PhD-FSTC-2016-31
The Faculty of Sciences, Technology and Communication

Universitá di Torino
Dipartimento di Informatica

DISSERTATION

Defence held on 25/07/2016 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN INFORMATIQUE
AND
DOTTORE DELL'UNIVERSITÁ DI TORINO IN INFORMATICA

by

Llio Bryn HUMPHREYS
Born on 30 November 1974 in Bangor (United Kingdom)

POPULATING LEGAL ONTOLOGIES USING INFORMATION EXTRACTION BASED ON SEMANTIC ROLE LABELING AND TEXT SIMILARITY

Dissertation defence committee

Dr Leendert van der Torre, dissertation supervisor
Professor, Université du Luxembourg

Dr Guido Boella
Professor, Universitá di Torino

Dr Christoph Schommer, Chairman
Professor, Université du Luxembourg

Dr Giovanni Sartor
Professor, European University Institute (Florence)

Dr Monica Palmirani
Professor, Universitá di Bologna

Dr Matteo Baldoni, Vice Chairman
Professor, Universitá di Torino

The love of a family is life's greatest blessing.

— Anon.

This thesis is dedicated to my husband Alessio and son Elwyn.

ABSTRACT

This thesis seeks to address the problem of the 'resource consumption bottleneck' of creating (legal) semantic technologies manually. It builds on research in legal theory, ontologies and natural language processing in order to semi-automatically normalise legislative text, extract definitions and structured norms, and link normative provisions to recitals. The output is intended to help make laws more accessible, understandable, and searchable in a legal document management system.

Key contributions are:

- an analysis of legislation and structured norms in legal ontologies and compliance systems in order to determine the kind of information that individuals and organisations require from legislation to understand their rights and duties;
- an analysis of the semantic and structural challenges of legislative text for machine understanding;
- a rule-based normalisation module to transform legislative text into regular sentences to facilitate natural language processing;
- a Semantic Role Labeling based information extraction module to extract definitions and norms from legislation and represent them as structured norms in legal ontologies;
- an analysis of the impact of recitals on the interpretation of legislative norms;
- a Cosine Similarity based text similarity module to link recitals with relevant normative provisions;
- a description of important challenges that have emerged from this research which may prove useful for future work in the extraction and linking of information from legislative text.

No problem is insurmountable.
With a little courage,
teamwork and determination
a person can overcome anything.

—Rufus B. Dodge, Jr

ACKNOWLEDGEMENTS

This thesis is the result of research conducted during my period as a PhD candidate, and would not have been possible without a rich academic environment and collaboration with colleagues at the University of Luxembourg and University of Torino. I would especially like to thank my two supervisors Prof. Guido Boella and Prof. Leendert van der Torre for their excellent guidance and support. Many thanks also to Dr. Milen Kouylekov for helpful initial discussions and to the following co-authors for fruitful collaborations: Alessio Antonini, Silvano Colombo Tosatto, Loredana Cupi, Luigi Di Caro, Sepideh Ghanavati, Michele Graziadei, Joris Hulstijn, Marijn Janssen, Hristo Konstantinov, Kornel Marko, Marco Martin, Robert Muthuri, André Rifaut, Livio Robaldo, Piercarlo Rossi, Claudio Ruffini, Cristiana Santos, Carlo Emilio Salaroglio, Kiril Simov, Veli Stroetmann and Andrea Violato. I have learned a great deal from all. Last but not least, I thank my family, friends and babysitters for their considerate practical and moral support.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CoNLL: Conference on Computational Natural Language Learning

COMPAS: Compliance-driven Models, Languages, and Architectures for Services

DOLCE: Descriptive Ontology for Linguistic and Cognitive Engineering

ECJ: European Court of Justice

EMTREE: Excerpta Medica Tree

ITTIG: Institute of Legal Information Theory and Techniques

GATE: General Architecture for Text Engineering

GORE: Goal-Oriented Requirements Engineering

HTML: HyperText Markup Language

IE: Information Extraction

LTS: Legal Taxonomy Syllabus

NLP: Natural Language Processing

NLTK: Natural Language Toolkit

OWL: Web Ontology Language

RAWE: Rules Advanced Web Editor

RDF: Resource Description Language

RE: Requirements Engineering

ROC: Receiver Operating Characteristic

SIE: Simple Information Extraction

SRL: Semantic Role Labeling

SVM: Support Vector Machines

tf-idf: Term Frequency – Inverse Document Frequency

TFEU: Treaty on the Functioning of the European Union

URL: Unique Resource Locator

URN: Unique Reference Number

URN: User Requirements Notation

XML: Extensible Markup Language

WEKA: Waikato Environment for Knowledge Analysis

1

INTRODUCTION

This thesis is about populating legal ontologies with information extraction based on semantic role labeling and text similarity. Section 1.1 of this chapter describes the motivations behind this research: why there is a need for legal ontologies, and why there is a need for automated processes to assist their population. Section 1.2 provides the research questions which this thesis seeks to address. Section 1.3 describes the methodology. Section 1.4 provides an outline of the remainder of the thesis. Finally, section 1.5 lists publications of most relevance to the thesis.

1.1 MOTIVATIONS

1.1.1 *Modern Regulatory Environment*

We live in a complex regulatory environment. The body of law to which citizens and businesses have to adhere to is increasing in volume and complexity as our society continues to advance. Italy now produces thousands of laws every year, while European legislation is estimated to be 170,000 pages long. To these figures we should add the internal regulations of firms. In Italy, each bank employee is expected to know 6,000 pages of internal regulations[1]. The extent of the law over our lives is increasing as the administrative and technological instruments at the disposal of the State allows for more control of individual and business behaviour. Laws have become more dynamic, more specialised and cover more and more areas of our lives. The law is increasing in level of specialisation as advanced multi-level societies require domain-specific laws for different areas of our lives. Another development is that we are becoming increasingly subject to multi-level jurisdictions. In Europe, due to subsidiarity, laws are applicable from European, national, regional and municipal levels. In the United States, "large corporations operating in multiple jurisdictions often need to conduct a so-called '50 state survey of the law' to identify and analyze different legal requirements on different topics" [114].

1.1.2 *Accessibility of the Law*

Paper-based laws and regulations are not accessible to most of the population. In many regions in Europe and beyond, there are now official online portals making laws and decrees available to all, due in

1 Source: ABILab.

no small part to the momentum gained by the Open Government Data and Linked Open Data initiatives. However, publishing laws online does not make them truly accessible.

The specialist and sometimes obscure terminology that is prevalent in 'legalese' can create much confusion and misunderstanding. Perhaps more insidious is the intense contextualisation of legal instruments: to understand one piece of legislation requires understanding the laws that preceded it, jurisprudence, and legal doctrine. Some of these influences are signposted with cross-references, but they are rarely fully explained. Navigating the laws to find and bring together all the relevant information is laborious, but extracting the relevant information automatically is equally difficult since cross-references may be partial, vague or non-existent. In many legal jurisdictions there are explanatory notes or recitals which seek to place a piece of legislation in the context of other laws. In EU directives, recitals provide useful information on the intentions and reasoning of the legislature with regard to individual normative provisions, but do not explicitly state which ones.

Sartor [162, page 7] envisages a future legal semantic web where legal contents on the web will be enriched with machine processable information: "This information will then be automatically presented in many different ways, according to the different issues at stake and the different roles played by its users (legislators, judges, administrators, parties in economic or other transactions)" . This thesis endorses the vision of providing machine processable information about laws suited to the needs of different users, and focusses on extracting and linking structured legal information for compliance purposes.

1.1.3 *Resource Bottleneck*

Legal informatics is a mature research area with decades of research, but solutions for creating machine processable information in the legal domain are yet to be widely adopted in the commercial world, despite some interest in specialist [152] and even mainstream [179] press. One of the reasons is that many resources developed, such as ontologies and automated reasoning systems, require extensive manual annotation. Advances in natural language processing tools such as part-of-speech taggers and parsers, the growing usage of statistical algorithms for handling uncertainty and the availability of semantic resources such as WordNet and FrameNet potentially provide opportunities for automated information extraction to help develop such resources. But legal language is not natural language, and the same issues that pose problems for human understanding also create difficulties for machine processing of legal text. Building user-friendly, sustainable and reliable applications for managing legal information is not easy. It requires real understanding of legal research

and discrimination in the use of legal informatics technology to ensure that solutions are useful, reliable and cost-effective.

1.1.4 *Implications*

The lack of real accessibility to the law has significant consequences for society. It affects citizens' freedom, organisational efficiency and business compliance. The cost of clerical, research and professional legal work is high for law firms, financial institutions and public administrations. There is a real risk of missing important information and misinterpreting the law, resulting in significant costs in legal payments and reputation. Organisations are motivated to comply with legislation since failure to do so leads to undesirable consequences such as lawsuits, loss of reputation and financial penalties. With the rapid increase and evolution of regulations and policies relevant to business processes, it is difficult for organisations to constantly keep their goals, policies and business processes compliant with applicable legislation.

1.2 RESEARCH QUESTIONS

This thesis addresses a major obstacle to the goal of ensuring real accessibility to the law: the 'resource consumption bottleneck' [89, page 19] of creating semantic technologies manually. The use of automated information extraction techniques could significantly reduce this bottleneck. However, Lenci et al. [116, page 77] argue:

> Technologies in the area of knowledge management and information access are confronted with a typical acquisition paradox. As knowledge is mostly conveyed through text, content access requires understanding the linguistic structures representing content in text at a level of considerable detail. In turn, processing linguistic structures at the depth needed for content understanding presupposes that a considerable amount of domain knowledge is already in place.

This paradox is addressed in this thesis by combining state-of-the-art general-purpose natural language processing (NLP) modules with pre- and post-processing using rules based on domain knowledge.

The thesis is mainly concerned with the analysis and transformation of legislative text into a suitable representation to facilitate human and machine readability. More specifically, the research questions are as follows:

- What kind of information do individuals and organisations need from legislation to understand their rights and duties? How can

this information be structured in the most useful way to improve accessibility to the law? (RQ1) Specifically:

  - What are norms and what are the elements of norms in real legislative text? (RQ1a)
  - How to represent norms in a structured format that is human and machine readable? (RQ1b)
  - Which elements would be useful to represent in a structured format? (RQ1c)

- What are the challenges of legislative text for machine understanding? How to normalise legislative text to facilitate natural language processing? (RQ2)
- How to address the resource bottleneck problem of creating specialist knowledge management systems? In particular, how to semi-automate the extraction of definitions, norms and their elements to populate legal ontologies? (RQ3)
- What are recitals and how do they relate to norms? How to automate the linking of recitals and norms? (RQ4)
- How to evaluate the extraction of definitions and norms and the linking of recitals and norms? (RQ5)

The output is intended to help make laws more accessible, understandable and searchable in a legal document management system. This research is not concerned with automated legal interpretation (e.g. Rissland et al. [153]) or automated legal reasoning (e.g. Robaldo et al. [154].) Another limitation is that this thesis is concerned with legislation, and not with case law and legal scholarship, although the approaches proposed could be adopted for such sources in the future (e.g. Boella et al. [33]).

1.3 METHODOLOGY

This multi-disciplinary research acknowledges the important theoretical and practical achievements in relevant fields (legal informatics, ontology building, information extraction, natural language processing) and seeks, where possible, to exploit proven approaches, whilst at the same time, reviewing selected approaches in relation to findings from whole legislative text. The starting point was the terminological and frame-based legal ontologies of Eunomos legal document management system [30] and LEGAL-URN compliance management system [68]. These systems show the kind of information required by individuals and organisations from legislation, and provide suitable representations for structuring norms (RQ1). These representations were then analysed and reviewed in light of findings from manual

corpus analysis as well as EU drafting guidelines, legal theory and deontic logic (RQ1a, RQ1b, RC1c).

Notwithstanding semantic challenges, normative provisions are structured in a way that is significantly different from standard written language on which most NLP systems have been trained. To avoid the arduous task of building a dedicated system for extraction tailored to legal language structural complexities, legislative text were normalised, transforming lists into proper grammatical sentences (RQ2).

The definition and norm extraction system is based on abstract semantic representations from a general-purpose SRL module [25], as well as dependency parse trees from the same system. This simplifies the sets of rules required which in cascade identify possible norms and definitions, classify their types, and then on the basis of their types, map arguments in a semantic role tree to domain-specific slots in a legal ontology. The use of an abstract semantic representation also allows flexibility in the determination of how norms are structured so that the system can easily be modified for different applications (RQ3). To evaluate the definitions and norm extraction system, the system was tested on both seen and unseen data from a corpus of EU Directives (RQ5).

Normative provisions are subject to legal interpretation involving a number of sources, not least the recitals in the preamble of the same legislation. To understand the legal relationship between recitals and normative provisions, recourse was made to drafting guidelines, legal scholarship and jurisprudence (RQ4). Whilst highly influential, recitals are not explicitly linked to the normative provisions they relate to in the legislation itself. Manual annotation by a legal expert suggested that many links could be made on the basis of textual similarity. To test this theory, experiments were conducted based on Cosine Similarity algorithm with Term Frequency - Inverse Document Frequency (tf-idf) [174, 148] (RQ4). The results were then evaluated against the gold-standard annotations of the legal expert (RQ5).

1.4 OUTLINE OF THE THESIS

The remainder of the thesis is structured as follows. Chapter 2 provides a literature review of technologies used in this research: systems for legal knowledge and compliance management, ontologies, information extraction and ontology population, text normalisation, semantic role labeling, and text similarity. Chapter 3 discusses why laws are difficult to understand, examining the semantic and structural challenges of working with legislative text. Chapter 4 describes systems developed to help organisations understand laws, namely Legal Taxonomy Syllabus, Eunomos and Legal-URN, including how definitions and norms are represented within these systems. Chapter 5, on the other hand, analyses what kind of information is contained in legislation

from a theoretical and practical perspective, based on legal theory and manual analysis. The analysis encompasses recitals, definitions, norms and meta-norms. The deep analysis required for extracting the elements of definitions and norms requires some pre-processing of the text in order to overcome some of the structural challenges described in Chapter 3. Chapter 6 describes a module for automating the normalisation of legislative text. Chapter 7 then proceeds to describe the automated extraction of different kinds of definitions using semantic role labeling and domain-specific rules for ontology population. Chapter 8 describes the automated extraction of different kinds of norms and meta-norms using the same approach. Chapter 9 shows how an ontology of structured norms can be enriched with background information about the motivations and history behind the norms by linking implicitly related recitals to normative provisions based on textual similarity. Chapter 10 describes the evaluation of the normalisation system, the system for extracting definitions and norms, and the system for linking recitals and norms. Chapter 11 ends the thesis with conclusions and future work.

1.5 PUBLICATIONS

This thesis builds on papers co-authored on Eunomos and LegalURN, norm extraction and recital analysis. Some ideas and figures have appeared previously in the following publications:

- *Gianmaria Ajani, Guido Boella Luigi Di Caro, Llio Humphreys, Livio Robaldo, Piercarlo Rossi and Andrea Violato.* ***European Legal Taxonomy Syllabus: A multi-lingual, multi-level ontology framework to untangle the web of European legal terminology.*** *Submitted to the journal Applied Ontology, An Interdisciplinary Journal of Ontological Analysis and Conceptual Modeling.*

  This article provides a thorough description of Legal Taxonomy Syllabus and the motivations behind this work. The article informs Chapter 1, Chapter 3 and Chapter 4 of this thesis.

- *Gianmaria Ajani, Guido Boella, Luigi di Caro, Llio Humphreys, Livio Robaldo, Piercarlo Rossi and Andrea Violato.* ***Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the Law*** *Accepted to the journal Artificial Intelligence and Law*

  This article provides a thorough description of Eunomos and the motivations behind this work. The article informs Chapter 1, Chapter 3 and Chapter 4 of this thesis.

- *Guido Boella, Luigi Di Caro, Michele Graziadei, Loredana Cupi, Carlo Emilio Salaroglio, Llio Humphreys, Hristo Konstantinov, Kornel Marko,*

Livio Robaldo, Claudio Ruffini, Kiril Simov, Andrea Violato, Veli Stroetmann. ***Linking legal open data: breaking the accessibility and language barrier in European legislation and case law.*** *ICAIL 2015: pages 171-175*

This paper describes how the EUCases FP7 project seeks to break the accessibility barrier to legal information in the EU by developing new applications for the legal information provision market to enrich legal documents structurally and semantically. The paper presents the most current research based around the Eunomos legal document information system, including work presented in this thesis in Chapter 4.

- *Llio Humphreys, Cristiana Santos, Luigi di Caro, Guido Boella, Leon van der Torre and Livio Robaldo.* ***Mapping recitals to normative provisions in EU legislation to assist legal interpretation.*** *In Legal Knowledge and Information Systems: JURIX 2015: The Twenty-Eighth Annual Conference, volume 279, page 41. IOS Press, 2015.*

  This paper looks at the use of recitals in the interpretation of EU legislation, and mechanisms for connecting them to normative provisions to help legal professionals and lay end-users interpret the law. The work is elaborated in this thesis in Chapter 5 and Chapter 9.

- *Llio Humphreys, Guido Boella, Livio Robaldo, Luigi di Caro, Loredana Cupi, Sepideh Ghanavati, Robert Muthuri and Leendert van der Torre.* ***Classifying and extracting elements of norms for ontology population using semantic role labeling.*** *In ICAIL Semantic Analysis of Legal Documents Workshop, San Diego, USA, 2015.*

  This paper discusses the population of structured legal ontologies using information extraction based on Semantic Role Labeling. The focus on actions in norms renders this work particularly suitable for a dependency parsing and Semantic Role Labeling approach. The work is elaborated in this thesis in Chapter 7 and Chapter 8.

- *Guido Boella, Silvano Colombo Tosatto, Sepideh Ghanavati, Joris Hulstijn, Llio Humphreys, Robert Muthuri, André Rifaut and Leendert van der Torre.* ***Integrating Legal-URN and Eunomos: towards a comprehensive compliance management solution****. In AI Approaches to the Complexity of Legal Systems, pages 130-144. Springer, 2014.*

  This paper brings together two leading systems, Legal-URN, a Requirements Engineering based framework for business process compliance, and Eunomos, a legal knowledge and document management system, for a comprehensive compliance management solution. The work is discussed in this thesis in Chapter 4.

- *Sepideh Ghanavati, Llio Humphreys, Guigo Boella, Luigi Di Caro, Livio Robaldo and Leon van der Torre.* ***Compliance with multiple regulations****. In 33th International Conference on Conceptual Modeling (ER 2014), USA, 2014.*

  This paper discusses how multiple regulations cover the same domain and can interact with, complement or contradict each other and shows how integrating Legal-URN and Eunomos can help users identify relevant regulations and ensure that their business processes comply with the legal provisions. The work is discussed in this thesis in Chapter 4.

- *Robert Muthuri, Sepideh Ghanavati, André Rifaut, Llio Humphreys and Guido Boella.* ***The role of power in legal compliance****. In Proceedings of IEEE 7th International Workshop on Requirements Engineering and Law (RELAW) 2014, pages 23-24*

  This paper explores different kinds of power and crucial factors to be considered for modeling them in Requirements Engineering (RE). The paper informs Chapter 5 of this thesis.

- *Guido Boella, Llio Humphreys, Robert Muthuri, Piercarlo Rossi and Leendert van der Torre.* ***A critical analysis of legal requirements engineering from the perspective of legal practice****. In Proceedings of IEEE 7th International Workshop on Requirements Engineering and Law (RELAW) 2014, pages 14-21*

  This paper reviews existing approaches to representing legal knowledge for legal requirements engineering. The paper informs Chapter 2 of this thesis.

- *Guido Boella, Robert Muthuri, Llio Humphreys and Leon van der Torre.* ***Managing Legal Resources in Open Governance and E-Democracy: Eunomos - an AI and Law Response*** *in Proceedings of the Conference for E-Democracy an Open Government (CEDEM) 2014*

  This paper describes how Eunomos can encourage more active participation in e-democratic processes by enabling public administration and citizens to access laws in an Internet of Social Things environment where laws have unique identities and are enriched with interpretations. The paper is relevant to Chapter 4 of this thesis.

- *Guido Boella, Marijn Janssen, Joris Hulstijn, Llio Humphreys and Leendert van der Torre.* ***Managing legal interpretation in regulatory compliance****. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, pages 23-32. ACM, 2013.*

  This paper discusses compliance management with the Eunomos legal document management system, with a particular focus on

managing different interpretations of norms. The approach is compared with leading compliance management systems. This paper is relevant to Chapter 2 and Chapter 4 of this thesis.

- *Alessio Antonini, Guido Boella, Llio Humphreys and Joris Hulstijn:* ***Requirements of legal knowledge management systems to aid normative reasoning in specialist domains****. In JSAI International Symposium on Artificial Intelligence 2013, pages 167-182*

  This paper discusses the information gap between legal and specialist domains and the interplay between industry/professional standards and legal norms. Extensions are proposed to the Eunomos legal knowledge management tool for aligning norms with operational procedures, and use of domain-specific specialist ontologies. This paper is relevant to Chapter 4 of this thesis.

- *Guido Boella, Luigi Di Caro, Llio Humphreys, Livio Robaldo and Leon van der Torre.* ***NLP challenges for Eunomos, a tool to build and manage legal knowledge****. In Language Resources and Evaluation (LREC), pages 3672-3678, 2012.*

  This paper discusses NLP solutions for semi-automating some routine tasks currently performed by knowledge engineers for the Eunomos legal document management system, such as classifying norms, or linking key terms within legislation to ontological concepts. This thesis further addresses the resource bottleneck problem of creating specialist knowledge management systems.

- *Guido Boella, Luigi Di Caro, Llio Humphreys and Livio Robaldo.* ***Using legal ontology to improve classification in the Eunomos legal document and knowledge management system****. In Semantic Processing of Legal Texts Workshop (SPLeT 2012) at LREC, 2012.*

  This paper discusses the classification of legislative text into different topics, using an Support Vector Machines classifier with the addition of ontological definitions and relations as factors. The paper is of relevance to Chapter 4.

- *Guido Boella, Joris Hulstijn, Llio Humphreys, Marijn Janssen and Leendert van der Torre.* ***Towards Legal Knowledge Management Systems for Regulatory Compliance****. In IX Conference of the Italian Chapter of AIS, 2012.*

  This discussion paper argues that Legal Knowledge Management should consist of four pillars: (1) a legal ontology, (2) natural language processing techniques, to semi-automatically populate the ontology, (3) a systematic method for mapping legal concepts onto the actual data and business processes of a company, and (4) a method to construct an audit trail. The paper is relevant to Chapter 1, Chapter 4, Chapter 8 and Chapter 7.

- *Guido Boella, Llio Humphreys and Leendert van der Torre.* ***The Role of Roles in Eunomos, a Legal Document and Knowledge Management System for Regulatory Compliance****. In IX Conference of the Italian Chapter of AIS, 2012*

  This paper discusses the application of a domain-specific ontology building tool used for compliance monitoring with suitable extensions for modelling duties and roles, so that compliance officers can research laws and monitor compliance within the same web environment. The paper is of relevance to Chapter 4.

- *Guido Boella, Llio Humphreys, Marco Martin, Piercarlo Rossi, Leendert van der Torre and Andrea Violato.* ***Eunomos, a legal document and knowledge management system for regulatory compliance****. In Information systems: crossroads for organization, management, accounting and engineering, pages 571-578. Springer, 2012.*

  This paper introduces the Eunomos legal document management system. The paper describes the challenges of legal research in an increasingly complex, multi-level and multi-lingual world and how Eunomos helps users keep track of the law. The paper is relevant to Chapter 1 and Chapter 4.

- *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, pages 131-146. Springer, 2011.*

  This paper discusses the Eunomos legal document management system with terminology management. It describes in particular the editorial process for building legal knowledge. This paper is summarised in Chapter 4.

- *Guido Boella, Llio Humphreys, Piercarlo Rossi, Leendert van der Torre, A. Violato.* ***Eunomos, a Web 3.0 Legal Knowledge Management System that helps people make sense of the law****. In 'From Information to Knowledge Online access to legal information: methodologies, trends and perspectives', M. A. Biasiotti and S. Faro (eds), Publications Office of European Union, 2011.*

  This paper discusses how the Eunomos legal document management system is suitable for public organisations and citizens, particularly in view of the challenges and opportunities presented by Web 3.0 and the Internet of things. This paper is relevant to Chapter 4.

- *Guido Boella, Luigi di Caro, and Llio Humphreys.* ***Using classication to support legal knowledge engineers in the Eunomos legal***

document management system. *In Fifth International Workshop on Juris-informatics (JURISIN), 2011.*

This paper discusses the classification of legislative text into different topics, improving the performance of a Support Vector Machines classifier by adding the TULE parser for selecting syntactically significant features and Information Gain for selecting the most informative units. This paper is relevant to Chapter 4.

2

LITERATURE REVIEW

This chapter describe the evolution of technologies of relevance to this thesis. Section 2.1 describes systems for legal knowledge and compliance management, section 2.2 describes ontologies, section 2.3 ontology population and information extraction, section 2.4 semantic role labeling, section 2.5 text normalisation and section 2.6 text similarity.

2.1 SYSTEMS FOR LEGAL KNOWLEDGE AND COMPLIANCE MANAGEMENT

There are different legal knowledge management systems for different types of users. LexisNexis and Westlaw are popular with American lawyers and legal researchers [13]. These systems store publicly available information such as court decisions, statutes, legislative materials, and regulations. The West Publishing Company back in 1876 became a successful legal publishing company by gathering together and publishing decisions from multiple jurisdictions in the U.S. LexisNexis provided a commercial full-text online information service and news from the 1970s, and the West Publishing Company countered with the Westlaw online service. A crucial added value of these systems are sophisticated search tools to assist in navigating the law. For instance, Lexis has a patented system [135] for classifying legal documents into a hierarchy without manual intervention.

In the academic field, there are several systems for enriching legal text with semantic knowledge. NavigaNorme [22] is a system for comparing laws and creating legal ontologies. Given a paragraph of legal text, it returns related paragraphs from other legislation in order of similarity. Similarity is calculated based on textual similarity and references. NavigaNorme contains the STIA web-based tool for creating ontologies, as an extension of the Semantic Turkey Firefox-based ontology tool [79]. Eunomos (see Chapter 4) provides similar functionalities, albeit with legislative XML (Extensible Markup Language) and a multi-level bottom-up approach to ontology development. The RAWE (Rules Advanced Web Editor) editor [147] is an integrated environment that enables users to mark up legal text in accordance with the Akoma Ntoso legislative XML standard [16] and convert legal rules into LegalRuleML, an OASIS XML standard. Linked Open Data interoperability is ensured with the facility to export the LegalRuleML into RDF (Resource Description Framework).

In the field of compliance management, there are software suites used by corporations and government agencies to handle knowledge management for complex administrative processes. This approach is declarative rather than procedural: if-then rules specify what should be the case, not how it is to be achieved. In van der Pol [184], the slogan is to "separate the know (domain knowledge) from the flow (sequence of activities)". In practice, knowledge repositories can get large and unwieldy, and people use the software to model specific business rules rather than generic legal knowledge. The most comprehensive research project in this area is arguably the COMPAS (Compliance-driven Models, Languages, and Architectures for Services) project [167], which aims to support the entire compliance life-cycle. Much of this work is inspired by the success of conformance checking: verifying whether a business process description conforms to some specific set of compliance requirements. The COMPAS project architecture maintains a Compliance Requirements Repository, separate from the rules that implement them, but the project does not state how this repository should be filled.

Much academic research in compliance have sought to develop a notation to represent norms and annotate business process models. Lu et al. [125] and El Kharbili et al. [57], have sought to develop a sophisticated notation for norms and business process models with the unfortunate drawback that the models are too general for use in legal settings and the notation difficult for legally trained people. El Kharbili et al. [56] propose a framework for semantic policy-based compliance management for business processes. The authors integrate the business process ontology with a legal ontology and develop a new ontology called compliance ontology. The compliance ontology is used to verify compliance of business rules.

Requirements Engineering (RE) approaches to legal compliance focus on the fact that legal statements can be treated as a type of requirement. These approaches aim to integrate laws with other types of requirements so that they can be modelled using the same notation or language to aid comprehensibility and integration. A recent systematic literature review [70] shows that requirements engineering techniques, especially Goal-Oriented Requirements Engineering (GORE) methods, have been used to extract and model legal requirements or build business process compliance frameworks. These focus mainly on the similar behaviour of legal requirements compared to other types of requirements (such as system, business or technical requirements) and try to bind the concepts of legal goals and intentions to stakeholders' goals and intentions. Ishikawa et al. [97] model legal interpretations as high-level goals and then refine them to more concrete goals. They perform a gap analysis between the expected instances and the actual concepts to identify instances of non-compliance. Some other approaches provide formalised methods to extract and model legal

requirements with the current goal models or their extensions via deontic logic concepts or a Hohfeldian ontology. Siena et al. [173] introduce a new language called Nomos, which models normative statements in terms of eight classes of rights (from a Hohfeldian ontology). Legal requirements modelled with Nomos are linked to organisational goals through a realization class.

Beside GORE approaches, some work in RE has been done to integrate regulatory compliance with business processes to ensure compliant business processes. Karagiannis [106] uses a meta-modelling platform called ADOxx to integrate Business Process Management (BPM) and Enterprise Risk Management (ERM) meta-models into one single meta-model. The approach evaluates risks linked to business processes and provides control processes resulting from risk management. Schleicher et al. [166] define a refinement process based on compliance templates. With this approach, the compliance templates, which are abstract business processes, are refined in several steps by satisfying a set of constraints with a set of activities. This refinement is done until they reach executable legally compliant business processes.

Chapter 4 describes Legal-URN, a system which attempts to to cover gaps identified in other systems: the lack of automatic or semi-automatic compliance analysis or structured guideline on how to model legal requirements with goal models.

2.2 ONTOLOGIES

Ontology as a branch of philosophy is described by Smith [63, page 155] as "the science of what is, of the kinds of objects, properties, events, processes and relations in every area of reality". In computer science, an oft-quoted definition of an ontology is "an explicit specification of a conceptualization" [81, page 1]. The objective is typically to provide humans or more often artificial agents with knowledge about objects in the real world; in the case of the latter, to facilitate machine reasoning and problem solving, semantic searches and interoperability between systems. Upper ontologies (or foundational ontologies) such as Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [66] describe the most general entities, such as entity, object, situation, while domain ontologies describe a specific domain. 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 the 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 concept found in natural language - a difficult barrier to machine understanding and reasoning.

Ontologies work especially well in domains that have clear, explicit and unambiguous conceptualisations. For instance, DogOnt [34] 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 (Excerpta Medica Tree) [12][page 180]. Ontologies have also proven to be useful in scientific and medical domains for overcoming terminological differences. For instance, the Gene Ontology was developed to address a common problem [181]:

> [I]f 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.

It is very difficult to create an explicit specification of the law. 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 (see chapter Chapter 3 for further discussion). There are two main approaches to ontology development in the legal domain. Top-down ontologies [187] start from fundamental legal concepts defined in legal jurisprudence and proceed to narrower concepts. Bottom-up ontologies start from terms extracted from legislation or case law in specific domains. There are problems with both approaches: with the former, superficial definitions that do not reflect actual meanings in practice. With the latter, the resource-intensive approach can be prohibitive to development beyond proof-of-concept prototypes.

There are now notable real-world legal projects that use legal ontologies, but they are limited to information retrieval rather than necessarily providing an accurate representation of the law for human or machine understanding. Fernandez-Barrera and Casanovas's ONTOMEDIA project [59] adopts a bottom up approach, providing basic legal and judicial resources to citizens involved in consumer mediation processes. Users select their region and can query relevant norms on consumer law for their region. Citizens will be able to present their problem in natural language and be directed to relevant information available online. This functionality is based on mapping an user representation of a problem to a regulative representation of the problem, using information leaflets that explain regulations in normal

language as an intermediary conceptual system. Their methodology is based on extraction of terms in everyday language from a corpus of consumer queries and enrichment of specialist ontologies on mediation and consumer law with the extracted terms from the consumer queries. Cherubini and Tiscornia's Pubblica Ammistrazione e Stranieri Immigranti (P.A.eS.I.) [134] is a portal on immigration procedures. The ontology-based computable model of the normative framework helps immigration services as well as non-Italian citizens to find the information they need. Information is organised along 'life events' in which the individual or business is involved e.g. gaining citizenship, employment and access to health services, with information sheets on each topic written in clear and plain language. About 230 procedures are mapped to related legislative norms, allowing citizens and organisations to query what they must do on the basis of which norms.

Legal ontologies include not only ontologies of legal terms, but also ontologies of norms. Van Kralingen's [110] frame-based legal ontology contains three frame structures: the norm frame, the act frame and the concept-description frame. The norm frame has the fields norm identifier, norm type, promulgation, scope, conditions of application, subject, legal modality. The act frame has the fields act identifier, promulgation, scope, agent, act type, means, manner, temporal aspects, spatial aspects, circumstances, cause, aim, intentionality and final state. The concept-description frame has the fields concept, priority, promulgation, scope, conditions and instances. The frame-based ontology is a seminal piece of work which has been applied to the representation of the Dutch Unemployment Benefits Act. Francesconi [65], provides a modern system for annotating laws with functional (procedure, duty, right etc), semantic (e.g. consumer as bearer of right) and logical profiles (representing the relationship between provisions). The RDF/OWL (Resource Description Language / Web Ontology Language) implementation allows advanced access services. Carneades, combining ontologies and rules, studies open source compatibility issues [77]. The LKIF ontology models basic concepts of law identified by citizens, legal professionals and legal scholars with a reasoning mechanism. However, the system finds its limits on EU Directive 2006/126 on driving licences, a relatively straightforward regulation. The problem with all these systems is the sheer amount of basic knowledge and interconnections that need to be provided, much of it manually.

