legifrance is created by decree of the Prime minister n° 2002-1064 entered into force on 15th of September 2002.

**Completeness**

Together with the Austrian Rechtsinformationssystem, Legifrance is one of the most comprehensive national legal collections in Europe. Its databases cover:

- **French legislation** – the full body of French legal instruments in force – constitution, codes, laws, decrees, decisions, collective agreements, deliberations, circulars, instructions and acts of the National commission of computing and liberties as well as other regulations in their applicable up-to-date version incorporating all amendments (i.e. consolidated version) with the opportunity to access previous (point-in-time) versions of the legal acts. In the same time users can access the officially published initial versions of these instruments in the Official Journal. The collection of all acts published in the Official Journal covers the period from 02.01.1990 to present days.

---

48 More information about Legifrance contents is available on [http://www.legifrance.gouv.fr/Bases-de-donnees/Contenus](http://www.legifrance.gouv.fr/Bases-de-donnees/Contenus).
International law – selection of basic international treaties like the Nord Atlantic Treaty, the Vienna Convention on Diplomatic Relations, the Vienna Convention on Consular Relations, the European Convention on Human Rights, the Geneva Convention on International Commercial Arbitration and other international acts in PDF format.

Bulletins of French ministries – each ministry is obliged to issue an official bulletin every three months by decree no 2005-1755 entered into force on 30th of December 2005. There are links to them on the website Legifrance.

French case law – decisions of the Constitutional Court (since 1959), Supreme Administrative Court (since 1959), Supreme Court of Cassation (since 1941), decisions of appeal courts, administrative appeal courts and other tribunals in civil, administrative and criminal matters. It is important to mention that case law of the supreme courts before 1988 is not complete and that the decisions of the first instance courts are quite insufficient.

Translations – translations in other languages of the French constitution and most important codes and laws.

Legal news.

Score: Very Good.

Legal Added Value

Legifrance portal offers many value-added features like consolidation of legislation, classification of regulations related to companies, useful legislative materials, information about French legislation applying EU directives as well as translations of selected codes and laws in English, German, Spanish or Arabic language:

- Legifrance provides access to a comprehensive collection of consolidated versions of all French legislative acts in force with the option to access previous (historic) versions of them. For each document the user can choose a date and receive the full text of the act to this specific point in time.
- Legislative materials represent a classification of laws issued by the French Parliament from June 2002 to present days accompanied by a bill, motives and press release;
- Legifrance offers a search form for finding information about the adopted legislative measures in transposition of EU directive. User can search by directive number, date of directive or date of publication in Official Journal of the EU or by key words. For each directive from the result list user can see the adopted French legislative acts and read a short expert summary about the transposition. Access to the texts of the directives is ensured via links to EUR-Lex portal.
- The consolidated up-to-date versions of more than 20 most important French codes and laws are provided in English, German, Spanish or Arabic translation;
- There is a rubric on the portal called <Companies-entry into force to the texts>. It represents a classification of the regulations not entered into force yet which are related to companies and other business organizations. This legal-added feature includes the following elements: date of entry into force, nature of the text with link to its official publication in the Official Journal, subject area, other conditions and observations.

Score: Very Good.
Legal links

Legifrance portal provides a number of hypertext links between its numerous legal collections. There are two types of links – in-line links and links to external resources:

- **in-line links** – such hypertext links are established between French legal instruments on article level. User can see also where every article is cited. Legifrance does not provide in-line hyperlinks when legal provisions are cited in case law, but it is possible to search for case law by entering a key word(s).

- **links to external resources** – the portal supports a few external links. Consultation to bulletins of French ministries mentioned above is provided by external link to every ministry. User can review an international legislation and case law by external links to EUR-Lex, Council of Europe, Court of Justice of the European Union and the European Court of Human rights. There are also links to other official websites of the French government.

Thanks to the well-established data structure, Legifrance portal offers to its users simple and expert search forms to perform more or less complex searches in the different legal collections.\(^{49}\) It is worth to be noticed here that users can search case law not only by traditional national decision number, but also by the recently established ECLI-number.\(^{50}\)

**Score:** Good.

Data structure and formats

Legal documents are split into articles (legal paragraphs) which form the basic logical information unit. For each legal paragraph Legifrance maintains connections to other related articles and legal acts. There is an option to receive the full text of the legal act in its consolidated up-to-date version or to choose a date and to receive the consolidated historic version valid to this specific point in time. For each legal act and for each of its provision user can obtain information about the initial publication and subsequent amendments. Each legal paragraph and each legal act can be opened and exported in HTML format. The data is offered for download/export also in XML format according to the conditions described below.

**Score:** Very Good.

Open data and licensing policy

The data contents of the Legifrance portal are opened for re-use with getting a paid license from the Directorate of the legal and administrative information. Download of bulk data is done through FTP-server. Free of charge forms of re-use are also possible.

The licensing policy was changed towards extension of the free re-use just a few days before end of 2013.\(^{51}\) The existing limits for free download of HTML documents have been waived. A new free license for XML bulk data load through FTP server will be introduced. New lower prices for the paid licenses are expected to be announced in the first quarter of 2014.

Data re-users may acquire the data in one of the following ways:

---

\(^{49}\) Detailed description of the various search options can be found on [http://www.legifrance.gouv.fr/Bases-de-donnees/Rechercher](http://www.legifrance.gouv.fr/Bases-de-donnees/Rechercher).


- **Free of charge via online link queries** – Legifrance allows making various search queries via URL links, because this form of re-use is not connected with export of data. The only conditions are to mention the date of the last update and the source.\(^52\)

- **Paid license** – there is a number of different prices for each data collection. Re-users may require single bulk data download of one dataset or in addition to order annual subscription for regular updates. The data is provided in XML format. Below we quote the aggregate prices for the most important data contents which were valid till the end of 2013:
  - application for license – €400;
  - French legislation – €4707.99 plus €1427 per year for updates;
  - case law of the three supreme courts – €2944.56 plus €3413.94 per year for updates;
  - case law of first instance courts and administrative jurisdictions – €1436.80 plus €1895 per year for updates.

- **Free license** – according to the aforementioned change in licensing policy users can download as of 2014 all documents published on the portal in HTML format for free. The following download limitations were valid before this change:
  - 2400 articles from French legislation;
  - 15 decisions of the Constitutional court;
  - 1300 decisions of the Supreme Court of Cassation;
  - 2200 decisions of different administrative jurisdictions.\(^53\)

It is obvious that the prices for the paid service in 2013 were quite expensive if compared with the paid services for similar contents of the RIS portal in Austria. It could be expected that the new prices to be announced in the first quarter of 2014 will be more or less lower than those in 2013. The new licensing policy will be of interest for all data re-users in France, but also for the partners in the EUCases project, because in the project budget are foreseen expenses namely for Legifrance licences.

*Score*: Very Good.


The portal was launched in the end of 2011. It brings together data from many public agencies among them the French institute for statistics INSEE, most of the ministries (Finance, Culture, etc.), and several state-owned companies (like the French railway company SNCF). Currently the portal offers more than 12 900 datasets of 230 public organisations. However, legislation and case law are not published, because the only official source for obtaining legal data remains the paid license of Legifrance portal.

**Germany**

In Germany access to public legal data sources is highly fragmented. Federal legislation is published in *Bundesgesetzblatt* (the Federal Law Gazette) whereas consolidated texts are available via two portals with identical architecture maintained by the Federal Ministry of

---

52 Detailed information is available on [http://www.legifrance.gouv.fr/Aide/Etablir-un-lien](http://www.legifrance.gouv.fr/Aide/Etablir-un-lien).
Justice – Gezetze im Internet (Laws on Internet) for primary federal legislation, and by the Federal Ministry of Interior – Verwaltungsvorschriften im Internet (Administrative regulations on Internet) for non-legislative acts and guidelines of the federal institutions. Legislation of the 16 German States (Länder) is published in the respective state law gazettes and on the web portals of the states. Access to case law is also more complicated. Decisions of the five German federal courts are published on their web sites, whereas case law of the lower instance courts is available either on their web sites or more often on portals developed by the ministries of justice of the German states.


The portal was developed by the Federal Ministry of Justice in cooperation with the private legal information provider Juris GmbH. As of 1 January 2013 the maintenance is performed by the Document service unit of the Federal Agency of Justice.

**Completeness**

Gesetze im Internet portal provides public access to the last consolidated versions of the federal laws and regulations. Originally published and historic (point-in-time) versions of the legislative acts are not available. The same is also valid for the texts of the recently repealed acts. Thus the usefulness of the portal is reduced significantly. In addition, the consolidation process after larger amendments could take up to several months. For this reason the portal offers the so called “Aktualitätendienst” (Actuality service) – in fact a page with a list of links to the official texts of the legislative acts published in Bundesgesetzblatt in the last 6 months.

The portal offers also translations in English of selected pieces of German legislation (47 translated acts to December 2013).

**Score:** Good.

**Legal Added Value**

Gesetze im Internet portal allows two very simple search forms: the first – to search in document titles, the second – to search in document texts. Other searches typical for legal databases (like by fate of publication, type of the act, etc.) are not available, because documents are published without metadata. Thus the only means of accessing legislation remain searching in titles or texts and browsing the alphabetical list of all documents. As mentioned above, more complex consolidations appear with a delay from several weeks to several months.

Documents are provided with concise notes informing users about the latest amendments and whether these have been applied in the text as well as (where needed) with footnotes pointing at the entry into force date of specific provisions or at EU directives which requirements have been implemented with act of question.

**Score:** Fair.

**Legal Links**

Documents in Gesetze im Internet portal are published without legal links. The only hyperlinks used in the portal are 1) between the table of contents reproducing the act structure (legal paragraphs and articles, annexes, etc.) and the referred elements in this table, and 2) from the list with the latest published in Bundesgesetzblatt legislative acts “Aktualitätendienst” (Actuality service) section of the portal to the PDF-copy with the texts of the acts on the Bundesgesetzblatt web page.

**Score:** Fair.
Data structure and formats

The legislative acts are available in full text (in HTML, PDF, XML and EPUB format) and also broken into separate provisions (in HTML format). The architecture of the portal allows for establishment of the so called “deep links” from external sources to the provided legal contents at act or provision level. The HTML and XML versions of the acts contain a limited amount of metadata like title, abbreviation, long title, date, status, notes and footnotes.

Score: Good.

Open data and licensing policy

Contents of Gesetze im Internet portal are free for use and re-use. No specific licensing conditions for data re-users are raised. The legal data is also published in HTML and XML format on the recently established Government Open Data portal (https://www.govdata.de).
All public data on the portal is re-usable under the Open Data Commons Attribution License 3.0.

Score: Very Good.

Case Law Portals of the Federal Supreme Courts and of the Federal Constitutional Court

In this section we will examine together the web portals of the Federal Constitutional Court (Bundesverfassungsgericht: http://www.bundesverfassungsgericht.de) and of the five supreme German federal courts – the Federal Court of Justice (Bundesgerichtshof: http://www.bundesgerichtshof.de), the Federal Administrative Court (Bundesverwaltungsgericht: http://www.bverwg.de), the Federal Fiscal Court (Bundesfinanzhof: http://www.bundesfinanzhof.de), the Federal Labour Court (Bundesarbeitsgericht: http://www.bundesarbeitsgericht.de) and the Federal Patent Court (http://www.bundespatentgericht.de).

Completeness

All above listed federal courts provide free access for non-commercial use to comprehensive electronic collections of their case law from the recent years and the option to order paid copies of decisions rendered prior to the initial date of the respective collection. Unpublished remain only insignificant for legal professionals court acts (like orders about costs, orders not having reasoning part, a. o.) and decisions in cases being not public for security reasons. More numerous is the number of unpublished decisions of the Federal Fiscal Court (the so called “UP-decisions”).

The starting point for provision of decisions in electronic form is different for each court:

- Federal Constitutional Court – 01.01.1998;
- Federal Court of Justice – after 01.01.2000;
- Federal Administrative Court – after 01.01.2002;
- Federal Fiscal Court – after 01.01.2009;
- Federal Labour Court – after 01.01.2009;
- Federal Patent Court – after 01.01.2006. In the period from 01.01.2000 to 31.12.2005 the database contains decision of some of the court chambers only.

Score: Very Good.

Legal Added Value

Typical value added feature for the all case law collections of the six federal courts is the provision of annotations (case summaries) of selected most important decisions. The annotations (the so called “Leitsätze”, i.e. basic rules) are produced by documentation units
to the courts, often with the help of experts of the private legal information provider “Juris” GmbH as is the case with the decisions of the Federal Court of Justice. In addition to the annotations for cases of public interest press releases are published together with the texts of the decisions. Unlike the annotations, these are not targeting the legal professionals, but the general public.

Decisions are published without metadata. Therefore, all six portals offer quite simple search forms that allow the user to search decisions by date, act number and text only. In addition, the user can browse the case law collections by year, month and day of the decisions.

The Federal Constitutional Court and the Federal Administrative Court have published on their web sites a small number of translations in English of the most important decisions of international significance.

**Score:** Very Good.

**Legal Links**

Case law is provided without any hyperlinks to legislation. In-line links between court decisions published on the same web site are established only on the portal of the Federal Administrative Court, and a similar functionality is available on the web page of the Federal Labour Court where links to related court acts are provided within the text of the annotations.

**Score:** Fair.

**Data structure and formats**

In practice, court decisions are published without metadata. Such can be extracted from the document titles and the introductory part of their texts, but they are provided as plain text only.

Decisions are available:

- a) in PDF format – on the portals of the Federal Court of Justice and Federal Patent Court;
- b) in HTML format – on the portals of the Federal Constitutional Court, Federal Administrative Court, Federal Fiscal Court and Federal Labour Court. The Federal Administrative Court offers its decisions also in PDF format.

Document texts have a clear and understandable logical structure (tenor, facts of the case, reasoning part / findings of the court) which is marked-up with relevant tags in the HTML presentation, but understandably cannot be reproduced in the PDF version of the decisions.

**Score:** Fair.

**Open data and licensing policy**

The situation in Germany regarding the reuse of databases with case law is quite confusing. On the one hand, there is no copyright protection of judgements according to §5 of the German Copyright Act. On the other hand, the database as such might be protected which gives rise to case law portals to introduce a number of restrictions on re-use. Not only case law is unavailable for bulk download, but any commercial re-use without the prior permission of the federal courts is strictly prohibited. All portals contain exclusive copyright statements and instructions to contact their documentation centres if data re-users would like to download decisions for commercial purposes. Actually, the commercial re-use is regulated from the German legislator with the Administration of Justice Expenses Act (Justizverwaltungskostengesetz). The Annex to § 4, Par. 1 (Cost table) governs the fees for


use and re-use of document copies. The fee for delivery or download of electronic copies of
court decisions is €1.50 for each file. This fee was previously €2.50 per file, but even now it
could be regarded as a prohibitive fee in respect to the present EU open data policy.

§ 20 of the Administration of Justice Expenses Act foresees another option for electronic
data delivery of court decisions. It concerns the conclusion of public-private partnership
contracts, in which case the data delivery is performed against quid pro quo services of the
data re-user. Probably, exactly this provision is used by German legal information providers
for obtaining case law as bulk data from the courts, but its application is also problematic in
view of the upcoming transposition of the new EU open data directive.

Score: Fair.

Government Open Data Portal (https://www.govdata.de)
The portal was launched in February 2013 on an initiative of the Federal Ministry of Interior.
Presently it offers public access to more than 4800 datasets. For around 800 from them
however the re-use is restricted to non-commercial purposes. Legal data is presented by 9
datasets among which the legislation portal Gesetze im Internet and the portal
Verwaltungsvorschriften im Internet (Administrative regulations in Internet). The open data
policy of the German government was hardly criticised in the recent years. Thus to the
launch of the portal Not-your-GOVDATA http://not-your-govdata.de housing a declaration of
more than 800 open data activists urging the government to really open public data in
Germany was not a surprise.56

Italy

Normattiva (http://www.normattiva.it)
The main initiative in Italy concerning the access to legislation is the Normativa portal of
Presidenza del Consiglio run by Istituto Poligrafico e Zecca dello Stato, responsible of the
Official Gazzette, started on 2008. It contains about 75,000 laws and regulations and update
almost at real time with new laws.

Completeness

The portal is complete concerning the legislation by the Parliament and contains also some
other regulations. It is planned to index also regional legislation.

Score: Good.

Legal Added Value

The main value of the Portal rests in the fact that it provides coordinated texts for all points of
time: the current version of the norm and also all the intermediate ones after its original
version. The coordinated texts are created manually in a few days. However, since the
legislation is stored article by article in a relational database, the portal cannot offer such
wealth of information in more useful formats like legal XML annotating modifications in a
detailed way. Another useful feature is the list of changes for each article and the list of
secondary legislation (but it is not clear the extent of the coverage of such feature).

Score: Good.

56 For more details about the problems with the re-use of public sector information in Germany and in other EU Member States
see the MICUS Study: Assessment of the Re-use of Public Sector Information (PSI) in the Geographical Information,
Meteorological Information and Legal Information Sectors, MICUS Management Consulting GmbH, 2008,
http://www.micus.de/51a_publicsectorinformation_en.HTML.
**Legal Links**

The legislation is accessible via URN and the text is enriched by links to other legislation. However, the format of the references is not standard and even is not the same as the Normainrete project which defined an XML standard from Italian legislation. Moreover, the references are generated automatically with a parser with limited capabilities, so that many links are missing, e.g. internal references.

*Score: Good.*

**Data Structure and Formats**

The legislation is published in HTML, with the possibility to select single articles. It is possible to download the legislation also in XML according the Normattiva legal XML standard, but:

1. The access is restricted to download of a single legislative act due to the use of a Captcha filter to prevent automatic download by crawlers. The likely reason is avoiding the information to be used by publishers, since they already enjoy a nearly monopolistic position.
2. The XML is generated from a representation of legislation as rows of a database table, so the quality is very poor.

*Score: Good.*

**Open data and licensing policy**

The use of those contents are permitted when is mentioned the source (via Permalink), the non-authenticity of text and the free use of service. Nothing is mentioned concerning legal open data policy.

*Score: Good.*

**Gazzetta Ufficiale (http://www.gazzettaufficiale.it/)**

This Gazette is the official legal source in Italy. It is published by the Printing Office and Mint State in Rome, with the collaboration of the Ministry of Justice and Ministry for economic, financial and budget policy.

Since 1st January 2013 is available the digital version for free online consultation for a period of 30 days. To access all contents, is required a subscription.

**Completeness**

The Official Gazette contains all acts of the Italian Parliament and Decrees of the President of Republic since 1946, all judgments and orders of the Constitutional Court, announcements and contracts with public administrations.

The digital version aims to provide an electronic version of acts, but it is non-authentic legal source.

The Official Gazette is sorted in two parts:

- Part I:
  - General section (published every day);
  - 1st Special section – Constitutional Court (published on Wednesday);
  - 2nd Special section – European Union (published Monday and Thursday);
  - 3rd Special section - Regions (published on Saturday);
  - 4th Special section – Announcements (published on Tuesday e Friday);
- 5th Special section – Contracts with Public Administrations (published on Monday, Wednesday and Friday).

- Part II:
  - Insertions (published on Tuesday, Thursday and Saturday).

The General Section in Part I include the same legal database of Normattiva.

*Score*: Good.

**Legal added value**

The free online consultation is available only on 30 days since the publishing of the act.

The full archive is available under subscription and it’s able to allow the search of documents with keywords, for years, via references.

The General section in Part I is update every day.

*Score*: Good.

**Legal links**

All contents have a “permanent link” (Permalink) with a webpage. It’s a stable connection always valid which allows you to access and view one full document via URL.

However in the plain text of acts there aren’t internal hyperlinks to other acts or legislations.

*Score*: Good.

**Data structure and formats**

Full archive is published online in HTML and, under subscription, it is also available in txt or pdf format for download.

For acts previous to 1986, the format is available only in pdf image.

There is no clarity regarding the possibility to download the data in XML format.

*Score*: Good.

**Open data and licensing policy**

The portal notifies that the official text of acts is printed on the press version of Gazette. The digital version is aimed to simplify the searching on legal context via web.

The use of those contents are permitted when is mentioned the source (via Permalink), the non-authenticity of text and the free use of service.

Nothing is mentioned concerning legal open data policy.

*Score*: Good.

**Arianna** ([http://arianna.consiglioregionale.piemonte.it/](http://arianna.consiglioregionale.piemonte.it/))

Arianna is the legal database of Piedmont Region. It contains texts of local laws, regulations and other legal acts promulgated by Regional Council.

Since 30th September 2013 the dataset is available for free consultation and reuse via dati.piemonte.it, the open data repository of Piedmont Region.

**Completeness**

The portal provides the online consultation of local laws and regional regulations promulgated since 1971.
Users can find the full text of acts, their drafting versions, Commission projects and the occurred abrogations.
The service covers all legal topics in Piedmont jurisdictional area.

*Score:* Very good.

**Legal added value**

Each act is available in coordinated form with chronological time line of changes and integrations. Users can choose the navigation criteria: years, topics, key words, Commissions.

The text is completed with tabs of metadata, like legal references, legal indicators and iter data.

The consultation of full texts is ever for free and the export in MSWord is allowed.

There are many other functionalities like “Virtual Dossier”, but at the moment the service does not work.

The system is subject to continuing updates and monitoring of data.

*Score:* Very good.

**Legal links**

The plain text of all acts are without hyperlinks to other legal references. However this links are available in tabs of metadata.

Arianna uses Normattiva to link its resources with national legislation, but it don’t uses URN to make accessible those data.

*Score:* Good.

**Data structure and formats**

All resources of Arianna are available in XML format or HTML format also via dati.piemonte.it portal, the open data repository of Piedmont Region. XML format is used for original texts, without changes and integrations; HTML is the format for all other acts after modifies.

The procedure to access XML format isn’t simple: to download the dataset is needed select the single act and view the source code in the browser.

*Score:* Very good.

**Open data and licensing policy**

The data-release for free reuse is under Creative Commons licensing ([CC BY SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)).

*Score:* Very good.

---

**Corte Costituzionale ([www.cortecostituzionale.it](http://www.cortecostituzionale.it))**

The Constitutional Court is the supreme court of Italy in charge on controversies on the constitutional legitimacy of laws and legal acts, on conflicts between powers of State, charges brought against the President of Republic.

Since January 2013 the new site has made available a legal open data section.

**Completeness**

The portal of Constitutional Court is very complete.

However, the legal area offers by this site is limited on constitutional case law and this represents an important restriction in data acquisition.
At this moment are available legal open data on:

- 18,000 judgments or orders (since 1956 to today)
- Full archive of principles (since 1956 to today)
- Registry of Judges (since 1956 to today)
- Laws and legal acts now pending in constitutional legal actions.

**Score:** Good.

**Legal added value**

The portal offers many added values, like the free search page to find case law through data input (key words, typology, legal references) and the monthly or weekly data monitoring.

Users can search case law by title, year, number and view full text plus legal references.

Each item is classified by judges (President and judge rapporteur), document metadata, topics.

**Score:** Good.

**Legal links**

All acts found online via search form are available with links to connected principles, legal references and list of similar results.

The legal references are linked with legal database of Normattiva. However, at the moment the service is not functioning.

There is not a direct link to recover one legal resource via URL.

**Score:** Good.

**Data structure and formats**

All acts online are available in HTML format.

Furthermore, the data are also provided in XML format, like described by DTDs published in open data section.

**Score:** Good.

**Open data and licensing policy**

The data-release for free reuse is under Creative Commons licensing ([CC BY SA 3.0](http://creativecommons.org/licenses/by-sa/3.0/)).

**Score:** Good.

**Consiglio di Stato** ([http://www.giustizia-amministrativa.it/](http://www.giustizia-amministrativa.it/))

The Consiglio di Stato and Regional Courts aim to protect the interests of citizens in respect to the public administrations.

Today the administrative tribunal in Italy has a small portal for judgements and other documents. It is not clear the extent of the coverage (probably since the late ‘80s and it covers also the regional level) and automated crawlers can access the data only after authorisation. Documents are published in HTML only.

However, since 27th November 2013 a new web site is available online, at the moment in beta-testing ([https://94.86.40.196/cdsintra/cdsintra/index.html](https://94.86.40.196/cdsintra/cdsintra/index.html)).

**Completeness**

In the new site are available all acts of Consiglio di Stato, Consiglio di Giustizia Amministrativa and documents from Sicily Region and Regional Courts.
At the moment, the database counts over 1,700,000 items.

**Score:** Very Good.

**Legal added value**

Users can use a new search engine to browse through decisions of the courts and other documents, including also papers and academic publishing.

For each item like decisions or judgements it is possible to verify the appeal acts.

The update of contents is daily and it is possible to refine the searching with concepts topic, document type, promulgation date.

**Score:** Good.

**Legal links**

Court decisions are not supplied with links to other documents or legal references.

However, all contents are accessible via URN:NIR.

**Score:** Fair.

**Data structure and formats**

The data available in new site are in HTML format. However, when users are searching a concept, in the plain text is active the underline option to identify the items.

**Score:** Good.

**Open data and licensing policy**

The portal is subject to big limitations:

- the reuse of data must be authorised;
- it is forbidden the use of spiders for contents downloading;
- the administration monitors the users activity to decide all possible legal action.

**Score:** Fair.

---

**Corte dei Conti** ([http://www.corteconti.it/](http://www.corteconti.it/))

The accounting tribunal has a similar service.\(^{57}\)

The Italian Corte dei conti is an Institution with the role of safeguarding public finance and guaranteeing the respect of jurisdictional system.

The Court pursues these two aims through two functions: the audit function and the jurisdictional function.

**Completeness**

The website is a single access point to two public databases of the Court.

The database of judgments, updated weekly, contains the measures taken by courts sections since 1\(^{st}\) January 2000. It contains also recovered significant decisions from previous years.

The database of the control and report, updated on a daily basis, collects, however, the measures taken by regional sections of Court since 1\(^{st}\) January 2009.

---

\(^{57}\) [https://servizi.corteconti.it/bds/doStart.do?pagina=semplice](https://servizi.corteconti.it/bds/doStart.do?pagina=semplice)
**Score:** Good.

**Legal added value**

Users can search acts in semantic databases: each piece of data is classified using legislative references, concepts topic, number of act, promulgation time.

Download of the contents is allowed.

**Score:** Good.

**Legal links**

Documents have not hyperlinks to other acts or legal references within the text.

However, in the detail page it is present one specific tab with link to "Linked documents".

**Score:** Good.

**Data structure and formats**

All acts available online are in HTML format and it is possible to export each item in MSWord.

**Score:** Good.

**Open data and licensing policy**

Acts and documents are available only for personal use and are subject to copyright.

The free reuse of data is allowed for informational purposes. The links from or to a webpage are subject to authorisation of Corte dei Conti.

**Score:** Fair.

**Corte di Cassazione (http://www.cortedicassazione.it/)**

The Corte di Cassazione manages a commercial service called ItalJureWeb\(^{58}\) with judgements of the court and of other courts but it is only meant for personal use and not for reuse.

**United Kingdom**

In recent years UK government has made significant efforts to improve access to legislation and case law and to open public legal materials for re-use. The launch of the portal legislation.gov.uk in July 2010 was a major event in the world of legal open data and an important contribution to the Web of Linked Data. However, the public services ensuring access to case law are still far behind those delivered by private commercial publishers or by non-governmental organisations.

The most important public case law portals are judiciary.gov.uk \(^{58}\) (http://www.judiciary.gov.uk – selected cases of predominantly after 2009), the web site of The Supreme Court (http://www.supremecourt.gov.uk – cases after 31.07.2009) and of the UK Parliament (House of Lord judgements from 14.11.1996 to 30.07.2009 – http://www.publications.parliament.uk/pa/ld/ljudgmt.htm). The case law published on these web sites is free for re-use under the conditions of the Crown copyright and the Open Government License. Much more extensive are the case law collections of the British and Irish Legal Information Institute (BAILII – http://www.bailii.org) and the Incorporated Council

---

\(^{58}\) [http://www.italgiure.giustizia.it/](http://www.italgiure.giustizia.it/)
of Law Reporting (ICLR – http://cases.iclr.co.uk). However, these are not open for data re-use.

Hereafter we will present the portals legislation.gov.uk and judiciary.gov.uk.

**Legislation Portal (http://www.legislation.gov.uk)**

Legislation.gov.uk is an on-line free accessible legal database which is managed by The National Archives on behalf of Her Majesty’s Government. The National Archives is a United Kingdom government department and an executive agency of the Ministry of Justice. The portal provides a comprehensive collection of UK legislation.

**Completeness**

The portal legislation.gov.uk covers the four jurisdictions that make up the United Kingdom (England, Scotland, Wales and Northern Ireland) and over 800 years of history. It comprises most (but not all) types of UK legislation and their accompanying explanatory documents. Most types of primary legislation (e.g. Acts, Measures, Orders in Council) are held in “revised” (consolidated) form which means that amendments made by subsequent legislation are incorporated into the text. Revised versions of some secondary legislative instruments (e.g. Statutory Instruments) are also published on the website. Legislation data covers:

- Public General Acts of the United Kingdom Parliament (1801 to date)
- Acts of the Scottish Parliament (1999 to date)
- Acts of the National Assembly for Wales (2012 to date)
- Measures of the National Assembly for Wales (2008 – 2011)
- Measures of the Northern Ireland Assembly (1974)
- Orders in Council made under the Northern Ireland Acts (1972 to date) (effectively the primary legislation for Northern Ireland under direct rule, though in the form of Statutory Instruments)
- Church of England Measures (1920 to date).

In addition, legislation.gov.uk provides also access to draft legislation.

**Score:** Very Good.

**Legal Added Value**

Legislation.gov.uk portal offers many value-added features like consolidation of legislation, classification of law by subject heading, year or type, document metadata, explanatory notes, expert annotations and translation of selected laws in Welsh language. The portal provides access to a comprehensive collection of consolidated versions of all British legislative acts in force with the option to access previous (historic) versions of them. For each document the user can choose a date and receive the full text of the act to this specific point in time. This
feature is presented as a timeline of the changes. Each legal act is supplied with valuable expert explanations and metadata. These inform users about specific conditions related to the entry into force of the document, associated documents, impact assessments, geographical extent and timeline of changes with textual amendments.

Users can search legislation by using simple and advanced search options. By simple search queries legislations is searched by type, title, year and number. The advanced searches include also options to search by key words within the document texts, to choose the language (English or Welsh), to give the range of years of publication, to define the geographical extent, to choose a specific point in time, to search in draft legislation only or within the information about impact assessment.

Score: Very Good.

**Legal links**

The well-established documents structure enables the provision of hypertext links at the level of each structural element – part, chapter, section or schedule. However, such links are not established within the text of the documents (in-line links), but in the supplementary expert annotations and explanatory notes about legislative changes. Even here the number of provided links is often insufficient. Legislation.gov.uk provides also links to external legal resources. For instance links to the House of Lords or to local authorities are established for some regulatory instruments. When EU legislation is cited, there is a link to EUR-Lex portal.

Score: Good.

**Data structure and formats**

Legislation.gov.uk is characterised by a well-established data structure. Legal documents are split into sections (the equivalent of articles, legal paragraphs) which form the basic logical information unit. For each legal paragraph legislation.gov.uk maintains a set of metadata. When opening a legal paragraph from the list, users can see the text and the metadata related to this paragraph. In addition there is an option to receive the full text of the legal act in its consolidated up-to-date version or to choose a date and to receive the consolidated historic version valid to this specific point in time (see above). The portal provides the content of legislation in XML format. The XML conforms to the Crown Legislation Markup Language (CLML) and associated schema. Legislation.gov.uk uses Dublin Core standard for metadata, XHTML for tables and MathML for formulae. Each document and legal paragraph can be exported / downloaded in PDF format.

Score: Very Good.

**Open data and licensing policy**

Legislation.gov.uk portal is definitely the flagship of the legal open data in Europe. The data contents of the portal are opened for free re-use subject to a few conditions described in the Open Government Licence.

Users are free to:

- copy, distribute and transmit the information;
- adapt the information;

---


exploit the information commercially whether by sub-licensing it, combining it or mixing with other information or data, or by including it in own product or application.

Users must:
- acknowledge the copyright and the source of the information
- include the same acknowledgment requirement in any sub-licences of the information
- ensure that user do not misrepresent the information or its source.

The re-users are provided with an API for downloading the contents of the portal. The data is available in XML, HTML, RDF/XML and ATOM format. Thus the portal is the first public legal data source in Europe offering its contents as linked data in RDF and ATOM format. To present moment, the RDF from legislation.gov.uk is limited to largely bibliographic information by making use of the Functional Requirements for Bibliographic Records (FRBR), the MetaLex vocabularies and the Dublin Core Terms.

Score: Very Good.

**Judiciary Portal** ([http://www.judiciary.gov.uk](http://www.judiciary.gov.uk))

Judiciary.gov.uk is a free on-line portal which is managed by Judicial Office Communications Team subject to the Her Majesty's Courts and Tribunals Service. The portal provides a variety of useful information about the judiciary system of England and Wales and an extensive collection of case law of the British courts and tribunals.

**Completeness**

The portal provides access to selected high profile judgments and particularly important precedents from the recent years. Court judgments are available for the most senior courts, typically Court of Appeal decisions. There are also many Tribunal decisions. Judgments are usually made available within 48 hours of the judgment being handed down. The database covers:

- **Judgments of the British courts since 2009** – Court of appeal (all divisions), High Court of Justice (Queen's Bench Division), House of Lords, Crown Court, Court of Protection and other courts in civil and criminal matters.
- **British Tribunal Decisions** – Information Tribunals, Employment Tribunals and Employment Appeal Tribunal, First-Tier Tribunal (all chambers), Upper Tribunals.

Score: Fair.

**Legal Added Value**

Judgements are provided with a summary and in full text. The entry point is a short notice in HTML format quoting the number of the case, names of the parties and date of pronouncement. Tribunal decisions are annotated by key words and phrases.

Score: Fair.

**Legal links**

No links to legislation or between judgements are available. There are some external links to Her Majesty's Courts and Tribunals Service and Land Registration division for some legal acts.

---

Score: Fair.

Data structure and formats
All judgments of the British courts can be viewed or exported/downloaded in their summary or full version in PDF format. On the other side, tribunal decisions are visible and downloadable in MS Word (DOC) and PDF format. User can search a court judgment by key words and a tribunal decision by subject matter, date, party and/or number.

Score: Good.

Open data and licensing policy
Judiciary.gov.uk is a legal data source open for use and re-use. Free re-use of the information published on the website is placed under the same Open Government Licence that was described in details for the portal legislation.gov.uk (see above). Hyperlinking to Judiciary portal is allowed, however, it is not permitted to open the linked web pages into frames. The Judiciary pages must be displayed in the user's entire browser window.

Score: Very Good.

Government Open Data Portal (http://data.gov.uk)
The portal was launched in January 2010 as a key part of the Government's work on Transparency which is being led by the Transparency and Open Data team in the Cabinet Office, working across government departments to ensure that data is released in a timely and accessible way. The work on the portal is being supported by the inventor of Internet Sir Tim-Berners Lee and Professor Nigel Shadbolt as well as by a number of technical partners among which the Comprehensive Knowledge Archive Network (CKAN), a project created by the Open Knowledge Foundation.

Currently the portal contains more than 13 570 published datasets, and more than 10 570 of them are available to data re-users under the Open Government License. Legal open data is presented on the portal with the legislation.gov.uk API and with SPARQL Endpoint provided for searching the linked data of legislation.gov.uk.

2.2.3 Public legal and open data portals of the EU
In this section we examine legal and open data portals operated by EU institutions which play a significant role in bridging EU and national legislation and case law – EUR-Lex, N-Lex and ECLI. More attention is paid to the EUR-Lex portal, which is considered in accordance with the evaluation criteria used in respect of national legal portals. Other portals with legal data, such as the web site of the EU Court of Justice in Luxembourg, the European Parliament, the Council of the EU, the European Commission and its Directorates-General, are not presented, because their legal contents are more or less their incorporated in EUR-Lex, being a single entry point to EU law and other documents considered to be public. Finally, the newly established open data portals of the Union are briefly reviewed.

EUR-Lex Portal (http://eur-lex.europa.eu)
Launched on 1st November 2004, the new EUR-Lex portal is the result of a merger of the former EUR-Lex site, free legal information service available online from 1998, with CELEX, a fee-based system providing additional value-added contents and functionalities. Since then the new portal is freely available to all EU citizens and businesses. EUR-Lex is operated by
the Publications Office in Luxembourg in cooperation with the Legal services of the European Commission, the Court of Justice of the EU and other EU institutions.

**Completeness**

The portal EUR-Lex is established as a single entry point to EU law. Thus it provides free online access to the most comprehensive EU legal database containing EU legislation and case law, international agreements and preparatory acts as well as other public documents considered to be of legal significance. In addition to this the portal supplies information about (but still not access to) national execution measures of the Member States transposing EU directives and national case law applying EU law. Currently, the contents of the site amount to some 2,815,000 documents in 24 languages with texts dating back to 1951. The database is updated daily and every year around 12,000 documents are added.

The various data collections (except Official Journal) are organised in sectors. The EUR-Lex portal comprises the following data collections:

- **Official Journal of the EU:**
  - scanned copies in PDF format from the first issue in December 1952 to 1997;
  - machine readable electronic copies in PDF format from 1998 to 30th June 2013;
  - authentic electronic versions from 1st July 2013 to date.

- **Treaties (Sector 1)** – original and consolidated versions of the founding treaties of the EU, accession treaties, other treaties and protocols.

- **International agreements (Sector 2):**
  - agreements concluded by the European Communities with non-member countries or with international organisations in their specific areas of responsibility;
  - agreements concluded jointly by the Member States and the European Communities in areas of shared responsibility (‘mixed type’ agreements);
  - decisions of joint committees set up pursuant to an international agreement and comprising representatives of the signatories for the purpose of administering the agreement.

- **Legislation (Sector 3)** – regulations, directives, decisions, opinions, recommendations and other acts of the institutions.

- **Consolidated legislation (Sector 0)** – consolidated up-to-date and historic (point-in-time) versions of the regulations, directives and decisions.

- **Complementary legislation (Sector 4)** – agreements between Member States, decisions of the representatives of the governments of the Member States and other acts.

- **Preparatory acts (Sector 5)** – legislative proposals and other acts of the European Commission (COM, SEC, JOIN and a. o. documents), legislative resolutions and other acts of the European Parliament, common positions or assents of the Council of the EU, opinions of the European Court of Auditors, the European Central Bank, the European Economic and Social Committee, the Committee of the Regions and other documents.

- **Case law of the Court of Justice of the EU (Sector 6)** – judgements, orders and decisions of the Court of Justice, the General Court and the Civil Service Tribunal, opinions of the Advocates-General and other documents.

- **National execution measures (Sector 7)** – lists of the national legislative and administrative acts adopted from the Member States to implement EU directives and
notified to the Commission. The information includes the name of the act(s) and textual reference (without hyperlink) to the publication in the official journal of the Member State.

- **National case law (Sector 8)** – metadata information about selected decisions of national courts in the field of EU law. Actually, Sector 8 of EUR-Lex currently houses the DEC.NAT database created by the Association of the Councils of State and Supreme Administrative Jurisdictions of the European Union. The database provides information about more than 27,300 decisions of predominantly administrative jurisdictions of the Member States.

- **Parliamentary questions (Sector 9)** – metadata information about written and oral questions and answers in the European Parliament.

- **EFTA documents (Sector E)** – agreements between Member States of the European Free Trade Association (EFTA), Acts of the EFTA Surveillance Authority and of the EFTA Standing Committee, Decisions, orders, consultative opinions of the EFTA Court and other documents.

- **Official Journal C Series documents (Sector C)** – documents published in the C Series of the Official Journal which do not fall within the above categories.

**Score**: Very Good.

**Legal Added Value**

EUR-Lex portal offers a number of value added contents and functionalities to its multi-million audience – consolidated legislation, detailed document metadata, case summaries, several document classification schemes, advanced search features, alerting service, multilingual interface, multilingual display of different language versions of the same document, export of documents in HTML and PDF formats, and others. Since March 2013, a new, further improved version of the portal – the new EUR-Lex, runs in parallel with the old one. The main additional functionalities of the new EUR-Lex are:

- A new design for easier navigation;
- “Google-like” quick search;
- View of the complete life cycle of a document throughout the legislative process;
- Personalisation of the system – home page, searches, result lists;
- RSS feeds based on saved search queries;
- Interaction with other users via the user community;
- Web services and API facilitating the free re-use of the information.

Consolidation in EUR-Lex is performed for the founding treaties and the most important legislative acts – regulations, directives and decisions. Consolidated texts are published within one month after the publication of the changes in legislation. Next to maintenance of the actual current versions of legislative acts, it includes provision of historic (point-in-time) versions.

Case law in EUR-Lex is published with concise case summaries (annotations) composed of key words and phrases. These summaries are placed in the “title and reference” part of the decisions allowing in this way the user to find case law by performing key word searches in the titles.

EU legislation is classified in EUR-Lex according to:

---

- Directory of the EU legislation – hierarchical analytical structure including 20 chapters (classification headings) with divisions into further sub-sections;
- Subject matter – alphabetical linear list of more than 150 legal terms;
- Eurovoc – terms from the multilingual and multidisciplinary thesaurus of the EU.

Subject matter and Eurovoc are used also for classification of the EU case law, but the latter is systematised also according to an own classification scheme known as Digest of the case law.

The portal offers also varied search facilities. The old EUR-Lex provides simple and advanced search forms. The simple search is actually quite sophisticated – by terms, date or time span, classification scheme, document number, document type, publication reference, a. o. Therefore the new EUR-Lex offers three search options: quick ("Google-like", really simple), advanced and expert search. Lists of search results can be ordered by date, classification headings, Eurovoc domain, CELEX sector, type of procedure.

The multilingualism of EUR-Lex is one of its great achievements. Documents are published in the 23 official languages of the Union, whereas the interface of the portal is maintained in 24 languages (the 23 official languages plus Gaelic).

Score: Very Good.

Legal Links

EUR-Lex provides an extensive net of metadata hyperlinks between its various collections of legal documents. The metadata links are available in a separate subsection of the so called "Bibliographic notice" 66 namely "Relationship between documents". For legislative acts this subsection includes:

- Legal basis – reference(s) to the provision(s) which is(are) the legal basis for the adoption of the current act;
- Select all documents based on this document – search function producing a list of documents which legal basis is the current act;
- Amendment to – linked list of the acts and their provisions which are amended (modified) by the current act;
- Amended by – linked list of the acts which amend (modify) the current act and its provisions;
- Consolidated versions – linked list of all consolidated historic versions of the act, incl. the latest actual version;
- Earlier related instruments – linked list of all earlier legislative proposals related to the current act;
- Subsequent related instruments – linked list of all subsequent legislative proposals related to the current act;
- Affected by case – linked list of all decisions of the Court of Justice of the EU which interpret or affect in other ways provisions of the current act;
- Display the national execution measures – link to a plain text list of the national execution measures notified by the Member States, but without external hyperlinks to the cited legal instruments published on portals of the Member States;

---

• *Instruments cited* – linked list of all legal documents and their provisions cited by the current act;

• *Select all documents mentioning this document* – a search function producing a list of documents which cite the current act and its provisions.

For judicial acts the subsection “Relationship between documents” includes:

• *Case affecting* – linked list of all legal instruments which are (which provisions are) interpreted or in other ways affected by the current court decision;

• *Instruments cited in case law* – linked list of all other legal instruments which are (which provisions are) cited (however not affected) by the current court decision;

• *Select all documents mentioning this document* – a search function producing a list of documents which cite the current court decision and its numbered paragraphs.

The metadata links consist of two parts: CELEX-number of the linked document with a hyperlink to it and an explanatory text indicating the type of the relation between the documents and to which structural parts of these documents the link refers. This means that links are established at document level, but the accompanying textual reference provides information about the referred structural parts of the documents (chapters, section, articles, paragraphs, etc.). Only links to the founding treaties are established at article level because of the document structure chosen for documents in Sector 1 of EUR-Lex, where each article is established as a separate document with its own metadata.

Here are some examples of links in EUR-Lex:

• *Interprets* 32004R0883 A87P8 – i.e. the current court decision interprets Article 87, Paragraph 8 of Regulation (EC) No 883/2004;

• 12008E045: N 1 37 46 – i.e. numbered paragraphs 1, 37 and 46 of the current court decision cite Article 45 of the Consolidated version of the Treaty on the Functioning of the European Union;

• *Amended by* 32005R0647 Amendment Article 3.1 from 05/05/2005 – i.e. the current act is amended by Regulation (EC) No 647/2005 and more specifically Article 3, Paragraph 1.

Legal links that are still missing in EUR-Lex are links between EU directives and the national execution measures listed in EUR-Lex, Sector 7, and between EU legislation and case law and the metadata of the selected national court decisions in EUR-Lex, Sector 8. The reason for the lack of such hyperlinks is that there are still no reliable means for establishing links between EU and national legal portals. The idea behind the introduction of the ELI and ECLI standards is namely to ensure a uniform technical infrastructure which will make possible the establishment of steady and reliable links between EU and national legal portals.

The EUR-Lex allows and welcomes the linking of document on the portal from external web sites. To this end a URI server is set up to make it possible to create logical links to the EUR-Lex portal. Detailed guidelines about the syntax of the links are published on: [http://eur-lex.europa.eu/en/tools/help_syntax.htm](http://eur-lex.europa.eu/en/tools/help_syntax.htm).

*Score: Very Good.*

**Data structure and formats**

Documents in EUR-Lex are provided in plain HTML text, but accompanied with very detailed metadata information, the so called “Bibliographic notice”. The text and the notice could be displayed separately or together depending on user’s choice. The metadata in the bibliographic notices comprises a variety of document properties and value added legal
characteristics which depend on the type of the document. For legislative acts, for instance, these metadata are systematised in the following rubrics:

- **Title and reference** – full title of the document and reference to the publication in Official Journal (OJ): series, number, date, pages, language versions, special editions of OJ;
- **Languages and formats** – in which languages and file formats is the document available;
- **Authentic language**;
- **Dates** – of document, of effect, entry into force, end of validity, of transposition;
- **Classifications** – Eurovoc descriptor, Directory code, Subject matter;
- **Miscellaneous information** – author, form, addressee, additional information;
- **Procedure** – procedure number, legislative history, link to the Legislative observatory of the European Parliament;
- **Relationship between documents** – see previous subsection about legal links.

All fields in document metadata are searchable and thus used by performing advanced and expert search queries.

Dating back to the old CELEX system, the predecessor of EUR-Lex, the HTML text is published as simple plain text without any structuring and formatting. Therefore, there is no structuring of legal documents in typical elements like parts, chapters, sections, articles, paragraphs, tables are not displayed correctly, images are missing and footnotes are not functioning as links. These shortcomings are already overcome in the new EUR-Lex which is expected to replace entirely the old portal in 2014.

More specific is the structure of the treaties in Sector 1 of EUR-Lex. Each article is available as a separate document with an own bibliographic notice. This splitting of the document text in articles is very useful when users would like to study the case law on a specific provision of the founding treaties. Of course, the treaties are also available in full text – basic enacted version and consequent consolidated versions.

Next to HTML format, documents in EUR-Lex can be obtained in PDF and TIFF format. The PDF file is a machine readable copy of the document as published in Official Journal, whereas the TIFF file is a scanned image of the document as printed in the paper edition of the Official Journal. TIFF files cannot be directly downloaded from the portal and are delivered per e-mail, because they are stored on an external archive.

Documents and bibliographic notices are available also in XML format for data re-users. By using URIs and implementing the Functional Requirements for Bibliographic Records (FRBR) the EUR-Lex portal is almost ready for a publication of the EUR-Lex data as linked data in XML/RDFa format in near future.

**Score**: Very Good.

**Open data and licensing policy**

Until 2013 the EUR-Lex data was opened for re-use:

- for free by downloading documents from the portal, or
- under paid license when the re-user would prefer to receive on an yearly subscription basis the data by initial bulk download from a FTP-server of the Publication Office and consequent download of weekly or monthly updates. The fee for the paid license was 10.000 Euro for receiving the data in one language version plus 3.000 Euro for each further language version.
As of 2014 the commercial and non-commercial re-use of EUR-Lex data is entirely free under a few conditions. Data re-users may choose one of the following two options:

- Webservice – in this case they have to register for a free WSDL webservice on the new EUR-Lex portal, to call the webservice via expert queries and to receive the data in XML format. However, some daily limits of the calls to the webservice have to be observed;
- License – the holder of the free license may deliver the licensed information as a product or service. Data in XML format is provided in bulk from a FTP server of the Publication Office and after that is updated weekly or monthly. The license holder is informed if new data is uploaded.

There are only a few conditions to be observed by data re-users. They have to give an appropriate copyright acknowledgement and a disclaimer when reproducing legal acts which states that only documents published in the Official Journal of the European Union are deemed authentic, respectively only decisions of the Court of Justice published in the “European Court Reports” are deemed official sources.

Score: Very Good.

N-Lex Portal (http://eur-lex.europa.eu/n-lex)

N-Lex was launched on 28 April 2006 as a common gateway to national law of EU Member States. In fact, the portal offers only a common search interface that enables users to make search queries in their own language by using terms of the multilingual Eurovoc thesaurus. The queries are sent to the national legal portals which return the search results to N-Lex. The result list provides links to the legal instruments containing the search terms. By following the links users are either directed to view the documents on the national legal portal or text of the document is sent by the national portal to EUR-Lex and viewed there. The communication between N-Lex interface and national legal databases is performed through the so called N-Lex connector.

To date there are still no technical means to establish links between EUR-Lex and N-Lex or between EUR-Lex and national legal portals in order to provide cross-links between EU directives and the national execution measures adopted by the Member States. The idea behind the recently elaborated ELI standard[^67] is namely to allow such linking between EU and national legislation.

ECLI portal

The ECLI portal, being part of the European e-Justice portal, is expected to be launched in 2014 as a common gateway to national case law. Similar to the N-Lex portal it will follow a decentralised approach, i.e. no database on European level will be built. The portal should offer only a common search interface for searching national case law by ECLI identifier and (some of) the metadata as defined in the ECLI standard.[^68]


[^68]: See Section 5 of Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law: [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52011XG0429%2801%29:EN:NOT](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52011XG0429%2801%29:EN:NOT)
Pan-European data portal ([http://publicdata.eu](http://publicdata.eu))

The Pan-European data portal PublicData.eu provides a single point of access to official open datasets from local, regional and national public bodies across Europe. The portal was developed by the Open Knowledge Foundation as part of the LOD2 project[^69] funded by the EC’s 7th Framework Programme. The vision of PublicData.eu is to harvest and federate the information about different national data catalogues, open data portals and institutional websites, thus to enable users to search, query, process, cache and perform other automated tasks on the data from a single place. Currently, the portal provides access to more than 29,000 datasets from 14 European countries. Legal data is presented on the portal with the UK legislation API.


The European Union Open Data Portal was launched in December 2012. It is established as a single entry point to the growing range of data from the institutions and other bodies of the European Union (EU). Data are free for use and re-use for commercial or non-commercial purposes. As a general principle, data can be re-used free of charge, provided that the source is acknowledged.

### 2.2.4 Open Access journals

According to Peter Suber, open access journals are scholarly journals that are available online to the reader without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself[^70]. The first digital-only, free journals were published on the Internet in the late 1980s. Greater spread, however, open-access journals received in the first decade of this century when the Directory of Open Access Journals (2003) and the Open Access Scholarly Publishers Association[^71] (2008) were founded.

The most authoritative portal providing facilities to search open access journals and articles is the Directory of Open Access Journals (DOAJ): [http://www.doaj.org](http://www.doaj.org). Currently, it houses more than 9,800 journals from 124 countries with more than 1,570,000 articles.

A detailed list with open access journals classified by subject is available on [http://www.oajse.com/index.HTML](http://www.oajse.com/index.HTML). The research performed by EUCases partners Apis and Averbis has proved that on the above sites can be found a number of open access articles in the field of EU law (unfortunately, not too much) and that there are no bans imposed for creating links to these articles.

[^69]: [http://lod2.eu](http://lod2.eu)
[^70]: Suber, Peter. Open Access Overview: [http://legacy.earlham.edu/~peters/fos/overview.htm](http://legacy.earlham.edu/~peters/fos/overview.htm)
[^71]: [http://oaspa.org](http://oaspa.org)
2.3 Pan-European legal information services

In this section we examine existing Pan-European legal information services that aim at provision of legal resources from various Member States linked to EU law. This means that we focus only at cross-border services which 1) offer legal contents from at least two Member States, and 2) these contents are closely related to the application of EU law. Outside the scope of the study remain purely national suppliers, despite the fact that they provide information about EU legislation and case law.

2.3.1 Non-commercial services

Altogether five significant non-commercial cross-border legal information services have been identified during the study performed by the project partners. As a rule, these have been developed either by the European Commission or by non-profit supranational organisations with the financial support or direct involvement of EU institutions.

DEC.NAT Database

DEC.NAT Database\(^{72}\) (from French: “Décisions nationales”) is a collection of metadata information and references to national court decisions of the Member States on the application of EU law available as an online service on the web site of the Association of the Councils of State and Supreme Administrative Jurisdictions of the European Union i.n.p.a. (ACA-Europe). The references and the juridical analysis of the national decisions have been supplied by the Research and Documentation Service of the Court of Justice of the European Union. Following an agreement between the Court of Justice and the ACA-Europe, this information was passed on to the association, which has developed an interface allowing the public to consult the data over the Internet using a number of search tools. The database comprises metadata in English and French (but not the full original texts) of more than 26 600 decisions rendered since 1959. It has been lastly brought up to date on 4\(^{th}\) December 2012. For each decision, DEC.NAT contains:

- national references – the names of the parties, an indication of the national provisions applied and the relevant doctrine;
- international or European references with an indication of the Community provisions applied;
- an analysis of the decision – keywords, i.e. a summary of the subject of the decision;
- in some cases, a fourth section indicating the preliminary ruling of the Court of Justice.

Presently, DEC.NAT database is an integral part of ‘Sector 8’ of the EUR-Lex service (http://eur-lex.europa.eu/RECH_jurisprudence-nat.do) and is updated regularly with new cases (see above the section about the EUR-Lex portal). In the end of 2013 the database contains more than references to 27 300 decisions, incl. a limited number of decisions of the Supreme Court of the United States, the European Court of Human Rights and the EFTA Court. The only shortcomings of this high quality value added service is the fact that the database does neither offer the full text of the decisions nor links to national sources where these texts can be consulted.

**JuriFast Database**

JuriFast Database\(^{73}\) includes the original full texts supplied with metadata and summaries in English and French of more than 1570 national court decisions. The aim of this online service is to offer immediate access to the latest developments in national case law which is linked to EU law.

As in the case with the DEC.NAT database, the search interface of the online service was also developed and has been further updated by ACA-Europe. Most of the decisions have been judged on the grounds of a preceding preliminary ruling of the Court of Justice and are supplied in a “preliminary files” together with the preliminary questions submitted to the Court of Justice and the Court’s answer to these questions. The rest part of the JuriFast database is formed by decisions of the national jurisdictions which are not connected with preliminary proceedings, but nonetheless are acts on the application of EU law. The decisions are directly supplied by the research and documentation services of the ACA-Europe's member institutions. For each decision, the database contains:

- the date and a short description of the subject;
- the full text of the ruling or judgement;
- a summary in French and English;
- a link to the relevant provisions of European law.

In the case of preliminary questions, additional information is added in due course, namely:

- a link to the response of the Court of Justice;
- the decision resulting from this response.

**JURE database**

JURE Database\(^{74}\) (the abbreviation comes from **JUrisdiction, Recognition, Enf orcement**) was developed and is further maintained by the European Commission. It contains case law of the Court of Justice of the EU and of the Member States’ courts on the interpretation of the EU legislation in the field of international private law: Council Regulation (EC) No 44/2001 (Brussels I Regulation), Regulation (EC) № 2201/2003 (new Brussels II Regulation), Regulation (EC) № 1347/2000 (Brussels II Regulation), the 1968 Brussels Convention and the 1988 Lugano Convention.

The database is constantly updated through contributions made by experts in this field and reviewed by the European Commission. However, the updates after 2009 are published in the JURE subdomain of Sector 8 (national case law) on the new EUR-Lex portal\(^{75}\) only. Therefore, the up-to-date collection on the New EUR-Lex contains more than 5740 decisions (around 500 of them are judgements of the Court of Justice and opinions of Advocates-General, and 5240 are national court decisions) whereas the search interface of the JURE Database offers access to about 140 decisions of the Court of Justice and 2400 decisions of national courts (supreme courts, courts of appeal and first instance courts of the Member States). Next to metadata in English and French (national case number, year, country, court name, court type, language, parties, publication reference, related legal instrument and


\(^{75}\) See the advanced search form on [http://new.eur-lex.europa.eu/advanced-search-form.HTML](http://new.eur-lex.europa.eu/advanced-search-form.HTML) and the “JURE” subdomain option from the “Domain selection” drop down box.
referred provisions), JURE database contains key terms and summaries of the cases in English, French and German as well as in the original language of the judgment.

**Common Portal of National Case Law**

The web site of the Network of the Presidents of the Supreme Judicial Courts of the European Union provides a service named "Common Portal of National Case Law", in fact a meta-search engine which enables users to search in the national case law databases through a common search interface. This project began in June 2006. It is supported by the European Commission and is monitored by the Legal Information Group of the European Council. In July 2008, the first version of the Portal has been released. It ensured a privileged access to all the case law databases (with login and password) for national Supreme Courts, the European Commission and the Court of Justice of the European Union. The Portal is also accessible to the general public but restricted to the free case law databases.

The search can be done in twenty languages: Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Lithuanian, Norwegian, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish. The user can enter key words in one these languages. However, search can be done in most 5 case law databases chosen by the user from the “List of search engines available in the Portal”. The query is first translated by the translation Google database, Eurovoc and the EU’s multilingual term database IATE (InterActive Terminology for Europe). Therefore, users are advised to use common keywords and to avoid words in plural in order to have a higher probability that the chosen keywords will be translated correctly by Google, Eurovoc or IATE. Users can modify the proposed translation. The list of the results can be ordered by country, relevance and date. For each decision in the list user can see the following data: country, source, title, date (if available), summary (if available), link to full text. The text is available in the original language. However, registered users can obtain machine translation of the text by using the DGT (Directorate-General for Translation) online translation database where decisions are available in HTML format.

**FRA Case Law Database**

FRA Case Law Database is a case law collection of the Fundamental Rights Agency, containing approximately 100 decisions of the Court of Justice of the EU with a direct reference to the Fundamental Rights Charter and more than 500 decisions forming national leading case law from all EU member states in the area of anti-discrimination law. The data include a formal reference to the decisions, judgments and opinions, an abstract in English, and if available a link to the full text document.