2.3 INFORMATION EXTRACTION AND ONTOLOGY POPULATION

Research on ontology learning (creating new ontologies) and ontology population (populating existing ones) is an important field of ontology engineering, albeit not without limitations. Biemann [23] states that "none of the methods used today are good enough for

creating semantic resources of any kind in a completely unsupervised fashion, albeit automatic methods can facilitate manual construction to a large extent". While ontologies can be learned from structured and unstructured data, most research on ontology population is based on extracting data from unstructured text.

Research on extracting simple concepts and their inter-relationships (rather than definitions) can involve rules or machine learning. Many concepts and ontological relations can be extracted based on simple patterns. Hearst [88] used the patterns "X, Ys and other Zs" and "Ws such as X, Y and Z" to extract is-a relations. Berland and Charniak [18] used similar patterns to find part-of relations. An alternative approach is semantic clustering. Following from Harris's [86] distributional hypothesis that "the meaning of entities, and the meaning of grammatical relations among them, is related to the restriction of combinations of these entities relative to other entities" , Hindle's system classify words according to their predicate-argument structures, i.e. nouns can be classified according to the verbs they occur with, and verbs according to the nouns they occur with.

For frame-based ontologies, we need to look at information extraction, in particular template filling. Information extraction from natural text is challenging because of language variability: the fact that the same information can be expressed with different words and syntactic constructs. Examples of language variability are lexical synonymy, syntactic word order and world knowledge inferences.

Traditionally, information extraction is approached in a supervised manner based on a set of examples expressing the relations or entities and constructed manually. Giuliano et al.'s [73] Simple Information Extraction (SIE) is a modular information extraction system designed to be easily and quickly portable across tasks and domains. SIE is composed of a general purpose machine learning algorithm, Support Vector Machines (SVM), combined with several customisable modules. A crucial role in the architecture is played by an instance filtering module, which is based on the assumption that entities to be recognised are unlikely to have low information content [74].

The core of many unsupervised information extraction systems, e.g. Yangarber et al. [199] and Stevenson et al. [175], are 'paraphrasing' modules to generate semantically equivalent components with lexical or syntactic variation. The synonym sets in the WordNet general-purpose lightweight ontology are also very useful for this purpose, and have been used by Moldovan & Rus [143], and Mihalchea & Moldovan [138]. Syntactic word order patterns, such as active/passive formulations can be generated according to standard template rules and grouped together in equivalence classes. Szpektor et al.'s [180] TEASE system is a generic paraphrasing extraction system that extracts relations between a pivot (lexical entry) and a template (dependency

parse fragment). The surrounding words of the lexical entries are used as anchor sets to extract templates.

An alternative approach to handling language variability is the transformation of text into logical form. Rus's [158] research transformed WordNet concept definitions into logical forms designed to be as close as possible to natural language. The notation module was developed for a question-answering system. The answer extraction procedure consisted of four steps: transforming questions and answers into logic forms, forming WordNet-based lexical chains between pairs of concepts, unifying lexical chains, and extracting inferences. The advantage of using logic form transformation as part of an information extraction system is that the knowledge could be used by various applications that require different information to be structured in different ways. However, this approach has already been tried and tested on legal text. Wyner [194] explored the use of C&C/Boxer [50] to extract norms from UK citizenship legislation, concluding that such systems perform better on Controlled English than the natural language constructions typically found in legislative text.

Wyner and Peters [197, 196] have researched information extraction for legislation and case reports. Their approach is semantic annotation of text, creation of a gold standard, and development of automated annotation tools. The gold standard requires annotations from domain experts, and they use the Teamware unified environment for these tasks. The Teamware system uses the open source GATE (General Architecture for Text Engineering) tool for information extraction [49] to pre-annotate the text, thereby removing some aspects of the annotation task for domain experts. The GATE platform enables template-based extraction on the basis of lookup lists (gazetteers), heuristic pattern-based grammars as well as a pipeline of standard natural language processing components such as tokenization, sentence splitting, part-of-speech tagging, morphological analysis (lemmatisation), verb and noun phrase chunking, and a parser.

There is much information extraction research involving machine learning in the legal domain. Of particular relevance to this thesis is the extraction of active roles, passive roles and involved objects in norms by Di Caro et al. [31]. Based on the idea that a semantic tag may be characterised by limited sets of syntactic contexts, their supervised machine learning approach involves the use of syntactic dependencies extracted with the TULE parser as factors in a Support Vector Machines classifier [48]. Gao and Singh [67] use pattern-matching and machine learning to extract commitments, authorizations, powers, prohibitions and sanctions from business contracts. The objective is similar to Chapter 8 of this thesis, in that they first classify the norm type, and then extract the elements of the norm. They identify norms based on use of modal verbs. They use a classifier for identifying norms using verb and clause conjunctions. Elements of norms are extracted based

on heuristics such as 'If a norm sentence has a subordinate clause led by conjunction words such as "if" and "unless", the subordinate clause expresses the antecedent. Grabmair et al. [78] and Biagoli et al. [20] also use machine learning approaches for classifying norms and extracting elements of legislation. Their methodology relies on the costly labour-intensive task of annotating legal corpora, and is less sensitive to specific linguistic expressions that are not commonly used in the corpus. Moreover, this makes it more difficult to fine-tune the extraction process.

Semantic Role Lsabeling (see below) has emerged as a suitable intermediary for unsupervised information extraction [177, 105]. To be clear, while the identification of general thematic roles is based on a general-purpose system trained on a general corpora, the information gained can then be used in a simpler system for domain-specific template filling.

2.4 SEMANTIC ROLE LABELING

Semantic Role Labeling (SRL) is the task of detecting basic event structures in a sentence such as "who" did "what" to "whom", "when" and "where" [128]. A semantic role (also known as a semantic case, thematic role, theta role or case role) is the underlying relationship that a participant has with the main verb in a clause [124, 149].

Many verbs allow a variable number of semantic roles to be realized in various syntactic positions (diathesis alternations). For instance [131]:

```
[John/AGENT] broke [the window/THEME].
[John/AGENT] broke [the window/THEME] with
[a rock/INSTRUMENT].
[The rock/INSTRUMENT] broke [the window/THEME].
[The window/THEME] broke.
[The window/THEME] was broken by [John/AGENT].
```

While an agent or instrument can both be the grammatical subjects of a sentence, a sentence such as the following is grammatically unacceptable:

```
John and a hammer broke the window
```

Fillmore [60] explained why: only noun phrases representing the same *case* may be conjoined. Fillmore used the term 'case' with reference to the case system that exist in many languages. For instance, in ancient Greek or Russian, the phrase 'with a rock' in the sentence "John broke the window with a rock" would be expressed with a single noun with the instrumental case marker, which is different from the nominative case used for "John". In English, the instrumental case is 'flagged' by the preposition 'with'. However, there is no rigid one-to-one mapping

between flags and cases –'with' can also flag the cases 'Manner' ('with glee') and 'Accompanier' ('with Nadia') [92]. Levin [120] noted that syntactic constraints on verbs and the arguments they may take are semantically determined, and created verb classes whose members pattern together with respect to diathesis alternations.

There is no consensus on a definitive list of semantic roles [129] that should be used for semantic role labeling. In FrameNet [62] (based on frame semantics [61]), arguments are related to deep roles related to specific scenarios or frames, such as Suspect, Authorities and Offense. PropBank [107], on the other hand, uses general roles and verb-specific roles based on Levin's [120] verb classes. The roles are numbered, rather than given semantic names, although in general, Arg-0 corresponds to Agent while Arg-1 corresponds to Patient. Both FrameNet and PropBank use extra-thematic elements such as Time, Manner and Place. Màrquez et al. [129] assert that most research on SRL is now conducted on PropBank, mainly because of its greater coverage. While the FrameNet corpus is a selection of illustrative sentences, PropBank has annotated the semantic roles of all verbs in the Penn Treebank II Wall Street Journal Corpus [126].

SRL systems rely on automated part-of-speech tagging and parsing. Most SRL systems use constituency parsers, probably because they have traditionally been better resourced for the English language. However, Johansson and Nugues [103, 102] argue that dependency structures offer a more transparent encoding of predicate argument relations (e.g. grammatical function such as subject and object is an integral concept in dependency syntax) and thus dependency structures are more suitable for explaining the syntax-semantics interface [136, 95]. Moreover, in their comparison of constituent-based and dependency-based SRL systems for FrameNet, they found that their performance was roughly the same, except that dependency-based systems outperformed constituent-based systems when using out-of-domain test sets, due to their lesser reliance on lexical features [103]. Another reason for choosing SRL systems based on dependency parsers is that they can be more efficient [43], rendering them more suitable for real-life applications [178].

Most automated SRL systems follow this three step architecture:

- filtering (or pruning) the set of argument candidates for a given predicate;
- local scoring of argument candidates for possible role labels, including a 'no-argument' label;
- joint scoring to combine the predictions of local scorers and ensure that the eventual labelling satisfies structural and SRL-dependent constraints.

The Mate Tools Semantic Role Labeler [25] used for this thesis follows this architecture. The first stage consists of a pipeline of inde-

pendent classifiers. It carries out the predicate disambiguation with a set of greedy classifiers, where one classifier is applied per predicate lemma. It then uses a beam search to identify the arguments of each predicate and to label them, yielding a pool of candidate propositions. The second stage consists of a reranker that is applied to the candidates using local models and proposition features. It combines the score of the greedy classifiers and the reranker in a third stage to select the best candidate proposition. Mate Tools came second in the 2009 Conference on Computational Natural Language Learning (CoNLL) shared SRL task on the joint parsing of syntactic and semantic dependencies in the closed challenge [83], and is freely available to download or use via an online demo.

Semantic Role Labeling has been used for information extraction [177, 105], question answering [144], automatic summarisation [137], textual entailment [82, 90] and machine translation [26].

There is little work on SRL for information extraction in legal informatics. Venturi et al. [186] use a FrameNet based methodology for extracting norms. Their methodology relies on a laborious methodology of building a legal domain extension to the general FrameNet, and producing a legal corpus annotated with frame information. Palmirani et al. [146] use a FrameNet approach to extract suspensions of norms in a three-step process: identification of relevant sentences, syntactic analysis, and semantic annotation using tree matching. Bertoldi and Chishman [19] investigated the use of FrameNet for building a legal ontology for Brazilian law, but decided against it, as they found that the legal frames were neither language nor jurisdiction independent as intended. The work presented in this thesis (and paper mentioned in the Introduction) is, to my knowledge, the first attempt to extract norms using a PropBank Semantic Role Labeler. It is submitted that this approach could achieve wider coverage than FrameNet, as has been the observation in other information extraction research [105]. Moreover, the use of shallow rather than deep roles allows for greater flexibility in the selection and classification of data extracted, tailored to the requirements of the relevant application.

2.5 TEXT NORMALISATION

Text normalisation is a mature research area with renewed interest due to the challenges of very irregular text on social media [55] and text messaging [14, 123, 85, 46]. Normalisation essentially involves making text suitable for part-of-speech tagging and parsing. This includes sentence segmentation, word segmentation, treatment of non-standard words such as abbreviations, URLs (Unique Resource Locators) and non-standard spellings. Techniques can involve rule-based pattern-matching or statistical approaches, depending on the task. Rule-based pattern-matching, particularly regular expressions, is the most pop-

ular technique for tokenization and has the advantage of being fast, which is an important consideration for real-life applications [130]. Probabilistic techniques, on the other hand, are suitable for spelling errors. Proven techniques include Edit Eistance [119] (finding the most appropriate standard word to replace a non-standard one based on the per-character number of modifications required) and Hidden Markov Models [42] (calculating joint probability distributions over sequences of observations). Identification of abbreviations, dates and URLs are essential both for sentence segmentation (they contain the same symbol as a full stop) and involve the use of pattern-based or statistical classifiers and/or an abbreviation dictionary [130].

Text normalisation research is dominated by the need to normalise non-standard language in social media, while elliptical lists (which are heavily used in legislation) is a subject that has received little attention in the literature. This is despite the fact that the default sentence segmentation approach in most systems, whereby each list item is regarded as a standalone sentence, leads to significant parsing errors [8]. Most solutions [8, 58, 38] involve joining the introductory clause with each list item to make standard sentences (the number of new sentences then corresponding to the number of list items). However, where the list items contain only a few words, Aıt-Mokhtar et al. [8] conjoin all the list items with the introductory clause into one sentence. Aıt-Mokhtar et al. [8] and Falco et al. [58] transform lists from proper HTML (HyperText Markup Language) markup. However, in reality, many documents containing lists are badly marked up with non-semantic markup.

There is little work on normalisation in the field of legal informatics. Wyner [198] discusses normalisation but in the context of transforming language into controlled English language. Francesconi [64] mentions segmentation-type normalisation only briefly. Surely the use of the ITTIG (Institute of Legal Information Theory and Techniques) parser for creating legislative XML documents from text helps to overcome the problem of lists. However, documentation on the ITTIG parser [21] does not mention how the lists are identified from source files and marked up. The idea of special processing for lists to make legal text more machine processable was pioneered by Buabuchachart et al. [38]. This work processes lists from text files of six Hungarian acts, where 28% of the text are lists. Their work has interesting revelations about the use of conjunctions in linking list items to determine whether none of the conditions have to be fulfilled, all have to be fulfilled, at least one has to be fulfilled, exactly one has to be fulfilled, or none can be fulfilled. The 'and' and 'or' connectives have different meanings for positive and negative conditions. For example, the famous De Morgan's law states that 'NOT (A or B)' is equivalent to 'NOT A and NOT B'. Moreover, the conjunction 'or' may take the meaning 'and' in the context of conditions: citing Jennings [100], they illustrate by

way of example the proposition 'Fred or Bill may come', which is equivalent to the proposition 'Fred may come and Bill may come'. All the work found on normalising lists [8, 58, 38] handle introductory clauses, but they do not mention end-of-list clauses. Moreover, none mention nested lists or lists without any bullet or index.

Also important to this section is the normalisation of legal references in order to simplify natural language processing. In Palmirani et al. [146], identified references are rewritten as RIF1, RIF2 and so on before sentences are provided as input to their parser.

2.6 TEXT SIMILARITY

Text similarity is an important task used for a wide range of applications including "information retrieval, text classification, document clustering, topic detection, topic tracking, questions generation, question answering, essay scoring, short answer scoring, machine translation, text summarization and others"[75]. String-based similarity measures look for similar character sequences. Semantic similarity measures use information gained from large corpora or semantic networks to look for synonyms, antonyms or words having the same context.

Mihalcea et al. [139] state that "[w]ith few exceptions, the typical approach to finding the similarity between two text segments is to use a simple lexical matching method". String-based similarity measures can be per-character or per-word. Edit Distance [119] involves calculating the per-character number of modifications required to transform one piece of text into another. N-gram Similarity [17] measures divide the number of same n-grams (word or character sequences) by maximal number of n-grams. Cosine Similarity looks at the similarity between two vectors of words representing each document, by measuring the cosine of the angle between the two vectors. The Dice Coefficient [54] is the number of common terms in the compared strings multiplied by two and divided by the total number of terms in both strings, while Jaccard Similarity [99] is the number of common terms divided by the the number of all unique terms in both strings and Euclidean distance [52] is the square root of the sum of squared differences between corresponding elements of the two vectors. Lexical similarity approaches can be improved with stemming, stop-word removal, part-of-speech tagging, longest subsequence matching, and various weighting and normalisation factors [161].

Semantic similarity measures have gained in popularity in recent years due to the availability of large corpora and semantic resources such as WordNet [141], a large lexical database and thesaurus of English. Latent Semantic Analysis [112] is the most popular corpus-based approach, and is based on the assumption that words that are close in meaning will occur in similar word contexts. The Google

Distance [44] approach uses Google hits to evaluate word similarity: two words are related if they have a higher probability of appearing on a webpage together than by themselves.

3

WHY LAWS ARE DIFFICULT TO UNDERSTAND FOR MACHINES

This chapter concerns the semantic and structural challenges of legislative text. Section 3.1 discusses the social perspective and semantic particularities of legal text. Section 3.2 describes the structural challenges - long sentences, lists, and so forth.

3.1 SEMANTIC CHALLENGES

Legal language (or legalese) is a natural (rather than formal) sublanguage. 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 and reasoning. 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. Legalese is notoriously difficult to understand since it inherits all the expressivity and ambiguity of natural language with additional terms of its own whose meaning are often obscure.

Understanding legal terminology requires deep understanding of legal culture and societal values. These problems have barely been addressed in the development of most legal ontologies, which may explain their low acceptability by legal professionals: Wahlgren states [190, pages 80–81]:

> [S]everal 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.

We will now look in greater detail at the reasons for the difficulties of understanding legal text, followed by the most common problems that occur as a result.

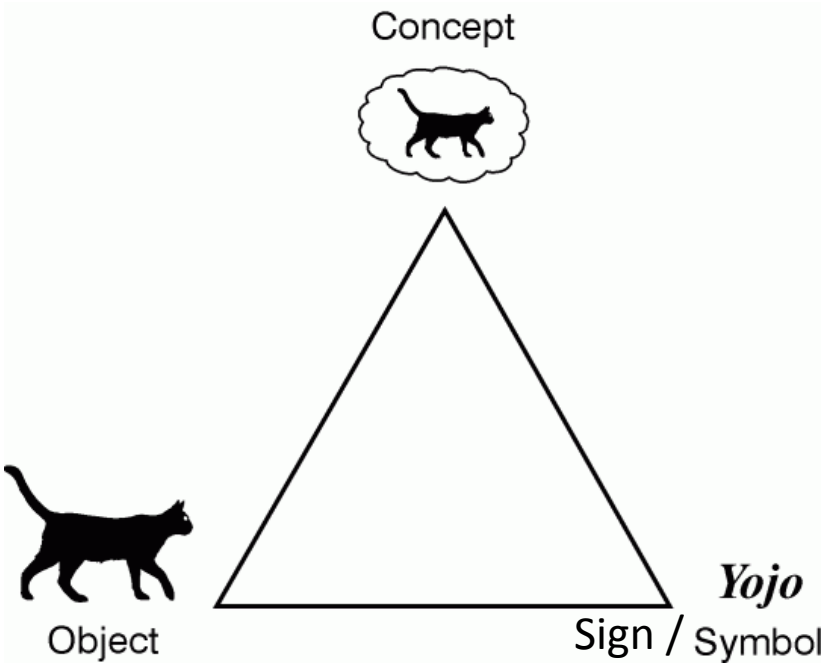


Figure 1: De Saussure's semiotic triangle

3.1.1 *The Social Perspective*

3.1.1.1 *Social Objects*

De Saussure [53], regarded as the father of modern linguistics, saw language as a 'code', a set of 'signs' that combined together, according to precise rules, to allow communication between two or more parties. He invented a relational model (the 'semiotic 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 social objects. Searle [169] stated:

> A peculiarly puzzling feature of social reality is that it exists only because we think it exists. It is an objective fact that the piece of paper in my hand is a twenty dollar bill, or that I am a citizen of the United States, or that the Giants beat the Athletics 3-2 in yesterday's baseball game. These are all objective facts in the sense that they are not matters of my opinion. If I believe the contrary, I am simply

> mistaken. But these objective facts only exist in virtue of collective acceptance or recognition or acknowledgment.

The power to *clarify* the meaning of legal social objects lies with the legislature and the courts. The difficulty is that, for reasons explained below, there is uncertainty about the meaning of legal terms due to the existence of a number of different but simultaneously valid conceptualisations.

3.1.1.2 *Dynamic Environment*

Legal norms cannot be fully defined for all possible situations because the law is applied to dynamic environments, so that meaning is neither static nor logically deductible: Difficulties can arise even for seemingly straightforward constitutive norms, as argued by Hart [87, 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 penumbral 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.

Rawls [151] 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. If clarification is required, the process of 'legal interpretation' usually involves analogy and invention, which are highly subjective activities, rendered all the more difficult because, as Sacco states, norms are not "legal flowers without stem or root" [159, page 27].

3.1.1.3 *Social and Cultural Background*

The comparative lawyer Rodolfo Sacco [159] identified several factors that make it impossible to understand the law at face value. One of these factors was what he called *cryptotypes* - the beliefs or mentality, the social and cultural background of those who have the power to interpret the law. 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 question 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 [159, page 345].

Sacco noted that the choice of language used can have important consequences for legal interpretation. 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, 'freedom of contract' is more ideologically connected to the ideal of liberty than 'autonomy of the contracting parties'. [159, pages 15–16]

3.1.1.4 *Multiple Sources of Law*

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 [159, 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 was no legally valid mechanism for trade unions to register, collective labour agreements should not be binding according to the Constitution. Nevertheless, the courts 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.

In reality, 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 [121] 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': corresponding to the grammatical rule above,
- the 'mischief rule': corresponding to the historical rule above, 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 can be used at will to best serve the demands of justice or the prejudices of the interpreter. The European Court of Justice (ECJ) favours the purposive approach to interpretation where an EU law provision is ambiguous or incomplete; i.e. the courts should always seek to give effect to the legislative purpose/objective behind the law. In order to ascertain the purpose of the legislation, the ECJ analyses relevant recitals in the preamble of the relevant legislation (along with preparatory documents and legislative proposals).

3.1.2 *Particularities of Legal Text*

3.1.2.1 *Obscure Terms of Art*

Some terms, understood as "terms of art", have acquired meanings from statutory definitions and scholarly or judicial interpretations that differ from their meaning in ordinary language. It is not always clear where to find the correct meaning for the term because legal interpretations often gain acceptance with professionals before influencing subsequent definitions in legislation. It is perhaps most striking in

borrowings from other languages. For example the French word *chose*, which simply means 'thing', has acquired specific technical meanings in English and Welsh Common Law e.g. '*chose* in action' is a right to sue. A more subtle example is the frequent use of the word 'failure' in expressions such as 'failure to comply' to mean 'unable or unwilling for whatever reason' rather than the more restricted meaning of 'unable' in ordinary language.

3.1.2.2 *Polysemy and Evolving Conceptualisations*

Polysemy is a significant problem in legal terminology, because legal terms can have significantly different meanings in different legislation, within different legal domains, across jurisdictions and over time.

Liebwald [121] described as a case in point the evolution of the meaning of 'essence of marriage'. The concept derives from The Marriage Act of 1938 in Germany, used to assess whether divorce or annulment is morally justified. During the Nazi 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.

3.1.2.3 *Neologisms*

While terms of art are redefined all the time, new ones are also added. Tiersma [182, page 31] 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 [sic] box or icon, indicating acceptance of the terms of the license)."

3.1.2.4 *Ambiguity*

Legislation is sometimes unintentionally ambiguous due to poor drafting. The Supreme Court [4, page 9] advises that in cases of attributive ambiguity, legislative intent may override literal interpretation:

> Ordinarily, as in everyday English, use of the conjunctive "and" in a list means that all of the listed requirements

> must be satisfied, while use of the disjunctive "or" means that only one of the listed requirements need be satisfied...however; if a "strict grammatical construction" will frustrate evident legislative intent, a court may read "and" as "or" , or "or" as "and".

Such was the case when the Appellate Court of Illinois concurred with the lower court that the restrictive covenant that the plaintiff, a licensed physician, "shall not (a) engage in the practice of medicine within a restricted area and (b) solicit patients of the District to become private clients" meant that he should not engage in either (a) or (b) (as opposed to the plaintiff's argument that the restrictive covenant merely prohibited him undertaking both (a) and (b) simultaneously) [5].

Such interpretations can have grave consequences. In the English case of R v Casement (1917) [39, 11], the defendant, who during the First World War tried to persuade Irish prisoners of war to fight for Home Rule against the British, was convicted of treason. The case revolved around different interpretations of the following clause (in Norman French, without punctuation) of the Treason Act 1351:

> *Si homme leve de guerre contre notre dit Seigneur le Roi en son Roialme ou soit aherdant as enemys notre seigneur le Roi en le Roialme donant a euz eid ou confort en son Roialme ou per aillours*

which translates as:

> If a man do levy war against our said Lord the King in his realm or be adherent to the enemies of our Lord the King in his realm giving to them aid and comfort in the realm or elsewhere

The defence argued that to 'be adherent to the enemies of our Lord the King' outside the realm is not treason since 'or elsewhere' applies only to 'giving aid and comfort' . The Prosecution argued that punctuation should be inferred, as per below, with the effect that 'or elsewhere' is read as appended to all three instances of "in his Realm":

> If a Man do levy War against our Lord the King in his Realm, or be adherent to the King's Enemies in his Realm, giving to them Aid and Comfort in the Realm, or elsewhere.

The interpretation of the Prosecution was accepted and the defendant was said to have been 'literally hanged on a comma'.

3.1.2.5 *Vagueness*

Liebwald [121] argues that while natural language is inherently imprecise, much of the vagueness found in legislative drafting and judgments is 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' and speed limits, it is also full of indeterminate and malleable concepts such as 'good faith' and 'reasonable discretion'.

Breaux [35] provides a typical example in HIPAA 164.512(e)(1)(iv) which "states that an entity must make 'reasonable' efforts to notify individuals of certain requests for their protected health information. The word "reasonable" is an intended ambiguity: exactly which mechanisms are considered reasonable, (e.g. postal mail, secure electronic mail or websites, etc.) varies depending on the type of communities served and the prevalence of relevant, existing technologies".

3.1.2.6 *Partial Articulation*

Sacco noted that it is quite common to find legal rules that are partially articulated. A synecdoche (partial articulation) occurs when only part of a phenomenon is indicated when referring to the whole. He gives as an example that the legal definition of contract in French law refers to the 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.

3.1.2.7 *Metaphors*

A legal sublanguage is in constant and continuous interaction with ordinary language, and uses all its linguistic baggage. Metaphors are an important linguistic feature often used to represent and navigate conceptions. Lakoff [111, 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.

While metaphors are not frequently found in legislation, they are found in other legal sources which impact upon interpretations of legislation. Winter [193] used the American case of NLRB v Jones & Loughlin Steel (1937)[1] to show how the power of metaphors is used

1 NLRB v Jones & Loughlin Steel (1937) 301 U.S. 1 1937.

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)[2] 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)[3] 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 a 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 [193] concludes that Hughes's reasoning was a 'metaphoric tour de force' that was neither predetermined by the materials nor completely arbitrary.

3.2 STRUCTURAL CHALLENGES

Legislation is difficult for NLP (or even humans) to parse because of the way the material has been laid out. NLP has traditionally been performed on text taken from news articles written in clearly identifiable sentences, but legislation is a great deal more intricate. Venturi [185, page 1], referring to Gildea [71], points out that "[a]pplying [statisticially-trained NLP] tools to out-of-domain corpora is known to be problematic" and that "when applied to domain-specific texts (e.g. bio-medical literature, law texts) their accuracy decreases significantly".

Legal text has some peculiarities, and additional work is required to obtain useful results. Some key features of legislative language, as highlighted by Wyner & Peters [195], are :-

- long sentences with several clause dependencies;
- lists, where each item is not a standalone sentence;
- difficulties for inter and intra-sentential anaphora resolution;
- references to other articles, the content of which is not quoted within the referring article.

To this list can be added the effect of implicit and explicit modifications on previous legislation, while many of the usual challenges of natural text remain, such as non-standard 'words' such as abbreviations, URLs, reference numbers, and dates. On the other hand, legal language does have the advantage that modal verbs are less ambiguous e.g. 'must' and 'shall' have deontic meanings and almost never

2 Kidd v Pearson (1888) 128 U.S. 1, 20.
3 Stafford v Wallace (1922) 128 U.S. 495, 521.

the inferential or predictive alternatives (see [45] for a comprehensive discussion of alternative modal meanings in natural language).

Below we discuss the challenges of processing legal text in greater detail. Chapter 6 discusses attempts to solve some of these problems for the purpose of information extraction.

3.2.1 *Long sentences*

Laws are well-known to contain long sentences, for reasons we can only speculate, e.g. "the subject matter of most legal writing lends itself to qualifications, modifiers, asides, and lists" [165].

Sentences such as the one below (66 tokens) from Directive 98/5/EC are by no means unusual:

> Notwithstanding points 1 to 4, a host Member State, insofar as it prohibits lawyers practising under its own relevant professional title from practising the profession of lawyer within a grouping in which some persons are not members of the profession, may refuse to allow a lawyer registered under his home-country professional title to practice in its territory in his capacity as a member of his grouping.

Venturi [185] analysed the linguistic differences between the TEMIS corpus of legislation in Italian (regional, national and European) with the ISST-TANL corpus of news reports in Italian. She found that laws on average contain sentences of 31.36 in length compared with 21.87 in the news reports (the regional laws averaged 41.95 tokens, national laws 39.04 tokens and EU laws 24.56 tokens). Moreoever, "i) legislative sentences contain dependency links much longer on average (14.5) than the ones of the general-Italian sentences (8.61) and ii) the average parse tree height of TEMIS (7.44) is higher than the one characterizing the ISST-TANL sentences (5.28)" [185, page 3].

Long sentences have traditionally created problems for parsing as attested by Tomita [183, page vii]:"I sometimes wondered whether it was ever possible to build a natural language parser that could parse reasonably long sentences in a reasonable time without help from large mainframe machines." Su et al. [176, page 94] stated that "a real system usually set a time limit to stop the parsing process when a sentence is taking too long to parse because of its long sentence length or complicated structure." Modern parsers are capable of providing parses that are good enough for many purposes, even if not entirely accurate e.g. "MaltParser ... typically performs worse on long sentences, long dependency arcs and arcs higher in the graphs." [133, pages 129–130].

3.2.2 *Lists*

Most legislation contain a great number of lists e.g. there are norms that apply to many different scenarios or a list of requirements are necessary to satisfy a norm. We may presume that this style may have been developed to save space in printed text. Moreoever, lists are regarded by Goody [76] as a grapholinguistic technique (other such techniques are tables and graphs), which provides additional information visually in a way that cannot be so easily transmitted orally. There is surprisingly little work on handling lists in processing legal text. The exception is Markovich et al. [127], who render lists similar to standard sentences so that they can be processed accurately. A similar approach is adopted in Chapter 6.

Lists in legal text can involve several degrees of nestedness as can be seen in this example from Statutory Instrument 640/2010 (Ireland)[4]:

```
(a) A solicitor in respect of whom a practising certificate
(within the meaning of the Solicitors Acts 1954 to [1994])
is in force shall be an investment business firm -
(i) where the solicitor provides investment business
services or investment advice in a manner which is not
incidental to the provision of legal services, or
(ii) where the solicitor holds himself or herself out as
being an investment business firm, or
(iii) where, when acting as an investment product
intermediary in a manner incidental to the provision of
legal services, the solicitor holds an appointment in
writing other than from -
(I) an investment firm authorised in accordance with the
Investment Services Directive by a competent authority
of another Member State, or an authorised investment
business firm (not being a restricted activity investment
product intermediary or a certified person), or a member
firm within the meaning of the Stock Exchange Act 1995
, or
(II) a credit institution authorised in accordance with
Directives 77/780/EEC of 12 December 1977 and
89/646/EEC of 15 December 1989, or
(III) a manager of a collective investment undertaking
authorised to market units in collective investments to
the public, which is situate in the State or the relevant
branch of which is situate [sic] in the State, and shall be
required to be authorised as an authorised investment
business firm pursuant to the provisions of the
[Investment Intermediaries Act] 1995.
```

4 Statutory Instrument 640/2010 - European Communities (Lawyers' Establishment) Regulations 2003 (Qualifying Certificate 2011) Regulations.

3.2.3 *Anaphora*

Anaphoras are words (usually pronouns) that refer to entities mentioned (usually) earlier in the sentence or in a previous sentence. Natural text often contains anaphoras to avoid repetition, and legal text is no exception. Inter- and intra-sentential anaphora resolution is an important NLP research area. Mitkov [142] states that '[a]pproaches to anaphora resolution usually rely on a set of "anaphora resolution factors". Factors used frequently in the resolution process include gender and number agreement, c-command constraints, semantic consistency, syntactic parallelism, semantic parallelism, salience, proximity etc. These factors can be "eliminating" i.e. discounting certain noun phrases from the set of possible candidates (such as gender and number constraints, c-command constraints, semantic consistency) or "preferential", giving more preference to certain candidates and less to others (such as parallelism, salience).'

Modern systems can solve many anaphora but certainly not all. The well-known Stanford Core NLP system [1] has no problem resolving the following anaphoric references from S.I. 2000 No. 1119 (UK)[5]:

> Where, on 22nd May 2000, **a European lawyer** is practising professional activities under **his** home professional title on a permanent basis in England and Wales or Northern Ireland or commences such practice by 21st November 2000, **he** shall apply to be registered in accordance with regulation 16 by 21st November 2000 where **he** intends to continue to practise those activities on a permanent basis after that date.

The system is also able to resolve an anaphora referring back to an earlier sentence from the same legislation:

> Subject to paragraphs 2 and 3, **a lawyer practising under his home-country professional title** carries on the same professional activities as a lawyer practising under the relevant professional title used in the host Member State and may, inter alia, give advice on the law of **his** home Member State, on Community law, on international law and on the law of the host Member State. **He** shall in any event comply with the rules of procedure applicable in the national courts.

Anaphoric references are not only pronouns, and the sentence below from Directive 98/5/EC[6] is too difficult for most systems, including

5 Statutory Instrument 2000 No. 1119 The European Communities (Lawyer's Practice) Regulations 2000.

6 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

that of Stanford, as it requires the system to understand what constitutes an activity:

> Member States which authorise in their territory a prescribed category of lawyers to **prepare deeds for obtaining title to administer estates of deceased persons and for creating or transferring interests in land which, in other Member States, are reserved for professions other than that of lawyer** may exclude from **such activities** lawyers practising under a home-country professional title conferred in one of the latter Member States.

The following references from Directive 98/5/EC[7] are also missed by the system. The 'latter' reference is particularly problematic as there are several preceding entities, and we need semantic knowledge to understand which one it refers to:

> For the pursuit of **activities relating to the representation or defence of a client in legal proceedings** and insofar as *the law of the host Member State* reserves **such activities** to lawyers practising under the professional title of that State, *the latter* may require lawyers practising under their home-country professional titles to work in conjunction with a lawyer who practises before the judicial authority in question and who would, where necessary, be answerable to that authority or with an 'avoué' practising before it.