**EC Consumer Law Compendium (http://www.eu-consumer-law.org)**

www.eu-consumer-law.org is a web site providing free access to expert information on eight consumer law directives, their transposition into the laws of 27 EU Member States, including case-law, bibliography and a comparative study. The database is an output of a

---

research project called "EC Consumer Law Compendium", which is being conducted by an international research group on behalf of the European Commission since 2004. The website contents cover eight directives, namely:

- the Doorstep Selling Directive 85/577
- the Package Travel Directive 90/314
- the Unfair Contract Terms Directive 93/13
- the Timeshare Directive, 2008/122/EC (according to the prevailing state of transposition in the Member States, otherwise Directive 94/47/EC)
- the Distance Selling Directive 97/7
- the Price Indication Directive 98/6
- the Injunctions Directive 98/27 and
- the Consumer Sales Directive 99/44.

The EU materials and the annotations made by the contributors to the database are provided in English, French and German languages. Case law about the above listed directives can be found by directive, date, court type, keyword or a Member State. In January 2008 a follow-up project called "Update of the EU Consumer Law Acquis Database" started to keep the database up-to-date. In 2008 more than 130 acts and statutes have been updated. Moreover, about 150 new judgements have been collected and inserted into the database. The follow-up project was completed in the first quarter of 2012. Therefore the Database contains information on the transposition of the above Directives into all 27 EU Member States as of March 2012.

**LATC Project**

The only EU funded project aiming at publishing legal open data as linked data that was identified by the present state-of-the-art study is the LATC project (Linked Open Data Around-The-Clock).\(^{80}\) LATC is a support action funded under the European Commission FP7 ICT Work Programme, within the Intelligent Information Management objective (ICT-2009.4.3). One of the tasks of the project (work package 2, task 2) included publication of four legal datasets produced by European institutions as linked data on the Web – N-Lex (only German legislation dataset), EUR-Lex, EPO and Pre-Lex. The data is published in RDF format on the following sites: [http://n-lex.publicdata.eu](http://n-lex.publicdata.eu), [http://eur-lex.publicdata.eu](http://eur-lex.publicdata.eu), [http://epo.publicdata.eu](http://epo.publicdata.eu) and [http://prelex.publicdata.eu](http://prelex.publicdata.eu). The datasets are interlinked with DBpedia resources.

**2.3.2 Commercial services**

The huge investment costs for digitalisation of legal archive materials, data processing, consolidation, case summarisation, etc. of national legislation and case law have prevented the development of multilingual and cross-border legal information services in the early years of the creation of legal databases. The first viable international commercial services in this field emerged in the 1990s and the first decade of 21st century simultaneously with the imposition of the Internet as a global medium. In the common law countries their appearance was connected with the rise of the “Big Three” – Thomson Reuters (Westlaw), Reed Elsevier

---

80 [http://latc-project.eu](http://latc-project.eu).
(LexisNexis) and Wolters Kluwer who created a variety of cross-border legal information products and services for the English-speaking world. In Europe, additional impetus for developing such contents was given with the establishment of the European Union and increasingly intense process of harmonization of national law of Member States with EU law.

Nowadays, there are a number of commercial Pan-European cross-border legal databases. Next to the multinational legal contents of the Big Three, some other innovative international services like vLex and Darts-IP have gained success in the recent years. In the following we present them briefly.

**Thomson Reuters**

Thomson Reuters is a multinational corporation created by the purchase of the British-based Reuters Group from the Canadian content provider Thomson Corporation in 2008. The company is operating in more than 100 countries, and has more than 60 000 employees. One of the five divisions of the firm is in the field of legal information provision, famous with the world-leading brand Westlaw, an online legal research services ensuring access to more than 40 000 databases with U.S. state and federal statutes, case law, administrative codes, public records, law journals, legal forms and other value added information.

In Europe Thomson Reuters is operating through Ellis Publications based in the Netherlands, Sweet & Maxwell and Lawtel in the United Kingdom and Thomson Reuters Round Hall in Ireland.

**Ellis Publications** is claimed to be the global leader in EU law. It provides online services offering EU law materials in English (OJ OnlinePlus), Dutch (EU Recht OnlinePlus) and French (Droit UE Online). However, the service which is combining cross-border legal contents with EU relevance from the Member States is “EU Competition Law Online”. This service (in English only) provides access to full official EU competition materials and the leading commentary modules from Sweet & Maxwell including EU Competition Law Handbook and Rowley & Baker’s International Mergers – The Antitrust Process. The Handbook is actually an expert guide to EU and national competition law cases. The Rowley & Baker module, known as the ‘Bible’ for international antitrust lawyers contains expert explanations of merger regulations in nearly 50 countries.

Another product with cross-border legal contents is “European Current Law” provided by Sweet & Maxwell. It covers European Union case law, legislation and its implementation by Member States. It also reports on the most important national and international material from all other European countries, both East and West. Extensive coverage provided by a team of practising lawyers and academics monitoring over 120 published sources each month. However, the product is available only on paper – as a monthly journal or a hardback. The same is valid also for EU the Competition Law Handbook (see previous paragraph) and the Fleet Street Reports: Cases on Intellectual Property Law.

The most prominent service of Thomson Reuters offering cross-border legal information from various jurisdictions in the world is Westlaw International. This online service provides simultaneous access to the tens of thousands databases of Thomson Reuters’ subsidiaries throughout the world. Legislation, case law, law reviews, treaties, news and business information from US, Canada, EU, UK, Hong Kong and Australia are combined in a global library. European legal contents are presented by the legal resources of Sweet & Maxwell

---


82 [http://westlawinternational.com](http://westlawinternational.com).
and Ellis Publications, including the above discussed cross-border legal information services and products.

**Reed Elsevier**

Reed Elsevier is a publisher and information provider operating in the science, medical, legal, risk and business sectors with headquarters in London and Amsterdam. The company is operating in more than 30 countries, and has more than 30 000 employees. Its most famous brand in the field of legal information provision is LexisNexis. In Europe LexisNexis is presented with subsidiaries in Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Luxembourg, the Netherlands, Norway, Poland, Russia, Spain, Sweden, Switzerland, Turkey and United Kingdom. Despite this significant presence in Europe LexisNexis does not offer any substantial cross-border legal contents. Its subsidiaries in the respective countries remain keen on offering only national and (in most cases) EU legal contents from EUR-Lex.

There are only two services ensuring access to European multinational legal contents: EU Tracker and Juris Classeur.

**EU Tracker** is a comprehensive web-based product that tracks implementation of key European Union directives across 12 practice areas. The tracking includes monitoring by “Traffic light” and alerting system of the implementation, coupled with analysis by a team of EU law specialists from UK, Germany, France, Spain and Poland. The expert contents deliver a general overview of a Directive plus specific country information focused on key date-by-date analysis charting progress towards implementation.

**Juris Classeur** is a French-language legal research online database including legislation, jurisprudence and journals from France, Belgium and Luxembourg. The service provides access to more than 35 000 laws and regulations, 1.2 million decisions and more than 130 journals with 150 000 bibliographic records. However, EU legislation and case law is not included. The product is also available in print and on CD-ROM.

**Wolters Kluwer**

**Wolters Kluwer** is a global information services and publishing company based in the Netherlands with operations across Europe, North America, Asia Pacific and Latin America and employing more than 18 000 people worldwide. The company provides products and services for professionals in the health, tax, accounting, corporate, financial services, legal and regulatory sectors.

In the field of legal information services Wolters Kluwer has an impressive presence in Europe – the Netherlands, Germany, France, Italy, Belgium, Spain, UK (tax and accounting regulations only), Poland, Czech Republic, Hungary. The company has developed also a number of products and services offering international legal contents. Most of them are produced by Kluwer Law International, part of Wolters Kluwer Law & Business. They offered through the Kluwer Law Online platform for journals and looseleafs/manuals. The paid online access is ensured on a subscription base or at journal article / looseleaf chapter level provided as searchable PDFs. **Kluwer Law Online Manuals** contain authoritative and up-to-date country-by-country overviews in a specific area of European / international law. Here we

---

will list only some of them: European Air Law, European Direct Taxation, Handbook of EU VAT Legislation, European Environmental Law, Customs Law of the EU, IEL (International Encyclopedia for) Private International Law, IEL Competition Law, IEL Intellectual Property, IEL Commercial and Economic Law, a. o. Kluwer Law Online Journals offer users who are not subscribers quick and easy to browse the journal contents and to purchase individual articles. The articles provide expert commentaries and surveys of EU and national legislation and case law in a specific legal area addressed by the respective journal. The most important journals connected with EU law and its implementation in the Member States are: EC Tax Review, European Business Law Review, European Company Law, European Energy and Environmental Law Review, European Foreign Affairs Review, European Public Law and European Review of Private Law.

Specialised online platforms are developed in the field of international commercial arbitration, international intellectual property law and international competition law.

Finally, we can conclude that the despite its presence in most EU Member States Wolters Kluwer does not offer a common database with national legislation, case law and other legal materials. The elaborated legal products and services in the field of European and international law are based predominantly on research materials and expert overviews written by authoritative legal correspondents from the respective countries, and therefore contain only a limited selection of national case law digests.

**vLex**

vLex is a Spanish company with headquarters in Barcelona founded in 1998. Based on legal materials from the Spanish-speaking community, soon it turned into a global legal information provider. Currently, the online search platform vLex offers legal contents from more than 130 countries and 1140 publishers around the globe. Next to the usual for legal information providers production of own value added legal contents, the business model of vLex includes direct integration of official public resources and conclusion of licensing agreements with legal publishers for revenue sharing. Thus its database covers legislation, case law, books, journals, articles, forms, contracts, news and other legal materials from various national and international jurisdictions (e.g. EU legislation and case law).

Despite the impressive global content coverage, incl. EU legal materials, the vLex service reveals some serious gaps when it is examined for completeness country by country. For instance, the French case law is presented only by decisions of the Supreme Court of Cassation, the German – by cases of the Federal Constitutional Court, the Austrian – by decisions of the Constitutional Court and the Supreme Court. For some countries, such as Bulgaria, there is even no case law included. Another disadvantage of the service is the lack of legal links between connected data resources.

**Darts-ip**

Darts-ip is a global intellectual property case law online database maintained by the homonymous company based in Brussels. It is updated daily and currently has over 1 350 000 cases from over 2 600 courts worldwide. Each individual decision is analysed locally by

---

trade mark, patent, design or domain name specialists and classified according to a “points of law” hierarchical classification structure. Cases are collected predominantly from European jurisdictions, but also from the United States, Brazil, South Africa, Australia, India, China, Hong Kong, Japan and South Korea. The European contents cover case law of the Court of Justice of the EU and of most national jurisdictions as follows:

- Trademark focused cases – over 307,000 decisions from 722 courts in 32 European countries;
- Patent focused cases – over 155,800 decisions from 422 courts in 23 European countries;
- Design focused cases – over 9,900 decisions from 355 courts in 26 European countries;
- Domain names focused cases – over 30,800 decisions from 349 courts in 31 European countries.

The success of the Darts-ip service was recently recognised by Thomson Reuters. In May 2012 Thomson CompuMark, part of the Intellectual Property & Science business of Thomson Reuters and one of the global leaders in trademark searching and brand protection solutions, announced a partnership with Darts-ip which will allow the customers of the SERION Search Analysis and Watch tools of Thomson CompuMark to conveniently access case law information of Darts-ip as a seamless part of their workflow.\(^{90}\)

### 2.4 Concluding remarks

The performed research on public legal and open data portals in the EU and in the six project Member States clearly shows the following two trends in the current development of legal open data in Europe:

- Improvement of the quality of public access to legislation and case law by continuous enrichment of contents and functionalities of public legal databases;
- Gradual elimination of impediments for the re-use of public legal data and the introduction of open data standards and formats.

The **first trend** is characterised by the efforts of public institutions to ensure **completeness** and to **add value** to legal data. With the exception of Bulgaria, in all other examined jurisdictions legislation on public legal portals is being offered not only in the officially enacted, but also in consolidated form. However, in Germany, consolidation process does not include historic (point in time) versions of legislative acts. In comparison with legislation data, public case law databases are not fully complete. The efforts are focused on covering all decisions of the national supreme courts from the recent years which are available in machine readable format. Case law of lower instance courts is presented mainly by collections of selected “most important” decisions. Only in Bulgaria case law of all courts is published on a full scale. The establishment of **legal links** has still not become a priority and can be observed mainly in legislation databases (RIS, Legifrance, legislation.gov.uk). Only the EUR-Lex portal provides a comprehensive net of metadata links between various types of legal documents, including the highly valuable links between legislation and case law.

The **second trend** is characterised by the gradual introduction of:

---

- Legal XML standards (RIS, Legifrance, Gesetze im Internet, Normattiva, legislation.gov.uk and EUR-Lex);
- FRBR conceptual entity-relationship model (EUR-Lex and legislation.gov.uk);
- URN, URI or other schemes allowing for establishing persistent external links to the published legal documents (RIS, Legifrance, Gesetze im Internet, Normattiva, legislation.gov.uk and EUR-Lex);
- ECLI standard (Legifrance, EUR-Lex);
- Free licenses for download and re-use of legal data (RIS, Legifrance, Gesetze im Internet, Normattiva, legislation.gov.uk and EUR-Lex);
- Reduction (RIS, Legifrance), abolishment (EUR-Lex) or refusal (legislation.gov.uk) of fee collection for paid data delivery services allowing bulk data download and regular updates in XML or other open data formats for data re-users;
- Publication of legal databases as open data (RIS, Gesetze im Internet, legislation.gov.uk) on national public open data portals.

The growing willingness of public bodies to open their legal resources for free re-use was observed even in the course of the present state-of-the-art study (October – December 2013) during which period the Publication Office officially announced the abolishment of the paid license for EUR-Lex as of 2014, Legifrance and RIS reduced the annual subscription prices for XML delivery services in 2014, and, finally, Legifrance portal communicated the authorisation of data re-users to download free of charge the HTML documents published on the web site. However, the open data community is still waiting for further good news connected with legal data, especially with regards to the restrictive conditions for re-using public case law resources in Germany and Italy. The imposed fees in Germany have a prohibitive effect, and in Italy only the portal of the supreme administrative jurisdiction – Consiglio di Stato, is opened for re-use. In the UK the content coverage of the case law published on the Judiciary portal is still insufficient in comparison with the databases of the British and Irish Legal Information Institute (BAILII).

The performed research on the public legal portals makes it possible to compare the developments in the six project Member States and in the EU from the point of view of the assessment criteria set out in section “Scope and assessment criteria” above. Although the arbitrary and rather subjective character of the examination narrowly focused on the project objectives, the following table with the portals, scores given for each criterion and the total scores give us an interesting comparative overview:

<table>
<thead>
<tr>
<th>Legal Portal (*)</th>
<th>Completeness</th>
<th>Legal Added Value</th>
<th>Legal Links</th>
<th>Data Structure &amp; Formats</th>
<th>Open Data &amp; Licensing Policy</th>
<th>Total Score (**)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rechtsinformationssystem (AT)</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>State Gazette Portal (BG)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Case Law Portal of the SJC (BG)</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Case Law Portals of Supreme Courts + Constitutional Court (BG)</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Legifrance (FR)</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Gesetze im Internet (DE)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Case Law Portals of Federal Supreme Courts and of Federal Constitutional Court (DE)</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Normattiva (IT)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Consiglio di Stato (IT)</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Corte Costituzionale</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>
The comparison clearly shows the leading position of the public legal portals of Austria, France, UK and EU followed – with confident steps – by those of Italy and Germany. Bulgaria as one of the few countries in Europe providing access only to the legal contents of its official journal web pages is at the bottom of the list. The insufficient developments of the case law portals in comparison with the legislative ones also become obvious.

Based on the overview and analysis in the preceding sections, the following conclusions from the point of view of EUCases objectives and future work can be drawn:

- Public legal resources in Europe are becoming more and more open for use and re-use. Thus, no hindrances in obtaining legal data from the portals initially listed in the Description of Work have been identified during the performed research.
- The project budget for data licenses and fees still needs to be clarified. The sum of 10 000 EUR that was initially foreseen for EUR-Lex license could be spent for national licenses, especially for downloading XML data in bulk from Legifrance portal. However, the new prices of the subscription service will be announced in the first quarter of 2014.
- The requirement for prior written consent and the high fees for commercial re-use of the German case law will not affect the project work, because no commercial exploitation is foreseen. However, the future plans of the SME-partners to exploit commercially the project results could be seriously jeopardised if this licensing policy will not be changed during project execution. Therefore the consortium will not expand the case law coverage beyond the five supreme German courts as foreseen in the Description of Work. Only cases of the Constitutional Court will be processed additionally.
- UK case law also cannot be covered fully, because the public sources do not provide full content coverage for old cases.
- The conversion of the court decisions of some German and Bulgarian courts from PDF format to Legal XML will cause a number of problems for their automated processing by LT2XML converters to be developed within WP2. In these cases some additional data processing and data curation of the texts by qualified operators have to be foreseen.
- Legal data (except on the EUR-Lex portal) is published in most cases without any or with insignificant number of links. Case law is not linked to legislation as a rule. The same is also valid for the establishment of hyperlinks between national and European legal contents. The research has confirmed the existing gap and the need for linking national with EU legislation and case law.
- The ELI and ECLI standards outline the way to the legal linked open data and the future Legal Semantic Web. It could be expected that, like France, some other project countries will introduce these standards in the next two years, thus indirectly facilitating the linking tasks foreseen within the project. But even without having these standards introduced, the linking of legal data is aided by the URN, URI or other schemes implemented by most legal portals allowing the establishing of persistent external links to their legal resources.
Despite the fact that there are some commercial and non-commercial services linking EU and national legislation and case law, there still does not exist a service which fully implements the idea of the EUCases project.
3 International standards in the legal information field

3.1 Legal XML

3.1.1 Introduction

XML (eXtensible Markup Language) is a markup language developed by W3C as an open standard renown for representing data and documents in a manner which is both human-readable and machine readable. It is therefore characterised by strong features such as hierarchies, extensibility, accuracy and flexibility. This empowers any fraternity to develop its own vocabulary of terms depending on the problem at hand. Consequently, its innovatory features lies in its use of a lexicon, syntax and grammar that defines its rules. These rules then define the behaviour of a tag to proscribe any deviation by the user e.g. that all sections should be numbered.

From the foregoing, Legal XML refers to technical standards for structuring legal documents and information using XML and related technologies. This extends XML technology to the documents, data systems, and information exchanges in the legal domain consisting of legislatures, attorneys, publishers, court administrators, IT staff, academics, criminal justice agencies, companies, legislatures etc. Many domain standards have been deployed thus far in this regard including standards for legislative drafting, court filing, eContracts, eNotary, ODR, logic rules, integrated justice etc. The rationale is that effective standards are a prerequisite for extensive development and deployment of information exchange technology.

To summarise the foregoing, below are the key benefits of Legal XML.

- open-source for data and documents;
- machine readable;
- technology-neutral;
- applicable to identify the purpose and content of the information (beyond HTML);
- hierarchical, rigorous, extensible, accurate and readable; and
- guarantees the long term preservation and accessibility of the document.

3.1.2 Legislative XML

Legislative XML, which is currently used across many jurisdictions, is an icon of great success in legal informatics research. Its added advantages in this context include:

- links between norms;
- better search capabilities;
- multiple platforms;
- dialogue between systems;

---

91 http://www.w3.org/XML/
• diverse applications; and
• reuse of information.

Nevertheless, not all XML standards are suitable for representing legislative or parliamentary
documents. Legislative XML standards are specifically built to depict the complete legal
document to avail the full potential of XML to refine the legislative process.\textsuperscript{93} In this regard, a
suitable XML standard needs to feature the following:

• include information that contributes to the direction of workflow;
• allow for interoperability with other institutions and resources;
• promote collaboration e.g. by citizens in Open Government and Open Data initiatives;
• preserve legal validity over long periods of time;
• permit accessibility through multiple channels;
• allow for inspection by anyone while balancing privacy and security;
• ensure dissemination and usability;
• integrate effectively with common tools and document management systems;
• ensure compatibility with other W3C standards to safeguard robustness over time;
• use abstract models as the basis for all other tags to allow for easier maintenance
  and expansion;
• include a robust naming model for legal resources on the Web to withstand
  technological changes;
• Differentiate official authorship from the publisher and content from the ontological
descriptions of concepts.

3.1.3 Existing standards

There are many XML standards that have been applied to the law for different purposes
following the above criteria. For instance, formalised Exchange of Electronic Publications
(Formex) was created for web publishing and therefore does not distinguish between
representation and presentation of content. Some standards such as Denmark’s LexDania,
Italy’s NormeInRete (NiR), Switzerland’s ChLeXML, Austria’s eLaw are specific to one
country and are difficult to adapt to other legal systems. On the other hand, the US House of
Representatives XML standard does not differentiate between procedural descriptions and
the basic representation of the document’s content. In a similar vein, at its Library of
Congress, Chile has embraced an XML standard similar in some respects to Akoma Ntoso
discussed below. However, it only models the normative document in its final phase
(enactment) thereby disregarding the entire legislative procedure.

Although the foregoing country-specific implementations model their legislatures distinct
characteristics, parliamentary features are quite commonly shared in regard to actors,
structures, procedures, documents, and information. Accordingly, Europe has the Metalex
interchange format\textsuperscript{94} while the Akoma Ntoso Legislative XML standard\textsuperscript{95} is sufficiently flexible

\textsuperscript{93} Palmirani et al., Legislative XML: Principles and Technical Tools, 2012. \url{http://ideas.repec.org/p/idb/brikps/71358.HTML}.
\textsuperscript{94} \url{http://www.metalex.eu/}.
\textsuperscript{95} M. Palmirani. Legislative change management with akoma- ntoso. In G. Sartor, M. Palmirani, E. Francesconi, and M.A.
Blasiotti, editors, Legislative XML for the Semantic Web, volume 4 of Law, Governance and Technology Series, pages 101–
for adoption by all African legislative bodies at national and regional levels. The African term meaning “linked hearts” in the Akan Language was developed in 2004 under the inspiration of the influential NiR standard above. It has become popular beyond Africa and is the basis of LeXML currently in use by the Senate in Brazil. This standard is extensible, customisable and adaptable to each local situation without sacrificing interoperability between systems. It is therefore applicable to all legislative documents coming from a legislative body, e.g. proposed legislation, registration of debates, drafts, reports, and agendas. Moreover, Akoma Ntoso is the only Legislative XML that currently meets all the criteria listed above. Its development was funded by UN/DESA as part of a project called “Africa i-Parliaments Action Plan: Strengthening the Role of African Parliaments in Fostering Democracy and Good Governance through Knowledge and Information Management.” Appendix II and III shows a simplified example of legislation and of judgement in the Akoma Ntoso standard.

Ultimately, the last decade has produced many Legal XML standards for representing legal resources. It may therefore be necessary to group these developments on the basis of their characteristics as they developed over time.

First Generation

Examples in this category include EnAct developed in the late 1990s by the Tasmanian Government in Australia and Formex discussed above. The motivation was mainly to describe legal text and its structure in a manner akin to the database model or the typography-based word processing models.

Second Generation

The second generation standards feature Italy’s NiR or Denmark’s Lex Dania concentrated on modelling and describing texts, their structure, and metadata. The relevant elements were however represented with no prior abstract analysis of the data classes. This resulted in a long list of tags, without complex inclusions of DTDs or XML schemas and many overlaps between metadata and text definition. Furthermore, the instruments offered for linking text to layers from this generation were weak.

Third Generation

Instances of this generation include Cen/Metalex and Akoma Ntoso already discussed. They are based on patterns which fix the properties of a class and its grammar. This defines a content model demarcating the class’ behaviour and hierarchy in relation to other classes. This way any new tags are pegged to an existing abstract class which allows for consistency over time. The idea is to separate text from its metadata and accompanying ontology in order to easily detect any new layers introduced over the basic text. A pattern therefore guarantees

96 http://projecto.leXML.gov.br/esquesmass/XML.xsd/view
97 http://www.akomantoso.org/
98 See http://www.parliaments.info "Africa i-Parliaments Action Plan: Strengthening the Role of African Parliaments in Fostering Democracy and Good Governance through Knowledge and Information Management".
101 https://www.ministerialtidende.dk/Forms/L0500.aspx?page=5
a clear design by specifying the general rules which no longer enforce real constraints on the markup. However, this means that the standards lack prescriptive constraints.

**Fourth Generation**

Finally, the fourth generation standards include REgular LAnguage for XML Next Generation (RELEX NG)\textsuperscript{102}, Schematron\textsuperscript{103}, and Document Structured Description (DSD). These standards use patterns jointly with co-constraint grammars to fix the lack of prescriptiveness featured in the third generation above.

### 3.1.4 Conclusion

Akoma Ntoso is a third generation standard well on its way to becoming a fourth generation Legal XML standard which testifies to its great success in entrenching XML markup in the legislative domain. We have seen that its forte lies in its ability to advance tools for modelling all parliamentary documents within a single schema. Some extensions maybe needed to tailor the standard to particular document types and formal characteristics e.g. Uruguay’s Parliament and its legislative procedure. However, as evidenced in the Brazilian implementation, this is not a limitation but a mark of Akoma Ntoso’s resilience.

### 3.2 EU and national standards for identification of legal acts, metadata and citations: ELI, ECLI

#### 3.2.1 Motivations

The main motivation of ELI and ECLI is to connect European legislation with relative national implementations. For this reason, legislation needs formal ways to be identified, and so accessed. Still, the presence of an ontological level is useful to avoid misunderstandings between terms.

In fact, knowledge about the European Union law cannot be solely acquired from EU legal sources, but also from national sources, in particular from national legislation implementing European Union law.

These standards can serve as a useful tool for citizens, legal professionals as well as Member States’ authorities. Then, experience has shown that searching on legal databases is often very complex and not user-friendly. In order to facilitate the further development of interlinked national legislations and to serve legal professionals and citizens in their use of these databases, a common system for the identification of legislation and its metadata can be useful.

\textsuperscript{102} http://relaxng.org/

\textsuperscript{103} http://XML.ascc.net/resource/schematron/
3.2.2 ELI

ELI (European Legislation Identifier) is a way to access and thus identify the European legislation by means of HTTP URIs. Each URI is made up of optional components, which do not need to be put in a specific order. URIs are machine-readable and carry semantics from a legal and an end-user's point of view.

A template is the following one:

/el/{jurisdiction}/{agent}/{sub-agent}/{year}/{month}/{day}/{type}/{natural identifier}/{level 1...}/{point in time}/

Where jurisdiction indicates the country code, the agent is the administrative hierarchical structure, and so forth.

Even if a URI can already identify acts, there exist a set of additional metadata with a fixed syntax that can be used to better describe the resources. They can belong to three different levels: work-level, expression-level, and manifestation. As an example, a work is a EU directive, an expression is one single interpretation in one specific nation, and a manifestation is the actual file containing the data.

Additionally, ELI provides an ontology to describe the properties of legislation and their relationships between different concepts in an unambiguous way.
3.2.3 **ECLI**

ECLI (European Case Law Identifier) is a set of five components in the following order: the abbreviation “ecli”, the country code, the court or tribunal, the year of the decision, and an ordinal number which must be unique and contains 25 characters. All these components are separated by a colon “:”.

As in ELI, there are several other metadata that can better describe the case law. Some examples are coverage, date, language, publisher, and so on. Then, each document which is an instance of judgment may also contain the following metadata: title, subject, description, references, and so forth.

3.3 **National standards**

3.3.1 **Italy**

*Italian standardization URN:NIR*

Norme in rete (NIR) was a national project aims at creating a unique access point on web to make easier the retrieval and the navigation between legal documents.

In order to obtain this, two standards had been defined:

- a URN standard, to identify these materials through uniform names, and;
- XML-DTDs to describe legislative documents within the NIR domain.

*URN schema*

Within the NIR project, documents are identified through a Uniform Resource Names: URNs. URNs are able to provide unambiguous and lasting identifiers, independent of physical location, of network resources.

```xml
```

Details:

- *authority*: issuing authority of measure (State, Minister, Region, Court, etc.), within internal departments or offices;
- *measure*: type of measure (laws, decrees, judgements, decisions, regulations, etc.);
- *reference*: references of act (date and number);
- annex: if necessary, id of annex (Allegato #).

Example:

Legislative Decree 8 June 2001, No. 231 → `urn:nir:stato:decreto.legislativo:2001-06-08;231`
**XML-DTDs ID**

NIR project has defined a standard based on XML, aimed at describing the content of legislative documents.

The NIR-DTDs basically describe a legislative text under one formal profile (which considers a legislative text as made up of divisions) and under one functional profile (which considers a legislative text as composed by elementary components called provisions: fragment of a regulation).

Details:

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>libro</td>
<td>book</td>
<td>lib</td>
</tr>
<tr>
<td>parte</td>
<td>part</td>
<td>prt</td>
</tr>
<tr>
<td>titolo</td>
<td>title</td>
<td>tit</td>
</tr>
<tr>
<td>capo</td>
<td>cape</td>
<td>cap</td>
</tr>
<tr>
<td>sezione</td>
<td>section</td>
<td>sez</td>
</tr>
<tr>
<td>paragrafo</td>
<td>item</td>
<td>prg</td>
</tr>
<tr>
<td>articolo</td>
<td>article</td>
<td>art</td>
</tr>
<tr>
<td>comma</td>
<td>comma</td>
<td>com</td>
</tr>
<tr>
<td>lettera</td>
<td>Litteral (list)</td>
<td>let</td>
</tr>
<tr>
<td>numero</td>
<td>Numeric (list)</td>
<td>num</td>
</tr>
<tr>
<td>annesso</td>
<td>annex</td>
<td>ann</td>
</tr>
<tr>
<td>preambolo</td>
<td>Foreword, introduction</td>
<td>pre</td>
</tr>
<tr>
<td>riferimenti</td>
<td>rif, mod, a</td>
<td>rif</td>
</tr>
<tr>
<td>Testo</td>
<td>body of text</td>
<td>txt</td>
</tr>
</tbody>
</table>

Example:

Legislative Decree 8 June 2001, No. 231, article 23, comma 2 -
> urn:nir:stato:decreto.legislativo:2001-06-08;231#art23-com2

<br />

<articolo id="art23">Articolo 23  
<comma id="com2">2  
</articolo>
4 State-of-the-art NLP technologies and legal ontologies used by project partners

In this section we first list the Natural language processing tools at disposal of the consortium. Then we describe open source tools for summarization and machine translation that the partners have good command of, and we describe the tools of the partners for multilabel hierarchical classification that can be used to classify legislation according to the Eurovoc thesaurus and the tools for ontologies.