3.2.4 *Cross-references*

The ubiquitous use of cross-references in legislative text can lead to problems, not only in readability, but also in determining which parts of a referenced article are relevant, particularly for automated systems.

For example, Article 6(1)(1) of Directive 98/41/EC[8] states:

> Each Member State shall, as regards every passenger ship that flies its flag and departs from a port located outwith the Community and is bound for a port located within the Community, require the company to ensure that the information specified in Articles 4(1) and 5(1) is provided as laid down in Articles 4(2) and 5(2).

Extracting the information required in Article 5(1) is quite feasible as the paragraph contains the words 'following information' followed by a list:

7 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

8 Council Directive 98/41/EC of 18 June 1998 on the registration of persons sailing on board passenger ships operating to or from ports of the Member States of the Community.

> The **following information** shall be recorded regarding every passenger ship that departs from a port located in a Member State to undertake a voyage of more than twenty miles from the point of departure:
> - the family names of the persons on board,
> - their forenames or initials,
> - their sex,
> - an indication of the category of age (adult, child or infant) to which each person belongs, or the age, or the year of birth,
> - when volunteered by a passenger, information concerning the need for special care or assistance in emergency situations.

It is far more challenging (for a machine) to ascertain the information that is referred to in Article 4(1):

> All persons on board any passenger ship which departs from a port located in a Member State shall be counted before that passenger ship departs.

Similarly, semantic knowledge is required to ascertain that providing information in Article 6(1)(1), the referring article, equals communicating and collecting in Article 4(2) and communicating in Article 5(2) below:

> Article 4(2):
> Before the passenger ship departs the number of persons on board shall be communicated to the master of the passenger ship and to the company's passenger registrar or to a shore-based company system that performs the same function.

> Article 5(2):
> That information shall be collected before departure and communicated not later than thirty minutes after the passenger ship's departure to the company's passenger registrar or to a shore-based company system that performs the same function.

3.2.5 *Legislative Updates*

Another problem is legislation that modifies other legislation. Some laws state explicitly which articles of other legislation are modified, others modify them only implicitly. The parliamentary principle of 'implicit abrogation' means that in the case of conflict between two norms, the more recent legislative norm will prevail, if it applies to

the same subject, whether or not it mentions the norm it effectively overrules.

Seta [150, page 81] commented on this issue:

> In the Italian legal system what is really difficult for citizens, as well as for the interpreter (the judge), is to recognize the final legislation resulting from the continuous, fragmentary and sometimes dispersed law-making process. This activity may involve the comparison of many acts and of explanatory notes, given that in the Italian legislation only very few consolidated codes are present.

The Italian Parliament occasionally does produce official consolidated codes. But most of the time, this work is left to independent agencies, whose interpretation does not have official status.

Even legal drafters have difficulty keeping up. In the US, failures in the legislative drafting process resulted in legislation that continues to refer to a norm that has since been overridden:

> ADAAG references the A17.1 elevator code for conformance. Since 2000 there has been no section of the A17 that references lifts for the disabled. Therefore ADAAG references a non-existent standard.
> *Balmer [15, page 10] cited by Lau [114, page 676]*

Solutions have been developed to address the problem of legislative updates. The Norma-Consolidation software [145] stores all legislative updates, and makes it possible to retrieve the norms in force at any date supplied to the system. There has also been research on detecting and extracting modifications in legislation [118, 155]. Unfortunately, such solutions are not known to be currently implemented by legal drafters and publishers.

4

SYSTEMS FOR HELPING ORGANISATIONS UNDERSTAND LAWS

This chapter describes the development of existing systems and technologies of relevance to this thesis. Section 4.1 describes Legal Taxonomy Syllabus, a multilingual legal ontology framework. Section 4.2 describes Eunomos, a legal document management system. Finally, section 4.3 describes Legal-URN, a Requirements Engineering based framework for business process compliance, and its integration with Eunomos.

4.1 LEGAL TAXONOMY SYLLABUS

The motivation for the development of a multilingual legal ontology framework was that anyone involved in transnational activities would benefit from access to user-friendly specialist ontologies to manage the deep, complex and interconnected terminology required for understanding laws in different jurisdictions. None of the pre-existing legal ontologies, taxonomies or thesauri available, for reasons explained in Chapter 2, were able to 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. A legal ontology that handles these issues could be useful for many different types of users e.g. lawyers who deal with cross-border issues, international financial institutions, or EU translators.

The system does not assume the existence of a single taxonomy covering all languages since different national systems may organize legal concepts in different ways. Figure 2 provides an example. *Withdrawal* and *recesso* are used as equivalent terms (concept *EU-2*) in some European legislation, such as Directive 90/314/EEC.[1] In that context, the term involved an act having as its purpose consumer protection (*difesa del consumatore* in Italian). In the British legal system, however, only a subtype of *withdrawals* have this goal, which the code refers to as *cancellation* (concept *Eng-3*). In the Italian legal system, the term *diritto di recesso* is ambiguous, since it can be used with reference to *risoluzione* (concept *Ita-4*), or *recesso* proper (concept *Ita-3*).

Legal Taxonomy Syllabus (LTS), informed by studies in comparative law [157] and ontologies engineering [108], was developed based on the following assumptions about the domain:

1 Council Directive 90/314/EEC of 13 June 1990 on package travel, package holidays and package tours.

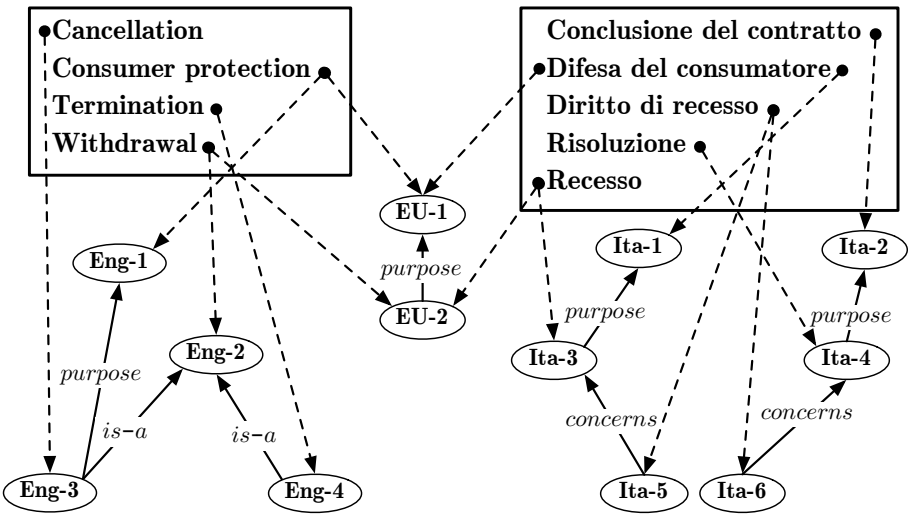


Figure 2: The mapping in LTS of European legal terms to national legal terms in Italy and England & Wales

- 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.
- 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.
- 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 European directive necessarily introduces in a national ontology the same concepts that are present at the EU level.

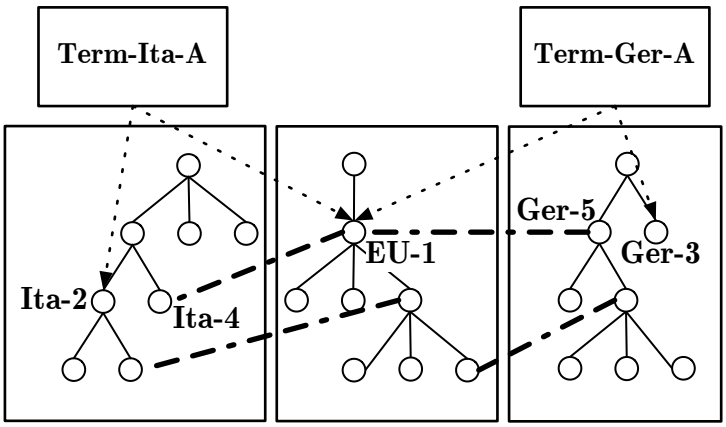


Figure 3: Mapping terms in the LTS European level ontology to terms in national ontologies

The challenge in developing Legal Taxonomy Syllabus was to design an ontology tool that captures the complex factors that influence the meaning of legal terms, but at the same time structures that knowledge in a way that leads to clarity rather than confusion.

LTS is based on a clear distinction between a legal term and a legal concept. The concepts are arranged into ontologies. Lexical terms are then linked to all concepts that apply. This organisation is evocative of de Saussure's [164] semiotic triangle referred to in Chapter 3. For the end-user, each term in the ontology is presented in an associated information table with the following information:

- language
- jurisdiction
- domain
- description in natural language
- references to legislative definitions
- links from national terms (and their definitions) to European ones and vice versa
- links to equivalent terms in other languages
- a list of other terms having the same meaning
- an ontology graph showing parent and child concepts
- notes

Legal Taxonomy Syllabus is not just one ontology. It is a framework of interlinked but separate ontologies for different languages and jurisdictions in Europe. Within LTS, we can talk about direct EU-to-national translations of terms, and about implicit national-to-national

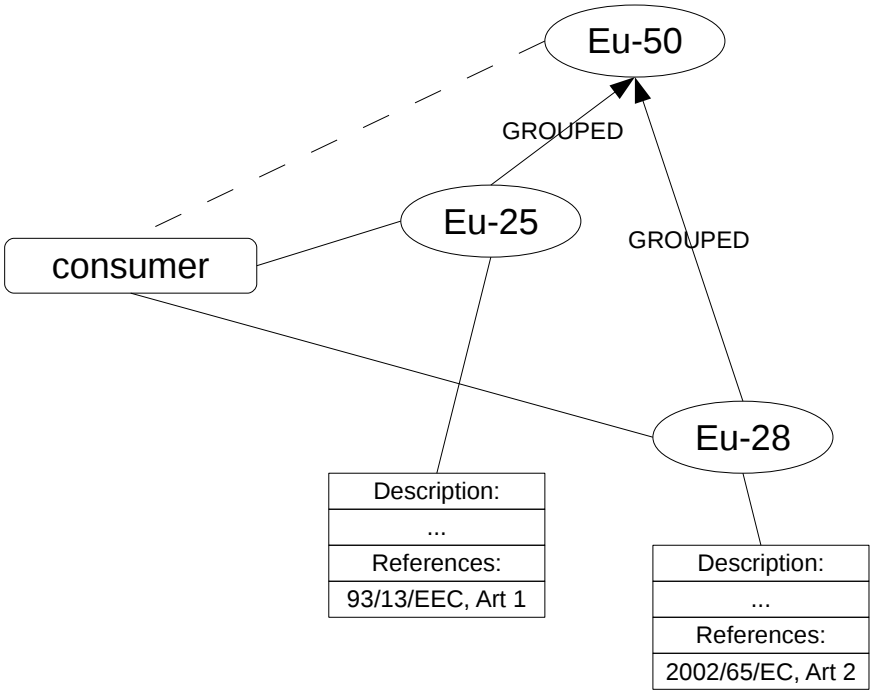


Figure 4: Grouping similar concepts into a generalised concept in LTS

translations of terms i.e. if we start from a concept at a given national level, by following a direct link to the same concept at the European level we will be able to see how that concept is mapped onto concepts at various national levels. Ajani et al. [10] shows how concepts from different ontologies in LTS are interlinked to help users find similar terms in other languages and jurisdictions.

In Figure 2 we see that beyond the usual *is-a* (linking a category to its super-category), there are a number of other ontological relations designed to help users understand the inter-relationship between concepts. The *purpose* relation links a concept to the legal principle behind it, while *concerns* refers 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* and *recesso*, *difesa del consumatore* and *consumer protection*).

The LTS bottom-up approach is suitable for dealing with legal interpretation and the evolution of terms. As more legislation-specific definitions about terms are collected, it becomes possible to find common attributes and derive a more general definition. The general definition is then linked to all the definitions that informed it via a *group_by* relation. This is particularly useful in cases where national legislatures transpose sets of European directives into one national 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.

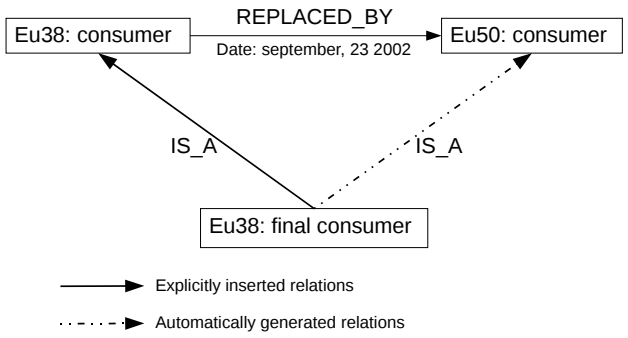


Figure 5: Relating an expired concept to a concept that replaces it in LTS

Another important issue is that when new legislation is approved and enacted, it can introduce a number of new definitions that render old definitions obsolete. LTS handles this issue 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).

This situation was made even more complex during the efforts to harmonise an European glossary, and a draft *Acquis Communitaire* Principles glossary was issued intended to replace existing European terminology in the future. While not yet officially in force, the draft glossary was beginning to have some influence on legal interpretation. This situation was handled by creating a separate ontology for *Acquis Communitaire* Principles, which was closely linked to the EU ontology of current definitions, with explicit associations between *Acquis* Principles concepts and EU-level concepts.

4.2 EUNOMOS

The second system of relevance to this thesis is Eunomos, a legal knowledge document management system to help professional users keep track of the law and ever-changing legal obligations. The basic idea of Eunomos is to create a strict coupling between legal knowledge and its legislative sources, associating the concepts of its legal ontology with regulations structured using legislative XML. The system takes inspiration from technologies developed in the related fields of legal ontologies, extending the tool for building legal ontologies called Legal Taxonomy Syllabus [9] as described above, and legislative drafting for parliaments (so-called legislative XML). Eunomos provides a web-based interface for users and Eunomos knowledge engineers to find information about laws and legal concepts in different sectors and different jurisdictions. The Eunomos system is envisaged as being useful to a wide range of user groups but is targeted towards compliance officers in the first instance, because they have the greatest need

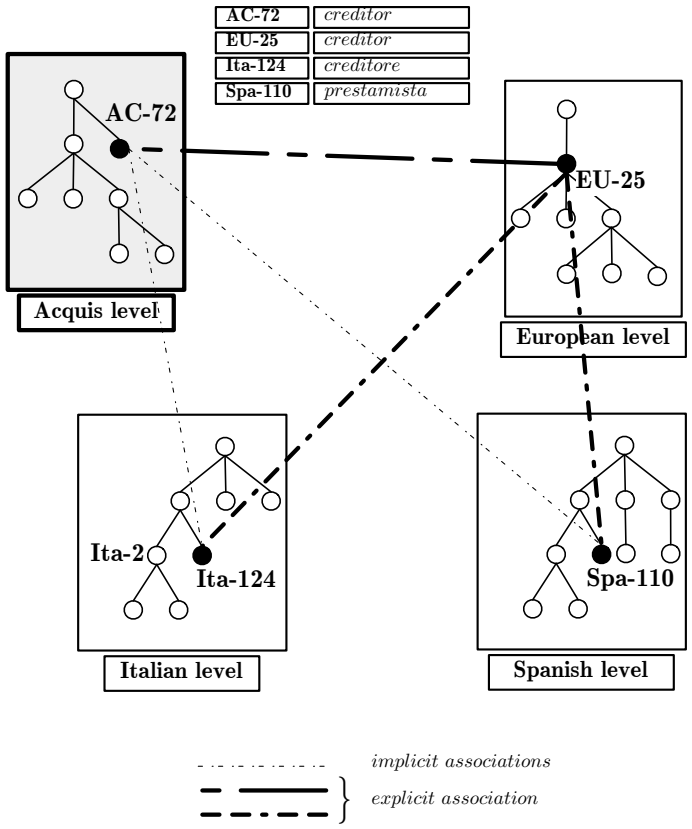


Figure 6: Mapping LTS Acquis Principles concepts to existing European level concepts, which are in turn mapped to national level concepts

and enthusiasm for a system of this kind. Eunomos can be employed as in-house software that enables expert users to search, classify, annotate and build legal knowledge, and keep up to date with legislative changes. Alternatively, Eunomos can be offered as an online service so that legislation monitoring is effectively outsourced. The software and related services can be provided to several clients, which means that information and costs are shared. The Eunomos system is the basis of the Menslegis commercial service for compliance distributed by Nomotika s.r.l., a spinoff of University of Torino.

The legal document management part of the system is composed of a legal inventory database of norms (about 70,000 Italian national laws in the current demo) converted into legislative XML format using the Institute of Legal Information Theory and Techniques (ITTIG)'s XML parser. Maintaining laws in the NormeInRete (NIR) XML format makes it easier for Eunomos to extract elements such as paragraphs, articles and references, so that knowledge engineers can categorise and annotate the elements and lawyers can view specific relevant information. Eunomos uses the XML Leges Linker tool developed by ITTIG[2] to find cross-references, an URN (Unique Reference Number)

2 www.xmleges.org

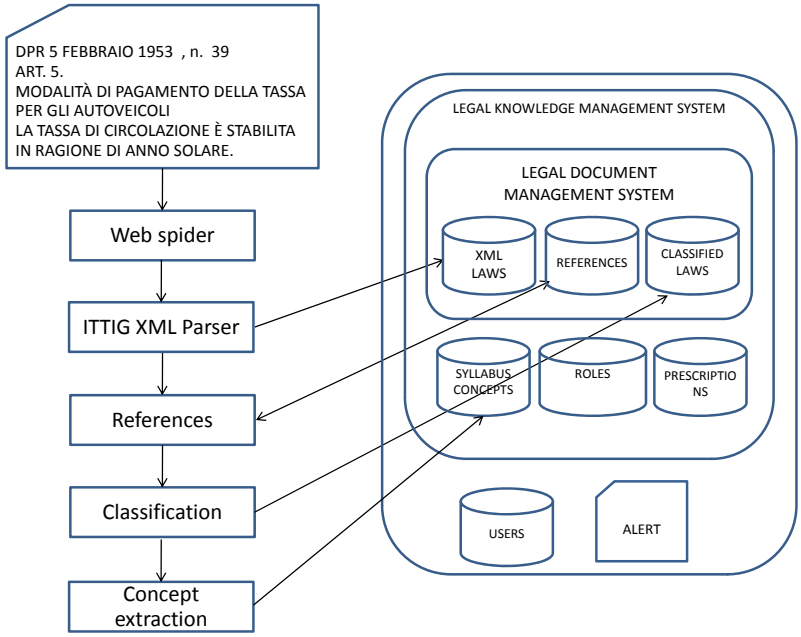


Figure 7: Key components of the Eunomos system

name resolver to obtain actual addresses of legislative articles, and XSLT (Extensible Stylesheet Language Transformations) to find and display outgoing and incoming hypertext links. Links between related legislation are created by automated analysis of in-text references. Each article is semi-automatically classified into legal domains. Most laws are collected from portals by means of web spiders on a daily basis, but they can also be inserted into the database via a web interface. Currently the system harvests the Normattiva national portal[3], the Arianna portal of Regione Piemonte[4] and a portal of regulations from the Ministry of Economy. For each legislation, Eunomos stores and timestamps the original and most up-to-date versions, but nothing prevents the inclusion of multiple versions of the coordinated text for users such as lawyers and judges who wish to see the law as it was on a given date in the past. An alert message is generated by the system to users if a newly downloaded legislation appears to be relevant to users' domains of interest. After legislation are converted into legislative XML, cross-references are extracted to build a network of links between norms citing other norms. The editorial process of the norm proceeds in a manual manner with a classification phase which is supported by tools suggesting categories on the basis of different clues. Finally, relevant concepts can be extracted and modelled using the Legal Taxonomy Syllabus ontology framework. The only difference with Eunomos is that there is a link to the XML version of the legislation

3 http://www.normattiva.it
4 http://arianna.consiglioregionale.piemonte.it/

via URN identifiers, while in the original LTS system there was only a hyperlink to the textual version of the norm.

In Eunomos, the Legal Taxonomy Syllabus ontology framework is extended to include prescriptive norms (as opposed to merely terminological 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.), but with additional characteristics. Each prescription is necessarily connected to relevant concepts defined in the terminological 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).
- description: the prescription reworded as necessary to aid comprehension.
- norm identifier: a hyperlink to the relevant provision in the source document.
- violation: the crime or tort resulting from violation (often defined in other legislation such as a Penal Code).
- sanction: the sanction resulting from violation (e.g. a fine of 1 'quote', where a 'quote' is defined in other legislation).
- notes: information about court decisions, scholarly interpretations or other information of interest.

Prescriptions are 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. For instance, a semantic search for prescriptions that relate to sellers would also bring up prescriptions for traders (the grandparent of seller in the terminological ontology).

Knowledge engineers are essential to maintain a reliable service and provide additional information where needed. They are responsible for:

- checking the output of the legislative XML parser and correcting any errors arising out of irregular patterns in the text;
- inserting missing legislation in the database;
- classifying the domain of legislative norms by selecting among the suggestions proposed by the automated classifiers;

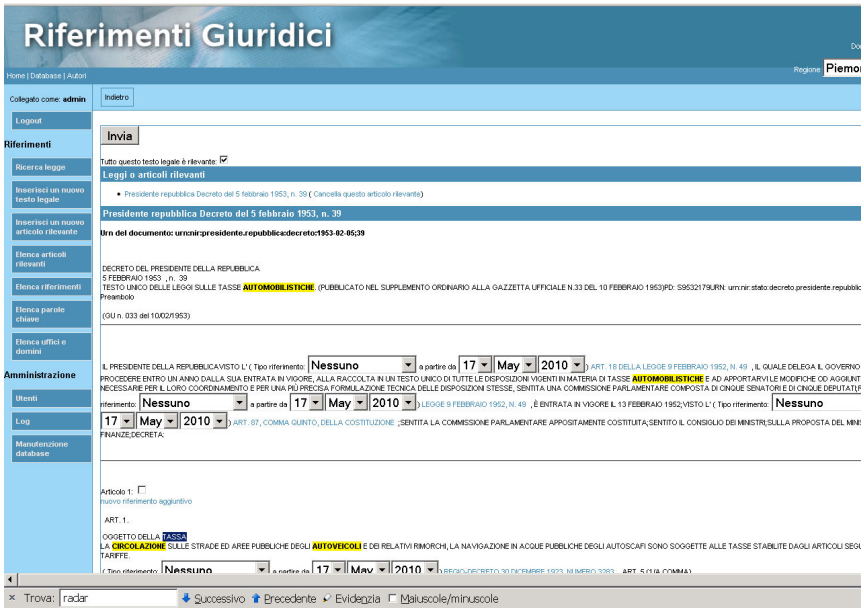


Figure 8: Annotating legislation in Eunomos

- classifying the type of modificatory references;
- adding concepts and terms to the ontology;
- adding explanations in plain language of terms and legal obligations;
- adding relevant information from case law or scholarly interpretation;
- checking for cross-references missed by the parser due to irregular textual patterns;
- manually inserting implicit cross-references where legislation fails to mention the legislation it modifies or overrides.

In Figure 8, we can see annotated articles from a piece of legislation. The knowledge engineer uses this interface to specify whether an article is relevant to the domain under consideration. (S)he can add a type (modification, suspension, etc.) for each reference to other legislation. Terms which are linked to concepts in the ontology of the relevant domain are highlighted to help the engineer understand the relevance of the article for the domain.

To resolve the resource bottleneck, natural language technologies are increasingly used at most of the above steps. A statistical classifier is used to determine the domain of each article (useful in Italy where legislation are sometimes issued with norms covering a range of unrelated domains). Usually, and particularly for well-populated domains, the classifier will select the correct domain for each article.

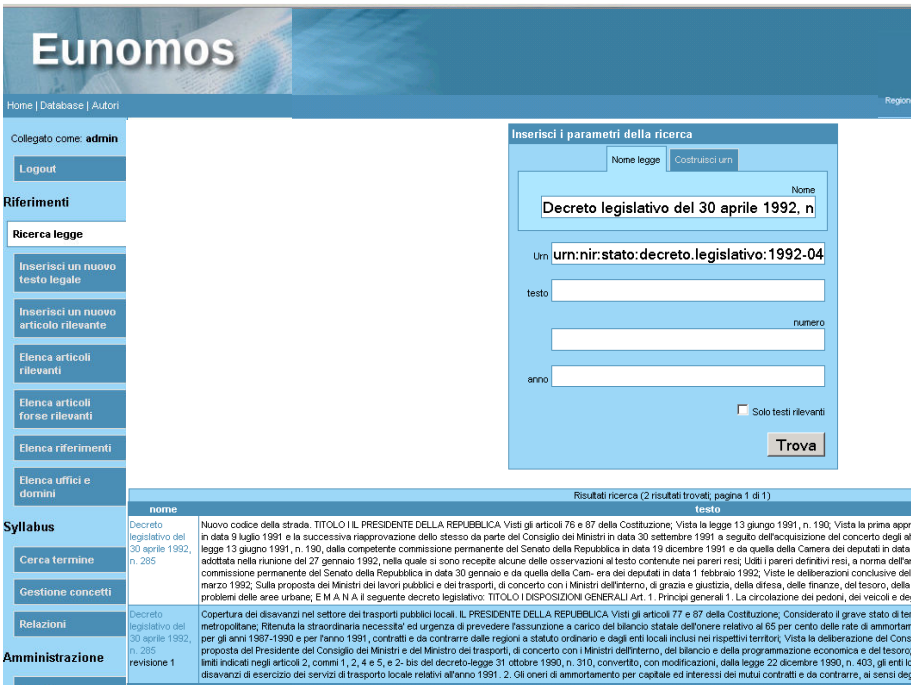


Figure 9: The search interface of Eunomos

However, due to overlaps in vocabulary and articles which contain no real content except cross-references, the knowledge engineer may need to resort to other supporting tools for this task: text similarity, prevalence of domain-specific terminology, and analysis of incoming and outgoing references.

Eunomos uses as a text similarity algorithm the Cosine Similarity to find the most similar pieces of legislation in the database. Since each piece of legislation contains a lot of text, they are indexed with the PostgreSQL internal inverted index facility in order to enable fast full text searches and ranking for document similarity. The Cosine Similarity metric uses the Term Frequency-Inverse Document Frequency (tf-idf) measure to gauge the relative weight to be apportioned to various key words in the respective documents. The Cosine Similarity metric is particularly useful for finding similar single-domain legislation. However, legislation that contains norms on different topics can introduce noise into the comparative process.

As such, Eunomos contains a classifier to identify which domain each article belongs to. This enables users to view, in each piece of new legislation, only articles relevant to a particular domain. Eunomos uses Support Vector Machines (SVM) for this task, since it frequently achieves state-of-the-art Accuracy levels [47, 101]. The association between legislation articles and a category label are fed to an external application based on the WEKA (Waikato Environment for Knowledge Analysis) toolkit [84] and incorporated into Eunomos, creating a model that can be used to classify new laws inserted on a daily basis

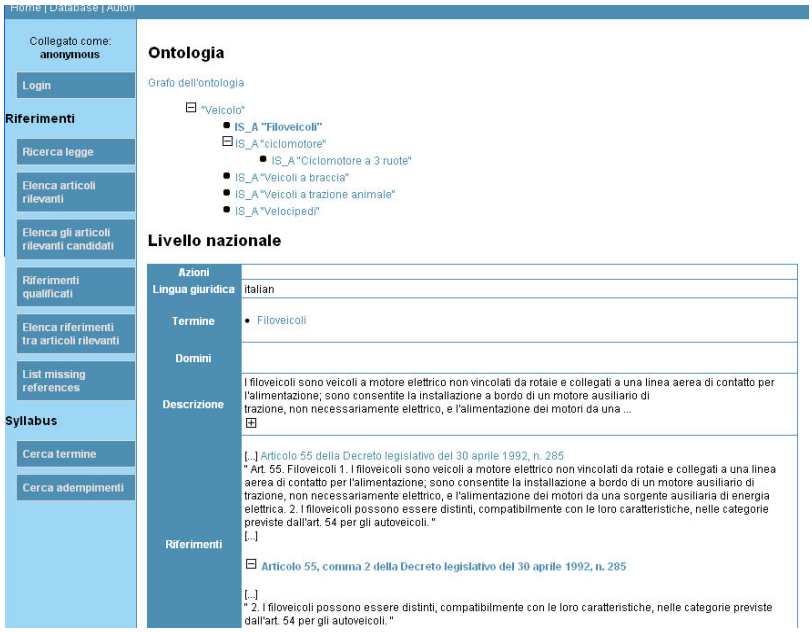


Figure 10: Legal Taxonomy Syllabus within the Eunomos system

into the database by web spiders or users. The process of transforming text into vectors requires selection of suitable terms, and use of a weighting function as part of the frequency calculations. Eunomos uses the TULE [117] dependency parser for Italian to select the informative units, i.e. lemmatised nouns. This works better than the more common practice of using WordNet [140] to eliminate stop-words and lemmatise informative units, since the latter is unable to recognise and lemmatise many legal domain-specific terms. The Eunomos ontology is also used to identify relevant phrases in the text of norms thereby adding further features to the classifier [28]. This improves the performance of the classifier, particularly for classes with fewer documents. The weighting function used is tf-idf, as proposed by Salton & Buckley [160], since it takes into account both the frequency of a term in a text and how characteristic it is of text belonging to a particular class.

4.3 LEGAL-URN

The final system of interest is LEGAL-URN [68], a sophisticated system for systematic modelling and analysis of compliance issues that enables business users to factor in legal requirements as part of their strategic planning. LEGAL-URN and Eunomos are complementary systems for compliance monitoring. Boella et al. [32] describe how both systems can be integrated to provide a comprehensive compliance management solution.

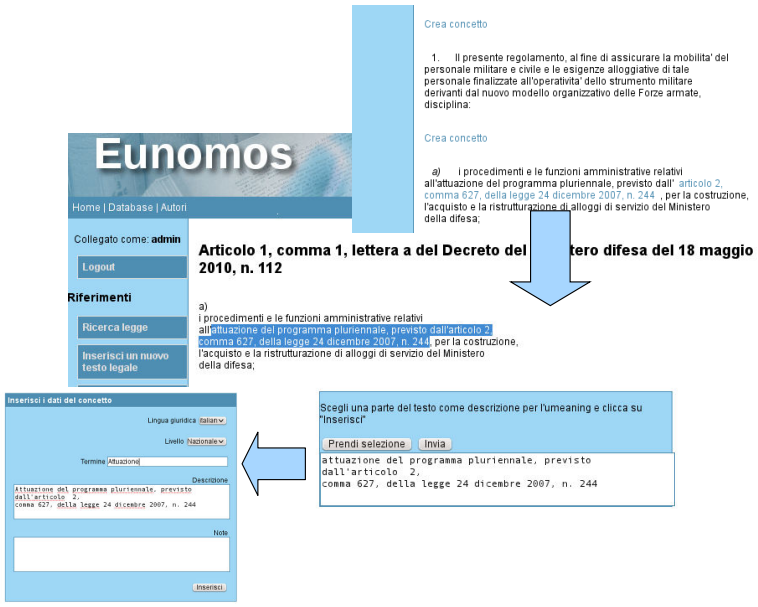


Figure 11: Creating concepts in Eunomos.

The LEGAL-URN framework helps model legal and organisational documents, analyse and manage compliance and the evolution of laws and business processes, identify non-compliant instances and prioritise non-compliant issues. The system is based on a Requirements Engineering (RE) approach to legal compliance, and models legal norms in the same notation as goal and business process management, albeit with deontic extensions. The unique characteristics of this framework are:

- using the same, standardised language for modelling business processes, goals and legislation;
- providing traceability between business processes, goals and legal documents;
- identifying instances of non-compliance through traceability links;
- managing the evolution of both law and organisational goals and processes via traceability links.

Requirements Engineering (RE) approaches to legal compliance focus on treating legal statements as a type of requirement. One approach in RE that has attracted much attention in the domain of legal compliance is goal-oriented modeling [36, 69, 173, 172]. In Goal-Oriented Requirements Engineering (GORE), goals are assigned to actors and range from high-level strategic objectives to low-level technical tasks. Goal models are used: to refine high-level goals into other goals and low-level operationalised tasks, to find alternatives

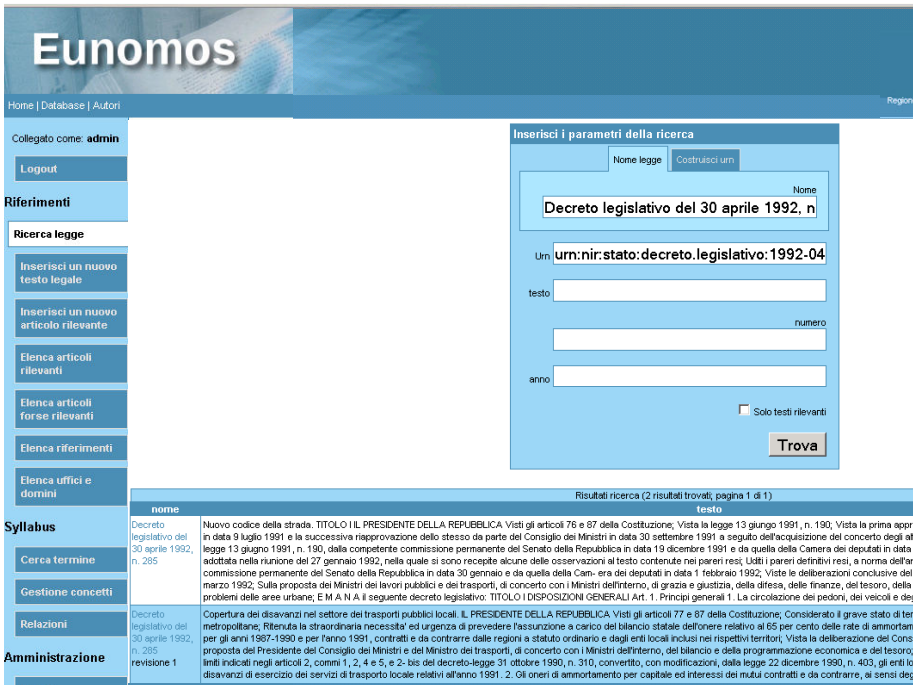


Figure 12: Eunomos system architecture

for achieving goals, and to assign tasks to actors [113]. Since goal-oriented modeling notations are used to capture goals that are high-level, generic and abstract, they are well-suited for modelling legal statements.

LEGAL-URN supports business process compliance by extending the model-based compliance framework based on the User Requirements Notation (URN) Language [98]. URN is the first and currently only standard (International Telecommunication Union (ITU-T Z.150 series)) that explicitly addresses goals and scenarios in a graphical way in one unified language. RE languages have not typically been developed to support the level of complexity that exists in the law. This is because requirements are usually intended to be written as single clear statements whereas laws are intended to cover all relevant circumstances with as few specifics as possible. The LEGAL-URN framework bridges the gap between the abstract level of legal prescriptions and the concrete level of business processes.

The LEGAL-URN framework has four layers of legal and organisational models, which are shown in the left-hand side of Figure 13:

1. **official source documents** that define the legislation on one side and organisational structures, policies and processes on the other side.

2. **a Hohfeldian model** (as expressed by Siena [171]) which consists of a set of Hohfeldian statements [192] of norms including structured elements of legal statements.

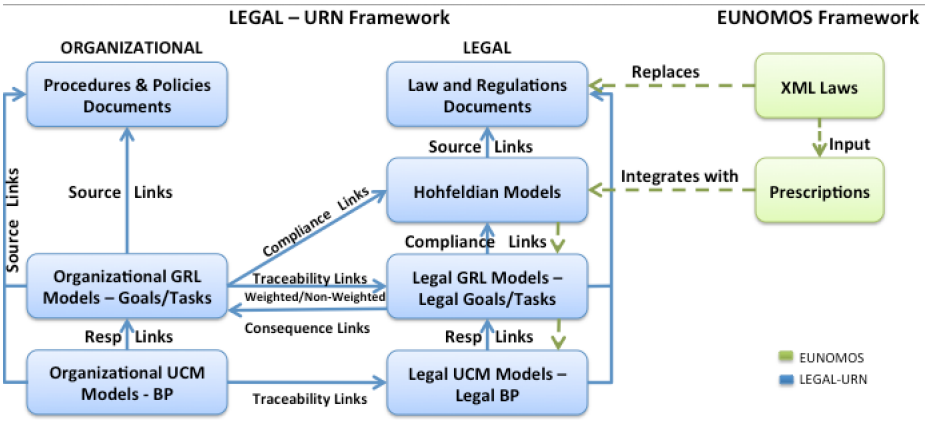


Figure 13: LEGAL-URN Framework Overview and Eunomos Integration

3. **goal models** based on URN's Goal-oriented Requirement Language (GRL), which capture the objectives and requirements of both organisation and legislation.

4. **business process models** based on URN's use case maps (UCMs), which define the business processes that implement organisational policies on the one hand and represents steps mandated by legislation on the other hand.

To build this framework, first, the relevant regulations, organizational policies and procedures are identified manually. This step is usually done by the legal expert in the organization. Next, the Hohfeldian model for the legal documents is created. For this, first, each legal statement in each legal document is annotated with one of the Hohfeldian correlative classes of rights: *duty-claim*, *privilege-no-claim*, *power-liability*, or *immunity-disability* (see Chapter 5 for a discussion of Hofheldian classes). Each legal statement must be atomic. This means that each legal statement contains one <actor> (the subject), one <modal verb>, one-to-many <clause> (<verb> & <action>), zero-to-many optional <crossreference>, zero-to-many optional <precondition> and zero-to-many optional <exception>. This layer provides the basis for goal modeling in Legal-GRL.

In the next step, the Hofeldian modalities (i.e. duty-claim, privilege-noclaim, etc.) are transformed into permissions and obligations for modelling in the Legal-GRL. Power-liability and immunity-disability statements are also of type permission and obligation with additional conditions and priorities. GRL's main concepts come from management and from socio-technical systems and include actors with intentional elements and indicators, linked through various relationships. GRL intentional elements are softgoals, goals, tasks, beliefs or resources. Softgoals differ from goals in that a goal is quantifiable (often in a binary way) and can be fully met while there is no clear, objective measure of satisfaction for a softgoal. GRL intentional elements are

HOHFELDIAN MODEL
SECTION
ARTICLE #
SUBJECT
MODAL VERB
CLAUSE
PRECONDITON
POSTCONDITION
EXCEPTION
XREF

a) Hohfeldian Structure

HOHFELDIAN MODEL	LEGAL-GRL MODEL
SECTION	-
ARTICLE #	-
SUBJECT	ACTOR, EXPECTIONACTOR
MODAL VERB	OBLIGATION, PERMISSION STEREOTYPE
CLAUSE	INTENTIONAL ELEMENT
PRECONDITON	PRECONDITION INTENTIONAL ELEMENT
POSTCONDITION	POSTCONDITION INENTIONAL ELEMENT
EXCEPTION	EXCEPTION INTENTIONAL ELEMENT
XREF	CROSSREFERENCE IE

b) Mapping between Hohfeldian Model and Legal-GRL Model

Figure 14: Hohfeldian Model Structure and Mapping with Legal-GRL in LEGAL-URN

connected to each other through decomposition, contribution, correlation or dependency links [40]. The GRL notation is extended to capture permission, obligation, precondition, exception and XRF goals/softgoals. The quantitative and qualitative analysis algorithms of URN's Goal-oriented Requirement Language are extended to help analyze an organisation's compliance to the legal models and prioritise which non-compliant instances to address first.

While the Goal-oriented Requirement Language (GRL) models the "why" aspect of requirements, Use Case Maps (UCM) model "what" aspects with scenarios (use cases). A scenario describes a specific path through the UCM model where only one alternative at any choice point is taken. Paths contain responsibilities (e.g. AskForData) which indicate where actions, activities, transformations, or processing is required. They can be performed in sequence, concurrently, or as alternatives. The UCM notation supports a simple but formal data model that can be used to formalise conditions at selection points. The benefit of using UCM over other business process modeling notations is that it has the ability to link its elements to GRL elements, since both views are part of URN. As such, tasks and actors in GRL can be linked to responsibilities and components in UCM maps. Having such business processes to represent legal clauses helps to capture the sequential aspects of laws and thus identify violations of procedural laws. The details of how to build the UCM models are provided by Weiss & Amyot [191].

LEGAL-URN has some limitations. The framework does not include a regulations repository. Furthermore, developing the Hohfeldian model is currently manual. It also lacks legal interpretations [27] to help identify sets of business process patterns which can be legally compliant. Integration with the Eunomos system described above solves these problems, as described by Boella et al. [32]. The Eunomos repository of laws - with legislative XML for clickable cross-references, definitions of terms and their inter-relationships in specialist ontologies - replaces the LEGAL-URN "Law and Regulation Documents" level. At the legal provisions level, there is a new representation that integrates Eunomos prescriptions and LEGAL-URN Hohfeldian models. Table 1

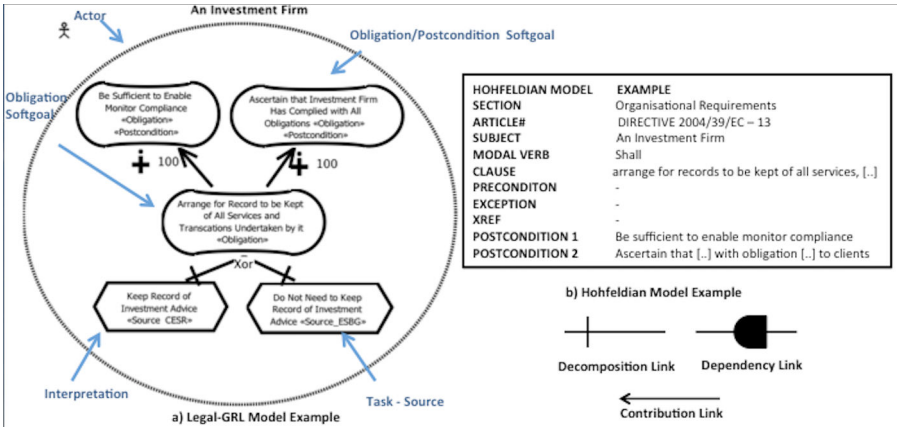


Figure 15: Hohfeldian Model Structure and Mapping with Legal-GRL in Legal-URN

shows the mapping of fields and relations between the two representations. Many fields can be mapped directly, some require adaptation, and others are taken from one representation.

The integrated solution classifies provisions according to Hohfeldian modalities rather than deontic logic because they allow a more refined characterisation of legal provisions with an explicit way to represent the hierarchy of norms. The active role, or subject, are essentially the same, and can be more clearly expressed as the responsible actor - who has the responsibility for ensuring the provision is fulfilled. This field is essential at the GRL or UCM level. The passive role here is renamed as beneficiary for clarification. Beneficiaries do not need to be represented at the next levels, unless they also have legal responsibilities that need to be modelled. The question of what is violated and what are the possible sanctions are important considerations in compliance decisions, and are represented at the legal provisions level. In Legal-URN, sanctions are modelled as "consequence" goals which have links from Legal-GRL to organisational models. The modal verb can provide useful clues for the knowledge engineer to classify legal provisions, but is not required as information about the provision in the final analysis. The description in Eunomos corresponds to the clause in the Hohfeldian model - simplifying the syntax and adding information from citations. The precondition from the Hohfeldian model is maintained as it is useful for describing applicability and sequential information. Postcondition is the correlative. The ontological relations from Eunomos - is-a, part-of and exception - are used to show the interaction between legal provisions. Clickable hyperlink norm identifiers are used instead of textual citations (section and article fields) to enable easy referencing to legal sources. The major innovation in the integrated solution is the addition of a stakeholder field which classifies the source of the legal provision as constitutional law, legislation, case law, subsidiary laws, ministerial decrees, legal

Table 1: Integration of Eunomos Prescriptions and LEGAL-URN Hohfeldian Models

Eunomos Prescriptions	LEGAL-URN Hohfeldian model	Integrated representation
Deontic clause	Hohfeldian modality	Hohfeldian modality
Active role	Subject	Responsible Actor
Passive role	-	Beneficiary
Violation	-	Violation
Sanction	-	Sanction
-	Modal verb	-
Description	Clause	Clause
-	Precondition	Precondition
-	-	Postcondition
Is-a relation	-	Is-a relation
Part-of relation	-	Part-of Relation
Exception relation	Exception	Exception relation
Norm identifier	Section + Article	Norm identifier
-	Cross-reference	Cross-reference
-	-	Stakeholder

scholars, self-regulatory bodies, industry bodies, internal regulator or external regulator. Different stakeholders have different levels of authority and/or persuasiveness in different jurisdictions and different domains, which is important to take into account in compliance decisions.

5 THE CONTENT OF LEGISLATION

This chapter describes the content of legislation: what kind of information is contained, what kind of information is useful to extract. A corpus of legislation on a variety of topics was analysed manually. Norm and norm element types were identified, using as a starting point the Eunomos prescriptions ontology, with a view to extending its coverage with well-known concepts from legal theory to reflect the realities of the corpus. At this stage, common problems for NLP processing were also identified which informed the normalisation module (see Chapter 6).

EU directives and regulations are structured in the following way: preamble, followed by enacting terms, followed by appendixes/annexes. The preamble contains citations (to treaties that legitimize the legislation), solemn procedural or principle-based expressions effectuating the law [7], and a numbered list of recitals. The "enacting terms" are the legislative part of the act. These normative (or operative) provisions are composed of articles, which may be grouped into parts, titles, chapters and sections. Enacting terms usually contain definitions followed by norms and meta-norms. Appendixes contain further information of a varied nature (tables, forms, explanations). Below we discuss the usefuleness of the different types of information contained in legislation for the purpose of compliance. The discussion is mainly centred on EU legislation, and some comparisons are also made with UK legislation.

Section 5.1 describes proclamations. Section 5.2 describes recitals and their relevance to normative provisions. Sections 5.3 and 5.4 describe constitutive and deontic norms respectively from a theoretical and observational perspective. Section 5.5 describes meta-norms and other legal statements less commonly discussed in legal theory. Section 5.6 describes elements of norms that should be reflected in structured legal ontologies.

5.1 PROCLAMATIONS

Legislation usually start and end with proclamations, or procedural norms. Such proclamations tend to follow exactly the same wording. Here are the main ones in EU legislation - the items in square brackets are generalised concepts, otherwise, the wording is as found in the legislation:

```
THE EUROPEAN PARLIAMENT AND THE COUNCIL
OF THE EUROPEAN UNION,
```

```
Having regard to [various treaties]
Acting in accordance with [a certain procedure]
[list of recitals]
HAVE ADOPTED THIS DIRECTIVE
[list of normative provisions]
This Directive shall enter into force on [date]
This Directive is addressed to [addressees]
Done at [location] [date]
[signatures of presidents of EU bodies]
```

Such sentences are classed by Sartor [163, p. 127]as result-declarations i.e. "statements of a legal outcome, which is intended to produce that very outcome". Result-declarations are not confined to legislation. Other examples are the proclamations of marriage by a priest or mayor, raising one's hand in an auction, or any other contractual offer which upon acceptance obliges the offeree to carry through. The procedural result-declarations that are required to make the legislation valid are beyond the scope of this thesis, concerned as it is with obtaining only the information required by addressees to ensure compliance with the law.

5.2 RECITALS

Approximately half of the text of some EU legislation are recitals. Recitals are compulsory in EU directives (legislation to be implemented by member states as they see fit) and optional in EU Regulations (legislation that have direct application in member states). Recitals contain objectives, references to other relevant legislation and occasionally definitions, but mainly consist of a principle or justification, followed by a concise norm-like element. The phrase norm-like element is used here because despite their normative language, recitals do not have the normative status of the enacting terms. Which begs the question: what exactly are recitals and what are they for?

Directives and regulations implement EU treaties, also called primary EU law. Treaties "are binding agreements between EU member countries. They set out EU objectives, rules for EU institutions, how decisions are made and the relationship between the EU and its member countries". They contain very few concrete rules and often general notions, which are then expanded upon and made more concrete in secondary legislation [115, page 13]. For example, the Treaty on the Functioning of the European Union (TFEU) (amended by the Lisbon Treaty in 2009) encompasses topics such as consumer protection, competition, tax, etc.). The EU air transport policy is provided in Article 100 of the TFEU. This has led to EU legislation on air transport (covering aspects such as airspace management, safety and security standards, passenger rights, environmental matters, etc.) such as EU Regulation 261/2004 establishing common rules on compensation

and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights (a piece of legislation analysed in Chapter 9). Many recitals refer to specific articles in treaties to justify the normative provisions; others explain the motivation behind the normative provisions or even a summary. This is why recitals are fundamental to legal reasoning in the EU.

According to the Guide [7, clause 10], a recital's purpose is "(...) to set out concise reasons for the chief provisions of the enacting terms, without reproducing nor paraphrasing them" in order "to give the parties concerned in a possible dispute the opportunity of defending their interests, to enable the Court to exercise its review jurisdiction and to allow MS and any national of these States who so wishes, to ascertain how the enacting institution has applied the Treaty" [7, clause 10.2]. Moreover, recitals "are of particular importance in order for the ECJ to assess whether the Community legislator has not made manifest errors in areas where it enjoys a margin of appraisal" [3, clause 4.1.4(c)][1]. Indeed, "if the reasoning set out in an act for which a statement of the reasons is compulsory is wanting or is not sufficient to fulfill the requirements (...), the Court can annul the regulation for breach of essential procedural requirements" [3, clause 4.1.4(a)].

Concerning the legal importance of recitals [3, clause 4.1.4], there are different doctrinal positions [109] on the relationship between recitals and normative provisions:

1. recitals have no effect;
2. recitals are dominant over normative provisions;
3. recitals have an equal position in relation to normative provisions;
4. recitals encompass a subordinate position towards normative provisions.

The ECJ has assumed both positions 3 and 4 in its judicature.

Supporting position 3, the ECJ has stated that recitals are used to "interpret the enabling provision of an act" [3, clause 4.1.4(b)], and that recitals are "necessary for courts to perform supervision"[2]. It is worth citing the following doctrinal interpretation [109]:

> [T]he law of recitals in EC Legislation can be summarized thusly: A) Where both the recitals and the operative [normative] provisions are clear but inconsistent, the operative provision will control. Corollary: recitals have no positive

1 Recital 9 of the Data Protection Directive provides an illustrative example of this margin of appraisal: "(...) whereas Member States will be left a margin for manoeuvre, which may, in the context of implementation of the Directive, also be exercised by the business and social partners(...)".

2 Case 24/62, F.R.G. v. Commission of the Eur. Econ. Cmty., 1963 E.C.R., paragraph 18.

> operation of their own. B) Where the recital is clear, it will control an ambiguous operative provision. This means that the operative provision will be interpreted in light of the recital. There have been cases wherein the nature of the operative provision is affected by a recital, and others where the scope of the operative provision is affected.

Substantiating position 4, the ECJ has ruled that the recitals "cannot be relied on as a ground for derogating provisions of the act"[3]. It is moreover stated that "if a recital is irredeemably inconsistent with the operative text, then the ECJ will ignore the recital and give effect to the text of the operative provisions" [132]. Recitals can be used to interpret only provisions which are ambiguous[4], but "they cannot, however, restrict an unambiguous provision's scope" [109, page 3], i.e. "the terms of a recital cannot be used to give a particular construction to a provision which the terms of that provision would not otherwise bear".[5]

In practice, recitals have been used exactly in this way. For instance, article 5 of Regulation EC 261/2004[6], headed "Cancellation", provides that an operating air carrier shall not be obliged to pay compensation if it can prove that the cancellation is caused by extraordinary circumstances which could not have been avoided even if all reasonable measures had been taken. The term "extraordinary circumstances" is not defined in any of the articles of the Regulation (not even in Article 2 which is devoted to "definitions"). However, recitals 14 and 15 of the Regulation give a few examples, by way of illustration, of events which may be regarded as extraordinary circumstances, namely cases of political instability, meteorological conditions incompatible with the operation of the flight concerned, security risks, unexpected flight safety shortcomings and strikes which affect the operation of an operating air carrier. These cases have been used by the Court to determine to which extent the air carrier is exempted from paying compensation [2].

It is suggested that recitals are an essential component in EU legal interpretation (as are Explanatory Notes in UK legislation). It would be useful if links between recitals and relevant normative provisions were made explicit where possible. If EU legal drafters explicitly linked normative provisions with recitals and made this information available

3 Case C-162/97, Nilsson et al., paragraph 54, 1998, E.C.R. I-07477; and Case C-344/04, IATA, ELFAA v Department for Transport, # 76 (specifically addressed to the air transport passenger domain).

4 Case C-244/95, P. Moskof AE v. Ethnikos Organismos Kapnou, 1997 E.C.R. I-06441.

5 Case C-412/93, Sociee d'Importation Edouard Leclerc-Siplec v TF1 Publicite SA and M6 Publicite.

6 Regulation 261/2004/EC of the European Parliament and of the Council of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, OJ L 46 of 17.2.2004

to the public, this would help render EU legislation more accessible and more certain. In the absence of any authoritative connections forthcoming from EU institutions, connections will be made by those who have to interpret the law (judges, lawyers, advisory bodies etc.). Legal knowledge-based systems could support linking normative provisions to related recitals just as they can support linking to case law and legal doctrine so that this 'hidden' knowledge becomes more widely available.

5.3 CONSTITUTIVE NORMS

When we think of norms, we usually think of deontic norms that tell us what we may or must do and when. However, such norms rely on clarification as to the meaning of ordinary and legal concepts. In EU legislation, the normative provisions sections usually start with definitions. The official definitions of legal concepts in legislation have the force of law, and are as such called constitutive norms.

Any fact that fits within the official description of a certain legal concept 'counts as' that concept for the purpose of the law. Legal objects, like all social objects, assume a supra-physical reality due to collective recognition and acceptance, which Searle [169, page 18] generalised as: 'X counts as Y in context C'. Searle[page 18] [169] provides as examples of non-legal social objects "such and such counts as a $20 bill in our society. George W. Bush counts as President of the United States. Such and such a move in chess counts as a legal knight move." Lindahl & Odelstad [122] provide as examples of legal social objects "property, tort, contract, trust, possession, guardianship, matrimony, citizenship, crime, responsibility, punishment".

In what way can definitions be regarded as norms? Searle [168, page 33] explains: "[R]egulative rules regulate antecedently or independently existing forms of behavior [...]. But constitutive rules do not merely regulate, they create or define new forms of behavior." Lindahl and Odelstad [122] state that intermediate legal concepts serve as links between facts on the ground (events, circumstances, actions) and normative consequences. Moreover, Grossi & Jones [80, page 411] state that "the way constitutive norms define new forms of actions or new states-of-affairs is by relating them to something already existing or established." The linking between "brute facts" and legal concepts takes place at different times by different actors: by legislators when legislation is drafted, by individuals or organisations when regulating their affairs in accordance with the law, and by law enforcers and lawyers in the case of a legal dispute.

Due to the challenge of linking the particular to the general (particularly when considering unforseen events), legislative constitutive norms are usually general descriptions, but also not uncommon are definitions by example, which allow extension by analogy, as well

as definitions that explicitly include or exclude certain items from counting as the legal concept in question. Such definitions are often outside the Definitions section, but are nevertheless influential. Here is one example from Directive 2009/12/EC[7]:

```
[Additional regulatory measures] may include economic
oversight measures, such as the approval of charging
systems and/or the level of charges, including
incentive-based charging methods or price cap  regulation.
```

Such examples require re-evaluation of the elements of definitions for the purpose of term-based ontologies as they do not fit easily into the definiendum-definiens model. For example the above example could be expressed in XML as follows:-

```
<Norm>
   <NormType> Definition </NormType>
   <Definiendum> Additional regulatory measures
   </Definiendum>
   <Includes> economic oversight measures </Includes>
</Norm>
<Norm>
   <NormType> Definition </NormType>
   <Definiendum> economic oversight measures </Definiendum>
   <Example> approval of charging systems </Example>
</Norm>
<Norm>
   <NormType> Definition </NormType>
   <Definiendum> economic oversight measures </Definiendum>
   <Example> the level of charges </Example>
</Norm>
<Norm>
   <NormType> Definition </NormType>
   <Definiendum> the level of charges </Definiendum>
   <Includes> incentive-based charging methods </Includes>
</Norm>
<Norm>
   <NormType> Definition </NormType>
   <Definiendum> the level of charges </Definiendum>
   <Includes> price cap regulation </Includes>
</Norm>
```

Most definitions apply to any instance of the term in the legislation under consideration. However, the scope can be even more local, as this definition shows from the UK legislation 'The Airport Charges Regulations 2011':

7 Directive 2009/12/EC of the European Parliament and of the Council of 11 March 2009 on airport charges (Text with EEA relevance.)

```
For the purposes of this Schedule, '' Member'' means a
member of the CAA appointed by the Secretary of State
under section 2(17) of the Civil Aviation Act 1982.
```

5.4 DEONTIC NORMS

Deontic norms are used here to mean generalised types of behaviour expected or permitted of addressees issued by a sovereign authority. The norms must come from a sovereign authority, otherwise they are not norms but norm-statements i.e. descriptions of a norms, according to Hilpinen & McNamara [91, page 26], citing von Wright [188, viii. page 105]. Unlike the laws of nature, there is nothing inevitable about deontic norms, as the addressees have a degree of autonomy to obey or disobey, particularly since the norms may or may not align with the addressees' interests. As such, sovereign authorities have to rely on the 'force of law' (policing and adequate punishment) to ensure compliance.

The simplest and most intuitive classifications of deontic norms are those of Leibniz's deontic categories: the permitted is "what is possible for a good person to do", and the obligatory is "what is necessary for a good person to do", according to Hilpinen & McNamara [91, page 26], citing Hruschka et al. [94, pages 35–36]. These categories (obligatory and permitted) were explored in a logic framework by von Wright, with the addition of a new category, prohibition. He defined them as follows: "obligatory (that which we ought to do), the permitted (that which we are allowed to do), and the forbidden (that which we must not do)". This recalls the work of fourteenth century philosophers [91, p. 6], who defined the interrelationship between the three categories as follows[8]:

- $\textit{Permitted}\ A \equiv \textit{NOT Obligatory NOT}\ A$
- $\textit{Obligatory}\ A \equiv \textit{NOT Permitted NOT}\ A$
- $\textit{Obligatory}\ A \equiv \textit{NOT Prohibited NOT}\ A$
- $\textit{Prohibited}\ A \equiv \textit{Obligatory NOT}\ A$

Von Wright's seminal work laid the groundwork for Standard Deontic Logic (SDL), which now includes omissible and optional, although these are less used than the other categories. They are defined as follows:

- $\textit{Omitted}\ A \equiv\ \textit{NOT Obligatory}\ A$
- $\textit{Optional}\ A \equiv (\textit{NOT Obligatory}\ \&\ \textit{NOT Obligatory NOT}\ A)$

8 *equiv* means material equivalence.

In common parlance, permission implies that the addressee has a choice whether to do something or not i.e. that permission is equal to optional in the above definition. However, Sartor [163, page 5] argues that this is not necessarily the case, that "when we only know that an action is permitted, we do not know the status of its complement" , since the categories obligatory and prohibited entail permission. He gives as an example the action of wearing a veil, obligatory in Iran, prohibited in France and permitted in the UK. It is permitted in Iran to wear the veil, it is permitted in France to not wear the veil, it is only in the UK where it is permitted both to wear and not to wear the veil. A similar point is made by Boella & van der Torre [29, page 3]: "[I]t is forbidden to have guns, but it is permitted for policemen to have guns. Is it permitted to policemen not to have a gun? It depends on other norms." Another important point is that traditional deontic logic does not distinguish between weak and strong permissions i.e. between that which is not prohibited, and that which is explicitly permitted by law. Ross [156, page 122] observes: "I know of no permissive legal rule which is not logically an exemption modifying some prohibition, and interpretable as the negation of an obligation". However, Boella and van der Torre [29] argue that in dynamic settings, a redundant permission may be introduced by a higher authority to effectively prevent lower authorities from introducing a prohibition.

The above conceptualisation of norms is essentially of a powerful state commanding individuals. An equally influential conceptualisation is that of Hohfeld [93], who considered norms from the point of view of individuals with different interests, and the role of the state is to safeguard one individual's legitimate rights against another individual (here, individual can also mean companies or institutions). For Hofheld, right is the correlative of duty or obligation: "if X has a right against Y that he shall stay off the former's land, the correlative (and equivalent) is that Y is under a duty toward X to stay off the place." Hofheld had no separate category for prohibition, regarding it simply as the obligation to not do something. Privilege is the negation of duty (or the permission to not do something): "whereas X has a right or claim that Y, the other man, should stay off the land, he himself has the privilege of entering on the land; or, in equivalent words, X does not have a duty to stay off." No-right is its correlative: in the above example, Y has no right to claim that X should stay off the land.

The above four categories are what Sartor [163] calls "obligative statements". But Hofheld defines a further four categories, "potestative statements". Power is the ability to change legal relations. For example, a land-owner has the power to sell his/her land, thereby relinquishing his/her legal interests and enabling another to use and sell the property [93, page 18]. Another example of power is the power to discharge a debt. In the corpus of EU legislation, we find the exercise of powers to delegate the creation of obligations to national

parliaments and beyond i.e. the power to grant power. This is an example from Directive 2009/12/EC:

```
Member States may allow the airport managing body of
an airport network to introduce a common and
transparent airport charging system to cover the
airport network.
```

Such a norm may be expressed structurally (here in XML) as follows:

```
<Norm>
   <NormType> Power </NormType>
   <ActiveRole> Member States </ActiveRole>
   <PassiveRole> the airport managing body of an airport
   network </PassiveRole>
   <Action> allow the introduction of a common and
   transparent airport charging system to cover the airport
   network </Action>
</Norm>
<Norm>
   <NormType> Power </NormType>
   <ActiveRole> the airport managing body of an airport
   network </ActiveRole>
   <Action> introduce a common and transparent airport
   charging system to cover the airport network </Action>
   <Condition> subject to permission by Member State
   </Condition>
</Norm>
```

The correlative of power is liability. A simple example is when a person is liable to serve on juries, that liability that comes into play from adulthood and becomes an obligation when the courts exercise their power to summon a person to serve. Power and liability are also at play in the doctrine of offer and acceptance in Common Law countries. If A mails a letter to B offering to sell land for a specific sum, A is liable (but not yet obliged) to sell to B for a reasonable time until B accepts the offer. However, if A revokes his/her offer before B accepts, his/her liability ceases along with B's power. Disability is when one has no legitimate power to oblige another. Unless B has received an offer, B is under a disability to oblige A to sell the land to him/her, and A has immunity from any such claims. Another example of immunity is diplomatic immunity where host nations are disabled from prosecuting diplomats who violate national laws in the normal way (their only recourse is to expel the diplomats).

In a slightly modified notation of Sartor's [163], each Hohfeldian category can be defined in terms of deontic categories directed towards agents (j) and patients (k):

- $Right\ k\ (Does\ j\ A) \equiv Obligation\ j\ (Does\ j\ A)\ For\ k$

- *NoRight k (Does j A)*
  ≡ *Privilege j (Does j A) For k*
  ≡ *Permission j NOT (Does j A) For k*
  ≡ *NO Obligation j (Does j A)*

- *Power k Towards j* ≡ *Liability j Towards k*
  ≡ *Permission k(Bestow Obligation / Right / Privilege / NoRight / Power / Liability / Disability / Immunity Towards j)*

- *Disability k Towards j*
  ≡ *Immunity j Towards k*
  ≡ *NO Permission k(Bestow Obligation / Right / Privilege / NoRight / Power / Liability / Disability / ImmunityTowards j)*

Do we need all these categories? Hofheld motivated his essay by saying that '[o]ne of the greatest hindrances to the clear understanding, the incisive statement, and the true solution of legal problems frequently arises from the express or tacit assumption that all legal relations may be reduced to "rights" and "duties," and that these latter categories are therefore adequate for the purpose of analyzing even the most complex legal interests, such as trusts, options, escrows, "future" interests, corporate interests, etc.' [93, page 8]. Husik counter-argued that a concept "is fundamental and requires a specific technical term only if it can not be expressed completely in the terms we already have. Otherwise we are merely encumbering our nomenclature without improving our insight into the conceptual bases of law." [96, pages 266–267] He added: 'the word "duty" does not really add anything substantial, it merely views the right from another angle.' [96, page 264]. Moreover, power can be regarded merely as a specific kind of right: "[t]he specific character of a power is that it denotes primarily, as used by Hohfeld, the ability to control legal relations, and through these to control acts, whereas rights which are not also powers denote directly the control of acts" [96, page 264]. The concept of privilege is also attacked: '[F]or me to have a privilege of doing a thing, means as mentioned before, (i) to have no duty of doing the thing, (2) to have no claim or right against others that they should refrain from interfering with my doing the thing, and (3) to be under no duty not to do the thing. In other words, the relation contained in the term "privilege" is completely expressed, by using the terms "right" and "duty." ' [96, page 267]. Whilst this analysis may be correct, that does not mean that the Hohfeldian concepts are not useful. Perhaps the only fundamental concept is obligation, or its correlate right. If the other 'fundamental' concepts are common, surely it is useful to retain them rather than having to define each one in terms of obligation every time it arises.

For our purposes, the critical question is what kind of legal statements are commonly found in legislation. Manual analysis was con-

ducted on Directive 2009/12/EC[9] and its UK implementation to get some idea of the most dominant categories. For this analysis, to simplify, one category was assigned for each sentence. As the list below shows, Directive 2009/12/EC is dominated by obligations.

- obligation: 46 (including 1 prohibition)
- definition: 13
- power: 12
- hierarchy: 5
- proclamation: 3
- permission: 2
- scope: 2
- exception: 1
- legal effect: 1

The same is true of the UK implementation of Directive 2009/12/EC, namely 'The Airport Charges Regulations 2011':

- obligation: 83 (including 3 prohibitions)
- definition: 30
- power: 17
- hierarchy: 2
- proclamation: 11
- permission: 4
- scope: 8
- exception: 5
- legal effect: 6
- sanction: 2
- amendment: 36
- condition: 3
- timeframe: 4

Only a small subset of the Hohfeldian categories are commonly found in legislation. Moreover, and notwithstanding the amendments and proclamations which are beyond the scope of this thesis, there are also a fair amount of meta-norms (norms about norms) and other legal statements. These are discussed in section 5.5.

5.5 META-NORMS AND OTHER LEGAL STATEMENTS

The normative provisions of legislation contain many statements which are not norms in the traditional sense. Omitted from this discussion are amendments, which are already well covered by Lesmo et al. [118] and Robaldo et al. [155], and in any case, once the amendments are made, we are left with norms of the usual variety. Also omitted are ceremonial proclamations, which are mandatory for creating legitimate laws, but have no bearing on the content of the norms which are of interest to its addressees.

9 Directive 2009/12/EC of the European Parliament and of the Council of 11 March 2009 on airport charges (Text with EEA relevance).

Most legal norms accordiong to von Wright [189] are behaviour rather than results oriented: they involve a normative determination governing some action on the part of the addressee (*tun-sollen*) rather than the achievement of a desirable state-of-affairs (*sein-sollen*). The action under consideration is subject to some kind of pre-condition. For instance, Sartor [163] states that the norm below:

```
Anyone below 18 years of age is forbidden to buy alcoholic
drinks.
```

can be expressed as an if-then normative conditional:

```
For any x,
IF x is below 18 years of age,
THEN it is forbidden that x buys alcoholic drinks
```

Usually, a condition is part of the same sentence as the norm, but occasionally a condition is found in a separate sentence. An exception can be regarded as a negative condition which, if satisfied, means the norm does not come into play. Exceptions are usually in a separate sentence. In a legal ontology, the exceptions and conditions need to be linked to the relevant norm or incorporated inside the structure of the norm.