4.1 NLP tools available to the consortium

NLP technologies are indispensable part of any endeavours, which have to do with processing texts in Natural Languages. Without the basic and also some more advanced NLP machinery it is impossible to perform any complex and real-case tasks, such as information extraction, retrieval, summarization, translation of data in the various domains of life. Current approaches in NLP have made explicit several facts:

- The processing heavily depends on the domains. In-domain solutions are more stable than out-of-domain ones.
- The basic steps, such as tokenization, truecasing, sentence splitting and NER are crucial for the success of the next steps. Although they are not considered a problem, there are still improvements that are needed per language.
- The automated language processing pipelines have become more reliable and include more and more sophisticated steps (such as, semantic annotation and coreference resolution).
- Adequate language resources and linguistic models are needed for achieving state-of-the-art results for analysis, generation and web services.
- Big data that exists in many domains of our life has become a real challenge for processing and utilizing with respect to the user’s needs.

In the EUCases projects the challenge will not be the availability of the Language Technology per se, but the adaptation of the tools and resources for the legal domain; their synchronization within a common architecture for a number of languages and for mapping EU law to national laws.

Three partners of the consortium have NLP tools that may be used in the project: UNITO, AVERBIS, and IICT-BAS. The tools are briefly described below, while their performances and the languages they are able to deal with are summarized in Table 1.

4.1.1 UNITO

The department of Turin in UNITO developed TULE, that stands for “Turin University Linguistic Environment”. TULE includes a Tokenizer, a Sentence splitter, a POS Tagger, a Chunker and a Dependency Parser. TULE works for Italian, English, and French, and also other languages that are not relevant for EUCases. For Italian it has been trained on the TUT corpus, which is the biggest syntactic corpus (treebank), freely available for Italian: it contains more than 3500 sentences annotated via TULE dependency format. TUT sentences have been mainly taken from newspaper, Italian civil code, Wikipedia entries. For French and
English (and also Italian), Tule has been trained on legal texts taken from Jrc-Acquis, UDHR: (Universal Declaration of Human Rights), Europarl, and others.

TULE performances are rather good for Italian, good for English and sufficient for French.

4.1.2 AVERBIS

Averbis developed NLP components and language resources for six languages - English, French, German, Italian, Portuguese, and Spanish. Most of the components use a statistical approach, some are rule-based. The foundations of the statistical components are large document collections available at Averbis (see table below).

All components for linguistic and semantic analysis of unstructured data are completely integrated into UIMA\(^\text{104}\), a framework developed by IBM which also plays an important role in supercomputers such as IBM’s Watson. UIMA allows for the flexible integration of use case specific analysis engines and guarantees highest performance due to the parallel computing processes. The input and output format of these components is a Common Analysis Structure (CAS)\(^\text{105}\) as an open interface for all functionalities. Based on these components, a series of tools is available\(^\text{106}\), as described below.

**Terminology Platform**

The Averbis Terminology Management System offers innovative tools for handling terminologies and ontologies. These term hierarchies aid in the formal sorting of the individual fields of expertise and establish rules on the contexts of the corresponding expressions, thus enabling conclusions to be drawn from the existing data, contradictions to be detected and missing knowledge to be added.

- Via a web-based editor, existing terminologies and other term catalogues can be imported, edited and made available for the extraction of information and term indexing.
- Multilingualism is supported, as is the enhancement of word synonyms and cross references to other terminologies.
- The editor supports the entry of new terms by means of automatic validation and consistency checks, and helps with the enhancement of information from various external sources.

Currently, the system allows for the following input formats: csv-lists, XML, OBO and SKOS.


\(^{105}\) [http://uima.apache.org/downloads/releaseDocs/2.3.0-incubating/docs/api/org/apache/uima/cas/CAS.HTML](http://uima.apache.org/downloads/releaseDocs/2.3.0-incubating/docs/api/org/apache/uima/cas/CAS.HTML)

Text Mining with the Extraction Platform

Averbis Extraction Platform, the text mining platform of Averbis, offers tools for the specific extraction of information from documents. The solution identifies single information units, as well as relevant facts and contexts which are of relevance. Here, many modular software components are implemented which are integrated in a comprehensive solution.

- **Document classification**: articles and text documents are automatically classified for indexing using a freely-definable category system. For example, agency notifications can be assigned to the corresponding departments (e.g. "Economy", "Politics")

- **Concept recognition with the use of terminologies**: here, the lexicon structure is flexible and enables the entry of synonyms and various attributes which play a role in the annotation. The lexicon matching procedure can be carried out for associated or individual text blocks.

- **Recognition of entities**: by mere statistical calculation of scores of different pieces of information and attributes of context words, thus precisely identifying persons, product names, organizations or geographic information.

- **Sentiment Analysis and Opinion Mining**: Qualitative value assessments are recognized reliably and evaluated on sentence level.
Figure 3: Averbis Text Mining platform

Semantic Search Engine

For finding specific relevant documents or for evaluating document collections in great amounts of data, a scalable semantic search engine based on Solr\textsuperscript{107} is available.

- Due to the integration of special components, the search engine offers comprehensive treatment of linguistic phenomena. Even phrases, synonyms or single components of compound words are recognized, and laymen and expert language are normalized
- Faceted search: to reasonably limit the amount of hits, the search engine shows the user related search terms which are semantically associated with the search query.
- Based on text similarities, the search engine automatically calculates recommendations for articles which likewise could be relevant to the user.
- Existing concepts for rights management (e.g. LDAP user groups) can be adopted, whereby the solution supports both the storing of authorizations in the search index and the integration of existing authorization services.

\textsuperscript{107} http://lucene.apache.org/solr/
4.1.3 IICT-BAS

IICT-BAS developed in-house tools for processing Bulgarian. Tools have been developed and tested on the relevant parts of the BuiTreeBank corpus. The BuiTreeBank tools include: a Tokenizer, a Sentence Splitter, a POS tagger, a Lemmatizer, a Chunker, a Syntactic Parser, Semantic parser, and a Coreference resolver. Tokenizer, Sentence Splitter, and Chunker are rule-based. The POS tagger is hybrid. It combines rule-based component with statistical techniques. The Syntactic parser is statistical. The Semantic Tagger is an Ontology Annotation System which implements related ontology-mapped lexicons as chunk grammars and applies it over texts. Finally, the IICT-BAS developed a Machine Translation System, implemented in Moses, from English to Bulgarian and from Bulgarian to English.

Table 1. Tools available to the Consortium

<table>
<thead>
<tr>
<th>Partner</th>
<th>Tool Type</th>
<th>Language(s)</th>
<th>Accuracy</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVERBIS</td>
<td>Language Guesser</td>
<td>27 languages</td>
<td>Very good</td>
<td>-</td>
</tr>
<tr>
<td>UNITO</td>
<td>Tokenizer</td>
<td>Italian, English, French</td>
<td>Italian (very good), English (good), French (good)</td>
<td>Italian (mixed domains), English (legal domain), French (legal domain)</td>
</tr>
<tr>
<td>Partner</td>
<td>Tool Type</td>
<td>Language(s)</td>
<td>Accuracy</td>
<td>Domain</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>AVERBIS</td>
<td>Tokenizer</td>
<td>Spanish, Portuguese, French, Italian, German, English</td>
<td>F1 score between 0,990 and 0,998, depending on language and corpus.</td>
<td>Corpora: ELRA Crater, ELRA PTBAR, ELRA Multex, FRAMED, Tiger, Genia, OANC-Slate, MASC-3</td>
</tr>
<tr>
<td>IICT-BAS</td>
<td>Tokenizer</td>
<td>Bulgarian</td>
<td>Not estimated</td>
<td>News media; Literature</td>
</tr>
<tr>
<td>UNITO</td>
<td>Sentence Splitter</td>
<td>Italian, English, French</td>
<td>Italian (very good), English (good), French (good)</td>
<td>Italian (mixed domains), English (legal domain), French (legal domain)</td>
</tr>
<tr>
<td>AVERBIS</td>
<td>Sentence Splitter</td>
<td>Spanish, Portuguese, French, Italian, German, English</td>
<td>F1 score between 0,4 and 0,986 depending on language and corpus.</td>
<td>Corpora: ELRA Crater, ELRA PTBAR, ELRA Multex, FRAMED, Tiger, Genia, OANC-Slate, MASC-3</td>
</tr>
<tr>
<td>IICT-BAS</td>
<td>Sentence Splitter</td>
<td>Bulgarian</td>
<td>Not estimated</td>
<td>News media; Literature</td>
</tr>
<tr>
<td>UNITO</td>
<td>POS Tagger</td>
<td>Italian, English, French</td>
<td>Italian (very good), English (good), French (sufficient)</td>
<td>Italian (mixed domains), English (legal domain), French (legal domain)</td>
</tr>
<tr>
<td>AVERBIS</td>
<td>POS Tagger</td>
<td>Spanish, Portuguese, French, Italian, German, English</td>
<td>F1 score between 0,949 and 0,992 depending on language and corpus.</td>
<td>Corpora: ELRA Crater, ELRA PTBAR, ELRA Multex, FRAMED, Tiger, Genia, OANC-Slate, MASC-3</td>
</tr>
<tr>
<td>IICT-BAS</td>
<td>POS Tagger</td>
<td>Bulgarian</td>
<td>Hybrid (statistical; disambiguation rules and dictionary)</td>
<td>97 %</td>
</tr>
<tr>
<td></td>
<td>Lemmatizer</td>
<td>Bulgarian</td>
<td></td>
<td>News media; Literature</td>
</tr>
<tr>
<td>IICT-BAS</td>
<td>Chunker</td>
<td>Italian, English, French</td>
<td>Italian (very good), English (good), French (sufficient)</td>
<td>Italian (mixed domains), English (legal domain), French (legal domain)</td>
</tr>
<tr>
<td>AVERBIS</td>
<td>Chunker</td>
<td>English, German</td>
<td>F1 score between 0,939 and 0,96 depending on language and corpus.</td>
<td>FRAMED, Tiger, Genia, OANC-Slate, MASC-3</td>
</tr>
<tr>
<td>IICT-BAS</td>
<td>Chunker</td>
<td>Bulgarian Rules-based</td>
<td>Roughly 80 %</td>
<td>News media; Literature</td>
</tr>
<tr>
<td>UNITO</td>
<td>Syntactic Parser</td>
<td>Italian, English, French</td>
<td>Italian (very good), English (good), French (sufficient)</td>
<td>Italian (mixed domains), English (legal domain), French (legal domain)</td>
</tr>
<tr>
<td>IICT-BAS</td>
<td>Syntactic Parser</td>
<td>Bulgarian</td>
<td>92 % (UAS)</td>
<td>News media;</td>
</tr>
</tbody>
</table>
### 4.1.4 Conclusions

This section shows that the partners have all the different tools covering the NLP stack in the 5 languages covered by the project.

### 4.2 Tools for integrating Human language technologies: Comparison between GATE and UIMA

The General Architecture for Text Engineering (GATE) and the Unstructured Information Management Architecture (UIMA) are information extraction frameworks and both are principally applicable to the text processing tasks of the EUCases project. Yet they differ in a basic principle: GATE was specifically designed for doing text analysis and evaluating the results. It is a complete distribution of tools ready for use including a graphical user-interface to perform whole tasks. In contrast, UIMA's understanding of unstructured information is not restricted to any data format in particular and covers also video broadcasts, for instance. Therefore, UIMA essentially is an execution framework for the orchestration of different analysis engines with the engines themselves not being part of UIMA. A major feature of the framework is its powerful support for horizontal scaling.

In summary, it can be stated that GATE and UIMA are too distinct for being competitive. GATE is a rather off-the-shelf software. Of course, it is possible to run other than the pre-contained processing resources with its pipeline tools. However, this is not primarily intended like it is in UIMA. Additionally, GATE does not have native support for distributed processing. On the other hand, UIMA is a core framework for unstructured information processing. It is oriented towards robust application deployment by providing the aforementioned scale-out ability and employing strong typing within the used annotation terminologies, for example. However, analysis engines and configuration, annotation and other tools are also necessary for a useful application of UIMA.

GATE is licensed under the GNU Library General Public License, UIMA is licensed under the Apache License. Both open-source-licenses seem to be applicable to the EUCases project. Both frameworks use the Java programming language. UIMA additionally is available for C++ and analysis engines that use different languages can be run even together.

UIMA at Averbis: Averbis has expertise in computational linguistics and offers professional large-scale text analysis based on UIMA. Averbis has developed various core analysis
engines that are used in almost every project. Additionally, several publicly available analysis engines are employed. Furthermore, there are many particular analysis engines being developed and used for project-specific tasks. Apart from that, Averbis has built a set of tools to support the configuration and deployment of UIMA-based software.

Available analysis engines at Averbis are – amongst others:

- Document Zoning
- Language Detection
- Document Classification
- Sentence Splitting, Rule Based
- Sentence Splitting, Trainable
- Tokenization, Rule Based
- Tokenization, Trainable
- Part-Of-Speech Recognition
- Shallow Parsing / Chunking
- Stemming
- Morphological Analysis
- Decompounding
- Stopword Recognition
- Invariant Recognition
- Acronym and Abbreviation Resolution
- Regular Expression Annotator
- Lemmatizer, Lexicon Based
- Concept Recognition
- Named Entity Recognition, Trainable
- Concept Disambiguation
- Keyword-Extraction, Controlled and Uncontrolled
- Evaluation Modules
- Table Format Recognition
- (HMM Tagger, BSF Annotator, Alchemi, OpenCalais)
- Drools Annotator
- Relation Extraction, Trainable.
4.3 Automatic text summarization

4.3.1 State-of-the-art in summarization

Automatic text summarization is a field of research based on methods developed in the natural language processing communities. Pieces of the text, sentences or paragraphs, are scored according to their relative relevance and importance by using heuristic or statistical techniques. These values are used for the decision, which parts of the text (keyphrases) are extracted and collected for generating a shorter version of the original text, keeping the most important content as an overview of the content. According to Sparck-Jones (1999, 2007), this approach may yield to incoherent extract and may be difficult to understand. This extraction-based approach, on which most research is focusing on due to computationally simpler methodologies, is contrasted by approaches based on so-called abstraction, where an internal linguistic and semantic representation is built from the text and natural language generation techniques (which are still an area of research) are used to create a summary of the original document (Leskovec et al., 2005a/b and Vanderwende et al., 2004).

The assessment of the quality of the results and an objective evaluation of the outcomes of such methods is, unfortunately, not that straightforward. This particularly came out at the Document Understanding Conference 2002 where an automatic evaluation tool ROUGE108 was proposed to make systems comparable in an objective way (Lloret & Palomar, 2010). Nevertheless, there is evidence that current systems based on extraction-based approaches combined with a combination of statistical classifiers such as maximum-entropy models, support-vector machines, naïve-bayes, decision trees, neural nets etc. (Suneetha, 2011) outperform the more sophisticated (but more complex and difficult to handle) techniques based on a formal semantic representation (Svore et al. 2007).

One particularity of EUCases, at least from the natural language processing perspective, is the characteristics of the different text sources to be analyzed. Usually, the flow of information in a given document is not uniform, which means that some parts are more important than others (e.g. the very beginning and end in a newspaper article, cf. Soneetha, 2011). This assumption is not necessarily true for legislative texts and corresponding documents, thus making the task of automatic summarization even more challenging. Related, but more pragmatic approaches such as automatic classification and keyword (descriptor) extraction can therefore be seen as an adequate alternative, as described in the subsequent chapter.

4.3.2 Available Libraries and Services

Open Text Summarizer

The Open Text Summarizer109 (OTS) summarizes texts in English, German, Spanish, Russian, Hebrew, Esperanto and other languages. To support more languages or tweak existing languages one can edit an XML file of rules. OTS is both a library and a command line tool. A simple web interface110 for testing the functionalities is available.

---

108 http://www.berouge.com/Pages/default.aspx
109 http://libots.sourceforge.net/
110 http://www.splitbrain.org/services/ots
**MEAD**

MEAD\(^{111}\) is a publicly available platform for multi-lingual summarization and evaluation. The platform implements multiple summarization algorithms such as position-based, centroid-based, largest common subsequence, and keywords. The methods for evaluating the quality of the summaries are both intrinsic and extrinsic. MEAD implements a battery of summarization algorithms, including baselines (lead-based and random) as well as centroid-based and query-based methods.

**CNGLSummarizer**

CNGLSummarizer\(^{112}\) is a tool that uses sentence extraction to generate summaries. The selection techniques are based on Information Retrieval and Natural Language Processing methodologies.

Other summarization services that can be integrated through a web API are: Text Teaser\(^{113}\), Sumplify\(^{114}\) or TextCompactor\(^{115}\).

**References for section 4.3**


---

\(^{111}\) http://www.summarization.com/mead/

\(^{112}\) https://github.com/CNGLdlab/CNGLSummarizer

\(^{113}\) http://www.textteaser.com/

\(^{114}\) http://sumplify.com/

\(^{115}\) http://www.textcompactor.com/
4.4 Multi-label classification of legislative text into EuroVoc

4.4.1 Introduction

Classification of legal text is an important task when dealing with large amounts of documents and information extraction processes, in some cases also amplified by high risks of not finding all relevant documents. It usually involves knowledge intensive manual work which is slow and costly. Knowledge engineers specify the domain to which each norm belongs, selecting key terms within a domain-specific ontology. In particular, many parliaments and other institutions are increasingly using the EuroVoc descriptors to classify documents. EuroVoc is a large multidisciplinary multilingual hierarchical thesaurus of more than 7,000 classes covering the activities of EU institutions. Given the amount of legal documents produced everyday and the huge mass of pre-existing documents to be classified, high-quality automated or semi-automated classification methods are most welcome in this domain.

In general, automated text classification is a well-studied task that most of the times works on flat organizations of the categories. Support Vector Machines (SVM) is known to achieve state-of-the-art levels of accuracy in several domains in that sense. Despite this, sometimes categories are structured in hierarchical organizations. While this situation can be generally faced by flattening out the hierarchy, an accurate exploitation of such a knowledge can lead to significant improvements of the used approach. Moreover, some labels may have very few positive examples, leading to highly skewed data that can be problematic for classifiers. Another increase of complexity in text classification tasks is represented by multi-labeled datasets where each piece of text can belong to more than one category. Finally, the large amount of data poses further complications.

Current state-of-the-art approaches rely on the use of supervised Machine Learning techniques like the above-mentioned SVM, that typically achieve state-of-the-art accuracy in several domains. Nevertheless, SVM cannot handle multi-labeled documents by their intrinsic mathematical nature, thus it has been introduced specific workarounds that try to preserve all the information while being suitable for SVM. The same conceptual problem is also present with the management of the hierarchical organization of the labels, where the data and the standard algorithms need to be elaborated in order to achieve the goal.

4.4.2 Background and Related Work

In this section we present the main works that have been done in automatic classification of legislative text and classification methods in case of multi-label texts. In general, Text Classification on non-domain specific data is a well-studied task and there is a plenty of works focusing on it. A complete and up-to-date survey of these approaches is contained in [1]. In the next sections we give an overview over specific subtasks related to the problem of classifying documents in the legal domain and with the discussed complexity.

---

4.4.3 Classification of legislative text

In [2], the authors proposed JEX, a system for the classification of texts into Eurovoc categories. The core of their approach is based on the construction of class profiles, i.e., sets of terms that define the related categories. The classification step is done by computing the cosine similarity of the documents with the class profiles. Although this is a recent contribution, the fact that it still uses well-known classic techniques like Cosine Similarity on Vector Space Models [3], well reflects the high complexity of the task.

Multi-label Classification

The problem of classifying text documents associated to more than a single category is currently met in several domains and applications. In the case of legislative documents, it is common to think about some legal text that can be associated to different concepts belonging to bureaucracy (e.g., "European contract"), economy (e.g., "inflation rate"), geography (e.g., "Italy"), and so forth. From a technical point of view, dealing with this kind of information leads to more complex classification systems. For instance, SVM, in spite of the fact that it is one of the most used Machine Learning algorithm to learn from text corpora, it needs the documents to be mono-label. There are mainly two approaches to face such complexity: transforming multi-label data into mono-label, or adapting existing algorithms to work with mono-label data, and here we will briefly go over them. In mono-label scenarios, the classification task is only concerned with learning from a set of examples that are associated with a single label. The easiest case is when the possible labels are just two, so that the learning problem is called a binary classification problem. In presence of more than two labels then it is called a multi-class classification problem.

Problem transformation Methods

The first approach to tackle multi-label classification problems regards the transformation to standard mono-label data, suitable for any supervised learning algorithms. The easiest way consists of removing all the documents that have more than one label from the dataset [4]. This is obviously ineffective, especially when all (or almost all) documents are multi-label. Another method randomly selects one of the multiple labels for each document, discarding the rest [4]. This allows the system to learn from multi-label data, but it actually does not represent a real treatment to the problem. Moreover, the accuracy of such solution is likely to be very low. A third approach called power set considers each different set of labels that is present in the multi-label data as a single label [5]. For instance, if the labels of a document are "A", "B", and "C", the system transforms the labels of the document into a single label "ABC". One of the negative aspects of this method is that it may lead to data sets with a large number of classes and few examples per class. Then, another transformation method learns one binary classifier for each label in the data. To classify a new document, it needs to pass over all the classifiers to determine its associated set of labels. Of course, in case of thousands of categories (as in the EuroVoc case), this strategy becomes unsustainable.

The state-of-the-art idea has been firstly described in [6] and extensively developed in [7]. The intuition is to consider each n-labeled document as a collection of n minor documents, each one associated to only one label. The main issue to deal with is about how to segment the original document, i.e. how to choose the features to maintain for each of the new mono-label documents. One solution is to only consider the creation of n identical mono-label documents from one multi-label document. However, this can be an inefficient approach in case of large corpora, and it can also lead to extremely overlapping classes from which it could be difficult to learn any model. In [7], the authors present an approach based on Pointwise-Mutual Information calculated over the co-occurrence matrix of the features (i.e.,
the terms) with the labels (i.e., the categories), achieving promising results with low computational requirements.

Algorithm Adaptation Methods

The second approach to cope with multi-label data is to adapt standard learning algorithms that usually work with mono-label data. Our technique, however, is completely different in its purpose, since data transformation leads to datasets that can be used for training any supervised learning algorithm. In [8], the authors adapted the C4.5 algorithm [9] for multi-label data by modifying the formula of the entropy calculation. In [10], the authors proposed two extensions of AdaBoost [11] for multi-label classification. The authors of [12] presented a technique called ML-kNN which adapts the simple K-nearest neighbours learning algorithm [13]. Then, [14] presented two improvements for the SVM classifier by using a different problem transformation method that creates one binary classifier for each category and it thus has strong limitations with respect to our system, as already mentioned in previous sections. Finally, [15] presented an algorithm that treats the multi-label classification task as an association rules mining approach.

References for section 4.4

4.5 Machine Translation

The methodology employed for the machine translation task can be divided into two main types: rule-based and statistical approaches.

Classic rule-based MT (RBMT) systems make use of a set of rules defined by a language expert. This method has not had much popularity in recent years, since the design of such systems requires a lot of time and manual effort, and they are still not able to achieve very broad coverage. Furthermore, these kinds of systems are not language-independent, and need to be adapted for every language pair. RBMT approaches have, however, been successfully applied for translations in limited domains, and when dealing with low resource languages.

Phrase-based statistical machine translation (PB-SMT) systems represent the current state-of-the-art approach for machine translation. They are superior to the previously used word-based models for SMT, which consider the word to be the fundamental unit of translation. Word-based approaches are limited because languages usually cannot be represented as a direct mapping of words. It cannot be safely assumed that each word in the source language covers the same concept in the target language, or the other way around. The problem becomes evident when several tokens in the source language need to be mapped to a single token in the target language, as is often the case when dealing with the translation of multiword expressions, for instance. Phrase-based approaches for MT are able to better handle this kind of issues. In them translation is designed as a mapping of small text chunks referred to as phrases, not in the linguistic sense, but rather as n-grams which represent good statistical descriptions of the parallel data. The limitations of word-based approaches are partially overcome in that the mapping of whole sequences of words is now considered during the interpretation process.

Factored translation models are an extension of the standard PB-SMT models which allow for additional linguistic information to be easily integrated into the translation process. The basic unit of translation in these models has a more general representation as a vector of linguistic factors, such as surface form, lemma, part-of-speech tag, morphological specifications, etc. Such models have been shown to overcome data sparsity issues when dealing with languages with rich morphology.

A further level of abstraction is introduced into SMT with the hierarchical phrase-based translation models. This approach incorporates syntactic knowledge in the form of synchronous context-free grammars. Hierarchical SMT systems are able to better handle certain phenomena, which standard PB-SMT systems cannot model, due to their restriction on n-gram phrase length. One example is long-distance dependencies.

In recent years, MT systems have achieved significant improvement in terms of translation quality. However, the final results produced by them are still far from satisfactory. Fortunately, recent error analysis shows that the two trends of the MT approaches tend to be complementary to each other, in terms of the types of the errors they make (Thurmair 2005; Chen et. Al 2009), which motivates for the exploration of hybrid approaches for MT. While RBMT systems often lack robustness, they are able to handle better the linguistic phenomena that require syntactic information. SMT systems, on the other hand, are more robust in general, but often output ungrammatical sentences.
Within the EUCases project we will investigate all of these models for supporting query translation. Here we present systems that are used to support MT and related activities.

### 4.5.1 Machine Translation Systems

**MOSES**

http://www.statmt.org/moses/

Moses is a data-driven machine translation system. It provides three types of translation model implementations: phrase-based models, where n-grams ("phrases") are the basic units of translation, hierarchical or syntax-based models, where information about the structure of the parallel data can be incorporated, and factored translation models, where additional linguistic information (e.g., lemma, part-of-speech tag) can be integrated into the translation process.

Moses has two main components – a training pipeline and a decoder. The training pipeline includes various tools for data preprocessing (e.g., for tokenisation, lowercasing, removing very long sentences on the source or target side of the training data, etc.), for accessing external tools for data alignment (GIZA++), language model building (SRILM, KenLM, IRSTLM, RandLM), and implementations of popular tuning algorithms. The Moses decoder tries to find the highest scoring sentence during translation, or outputs a ranked list of translation candidates with additional information. Standard tools for the evaluation of translations (e.g., BLEU scorer) are also available.

In addition to parallel data in the form of plain text, Moses can be used to decode data represented as confusion networks or word lattices. In this way, ambiguous input data, such as the output of an automatic speech recognizer, or a morphological analyzer, can be processed to reduce erroneous hypothesis.

**JOSHUA**

http://joshua-decoder.org/

Joshua is an open-source SMT system for hierarchical and syntax-based machine translation with synchronous context-free grammars. It includes a decoder component and a tool chain for data preprocessing (tokenization, normalization), alignment (GIZA++ or Berkeley aligner), building translation and language models (KenLM, BerkeleyLM, or SRILM), tuning, and evaluation, similarly to the Moses training pipeline. Joshua has a built-in grammar extractor. The decoder component is an implementation of the CYK+ algorithm.

**APERTIUM**

http://www.apertium.org/

Apertium is an open-source rule-based machine translation system. It provides a language-independent machine translation engine, tools for data processing, and linguistic data for several language pairs. The pipeline of Apertium includes the following processes: de-formatting, morphological analysis, part-of-speech disambiguation, shallow structural transfer, lexical transfer, morphological generation, and re-formatting. It uses finite-state methods for the lexical processing operations (morphological analysis and generation, lexical transfer), Hidden Markov Models for part-of-speech tagging, and multi-stage finite-state based chunking for structural transfer.
4.5.2 Aligners

**HUNALIGN**

http://mokk.bme.hu/resources/hunalign/

Hunalign is a tool which aligns parallel data on the sentence level. The input to the system should be tokenized and sentence-segmented. A bilingual dictionary can be provided to aid the alignment process. In the absence of such, a dictionary is constructed automatically. Gale-Church sentence length information is used in this process.

**GIZA++**

http://code.google.com/p/giza-pp/

GIZA++ is a tool for aligning words or n-grams from sentence-aligned parallel data. It includes implementations of IMB Models 4 and 5, alignment models using word classes, the HMM alignment model (Baum-Welch training, Forward-Backward algorithm, empty word, dependency on word classes, transfer to fertility models), a version of IMB Model 3 and 4, various smoothing techniques for fertility, distortion/alignment parameters, and other functionality.

**MGIZA**

http://www.kyloo.net/software/doku.php/mgiza:overview

MGIZA (Multi-threaded GIZA++) is an implementation of the GIZA++ word alignment toolkit which runs multi-threaded on multi-core machines. In addition to multi-threading, it offers several extensions to the original: memory optimization, functionality to resume training, running on clusters using Chaski, and native Windows support.

**ANYMALIGN**

http://anymalign.limsi.fr/

Anymalign is an open-source tool which can align any number of languages simultaneously from sentence-aligned parallel data. It employs a different, simpler alignment strategy than GIZA++, which relies on the use of low-frequency terms. Thus Anymalign is more flexible regarding the amount of input data. Experiments reveal that the system achieves state-of-the-art performance.

**Berkeley Aligner**

http://code.google.com/p/berkeleyaligner/

The Berkeley aligner is a tool for unsupervised word alignment implemented in Java.
**FAST ALIGN**

https://github.com/clab/fast_align/

Fast Align is a tool for unsupervised word alignment. It is a modification of IBM Model 2. The model has been shown to outperform IBM Model 4 in terms of speed and quality.