There are certain legal statements that state that when a condition is satisfied, a certain legal effect takes place without any particular action on the part of any addressee. Here are a couple of examples[10]:

```
Where the professional body fails to take a decision and
notify the registered European lawyer within four months
in accordance with paragraph (1), it shall be deemed to
have taken a decision to reject his application and to
have notified it to him on the last day of that period.
```

```
Although it is not a prerequisite for the decision of the
competent authority in the host Member State, the
temporary or permanent withdrawal by the competent
authority in the home Member State of the authorisation
to practise the profession shall automatically lead to the
lawyer concerned being temporarily or permanently prohibited
from practising under his home-country professional title
in the host Member State.
```

It is difficult to find an accurate description of this type of legal statement in the literature. There is some similarity to generic power as defined by Sartor [72, page 578]: "We say that there is the generic power

10 from 'The European Communities (Lawyer's Practice) Regulations 2000' and 'Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained'.

to achieve B via A, and write GenericPower (B VIA A), whenever B is normatively determined by A" (as opposed to action power when the antecedent of a normative connection is formed by the action of an agent or enabling power, where the agent has the intention of bringing about the consequence). In fact, the notion of generic power is very different to the Hohfeldian notion of power (which corresponds to enabling power), and in a sense does not appear to be a real power at all. Sartor [72, page 581] concedes that "[w]e tend to use the notion of a power only to cover those cases where the law, by linking a certain result to one's action, aims at enabling one to achieve that results". Alternatively, the examples above could be viewed as a kind of count-as statement: *'X counts as Y in context C'* [169]. However, this suggests a static state of affairs, which is not appropriate. The key characteristic of this type of statement is that a certain legal condition leads to a certain legal effect. For want of a better term, therefore, such statements are called legal effects in this thesis. They may be represented structurally (here in XML) as follows:

```
<Norm>
   <NormType> Legal Effect </NormType>
   <Situation> the professional body fails to take a
   decision and notify the registered European lawyer
   within four months in accordance with paragraph (1)
   </Situation>
   <Result> the professional body shall be deemed to have
   taken a decision to reject the registered European lawyer's
   application and to have notified it to him on the last day
   of that period </Result>
</Norm>
<Norm>
   <NormType> Legal Effect </NormType>
   <Situation> the temporary or permanent withdrawal by the
   competent authority in the home Member State of the
   authorisation to practise the profession </Situation>
   <Result> the lawyer concerned being temporarily or
   permanently prohibited from practising under his
   home-country professional title in the host Member
   State </Result>
</Norm>
```

Another important type of legal statement is about the scope or applicability of norms. For instance:

```
This Directive shall apply both to lawyers practising in a
self-employed capacity and to lawyers practising in a
salarial capacity in the home Member State and, subject to
Article 8, in the host Member State.
```

Such statements often contain a precondition as to when norms are applicable:

```
The first subparagraph shall apply mutatis mutandis where
disciplinary proceedings are initiated by the competent
authority of the home Member State, which shall inform the
competent authority of the host Member State(s) accordingly.
```

Such statements are called scope in this thesis. Both the above examples come from Directive 98/5/EC.

The final meta-norm encountered is that governing the relative hierarchy among norms. While no legal system can be entirely free of conflicts as there are unprecedented scenarios, we do find in legislation attempts to resolve the most obvious conflicts. Wordings such as 'subject to' and 'without prejudice' are common in such statements to indicate lower priority. For example (from Directive 2009/12/EC):

```
The decisions of the independent supervisory authority shall
have binding effect, without prejudice to parliamentary or
judicial review, as applicable in the Member States.
```

5.6 THE ELEMENTS OF NORMS

In order to fully understand a norm, it helps to break it down into its constituent elements. Jørgensen [104] states that a norm has two parts - the 'imperative factor' which identifies the deontic element of the norm, and the 'indicative factor' which identifies the content of the norm. As previously mentioned, the content of the norm is usually an action of some kind, but there are also other elements - condition, exception, whom the norm applies to, sanction. It is rare to find all these elements mentioned together in one sentence, but some will be present, and then legal expertise is required to make connections and fill in the gaps where possible.

6 NORMALISING LEGISLATIVE TEXT

Chapter 3 discusses the challenges of legal text from a semantic and structural point of view. The semantic challenges (vagueness, 'terms of art' etc.) are complex and are best handled with expert annotation. The Eunomos system (see Chapter 4) enables such annotations to be provided in a well-organised and user-friendly way. Structural textual challenges, on the other hand, can be realistically addressed with automated processes.

There are two ways to bridge the differences between legislative text and standard written text. One is creating bespoke components such as part-of-speech taggers, parsers, and so forth. The other approach is to avoid building a dedicated system for extraction tailored to legal language structural complexities by normalising legal text and using standard NLP tools. This chapter is based on the assumption that the second approach should be the most satisfactory because application-independent standards and notations are more sustainable in the long term, and components can be upgraded in due course.

Section 6.1 of this chapter describes the scope of the normalisation module, section 8.1 the methodology, and section 6.3 the output.

6.1 SCOPE

The normalisation module is limited to making legislative text akin to standard written text, so that information can be extracted to the norm ontology using the Mate Tools Semantic Role Labeler [25]. The normalisation module uses pattern-matching and the Python Natural Language Toolkit (NLTK) Brill part-of-speech tagger [37, 24] to perform the following tasks:

- sentence and word segmentation (with identification of abbreviations, URLs, reference numbers, and dates)
- identify titles (uppercase or lacking verb)
- identify and transform references into processable units
- transform certain words to enable SRL handling
- identify lists (and their nestedness) and transform the lists into proper grammatical sentences, by adding introductory clauses and endings to each list item

An important problem that is not addressed in this work is the extraction of data via references. Chapter 3 discusses the challenges

involved in extracting relevant information from articles referenced. This is a major area of work that would be interesting to research as future work. However, for this thesis, structural titles such as 'Article 2' are transformed into the more manageable format S_A2 and serve as an identifier for extracted norms that may be reconstructed as an Uniqure Reference Number to facilitate linking with norms in legislative XML. In the text of norms, references are simply joined together e.g. 'S_Article_2' to avoid parsing errors, particularly useful for legislation with long names, and easily reconstructable to the original form when required.

Another challenge that is not addressed in this work is inter- and intra-sentential anaphora resolution, as this also a major topic in itself which would be interesting to research properly in the context of legislation. It should be noted that for the information extraction work undertaken for this thesis, not all anaphora need to be resolved e.g. if the element referred to is inside the same norm element.

In the norm ontology entry below from Directive 98/5/EC[1] based on information extraction using Semantic Role Labeling (see Chapter 8), the field for passive role is incomplete and the reference to 'that State' is only understandable when reading the original sentence in full. However, the anaphora 'his' in the field for active role does not need to be resolved as the entity it refers to is mentioned in the same field:

```
NORM ONTOLOGY ENTRY:
<Norm>
  <NormType> Obligation </NormType>
  <ActiveRole>
        A lawyer who wishes to practise in a Member State
        other than that in which he obtained his professional
        qualification
  </ActiveRole>
  <Action> register </Action>
  <PassiveRole>
       the competent authority in that State
  </PassiveRole>
</Norm>

ORIGINAL SENTENCE:
A lawyer who wishes to practise in a Member State other
than that in which he obtained his professional
qualification shall register with the competent authority
in that State.
```

The corpus used for information extraction (and therefore the pre-processing module) is restricted to normative provisions - preambles,

1 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained

annexes and appendices are excluded. The preamble, while often containing norm-like elements, are not proper normative provisions (see Chapter 9). Annexes generally contain further or procedural information and are also not considered normative provisions. Furthermore, they often contain tables, which are more unpredictable and difficult to process in an uniform way. Tables are rarer in normative provisions (there is one in the corpus) and they are not handled in this research. Titles are also excluded and are identified by automated grammatical analysis.

6.2 METHODOLOGY

6.2.0.1 *Preliminary Sentence Segmentation*

The input files are text files of legislation. The file is read line by line, and each line is stored in a Python list. Then preliminary sentence segmentation is performed resulting in another Python list. During this phase:

- non-standard apostrophes and quotation marks are replaced with standard ones
- list indicators (bullet points, dashes, numbers, letters or roman numerals) are stored as separate 'sentences'
- colons, semi-colons and full-stops are treated as end-of-sentence indicators
- dots in known abbreviations (based on analysis of the corpus) are exempted from being treated as full-stops. Also exempted are dots within URLs, reference numbers, and dates (via pattern-matching).
- text within quotations are treated as one sentence whether or not they actually contain more than one or span over more than one line. This is because quotations in legislation are usually used reserved for legislative amendments and are not used for definition or norm extraction for our purposes.
- lines that do not end with an end-of-sentence indicators are joined with the next line and the segmentation is made at the next end-of-sentence indicator.

6.2.0.2 *Identifying Sentence Types*

After the preliminary sentence segmentation begins the serious work of identifying lists and rearranging them so that they read as proper sentences (as described above). This work is complicated by the fact that many legislation contain nested lists - sometimes up to three levels of nestedness. In this phase, each 'sentence' from the preliminary

sentence segmentation phase is analysed and classified as a certain 'type' and handled accordingly. The rules for identifying each type are prioritised as follows (high to low):

- title: the identification of titles is based upon part-of-speech i.e. they do not contain modal or other titles - a hypothesis that works well most of the time. Such sentences are removed from consideration in the information extraction phase.
- sentence before list: this is determined by analysing the characteristics of the sentence that comes afterwards. Negative indicators are that the 'sentence' that follows is a title, table, graphic, quote, a list item that is part of an unfinished list, or a list item that is part of an outer list. Otherwise, a positive indicator is that the sentence that follows is an index.
- introductory clauses: sentences before lists can be 'common beginnings' to be appended to list items to form proper sentences if they conform to certain pattern-matching rules. Positive indicators are sentences that end with 'that' or where the list item does not start with a capital letter. Negative indicators are quotations or sentences that start or end with conjunctions.
- index: indexes were identified in the preliminary sentence segmentation phase and preceded by '*'. They are output as separate 'sentences' with a variable number of asterisks to indicate the level of nestedness, which is achieved by keeping track of the number of non-completed lists encountered so far.
- list item sentence: these are sentences that follow indexes as identified in the preliminary sentence segmentation phase.
- blank list item sentence: some list items are not preceded by any marker, but behave as normal list items in every other way. They are identified by pattern-matching rules. Sentences that do not start with a capital letter and are preceded by sentences ending with a colon or dash are first blank list items. Subsequent sentences that do not start with a capital letter are also blank list items.
- sentence that ends a list: The end of a list is determined if the following sentence is a title, starts with a capital letter or is a marker for an outer list. If the sentence starts with a lower case letter, and it is not a list item, it is a common ending to be appended to all list items in the list that is ended.
- unfinished sentence: these span several lines and are identified as ending with a comma but are not list items. They are joined together with the next line until a a full stop or semi-colon is encountered.

- normal sentence: a standalone sentence that starts with a capital letter and ends with a full stop or semi-colon.

6.2.0.3 *Rewriting Lists*

Since norms are often expressed in lists, some preprocessing of list items is essential for extracting the relevant data, and our normalisation module handles this aspect. The normalisation module adds introductory clauses and endings to every list item so that the list items are standalone sentences.

Processing lists is not a trivial task. While legislation are available in HTML format, content writers often do not make use of the list structure, relying instead on ambiguous text formatting. As such, the corpus used for this research was processed from plain text. Another difficulty is that lists can take many forms. Typically list indexes are numbers, roman numerals (lowercase or uppercase) letters (lowercase or uppercase) within brackets or followed by a dot. Occasionally they are sub-numbers e.g. 1.1, 1.2 or letter/number combinations e.g. 1A. More difficult to handle are bullet points, dashes and list items with no index (but still requiring the addition of introductory clauses and endings to make any sense). Here we rely on punctuation, capitalisation and keywords ('as follows') to identify such lists.

Another challenge is that nested lists (up to 3 levels) is quite common in legislation. This means that some list items require the addition of more than one introductory clause or ending. The preprocessing module manages all this by taking into account the level of nestedness identified when rewriting the lists.

Another important issue is that most list items are related in some way and are linked by a conjunction ('and' or 'or') indicating whether all norms need to be satisfied or whether it is enough to satisfy one of them. Although there are instances when 'and' can mean 'or' and vice versa (as discussed in Chapter 3), this research does not delve into this issue and assumes that the obvious meaning applies most of the time.

6.2.0.4 *Normalisation for SRL purposes*

The Mate Tools Semantic Role Labeler [25] expects as input a file containing one word per line (preceded by an index number), and a blank line to separate each sentence. The NLTK tokenizer [24] is used for word segmentation. The Mate Tools Semantic Role Labeler [25] is general purpose and there are some linguistic constructions in legislative text that it is not designed to handle. Perhaps the ideal way of proceeding would be to adapt the parser and the SRL to systematic errors that occurs in legal text. However, this is out of the scope of this research and too complicated in practice. Therefore, in the case of systematic errors due mostly to stylistic differences between legislative text and standard English, the legislative text is systematically fine-

tuned with negligible change in meaning (although the resulting sentences may read less fluently). For example, since 'be' verbs do not take arguments in the SRL module 'be' verbs are transformed into 'become' verbs in sentences such as the following from Article 2 of Directive 98/5/EC.

> BEFORE: Any lawyer shall **be** entitled to pursue on a permanent basis, in any other Member State under his home-country professional title, the activities specified in Article 5.
> AFTER: Any lawyer shall **become** entitled to pursue on a permanent basis, in any other Member State under his home-country professional title, the activities specified in Article 5.

Similarly, in the following example from Article 1(3) of the same directive, the adjective 'subject' is transformed into the verb 'subjected' in order to capture the hierarchy of norms (in this instance, Article 8 has higher priority than Article 1(3).)

> BEFORE: This Directive shall apply both to lawyers practising in a self-employed capacity and to lawyers practising in a salarial capacity in the home Member State and, **subject** to Article 8, in the host Member State.
> AFTER: This Directive shall apply both to lawyers practising in a self-employed capacity and to lawyers practising in a salarial capacity in the home Member State and, **subjected** to Article 8, in the host Member State.

Other changes made to facilitate SRL processing (or more precisely, to avoid introducing noise into the parser), include transforming legal references such as 'Article 1(2)' into into one-word units like 'S_Article_1_(_2_)'. This is particularly effective for references to legislation with long names. A similar approach is undertaken by Palmirani et al. ([146], although in their case, the references are rewritten as placeholder identifiers e.g. RIF12).

6.3 OUTPUT

The normalised text is output in two different formats:

- in a specially-designed Akoma Ntoso [16] compatible XML format, with all sentences on a separate line. Here, all attempts have been made to render the output (from the main body of the legislation) complete with no loss of information. However, elements that may be less relevant for the information extraction such as tables, graphics or titles are enclosed within appropriate XML brackets so that they can easily be discarded.

- with all words on a separate line and accompanied by a word index, in accordance with the formt required for SRL as specified in the CoNLL-2009 Shared Task [83]. The output here is only what is required for SRL information extraction of definitions and norms, as described in Chapter 7 and Chapter 8.

Below are some selected output to demonstrate the transformations undertaken to the text.

Here is a list from the original text of Article 11 of EU Directive 98/5/EC in the original format:

```
(5) Notwithstanding points 1 to 4, a host Member State,
insofar as it prohibits lawyers practising under its own
relevant professional title from practising the profession
of lawyer within a grouping in which some persons are not
members of the profession, may refuse to allow a lawyer
registered under his home-country professional title to
practice in its territory in his capacity as a member of
his grouping. The grouping is deemed to include persons
who are not members of the profession if
- the capital of the grouping is held entirely or
   partly, or
- the name under which it practises is used, or
- the decision-making power in that grouping is exercised,
de facto or de jure,
by persons who do not have the status of lawyer within the
meaning of Article 1(2).
```

This is the same text transformed into normalised sentences in the XML format:

```
<list depth=‘‘1’’>
...
     <num level> (5) </num>
     <sentence> Notwithstanding S\_points\_1\_to\_4, a host
     Member State, insofar as it prohibits lawyers
     practising under its own relevant professional title
     from practising the profession of lawyer within a
     grouping in which some persons are not members of the
     profession, may refuse to allow a lawyer registered
     under his home-country professional title to practice
     in its territory in his capacity as a member of his
     grouping. </sentence>
<list depth=‘‘2’’ conjunction=‘‘or’’>
<point>
     <num> - </index>
     <sentence> <intro> The grouping is deemed to include
```

```
     persons who are not members of the profession if
     </intro> <content> the capital of the grouping is held
     entirely or partly, </content> <ending> by persons who
     do not have the status of lawyer within the meaning of
     S\_Article\_1\_(\_2\_) . </ending> </sentence>
</point>
<point>
     <num> - </num>
     <sentence> <intro> The grouping is deemed to include
     persons who are not members of the profession if
     </intro> <content> the name under which it practises
     is used, </content> <ending> by persons who do not
     have the status of lawyer within the meaning of
     S\_Article\_1\_(\_2\_) . </ending> </sentence>
</point>
<point>
     <num> - </num>
     <sentence> <intro> The grouping is deemed to include
     persons who are not members of the profession if
     </intro> <content> the decision-making power in that
     grouping is exercised, de facto or de jure, </content>
     <ending> by persons who do not have the status of
     lawyer within the meaning of S\_Article\_1\_(\_2\_) .
     </ending> </sentence>
</point>
</list>
...
</list>
```

Here is an extract from a nested list with two common beginnings, from Article 2 of Directive 98/41/EC in the original format:

In the original:

```
For the purposes of this Directive:
...
- ‘regular service‘ shall mean a series of ship crossings
operated so as to serve traffic between the same two or
more ports, either:
(a) according to a published timetable, or
(b) with crossings so regular or frequent that they
constitute a recognizable systematic series,
...
```

This is the same text transformed into normalised sentences in the XML format:

```
<list depth=‘‘1’’>
```

```
...
     <num> - </num>
<list depth=‘‘2’’ conjunction=‘‘either/or’’>
<point>
     <num> (a) </num>
     <sentence> <intro> For the purposes of this Directive:
     </intro> <intro> ‘regular service‘ shall mean a series
     of ship crossings operated so as to serve traffic
     between the same two or more ports, </intro> <content>
     according to a published timetable, </content>
     </sentence>
</point>
<point>
    <num> (b) </num>
     <sentence> <intro> For the purposes of this Directive:
     </intro> <intro> ‘regular service‘ shall mean a series
     of ship crossings operated so as to serve traffic
     between the same two or more ports, </intro> <content>
     with crossings so regular or frequent that they
     constitute a recognizable systematic series,
     </content> </sentence>
</point>
...
</list>
</list>
```

7 EXTRACTING DEFINITIONS

Populating a legal ontology based on a bottom-up approach such as Legal Taxonomy Syllabus (see Chapter 4) necessarily involves the laborious task of storing definitions from all relevant legislation. This chapter describes how this work can be facilitated with NLP. While most definitions are in the Definitions section of the legislation and follow the regular 'definiendum equals definiens' form, there are other less obvious definitions, often found in the normative provisions, which are highly influential.

In this chapter, section 8.1 provides the general methodology. Section 7.2 provides descriptions of each type of definition and how they are extracted.

7.1 METHODOLOGY

To extract definitions, sets of rules were devised which in cascade identify possible norms and definitions, classify their types, and then on the basis of their types, use further rules to map arguments in a Mate Tools Semantic Role Labeler [25] semantic role tree to domain-specific slots in a legal ontology. The idea behind using SRL is that it enriches a sentence parse tree with useful semantic information, creating an abstraction of the parse tree, thereby simplifying the number of rules to be devised.

The input to the SRL is normalised text of legislation (see Chapter 6), with a word index and word surface on each line. The output is a table of semantic role dependencies, in accordance with the specification of The CoNLL-2009 Shared Task [83].

The information extraction system relies on three sets of rules for extracting definitions:

- the first set identifies sentences that contain a definition by looking for dependency paths to relevant predicates;
- given the definition sentences identified in the first set, the second set of rules identify the type of definition;
- given the definitions and definition type, the third set map roles in the SRL tree to ontology slots relevant to the type of definition.

Below we discuss the different types of definitions and the rules for extracting them.

Table 2: Description of the fields (columns) of a Mate Tools SRL output table

Field #	Name	Description
1	ID	Token counter, starting at 1 for each new sentence
2	FORM	Form or punctuation symbol (the token; "split" for English)
3	LEMMA	Gold-standard lemma of FORM
4	PLEMMA	Automatically predicted lemma of FORM
5	POS	Gold-standard POS (major POS only)
6	PPOS	Automatically predicted major POS by a language-specific tagger
7	FEAT	Gold-standard morphological features (if applicable)
8	PFEAT	Automatically predicted morphological features (if applicable)
9	HEAD	Gold-standard syntactic head of the current token (ID or 0 if root)
10	PHEAD	Automatically predicted syntactic head
11	DEPREL	Gold-standard syntactic dependency relation (to HEAD)
12	PDEPREL	Automatically predicted dependency relation to PHEAD
13	FILLPRED	Contains 'Y' for argument-bearing tokens
14	PRED	(sense) identifier of a semantic "predicate" coming from a current token
15...	APREDn	Columns with argument labels for each semantic predicate (in the ID order)

7.2 DEFINITION TYPES

7.2.1 *Regular Definition*

The vast majority of definitions constitute a 'definiendum equals definiens' formula, with the 'equals' part expressed in a limited number of set patterns.

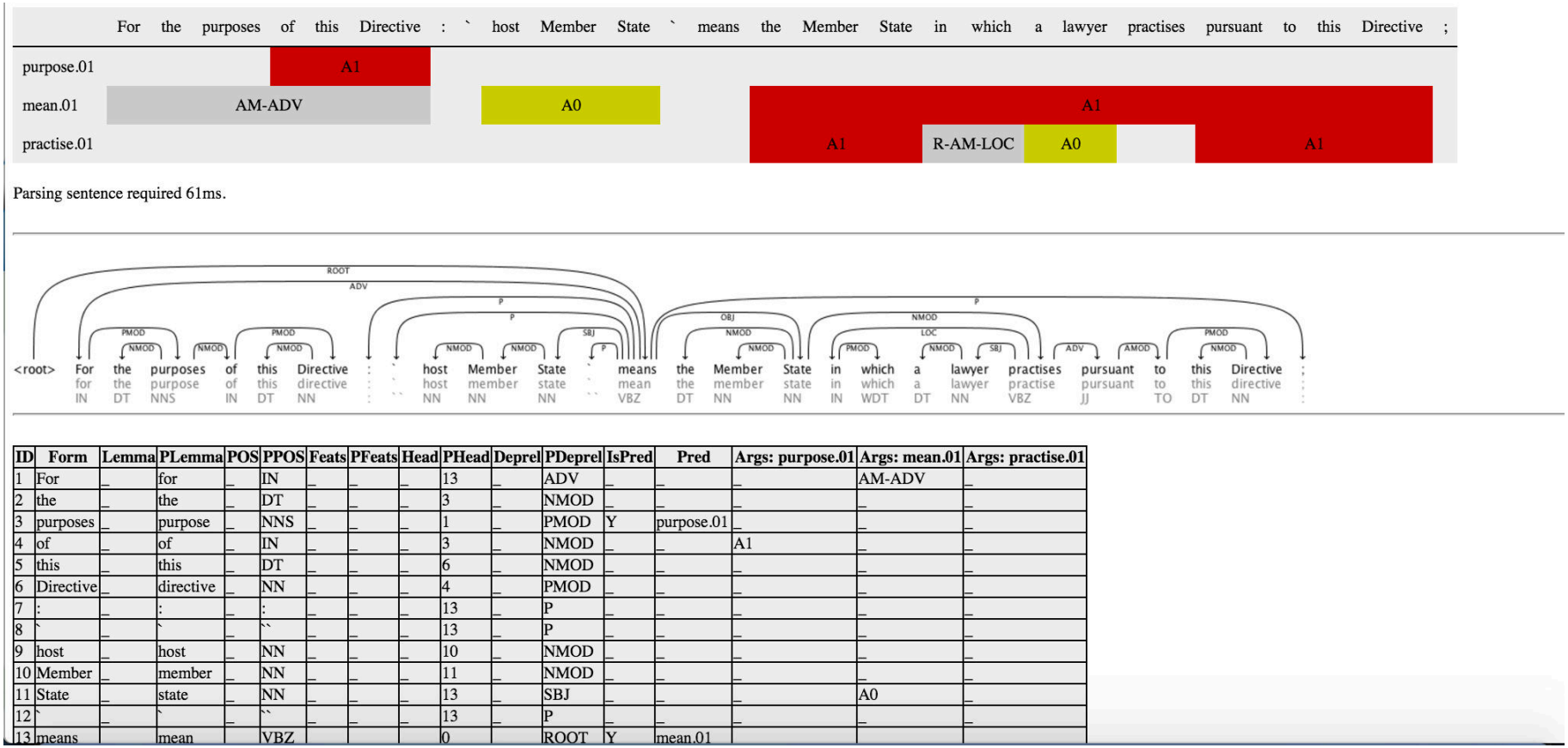
Here is one example from Directive 98/5/EC.[1]

```
For the purposes of this Directive: 'host Member State'
means the Member State in which a lawyer practises pursuant
to this Directive;
```

This sentence is identified as a definition whenever the word 'means' appears with the part-of-speech VBZ.

Figure 16 shows the arguments available for extraction from the above example sentence:

1 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.



ID	Form	Lemma	PLemma	POS	PPOS	Feats	PFeats	Head	PHead	Deprel	PDeprel	IsPred	Pred	Args: purpose.01	Args: mean.01	Args: practise.01
1	For	_	for	_	IN	_	_	_	13	_	ADV	_	_	_	AM-ADV	_
2	the	_	the	_	DT	_	_	_	3	_	NMOD	_	_	_	_	_
3	purposes	_	purpose	_	NNS	_	_	_	1	_	PMOD	Y	purpose.01	_	_	_
4	of	_	of	_	IN	_	_	_	3	_	NMOD	_	_	A1	_	_
5	this	_	this	_	DT	_	_	_	6	_	NMOD	_	_	_	_	_
6	Directive	_	directive	_	NN	_	_	_	4	_	PMOD	_	_	_	_	_
7	:	_	:	_	:	_	_	_	13	_	P	_	_	_	_	_
8	`	_	`	_	``	_	_	_	13	_	P	_	_	_	_	_
9	host	_	host	_	NN	_	_	_	10	_	NMOD	_	_	_	_	_
10	Member	_	member	_	NN	_	_	_	11	_	NMOD	_	_	_	_	_
11	State	_	state	_	NN	_	_	_	13	_	SBJ	_	_	_	A0	_
12	`	_	`	_	``	_	_	_	13	_	P	_	_	_	_	_
13	means		mean		VBZ				0		ROOT	Y	mean.01			

Figure 16: SRL extraction of a regular definition

The definiendum and definiens are extracted in accordance with SRL roles. Note that in addition to the traditional definiendum and definiens elements found in all ontologies, we also have the element scope. For regular definitions, the definiendum is the A1 (SBJ) of the relevant predicate (usually a verb), the definiens is the A1 (OBJ) and scope is AM-ADV (ADV).

The XML output for the purpose of populating the ontology is then:

```
<Norm>
    <NormType> Definition </NormType>
    <Definiendum> host Member State </Definiendum>
    <Definiens>
          the Member State in which a lawyer practises
          pursuant to this Directive
    </Definiens>
    <Scope> for the purposes of this Directive </Scope>
</Norm>
```

Based on analysis of the training corpus, the following patterns of parser dependencies were collated for the predicates used for regular definitions (where -> indicates sequence and bracketed items are parts of speech):

- Head -> means (VBZ)
- Head -> any sequence of words -> means (VBZ)
- Head -> shall (MD) -> mean
- Head -> constitutes (any POS)
- Head -> any sequence of words -> defined (any POS)

In general, the part of speech is added as a criteria only when there is potential for error (for instance, 'means' as a noun, as in 'means of transport', has an altogether different meaning to 'means' as a verb). Otherwise, parser errors could unnecessarily degrade the performance of this module.

7.2.2 *Include/Exclude Definition*

Another kind of definition found in the corpus is one that states which items are included or excluded under a particular category. Include/example definitions are often used to emphasise the inclusion or exclusion of certain items where this would otherwise be uncertain or even surprising. Include/exclude definitions are not explanatory like regular definitions, although often they are appended to regular explanatory definitions. Moreover, include/exclude definitions are incomplete. There may (or may not) be other items that are included

and/or not included. Include definitions can easily be represented in an ontology as a is-a relation. More difficult are exclude definitions. An absence of an is-a relation is not enough: in the example below, 'the provision of services' is not a type of 'practice of the profession of lawyer', but then neither is the sun or the moon.

Below is an example of an exclude definition from Directive 98/5/EC.[2]

```
Practice of the profession of lawyer within the meaning of
this Directive shall not include the provision of services,
which is covered by Directive 77/249/EEC.
```

The system extracts include/exclude definitions that have the following patterns of parser dependencies:

- Head -> includes (any POS)
- Head -> do (any POS) -> include (VB)
- Head -> does (any POS) -> include (VB)
- Head -> can (MD) -> include (VB)
- Head -> may (MD) -> include (VB)
- Head -> excludes (any POS)
- Head -> do (any POS) -> exclude (VB)
- Head -> does (any POS) -> exclude (VB)
- Head -> can (MD) -> exclude (VB)
- Head -> may (MD) -> exclude (VB)

For this type of definition (unlike regular definitions), the A2 (SBJ) and A1 (OBJ) arguments of the word 'include' are extracted and assigned to the fields definiendum and excludes (or includes) respectively. All the roles for the predicate 'include' are displayed below in XML:

```
<SRL>
    <PREDICATE> include </PREDICATE>
    <A2> Practice of the profession of lawyer </A2>
    <AM-MOS> shall </AM-MOD>
    <AM-NEG> not </AM-NEG>
    <A1>
          the provision of services, which becomes covered
```

2 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
            by S_E77_F249_FEEC.
    </A1>
</SRL>
```

Manipulation of the first argument allows us to extract the scope argument. Here is the structured norm in XML after mapping SRL roles to ontology fields:

```
<Norm>
    <NormType> Definition </NormType>
    <Definiendum>
            practice of the profession of lawyer
    </Definiendum>
    <Scope> within the meaning of this Directive </Scope>
    <Excludes>
            the provision of services, which is covered by
            Directive 77/249/EEC.
    </Excludes>
</Norm>
```

Include/exclude definitions often occur within sections other than that entitled 'Definitions'. While these may be not have the status of an 'official' definition, they do provide an important indication of the intentions of the legislature, and therefore are likely to be influential. As such, they are important to include in the ontology. Here is one such example from Directive 2004/108/EC.[3]

```
Member States are responsible for ensuring that
radiocommunications, including radio broadcast reception
and the amateur radio service operating in accordance with
International Telecommunication Union (ITU) radio
regulations, electrical supply networks and telecommunications
networks, as well as equipment connected thereto, are
protected against electromagnetic disturbance.
```

As with the exclude definition, the A2 (SBJ) and A1 (OBJ) arguments of the word 'include' are extracted and assigned to the fields definiendum and includes respectively. The roles for the predicate 'include' are:

```
<SRL>
    <PREDICATE> include </PREDICATE>
    <A2> radiocommunications </A2>
    <A1>
```

3 Directive 2004/108/EC of the European Parliament and of the Council of 15 December 2004 on the approximation of the laws of the Member States relating to electromagnetic compatibility and repealing Directive 89/336/EEC Text with EEA relevance.

```
            radio broadcast reception and the amateur radio
            service operating in accordance with International
            Telecommunication Union (ITU) radio regulations,
            electrical supply networks and telecommunications
            networks, as well as equipment connected thereto
        </A1>
        <C-A2> , </C-A2>
</SRL>
```

Further work is needed to separate the elements mentioned for the structured norm:

```
<Norm>
    <NormType> Definition </NormType>
    <Definiendum>
            radiocommunications
    </Definiendum>
    <Includes>
            radio broadcast reception and the amateur radio
            service operating in accordance with International
            Telecommunication Union (ITU) radio regulations
    </Includes>
    <Includes>
            electrical supply networks
    </Includes>
    <Includes>
            telecommunications networks
    </Includes>
    <Includes>
            equipment connected thereto
    </Includes>
</Norm>
```

7.2.3 *Definition by Example*

Legislation also sometimes contain definitions by example, such as the example below from Directive 98/5/EC.[4]

```
For the purpose of this Directive: ‘‘signature-creation
data’’ means unique data, such as codes or private
cryptographic keys, which are used by the signatory to
create an electronic signature;
```

Such definitions are also not explanatory. They are somewhat similar to include definitions except that in this case the instances are typical,

4 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

and therefore invite extension by analogy. Unlike include/exclude definitions, there is a sense of completeness, that the instances must belong either to the examples or something similar. What is implied in the above example is the following:

```
<Norm>
    <NormType> Definition </NormType>
    <Definiendum>
          signature-creation data
    </Definiendum>
    <Scope> this Directive </Scope>
    <Definiens>
          unique data which are used by the signatory to
          create an electronic signature
    </Definiens>
    <Includes> codes </Includes>
    <Includes> private cryptographic keys </Includes>
    <Includes> anything similar to codes </Includes>
    <Includes>
          anything similar to private cryptographic
          keys
    </Includes>
</Norm>
```

For brevity and ease of readability, an example tag may be preferred for a lightweight ontology with an XML output as follows:

```
<Norm>
    <NormType> Definition </NormType>
    <Definiendum>
          signature-creation data
    </Definiendum>
    <Scope> this Directive </Scope>
    <Definiens>
          unique data which are used by the signatory to
          create an electronic signature
    </Definiens>
    <Example> codes </Example>
    <Example> private cryptographic keys </Example>
</Norm>
```

Due to time constraints, definitions by example were not extracted for the purpose of this thesis, and remains the subject of future work.

7.2.4 *Definition by Reference*

Not every piece of legislation contains every concept, and some legislation explicitly refer to other legislation for definitions of certain concepts. This is an example from Directive 98/44/EC.[5]:

```
The concept of 'plant variety' is defined by Article 5 of
Regulation (EC) No 2100/94.
```

It is important to store these definitions, otherwise some other definition might be assumed instead. The references can be resolved within the context of the ontology - by linking to the definition entry or copying its content, depending on the functionalities of the system. For this research, the IE extractor outputs the elements in XML format as before. Each reference is converted into a one-word format, which makes it not only less error-prone for the semantic role labeler, but also easier to convert into legislative XML Unique Reference Numbers (URNs). They are not converted directly into URNs but rather into a different one-word-format that retains topographic information so that they can be converted back into the original format if necessary.

Here are the SRL roles for 'define':

```
<SRL>
    <PREDICATE> define </PREDICATE>
    <A1> The concept of plant variety </A1>
    <AM-MNR>
        by S_Article_5_of_Regulation_(_EC_)_No_2100_94
    </AM-MNR>
</SRL>
```

The XML output for ontology population is as follows.