### 4.5.3 Language Model Toolkits

**IRSTLM**

http://sourceforge.net/projects/irstlm/

The IRST Language Modeling Toolkit features algorithms and data structures suitable to create and access very large n-gram language models. The system uses model formats which permit to reduce both storage and decoding memory requirements, and to save time while loading. After the collection of n-grams and their frequency estimation, smoothing parameters are calculated for each entry. The available smoothing methods are Witten-Bell and an approximation of the modified Kneser-Ney smoothing. To decrease memory and time requirements, IRSTLM splits the training process into smaller independent steps, which can be distributed among independent processes.

**SRILM**

http://www.speech.sri.com/projects/srilm/

SRILM is a toolkit which supports the creation and evaluation of a variety of language model types: class-based models, cache models, disfluency and hidden event language models, skip language models, HMMs of n-grams, and dynamically interpolated language models. Additional functionality includes statistical tagging and manipulation of n-best lists and word lattices.

**RANDLM**

http://sourceforge.net/projects/randlm/

RandLM is a tool for language modeling, which employs a different strategy for storing n-grams than SRILM and IRSTLM. Language models are represented using a randomized data structure, which is a variation of the Bloom filters. Models can therefore be stored in a form which is about ten times smaller than those created using the other modeling systems, but at the cost of slower decoding, even with the multi-threaded version of the tool.

**KENLM**

http://kheafield.com/code/kenlm/

KenLM is a language modeling system which offers speed and lower memory consumption compared to other tools. The probability estimates obtained with it are the same as the ones from SRILM, up to floating point rounding. KenLM is distributed with Moses, and is thread-safe for use with multi-threaded Moses.
4.6 Ontology annotation

4.6.1 Introduction

Current development of Semantic Web technologies determines ontologies as primary means for representation of conceptual knowledge. Within the project we adopt Tim Gruber’s definition of ontology: “An ontology is a formal, explicit specification of a shared conceptualisation.” (Gruber 1993). We consider an ontology as a set of classes (concepts) which are interconnected via properties (relations). Within the EUCases project, ontologies will represent legal knowledge in a language independent way. On top of the ontologies we will build ontology annotation components for each project language to discover and annotated mentionings of conceptual elements (concepts, relations) in text.

Regarding the complexity and precision of a given ontology we follow the definition in (Guarino 2000). It represents the following classification of ontologies:

- **Lexicon**: machine readable dictionaries; vocabulary with natural language definitions.
- **Simple Taxonomy**: classifications.
- **Thesaurus**: WordNet; taxonomy plus related-terms.
- **Relational Model**: Light-weight ontologies; unconstrained use of arbitrary relations.
- **Fully Axiomatized Theory**: Heavy-weight ontologies.

The classification starts with a less formal and knowledge-poor ontology (hence - simple lexicons) and ends with heavily constrained theories about the world. Sometimes the first three elements of the classification are not considered ontologies, because the ontological information is represented mainly implicitly. This hierarchy shows also the transition from lexicon to ontology and plays a role in the process of creation of ontologies. In our view such a transition also supports the mapping between the ontology and lexicon. In the project our goal is to annotate primarily legal documents with light-ontologies in the legal domain.

4.6.2 The structure of a domain ontology

Our work is based on a model developed for the annotation of domain conceptual knowledge in text. In this model we assume that the ontology is the starting point for the creation of the ontology-to-text relation. The structure of a domain ontology can be defined (at least) with respect to: (1) the collection of concepts represented in the ontology; and (2) the complexity of the concept definitions. Independently from the methodology for ontology creation, the concepts represented in the ontology can be distributed on the following layers which reflect the generality of the conceptual information:

- **Domain layer**.

At this layer we have the domain concepts and relations representing the main notions in the domain. These concepts and relations are used for solving different tasks, such as the representation of domain knowledge, the representation of common conceptualization for information exchange in the domain, the semantic annotation of domain texts, etc.

- **Upper layer**.

The alignment of the domain layer to an upper ontology is an obligatory step in each ontology creation methodology. This alignment ensures several properties of the domain ontology: (1) its consistency with the design of the upper ontology; (2) inheritance of the knowledge represented in the upper ontology. The inheritance requires the imposition of more specific constraints reflecting the structure of the domain.
- **Middle layer.**

This layer contains concepts and relations which are neither part of the upper layer, nor of the domain one, but play an important role for the alignment between them. This layer is the result of the ontology creation practice and depends on the coverage of the domain and the range of concepts defined in the upper ontology. In our view it is a useful instrument for transition from the domain to the upper layer.

An additional layer related to the conceptual information is the linguistic information represented by a lexicon and a grammar. This information is necessary in all cases where the ontology interacts with natural language, for example in the analysis of texts, when navigating the ontology, in ontology based searches, etc.

- **Language layer.**

It is assumed that the ontology with its three layers is language independent, formalized in some ontology representation language. In practice, such an ontology would be incomprehensible to humans and therefore has to be mapped to some linguistic resources. This mapping is required for at least two reasons: to allow to present the ontology to users who are not ontology engineers, and to support semantic analysis and retrieval of texts. Thus, as a minimum it is necessary to have a lexicon mapped to the concepts and the relations in the ontology. The following figure shows the structure of a domain ontology:

![Figure 5: Domain ontology structure](image)

Here on the left side we have two domain ontologies (IT for the domain of Information Technology for End Users and LG for the Legal domain) aligned to the middle layer and the upper layer. The linguistic layer consists of lexical units, grammar rules, disambiguation rules, etc. The mapping between the ontology and the linguistic layer is the way to define the *ontology-to-text* relation. This relation supports the semantic annotation of text. It generally comprises a lexicon and a grammar. In our own work in the past the linguistic layer was implemented via domain lexicons, presented below, and concept annotation grammar,
described in (Simov and Osenova, 2007) and (Simov and Osenova 2008). Our approach draws in many respects on the work done on WordNet (Fellbaum, 1998), EuroWordNet (Vossen, 1998), SIMPLE (Lenci et al, 2000). With WordNet-like lexicons we share the idea of grouping lexical units around a common meaning and in this respect the term groups in our model correspond to synsets in the WordNet model. The difference is that the meaning is defined independently in the ontology. With the SIMPLE model we share the idea to define the meaning of lexical units by means of an ontology, but we differ in the selection of the ontology which in our case represents the domain of interest, and in the case of SIMPLE reflects the lexicon model of Generative Lexicon.

4.6.3 Ontology-to-text relation

Here we represent the two main components that define the ontology-to-text relation necessary to support the tasks within our projects. These components are: (terminological) lexicon and concept annotation grammar.

The lexicon plays twofold role in our architecture. First, it interrelates the concepts in the ontology to the lexical knowledge used by the grammar in order to recognize the role of the concepts in the text. Second, the lexicon represents the main interface between the user and the ontology. This interface allows for the ontology to be navigated or represented in a natural for the user way. For example, the concepts and relations might be named with terms used by the users in their everyday activities and in their own natural language. This could be considered the first step to a contextualized usage of the ontology in a sense that the ontology could be viewed through different terms depending on the context. For example, the color names will vary from very specific terms within the domain of carpet production to more common names used when the same carpet is part of an interior design.

Thus, the lexical items contain the following information: a term, contextual information determining the context of the term usage, grammatical features determining the syntactic realization within the text. In the current implementation of the lexicons the contextual information is simplified to a list of a few types of users (lawyers, judges, etc).

With respect to the relations between the terms in the lexicon and the concepts in the ontology, there are two main problems: (1) there is no lexicalized term for some of the concepts in the ontology, and (2) there are lexical terms in the language of the domain which lack corresponding concepts in the ontology, which represent the meaning of the terms.

The first problem is overcome by writing down in the lexicon also non-lexicalized (fully compositional) phrases to be represented. Even more, we encourage the lexicon builders to add more terms and phrases to the lexicons for a given concept in order to represent as many ways of expressing the concept in the language as possible. These different phrases or terms for a given concept are used as a basis for construction of the annotation grammar. Having them, we might capture different wordings of the same meaning in the text. The picture below shows the mapping varieties. It depicts the realization of the concepts (similarly for relations and instances) in the language. The concepts are language independent and they might be represented within a natural language as form(s) of a lexicalized term, or as a free phrase. In general, a concept might have a few terms connected to it and a (potentially) unlimited number of free phrases expressing this concept in the language. Some of the free phrases receive their meaning compositionally regardless their usage in the text, other free phrases denote the corresponding concept only in a particular context. In our lexicons we decided to register as many free phrases as possible in order to have better recall on the semantic annotation task. In case of a concept that is not-lexicalized in a given language we require at least one free phrase to be provided for this concept.
We could summarize the connection between the ontology and the lexicons in the following way: the ontology represents the semantic knowledge in form of concepts and relations with appropriate axioms; and the lexicons represent the ways in which these concepts can be realized in texts in the corresponding languages. Of course, the ways in which a concept could be represented in the text are potentially infinite in number, thus, we could hope to represent in our lexicons only the most frequent and important terms and phrases.

The second component of the ontology-to-text relation, the concept annotation grammar, is ideally considered as an extension of a general language deep grammar which is adopted to the concept annotation task. Minimally, the concept annotation grammar consists of a chunk grammar for concept annotation and (sense) disambiguation rules. The chunk grammar for each term in the lexicon contains at least one grammar rule for recognition of the term. As a preprocessing step we consider annotation with grammatical features and lemmatization of the text. The disambiguation rules exploit the local context in terms of grammatical features, semantic annotation and syntactic structure, and also the global context such as topic of the text, discourse segmentation, etc. In the project we will implement conceptual chunk grammars for several languages. The disambiguation model will be statistically-based.

For the implementation of the annotation grammar we rely on the grammar facilities of the CLaRK System (Simov et al., 2001). The structure of each grammar rule in CLaRK is defined by the following DTD fragment:

```xml
<!ELEMENT line (LC?, RE, RC?, RM, Comment?)>
<!ELEMENT LC (#PCDATA)>
<!ELEMENT RC (#PCDATA)>
<!ELEMENT RE (#PCDATA)>
<!ELEMENT RM (#PCDATA)>
<!ELEMENT Comment (#PCDATA)>
```
Each rule is represented as a line element. The rule consists of a regular expression (RE) and a category (RM = return markup). The regular expression is evaluated over the content of a given XML element and could recognize tokens and/or annotated data. The return markup is represented as an XML fragment which is substituted for the recognized part of the content of the element. Additionally, the user could use regular expressions to restrict the context in which the regular expression is evaluated successfully. The LC element contains a regular expression for the left context and the RC for the right one. The element Comment is for human use. The application of the grammar is governed by Xpath expressions which provide additional mechanism for accurate annotation of a given XML document. Thus, the CLaRK grammar is a good choice for implementation of the initial annotation grammar. In the project we will try also to implement concept annotation grammars as an extension of dependency parsing in order to have a better structure for relation identification. Also for chunk grammar can be used other frameworks.

The creation of the actual annotation grammars started with the terms in the lexicons for the corresponding languages. Each term was lemmatized and the lemmatized form of the term was converted into a regular expression of grammar rules. Each concept related to the term is stored in the return markup of the corresponding rule. Thus, if a term is ambiguous, then the corresponding rule in the grammar contains reference to all concepts related to the term.

The following picture depicts the relations between lexical items, grammar rules and the text:

![Figure 7: relation between lexical items, grammar rules and text](image)

The relations between the different elements of the models are as follows. A lexical item could have more than one grammar rule associated to it depending on the word order and the grammatical realization of the lexical item. Two lexical items could share a grammar rule if they have the same wording, but they are connected to different concepts in the ontology. Each grammar rule could recognize zero or several text chunks.

The relation ontology-to-text implemented in this way provides facilities for solving different tasks, such as ontology search (including crosslingual search), ontology browsing, ontology learning. In order to support multilingual access to semantic annotated corpus we have to implement the relation for several languages using the same ontology as starting point. In this way we implement a mapping between the lexicons in these languages and also comparable annotation of texts in them.
A new model for aligning lexicons to ontologies is the result from the European project MONNET. A presentation follows:

**LEMON (Lexicon Model for Ontologies)**

http://lemon-model.net/

Lemon is a proposed model for modeling lexicon and machine-readable dictionaries and linked to the Semantic Web and the Linked Data cloud. It was designed to meet the following challenges:

- RDF-native form to enable leverage of existing Semantic Web technologies (SPARQL, OWL, RIF etc.).
- Linguistically sound structure based on LMF to enable conversion to existing offline formats.
- Separation of the lexicon and ontology layers, to ensure compatibility with existing OWL models.
- Linking to data categories, in order to allow for arbitrarily complex linguistic description. In particular the LexInfo vocabulary is aligned to Lemon and ISOcat.
- A small model using the principle of least power - the less expressive the language, the more reusable the data.

Lemon was developed by the Monnet project as a collaboration between: CITEC at Bielefeld University, DERI at the National University of Ireland, Galway, Universidad Politécnica de Madrid and the Deutsche Forschungszentrum für Künstliche Intelligenz.

The Lemon API is a Java-based interface for using Lemon that is compatible with all major platforms. In addition, there is a Wiki-like site for manipulating and publishing Lemon data, called Lemon Source. This includes a service for generating lexica from ontologies. Downloads are available in the following formats: XML/RDF; Turtle; HTML+RDFa. The standard is well-documented at: http://lemon-model.net/learn/cookbook.HTML. Lemon API is an open source project, released under the BSD License. The core model is visualized below:
Our brief analysis shows the following positive points:

- Provides granularity of implementation. The model can be realized in its minimal part, but also extended further, if needed.
- Provides connections to other standards (SKOS, LMF, TBX, ISOcat).
- There are resources that already followed this standard (such as, Wordnet)
- There is a detailed guide for Lemon model implementation
- There is a formal model behind the standard
- Existing multilingual RDF extraction from Wictionary

and some negative points:

- There is no 1:1 reversibility among Lemon and other standards.

Multilinguality settings require additional effort by the interested groups. Within the project we keep also the mapping of lexicons to ontology in the format provided by the LEMON model.

**References for section 4.6**


Legal ontologies and the European legal taxonomy Syllabus

4.7.1 Introduction

Ontology as a branch of philosophy is described by Smith (Floridi [2008], page 155) as having a very general scope: “the science of what is, of the kinds of objects, properties, events, processes and relations in every area of reality”.

---


The section will eventually be the starting point for a journal paper as detailed in the dissemination plan.
In computer science, this quest for describing every aspect of reality takes the form of creating an "an explicit specification of a conceptualization" according to the oft-quoted definition of Gruber and others [1993].

The objective is typically to provide humans or more typically artificial agents with knowledge about different aspects of the world; in the case of the former, to facilitate the sharing of knowledge and intercommunication, also between different languages. In the case of the latter, to facilitate machine reasoning and problem solving, semantic searches and interoperability between systems. Ontologies are often seen as intermediaries between the world conceived by humans in natural language and systems that have no a priori knowledge of either concepts or the language used to express such concepts and their inter-relations. Conceptual specifications should be made explicit, clear and unambiguous, and aligned to all relevant terms that express such concepts, in order to overcome the redundancy of multiple expressions for the same thing found in natural language, a difficult barrier to machine understanding and reasoning.

Ontologies can work very well in certain scenarios. For instance, DOMOTICS (Groot [1999]) is a sophisticated system for manipulating multiple devices such as security, heating, air-conditioning, television and dishwashers in the home. The system uses an ontology to describe the systems in conceptual terms and overcome terminological differences by different suppliers. Another example is the Elsevier system for cross-journal querying, which addresses variability of search terms by grouping terms and providing a controlled vocabulary for indexing, using a single underlying ontology called EMTREE (Antoniou [2004], page 180). Ontologies have also proven to be useful in scientific and medical domains for overcoming terminological differences, again also for humans. For instance, the Gene Ontology was developed to address a common problem (the Gene Ontology [1999, 2013]):

If you were searching for new targets for antibiotics, you might want to find all the gene products that are involved in bacterial protein synthesis, and that have significantly different sequences or structures from those in humans. If one database describes these molecules as being involved in ‘translation’, whereas another uses the phrase ‘protein synthesis’, it will be difficult for you - and even harder for a computer - to find functionally equivalent terms.

These examples are successful because in their domains knowledge engineers can create clear, explicit and unambiguous conceptualisations ready for use by a system.

This is not possible in all domains, despite the apparent usefulness of ontologies for the problems of practitioners and systems in such fields. It is very difficult to create an explicit specification of the law, even if legal ontologies would be useful for many different types of users, in particular in multilingual contexts like European Union. From the human point of view, they can help law practitioners and scholars to understand specific legal sub-languages they are not expert in, and to keep up to date with the continuous changes in legislation. They can help to achieve clarity and consistency in legislative drafting at European and national levels. Currently, European translators are not always aware on the impact of their choice of terms on the way legislation will be interpreted in different jurisdictions. At the same time, drafters of national legislation are inconsistent in the use of terminology, sometimes using European terms unfiltered, and at other times transposing into terms that are more familiar in the national jurisdiction. A multi-jurisdictional legal ontology can help show the inter-relationship between national and European terms. The ontology can also serve as an useful tool to help find similar legislation, compare the implementation of laws in national jurisdiction and help translators. This can be most useful for lawyers who deal with cross-border issues, international financial institutions, as well as legal scholars in comparative law.

None of the existing legal ontologies, taxonomies or thesauri available could provide authoritative context-specific definitions of legal terms, or map terms from one jurisdiction to another to help with cross-lingual and cross-jurisdictional search.

From the machine point of view, legal ontologies can be useful in search, information retrieval, automatic translation, automated reasoning, regulatory compliance verification, etc.
The peculiarities of the legal domain are many. Laws are written in legalese, a domain-specific sublanguage that inherits all the expressivity and ambiguity of natural language with additional terms of its own whose meaning are often obscure. Their meanings are obscure either because the terms predate the modern state, are 'obvious' to anyone who shares the same culture, are ill-defined (i.e. some 'obvious' elements are not made explicit), or are defined in incompatible ways in different valid legal sources. Understanding legal terminology requires deep understanding of legal culture and societal values, a situation made more complex by the fact that there are several legal jurisdictions who each have their own set of concepts. Finally, concepts from one jurisdiction are sometimes adopted and adapted in other jurisdictions due to globalisation and subsidiarity, creating multilingual terminological complexities.

These problems have barely been addressed in the development of most legal ontologies, which may explain their low acceptability by legal professionals. It is still difficult to deny what Wahlgreen stated (Wahlgren [1984], page 80-81):

Several contributions in the field of AI and law appear to be incompatible with respect to the understanding of the law and legal work. It is no secret that very few systems have been accepted by the legal community. The problems, however, are not primarily of a technical nature. With little doubt, the difficulties are more closely related to a too shallow understanding of the requirements of the domain taken as whole.

The methodology we adopted is to look at the approach of comparative law, considered first of all that our aim is modelling European Law. Legal knowledge will be constructed in a bottom-up fashion, to avoid abstractions which are too far from the daily work of legal practitioners, using a lightweight ontology to reduce the costs of construction and to leave space for interpretation.

As a benchmark, we use the ability of our framework to model a detailed European wide legal terminology analysis developed in the European project Uniform Terminology\[^{118}\] in the domain of consumer law.

First we discuss philosophical and linguistic issues that have impacted on the design of LTS. Then we describe the requirements, features, platform and interface of the system. Finally we describe state-of-the-art legal ontologies and the similarities and differences to our approach.

4.7.2 Background

Legal Theory

Ontologies are a costly, resource intensive endeavour at the best of the times, and the law is a particularly difficult and untidy domain to model. Ontology engineers have responded to this challenge by focussing on abstract high-level concepts first. Such an approach to building ontologies is generally known as ‘top-down’ and is the opposite of the alternative ‘bottom-up’ approach, where ontologies are initially populated with atomic concepts. Uschold and Grüninger (Uschold and Grüninger [1996], page 13) explains this preference: “A bottom-up approach results in a very high level of detail. This in turn (1) increases overall effort, (2) makes it difficult to spot commonality between related concepts and (3) increases risks of inconsistencies which lead in turn to (4) re-work and yet more effort.”

In the legal ontology field, the top-down approach results in taking from classical legal theory such as ‘permission’ and ‘obligation’ rather than low-level concepts such as ‘contract’ and

\[^{118}\] www.uniformterminology.unito.it
‘consumer’, albeit with a view to extending to lower-level concepts in due course. For instance, FOLAW, (Engers et al. [2008]) is informed by Hart, Kelsen and Hohfield, has the following top-level categories.

The rationale for deriving specifications from legal theory, as opposed to legal practice, is that “they have the advantage of having been discussed and debugged in the course of years” (Valente and Breuker [1994], page 144). The top-down approach to legal ontologies bears a close resemblance to Kelsen’s doctrine of the unity of law: “As it is the task of natural science to describe its object – reality – in one system of natural laws, so it is the task of jurisprudence to comprehend all human law in one system of norms” (Kelsen [1941], page 70).

The universalist view of law has a long and distinguished history in legal philosophy. Natural law theory, whose proponents include Aristotle, Thomas Aquinas and Grotius, held that human law in all jurisdictions derive from one root - divine law. Indeed, the Canon Law of the Catholic Church was a pan-European element of many national and local legal systems for centuries. In the nineteenth century, the German Pandettist school sought to create and adapt new norms in response to changing times, or fill the gaps where none existed, in a way that was logically consistent with Justinian’s Civil Code, something that Sacco attributed to the ‘the medieval cult of Roman law, that the Corpus emphius civile was compiled by a sort of divine mandate’ (Sacco [1991], page 346).

Kelsen’s universalist point of view led him to the conclusion that contradictory norms within the same legal system cannot exist (Kelsen [1992], note 7, page 112):

If legal cognition encounters legal norms that contradict one another in content, it seeks, by interpreting their meaning, to resolve the contradiction as a mere pseudo-contradiction. If this effort fails, legal cognition disposes of the material to be interpreted, disposes of it as lacking in meaning altogether and therefore as non-existent in the legal sphere qua realm of meaning.

This is arguably an over-simplification of how the law handles problematic cases, and is disputed by some legal philosophers.

**Comparative Law**

Comparative law has emerged as a descriptive investigation of laws and legal systems, and its analysis can help bring out the meaning of legal terms used by practitioners. One of the most influential comparative lawyers is Rodolfo Sacco. Starting from the premise that norms are not “legal flowers without stem or root” (Sacco [1991], page 27), he identified factors that influence legal interpretation in various jurisdictions.

First, all legal systems have several legal formants, otherwise known as sources of law - codes and statutes, judicial decisions, legal scholarship and political ideologies.

“The civil lawyer may say that this rule comes, in principle, from the code; the common lawyer may say it comes from a particular statute or from judicial decisions; and yet they both will learn their law initially from the books of legal scholars.” (Sacco [1991], page 22)

The importance of these legal formants vary considerably in different jurisdictions and different areas of law - case law is more important in France than in Italy, some areas of English law are subject to more statutes than others - although all these legal formants have some influence, whatever the official model of the law might say. The existence of multiple legal formants creates uncertainty, since they are rarely in complete harmony on a point of law. And yet, this does not usually stop the law from functioning. Sacco cites as an example article 39 of the Italian Constitution, which provides that “duly registered trade unions may enter into collective labor agreements which are binding upon all”. Since there is no legally valid mechanism for trade unions to register, collective labour agreements should not be
binding according to the Constitution. Nevertheless, the courts have consistently enforced such agreements, which leads to the conclusion that in Italy, judicial cases have influence, if not binding precedence, although they are officially not a source of law at all.

Legal formants are not the only factors that influence legal interpretation. There are also the invisible factors of the beliefs or mentality of the interpreters, their social and cultural background. Sacco calls these factors emphcryptotypes. Such factors rarely need to be articulated in a mono-culture. Comparative law helps reveal hidden cryptotypes when a seemingly equivalent rule is interpreted in different ways in different legal jurisdictions, or when an implicit rule is made explicit in another legal system. Sacco cites as an example the issue of whether an heir can transfer property before possessing it. Belgian interpreters of the Code Civil have deemed such transfers invalid, but the French have upheld them. The discrepancy is explained by the fact that while the Code itself does not support such transfers, the old Roman law did, and the custom carried over into French law. A similar situation happened in Italy with the introduction of a new civil code in 1942. Legal scholars interpreted the law in accordance with the earlier doctrine of German Pandettists, convinced that the code was incomprehensible otherwise (Sacco [1991], page 345).

Sacco noted that it is quite common to find that not all legal rules are fully articulated. A synecdoche occurs when only part of a phenomenon is indicated when referring to the whole. He gives the example that the legal definition of contract in French law refers to will of the parties without mentioning the need for the will to be declared or that there needs to be a good reason for the parties to declare their will and for the law to respect it. Filling in the gaps requires knowledge of the legal culture and custom. Unwritten rules are passed on from one generation of jurists to another.

There is a fourth element that Sacco identified as influencing interpretation of law, and that is the choice of language used. Legal language is not always objectively descriptive, merely defining categories and their constituent features. It is also the language of political thought, and some legal expressions have certain connotations. For instance, the word ‘saving’ is more favourable than the French word ‘capitalisation’. ‘Freedom of contract’ is more ideologically connected to the ideal of liberty than ‘Autonomy of the contracting parties’. (Sacco [1991], page 15-16).

Sacco claims that identification of legal formants, cryptotypes, synedcohe and connotation were found “almost as a by-product” (Sacco [1991], page 388) in comparing different legal systems, but has led to more profound understanding of how the law functions than pure theories of law of Kelsen, Hohfeld and others. The reason is that limiting the study of law to a single legal system leads to ignoring certain features that appear ‘obvious’. Such features are not necessarily ‘obvious’ or common to all legal systems, and their discovery uncovers the unwritten rules and values that underpin the law in different legal systems. The comparative approach can go beyond the letter of the law to find its true meaning (Sacco [1991], pages 16-17).

An abstract idea finds concrete expression in a given legal language much as, in biology, a genotype or distinctive set of genes is expressed in the phenotype or outward form of a plant or animal. The jurist of an individual country studies the phenotype. The comparativist must study the genotype of which it is the expression.

Another important contribution of comparative law is the exploration of the interplay between legal formants, and the awareness that while the law constantly evolves, legal formants rarely move together in sync, so that conflicting valid interpretations are inevitable (Sacco [1991], page 394).

Sacco has had an immense influence on legal scholarship, most especially in his bottom-up approach to the analysis of legal concepts. Graziadei [2004] analyses the gradual acceptance of interest as a return on capital in European countries achieved through different paths, ‘civil fruit’ in civil law jurisdictions, and profit in common law jurisdictions. The study shows the influence of changes in religion, economy and philosophy on the evolution
of these concepts in different civil and common law jurisdictions as well as the practical implications of conceptual differences to this day.

**Logic and norms**

The difficulty of negotiating norms and the social values from whence they came were illustrated by Dworkin (Dworkin [1985], page 119):

Suppose the legislature has passed a statute stipulating that “sacrilegious contracts shall henceforth be invalid.” The community is divided as to whether a contract signed on Sunday is, for that reason alone, sacrilegious. It is known that very few of the legislators had that question in mind when they voted, and that they are now equally divided on the question of whether it should be so interpreted. Tom and Tim have signed a contract on Sunday, and Tom now sues Tim to enforce the terms of the contract, whose validity Tim contests. Shall we say that the judge must look for the right answer to the question of whether Tom’s contract is valid, even though the community is deeply divided about what the right answer is? Or is it more realistic to say that there simply is no right answer to the question?

Difficulties can arise even for seemingly straightforward constitutive norms, as argued by Hart (Hart [1958], page 607): ‘A legal rule forbids you to take a vehicle into a public park. Plainly this forbids an automobile, but what about bicycles, roller skates, toy automobiles? What about airplanes? We may call the problems which arise outside the hard core of penumbral instances “problems of the penumbra”. If a penumbra of uncertainty must surround all legal rules, then their application to specific cases in the penumbra area cannot be a matter of logical deduction, and so deductive reasoning, which for generations has been cherished as the very perfection of human reasoning, cannot serve as a model for what judges, or indeed what anyone, should do’.

Which brings the question: why do so many legal ontologies define legal concepts in formal logic? It would seem that they simply follow the predominant trend of ontology development in general. Hepp (Hepp [2007], page 5) states that “For a large fraction of ontology researchers, formal logic as a means (i.e., modality) for expressing the semantic account is a constituting characteristic of an ontology” before going on to say “In our opinion, it is highly arguable whether formal logic is the only or even the most appropriate modality for specifying the semantics of a conceptual element in an ontology.” This is particularly true for law, our domain of interest. Winter (Winter [1989], page 1134) states that “More abstract concepts are likely to show the greatest cultural variation because they are indirect, imaginative extensions of the physical and social experiences that ground them.” Legal conceptualisations are social objects (Searle [1995]) and are jurisdiction-, language-, domain-, and often statute-specific. The differences between such conceptualisations can only really be understood in the rich expressiveness of natural language since “like language generally, legal discourse can never escape its own textuality” (Peller [1985], page 1182).

**Legal sublanguages**

Natural language is a “natural metalanguage” for legal sublanguage, and provides a “preconstructed” linguistic structure on which to build every rule of law. The richness and diversity of natural language is such that it allows everyone, in every context, to find the best way to communicate their ideas, their feelings, their own reasons. But to fully understand what is being said, we have to take into account various semantic layers, something we do automatically and perhaps unconsciously in familiar terrain, but can overlook when wading through unfamiliar languages or sub-languages.