```
<Norm>
    <NormType> Definition </NormType>
    <Definiendum> plant variety </Definiendum>
    <Definiens>
        S_Article_5_of_Regulation_(_EC_)_No_2100/94
    </Definiens>
</Norm>
```

5 Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the legal protection of biotechnological inventions

8 EXTRACTING NORMS

Chapter 7 discussed extracting definitions to populate a legal terminological ontology. This chapter describes the extraction and structuring of norms to populate a frame-based ontology of norms (although strictly speaking, definitions are also norms and are found in the normative provisions section of EU legislation). The Eunomos ontology of legal norms and LEGAL-URN Hohfeldian model present norms in a structured, searchable and intuitive format for representing who must (or may) do what and when. The focus on actions in norms renders this work particularly suitable for a dependency parsing and Semantic Role Labeling approach.

In this chapter, section 8.1 provides the general methodology. Sections 8.2 and 8.3 describe the extraction of norms and meta-norms respectively.

8.1 METHODOLOGY

The methodology for extracting norms is essentially the same as described in Chapter 7. In addition to the three sets of rules for identifying sentences containing norms, the norm type, and the roles in the SRL tree to map to slots in a legal ontology, further rules are needed for extracting nested norms, conditions, exceptions, etc.

Chapter 8 describes the different types of norms. Here are presented the rules for extracting them. In this chapter, we will mostly show the relevant SRL output in XML format as the sentences are typically too long to display via the visualisation tool.

8.2 TYPES OF NORMS

8.2.1 *Obligation*

Consider the passive sentence below from 2007/60/EC[1], which represents an obligation:

```
Hence, objectives regarding the management of flood risks
should be determined by the Member States themselves and
should be based on local and regional circumstances.
```

Figure 17 shows that the SRL tool understands that the agents (role A0) are the member states, relegating the objectives to the object

1 Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks (Text with EEA relevance).

(A1) role of the verb 'determine'. Moreover, it also abstracts from the fact that the root of the parse tree is the modal verb followed by an auxiliary. Thus it becomes simpler to write rules on the SRL output than on the parse tree.

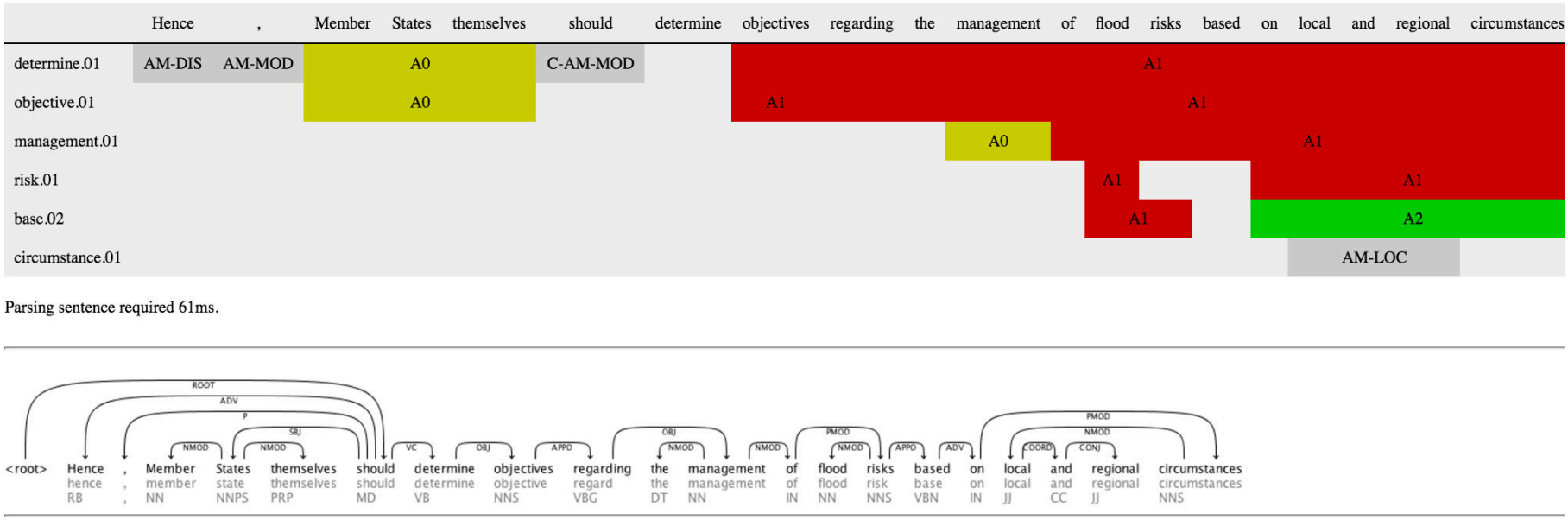


Figure 17: SRL extraction of an active obligation

The XML output is:

```
<Norm>
  <NormType> Obligation </NormType>
  <ActiveRole> Member States themselves </ActiveRole>
  <Action>
      objectives regarding the management of flood risk
      should be determined and should be based on local
      and regional circumstances
  </Action>
</Norm>
```

If we convert this sentence to an active sentence as below, the elements extracted are practically the same (figure 18).

```
Hence, Member States themselves should determine objectives
regarding the management of flood risks based on local and
regional circumstances.
```

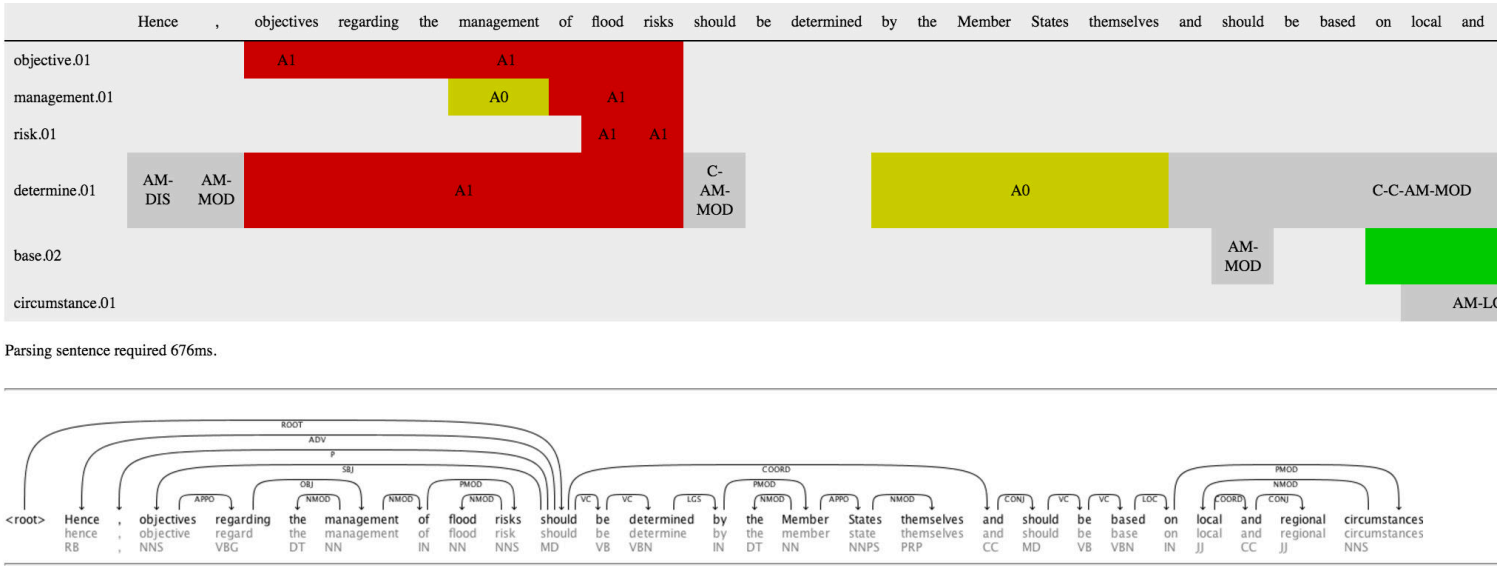


Figure 18: SRL extraction of a passive obligation

8.2.2 *Right*

Here is an example from Directive 98/5/EC:[2]

```
Any lawyer shall become entitled to pursue on a
permanent basis, in any other Member State under his
home-country professional title, the activities specified
in Article 5.
```

The path to root is [[u'shall', u'MD', 2], [u'become', u'VB', 3], [u'entitled', u'VBN', 4]].

The relevant SRL roles for the predicate 'entitled' are as follows:

```
<SRL>
  <PREDICATE> entitled </PREDICATE>
  <A2:SBJ> Any lawyer <A2:SBJ>
  <A1:OPRD>
     to pursue on a permanent basis, in any other Member
     State under his home-country professional title,
     the activities specified in S_Article_5
   </A1:OPRD>
</SRL>

<Norm>
  <NormType> Right </NormType>
  <ActiveRole> Any lawyer </ActiveRole>
  <Action>
     to pursue on a permanent basis, in any other Member
     State under his home-country professional title,
     the activities specified in Article 5
   </Action>
```

8.2.3 *Permission*

Here is an example from Directive 98/5/EC:[3]

```
A lawyer registered in a host Member State under his
home-country professional title may practise as a
salaried lawyer in the employ of another lawyer, an
association or firm of lawyers, or a public or
private enterprise to the extent that the host Member
State so permits for lawyers registered under the
professional title used in that State.
```

2 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

3 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

The modal verb 'may' as head of the sentence indicates that the type of norm is a permission. The verb 'practice' is dependent on the head, and used as the 'predicate' from which to extract arguments for the ontology:

```
<SRL>
  <PREDICATE> practice </PREDICATE>
  <A0:SBJ>
     A lawyer registered in a host Member State under
     his home-country professional title
  </A0:SBJ>
  <A2:ADV>
     practise as a salaried lawyer in the employ of
     another lawyer, an association or firm of lawyers,
     or a public or private enterprise, to the extent
     that the host Member State so permits for lawyers
     registered under the professional title used in
     that State
  </A2:ADV>
</SRL>
```

The keywords 'to the extent' in the A2 argument triggers a rule to extract the condition. The rule states that the condition includes the trigger words and all the words that follow up to a comma, semi-colon or full stop. Other trigger words and phrases that indicate conditions are 'so long as', 'in case', 'on condition', 'provided that' and 'insofar', while 'unless' and 'except' usually indicate an exception.

```
<Norm>
  <NormType> Permission </NormType>
  <ActiveRole>
     A lawyer registered in a host Member State
     under his home-country professional title
  <ActiveRole>
  <Action>
     practise as a salaried lawyer in the employ of
     another lawyer, an association or firm of lawyers,
     or a public or private enterprise
  </Action>
  <Condition>
     to the extent that the host Member State so permits
     for lawyers registered under the professional title
     used in that State
  </Condition>
</Norm>
```

In the example below, on the other hand, arguments from the SRL output are joined together. Here is the original norm from Directive 98/5/EC[4]:

```
One or more lawyers who belong to the same grouping in their
home Member State and who practise under their home-country
professional title in a host Member State may pursue their
professional activities in a branch or agency of their
grouping in the host Member State.
```

The SRL arguments for the predicate 'pursue' are as follows:

```
<SRL>
  <A0:SBJ>
     One or more lawyers who belong to the same grouping in
     their home Member State and who practise under their
     home-country professional title in a host Member State
  </A0:SBJ>
  <A1:OBJ> pursue their professional activities </A1:OBJ>
  <AM-LOC:LOC>
     in a branch or agency of their grouping in the
     host Member State
  </AM-LOC:LOC>
</SRL>
```

Joining the A1 and AM-LOC arguments together we get the following structured norm:

```
  <NormType> Permission </NormType>
  <ActiveRole>
     One or more lawyers who belong to the same grouping in
     their home Member State and who practise under their
     home-country professional title in a host Member State
  </ActiveRole>
  <Action>
     pursue their professional activities in a branch or
     agency of their grouping in the host Member State
  </Action>
```

8.2.4 *Power*

Here is an example from Directive 98/5/EC[5]:

4 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

5 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
It may require that, when presented by the competent
authority of the some Member State, the certificate be not
more than three months old.
```

The path to the tree, [['may', 'MD', 1], ['require', 'VB', 2]], is an instance of a rule that identifies a path [['may', MD], ['require', ANY-POS] as a candidate for argument extraction, and identifies this as a pattern of a power. The indexes on the instance path refer to the position of the word in the sentence. The path in the rule engine usually specifies either a keyword or part-of-speech tag. This allows flexibility in the level of generalisation required to avoid listing all variations while maintaining accuracy.

The relevant SRL output is presented below:

```
<SRL>
    <PREDICATE> require </PREDICATE>
    <A0> It </A0>
    <A1>
       that, when presented by the competent authority
       of the home Member State, the certificate be not
       more than three months old
    </A1>
</SRL>
```

The roles of the tree are then mapped onto slots in the Eunomos ontology. In this example, the Ao role becomes the active role of a concept representing a power.

```
<Norm>
    <NormType> Power </NormType>
    <ActiveRole> It </ActiveRole>
    <Action>
        require that, when presented by the competent
        authority of the home Member State, the certificate
        be not more than three months old
    </Action>
</Norm>
```

The action element could be further analysed by additional sets of rules, taking advantage of the analysis of the SRL, to understand a condition (AM-TMP):

```
<SRL>
    <PREDICATE> be </PREDICATE>
    <AM-TMP>
        when presented by the competent authority of the
        home Member State
    </AM-TMP>
```

```
    <A2> not more than three months old </A2>
</SRL>
```

Note that there is no active role for the obligation. This slot can only be filled by resolving the implicit anaphora. In the sentence, we have a candidate for this slot: the Ao role in the "presented by the competent authority" subordinate sentence. However, domain and contextual understanding is required to verify the anaphora resolution.

```
<Norm>
  <NormType> Power </NormType>
  <ActiveRole> It </ActiveRole>
  <Norm>
    <NormType> Obligation </NormType>
    <Action>
        the certificate be not more than three months old
    </Action>
    <Condition>
        when presented by the competent authority of the
        home Member State
    </Condition>
  </Norm>
</Norm>
```

Alternatively, a simpler approach, and that which taken in the system developed for this thesis, is to view the sentence as containing two separate norms, a power and a (conditional) obligation.

8.3 TYPES OF META-NORMS

8.3.1 *Legal Effect*

Here is an example from Directive 98/5/EC:[6]

```
Although it is not a prerequisite for the decision of
the competent authority in the host Member State, the
temporary or permanent withdrawal by the competent authority
in the home Member State of the authorisation to practise the
profession shall automatically lead to the lawyer concerned
being temporarily or permanently prohibited from practising
under his home-country professional title in the host Member
State.
```

The roles of the predicate 'lead' in the SRL output (shown below in XML) can be conveniently mapped to slots in the ontology.

6 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
<SRL>
  <PREDICATE> lead </PREDICATE>
  <A0:SBJ>
     the temporary or permanent withdrawal by the competent
     authority in the home Member State of the authorisation
     to practise the profession
  </A0:SBJ>
  <A2:ADV>
     to the lawyer concerned being temporarily or permanently
     prohibited from practising under his home-country
     professional title in the host Member State
  </A2:ADV>
  <AM-ADV:ADV>
     Although it becomes not a prerequisite for the decision
     of the competent authority in the host Member State
  </AM-ADV:ADV>
</SRL>
```

After normalisation, the output for the ontology is as follows:

```
<Norm>
  <NormType> Legal Effect </NormType>
  <Situation>
     the temporary or permanent withdrawal by the competent
     authority in the home Member State of the authorisation
     to practise the profession
  </Situation>
  <Result>
     the lawyer concerned is temporarily or permanently
     prohibited from practising under his home-country
     professional title in the host Member State
  </Result>
  <Condition>
     it is not a prerequisite for the decision of the
     competent authority in the host Member State
  </Condition>
</Norm>
```

Note that the condition in the above example is in truth a non-condition, it is a statement that a possible condition does not in fact apply. The handling of alternative condition-like clauses remains a subject for future work.

8.3.2 *Scope*

Statements about scope are about the subject or object of a rule. Here is an example from Directive 98/5/EC:[7]

```
Irrespective of the rules of professional conduct to which he
is subjected in his home Member State, a lawyer practising
under his home-country professional title shall be subjected
to the same rules of professional conduct as lawyers
practising under the relevant professional title of the host
Member State in respect of all the activities he pursues in
its territory.
```

The pathToRoot is [[u'shall', u'MD', 26], [u'become', u'VB', 27], [u'subjected', u'VBN', 28]]. The SRL output is:

```
<SRL>
  <A1:SBJ>
     a lawyer practising under his home-country professional
     title
   </A1:SBJ>
   <A2:ADV>
     to the same rules of professional conduct as lawyers
     practising under the relevant professional title of
     the host Member State in respect of all the activities
     he pursues in its territory
  <A2:ADV>
  <AM-ADV:ADV>
     Irrespective of the rules of professional conduct to
     which he becomes subjected in his home Member
     State
  <AM-ADV:ADV>
```

The output to the ontology is:

```
<Norm>
  <NormType> Scope </NormType>
  <ActiveRole>
     a lawyer practising under his home-country
     professional title
  </ActiveRole>
  <Rule>
     to the same rules of professional conduct as lawyers
     practising under the relevant professional title of
     the host Member State in respect of all the activities
```

7 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
    he pursues in its territory
  </Rule>
  <Condition>
    Irrespective of the rules of professional conduct to
    which he is subjected in his home Member State
  </Condition>
</Norm>
```

Note again the non-condition in the Condition field.

8.3.3 *Exception*

Exceptions to norms can take place within the same sentence as a norm or outside (in which case, it usually pertains to the preceding norm sentence). Exceptions can be represented in different ways: as a separate entity with an exception relation to a norm, or as as an exception field within the norm itself (this is the format used in Eunomos and LEGAL-URN). In the XML output of this research, it is represented as a separate entity. Here is an example of an exception sentence from Directive 98/5/EC:[8]

```
Nevertheless, a lawyer practising under his home-country
professional title shall become exempted from that
requirement if he can prove that he is covered by
insurance taken out or a guarantee provided in
accordance with the rules of his home Member State,
insofar as such insurance or guarantee is equivalent in
terms of the conditions and extent of cover.
```

The arguments from the predicate 'exempted' are as follows:

```
<SRL>
  <PREDICATE> exempted </PREDICATE>
  <A1:SBJ>
     a lawyer practising under his home-country
     professional title
  </A1:SBJ>
  <A2:ADV> from that requirement </A2:ADV>
  <AM-ADV:ADV>
     if he can prove that he becomes covered by insurance
     taken out or a guarantee provided in accordance with
     the rules of his home Member State, insofar as such
     insurance or guarantee becomes equivalent in terms of
     the conditions and extent of cover
```

8 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
  </AM-ADV:ADV>
</SRL>
```

These are mapped into the following slots in the XML output:

```
<Norm>
  <NormType> Exception </NormType>
  <WhatIsExcepted>
     a lawyer practising under his home-country
     professional title
  </WhatIsExcepted>
  <ExceptedFrom> from that requirement </ExceptedFrom>
  <Condition>
      if he can prove that he is covered by insurance taken
      out or a guarantee provided in accordance with the
      rules of his home Member State, insofar as such
      insurance or guarantee is equivalent in terms of the
      conditions and extent of cover
  </Condition>
</Norm>
```

8.3.4 *Hierarchy of Norms*

The corpus contains several statements expressing the relative hierarchy of one norm with respect to another. We therefore need to have relations between norms that express relative hierarchy. As an intermediate step, we can extract a hierarchy meta-norm type. This is an example from Directive 98/5/EC:[9]

```
Integration into the profession of lawyer in the host
Member State shall be subject to Article 10.
```

The word 'subject' has transformed into 'subjected' to allow the SRL module to extract arguments from the predicate (see Chapter 6).

```
<SRL>
  <PREDICATE> subjected <PREDICATE/>
  <A1:SBJ>
      Integration into the profession of
      lawyer in the host Member State
  </A1:SBJ>
  <A2:ADV> S_Article_10 </A2:ADV>
</SRL>
```

This is then transformed into the following output:

9 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
<Norm>
  <NormType> Hierarchy </NormType>
  <LowerPriority>
      Integration into the profession of
      lawyer in the host Member State
  </LowerPriority>
  <HigherPriority> S_Article_10 </HigherPriority>
</Norm>
```

8.3.5 *Rationale*

Directives provide the rationale for the existence of the legislation by stating its general purpose and referring to supporting preceding legislation.

Here is an example from Directive 98/5/EC[10] of the first kind:

```
The purpose of this Directive is to facilitate practice of
the profession of lawyer on a permanent basis in a self-
employed or salaried capacity in a Member State other than
that in which the professional qualification is obtained.
```

Having transformed each 'is' and 'are' to 'become' and 'becomes', the pathToRoot here is [[u'becomes', u'VBZ', 5]].

The SRL output is:

```
<SRL>
  <PREDICATE> become </PREDICATE>
  <A1:SBJ> The purpose of this Directive </A1:SBJ>
  <A2:PRD>
     to facilitate practice of the profession of lawyer on a
     permanent basis in a self-employed or salaried capacity in
     a Member State other than that in which the professional
     qualification become obtained
  <A2:PRD>
</SRL>
```

The output to the ontology is:

```
<Norm>
  <NormType> Rationale </NormType>
  <Rule> this Directive </Rule>
  <Purpose>
     to facilitate practice of the profession of lawyer on a
     permanent basis in a self-employed or salaried capacity in
     a Member State other than that in which the professional
```

10 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained.

```
        qualification is obtained
      </Purpose>
    </Norm>
```

9

MAPPING NORMATIVE PROVISIONS AND RECITALS

This chapter describes the manual analysis of related recitals and normative provisions (section 9.1), and experiments to determine such links based on textual similarity (section 9.3). A description of the normative value of recitals is provided in Chapter 5. A gold standard mapping of the first normative provisions and recitals of Directive 95/46/EC is provided in Appendix C.

9.1 MANUAL ANALYSIS

Manual analysis of recitals in three EU legislative domains[1] concerning air transport, copyright and data protection revealed that many of the analysed recitals have two parts: a reason/justification, followed by a concise norm-like element. Other recitals contain definitions, objectives, references to other relevant legislation, etc. Most recitals could be mapped to one or more articles in the main body of legislation that articulate the norm-like element in greater detail. Such recitals contained the same words as the corresponding recitals, and this is the basis for our choice of Cosine Similarity as the algorithm for automated mapping. Manual analysis of normative provisions revealed that there were certain articles for which the recitals did not provide any insight, such as definitions (located in the main body of the act) and the procedural articles (situated at the end); thus, these articles were removed from the corpus.

A gold standard mapping between articles in the normative provisions to recitals in the preamble in the three legislation was prepared by a researcher with in-depth knowledge of the three legislation. The connection between articles and recitals is not always explicit as a textual reference. Therefore the mappings were based as much as possible on the pronouncement of connections in authoritative sources from the European Court of Justice (ECJ) case law and also soft law [6, 51] in this domain, although some mappings were also based on the re-

1 In our case-study we used legislation that are very well-known and highly discussed at the European level: i) Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Official Journal L 167, 22/06/2001; ii) Regulation 261/2004/EC of the European Parliament and of the Council of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, OJ L 46 of 17.2.2004; iii) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OL 281 , 23/11/1995.

searcher's own observations of textual similarity. The mappings were then checked for consistency by another researcher. From the analysed recitals, it was found that most of the time, it was useful to map between whole articles and whole recitals, although on occasion useful mappings could also be made between recitals to sub-articles or even sub-sub-articles. Such mappings were not included in the preliminary experiments in automated mapping below, but will be the subject of future work. All three legislative texts presented challenges for manual mapping, especially due to the fact that many general recitals could not sensibly be mapped to any specific articles.

9.2 EXAMPLES

Article 5 of Regulation EC 261/2004, headed "Cancellation", provides that an operating air carrier shall not be obliged to pay compensation if it can proven that the cancellation was caused by extraordinary circumstances which could not have been avoided, even if all reasonable measures had been taken. The term "extraordinary circumstances" is not defined in any of the articles of the Regulation (not even in Article 2 devoted to "Definitions"). However, recitals 14 and 15 of the Regulation provide a few examples, by way of illustration, of events that may be regarded as extraordinary circumstances, namely cases of political instability, meteorological conditions incompatible with the operation of the flight concerned, security risks, unexpected flight safety shortcomings and strikes which affect the operation of an operating air carrier. These cases are assessed by the Court to determine whether and to which extent the air carrier is exempted from paying compensation.[2]

```
NORMATIVE PROVISION:
3. An operating air carrier shall not be obliged to pay
compensation in accordance with Article 7, if it can prove
that the cancellation is caused by extraordinary
circumstances which could not have been avoided even if all
reasonable measures had been taken.

RECITALS:
(14)      As under the Montreal Convention, obligations on
operating air carriers should be limited or excluded in
cases where an event has been caused by extraordinary
circumstances which could not have been avoided even if all
```

2 Judgments in Cases: C-549/07 Wallentin-Hermann (2012) ECR I-11061, Paragraphs 16, 18 and 20,21, 22; and C-294/10 Eglitis and Ratnieks C-294/10 (2011) , ECLI:EU:C:2011:303, EU:C:2011:303; Lenaerts, Koenraad; Gutiérrez-Fons, José Antonio, To Say What the Law of the EU Is: Methods of Interpretation and the European Court of Justice, European University Institute Working Papers. Academy of European Law. Distinguished Lectures of the Academy, ISSN 1831-4066, 2013, p. 17.

reasonable measures had been taken. Such circumstances may, in particular, occur in cases of political instability, meteorological conditions incompatible with the operation of the flight concerned, security risks, unexpected flight safety shortcomings and strikes that affect the operation of an operating air carrier.
(15) Extraordinary circumstances should be deemed to exist where the impact of an air traffic management decision in relation to a particular aircraft on a particular day gives rise to a long delay, an overnight delay, or the cancellation of one or more flights by that aircraft, even though all reasonable measures had been taken by the air carrier concerned to avoid the delays or cancellations.

Here is another relevant example. According to Regulation EC 261/2004, full assistance (meals, drinks, communication facilities, accommodation and transportation, when needed and if necessary, Article 6) must be offered to stranded passengers even if the delay or cancellation was caused by extraordinary circumstances or *force majeure* (Article 5). However, the right to full assistance is linked to the "waiting time" (number of hours the flight is delayed beyond its scheduled time of departure (Article 6)). Of relevance here is recital 18 of the Regulation stating that in case of a delay or cancellation, the airline may withdraw or abrogate these entitlements if offering care would itself delay the flight further. This recital may be considered as helping to "interpret the enabling provision" of Article 6 .

NORMATIVE PROVISION:
Article 9
Right to care
1. Where reference is made to this Article, passengers shall be offered free of charge:
(a) meals and refreshments in a reasonable relation to the waiting time;
(b) hotel accommodation in cases
- where a stay of one or more nights becomes necessary, or
- where a stay additional to that intended by the passenger becomes necessary;
(c) transport between the airport and place of accommodation (hotel or other).

RECITAL:
(18) Care for passengers awaiting an alternative or a delayed flight may be limited or declined if the provision of the care would itself cause further delay.

9.3 AUTOMATED MAPPING

Experiments were conducted on mapping (automatically) normative provisions to recitals. Each recital item contained all the text without its index number. Similarly, all index numbers were removed from the normative provision items. Otherwise, no further normalisation was taken to handle lists, references etc. (as described in Chapter 6). Substantive titles were included as if they were normal sentences, but non-informative structural titles such as 'Section II' and 'Article 1' were removed, although such terms were included when used as references within the sentences of the articles. For our experiments, we used the Cosine Similarity algorithm with Term Frequency - Inverse Document Frequency (tf-idf)[148]. Each normative provision and recital was presented as a vector of terms, and the Cosine Similarity between two vectors were quantified as the cosine of the angle between the two vectors [174]. Each term in the vector was weighted using tf-idf, a measure designed to evaluate the importance of each term in the vector, offsetting the frequency of a term in the vector with its frequency in the corpus as a whole.

From this baseline, experiments were conducted in order to improve the performance of our mapping tool. We had observed from the gold standard that many recitals that correspond to articles use the words of the substantive title in the recital text. We thus sought to give greater weight to terms appearing in substantive titles in determining similarity. However, the title words "scope" and "objective" were not given extra weight as they are effectively metadata rather than substantive terms. We also experimented with different threshold levels in order to ascertain whether the True Positives have a higher similarity than the False Positives, so that the threshold can be adjusted without compromising recall. We then experimented with giving extra weight to the title tokens (with a multiplication factor of 3), based on our observation that relevant recitals often use these terms. Our final experiment was to lemmatise words and remove from the vectors terms having what are generally considered to be non-informative parts of speech, retaining only the nouns, verbs and adverbs, using the Stanford part-of-speech parser [41]. The first experiment was to remove all tokens apart from the nouns, and then lemmatise those nouns. The results are not indicated here, but were the poorest, as the system lost some important features. By using the lemmas of all nouns and verbs and adjectives the system performs a little better. However, the best results of all was with the lemmatisation of all terms whatever their part of speech.

Our experiments supported our hypothesis that in the studied texts, there is often textual similarity between normative provisions and related recitals, such that automated similarity methods can be effective. The results are discussed in detail in Chapter 10.

10

EVALUATION

This chapter describes the evaluation of systems developed for this thesis. Section 10.1 provides results for normalising legislative text, focussing on the identification of titles and list elements. Section 10.2 provides results for automated extraction of definitions and norms based on Semantic Role Labeling and dependency parsing. Section 10.3 provides results for different experiments in mapping normative provisions and recitals with Cosine Similarity and tf-idf: adjusting the thresholds, giving extra weight to the title tokens, and lemmatising words and removing from the vectors terms having what are generally considered to be non-informative parts of speech.

10.1 NORMALISATION

The rule-based normalisation system was trained on twenty directives and tested on one unseen directive, Directive 07/46/EC[1].

The key challenges were to identify titles, and to identify lists and reconstruct them as standalone sentences. In the test directive, there were 79 sentences that could be considered as list sentences, as they were preceded by an index. However, only 43 of these were prefixed by a pre-list or end-of-list clauses. Most errors occurred in nested lists. The joining of list items with introductory and end-of-list clauses could be improved with greater attention to punctuation. For this research, colons or commas were added to introductory list clauses before they were joined with list items, but this occasionally created problems for semantic role labeling. Further work is required to determine when and how punctuation should be changed.

Several lists contained conjunctions and these were extracted, mostly correctly, to determine the relationship between list items. Future work will take into account the impact of negatives on conjuntions, as highlighted by Buabuchachart et al. [38].

Title identification is important not only to eliminate text that are not norms, but also to identify the beginning and ending of lists (titles do not occur within lists). Notwithstanding the trivial identification of structural titles ('Article 1' etc) or titles in capital letters, the system used a simple approach to identifying semantic titles: absence of verbs. This works reasonably well, as can be seen in tables 3 and 4.

1 Directive 07/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

Table 3: Accuracy of normalisation

Element	Fully Correct	Partially Correct	Wrong (False Positive)	Missing (False Negative)
Title	79	0	0	5
Non-capitalised semantic title	14	0	0	5
Reference	110	8	0	3
No-list sentence	126	0	0	6
List sentence	69	10	0	0
Start-of-list clause	42	0	0	1
End-of-list clause	34	0	1	8
Conjunction	73	0	5	0

Tables 3 and 4[2] also show very good performance for the identification of in-text references. Transformation of long names for legislation into one-word references can help reduce noise into the SRL system.

The system performs reasonably well in general, and is surely more efficient than manual manipulation. Moreover, the XML format allows flexibility of representation for different purposes, whilst allowing list sentences to be constructed easily. Any errors can be manually identified and corrected in a fraction of the time required to create these intermediate representations from scratch. An output of initial normalised sentences can be found in Appendix A.

10.2 EXTRACTING DEFINITIONS AND NORMS

Before evaluating the performance of the system for extracting definitions and norms, it is worth taking into account the performance of the Mate Tools Semantic Role Labeler on legislative text. The system was tested on 224 sentences from Directive 95/46/EC[3], 58 definitions and 166 norms. For each sentence, the arguments for all the verb predicates in the sentence were evaluated, and the overall sentence was evaluated as accurate if all the arguments for all the verbs were correct. 78.5% of definitions had correct arguments for all verbs. However, only 52% of norms had the same. Generally, norms are more complex, and therefore more errors are introduced. Nevertheless, not all errors have

2 F-measure is calculated using Precision and Recall decimal values to 17 decimal points

3 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

Table 4: Precision, Recall and F-measure of elements of definitions and norms

Element	Precision (Strict)	Recall (Strict)	F-Measure (Strict)	Precision (Lenient)	Recall (Lenient)	F-Measure (Lenient)
Title	100%	94.05%	94.05%	100%	94.05%	96.93%
Non-capitalised semantic title	100%	77.78%	87.50%	100%	77.78%	87.50%
Reference	93.22%	97.35%	95.24%	100%	97.52%	97.56%
No-list sentence	95.45%	100%	97.67%	95.45%	100%	97.67%
List sentence	87.34%	100%	93.24%	100%	100%	100%
Start-of-list clause	100%	97.67%	98.82%	100%	97.67%	98.82%
End-of-list clause	97.14%	80.95%	88.31%	97.14%	80.95%	88.31%
Conjunction	88.37%	100%	93.83%	88.37%	100%	93.83%

Table 5: Precision, Recall and F-measure of definitions and norms: preliminary experiment

Type	Number	Precision (Strict)	Recall (Strict)	F-Measure (Strict)	Precision (Lenient)	Recall (Lenient)	F-Measure (Lenient)
Obligation	36	75.24%	77.45%	76.33%	75.24%	77.45%	76.33%
Power	13	88.64%	61.90%	72.90%	95.35%	76.92%	85.15%
Legal Effect	10	53.33%	25.81%	34.78%	53.33%	25.81%	34.78%
Definition	8	100%	87.10%	93.10%	100%	87.10%	93.10%
Permission	6	52.94%	90%	66.67%	52.94%	90%	66.67%
Scope	6	73.91%	77.27%	75.55%	94.74%	90%	92.31%
Rationale	6	100%	100%	100%	100%	100%	100%
Right	2	100%	100%	100%	100%	100%	100%
Exception	1	100%	100%	100%	100%	100%	100%
Hierarchy	1	100%	100%	100%	100%	100%	100%

consequences for the definition and norm extraction system, since only certain predicates are used by the system.

The approach described in Chapter 7 and Chapter 8 was initially trained on fourteen directives and tested on Directive 98/5/EC[4], excluding the preamble and annexes. Three sentences were discarded with corrupted output from the normalisation module.

Table 5[5] shows the results. Here, the strict (S) results take partially correct results as wrong whereas the lenient (L) results take them as being correct (the latter is probably more appropriate since the work in any case should be checked by a knowledge engineer). These tests were conducted on legislation that was also used to develop the system, which accounts for the exceptionally high results. What this does show is the degree of linguistic consistency across one piece of legislation, which shows the potential for semi-automated definition and norm extraction.

4 Directive 98/5/EC of the European Parliament and of the Council of 16 February 1998 to facilitate practice of the profession of lawyer on a permanent basis in a Member State other than that in which the qualification was obtained

5 F-measure is calculated using Precision and Recall decimal values to 17 decimal points

The system was then trained on ten directives and tested on hitherto unseen legislation, Directive 95/46/EC[6], excluding the preamble and annexes. Here, almost all extracted norms were partially correct. To understand this better, table 6 and table 7[7] show the quantified results for each norm element. Again, the strict (S) results take partially correct results as wrong whereas the lenient (L) results take them as being correct, bearing in mind that the latter is probably more appropriate since the work in any case should be checked by a knowledge engineer. Appendix B contains evaluation of the first extracted definitions and norms.

The norm type field is relevant to all norms. Possible outputs were: definition, obligation, permission, power, scope, right, hierarchy, exception, legal effect and unknown. All the norms classified as unknown were in fact, actual norms of the relevant type, apart from one proclamation which were not sought in the program. The evaluation also revealed a number of norms that should be classed as liability, which are potential obligations arising from the power of another to impose an obligation. This reflects a level of uncertainty about whether such an obligation will arise. On the other hand, it could be argued that obligations arising from the obligation of another to impose an obligation have a greater level of certainty and should be (and have) been classed as obligations. Most of the errors in determining the norm type (36%) arose from mistaking powers for permissions. The problem is that both types of norms have the modal verb 'may'. The module sought to deal with this by identifying 'power' verbs that follow the modal, based on the corpus used to develop the system. However, the evidence of this evaluation shows that this is less than satisfactory. For EU legislation, it can be assumed that almost all norms involving the modal 'may' and having a member state as an active role are powers. But this is not a general solution to the problem. Also common were the misclassifying of obligations as unknown, or the extraction of too many obligations from the sentence. Some obligations were missclassified as legal effect or scope.

The other elements in Table 6 and table 7 are definiendum, definiens, includes and excludes, which are all elements that pertain to definitions. The elements action, active role, passive role, condition, timeframe, exception and reason pertain to norms of the type obligation, permission, power and right. The elements situation and result pertain to meta-norms of the type legal effect. The elements object, excludes object and active role pertain to meta-norms of the type scope. The elements higher priority and oower priority pertain to meta-norms

6 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

7 F-measure is calculated using Precision and Recall decimal values to 17 decimal points

Table 6: Accuracy of SRL extraction of elements of definitions and norms

Element	**Fully Correct**	**Partially Correct**	**Wrong (False Positive)**	**Missing (False Negative)**	**Misclassified as Different Element (False Negative)**
Norm Type	184	N/A	66	17	N/A
Definiendum	6	8	0	2	0
Definiens	5	3	0	1	0
Includes	5	1	0	0	0
Excludes	0	0	0	3	10
Action	46	113	56	40	2
Active Role	113	6	10	33	10
Passive Role	13	15	5	90	4
Condition	30	41	35	38	13
Timeframe	11	11	9	6	0
Exception	4	13	1	7	7
Scope	2	2	6	9	0
Situation	0	0	5	1	0
Result	0	0	3	1	0
Object	1	0	0	0	0
ExcludesObject	0	3	2	3	0
HigherPriority	0	0	3	0	3
LowerPriority	0	0	0	0	0

Table 7: Precision, Recall and F-measure of extraction of elements of definitions and norms

Element	Precision (Strict)	Recall (Strict)	F-Measure (Strict)	Precision (Lenient)	Recall (Lenient)	F-Measure (Lenient)
Norm Type	73.60%	91.54%	81.60%	73.60%	91.54%	81.60%
Definiendum	42.86%	75.00%	54.55%	100%	87.5%	93.33%
Definiens	62.50%	83.33%	71.43%	100%	88.89%	94.12%
Includes	83.33%	100%	90.91%	100%	100%	100%
Excludes	0%	0%	0%	0%	0%	0%
Action	21.40%	52.27%	30.36%	73.95%	79.10%	76.44%
Active Role	87.60%	72.44%	79.30%	92.25%	73.46%	81.79%
Passive Role	39.39%	12.15%	18.57%	84.85%	22.95%	36.13%
Condition	28.30%	37.04%	32.09%	66.98%	58.20%	62.28%
Timeframe	35.48%	64.71%	45.82%	70.97%	78.57%	74.58%
Exception	22.22%	22.22%	22.22%	94.44%	54.84%	69.39%
Reason	20.00%	18.18%	19.05%	40.00%	30.77%	34.78%
Situation	0%	0%	0%	0%	0%	0%
Result	0%	0%	0%	0%	0%	0%
Object	100%	100%	100%	100%	100%	100%
Excludes-Object	0%	0%	0%	60.00%	50.00%	54.55%
Higher-Priority	0%	0%	0%	0%	0%	0%
Lower-Priority	0%	0%	0%	0%	0%	0%

of the type hierarchy. There were few meta-norms in the legislation evaluated.

The results are very varied, and shows that further work is required to achieve acceptable results for a comprehensive norm extraction system. Nevertheless, the most important elements - norm type, active role - are obtained with good Accuracy, which in itself should help compliance officers or related personnel in their most important search i.e. which obligations need to complied with.

One significant weakness, however, is the poor performance of the system on identifying passive roles. Moreover, apart from their identification, there are two aspects that require further consideration. Firstly, how to distinguish between beneficiaries of norms and other passive roles, such as agents who play an active role in a condition or exception. Secondly, how to relate the passive roles to the relevant parts of the norm. An example of a particularly intricate norm from this point of view is the following:

```
Any person acting under the authority of the controller or
of the processor, including the processor himself, who has
access to personal data must not process them except on
instructions from the controller, unless he is required to
do so by law.
```

The high partially correct results for the action element reveals that it suffers the most from boundary errors. Boundary errors are also a problem for conditions, timeframes, exceptions and reasons. This is one particular area where the output of the SRL system is particularly disappointing. However, even when supplemented by pattern-matching, problems remain. 30 fully correct conditions were identified via the SRL output as opposed to 21 via pattern-matching (keywords such as 'where' or 'when'). 41 partially correct were identified via SRL, 34 via pattern-matching. 35 elements were wrongly classified as conditions via SRL, 27 via pattern-matching. 38 conditions were missed altogether and 13 conditions were wrongly classified as something else. The situation is similar for timeframes and exceptions, although there are fewer of those in the legislation evaluated. Some improvement could be made by deeper analysis of dependency trees. However, many of these problems arose due to problems with linking different elements of lists in the normalisation module, and this needs to be looked at further.

The low incidence of scope and hierarchy norm types in this particular legislation makes it difficult to provide a proper evaluation of relevant elements.

10.3 MAPPING NORMATIVE PROVISIONS AND RECITALS

Following manual mapping of normative provisions and recitals (see Appendix C for the gold standard of Directive 95/46/EC), experiments

Table 8: Baseline: mapping of three legislation with different thresholds

Measure	CD with CT 0.1	CD with CT 0.16	DPD with CT 0.1	DPD with CT 0.16	ATPR with CT 0.1	ATPR with CT 0.16
TP	24	19	48	36	15	13
TN	452	499	2134	2273	295	329
FP	66	19	182	43	58	24
FN	7	12	12	24	7	9
Accuracy	86.70%	94.35%	91.84%	97.18%	82.67%	91.20%
Precision	26.67%	50.00%	20.87%	45.57%	20.55%	35.14%
Recall	77.42%	61.29%	80.00%	60.00%	68.18%	59.09%
F-measure	39.67%	55.07%	33.10%	51.80%	31.58%	44.07%
TNR	87.26%	96.33%	92.14%	98.14%	83.57%	93.20%

were conducted to map them automatically. Table 8[8] shows the results of the first experiment: Cosine Similarity with tf-idf on all words in their surface forms. In the following tables, CD means the copyright directive, DPD means the data protection directive, ATPR means the air traffic regulation, CT means Cosine Threshold, TP means True Positive, TN True Negative, FP False Positive, FN False Negative and TNR True Negative Rate. We can see that the Accuracy level is very high, ranging from 82.67% to 97.18%. However, it is acknowledged that Accuracy is not a fair way to measure the quality of a system in case of unbalanced datasets (i.e. datasets that have very different numbers of "positives" compared to "negatives"). Other classic measures of performance for similarity are Precision, Recall, and the F-measure, which seeks a balance between the two.

Different threshold levels were experimented with in order to ascertain whether the True Positives have a higher similarity than the False Positives, so that the threshold can be adjusted without compromising Recall. However, it was found that this was not the case with our data. Indeed, this may lie in the nature of the relationship between Precision and Recall in general. Precision and Recall creates a curve (called Receiver Operating Characteristic (ROC) Curve) which is a way to evaluate them considering all the similarity threshold. There is a threshold point in this curve that maximises the F-measure, and it is sometimes the best combination of Precision and Recall one can have. However, it is suggested that the classic F-measure is also a poor mea-

8 F-measure is calculated using Precision and Recall decimal values to 17 decimal points

sure for evaluating performance in our case. Precision is the important measure for systems that require few and precise information without any manual analysis, whereas Recall is arguably more important than Precision to support legal knowledge engineers. For them, it is more important to identify as many of the relevant connections as possible, even within a noisy set of possibilities, rather than to identify very few precise connections.

It should be noted that the drawback of Precision and Recall is that they avoid evaluating the ability to identify "negatives" since they are calculated using only True Positives, False Positives and False Negatives (and not True Negatives). It is therefore useful also to consider the True Negative Rate (TNR). For instance, the baseline measure with a threshold of 0.1 on the copyright directive is able to identify around 24 positive connections, but within a quite large set of 66 False Positives. "Large", however, is not as large as the entire set of connections. This means in the case of the copyright directive that the knowledge engineer has to manually go over 24+66=90 connections instead of analysing the entire set of 549 connections, which works out at having to check 10 recitals per article, instead of 61.

Another experiment was made (Table 9[9]) giving extra weight to the title tokens (with a multiplication factor of 3), based on our observation that relevant recitals often use these terms. In two out of three legislation, the weighting of the title tokens produced some improvement in terms of Precision and F-measure. However, the weighting for the titles was arbitrary, and more experiments are required in this vein. This experiment could be followed up with strategies of automatically detecting the structural parts of the texts that need to be "boosted". For instance, the first sentences of the texts may also have a greater weight.

The final experiment (Table 10[10]) was to lemmatise words and remove from the vectors terms having what are generally considered to be non-informative parts of speech, using the Stanford part-of-speech parser [41]. The first experiment was to remove all tokens apart from the nouns, and then lemmatise those nouns. The results are not indicated here, but were the poorest, as the system lost some important features. There are important facts to consider behind the concept of Cosine Similarity. It follows a specific curve which depends on the number of features of the vectors. In general, the larger the number of features, the more the sensitivity of Cosine Similarity. By using the lemmas of all nouns and verbs and adjectives, the system performs a little better. However, the best results of all was had with the lemmatisation of all terms whatever their part of speech.

9 F-measure is calculated using Precision and Recall decimal values to 17 decimal points

10 F-measure is calculated using Precision and Recall decimal values to 17 decimal points

Table 9: Mapping with extra weight for title terms

Measure	CD with CT 0.1	CD with CT 0.1 with extra weighting to title terms	DPD with CT 0.1	DPD with CT 0.1 with extra weighting to title terms	ATPR with CT 0.1	ATPR with CT 0.1 with extra weighting to title terms
TP	24	21	48	46	15	13
TN	452	499	2134	2151	295	317
FP	66	69	182	165	58	36
FN	7	10	12	14	7	9
Accuracy	86.70%	86.81%	91.84%	92.47%	82.67%	88.00%
Precision	26.67%	23.33%	20.87%	21.80%	20.55%	26.53%
Recall	77.42%	67.74%	80.00%	76.67%	68.18%	59.09%
F-measure	39.67%	34.71%	33.10%	33.95%	31.58%	36.62%
TNR	87.26%	87.85%	92.14%	92.88%	83.57%	89.80%

Table 10: Mapping with lemmatisation and filtering on parts of speech

Measure	CD with CT 0.1 with lemmas of all the POS	CD with CT 0.1 with lemmas of nouns, adjectives and verbs	DPD with CT 0.1 with lemmas of all the POS	DPD with CT 0.1 with lemmas of nouns, adje= ctives and verbs	ATPR with CT 0.1 with lemmas of all the POS	ATPR with CT 0.1 with lemmas of nouns, adjectives and verbs
TP	27	21	49	51	14	15
TN	426	423	2057	2054	301	295
FP	92	95	259	262	52	58
FN	4	10	11	9	8	7
Accuracy	82.51%	80.87%	88.64%	88.59%	84.00%	82.67%
Precision	22.69%	18.10%	15.91%	16.29%	21.21%	20.55%
Recall	87.10%	67.74%	81.67%	85.00%	63.64%	68.18%
F-measure	36.00%	28.57%	26.63%	27.35%	31.82%	31.58%
TNR	82.24%	81.66%	88.82%	88.69%	85.27%	83.57%

In summary, these experiments support the hypothesis that there is often textual similarity between normative provisions and related recitals, such that automated similarity methods can be effective. The performance of the system depends on which evaluation metric is used, but with a bias towards Recall, good initial results are reached, which in practical terms means that a knowledge engineer can be presented with almost all the plausible connections without having to cross-check all possible connections.

11

CONCLUSIONS AND FUTURE WORK

This chapter provides conclusions and contributions of the thesis, as well as possibilities for future work.

This thesis endorses Sartor's [162, page 7] vision of a future legal semantic web where legal content on the web is enriched with machine processable information that can be "presented in many different ways, according to the different issues at stake and the different roles played by its users" by focussing on extracting and linking structured legal information for compliance purposes, building on research in legal theory, ontologies and NLP. In addition to a thorough description of methodology and results, the thesis highlights important challenges that have emerged from this research which will hopefully prove useful for future work in the extraction and linking of information from legislative text.

The terminological and frame-based legal ontologies of the Eunomos legal document management system [30] and LEGAL-URN compliance management system [68] described in Chapter 4 provided the starting-point for determining the kind of information required by individuals and organisations from legislation in a compliance setting, and how to structure norms for greater accessibility in terms of understanding and improved search capabilities (RQ1). These representations were then analysed and reviewed in the light of findings from manual corpus analysis as well as EU drafting guidelines, legal theory and deontic logic (RQ1a, RQ1b, RC1c), as described in Chapter 5. Chapter 3 explored the challenges of legislative text for machine understanding. Key structural challenges of legislative text were addressed by normalisation, as described in Chapter 6, including transforming lists into proper grammatical sentences (RQ2). To address the resource bottleneck problem of creating specialist knowledge management systems, in particular how to semi-automate the extraction of definitions, norms and their elements to populate legal ontologies (RQ3), Chapter 7 and Chapter 8 described a definition and norm extraction system using semantic representations from a general-purpose SRL module [25], as well as dependency parse trees from the same system. This solution was pursued in order to simplify the sets of rules required to identify possible norms and definitions, classify their types, and map arguments to domain-specific slots in a legal ontology. Chapter 9 investigated the impact of recitals on the interpretation of legislative norms (RQ4). Experiments based on the Cosine Similarity algorithm

with Term Frequency - Inverse Document Frequency (tf-idf) supported the hypothesis that many links between normative provisions could be made on the basis of text similarity. This work, as well as the systems developed for normalisation and extraction (RQ5), were evaluated in Chapter 10.

There are a number of observations that have emerged from this evidence-based research. While much theoretical work on norms have focussed on obligations, and there are indeed plenty of them in the legislation studied, there are also certain kinds of norms that are less prominent in the literature but are, nevertheless, important to cover in a comprehensive norm extraction system. EU legislation in particular are full of norms bestowing a power to impose certain obligations (as well as obligations to impose obligations). It is a moot point whether the secondary obligations then should be considered liabilities or obligations, particularly since there is a greater element of uncertainty where the secondary obligations derive from a power rather than an obligation. Legislation also contain a number of meta-norms indicating scope, hierarchy among norms and exceptions, and these are arguably essential information to extract for an ontology of legal norms. Statements indicating the legal effect of a particular situation may not be norms in the strictest sense, but have important implications and therefore should also be included. As for elements of norms, many statements refer to the scope of the norm. In addition to the condition element used in Eunomos and LEGAL-URN, it is submitted that a time-frame element can also be of value. One element that requires further investigation is passive role, since this research found that there are many different kinds of passive roles, not only beneficiaries, and it would be useful to distinguish between them. Another element that requires further consideration is the condition element. There are a number of constructions in legislative text regarding the applicability of norms that are somewhat similar to conditions, but have different effects. For instance, 'in particular' indicates that a norm applies in a particular scenario, but is not limited to that scenario. 'Notwithstanding' implies that what might be considered as a negative condition does not in fact apply.

Other future work that follows from this research are

- improving the normalisation methodology to provide better linking between different parts of a sentence, since a reliance on commas as a default linking mechanism can lead to parsing and SRL errors.
- integrating the normalisation work with a legislative XML module, either by using the normalisation output in XML as input to the legislative XML parser, or taking the legislative XML as input

to the normalisation module, thereby ensuring maximum connection, traceability and transparency between different views of the legislation.

- intra- and extra-sentential anaphora resolution specifically for legislative text. There are particular challenges and opportunities here. Anaphora can take the form of pronouns, abbreviated forms, or references to entities in other articles. A related issue is how to determine 'missing' elements of a norm e.g. for normative sentences that lack an active role, it is often the case that the role is implicitly carried over from the previous sentence.

- detection of is-a and part-of relationships among defined entities, and in particular, detection of 'example' type definitions found in sentences containing the construction 'such as'.

- generation of different views of a norm based on Hohfeldian correlatives, such as the rights that can be implicitly derived from obligations, since legislative text will usually provide only one view explicitly.

- rigorous analysis of the relationship between norms to determine norms that may be satisfied in alternative ways, or must be satisfied as a group, taking into account the impact of negatives on conjunctions, as highlighted by Buabuchachart et al. [38].

- the development of SRL-based information extraction modules for other jurisdictions and other languages;

- identification of related norms in different legislation (even different jurisdictions). This thesis suggests two possible approaches: firstly, text similarity based on the unstructured text of the original norms, secondly, comparison of structured norms. In both cases, semantic similarity approaches such as Latent Semantic Analysis [112] or use of WordNet [170] may help identify semantic similarities in the absence of terminological equivalence.

BIBLIOGRAPHY

[1] Stanford core nlp. http://nlp.stanford.edu:8080/corenlp/.

[2] Legislative drafting, a commission manual. http://ec.europa.eu/smartregulation/better_regulation/documents/legis_draft_comm_en.pdf, 1997.

[3] Manual of precedents for acts established within the council of the european union, office of official publications of the european communities. http://publications.europa.eu/code/en/en-120200.htm, 2002.

[4] Statutory interpretation: General principles and recent trends. http://www.fas.org/sgp/crs/misc/97-589.pdf, 2011.

[5] When 'and' really means 'or' in a restrictive covenant. http://www.drinkerbiddle.com/resources/publications/2013/-When-and-Really-Means-or-in-a-Restrictive-Covenant-, 2013.

[6] Handbook on european data protection law, european union agency for fundamental rights, publications office of the european union. http://fra.europa.eu/sites/default/files/fra-2014-handbook-data-protection-law-2nd-ed_en.pdf, 2014.

[7] Joint practical guide of the european parliament, the council and the commission for persons involved in the drafting of european union legislation. http://eur-lex.europa.eu/content/techleg/KB0213228ENN.pdf, 2014.

[8] Salah Aıt-Mokhtar, Veronika Lux, and Eva Bánik. Linguistic parsing of lists in structured documents. In *Proceedings of the EACL Workshop on Language Technology and the Semantic Web (3rd Workshop on NLP and XML, NLPXML-2003), Budapest, Hungary*, volume 15, page 96. Citeseer, 2003.

[9] Gianmaria Ajani, Leonardo Lesmo, Guido Boella, Alessandro Mazzei, and Piercarlo Rossi. Terminological and ontological analysis of european directives: multilinguism in law. In *The 11th International Conference on Artificial Intelligence and Law, Proceedings of the Conference (ICAIL)*, pages 43–48. ACM, 2007.

[10] Gianmaria Ajani, Guido Boella, Leonardo Lesmo, Marco Martin, Alessandro Mazzei, Daniele P Radicioni, and Piercarlo Rossi. *Multilevel legal ontologies*. Springer, 2010.

[11] Mark Anderson. Hanged on a comma: drafting can be a matter of life and death. https://ipdraughts.wordpress.com/2013/10/14/hanged-on-a-comma-drafting-can-be-a-matter-of-life-and-death/, 2013.

[12] Grigoris Antoniou. *A semantic web primer*. the MIT Press, 2004.

[13] Olufunmilayo Arewa. Open access in a closed universe: Lexis, westlaw, law schools, and the legal information market. *Lewis & Clark Law Review*, 10:797, 2006.

[14] AiTi Aw, Min Zhang, Juan Xiao, and Jian Su. A phrase-based statistical model for sms text normalization. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 33–40. Association for Computational Linguistics, 2006.

[15] DC Balmer. Trends and issues in platform lift. In *Proceedings of Space Requirements for Wheeled Mobility Workshop*, pages 9–11, 2003.

[16] Gioele Barabucci, Luca Cervone, Angelo Di Iorio, Monica Palmirani, Silvio Peroni, and Fabio Vitali. Managing semantics in xml vocabularies: an experience in the legal and legislative domain. In *Proceedings of Balisage: The markup conference*, volume 5, 2010.

[17] Alberto Barrón-Cedeno, Paolo Rosso, Eneko Agirre, and Gorka Labaka. Plagiarism detection across distant language pairs. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 37–45. Association for Computational Linguistics, 2010.

[18] Matthew Berland and Eugene Charniak. Finding parts in very large corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 57–64. Association for Computational Linguistics, 1999.

[19] Anderson Bertoldi and Rove Luiza de Oliveira Chishman. The limits of using framenet frames to build a legal ontology. In *ONTOBRAS-MOST*, pages 207–212. Citeseer, 2011.

[20] C. Biagioli, E. Francesconi, A. Passerini, S. Montemagni, and C. Soria. Automatic semantics extraction in law documents. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, ICAIL '05, New York, NY, USA, 2005. ACM.

[21] Carlo Biagioli, Enrico Francesconi, P Spinosa, and Mirco Taddei. A legal drafting environment based on formal and semantic xml standards. In *Proceedings of the 10th international conference on Artificial intelligence and law*, pages 244–245. ACM, 2005.

[22] Marco Bianchi, Mauro Draoli, Giorgio Gambosi, Maria Teresa Pazienza, Noemi Scarpato, and Armando Stellato. Ict tools for the discovery of semantic relations in legal documents. In *Proceedings of the 2nd International Conference on ICT Solutions for Justice (ICT4Justice)*, 2009.

[23] Chris Biemann. Ontology learning from text: A survey of methods. In *LDV forum*, volume 20, pages 75–93, 2005.

[24] Steven Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics, 2006.

[25] Anders Björkelund, Love Hafdell, and Pierre Nugues. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 43–48. Association for Computational Linguistics, 2009.

[26] Hans Christian Boas. Bilingual framenet dictionaries for machine translation. In *LREC*, 2002.

[27] Philip Bobbitt. *Constitutional Interpretation*. Oxford University Press,, 2006. ISBN 9780195189568. URL http://books.google.lu/books?id=UDZqAAAACAAJ.

[28] G. Boella, L. di Caro, L. Humphreys, and L. Robaldo. Using legal ontology to improve classification in the eunomos legal document and knowledge management system. In *Semantic Processing of Legal Texts (SPLeT) at LREC12*, 2012.

[29] Guido Boella and Leendert van der Torre. Permissions and obligations in hierarchical normative systems. In *Proceedings of the 9th international conference on Artificial intelligence and law*, pages 109–118. ACM, 2003.

[30] 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*, pages 131–146. Springer, 2012.

[31] Guido Boella, Luigi Di Caro, and Livio Robaldo. Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 218–225. Springer, 2013.

[32] Guido Boella, Silvano Colombo Tosatto, Sepideh Ghanavati, Joris Hulstijn, Llio Humphreys, Robert Muthuri, André Rifaut, and

Leendert van der Torre. Integrating legal-urn and eunomos: towards a comprehensive compliance management solution. In *AI Approaches to the Complexity of Legal Systems*, pages 130–144. Springer, 2014.

[33] Guido Boella, Luigi Di Caro, Michele Graziadei, Loredana Cupi, Carlo Emilio Salaroglio, Llio Humphreys, Hristo Konstantinov, Kornel Marko, Livio Robaldo, Claudio Ruffini, Kiril Simov, Veli Stroetmann, and Andrea Violato. Linking legal open data: breaking the accessibility and language barrier in european legislation and case law. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 171–175. ACM, 2015.

[34] Dario Bonino and Fulvio Corno. Dogont-ontology modeling for intelligent domotic environments. In *The Semantic Web-ISWC 2008*, pages 790–803. Springer, 2008.

[35] T. D. Breaux:. *Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems*. PhD thesis, North Carolina State University, Raleigh (NC), 2009.

[36] Travis D Breaux and Annie I Antón. Analyzing goal semantics for rights, permissions, and obligations. In *Requirements Engineering, 2005. Proceedings. 13th IEEE International Conference on*, pages 177–186. IEEE, 2005.

[37] Eric Brill. A simple rule-based part of speech tagger. In *Proceedings of the workshop on Speech and Natural Language*, pages 112–116. Association for Computational Linguistics, 1992.

[38] Alan Buabuchachart, Nina Charness, Katherine Metcalf, and Leora Morgenstern. Automated methods for extracting and expanding lists in regulatory text. In *Proceedings of the First JURIX Doctoral Consortium and Poster Sessions in conjunction with the 26th International Conference on Legal Knowledge and Information Systems, JURIX 2013, Bologna, Italy, December 11-13, 2013.*, 2013.

[39] Julian Burnside. R v roger casement. http://www.julianburnside.com.au/law/r-v-roger-casement/.

[40] Zhiming Cai and Eric S. K. Yu. Addressing performance requirements using a goal and scenario-oriented approach. In *CAiSE 2002*, volume 2348 of *LNCS*, pages 706–710. Springer, 2006.

[41] Danqi Chen and Christopher D Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, volume 1, pages 740–750, 2014.

[42] Monojit Choudhury, Rahul Saraf, Vijit Jain, Animesh Mukherjee, Sudeshna Sarkar, and Anupam Basu. Investigation and modeling of the structure of texting language. *International Journal of Document Analysis and Recognition (IJDAR)*, 10(3-4):157–174, 2007.

[43] Massimiliano Ciaramita, Giuseppe Attardi, Felice Dell'Orletta, and Mihai Surdeanu. Desrl: A linear-time semantic role labeling system. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 258–262. Association for Computational Linguistics, 2008.

[44] Rudi L Cilibrasi and Paul MB Vitanyi. The google similarity distance. *IEEE Transactions on knowledge and data engineering*, 19 (3):370–383, 2007.

[45] Cleo Condoravdi. Temporal interpretation of modals - modals for the present and for the past. In *The construction of meaning*. Citeseer, 2001.

[46] Paul Cook and Suzanne Stevenson. An unsupervised model for text message normalization. In *Proceedings of the workshop on computational approaches to linguistic creativity*, pages 71–78. Association for Computational Linguistics, 2009.

[47] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[48] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[49] H Cunningham, D Maynard, K Bontcheva, and V Tablan. Gate: A framework and graphical development environment for robust nlp tools and applications. In *Proc. 40th Anniversary Meeting of the Association for Computational Linguistics (ACL)*, 2002.

[50] James R. Curran, Stephen Clark, and Johan Bos. Linguistically motivated large-scale NLP with c&c and boxer. In *ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic*, 2007.

[51] Francesco Rossi Dal Pozzo. *EU Legal Framework for Safeguarding Air Passenger Rights*. Springer, 2014.

[52] Per-Erik Danielsson. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980.

[53] Ferdinand De Saussure and Wade Baskin. *Course in general linguistics [1916]*. Columbia University Press, 2011.

[54] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[55] Jacob Eisenstein. What to do about bad language on the internet. In *HLT-NAACL*, pages 359–369, 2013.

[56] Marwane El Kharbili and Elke Pulvermüller. A semantic framework for compliance management in business process management. In *BPSC*, pages 60–80, 2009.

[57] Marwane El Kharbili, Qin Ma, Pierre Kelsen, and Elke Pulvermueller. Corel: Policy-based and model-driven regulatory compliance management. In *Enterprise Distributed Object Computing Conference (EDOC), 2011 15th IEEE International*, pages 247–256. IEEE, 2011.

[58] Mathieu-Henri Falco, Véronique Moriceau, and Anne Vilnat. Kitten: a tool for normalizing html and extracting its textual content. In *LREC*, pages 2261–2267. Citeseer, 2012.

[59] M. Fernandez-Barrera and P. Casanovas. Towards the intelligent processing of non-expert generated content: Mapping web 2.0 data with ontologies in the domain of consumer mediation. In *Proceedings of International Conference on Artificial Intelligence and Law Workshop, Applying Human Language Technology to the Law*, pages 18–27, 2011.

[60] Charles J Fillmore. The case for case. 1968.

[61] Charles J Fillmore. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32, 1976.

[62] Charles J Fillmore, Josef Ruppenhofer, and Collin F Baker. Framenet and representing the link between semantic and syntactic relations. *Frontiers in linguistics*, 1:19–59, 2004.

[63] Luciano Floridi. *The Blackwell guide to the philosophy of computing and information*. Wiley.com, 2008.

[64] Enrico Francesconi. Legal rules learning based on a semantic model for legislation. In *Workshop Programme*, page 46, 2010.

[65] Enrico Francesconi. Axioms on a semantic model for legislation for accessing and reasoning over normative provisions. 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*, pages 147–161. Springer, 2011.

[66] Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening ontologies with dolce.

In *International Conference on Knowledge Engineering and Knowledge Management*, pages 166–181. Springer, 2002.

[67] Xibin Gao and Munindar P. Singh. Extracting normative relationships from business contracts. In *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, pages 101–108, 2014.

[68] Sepideh Ghanavati. Legal-URN framework for legal compliance of business processes. http://hdl.handle.net/10393/24028, 2013. Ph.D. thesis, UOttawa, Canada.

[69] Sepideh Ghanavati, Daniel Amyot, and Liam Peyton. Towards a framework for tracking legal compliance in healthcare. In *Advanced Information Systems Engineering*, pages 218–232. Springer, 2007.

[70] Sepideh Ghanavati, Daniel Amyot, and Liam Peyton. A systematic review of goal-oriented requirements management frameworks for business process compliance. In *Requirements Engineering and Law (RELAW), 2011 Fourth International Workshop on*, pages 25–34. IEEE, 2011.

[71] Daniel Gildea. Corpus variation and parser performance. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, pages 167–202, 2001.

[72] Sartor Giovanni. Legal reasoning: A cognitive approach to the law, 2007.

[73] Claudio Giuliano, Alberto Lavelli, and Lorenza Romano. Exploiting shallow linguistic information for relation extraction from biomedical literature. In *EACL*, volume 18, pages 401–408. Citeseer, 2006.

[74] Alfio Massimiliano Gliozzo, Claudio Giuliano, and Raffaella Rinaldi. Instance pruning by filtering uninformative words: an information extraction case study. In *Computational Linguistics and Intelligent Text Processing*, pages 498–509. Springer, 2005.

[75] Wael H Gomaa and Aly A Fahmy. A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13), 2013.

[76] Jack Goody. *The Interface between the Written and the Oral*. Cambridge University Press, 1987.

[77] Thomas F Gordon. Combining rules and ontologies with carneades. In *Proceedings of the 5th International RuleML2011@ BRF Challenge, CEUR Workshop Proceedings*, pages 103–110. Citeseer, 2011.

[78] Matthias Grabmair, Kevin D. Ashley, Rebecca Hwa, and Patricia M. Sweeney. Toward extracting information from public health statutes using text classification machine learning. In *Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference, University of Vienna, Austria, 14th-16th December 2011*, pages 73–82, 2011.