De Saussure, regarded as the father of modern linguistics, saw language as a ‘code’, a set of ‘signs’ that combined together, according to precise rules, allowing communication between
two or more parties. He invented a relational model (the ‘semitic triangle’) representing the ‘process of meaning’. Meaning is described as a triadic relation (or three dyadic relations) between a sign, a concept, and an actual object. For communication to occur, the conceptualisation of a concept must be understood in the same way by speaker and audience. For natural objects, their properties are ‘out there’ and shared conceptualisation is possible according to the scientific standards of the day. If any clarification is required, this can be achieved objectively by studying the object itself. Legal terms such as ‘negligence’ or ‘liability’ are socially constructed concepts and not fully defined for all possible situations. If clarification is required, the process usually involves analogy and invention, which are highly subjective activities. Moreover, the law is applied to dynamic environments, so that meaning is neither static nor logically deductable, (see Hart above). Rawls (Rawls [1955]) stated that legal rules are of necessity more general than the myriad present and future scenarios to which such rules must be applied. He compared this with practice rules in games like poker and chess, where rules are well defined and rarely need to be reinterpreted during or after games have taken place.

**Norm dynamics**

The legal domain is constantly subject to evolving conceptualisations and neologisms. Legal sublanguages, despite their conservative style, show remarkable flexibility and dynamism in redefining new words and coining new ones when the need arises. Tiersma (Tiersma [2006]) states that “the language of lawyers can sometimes be surprisingly creative and innovative. Lawyers are quick to coin a new word when their existing vocabulary is insufficient. Consider recent additions to the legal vocabulary, such as a shrinkwrap license (where a software user agrees to terms contained in software itself, or in a user’s manual, merely by opening the box), or a clickwrap license (where the user clicks on an box or icon, indicating acceptance of the terms of the license).”

Legal concepts may be defined by statute, but are often re-elaborated several times via scholarly or judicial legal interpretation. Legal interpretation is an indispensable part of the legal process, and established interpretation methodologies can modify or extend legal rules. Liebwald (Liebwald [2012]) states that civil law countries often refer to Savigny’s canons of interpretation: grammatical (a literal reading of the norm itself and nothing else), systematic (taking into account the domain or legal system in general), historical (based on the purpose of the norm as revealed in the preamble or preliminary discussions) and teleological (based on the ‘independent will of the norm’ which in practice means the will of the interpreter). Similar canons of interpretation exist in the Common Law tradition: the ‘plain meaning rule’ (which corresponds to the grammatical rule above), the ‘mischief rule’ (corresponding to the historical rule), and the ‘golden rule’ where a word’s usual meaning can be discarded to avoid an absurd result. There are no firm rules on when to use such canons of interpretation, which means that in practice, they are used at will to best serve the demands of justice or the prejudices of the interpreter.

Liebwald (Liebwald [2012]) argues that though natural language is inherently imprecise in general, there is even more vagueness in legislative drafting and judgments, and this vagueness is often intentional: to cover unexpected cases, to leave space for more specific rules in the future, or due to a reluctance to spell out explicit rules where there is political or social disagreement. While the law does contain determinate concepts such as age of consent or speed limits, it is also full of indeterminate and malleable concepts such as good faith or reasonable discretion.

Liebwald (Liebwald [2012]) described the evolution of the meaning of ‘essence of marriage’ as a case in point. The concept derives from The Marriage Act of 1938 in Germany, used to assess whether divorce or annulment is morally justified. During the NS era, marriages between Aryan and non-Aryan spouses, or between wives older than their husbands, were
considered contrary to the ‘essence of marriage’. The Act survived in West Germany and East Germany, but were interpreted quite differently. In West Germany, marriage was interpreted as an absolute, predetermined moral order, and therefore indissoluble. In East Germany, however, marriages were evaluated in terms of the spouses' ability to serve social ideals, particularly the ability to work. The Act has also survived in Austria, where the ‘essence of marriage’ is defined as the possibility of parenthood, whether or not the spouses want to, or are able to, have children. According to this definition therefore, same-sex couples cannot get married.

Not all legal reasoning revolves around legal concepts, or common words ‘technicalised’ and imbued with a specific legal meaning. A legal sublanguage is in constant and continuous interaction with ordinary language, and uses all its linguistic baggage. Metaphors are another linguistic feature often used to represent and navigate conceptions, in general language and in legal language. Lakoff (Lakoff [1993], page 206) showed how ‘journey’ is often used in every day life as a metaphor for relationships:

Look how far we’ve come. It’s been a long bumpy road. We can’t turn back now. We’re at a crossroads. We may have to go our separate ways. The relationship isn’t going anywhere.

Winter (Winter [1989]) used the American case of NLRB v Jones & Loughlin Steel (1937)\(^{119}\) to show how the power of metaphors is used in legal reasoning. The key question in that case was whether the federal government, under the commerce clause, could regulate labour relations in manufacturing. Kidd v Pearson (1888) 3 had held that manufacturing is not commerce (invoking the ‘container’ metaphor), since manufacturing is ‘purely local’. An alternative ‘stream of commerce’ metaphor was used in Stafford v Wallace (1922) 4 in which the “Court found that the stockyards were but a ‘throat’ through which the current of commerce flowed”. Judge Hughes took this metaphor as an 'illustration' before going on to reconceptualise commerce as a traveller going along well-defined path, whose journey should be protected from undue burdens, obstructions and dangers. In this way, Hughes overcomes the ‘container’ (P or not P) metaphor to allow federal law to extend its influence. Winter (Winter [1989]) concludes that Hughes's reasoning was a ‘metaphoric tour de force’ that was neither predetermined by the materials or completely arbitrary.

**Multilinguality and Multijurisdictionality**

Achieving shared conceptualisations of law is difficult in any legal system. The problem is confounded in Europe, which is increasingly governed by multiple jurisdictions - European, national and sometimes regional as well.

Whole sectors of diverse legal disciplines have experienced gradual ‘harmonization’, as national laws are increasingly derived from European law. In the case of Regulations the implementation is automatically ‘binding in its entirety and directly applicable in all Member States’, whereas directives are ‘binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and methods.’\(^{120}\)

Harmonisation of EU law is a complex matter, particularly for Directives, because absolute uniformity or equivalence is not required. The principle of subsidiarity means that Member States are obliged to achieve the objectives of EU Directives, but have some flexibility in how they do so. This inevitably leads to differences among EU norms and various national norms. The principle of subsidiarity is a compromise solution arising from opposing perspectives on

---

\(^{119}\) NLRB v Jones & Loughlin Steel (1937) 301 U.S. 1 1937 3Kidd v Pearson (1888) 128 U.S. 1, 20 4Stafford v Wallace (1922) 128 U.S. 495, 521

\(^{120}\) Article 249, Treaty of Rome 1957
Community law: those who hope for future standardization against those who prefer to maintain differences that reflect local contexts.

Regulation No. 1 of 15 April 1958 recognises twenty-three languages as having the status of official and working languages in European institutions. This poses a significant challenge, since each of the twenty-seven Member States that make up the European Community has its own cultural baggage that no one, let alone the Community legislature, is able to escape. The Sapir-Whorf hypothesis (Hoijer [1954]) maintains that it is impossible for a concept in one language to be imported wholesale into another language due to linguistic relativity. The European Union, on the other hand, operates on the basis that such things are perfectly manageable.

How do European legislative bodies manage the discussion, drafting and publication of European law in twenty-three languages at supra-national and national levels? Analysing procedural steps, it turns out that while debates on legislation in the European Parliament and Council can take place in all 23 official languages with the assistance of interpreters, the working draft legislation under discussion is usually only available in one language: English or French and occasionally German. At the end of the drafting phase, a team of specialist legal translators translate the text into the other official community languages, subject to consistency checks by the General Translation Team.

This way of working, along with a particular emphasis on consistency, has resulted in the development of an European legal language, commonly referred to as Eurocratese. Born almost spontaneously, Eurocratese has gained importance over the years, as it greatly facilitates one-to-one translations. It is made possible through the use of an Acquis Communitaire glossary, designed to be impartial, and which is widely shared at Community level. This glossary is designed to minimise conceptual differences and semantic ambiguity in the EU legal process, and is of great importance in the activities of translators as well as legal professionals.

However, it does not solve the problem of conceptual and terminological misalignment altogether, since Directives need to be transposed into national law using terms that make sense within the national legal system. In fact, it is precisely this second level of translation that causes most problems. The problem is that transposing a rule often means having to use and adapt a different, and sometimes conflicting, lexical baggage to the traditional national one, i.e. a product of an incoherent mix of different cultures and traditions, exacerbated by translations that are inaccurate or insufficiently precise form a legal point of view.

There are several possibilities in the transposition process, as depicted in Figure 9. The general scenario is composed by a concept ec1 in some European directive, which is expressed in the translation of the directive in some national jurisdiction with the term et1. When such directive is transposed in the national law, the concept ec1 will correspond (hopefully but not necessarily coherently) with some national concept nc1 expressed by the term nt1. Different situations arise depending on whether nc1 and nt1 already existed at the national level and if they were already associated together or with other terms and concepts, where only the first two do not raise misalignment problems:

1. The concept nc1 is new at the national level and expressed in the transposition with the same term used at the European level nt1 = et1, not yet used at the national level.
2. The concept nc1 was already present at the national level and expressed by the same term used at the European level nt1 = et1.
3. The concept nc1 is new at the national level and expressed in the transposition with the same term used at the European level nt1 = et1, which was already used at the national level to mean concept nc2. Polysemy is introduced at the national level.
4. The concept nc1 was already present at the national level but expressed by another term than the one used at the European level: nt2 et1. Synonymity is introduced at the national level.

5. The concept nc1 was already present at the national level and expressed by the same term used at the European level nt1 = et1, but the meaning of nc1 is more general or more specific than the corresponding European concept ec1.

The misalignment problems can occur also in case of complex terms, where even if the overall term is transposed coherently, it may happen that some component term has at the national level a different meaning than the one referred to at the European level. So, it may happen that a legal concept expressed in an European Directive is not present in a national legal system. To make sense in the national legal system, European legal terms have to be translated into comprehensible national terminology. For example, the concept corresponding to the word ‘reasonably’ in English, is translated literally into Italian as ‘ragionevolmente’ in the EUD, and as ‘con ordinaria diligenza’ in the transposition law.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>lauterkeit good faith buena fe</td>
<td>Treu und Glauben good faith buena fe</td>
</tr>
<tr>
<td>EN</td>
<td>loyalty</td>
<td>buona fede</td>
</tr>
<tr>
<td>ES</td>
<td>lealtad</td>
<td>correttezza e buona fede</td>
</tr>
<tr>
<td>FR</td>
<td>foi buona fede</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>fede</td>
<td></td>
</tr>
</tbody>
</table>


Terms for ‘good faith’ in Italian European Directives and their national implementation Literal or inexact translations are chosen ad hoc in Directives and transposition laws, as 2.6 and 2.6 show.

While such a terminological misalignment can lead to confusion, far more insidious is the effect of conceptual misalignment. For instance, in Directive 97/7/EC, Art. 4.2, the legal terms ‘klar und verständlich’ in German, ‘clear and comprehensible’ in English, and ‘chiaro e comprensibile’ in Italian, appear to have the same meaning until read with the filters of lawyers’ domestic legal thinking. ‘Klar und verständlich’ in the German system is considered by German commentators as referring to three different legal concepts:

1. The font must be clear and legible (gestaltung der information),
2. The information must be intelligible to the consumer (formulierung der information),
3. The language of the information must be in the national language of the consumer (sprache der information).
In Italy, the judiciary tend to check more for concepts 1 and 3, and less concept 2, while in England and Wales, more attention has been paid to concept 2, though interpreted as requiring a plain style of language (as opposed to technical legal jargon) due to the historical influences of the Plain English Campaign[121] in the United Kingdom.

These problems are confounded by the fact that in every jurisdiction, terms have different meanings in different contexts and over time. This can make it difficult to interpret national legislation that transpose a set of European Directives all at once. Which Directive provides the authoritative definition of a legal term? Unless one Directive overrides another, the definitions can be regarded as complementary, and the meaning of a term is the union of all the definitions in all the Directives being transposed.

Another issue to bear in mind is that definitions of terms in European Directives and national legislation are sector-specific, and different definitions will apply in other sectors. Consider, for instance the definition of ‘consumer’ in the English version of European Directive 1999/44/EC, Art. 1.2 (on sale of consumer goods):

consumer: shall mean any natural person who, in the contracts covered by this Directive, is acting for purposes which are not related to his trade, business or profession that has a different meaning with respect to the definition of consumer given in Council Directive 90/314/EEC, Art. 2.4 (on package holidays):

“consumer” means the person who takes or agrees to take the package (‘the principal contractor’), or any person on whose behalf the principal contractor agrees to purchase the package (‘the other beneficiaries’) or any person to whom the principal contractor or any of the other beneficiaries transfers the package (‘the transferee’).

Revising the European law

The European Commission has acknowledged these problems. In its technical report for the European Commission, the Mandelkern Group Mandelkern [2001] admitted “lack of simplicity, clarity and accessibility of European provisions — such as unclear, confusing terminology, incomplete or inconsistent regulations or use of vague terms”. Moreover, they stressed the need for internal coherence and consistency in the use of EU legal terms, as well as external coherence, consistency in the transposition of legal concepts into national law. In of the European Communities [2003], the Commission stated that certain market inefficiencies can be attributed to diverse interpretation of European contract law due to vague terms and rules, and to the fact that EU contract law consists of several highly sector-specific directives rather than a coherent approach to contracts in general. Their solution was a Common Frame of Reference (CFR) which would provide common principles, terminology, and legal rules to address gaps, conflicts, and ambiguities emerging from application of European law.

A research group aiming at consolidating existing EC law worked on the ‘Principles of the Existing EC Private Law’ or ‘Acquis Principles’ (ACQP) which would provide a common terminology as well as common principles to constitute a guideline for uniform implementation and interpretation of European law. The Acquis Principles were sketched by scholars in European Private Law from the so-called emphAcquis communitaire, a collection of the existing body of EU primary and secondary legislation as well as European Court of Justice decisions. Whether the Acquis Principles can truly address the considerable challenges it seeks to address is a moot question (Legrand [2006]).

[121] http://www.plainenglish.co.uk/
4.7.3 Legal taxonomy Syllabus

Legal Taxonomy Syllabus was designed to help legal professionals acquire deep understanding of laws in European and other national jurisdictions. The goal of the ontology necessitated a completely new design and level of detail in comparison to ontologies designed for information retrieval or other semantic web applications. The most important insight from lawyers, which informed our design, was that the meaning of a legal term depends on context (jurisdiction, domain, legislation, timeframe). We designed an ontology framework which makes these considerations explicit, so that users can determine which definitions are most relevant.

Given the analysis above, we design the Legal Taxonomy Syllabus, an ontology framework supported by a web based tool to support the work of legal practitioners and scholars in multilingual and multi-level contexts such as the European Union’s. Thus, the system is aimed at humans, to improve the possibility of sharing technical knowledge but also to analyse law in its complexity. As a secondary aim, the system can be used for building automated tools, e.g., for information retrieval and translation. Given our objective, the first requirement becomes the acceptability by legal experts of such a tool. Since we aim at a multi-jurisdictional analysis, resting on a comparative approach Rossi and Vogel [2004] seems the obvious choice. The benefit of this choice is also that comparative law is closer to the practice of law, and thus it fosters acceptability. The comparative law approach brings with it the need for a bottom-up approach: first lower level concepts have to be addressed in a contrastive manner among different systems, to get a deeper understanding of law. Starting from basic elements will also give rise to less disagreement. To highlight this focus, in the LTS the concepts are called unitary meanings (umeaning). Moreover, we adopt the view of comparative law that norms rest on different formants and not only on legislation. So each concept will be linked to different textual sources, including case law and doctrine. The link with the textual dimension will also constitute the context of applicability of the concept.

The second choice driven by acceptability is using a lightweight ontology rather than an axiomatic one. This choice at the same time enlarges the possible audience of users and reduces the costs of building an ontology. As discussed, it is also driven by the consideration that many peculiarities of law, like interpretation, penumbra, interaction with social values, metaphors, are far from having commonly accepted solutions in logic. On the other hand, the ontology is suited for modelling interpretation and temporal evolution of concepts on the basis of explicit (change in law) or implicit (e.g., interpretation) changes.

From the need of modelling European law, we derive the requirement of multilinguality and multi-level: for this reason we will keep terms associated with languages and distinct from concepts and will model separately the European and the different national levels, allowing links among the different ontologies.

Below we describe the LTS ontology framework, with examples taken from the study\textsuperscript{122}.

Legal Taxonomy Syllabus was developed based on the following assumptions about the domain:

- Law is a highly polysemous domain, and the best way to manage this is by modelling concepts and terms (lexical entries) separately, with links between the two type of elements;
- The meaning of terms and norms are highly context-sensitive and sometimes not universally agreed. Terms are defined in multiple sources, and legal professionals need access to all valid definitions in order to find the most appropriate definition for a particular context.

\textsuperscript{122} http://www.uniformterminology.unito.it/
Sometimes the purpose behind definitions can give additional clues as to its suitability. Sometimes generalised definitions develop from combinations of several context-specific definitions. An ontology that is able to model such complexities in a structured way would help professionals interpret legal terms appropriately.

Legal concepts are constantly evolving, particularly with the introduction of new legislation with new definitions. A legal ontology tool should provide some mechanism for keeping track of these changes, especially since deprecated definitions may still have some relevance.

Constitutive norms are a good source of object definitions. Prescriptive norms are complex and rarely fully articulated in one section in legislation. Modelling prescriptions requires complex object types.

The European Union and national jurisdictions have their own legal terminology and therefore should have their own legal ontology. Corresponding concepts at the EU and national levels can be denoted by different terms in the same national language, and there should be explicit connections between such terms. However, we do not assume that the transposition of an EUD necessarily introduces a national ontology the same concepts that are present at the EU level.

The best way to model law accurately is by comparative law methodology. Any sustainable ontology tool thus has to usable for legal experts in multiple jurisdictions, so that they can populate and update the ontologies on a regular basis.

For collecting and sharing knowledge across different jurisdictions, a web-based distributed workspace is better rather than a stand-alone application. In this way, different teams can work on different languages at the same time from different places.

Different types of contributors need specific interfaces for their work. Legal experts need an intuitive interface for annotating terms and relating them to concepts with basic semantic relations. Ontology experts can then use a different interface to add more complex relations among the concepts.

Contributors and end-users need user-friendly interfaces to be able to visualize and navigate ontologies at both European and national levels, and the relations between different ontologies. They also need links between ontologies and legislation in order to find patterns of usage and inconsistencies.

Populating and updating ontologies is labour-intensive and costly. An advanced ontology building tool should incorporate advanced technologies in legislative XML and natural language processing to help overcome the ‘resource bottleneck’ (Hepp [2007]) of creating and updating ontologies.

**Polysemy and Terminological Variation**

Legal terms are often defined differently in different legislation, and each definition can be regarded as an unique concept. We call these definitions *umeanings*. At the same time, some concepts can be expressed in alternative terms. It is therefore useful and natural to separate concepts and terms. The concepts are arranged into ontologies. Lexical terms are then linked to all concepts that apply. This organisation is evocative of de Saussure’s (Floridi [1912]) semiotic triangle referred to in 2.4.

For the end-user, each term in the ontology is presented in an associated information table with the following information:

- language
- jurisdiction
Doctrinal interpretation

While legal professionals need access to legislative definitions to have a precise model of European law, we accept the idea of several scholars (Sacco [1987], Vanderlinden [1995] and de Groot [1999]) that there is more to legal knowledge than the knowledge contained in statutory rules, since legal professionals also take into account case law and jurisprudence. While different detailed definitions is important during the interpretation of specific legal cases, it is equally important to have a sense that abstracts from the peculiarities of specific legislation by considering a broader fragment. As more legislation-specific definitions about terms are collected, it becomes possible to find common attributes and derive a more general definition. This endeavour is in keeping with our bottom-up philosophy.

Lawyers and legislators often combine highly sector-specific concepts into more abstract concepts with broader meaning, in order to describe (complex) entities, such as the consumer protection in all its aspects. This is particularly important when considering European Union terms, since different definitions may not all make their way into national legislation. For example, the Italian legislature has transposed sets of European Directives into unified Italian legislation, rather than transposing each one individually. The legal concepts are defined as the union of all the concepts provided by the individual Directives in a specific sector, as a result of doctrinal interpretation of the Directives.

Legal Taxonomy introduces a kind of ‘abstract’ concept to deal with this situation (see Figure 11). Abstract concepts are not linked to single pieces of legislation, but are conveniently recognized as a generalisation of legislation-specific concepts. We introduced a new kind of ontologic relation called GROUPED: it is a non-transitive relation where the more general meaning, that we will call group leader, represents the abstracted concept that groups the meaning of a number of more specific meanings that are the sectorial meanings defined in individual Directives or national laws. Users can navigate the ontology at different levels of detail, depending on their goals. However, when the user searches into the meaning database, more specific meanings are excluded from the results unless the user explicitly asks to show them, i.e. only the group leaders are shown in the results by default.

Evolution of terms

Legislation are written at different points in time. When new legislation is approved and enacted, it can introduce a number of new meanings which can change a number of old meanings defined by old laws, so that these old meanings become obsolete. Normative change (Palmirani and Brighi [2006]) is an open issue in building tools for describing legal frameworks. There are two types of normative change: explicit change and implicit change. In the first case, the new norm explicitly states the abrogation of a specific paragraph of an old law (for details, please refer to Spinosa et al. [2009]). Alternatively, the new law can
define a term in contradiction to previous laws without mentioning these laws explicitly. In this case, the definition provided by the new law renders the old definition obsolete.

We respond to this problem by introducing a temporal dimension into the ontology, allowing new concepts to replace old ones while still retaining the old concepts in the system for reference purposes (this information is valuable for understanding the evolution of terms). We are aware of the difficulties of modelling time in artificial intelligence and in formal ontology. Legal Taxonomy Syllabus adopts a naive solution similar to the one adopted for interpretation and abstraction of norms, that is to empower LTS with a new ontological relation called REPLACED BY.

The new ontological relation has some peculiar characteristics that distinguish it from the usual ontological relations (Figure 11):

- A REPLACED BY relation brings with it a new data field not present in the other relations: the substitution date.
- When the user searches in the umeanings database, replaced concepts will not be shown unless the user asks for a certain date in the past. This enables the user to obtain a snapshot of the legal ontology that was valid at any particular moment.
- When a new umeaning replaces an old one, all the ontological relations that applied to the old umeaning are automatically applied to the new umeaning. If some of them are no longer valid for the new umeaning, manual intervention from the user is required.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IS A quasiSyn</td>
<td>Something is a specific kind of another thing Two concepts are almost, but not quite, synonymous</td>
<td>Transitive Non-transitive Non-transitive Transitive Non-transitive</td>
</tr>
<tr>
<td>INVERSE Of</td>
<td>Something has the opposite meaning of another thing</td>
<td>Non-transitive Non-transitive</td>
</tr>
<tr>
<td>PART OF HAS</td>
<td>Something is part of another thing</td>
<td></td>
</tr>
<tr>
<td>VALUE</td>
<td>A legal principle behind a norm Links a term to the legal principle behind it</td>
<td></td>
</tr>
<tr>
<td>PURPOSE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ontological Relations in Legal Taxonomy Syllabus**

**Other Relations Among Terms**

The LTS ontology has a number of other ontological relations designed to help users understand the inter-relationship among concepts:

**Multi-level ontologies**

Legal Taxonomy Syllabus is not just one ontology. It is a framework of interlinked but separate ontologies for different languages and jurisdictions in Europe. So far, the data collected have been on the subject of consumer law. The actual number of annotated terms and concepts are provided in Tables 4 and 5, respectively. Terms were initially extracted from a corpus of 24 EC directives, and 2 EC regulations. Occurrences of such entries were detected from national transposition laws of English, French, Spanish, Italian and German jurisdictions, see Figure 12.

Within LTS, we can talk about direct EU-to-national translations of terms, and about implicit national-to-national translations of terms. In other words, we distinguish between explicit and implicit associations among concepts belonging to different levels. The former ones are direct links that are explicitly used by legal experts to mark a relation between concepts. The latter ones are indirect links: if we start from a concept at a given national level, by following a direct link we reach another concept at European level. Then, we will be able to see how that concept is mapped onto further concepts at the various national levels.
It is possible to use the LTS to translate terms into different national systems via the transposed concepts at the European level, i.e. by using the implicit associations.

<table>
<thead>
<tr>
<th>Language</th>
<th>National</th>
<th>European</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>8</td>
<td>47</td>
</tr>
<tr>
<td>Italian</td>
<td>28</td>
<td>52</td>
</tr>
<tr>
<td>English</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>Spanish</td>
<td>41</td>
<td>60</td>
</tr>
<tr>
<td>German</td>
<td>66</td>
<td>98</td>
</tr>
<tr>
<td>total</td>
<td>214</td>
<td>332</td>
</tr>
</tbody>
</table>

Number of concepts

For instance suppose that we want to translate the legal term ‘credito al consumo’ from Italian to German. In LTS, ‘credito al consumo’ is associated to the national meaning Ita-175. We find that Ita-175 is the transposition of the European meaning EU-26 associated with the term ‘contratto di credito’. EU-26 is also associated with the German legal term ‘Kreditvertrag’ at the European level. We then find that the national German transposition of EU-26 corresponds to the national meaning Ger-32 that is associated with the national legal term ‘Darlehensvertrag’. Then, by using implicit links in the European ontology, we can translate the Italian legal term ‘credito al consumo’ into the German legal term ‘Darlehensvertrag’.

Ajani et al. [2007] shows how concepts from different ontologies are interlinked to help users find similar terms in other languages and jurisdictions.

The Italian term Term-Ita-A and the German term Term-Ger-A are used as corresponding terms in the translation of an EU directive, as shown by the fact that both refer to the same EU-concept EU-1. In the Italian legal system, Term-Ita-A has the meaning of Ita-2. In the German legal system, Term-Ger-A has the meaning of Ger-3. The EU translations of the directive is correct insofar as no terms exist in Italian and German that characterize precisely the concept EU-1 in the two languages (i.e., the ‘associated concepts Ita-4 and Ger-5 have no corresponding legal terms).

A practical example of such a situation is reported in Figure 9, where we can see that the ontologies include different types of arcs. Beyond the usual is-a (linking a category to its supercategory), there are also the arcs purpose, which relate a concept to the legal principle motivating it, and concerns, which refer to a general relatedness. The dotted arcs represent the reference from terms to concepts. Some terms have links both to a National ontology and to the EU Ontology (in particular, ‘withdrawal’ vs. ‘recesso e difesa del consumatore’ vs. ‘consumer protection’).

The last item above is especially relevant: note that this configuration of arcs specifies that: 1) ‘withdrawal’ and ‘recesso’ have been used as equivalent terms (concept EU-2) in some European Directives (e.g., Directive 90/314/EEC). 2) In that context, the term involved an act having as purpose some kind of protection of the consumer. 3) The terms used for referring to the latter are consumer protection in English and difesa del consumatore in Italian. 4) In the British legal system, however, not all withdrawals have this goal, but only a subtype of them, to which the code refers to as cancellation (concept Eng-3). 5) In the Italian legal system, the term ‘diritto di recesso’ is ambiguous, since it can be used with reference either to something concerning the ‘risoluzione’ (concept Ita-4), or to something concerning the ‘recesso’ proper (concept Ita-3).
**Acquis Communitaire Principles ontology**

We have mentioned the issue of the evolution of terms over time, and how a REPLACED BY relation can be used to connect a definition that replaces another. As Legal Syllabus Ontology was being constructed, there were efforts to harmonise European Glossary and a draft Acquis Communitaire Principles glossary was issued. This glossary was intended to replace existing European terminology in the future. However, while the glossary had not exactly emphreplaced the terms and concepts in European law, it had some influence on legal interpretation.

We responded to this situation by creating a separate ontology for Acquis Communitaire Principles which is closely to the EU ontology of current definitions. The Acquis Communitaire Principles ontology was introduced into the LTS by defining explicit associations between Acquis Principles concepts and EU-level concepts.

For example, in 6 we have that the concept EU-25 (corresponding to the English legal term creditor) present in a EUD is explicitly associated with the national legal concepts Ita-124 (finanziatore) and Spa-110 (prestamista) for Italian and Spanish, respectively. We can add the term ‘creditor’ from the Acquis Level by inserting an explicit association between the Acquis legal concept AC-72 and the European legal concept EU-25. As a consequence, concept AC-72 is implicitly associated with concepts Ita124 and Spa-110.

This has deep consequences on the way one can build systems for reasoning, that are allowed to make paths passing through more than two levels, thereby offering new insights (and ready-to-use associations between terms) to scholars in comparative law.

**Ontology of Prescriptions**

A recent development to the Legal Taxonomy Syllabus framework has been to extend the type of concepts described to include prescriptive norms (as opposed to only the definitions found in constitutive norms). A prescription is treated as a complex type of concept, which is subject to all the relations of simple concepts (is a, part of, replaced by etc.). However, each prescription is also necessarily connected to other relevant concepts defined in the ontology via the following relations:

- Deontic clause: the type of prescription: obligation, prohibition, permission, exception.
- Active role: the addressee of the norm (e.g. director, employee).
- Passive role: the beneficiary of the norm (e.g. customer).
- Crime: the type of crime resulting from violation of the prescription, often defined in other legislation such as the Penal Code.
- Sanction: the sanction resulting from violation (e.g., a fine of 1 quote, where quote is defined in other legislation).

This complex type is a powerful way to structure information about norms that are usually scattered across different sections, not necessarily from the same piece of legislation. It makes it easy for users to quickly find norms according to various search criteria indicated by the above relations. The collection of such information in an ontology also enables semantic rather than mere keyword search, for instance, a search for prescriptions that apply to sellers will also bring up prescriptions for traders, its grandparent in the ontology graph. The prescriptions ontology has been populated with 500 Italian national legal terms in the domain of regulatory compliance in finance. This extension is discussed in more detail in Boella et al. [2012].
4.7.4 Implementation issues

Distributed platform

The LTS ontology framework was based on the Gene Ontology project\(^{123}\), from which it inherits the logical and graphical representation, and the ability to import ontologies and export them to representation languages such as OWL. The Gene Ontology consortium (often referred to as GO) was founded with the aim of creating a consistent terminology between different databases. The resulting tools and databases were then made freely available. The Gene Ontology consortium provides several tools: Go database, whose schema is public; Amigo, a web-application for consulting ontologies; and DAG-Edit.

In the Gene Ontology, GO terms are viewed as graph nodes and relations as arches with labels identifying the type of relation. The nodes and arcs are implemented by the term and term2term relations, while the types of relationships are also stored as terms in a special ontology for cataloguing types of relations. Legal Taxonomy Syllabus manages ontologies in a similar manner, but with two differences: first, the relations are not between terms, but between terms and their meanings, and secondly, the types of relations are considered not as terms but as separate concepts stored in a special table.

Key aspects of the management of Legal Taxonomy Syllabus ontologies would be used by legal experts who do not have specialist IT knowledge. Moreover, since the creation of multi-level ontologies is a coordinated effort between several groups, the system should ideally be easily accessible from different computers without the constraints of any particular operating system or the need to install special software. These two constraints led us to exclude applications for the management of ontologies such as the Gene Ontology’s own DAG-Edit or the leading Protégé system\(^{124}\), which although very complete, are highly complex and use local files making it difficult to collaborate between different stakeholders. It was decided that the most appropriate solution would be to design our own user-friendly web-based interface to manage ontologies via a database server offering the scalability, concurrency, and integrity checks required. Unlike Amigo, the LTS web interface not only displays the dictionaries and ontologies, but also provides a mechanism for creating and deleting terms, definitions and ontologies. The LTS web application is implemented in PHP.

In the web application, the level of logic is divided into two modules, User Interaction Handler and Business Logic Layer, both of which are divided into several modules designed to perform a specific function. The classes Smarty\(^{125}\) and GraphViz\(^{126}\) are two external projects used to connect the level of User Interaction Handler at the level of the user interface itself. The GraphViz class manages the generation of an image representative of the graph of the ontology of a given umeaning through the homonymous tool AT&T. The Smarty class is a vehicle for the interpretation of HTML templates. It comes with a very consistent user interface without having to affect the structure of the main application.

Security and Integrity

While users interact with LTS through a web browser, the system has to support concurrent access and modification of data to maintain its integrity. Authentication is required to access the editing part of the system, and any changes are recorded in a log file to track what changes were made and who made them. Complex queries that require the use of a cycle of


\(^{124}\) [http://protege.stanford.edu/](http://protege.stanford.edu/)

\(^{125}\) [http://www.smarty.net/](http://www.smarty.net/)

\(^{126}\) [http://www.graphviz.org/](http://www.graphviz.org/)
several SQL queries on the target ontology are made more efficient by ensuring that the path from each node to all its ancestors is precomputed by reflexive transitive closure.

The state of the system is always checked after operations that could potentially lead to an inconsistent state. Each term must always be defined by at least one umeaning, and each umeaning must always be associated with at least one term. When the user enters a new term, the system checks that there is at least one umeaning associated with it; likewise, if the user enters a new umeaning, the system checks that there is at least one term associated with it. If all is well, the system proceeds with the insertion, otherwise the operation is cancelled and an error message returned.

For the insertion and deletion of ontological relations and associations between national and European umeanings, the system checks that the two umeanings the user wishes to connect exist and are in the same language.

To enter any ontological relations between different umeanings, the following constraints must be met:

- The user must be authenticated and have the necessary permissions to enter an umeaning.
- The umeaning elements to be connected need to exist in the database.
- The type of ontological relationship specified must exist in the database.
- The two umeaning to be connected must be on the same level and have the same legal language.

There are also particular constraints and integrity checks to respect the semantics of the grouping concept:

- Each umeaning can only belong to a single group.
- A group leader cannot exist without group members.

**User Interface**

The LTS web interface enables the addition, editing and deletion of umeanings, equivalence relations between national and European umeanings as well as ontological relations between umeanings in the same ontology.

In the development of the content of LTS ontologies, we use a two-phase procedure. In the first phase, terms are collected in a database together with the legal sources where they appear, and the underlying concepts were identified. In this phase, polysemy is identified and different conceptualisations represented separately. Also, terms at the European level are associated with their transpositions in the national languages. In the second phase, for each different ontology, the set of concepts are organized in an way that is suitable for the relevant jurisdiction. The result is a principled lightweight, but not axiomatic, ontology.

The first phase of defining legal concepts in the context of European directives and national legislation is quite protracted, and requires contributors to search and read a large number of documents. In order to facilitate this process, the LTS system includes a database that contain full versions of the desired European directives and national laws. The legal expert can specify correspondences among the terms found after manual inspection of legal documents in a controlled way, and save the result of his/her analysis in an ontology structure. The workflow is as follows.

- The user creates a new umeaning linked with the term (s)he wants to define.
- (S)he selects relevant citation from legal text; and the main term attached to the umeaning is used as the default query.
• After choosing one of the search results, the full text of the legal document is displayed, with the search terms highlighted.

• Finally, the user selects the text that should go in the citation and confirms the insertion in the references database.

The link between the concepts, terms and original definitions in legislative sources remain thereafter. This means that should users wish to review concepts, they can the documents that informed the given legal term.

**Enhancement with language technologies**

In order to speed up the annotation process, LTS provides a number of facilities based on natural language processing. When defining a new concept it is important to collect all occurrences of terms that implement that concept. We devised an intelligent search procedure for retrieving the terms which is based on inverted index, stemming, function words, and document similarity.

The inverted index has a quite simple structure: it is a relational table that maps terms to documents that contain them, along with the position of their occurrence in the documents. Documents belonging to national and European levels are stored and indexed in separate relational tables.

The search for a given term in the documents database is performed on the root (stem) of a term, instead of searching for the exact term. For instance, when an user performs a search on the term ‘contracts’, documents containing ‘contract’ will also be found. This is commonly acknowledged to enhance information retrieval performance Krovetz [1996].

Inverted index based on stemmed words can be computationally expensive, but the index size can be reduced significantly by excluding function words such as prepositions and articles that lack real semantic content. For instance, the size of the index of the EUD 90/314/EC decreases from 2336 records to 1152.7.

By using the inverted index technique we obtain, as a useful side effect, a statistical distribution of the words in the documents. Such statistics allows us to define a document similarity measure based on the cosine document similarity technique (Lin [1998], Isaacs and Aslam [1999]). This is useful for many tasks, such as annotating a new Directive with the definitions of prefixed terms from previous Directives on the same topic (unless, of course, the terms are defined anew in the new Directive).

**4.7.5 Comparison with state-of-the-art**

The purpose of Legal Taxonomy Syllabus is to help legal and business professionals understand the meaning of terms used in European and national legislation, particularly to help compare laws and interpretation of laws in different countries in the European Union. Current instruments provided by the European Union have several limitations. In particular, the Eurovoc thesaurus is based on the EU Treatise structure rather than on a taxonomy of law. That structure denotes the origin of the European Union as an international organisation, but has no connection with the classifications implied in the legal discourse within the Member States and in the national enforcement of EU law. Moreover, Eurovoc is a multilingual terminology in one dimension, that of the European Union, while ignoring the difference in terminology used in national laws. This drawback that is also shared by IATE\(^{127}\).

\(^{127}\) iate.europa.eu/
It is clear that a more systematic way of organising legal terminology is needed, and the most advanced technology for this is ontologies.

Unfortunately, there simply aren’t many legal ontologies created for human understanding, let alone ontologies for comparing terminologies in European and national jurisdictions. The most well-known and influential legal ontologies are top-down abstract explorations of legal notions and their relations to real-world objects, designed for machine reasoning, and informed by legal philosophy and cognitive science. Notable examples are Valente’s functional ontology of law, (Valente [1995]), van Kralingen’s frame based ontology, (van Kralingen [1995]), Gangemi et al’s Core Legal Ontology (Gangemi et al. [2005]) and LKIF Core (Hoekstra et al. [2007]). Notable for their universal conception of law, it is questionable whether the data collected is precise enough for reasoning on legal concepts in practice. For instance the LKIF Core Ontology set out to model basic concepts of law identified by citizens, legal professionals and legal scholars, as well as top-level concepts (location, time etc) and intentional concepts (beliefs, roles). They struggled to reason on EU Directive 2006/126 on driving licences, a relatively straightforward regulation, due to the coverage that was realistically achievable due to the known problem of ‘resource bottleneck’ (Hepp [2007]). Their list of basic legal concepts initially came to 250, but had to be scaled down to 50. Perhaps the biggest challenge for creating ontologies for machine reasoning, as opposed to merely for human understanding, is the level of basic knowledge and interconnections a machine needs to be provided with. However, if it’s just about possible to model precise terms such as ‘at least 16 years old’ or ‘maximum design speed of not more than 45 km/h’ (Hoekstra et al. [2007]), open-textured norms are fraught with problems (Rissland and Skalak [1991]), making it difficult to determine what is prohibited or permitted by deduction even from one piece of legislation, let alone the effect of other legal formants.

The EU Employee Legal Ontology by Desprès and Szulman (Desprès and Szulman [2007]) is an example of a precise domain ontology. The topic is employment law in European Directives, and the ontology was developed with a bottom-up approach. Terms were selected from two European directives about employees, and structured into two different micro-ontologies. The ontologies contain terminological concepts and non-terminological concepts. Terminological concepts are the terms themselves, with legal definitions in natural language, and a list of synonyms. Non-terminological concepts are labelled relations between terminological concepts, either structural (classic relations such as hypernyms or meronyms) or functional (domain-specific relations e.g. ‘is protected by national employment law’). The two micro-ontologies were then merged, and connected by subsumption to higher-level concepts defined formally in the LRI-Core Legal Ontology (Breuker [2004]), and the DOLCE ontology (Borgo and Masolo [2009]). The use of textual and formal elements in this work is indicative of a dual purpose. The work was initiated by French jurists who wanted to compare the use of the concept ‘employee’ within the directives before their transposition into national law. However, its stated aim is to support the development of formal models of legislation to be used in legal knowledge-based systems. The way in which legal knowledge-based systems might use the ontology is not described, and neither is the issue of open-textured norms. More interestingly in the context of the focus of the present paper, the ontologies lack information which we consider important for human understanding of terminology in European directives. The relationship between European and national terminology was not explored, even though the directives themselves state that employees are subject to employment law in national jurisdictions. No language other than French was considered, even though the meaning of European terms can be influenced by terminology in other languages and jurisdictions. Neither were other other legal formants (jurisprudence, legal scholarship) consulted in exploring the meaning of concepts. Definitions for terms were taken only from the directives, which were regarded as ‘self-contained and cohesive’ (page 191, Desprès and Szulman [2007]), without any input from legal experts save for final evaluation. Moreover, while the ontology has the expected classical ontological relations, and is also rich in domain-specific relations, relations between different conceptions of the same terms are not described e.g. after merging definitions of terms, is it still possible to
access the more context-specific definitions in the original micro-ontologies? When terms are redefined due to amendments to directives, is it still possible to access earlier conceptualisations?

Recent years has witnessed legal ontologies designed primarily for information retrieval which happen to also provide textual explanations of terms. One could say they have the same goals as Legal Taxonomy Syllabus in reverse priority. They are developed with a bottom-up approach, since they first and foremost need to include terms that appear in the documents to be retrieved.

OntoPrivacy (Cappelli et al. [2007]) is an ontology based on a glossary of keywords extracted from the Italian Personal Data Protection Code and includes legal, public administration and generic words. The terms were organised into a lightweight ontology containing textual definitions while also identifying synonymy, hyponymy and hypernymy relations. The trees were then linked to a top-level ontology, – inspired by LRI Core, the Core Legal Ontology and Legal-RDF Vocabularies – with five main classes: events, scenes (where and when an event occurs), physical objects, non-physical objects and roles (played by individuals or groups). The work is interesting in that it bridges the gap between legal glossaries and legal ontologies. It is also similar to our approach in that the data was collated by legal experts. It is not clear which legal formants apart from legislation were used to define the contents. They do not mention the European dimension which has had a significant impact on this area of law in Italy. It is also not clear if and how they deal with polysemy, terminological variation, evolution in the meaning of legal terms or the issue of alternative legal interpretation of legal terms.

JurWordNet (Gangemi et al. [2003]) is a legal ontology of Italian legal terms also developed primarily for information extraction, other stated goals are semantic tagging of legislative texts, question answering, automatic tagging, norm comparison and knowledge sharing. The ontology is constructed as an extension of ItalWordNet (Roventini et al. [2000]) and EuroWordNet (Vossen [1998]), which are in turns extensions of Wordnet (Miller [1995]). Wordnet is an extensive general glossary which has been used as an ontology in information extraction due to its extensive coverage. Concepts and terms are separated, with terms organised into sets of synonyms (known as synsets) mapping to different contextual concepts (known as senses). Gangemi et al (Gangemi et al. [2002]) have created a more ‘principled’ ontology of Wordnet, correcting inflated concepts such as is-a which is used for instances as well as specialised concepts (e.g. Beethoven and songwriter) or conflating types and roles (e.g. dog and worker animal). Jurwordnet is organised in accordance with these more precise relations. It also carefully connects legal concepts to layman concepts, since legal concepts are often specialisations of generic layman terms.

The LOIS ontology (Tiscornia [2006]) is a multilingual version of Jurwordnet (Gangemi et al. [2003]). It contains monolingual wordnets linked by equivalence, near-equivalence or subsumption relations to an English synset in the Inter-Lingual-Index (ILI). Synsets in other languages refer to glosses in the English ILI to get their meaning. The LOIS ontology is the most similar ontology to LTS in its multilinguality, domain, separation of concepts and terms and handling of polysemy. Both ontologies contain terms from European and national jurisdictions. The main difference is that LTS has a separate ontological structure for every jurisdiction, with terms in different languages mapping to jurisdiction-specific definitions, whereas the LOIS ontology has one conceptual structure to which terms in different languages are mapped. This is an important limitation inherited from EurWordnet (Vossen [1998]) which can impede accurate modelling of differences in conceptions across different jurisdictions. Another limitation imposed by Wordnet is that relations are lexical rather than semantic, so that relations such as ‘purpose’, ‘implemented by’ or ‘replaced by’ are not allowed. The LOIS ontology works around this limitation by connecting to the DOLCE2.1-LitePlus + CLO (Gangemi et al. [2005]) where it duplicates LOIS concepts and connects them with semantic relations. Some lexical relations in Wordnet are practically the same as DOLCE and CLO semantic relations (e.g. hyponym and sub-class-of) which introduces the
risk of potential misalignment. Of course, ontologies “are dependent on the purposes they are made for” (Chandrasekaran et al. [1998]), and the strengths of the LOIS ontology may outweigh its weakness for its main purpose of monolingual and cross-lingual information retrieval.

The main structural weakness of the LOIS ontology is being addressed in later work (albeit the purpose of the work is different). The Consumer Protection Ontology within the DALOS Project (Francesconi et al. [2007]), which builds on the LOIS ontology, acknowledges that “in a legal domain one cannot transfer the conceptual structure from one legal system to another” (ibid, page 104) and, like LTS, creates one conceptual layer for EU legal concepts which is mapped to by lexical layers in different languages. The intended aim of the ontology is to support multilingual EU legislative drafting. Unfortunately, the system described does not include a conceptual layer for national jurisdictions. It could be argued that the multi-jurisdictional framework of ontologies in LTS offers a more comprehensive system for human understanding of legal terms in European and national jurisdictions. Legal experts can explore differences in different jurisdictions, and translators can gain greater awareness of different conceptions at European and national levels to help them understand the impact of their translation choices.

A recent development to LTS is the prescriptions ontology (Boella et al. [2012]). As such, our survey of related work would not be complete without mentioning (Kralingen [1995], Visser [1995]) frame-based ontology, which provides data structures for representing norms and acts containing their constituent elements in a stereotyped situation. The similarity ends there, as the ontology is based on theoretical assumptions, and analyses procedures more than normative content. More recent work in this vein is (Venturi et al. [2009] et al). This corpus-based resource aims to represent complex events and situations expressed within legal documents using many different kinds of frames and frame elements. For example, the sentence “Il venditore deve consegnare al consumatore beni conformi al contratto di vendita” (“The seller must deliver goods to the consumer which are in conformity with the contract of sale”) instantiates the Frame ‘Being obligated’ evoked by the Lexical Unit deve (‘must’) as well as the Frame Elements ‘Duty’ (in this case, to deliver goods), ‘Beneficiary’ to describe the addressee of the duty (in this case, the consumer), and ‘Responsible party’, the person who must perform the Duty (in this case, the seller).

The Ontology of Professional Judicial Knowledge (Casanovas et al. [2007]), though exploring a different topic (the judicial view of legal procedures in Spain), has important similarities to LTS in its emphasis on the collaborative involvement of legal experts in populating the ontology. OPJK explores differences in interpretation (in this case, via a wiki) before reaching generalised definitions, which can be compared to the grouping in LTS. Discarded conceptualisations are archived, and reasons for their archival noted. LTS also has a mechanism for rendering conceptualisations obsolete, with the ‘replaced by mechanism described above.

**References for section 4.7**


Guido Boella, Llio Humphreys, Marco Martin, Piercarlo Rossi, and Leendert van der Torre. Eunomos, a legal document and knowledge management system to build legal services. In AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents -


Tom Engers, Alexander Boer, Joost Breuker, Andr Valente, and Radboud Winkels. Ontologies in the legal domain. In Hsinchun Chen, Lawrence Brandt, Valerie Gregg, Roland Traum ‘


Figure 9: The mapping of European legal terms to national legal terms in Italy and England&Wales

Figure 10: relating an expiring concept to a concept that replaces it
Figure 11: concepts representing interpretation

Figure 12: mapping terms in the European level ontology to terms in national ontologies
**Figure 13: the search interface of Eunomos**

![Search Interface](image)

### Search results (39 Entries found; page 1 of 4)

<table>
<thead>
<tr>
<th>Meaning ID</th>
<th>Legal language</th>
<th>Level</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>English</td>
<td>National</td>
<td>adequate</td>
<td>Notes: Case law relating</td>
</tr>
<tr>
<td>0</td>
<td>English</td>
<td>National</td>
<td>advertising</td>
<td>Office of Fair Trading v Office</td>
</tr>
<tr>
<td>8</td>
<td>English</td>
<td>National</td>
<td>a readily comprehensible form</td>
<td>Bankers Insurance Co Ltd v Sou</td>
</tr>
<tr>
<td>1</td>
<td>English</td>
<td>National</td>
<td>cannot be excluded by any contractual term</td>
<td>SSSCR 2002; Since domestic law</td>
</tr>
<tr>
<td>0</td>
<td>English</td>
<td>National</td>
<td>charge</td>
<td>London North Securities Ltd v</td>
</tr>
<tr>
<td>7</td>
<td>English</td>
<td>National</td>
<td>circumstances</td>
<td>Director General of Fair Trade</td>
</tr>
<tr>
<td>8</td>
<td>English</td>
<td>National</td>
<td>clear and comprehensible manner</td>
<td>Bankers Insurance Co Ltd v Sou</td>
</tr>
<tr>
<td>8</td>
<td>English</td>
<td>National</td>
<td>clean and legible and in uniform lettering</td>
<td>Bankers Insurance Co Ltd v Sou</td>
</tr>
<tr>
<td>6</td>
<td>English</td>
<td>National</td>
<td>clear and unequivocal way</td>
<td>Bankers Insurance Co Ltd v Sou</td>
</tr>
<tr>
<td>8</td>
<td>English</td>
<td>National</td>
<td>clearly</td>
<td>Bankers Insurance Co Ltd v Sou</td>
</tr>
</tbody>
</table>

[Image of the search interface]
5 Identification of user needs for cross-border legal information services

In this section we first describe the questions of the online survey we submitted to possible users of the project outcomes, the results of the 95 answers received and the conclusions we can draw from this analysis.

5.1 User questionnaire / online survey

We devised an online questionnaire, which is available at this address.

The 18 questions are the following ones

1. Professional area
   - Lawyer / Solicitor / Barrister
   - In-house legal council
   - Judge
   - Lecturer in Law Sciences / Scholar
   - Fiscal expert / Accountant
   - Compliance manager / Auditor
   - Other

2. Type of information needed
   - Legislation
   - Case law
   - Doctrinal writings
   - Other

3. Percentage of monthly time in searching legal information
   - Less than 25%
   - 25%-50%
   - 50%-75%
   - More than 75%

4. Sources to search for information
   - Specialized magazines
   - Corpora / Body of Laws

128 The complete data are available in Appendix I
• Legal databases
• Online specialized magazines
• Online legal databases
• Information portal

5. Tools used to search legal information
• Commercial legal databases
• Free national legal portals
• Online legal journals
• National institutional portals
• Specialized magazines

6. Use of N-Lex portal
• Yes
• No

7. Frequency of use of the tools specified in #5 to find legal information from other EU countries
• One-two times per year
• Several times per year
• One-two times per month
• Several times per month
• Other

8. Adequacy of these tools with respect to user needs
• Yes
• No

8.1. If no - why?
• Difficult for orientation
• Not complete enough
• Contents in mother tongue only

9. EU countries from which legal information is used most often
• United Kingdom
• France
• Germany
• Italy
• Austria
10. Is such legal information from other EU countries always connected with the application of EU law?
   • Yes, until now it has always been connected with the application of EU law
   • No, in some cases it was not connected with the application of EU law
   • No, until now it was never connected with the application of EU law

11. Search frequency of case law in such situations
   • More than 75%
   • 50%-75%
   • 25%-50%
   • Less than 25%
   • Other (please specify):

12. Is the EU multilingual legal System an issue to you?
   • Yes
   • No

13. Do you use such legal information in case it is available in a foreign language only?
   • No, I use it only if it is officially translated into my mother tongue
   • Yes, I use it by translating its text with Google Translate or other machine translation tools
   • Yes, but I use it only if I have a good command of the language in which the legal document is drafted

14. Use of EUR-Lex Portal?
   • Yes
   • No

14.1. Reasons for not using the EUR-Lex portal
   • No need
   • Ignorance of the existence of such a portal
   • Use of other information resources

15. Areas of law, other than consumer protection law, in which legal information from other EU countries is mostly needed
   • Private international law
• Protection of human rights and fundamental freedoms
• Competition law
• Tax law
• Judicial cooperation in civil and commercial matters
• Judicial cooperation in criminal matters
• Other areas

16. Which of the following functionalities would you recommend to be integrated within the EUCases services, applications and tools?

• 16a - Simultaneous search in EU and national legislation and case law by using the terms of the multilingual thesaurus Eurovoc
• 16b - Simultaneous search in EU and national legislation and case law by using automated translation of the search query in the language of the documents
• 16c - Browsing of EU and national legislation and case law by using hierarchical classifications by subject matter or legal terms
• 16d - Links from EU directives to the full texts of the so called “National execution measures” and vice versa
• 16e - Links from national legal instruments and case law to the cited full texts and concrete provisions of the EU legislative and judicial acts
• 16f - For each cited EU provision or numbered paragraph in a decision of the EU Court of Justice – links to all referring national legislative or judicial acts

17. Would you like to test for free the EUCases web services, applications and tools when they are ready and to share with us your opinion about their contents and functionalities?

• Yes, I do
• No, thank you

5.2 Analysis of results from the survey

The survey was mainly of qualitative, not quantitative nature. Therefore, it was conducted among legal professionals who have good command of languages spoken in other EU Member States, i.e., among potential customers of EUCases services.

The questionnaire, sent to about 300 candidates, lead to 95 answers from 5 countries, distributed among different kinds of users: 26% lawyers/barristers/solicitors, 32% in-house lawyers, 8% law scholars, 6% compliance managers, 5% from finance and accounting, 4% judges.

It was difficult to convince judges to fill-in the questionnaire. Less than 5% of the respondents are judges despite the significant efforts made by the project partners. There are two consequences for the project when considering this fact:

• Partners should try further to approach judges during the project execution and to ask for their opinion, because they are one of the main targeted customer groups;
• The commercial partners should count that it will be difficult to present the future commercial services directly to judges. Such presentations could be organised only with the support of the court administrations or on IT (e-Justice) fora with the participation of judges.
The respondents say that for their work they need access to legislation (93%). Nearly 90% of all respondents have quoted case law as part of their everyday work which means that practically almost all legal professionals rely on it as a source of law and/or source of legal knowledge and inspiration.

Doctrinal writings are clearly less important legal information resource if compared with legislation and case law: two thirds of respondents say that they need access also to doctrinal writing, even if only 8% are law scholars. There are, however, other categories of information needed as well, some potentially interesting for us such as administrative regulations and draft of legislation.

Looking for legal information is a very time consuming task: 60% of respondents spend between 25% and 75% of their time in researching legislation; 9% even more than 75% of their time.

This means that the need for cross-border legal information services is relatively high. 25% of the respondents use legal information from other EU countries several times per month and 3% use it daily. This fact combined with the observation that 75% of the respondents use online legal databases and 67% of them use commercial legal databases makes us more optimistic about the market potential of the future EUCases online services.

However, traditional sources are still often used (by 50%). Other sources like magazines and information portals are also important. Commercial databases win the majority of users: 67% of respondents use them. Free national portals and institutional ones are used by less than 30% of respondents. Online legal journals have still to become popular (13%).

Surprisingly, only 14% of the respondents use the N-Lex portal to search information about legislation of other EU countries. This could be interpreted as meaning that they use other public or commercial legal information resources. EUR-LEX is more used but still not so extensively (65%), where the remaining users don’t have the need to consult it, don’t know it or have other sources of information. This means that they are using commercial or non-commercial legal databases as a substitution service for receiving information about EU law, i.e. there is another good reason to believe that there is a niche for the EUCases service.

About 1/3 of the respondents say that the tools used for searching cross-border legal information are not adequate to user needs. More than half of them reckon that the reason for this is the fact that these legal information services are not complete enough (56%), whereas 1/3 of them think that they are difficult to use (34%) and 6 % experienced difficulties with foreign languages. These two findings should be taken into account by the development of the EUCases service.

Quite interesting are the results from the question for which other (i.e. foreign) EU countries respondents search legal information most often. As expected, on the top are the biggest jurisdictions – UK with 35%, Germany – 34%, France – 33%. Surprisingly, 19% of the respondents search for legal information of Austria, which is even more than the search for Italian legal resources (13%). However, the reason is not that Austrian databases are used by users from Germany as it might be thought. Only one of the German respondents has checked Austria in his answer. Actually, the main interest comes from Italian legal professionals, perhaps connected with the common border and the minority German-speaking community in Italian South Tyrol. Around 7% of the respondents have interest for Belgium legal resources, which should be taken into account in case of possible future extensions of the content scope of EUCases service.

Surprisingly and most importantly for this project, the information from foreign countries is not only sought when discussing matters related to the application of EU Law. For 54% of respondents say that they use EU law even in matters not related to its application. Only 13% say that the cross-border legal information that was searched by them has never been connected with the application of EU law. At the same time 34% admit that such information has always been connected with the application of EU law. Combined with the fact that 24%
consider that in more than half of the cases the searched legal information was case law of the foreign country, this is another important acknowledgement of the vitality of the EUCases idea.

Another very interesting conclusion which could be drawn from the answers of Q13 (Do you use cross-border legal information in case it is available in a foreign language only?) and Q16b (suggestion to integrate within the EUCases service a functionality allowing simultaneous search in EU and national legislation and case law by using automated translation of the search query) is that only a few respondents are willing to rely on machine translation tools when the automatic translation concerns legal texts.

In other words, more than expected people started using automatic translation tools (19%), while only a minority relies on official translations. It will be interesting to know the opinion of the users on this issue at the end of the project after functionality tests of the multilingual access module to be developed by IICT-BAS.

EUCases partners should pay special attention to the results from the question in which areas of EU law (other than consumer protection law) users mostly need cross-border legal information. The high interest in private international law (almost 50%), judicial cooperation in civil matters (45%), tax law (41%) and competition law (33%) are not surprise. A little unexpected is the high percentage of the respondents (33%) who search for cross-border legal information in the area of protection of human rights and fundamental freedoms. This a good reason for EUCases partners to discuss on their next consortium meeting whether it will be viable to incorporate the case law of the European Court of Human Rights in Strasbourg into the EUCases service. This case law has been for long years incorporated in the EU legal doctrine and the jurisprudence of the Court of Justice.

A high percentage of the respondents are most interested in (in order of interest):

- 16e - Links from national legal instruments and case law to the cited full texts and concrete provisions of the EU legislative and judicial acts
- 16d - Links from EU directives to the full texts of the so called “National execution measures” and vice versa
- 16f - For each cited EU provision or numbered paragraph in a decision of the EU Court of Justice – links to all referring national legislative or judicial acts
- 16a - Simultaneous search in EU and national legislation and case law by using the terms of the multilingual thesaurus Eurovoc

5.3 Conclusions

In conclusion, the questionnaire says us some important information for the project. First of all that there is a demand for accessing legal information at an EU wide level, even that there a strong demand on legal information from other EU countries as a whole and not only on such information which is closely related to the application of EU law. Second, online resources are the predominant way of getting information, where the commercial sector gets the biggest share, while some free online resources like EUR-Lex are disregarded or not known at all. The latter fact highlights that there is probably a market niche for the EUCases future developments by providing legal added value services in between EU portals and commercial databases not specialised for accessing law at an European wide dimension.

The limited use of automatic translation tools is a challenge for the project, so that we should consider the possibility to explore which is the reaction of the users to the use of machine translation in the project.
# APPENDIX I: Answers to the survey

## 1. Professional area

<table>
<thead>
<tr>
<th>Answers</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lawyer / Solicitor / Barrister</td>
<td>25</td>
<td>26.32%</td>
</tr>
<tr>
<td>In-house legal council</td>
<td>30</td>
<td>31.58%</td>
</tr>
<tr>
<td>Judge</td>
<td>4</td>
<td>4.21%</td>
</tr>
<tr>
<td>Lecturer in Law Sciences / Scholar</td>
<td>8</td>
<td>8.42%</td>
</tr>
<tr>
<td>Fiscal expert / Accountant</td>
<td>5</td>
<td>5.26%</td>
</tr>
<tr>
<td>Compliance manager / Auditor</td>
<td>6</td>
<td>6.32%</td>
</tr>
<tr>
<td>Other</td>
<td>17</td>
<td>17.89%</td>
</tr>
</tbody>
</table>

## 2. Type of information needed

<table>
<thead>
<tr>
<th>Answers</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legislation</td>
<td>88</td>
<td>92.63%</td>
</tr>
<tr>
<td>Case law</td>
<td>86</td>
<td>90.53%</td>
</tr>
<tr>
<td>Doctrinal writings</td>
<td>60</td>
<td>63.16%</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>13.68%</td>
</tr>
</tbody>
</table>

## 3. Percentage of monthly time in searching legal information

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 25%</td>
<td>29</td>
<td>30.53%</td>
</tr>
<tr>
<td>25%-50%</td>
<td>35</td>
<td>36.84%</td>
</tr>
<tr>
<td>50%-75%</td>
<td>22</td>
<td>23.16%</td>
</tr>
<tr>
<td>More than 75%</td>
<td>9</td>
<td>9.47%</td>
</tr>
</tbody>
</table>

## 4. Sources to search for information

<table>
<thead>
<tr>
<th>Sources</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialized magazines</td>
<td>40</td>
<td>42.11%</td>
</tr>
<tr>
<td>Corpora / Body of Laws</td>
<td>50</td>
<td>52.63%</td>
</tr>
<tr>
<td>Legal databases</td>
<td>55</td>
<td>57.89%</td>
</tr>
<tr>
<td>Online specialized magazines</td>
<td>30</td>
<td>31.58%</td>
</tr>
<tr>
<td>Online legal databases</td>
<td>71</td>
<td>74.74%</td>
</tr>
<tr>
<td>Information portal</td>
<td>63</td>
<td>66.32%</td>
</tr>
</tbody>
</table>

## 5. Tools used to search legal information

<table>
<thead>
<tr>
<th>Tools used to search legal information</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial legal databases</td>
<td>64</td>
<td>67.37%</td>
</tr>
<tr>
<td>Free national legal portals</td>
<td>30</td>
<td>31.58%</td>
</tr>
</tbody>
</table>
### 6. Use of N-Lex portal

Yes 13 13,68%
No 82 86,32%

### 7. Frequency of use of the tools specified in #5 to find legal information from other EU countries

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-two times per year</td>
<td>18</td>
<td>18,95%</td>
</tr>
<tr>
<td>Several times per year</td>
<td>26</td>
<td>27,37%</td>
</tr>
<tr>
<td>One-two times per month</td>
<td>17</td>
<td>17,89%</td>
</tr>
<tr>
<td>Several times per month</td>
<td>24</td>
<td>25,26%</td>
</tr>
<tr>
<td>Daily</td>
<td>3</td>
<td>3,16%</td>
</tr>
<tr>
<td>Never</td>
<td>6</td>
<td>6,32%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1,05%</td>
</tr>
</tbody>
</table>

### 8. Adequacy of these tools with respect to user needs

Yes 63 66,32%
No 32 33,68%

#### 8.1. If no - why?

<table>
<thead>
<tr>
<th>Reason</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficult for orientation</td>
<td>11</td>
<td>34,38%</td>
</tr>
<tr>
<td>Not complete enough</td>
<td>18</td>
<td>56,25%</td>
</tr>
<tr>
<td>Contents in mother tongue only</td>
<td>2</td>
<td>6,25%</td>
</tr>
<tr>
<td>No answer</td>
<td>1</td>
<td>3,13%</td>
</tr>
</tbody>
</table>

### 9. EU countries from which legal information is used most often

<table>
<thead>
<tr>
<th>Country</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>33</td>
<td>34,74%</td>
</tr>
<tr>
<td>France</td>
<td>31</td>
<td>32,63%</td>
</tr>
<tr>
<td>Germany</td>
<td>32</td>
<td>33,68%</td>
</tr>
<tr>
<td>Italy</td>
<td>12</td>
<td>12,63%</td>
</tr>
<tr>
<td>Austria</td>
<td>18</td>
<td>18,95%</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>2</td>
<td>2,11%</td>
</tr>
<tr>
<td>Other</td>
<td>21</td>
<td>22,11%</td>
</tr>
<tr>
<td>Same country</td>
<td>71</td>
<td>74,74%</td>
</tr>
</tbody>
</table>
10. Is such legal information from other EU countries always connected with the application of EU law?

Yes, until now it has always been connected with the application of EU law 32 33,68%
No, in some cases it was not connected with the application of EU law 51 53,68%
No, until now it was never connected with the application of EU law 12 12,63%

11. Search frequency of case law in such situations

More than 75% 12 12,63%
50%-75% 11 11,58%
25%-50% 24 25,26%
Less than 25% 38 40,00%
Never 5 5,26%
Other (please specify): 3 3,16%
Other (see Notes): 2 2,11%

12. Is the EU multilingual legal System an issue to you?

Yes 47 49,47%
No 48 50,53%

13. Do you use such legal information in case it is available in a foreign language only?

No, I use it only if it is officially translated into my mother tongue 17 17,89%
Yes, I use it by translating its text with Google Translate or other machine translation tools 18 18,95%
Yes, but I use it only if I have a good command of the language in which the legal document is drafted 60 63,16%

14. Use of EUR-Lex Portal?

Yes 62 65,26%
No 33 34,74%

14.1. Reasons for not using the EUR-Lex portal

No need 12 36,36%
Ignorance of the existence of such a portal 9 27,27%
Use of other information resources 5 15,15%
15. Areas of law, other than consumer protection law, in which legal information from other EU countries is mostly needed

<table>
<thead>
<tr>
<th>Area</th>
<th>Number of Respondents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private international law</td>
<td>47</td>
<td>49.47%</td>
</tr>
<tr>
<td>Protection of human rights and fundamental freedoms</td>
<td>31</td>
<td>32.63%</td>
</tr>
<tr>
<td>Competition law</td>
<td>32</td>
<td>33.68%</td>
</tr>
<tr>
<td>Tax law</td>
<td>39</td>
<td>41.05%</td>
</tr>
<tr>
<td>Judicial cooperation in civil and commercial matters</td>
<td>43</td>
<td>45.26%</td>
</tr>
<tr>
<td>Judicial cooperation in criminal matters</td>
<td>18</td>
<td>18.95%</td>
</tr>
<tr>
<td>Other areas</td>
<td>20</td>
<td>21.05%</td>
</tr>
</tbody>
</table>

16. Which of the following functionalities would you recommend to be integrated within the EUCases services, applications and tools?

16a - Simultaneous search in EU and national legislation and case law by using the terms of the multilingual thesaurus Eurovoc

<table>
<thead>
<tr>
<th>Option</th>
<th>Number of Respondents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>41</td>
<td>43.16%</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>20.00%</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>12.63%</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>13.68%</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4.21%</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4.21%</td>
</tr>
</tbody>
</table>

16b - Simultaneous search in EU and national legislation and case law by using automated translation of the search query in the language of the documents

<table>
<thead>
<tr>
<th>Option</th>
<th>Number of Respondents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>23</td>
<td>24.21%</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>13.68%</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>22.11%</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>16.84%</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>10.53%</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>11.58%</td>
</tr>
</tbody>
</table>

16c - Browsing of EU and national legislation and case law by using hierarchical classifications by subject matter or legal terms

<table>
<thead>
<tr>
<th>Option</th>
<th>Number of Respondents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>33</td>
<td>34.74%</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>22.11%</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>28.42%</td>
</tr>
</tbody>
</table>
16d - Links from EU directives to the full texts of the so called “National execution measures” and vice versa

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>46</td>
<td>48.42%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>26.32%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>11.58%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>9.47%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3.16%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.00%</td>
<td></td>
</tr>
</tbody>
</table>

16e - Links from national legal instruments and case law to the cited full texts and concrete provisions of the EU legislative and judicial acts

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>53</td>
<td>55.79%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>21.05%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>11.58%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4.21%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4.21%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.05%</td>
<td></td>
</tr>
</tbody>
</table>

16f - For each cited EU provision or numbered paragraph in a decision of the EU Court of Justice – links to all referring national legislative or judicial acts

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>48</td>
<td>50.53%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>18.95%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>15.79%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7.37%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.05%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3.16%</td>
<td></td>
</tr>
</tbody>
</table>

17. Would you like to test for free the EUCases web services, applications and tools when they are ready and to share with us your opinion about their contents and functionalities?

Yes, I do 79 83.16%
No, thank you 16 16.84%

18. If you would like, you can write down here some other advice or recommendations to the partners in the EUCases project

More perfection in the translation of documents 1
Quite ambitious project! Good luck!

I will be very glad if the project will be successful and will ensure us new opportunities for comparative analysis of legislation and case law

Refine the legal translation in Patent and Trademark attorney, design and copyright.

It’s very important to develop a reseaching tool in European Patent Case Law (EN-DE-FR) to avoid competitiveness imbalance.

System to find the consolited text of EU Regulations and Directives

Hi. I am an expert on open data; a qualified computer scientist and an experienced open data lawyer (both in IP and in more general law). I strongly suggest that anyone trying to do open data law spends some time talking to me.

For example: some efforts to create a schema for legislation markup have assumed (wrongly) that numbering in legislation is purely hierarchical. This is not true and real legislation will break some well-known XML schemes that have been designed with the best of intentions but without talking to people (like me) with experience.

Eg. I spent some time - quite a few years ago now - writing some code to process UK legislation into a usable form before making it was available (as it is now) in a better XML format. I developed my own object model and realised that numbering was not hierarchical by experience with the data set.

Please drop me a line to talk.
7 APPENDIX II: Legal XML - example regulation

<?XML version="1.0" encoding="iso-8859-1" standalone="no"?>
<!--Akoma_Ntoso document-->
<akomaNtoso xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.akomantoso.org/2.0 ./akomantoso20.xsd"
xmlns="http://www.akomantoso.org/2.0">
<!--Illustration of a chapter of the French Labour Code in XML. Analysis and markup performed by Robert Muthuri, LAST-JD - CIRSFID, University of Bologna -->

<act contains="originalVersion">
<meta>
  <identification source="#somebody">
    <FRBRWork>
      <FRBRthis value="/fr/act/2008-03-01/nn/main"/>
      <FRBRuri value="/fr/act/2008-03-01/nn"/>
      <FRBRdate date="2008-03-01" name="generation"/>
      <FRBRauthor href="#parliament" as="#author"/>
      <componentInfo>
        <componentData id="wmain" href="#emain" name="main" showAs="Main document"/>
      </componentInfo>
    </FRBRWork>
    <FRBRcountry value="fr"/>
  </FRBRWork>

<FRBRExpression>
  <FRBRthis value="/fr/act/2008-03-01/nn/fra@/main"/>
  <FRBRuri value="/fr/act/2008-03-01/nn/fra@"/>
  <FRBRdate date="2013-11-15" name="generation"/>
  <FRBRauthor href="#somebody" as="#editor"/>
  <componentInfo>
    <componentData id="emain" href="#mmain" name="main" showAs="Main document"/>
    <componentData id="eschedule" href="#mschedule" name="schedule" showAs="SCHEDULE"/>
  </componentInfo>
  <FRBRlanguage language="eng"/>
</FRBRExpression>
<FRBRManifestation>
  <FRBRthis value="/fr/act/2008-03-01/nn/fra@/main.XML"/>
  <FRBRuri value="/fr/act/2008-03-01/nn/fra@.akn"/>
  <FRBRdate date="2013-11-15" name="generation"/>
  <FRBRauthor href="#somebody" as="#editor"/>
  <componentInfo>
    <componentData id="mmain" href="main.XML" name="main" showAs="Main document"/>
    <componentData id="mschedule" href="schedule.XML" name="schedule" showAs="SCHEDULE"/>
  </componentInfo>
</FRBRManifestation>

<!-- shows the publication details -->
<publication name="officialGazette" date="2008-03-01" showAs="French Gazette" number="Supplement No. xx"/>

<lifecycle source="#somebody">
  <eventRef id="e1" date="2008-03-01" source="#ro1" type="generation"/>
</lifecycle>

<analysis source="#somebody">
  <activeModifications>
    <!-- details the document lifecycle - various amendments/modifications that a document has undergone while retaining their fundamental nature -->
  </activeModifications>
</analysis>

<!-- defining local/external references (docs, concepts, people and places) -->
<references source="#somebody">
  <original id="ro1" href="/fr/act/2008-03-01/nn/fra@/main" showAs="Original"/>
  <TLCPerson id="somebody" href="/ontology/person/editors/somebody" showAs="somebody"/>
</references>
Chapitre Ier
- Objet et constitution.

Les syndicats professionnels ont exclusivement pour objet l'étude et la défense des droits ainsi que des intérêts matériels et moraux, tant collectifs qu'individuels, des personnes mentionnées dans leurs statuts.

Les syndicats ou associations professionnels de personnes exerçant la même profession, des métiers similaires ou des métiers connexes concourant à l'établissement de produits déterminés ou la même profession libérale peuvent se constituer librement.
<p>Par dérogation à ces dispositions, les particuliers occupant des employés de maison peuvent se grouper en syndicat pour la défense des intérêts qu'ils ont en commun en tant qu'employeur de ces salariés.</p>

<p>Les fondateurs de tout syndicat professionnel déposent les statuts et les noms de ceux qui, à un titre quelconque, sont chargés de l'administration ou de la direction.</p>

<p>Ce dépôt est renouvelé en cas de changement de la direction ou des statuts.</p>

<p>Tout adhérent d'un syndicat professionnel peut, s'il remplit les conditions fixées par l'article, accéder aux fonctions d'administration ou de direction de ce syndicat.</p>
Tout membre français d’un syndicat professionnel chargé de l’administration ou de la direction de ce syndicat doit jouir de ses droits civiques et n’être l’objet d’aucune interdiction, déchéance ou incapacité relative à ses droits civiques. 

Sous les mêmes conditions, tout ressortissant étranger âgé de dix-huit ans accomplis adhérent à un syndicat peut accéder aux fonctions d’administration ou de direction de ce syndicat.

En cas de dissolution volontaire, statutaire ou prononcée par décision de justice, les biens du syndicat sont dévolus conformément aux statuts ou, à défaut de dispositions statutaires, suivant les règles déterminées par l’assemblée générale.

En aucun cas les biens du syndicat ne peuvent être répartis entre les membres adhérents.
<p>Coded by: <ref id="ref 8" href="ef/gazetteList/2007-03-12/329/fr@/main"/></p>

<p>Ancient texts: <ref id="ref9" href="fr/act/2008-03-01/nn/main#artL411-9AbD"/> <ref id="ref10" href="fr/act/2008-03-01/nn/main#artL411-9M"/></p>
APPENDIX III: Legal XML - example judgement

<?xml version="1.0" encoding="iso-8859-1"?>

<akomaNtoso xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.akomantoso.org/2.0 ./akomantoso20.xsd"
xmlns="http://www.akomantoso.org/2.0">

<!-- Analysis and markup performed by Robert Muthuri, LAST-JD - CIRSFID, University of Bologna -->

<judgement>
  <meta>
    <identification source="#somebody">
      <FRBRWork>
        <FRBRthis value="/fr/judgement/2010-09-13/09-14418/main"/>
        <FRBRUri value="/fr/judgement/2010-09-13/09-14418"/>
        <FRBRdate date="2010-09-13" name="Hearing"/>
        <FRBRAuthor href="#Scoz" as="#Author"/>
        <FRBRCountry value="za"/>
        <FRBRNumber value="14-2007"/>
      </FRBRWork>
      <FRBRExpression>
        <FRBRthis value="/fr/judgement/2010-09-13/09-14418/fra@/main"/>
        <FRBRUri value="/fr/judgement/2010-09-13/09-14418/fra@"/>
        <FRBRdate date="2010-09-13" name="Delivery"/>
        <FRBRAuthor href="#somebody" as="#Editor"/>
        <FRBRLanguage language="eng"/>
      </FRBRExpression>
      <FRBRMannifestation>
        <FRBRthis value="/fr/judgement/2010-09-13/09-14418/fra@/main.XML"/>
        <FRBRUri value="/fr/judgement/2010-09-13/09-14418/fra@/main.akn"/>
        <FRBRdate date="2011-01-20" name="XMLConversion"/>
        <FRBRAuthor href="#somebody" as="#Editor"/>
        <FRBRFormat value="XML"/>
      </FRBRMannifestation>
    </identification>
  </meta>
</judgement>
<identification>

<pub><publication date="2010-09-13/09-14418" name="Law Report" showAs="somebodyLand Law Journal" number="555"/></pub>

<lifecycle source="#somebody">
  <eventRef date="2010-09-13/09-14418" id="e1" source="#ro1" type="generation"/>
</lifecycle>

<workflow source="#somebody">
  <step date="2010-03-26" id="a1" outcome="#hearingDate" actor="Scoz"/>
  <step date="2010-09-13/09-14418" id="a2"/>
</workflow>

<analysis source="#somebody">
  <judicial>
    <!-- shows the modification cycle -->
  </judicial>
</analysis>

<!-- defining local/external references (docs, concepts, people and places) -->

<references source="#somebody">
  <original id="ro1" href="/ak/judgement/2007-07-25/14-2007/eng@/main" showAs="Original"/>
  <TLCOrganization id="somebody" href="/ontology/organization/kn/somebody" showAs="somebody"/>
  <TLCOrganization id="Scoz" href="/ontology/organization/fr.CourDeCassation" showAs="COUR DE CASSATION"/>
  <TLCOrganization id="ag" href="/ontology/organization/fr.AvocatGeneral" showAs="AVOCAT GENERAL"/>
  <TLCOrganization id="avocat" href="/ontology/organization/fr.SCPGatineauFattaccini" showAs="SCP GATINEAU FATTACCINI"/>
  <TLCPerson id="mjacoupy" href="/ontology/person/fr.mejacoupy" showAs="ME JACOUPY"/>
  <TLCPerson id="rap" href="/ontology/person/fr.rapportuer" showAs="M. Gosselin"/>
  <TLCPerson id="président" href="/ontology/person/judges/fr.president" showAs="MME COLLOMP"/>
</references>

</identification>
Some of the content of the motivations was deleted in order to simplify the reading of the example. 

<notes source="#somebody">
  <note id="not1">
    <p>Some of the content of the motivations was deleted in order to simplify the reading of the example. </p>
  </note>
</notes>

<header>
  <p class="judgementNumber">Numéro d'arrêt : <docketNumber>09-14418</docketNumber></p>
  <span class="preface">Numéro d'arrêt : <docketNumber>09-14418</docketNumber></span>
</header>

<!-- Important the neutral citation: methodology for make a citation of law case-->
<neutralCitation>Cass. Soc., 13 octobre 2010, pourvoi n°09-14418, Bull. civ. 2010, V, n° 223</neutralCitation>

<p class="caseNumber">Numéro de décision: <docNumber id="judgmentNumber">51001921</docNumber></p>
France, Cour de cassation  
Chambre sociale

TRAVAIL REGLEMENTATION - DUREE DU TRAVAIL - Repos et congés - Repos hebdomadaire - Repos dominical - Dérogations - Accord intervenu entre les organisations syndicales de salariés et les organisations d'employeurs - Arrêté préfectoral de fermeture au public - Etablissements visés - Détermination - Portée.

TRAVAIL REGLEMENTATION - DUREE DU TRAVAIL - Repos et congés - Repos hebdomadaire - Repos dominical - Dérogations - Accord intervenu entre les organisations syndicales de salariés et les organisations d'employeurs - Validité - Accord signé par un organisme habilité à le signer - Portée

Si le juge civil n'a pas à statuer sur la légalité d'un arrêté préfectoral pris en application de l'article réf. du code du travail, il lui appartient, lorsqu'il est saisi d'une demande de sursis à statuer dans l'attente de la décision de la juridiction administrative, d'apprécier si la contestation de cet arrêté est sérieuse. Doit ainsi être approuvé l'arrêt qui estime que la contestation d'un arrêté préfectoral n'est pas sérieuse en relevant que l'organisme professionnel était habilité à signer l'accord exigé par l'article réf. du code du travail et que l'arrêté préfectoral peut viser des parties d'établissement, dès lors qu'il concerne, en termes généraux, les terminaux de cuisson et les dépôts de pain, qu'ils se situent dans des magasins spécialisés ou dans des magasins à succursales multiples.

Sur le n° 2 : Sur l'application, à des parties d'établissement, d'un arrêté préfectoral de fermeture pour respect du repos hebdomadaire, énoncé en terme généraux, dans le même sens que : Soc., 25 mars 1997, pourvoi n° 95-15248 Bull. 1997, V, n° 124 (rejet). Sur la nécessité pour une cour d'appel saisie d'une demande de sursis à statuer dans l'attente de la décision de la juridiction administrative d'un arrêté préfectoral, d'apprécier si la contestation de cet arrêté est sérieuse, à rapprocher : Soc., 25 octobre 1990,
Attendu que l'arrêt attaqué (Rouen, 12 mars 2009), que par arrêté du 9 octobre 1996, pris sur le fondement de l'article de l'article [ref id="ref4" href="fr/act/2008-03-01/nn/main#artL3132-29"></ref> du code du travail, le préfet de l'Eure, considérant que l'accord intervenu entre plusieurs organisations d'employeurs et de salariés le 7 octobre 1996, exprimait la volonté de la majorité des professionnels concernés, à titre principal ou accessoire, par la fabrication, la vente ou la distribution de pain, a décidé que dans les communes du département de l'Eure, les établissements ou parties d'établissements sédentaires ou ambulants employant ou non des salariés, et dans lesquels s'effectuent ces activités, y compris les terminaux de cuisson quelle que soit leur appellation, seront fermés au public un jour par semaine au choix des intéressés ; que la Maison de la boulangerie pâtisserie artisanale de l'Eure a fait attaquer la société Atac devant le tribunal de commerce afin qu'il ordonne à cette société de se conformer aux prescriptions de l'arrêté préfectoral sous astreinte de 1 500 euros par infraction constatée ;</p>

Sur le premier moyen : <eol/>

Attendu que la société fait grief à l'arrêt d'avoir dit que la Maison de la boulangerie pâtisserie artisanale de l'Eure avait la capacité d'ester en justice alors, selon le moyen, que lorsqu'un syndicat professionnel fait le choix de se constituer sous la forme d'une association, il ne dispose de la capacité d'ester en justice qu'en réunissant les conditions d'acquisition de la personnalité morale propres à ce type de groupement ; qu'une association ne dispose de la capacité d'ester en justice qu'à compter de sa déclaration en préfecture ou en sous-préfecture ; qu'en l'espèce, en énonçant que La Maison de la boulangerie pâtisserie artisanale de l'Eure était dotée de la capacité d'ester en justice dès lors qu'elle constituait un syndicat professionnel et qu'elle avait déposé ses statuts en mairie, peu important qu'elle ait pris la forme d'une association, sans constater que la demanderesse s'était déclarée en préfecture ou en sous-préfecture, la cour d'appel a violé les articles [ref id="ref5" href="fr/act/1901-07-1/main#art5"></ref> 5 et [ref id="ref6" href="fr/act/1901-07-1/main#art6"></ref> de la loi du 1er juillet 1901, ensemble l'article [ref id="ref7" href="fr/act/1804/main#art32"></ref> du code de procédure civile ;</eol/>

Mais attendu que les syndicats ou associations professionnels qui regroupent des personnes exerçant la même profession, des métiers similaires ou connexes, qui ont pour objet exclusif l'étude et la défense des droits ainsi que des intérêts matériels et moraux tant collectifs qu'individuels des personnes mentionnées dans leurs statuts, ont la capacité d'ester en justice, dès lors qu'ils ont satisfait à l'obligation de dépôt de leurs statuts en mairie ;

Et attendu qu'ayant vérifié que l'objet de la Maison de la boulangerie-pâtisserie artisanale de l'Eure, qui regroupe des professionnels exerçant un même métier, était celui d'un syndicat ou association professionnels, et qu'elle justifiait du dépôt en mairie de ses statuts, la cour d'appel a exactement décidé qu'elle avait, de ce fait, la capacité d'ester en justice ; que le moyen n'est pas fondé ;</eol/>

Sur le second moyen : <eol/>

Attendu que la société fait encore grief à l'arrêt de l'avoir condamnée à se conformer aux prescriptions de l'arrêté du Préfet de l'Eure en date du 9 octobre 1996 relatives à la fermeture hebdomadaire de 24 heures consécutives des boulangeries, boulangeries pâtisseries, dépôts de pain, ce, dans un délai d'un mois à compter de la
signification de l’arrêt attaqué et sous astreinte de 1 000 euros par infraction constatée alors, selon le moyen, que : <eol/>

1°/ l’accord entre les syndicats d’employeurs et de travailleurs sur les conditions dans lesquelles le repos hebdomadaire est donné aux salariés, préalable à l’arrêté par lequel le préfet peut ordonner la fermeture au public des établissements de la profession pendant toute la durée de ce repos, doit correspondre à la volonté de la majorité indiscutable de tous ceux qui, dans le département, exercent cette profession à titre principal ou accessoire et dont l’établissement est susceptible d’être fermé ; qu’en l’espèce, en affirmant que la légalité de l’arrêté préfectoral du 9 octobre 1996 n’était pas sérieusement contestable, sans rechercher, comme elle y était invitée, si l’accord sur la base duquel cet arrêté avait été pris reflétait la volonté de la majorité indiscutable de tous ceux qui, dans le département, assuraient la fabrication, la distribution ou la livraison du pain, la cour d’appel a privé sa décision de base légale au regard de l’article <ref id="ref8" href="fr/act/2008-03-01/nn/main#art221-17"></ref> devenu <ref id="ref9" href="fr/act/2008-03-01/nn/main#art3132-29"></ref> du code du travail ; <eol/>

2°/ l’arrêté par lequel le préfet peut ordonner la fermeture au public d’établissements de la profession pendant toute la durée du repos hebdomadaire ne peut viser une partie d’établissement ; qu’en affirmant en l’espèce que la légalité de l’arrêté préfectoral du 9 octobre 1996 n’était pas sérieusement contestable, dès lors qu’il visait tous les terminaux de cuisson et tous les dépôts de pain sous quelque forme que ce soit, ce qui permettait d’ordonner la fermeture de parties d’établissement, la cour d’appel a violé l’article <ref id="ref10" href="fr/act/2013-10-14/3/main#art221-17"></ref> devenu <ref id="ref11" href="fr/act/2013-10-14/3/main#art3132-29"></ref> du code du travail ; <eol/>

Mais attendu, d’abord, qu’après avoir rappelé qu’elle n’avait pas à statuer sur la légalité de l’arrêté préfectoral, la cour d’appel, qui n’avait pas à procéder à une recherche qui ne lui était pas demandée, a relevé que la fédération des entreprises du commerce et de la distribution était représentée à la réunion ayant donné lieu à l’accord critiqué et que la Maison de la boulangerie-pâtisserie de l’Eure était habilitée à signer, avec les autres organisations professionnelles et syndicales, l’accord exigé par l’article <ref id="ref12" href="fr/act/2008-03-01/nn/main#artL3132-29"></ref> du code du travail ; <eol/>

Attendu, ensuite, qu'après avoir exactement énoncé que l'arrêté pris en application de l'article L. <ref id="ref13" href="fr/act/2013-10-14/3/main#artL3132-29"></ref> du code du travail peut viser des parties d'établissement, dès lors que l'arrêté concerne, en termes généraux, les terminaux de cuisson et les dépôts de pain, qu'ils se situent dans des magasins spécialisés ou dans des magasins à succursales multiples, elle a pu en déduire que la contestation de la légalité de l'arrêté préfectoral du 9 octobre 1996 n'était pas sérieuse et qu'elle n'était pas tenue de surseoir à statuer ; que le moyen n’est pas fondé ;

</motivation>

<decision>

<p>PAR CES MOTIFS : </eol>

REJETTE le pourvoi ;</eol>

Condamne la société Atac aux dépens ;</eol>

Vu l’article 700 du code de procédure civile, condamne la société Atac à payer à la société La Maison de la boulangerie-pâtisserie artisanale de l’Eure la somme de 2 500 euros ;</eol>

Ainsi fait et jugé par la Cour de cassation, chambre sociale, et prononcé par le président en son audience publique du treize octobre deux mille dix. </p>

</decision>
<conclusions>

<!-- Composition du Tribunal -->

<p class="signature">
  <judge id="Président" refersTo="#Président">Mme Collomp</judge>, <eol/>
  <span class="signature">Président.</span>
  <eol/>
  <avocat id="Avocat général" refersTo="#ag">M. Lacan</avocat>, <eol/>
  <span class="signature">Avocat général.</span>
  <eol/>
  <rapporteur id="Rapporteur" refersTo="#rap">Rapporteur</rapporteur>, <eol/>
  <span class="signature">M. Gosselin.</span>
  <eol/>
  <avocat id="Avocat" refersTo="#Avocat">Avocat(s)</avocat>, <eol/>
  <span class="signature">Me Jacoupy, SCP Gatineau et Fattaccini.</span>
  <eol/>
</p>
</conclusions>
</akomaNtoso>