[79] Donato Griesi, Maria Teresa Pazienza, and Armando Stellato. Semantic turkey: A semantic bookmarking tool (system description). In *European Semantic Web Conference*, pages 779–788. Springer, 2007.

[80] Davide Grossi and A Jones. Constitutive norms and counts-as conditionals. *Handbook of Deontic Logic and Normative Systems. College Publications*, 80, 2013.

[81] Thomas R Gruber et al. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

[82] Aria D Haghighi, Andrew Y Ng, and Christopher D Manning. Robust textual inference via graph matching. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 387–394. Association for Computational Linguistics, 2005.

[83] Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, et al. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–18. Association for Computational Linguistics, 2009.

[84] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009. ISSN 1931-0145. doi: http://doi.acm.org/10.1145/1656274.1656278. URL http://doi.acm.org/10.1145/1656274.1656278.

[85] Bo Han and Timothy Baldwin. Lexical normalisation of short text messages: Makn sens a# twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 368–378. Association for Computational Linguistics, 2011.

[86] Zellig Harris. Mathematical structures of language. 1968.

[87] H. L. A. Hart. Positivism and the separation of law and morals. *Harvard Law Review Vol. 71 No. 4*, 1958.

[88] Marti A Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545. Association for Computational Linguistics, 1992.

[89] M. Hepp. Ontologies: State of the art, business potential, and grand challenges. In M. Hepp, P. De Leenheer, A. de Moor, and Y. Sure, editors, *Ontology Management: Semantic Web, Semantic Web Services, and Business Applications*, pages 3–22. Springer, 2007. ISBN 978-0-387-69899-1.

[90] Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. Recognizing textual entailment with lccÃŢs groundhog system. In *Proceedings of the Second PASCAL Challenges Workshop*, 2006.

[91] Risto Hilpinen and Paul McNamara. Deontic logic: a historical survey and introduction. *Handbook of Deontic Logic and Normative Systems. College Publications*, 80, 2013.

[92] Graeme Hirst. *Semantic interpretation and the resolution of ambiguity*. Cambridge University Press, 1992.

[93] Wesley Newcomb Hohfeld. Some fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal*, 23(1):16–59, 1913.

[94] Joachim Hruschka, Gottfried Achenwall, et al. Das deontologische sechseck bei gottfried achenwall im jahre 1767 zur geschichte der deontischen grundbegriffe in der universaljurisprudenz zwischen suarez und kant. 1986.

[95] Richard Hudson. A (1984) word grammar. *Basil Blackwell, Oxford*.

[96] Isaac Husik. Hohfeld's jurisprudence. *University of Pennsylvania Law Review and American Law Register*, pages 263–277, 1924.

[97] Fuyuki Ishikawa, Rihoko Inoue, and Shinichi Honiden. Modeling, analyzing and weaving legal interpretations in goal-oriented requirements engineering. In *Requirements Engineering and Law (RELAW), 2009 Second International Workshop on*, pages 39–44. IEEE, 2009.

[98] ITU-T. Recommendation Z.151 (10/12), User Requirements Notation (URN) - Language definition. http://www.itu.int/rec/T-REC-Z.151/en, 2012.

[99] Paul Jaccard. Etude comparative de la distribution dans une portion des alpes et du jura. *Bulletin de la Societe Vaudoise des Sciences Naturelle*, 4, 1901.

[100] RE Jennings. Can there be a natural deontic logic? *Synthese*, 65 (2):257–273, 1985.

[101] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: ECML-98*, pages 137–142, 1998.

[102] Richard Johansson and Pierre Nugues. Dependency-based semantic role labeling of propbank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 69–78. Association for Computational Linguistics, 2008.

[103] Richard Johansson and Pierre Nugues. The effect of syntactic representation on semantic role labeling. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 393–400. Association for Computational Linguistics, 2008.

[104] Jørgen Jørgensen. Imperatives and logic. *Erkenntnis*, 7(1):288–296, 1937.

[105] Michael Kaisser and Bonnie Webber. Question answering based on semantic roles. In *Proceedings of the Workshop on Deep Linguistic Processing*, pages 41–48. Association for Computational Linguistics, 2007.

[106] Dimitris Karagiannis. A business process-based modelling extension for regulatory compliance. In *Multikonferenz Wirtschaftsinformatik*, volume 2008, 2008.

[107] Paul Kingsbury and Martha Palmer. From treebank to propbank. In *LREC*. Citeseer, 2002.

[108] M. Klein. Combining and relating ontologies: an analysis of problems and solutions. In *Workshop on Ontologies and Information Sharing, IJCAI'01*, Seattle, USA, 2001.

[109] Tadas Klimas and Jurate Vaiciukaite. The law of recitals in european community legislation. *ILSA Journal of International and Comparative Law*, 15:61, 2008.

[110] Robert W Van Kralingen. *Frame-based conceptual models of statute law*. Kluwer Law Intl, 1995.

[111] G. Lakoff. The contemporary theory of metaphor. In A. Ortony, editor, *Metaphor and Thought*, pages 202–250. Cambridge University Press, 1993.

[112] Thomas K Landauer and Susan T Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.

[113] Alexei Lapouchnian. Goal-oriented requirements engineering: An overview of the current research. *University of Toronto*, page 32, 2005.

[114] G. Lau. *A Comparative Analysis Framework for Semi-Structured Documents, with Applications to Government Regulations*. PhD thesis, Stanford University, 2004.

[115] Koen Lenaerts and José A Gutiérrez-Fons. To say what the law of the eu is: methods of interpretation and the european court of justice. *Colum. J. Eur. L.*, 20:3, 2013.

[116] Alessandro Lenci, Simonetta Montemagni, Vito Pirrelli, and Giulia Venturi. Ontology learning from italian legal texts. In *Proceedings of the 2009 Conference on Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood*, pages 75–94, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press.

[117] L. Lesmo. The Turin University Parser at Evalita 2009. *Proceedings of EVALITA*, 9, 2009.

[118] Leonardo Lesmo, Alessandro Mazzei, Monica Palmirani, and Daniele P Radicioni. Tulsi: an nlp system for extracting legal modificatory provisions. *Artificial intelligence and law*, 21(2):139–172, 2013.

[119] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, volume 10, page 707, 1966.

[120] Beth Levin. *English verb classes and alternations: A preliminary investigation*. University of Chicago press, 1993.

[121] D. Liebwald. Law's capacity for vagueness. *International Journal for the Semiotics of Law Vol. 25 No. 4*, pages 29–50, 2012.

[122] Lars Lindahl and Jan Odelstad. Intermediate concepts in normative systems. In *Deontic Logic and Artificial Normative Systems*, pages 187–200. Springer, 2006.

[123] Fei Liu, Fuliang Weng, Bingqing Wang, and Yang Liu. Insertion, deletion, or substitution?: normalizing text messages without pre-categorization nor supervision. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 71–76. Association for Computational Linguistics, 2011.

[124] Eugene Emil Loos, Susan Anderson, Dwight Day, Paul Jordan, and J Wingate. *Glossary of linguistic terms*, volume 29. SIL International, 2004.

[125] Ruopeng Lu, Shazia Sadiq, and Guido Governatori. Measurement of compliance distance in business processes. *Information Systems Management*, 25(4):344–355, 2008.

[126] Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. The penn treebank: annotating predicate argument structure. In *Proceedings of the workshop on Human Language Technology*, pages 114–119. Association for Computational Linguistics, 1994.

[127] Réka Markovich, Gábor Hamp, et al. Elliptical lists in legislative texts. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, pages 192–195. ACM, 2015.

[128] Lluís Màrquez. Semantic role labeling: past, present and future. In *Tutorial Abstracts of ACL-IJCNLP 2009*, pages 3–3. Association for Computational Linguistics, 2009.

[129] Lluís Màrquez, Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. Semantic role labeling: an introduction to the special issue. *Computational linguistics*, 34(2):145–159, 2008.

[130] James H Martin and Daniel Jurafsky. Regular expressions, text normalization, edit distance. draft of september 1, 2014. In *Speech and language processing*. 2014.

[131] James H Martin and Daniel Jurafsky. Semantic role labeling. draft of june 26, 2015. In *Speech and language processing*. 2016.

[132] Slaugher & May. Introduction to the legislative processes for european union directives and regulations on financial services matters. https://www.slaughterandmay.com/media/1934583/introduction-to-the-legislative-processes-for-european-union-directives.pdf, 2014.

[133] Ryan T McDonald and Joakim Nivre. Characterizing the errors of data-driven dependency parsing models. In *EMNLP-CoNLL*, pages 122–131, 2007.

[134] M.Cherubini and D.Tiscornia. An Ontology-based model of procedural norms and regulated procedures. Technical Report 1/2010, ITTIG-CNR, 2010.

[135] Joseph P Mehrle. Computer-based system for classifying documents into a hierarchy and linking the classifications to the hierarchy, August 11 1998. US Patent 5,794,236.

[136] Igor? Aleksandrovič Mel?čuk. *Dependency syntax: theory and practice*. SUNY press, 1988.

[137] Gabor Melli, Yang Wang, Yudong Liu, Mehdi M Kashani, Zhongmin Shi, Baohua Gu, Anoop Sarkar, and Fred Popowich. Description of squash, the sfu question answering summary handler for the duc-2005 summarization task. *safety*, 1:14345754, 2005.

[138] Rada Mihalcea and Dan Moldovan. Semantic indexing using wordnet senses. In *Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 11*, pages 35–45. Association for Computational Linguistics, 2000.

[139] Rada Mihalcea, Courtney Corley, and Carlo Strapparava. Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI*, volume 6, pages 775–780, 2006.

[140] G.A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[141] George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4): 235–244, 1990.

[142] Ruslan Mitkov. *Anaphora resolution: the state of the art*. Citeseer, 1999.

[143] Dan I Moldovan and Vasile Rus. Logic form transformation of wordnet and its applicability to question answering. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 402–409. Association for Computational Linguistics, 2001.

[144] Srini Narayanan and Sanda Harabagiu. Question answering based on semantic structures. In *Proceedings of the 20th international conference on Computational Linguistics*, page 693. Association for Computational Linguistics, 2004.

[145] Monica Palmirani and Raffaella Brighi. Norma-system: A legal document system for managing consolidated acts. In *Database and Expert Systems Applications*, pages 310–320. Springer, 2002.

[146] Monica Palmirani, Marcello Ceci, Daniele Radicioni, and Alessandro Mazzei. Framenet model of the suspension of norms. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, pages 189–193. ACM, 2011.

[147] Monica Palmirani, Luca Cervone, Octavian Bujor, and Marco Chiappetta. Rawe: An editor for rule markup of legal texts. In *RuleML (2)*, 2013.

[148] Kishore Papineni. Why inverse document frequency? In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8. Association for Computational Linguistics, 2001.

[149] Thomas Edward Payne. *Describing morphosyntax: A guide for field linguists*. Cambridge University Press, 1997.

[150] Ismael Peña-López et al. World e-parliament report 2008. 2008.

[151] J. Rawls. Two concepts of rules. *The Philosophical Review Vol. 64*, pages 3–32, 1955.

[152] Robert Richards. Legal informatics blog. https://legalinformatics.wordpress.com/.

[153] Edwina L Rissland and David B Skalak. Cabaret: rule interpretation in a hybrid architecture. *International journal of man-machine studies*, 34(6):839–887, 1991.

[154] Livio Robaldo, Llio Humphreys, Xin Sun, Loredana Cupi, Cristiana Santos, and Robert Muthuri. Combining input/output logic and reification for representing real-world obligations.

[155] Livio Robaldo, Leonardo Lesmo, and Daniele P Radicioni. Compiling regular expressions to extract legal modifications. In *Legal Knowledge and Information Systems: JURIX 2012, the Twenty-fifth Annual Conference*, volume 250, page 133. IOS Press, 2012.

[156] Alf Ross. Directives and norms. 1968.

[157] P. Rossi and C. Vogel. Terms and concepts; towards a syllabus for european private law. *European Review of Private Law (ERPL)*, 12(2):293–300, 2004.

[158] Vasile Rus. *Logic Form for WordNet Glosses and Applications*. PhD thesis, Phd thesis, Southern Methodist University, May, 2002.

[159] R. Sacco. Legal formants: A dynamic approach to comparative law installments i and ii. *American Journal of Comparative Law Vol. 39*, 1991.

[160] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5): 513–523, 1988.

[161] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[162] G. Sartor. Access to legislation in the semantic web. In M.A. Biasiotti and S. Faro, editors, *From Information to Knowledge - Online Access to Legal Information: Methodologies, Trends and Perspectives*. IOS, 2011.

[163] Giovanni Sartor. Fundamental legal concepts: A formal and teleological characterisation*. *Artificial Intelligence and Law*, 14 (1-2):101–142, 2006.

[164] F de Saussure. Cours de linguistique générale. *Paris: Payot.(Primera edición, 1916)*, 1995.

[165] Wayne Schiess. Manage your sentence length. http://sites.utexas.edu/legalwriting/2014/08/25/manage-your-sentence-length/, 2014.

[166] Daniel Schleicher, Tobias Anstett, Frank Leymann, and David Schumm. Compliant business process design using refinement layers. In *On the Move to Meaningful Internet Systems: OTM 2010*, pages 114–131. Springer, 2010.

[167] David Schumm. Compliance-driven models. *Languages, and Architectures for Services (COMPAS): Achievements and Lessons Learned*, 2011.

[168] John R Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press, 1969.

[169] John R Searle. Social ontology: Some basic principles. *Anthropological theory*, 6(1):12–29, 2006.

[170] Nuno Seco, Tony Veale, and Jer Hayes. An intrinsic information content metric for semantic similarity in wordnet. In *ECAI*, volume 16, page 1089, 2004.

[171] Alberto Siena. *Engineering Law-Compliant Requirements: The Nòmos Framework*. PhD thesis, University of Trento, 2010.

[172] Alberto Siena, Neil Maiden, James Lockerbie, Kristine Karlsen, Anna Perini, and Angelo Susi. Exploring the effectiveness of normative i* modelling: Results from a case study on food chain traceability. In *Advanced Information Systems Engineering*, pages 182–196. Springer, 2008.

[173] Alberto Siena, John Mylopoulos, Anna Perini, and Angelo Susi. Designing law-compliant software requirements. In *Conceptual Modeling-ER 2009*, pages 472–486. Springer, 2009.

[174] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28 (1):11–21, 1972.

[175] Mark Stevenson and Mark A Greenwood. A semantic approach to ie pattern induction. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 379–386. Association for Computational Linguistics, 2005.

[176] Keh-Yih Su, Jong-Nae Wang, Mei-Hui Su, and Jing-Shin Chang. Glr parsing with scoring. In *Generalized LR Parsing*, pages 93–112. Springer, 1991.

[177] Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 8–15. Association for Computational Linguistics, 2003.

[178] Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 159–177. Association for Computational Linguistics, 2008.

[179] Richard Susskind. In the 2020s, artificial intelligence will transform the work of lawyers. http://www.thetimes.co.uk/tto/law/article4592380.ece, 2015.

[180] Idan Szpektor, Hristo Tanev, Ido Dagan, Bonaventura Coppola, et al. Scaling web-based acquisition of entailment relations. In *EMNLP*, volume 4, pages 41–48, 2004.

[181] the Gene Ontology. An introduction to the gene ontology, 1999. URL http://www.geneontology.org/GO.doc.shtml.

[182] P. M. Tiersma. Some myths about legal language. *Law, Culture and the Humanities, Vol. 2, No. 1*, pages pp. 29–50, 2006.

[183] Masaru Tomita. *Generalized LR parsing*. Springer Science & Business Media, 2012.

[184] J van der Pol. Rules-driven business services: Flexibility within the boundaries of the law. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law, ICAIL*, 2011.

[185] Giulia Venturi. Design and development of temis: a syntactically and semantically annotated corpus of italian legislative texts. In *Proceedings of the Workshop on Semantic Processing of Legal Texts (SPLeT 2012)*, pages 1–12, 2012.

[186] Giulia Venturi, Alessandro Lenci, Simonetta Montemagni, Eva Maria Vecchi, Maria Teresa Sagri, Daniela Tiscornia, and Tommaso Agnoloni. Towards a framenet resource for the legal domain. *LOAIT*, pages 67–76, 2009.

[187] P.R.S. Visser and T.J.M. Bench-Capon. A comparison of four ontologies for the design of legal knowledge systems. *Artificial Intelligence and Law*, 6:27–57, 1998.

[188] Georg Henrik Von Wright. Norm and action: a logical enquiry. 1963.

[189] Georg Henrik Von Wright. Norms, truth and logic. 1983.

[190] Peter Wahlgren. A general theory of artificial intelligence and law. *Legal knowledge based systems JURIX*, 94:79–83, 1984.

[191] M. Weiss and D. Amyot. Designing and evolving business models with URN. In *Proc. of Montreal Conf. on eTechnologies (MCETECH05)*, pages 149–162, 2005.

[192] Leif Wenar. Rights. http://plato.stanford.edu/entries/rights/, 2011. [Online; accessed 15-July-2011].

[193] S. L. Winter. Transcendental nonsense, metaphoric reasoning, and the cognitive stakes for law. *University of Pennsylvania Law Review Vol. 137 No. 4*, pages 1105–1237, 1989.

[194] Adam Wyner. Problems and prospects in the automatic semantic analysis of legal texts. In *SPLET Workshop at LREC*, 2012.

[195] Adam Wyner and Wim Peters. On rule extraction from regulations. In *Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference, University of Vienna, Austria, 14th-16th December 2011*, pages 113–122, 2011.

[196] Adam Wyner and Wim Peters. On rule extraction from regulations. In *JURIX*, volume 11, pages 113–122. Citeseer, 2011.

[197] Adam Wyner and Wim Peters. Semantic annotations for legal text processing using gate teamware. In *Semantic Processing of Legal Texts (SPLeT-2012) Workshop Programme*, page 34, 2012.

[198] Adam Wyner and Tom Van Engers. Web-based mass argumentation in natural language. *In Proceedings of Knowledge Engineering and Knowledge Management*, 2010.

[199] Roman Yangarber, Ralph Grishman, Pasi Tapanainen, and Silja Huttunen. Automatic acquisition of domain knowledge for information extraction. In *Proceedings of the 18th conference on Computational linguistics-Volume 2*, pages 940–946. Association for Computational Linguistics, 2000.

Part I

Appendices

12

APPENDIX A: NORMALISED NORMATIVE PROVISIONS: DIRECTIVE D07/46/EC

This appendix contains the first normalised sentences of the Directive in an Akoma Ntoso compatible format. The complete file is available upon request.

<?xml version="1.0" encoding="UTF-8"?>
<akomaNtoso
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.akomantoso.org/2.0"
xsi:schemaLocation="http://www.akomantoso.org/2.0 ./akomantoso20.xsd">
<act id="32007L0046-en">
<heading> CHAPTER I. <heading/>
<heading> GENERAL PROVISIONS. </heading>
<heading> Article 1. </heading>
<heading> Subject matter. </heading>
<sentence> This Directive establishes a harmonised framework containing the administrative provisions and general technical requirements for approval of all new vehicles within its scope and of the systems, components and separate technical units intended for those vehicles, with a view to facilitating their registration, sale and entry into service within the Community. </sentence>
<sentence> This Directive also establishes the provisions for the sale and entry into service of parts and equipment intended for vehicles approved in accordance with this Directive. </sentence>
<sentence> Specific technical requirements concerning the construction and functioning of vehicles shall be laid down in application of this Directive in regulatory acts, the exhaustive list of which is set out in Annex IV. </sentence>
<heading> Article 2. </heading>
<heading> Scope. </heading>
<list depth="1" conjunction="">
<point>
<num> 1. </num>
<sentence> This Directive applies to the type-approval of vehicles designed and constructed in one or more stages for use on the road, and of systems, components and separate technical units designed and constructed for such vehicles. </sentence>
<sentence> It also applies to the individual approval of such vehicles. </sentence>

<sentence> This Directive also applies to parts and equipment intended for vehicles covered by this Directive. </sentence>
</point>
<point>
<num> 2. </num>
<list depth="2" conjunction="and">
<point>
<num> (a) </num>
<sentence> <intro> This Directive does not apply to the type-approval or individual approval of the following vehicles: </intro> <content> agricultural or forestry tractors, as defined in Directive 2003/37/EC of the European Parliament and of the Council of 26 May 2003 on type-approval of agricultural or forestry tractors, their trailers and interchangeable towed machinery, together with their systems, components and separate technical units [10] and trailers designed and constructed specifically to be towed by them; </content> </sentence>
</point>
<point>
<num> (b) </num>
<sentence> <intro> This Directive does not apply to the type-approval or individual approval of the following vehicles: </intro> <content> quadricycles as defined in Directive 2002/24/EC of the European Parliament and of the Council of 18 March 2002 relating to the type-approval of two or three-wheel motor vehicles [11]; </content> </sentence>
</point>
<point>
<num> (c) </num>
<sentence> <intro> This Directive does not apply to the type-approval or individual approval of the following vehicles: </intro> <content> tracked vehicles. </content> </sentence>
</list>
</point>
<point>
<num> 3. </num>
<list depth="2" conjunction="and">
<point>
<num> (a) </num>
<sentence> <intro> Type-approval or individual approval under this Directive is optional for the following vehicles: </intro> <content> vehicles designed and constructed for use principally on construction sites or in quarries, port or airport facilities; </content> <ending> :to the extent that these vehicles fulfil the requirements of this Directive. </ending> </sentence>
</point>

<point>
<num> (b) </num>
<sentence> <intro> Type-approval or individual approval under this Directive is optional for the following vehicles: </intro> <content> vehicles designed and constructed for use by the armed services, civil defence, fire services and forces responsible for maintaining public order; </content> <ending> :to the extent that these vehicles fulfil the requirements of this Directive. </ending> </sentence>
</point>
<point>
<num> (c) </num>
<sentence> <intro> Type-approval or individual approval under this Directive is optional for the following vehicles: </intro> <content> mobile machinery, </content> <ending> :to the extent that these vehicles fulfil the requirements of this Directive. </ending> </sentence>
</list>
<sentence> Such optional approvals shall be without prejudice to the application of Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery [12]. </sentence>
</point>
<point>
<num> 4. </num>
<list depth="2" conjunction="">
<point>
<num> (a) </num>
<sentence> <intro> An individual approval under this Directive is optional for the following vehicles: </intro> <content> vehicles intended exclusively for racing on roads; </content> </sentence>
</point>
<point>
<num> (b) </num>
<sentence> <intro> An individual approval under this Directive is optional for the following vehicles: </intro> <content> prototypes of vehicles used on the road under the responsibility of a manufacturer to perform a specific test programme provided they have been specifically designed and constructed for this purpose. </content>
</sentence>
</list>
</list>
<heading> Article 3. </heading>
<heading> Definitions. </heading>
<list depth="1" conjunction="">
<point>
<num> 1. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided

therein: </intro> <content> ”regulatory act” means a separate directive or regulation or a UNECE Regulation annexed to the Revised 1958 Agreement; </content> </sentence>
</point>
<point>
<num> 2. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”separate directive or regulation” means a directive or regulation listed in Part I of Annex IV. </content> </sentence>
<sentence> This term includes also their implementing acts; </sentence>
</point>
<point>
<num> 3. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”type-approval” means the procedure whereby a Member State certifies that a type of vehicle, system, component or separate technical unit satisfies the relevant administrative provisions and technical requirements; </content> </sentence>
</point>
<point>
<num> 4. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”national type-approval” means a type-approval procedure laid down by the national law of a Member State, the validity of such approval being restricted to the territory of that Member State; </content> </sentence>
</point>
<point>
<num> 5. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”EC type-approval” means the procedure whereby a Member State certifies that a type of vehicle, system, component or separate technical unit satisfies the relevant administrative provisions and technical requirements of this Directive and of the regulatory acts listed in Annex IV or XI; </content> </sentence>
</point>
<point>
<num> 6. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”individual approval” means the pro-

cedure whereby a Member State certifies that a particular vehicle, whether unique or not, satisfies the relevant administrative provisions and technical requirements; </content> </sentence>
</point>
<point>
<num> 7. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "multi-stage type-approval" means the procedure whereby one or more Member States certify that, depending on the state of completion, an incomplete or completed type of vehicle satisfies the relevant administrative provisions and technical requirements of this Directive; </content> </sentence>
</point>
<point>
<num> 8. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "step-by-step type-approval" means a vehicle approval procedure consisting in the step-by-step collection of the whole set of EC type-approval certificates for the systems, components and separate technical units relating to the vehicle, and which leads, at the final stage, to the approval of the whole vehicle; </content> </sentence>
</point>
<point>
<num> 9. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "single-step type-approval" means a procedure consisting in the approval of a vehicle as a whole by means of a single operation; </content> </sentence>
</point>
<point>
<num> 10. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "mixed type-approval" means a step-by-step type-approval procedure for which one or more system approvals are achieved during the final stage of the approval of the whole vehicle, without it being necessary to issue the EC type-approval certificates for those systems; </content> </sentence>
</point>
<point>
<num> 11. </num>
<sentence> <intro> For the purposes of this Directive and of the

regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”motor vehicle” means any power-driven vehicle which is moved by its own means, having at least four wheels, being complete, completed or incomplete, with a maximum design speed exceeding 25 km/h; </content> </sentence>
</point>
<point>
<num> 12. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”trailer” means any non-self-propelled vehicle on wheels which is designed and constructed to be towed by a motor vehicle; </content> </sentence>
</point>
<point>
<num> 13. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”vehicle” means any motor vehicle or its trailer as defined in points (11) and (12); </content> </sentence>
</point>
<point>
<num> 14. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”hybrid motor vehicle” means a vehicle with at least two different energy converters and two different energy storage systems (on-vehicle) for the purpose of vehicle propulsion; </content> </sentence>
</point>
<point>
<num> 15. </num>
<list depth=“2” conjunction=“”>
<point>
<num> - </num>
<sentence> <intro> The manufacturer may choose one of the following procedures: </intro> <intro> ”hybrid electric vehicle” means a hybrid vehicle that, for the purpose of mechanical propulsion, draws energy from both of the following on-vehicle sources of stored energy/power: </intro> <content> a consumable fuel, </content> </sentence>
</point>
<point>
<num> - </num>
<sentence> <intro> The manufacturer may choose one of the following procedures: </intro> <intro> ”hybrid electric vehicle” means

a hybrid vehicle that, for the purpose of mechanical propulsion, draws energy from both of the following on-vehicle sources of stored energy/power: </intro> <content> an electrical energy/power storage device (e.g. battery, capacitor, flywheel/generator, etc.); </content>
</sentence>
</list>
</point>
<point>
<num> 16. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”mobile machinery” means any self-propelled vehicle which is designed and constructed specifically to perform work which, because of its construction characteristics, is not suitable for carrying passengers or for transporting goods. </content> </sentence>
<sentence> Machinery mounted on a motor vehicle chassis shall not be considered as mobile machinery; </sentence>
</point>
<point>
<num> 17. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”type of vehicle” means vehicles of a particular category which do not differ in at least the essential respects specified in Section B of Annex II. </content> </sentence>
<sentence> A type of vehicle may contain variants and versions as defined in Section B of Annex II; </sentence>
</point>
<point>
<num> 18. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”base vehicle” means any vehicle which is used at the initial stage of a multi-stage type-approval process; </content>
</sentence>
</point>
<point>
<num> 19. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”incomplete vehicle” means any vehicle which must undergo at least one further stage of completion in order to meet the relevant technical requirements of this Directive; </content>
</sentence>
</point>

<point>
<num> 20. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”completed vehicle” means a vehicle, resulting from the process of multi-stage type-approval, which meets the relevant technical requirements of this Directive; </content> </sentence>
</point>
<point>
<num> 21. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”complete vehicle” means any vehicle which need not be completed in order to meet the relevant technical requirements of this Directive; </content> </sentence>
</point>
<point>
<num> 22. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”end-of-series vehicle” means any vehicle that is part of a stock which cannot be registered or sold or entered into service owing to the entry into force of new technical requirements against which it has not been approved; </content> </sentence>
</point>
<point>
<num> 23. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”system” means an assembly of devices combined to perform one or more specific functions in a vehicle and which is subject to the requirements of any of the regulatory acts; </content> </sentence>
</point>
<point>
<num> 24. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”component” means a device subject to the requirements of a regulatory act and intended to be part of a vehicle, which may be type-approved independently of a vehicle where the regulatory act makes express provisions for so doing; </content> </sentence>
</point>
<point>
<num> 25. </num>

<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”separate technical unit” means a device subject to the requirements of a regulatory act and intended to be part of a vehicle, which may be type-approved separately, but only in relation to one or more specified types of vehicle where the regulatory act makes express provisions for so doing; </content> </sentence>
</point>
<point>
<num> 26. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”original parts or equipment” means parts or equipment which are manufactured according to the specifications and production standards provided by the vehicle manufacturer for the production of parts or equipment for the assembly of the vehicle in question. </content> </sentence>
<sentence> This includes parts or equipment which are manufactured on the same production line as these parts or equipment. </sentence>
<sentence> It is presumed unless the contrary is proven, that parts constitute original parts if the part manufacturer certifies that the parts match the quality of the components used for the assembly of the vehicle in question and have been manufactured according to the specifications and production standards of the vehicle manufacturer; </sentence>
</point>
<point>
<num> 27. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”manufacturer” means the person or body who is responsible to the approval authority for all aspects of the type-approval or authorisation process and for ensuring conformity of production. </content> </sentence>
<sentence> It is not essential that the person or body be directly involved in all stages of the construction of the vehicle, system, component or separate technical unit which is the subject of the approval process; </sentence>
</point>
<point>
<num> 28. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”manufacturer”s representativeâĂŹ means any natural or legal person established in the Community who is duly

appointed by the manufacturer to represent him before the approval authority and to act on his behalf in matters covered by this Directive, and where reference is made to the term "manufacturer", it is to be understood as indicating either the manufacturer or his representative; </content> </sentence>
</point>
<point>
<num> 29. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "approval authority" means the authority of a Member State with competence for all aspects of the approval of a type of vehicle, system, component or separate technical unit or of the individual approval of a vehicle; </content> </sentence>
<sentence> for the authorisation process, for issuing and, if appropriate, withdrawing approval certificates; </sentence>
<sentence> for acting as the contact point for the approval authorities of other Member States; </sentence>
<sentence> for designating the technical services and for ensuring that the manufacturer meets his obligations regarding the conformity of production; </sentence>
</point>
<point>
<num> 30. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "competent authority" in Article 42 means either the approval authority or a designated authority, or an accreditation body acting on their behalf; </content> </sentence>
</point>
<point>
<num> 31. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "technical service" means an organisation or body designated by the approval authority of a Member State as a testing laboratory to carry out tests, or as a conformity assessment body to carry out the initial assessment and other tests or inspections, on behalf of the approval authority, it being possible for the approval authority itself to carry out those functions; </content> </sentence>
</point>
<point>
<num> 32. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> "virtual testing method" means computer simula-

tions including calculations which demonstrate whether a vehicle, a system, a component or a separate technical unit fulfils the technical requirements of a regulatory act. </content> </sentence>

<sentence> For testing purposes, a virtual method does not require the use of a physical vehicle, system, component or separate technical unit; </sentence>
</point>
<point>
<num> 33. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”type-approval certificate” means the document whereby the approval authority officially certifies that a type of vehicle, system, component or separate technical unit is approved; </content> </sentence>
</point>
<point>
<num> 34. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”EC type-approval certificate” means the certificate set out in Annex VI or in the corresponding annex to a separate directive or regulation, the communication form set out in the relevant Annex to one of the UNECE Regulations listed in Part I or Part II of Annex IV to this Directive, being deemed to be equivalent thereto; </content> </sentence>
</point>
<point>
<num> 35. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”individual approval certificate” means the document whereby the approval authority officially certifies that a particular vehicle is approved; </content> </sentence>
</point>
<point>
<num> 36. </num>
<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”certificate of conformity” means the document set out in Annex IX, issued by the manufacturer and certifying that a vehicle belonging to the series of the type approved in accordance with this Directive complied with all regulatory acts at the time of its production; </content> </sentence>
</point>
<point>

<num> 37. </num>

<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”information document” means the document set out in Annex I or Annex III, or in the corresponding Annex to a separate directive, or regulation, that prescribes the information to be supplied by an applicant, it being permissible to supply the information document in the form of an electronic file; </content> </sentence>
</point>
<point>

<num> 38. </num>

<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”information folder” means the complete folder, including the information document, file, data, drawings, photographs, and so on, supplied by the applicant, it being permissible to supply the information folder in the form of an electronic file; </content> </sentence>
</point>
<point>

<num> 39. </num>

<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”information package” means the information folder accompanied by the test reports and all other documents added by the technical service or by the approval authority to the information folder in the course of carrying out their functions, it being permissible to supply the information package in the form of an electronic file; </content> </sentence>
</point>
<point>

<num> 40. </num>

<sentence> <intro> For the purposes of this Directive and of the regulatory acts listed in Annex IV, save as otherwise provided therein: </intro> <content> ”index to the information package” means the document listing the contents of the information package, suitably numbered or otherwise marked so as to identify clearly all the pages, the format of that document being such as to present a record of the successive steps in the management of the EC type-approval, in particular the dates of the revisions and updating. </content> </sentence>
</list>
<heading> CHAPTER II. </heading>
<heading> GENERAL OBLIGATIONS. </heading>
<heading> Article 4. </heading>
<heading> Obligations of Member States. </heading>

<list depth="1" conjunction="">
<point>
<num> 1. </num>
<sentence> Member States shall ensure that manufacturers applying for approval comply with their obligations under this Directive. </sentence>
</point>
<point>
<num> 2. </num>
<sentence> Member States shall approve only such vehicles, systems, components or separate technical units as satisfy the requirements of this Directive. </sentence>
</point>
<point>
<num> 3. </num>
<sentence> Member States shall register or permit the sale or entry into service only of such vehicles, components and separate technical units as satisfy the requirements of this Directive. </sentence>
<sentence> They shall not prohibit, restrict or impede the registration, sale, entry into service or circulation on the road of vehicles, components or separate technical units, on grounds related to aspects of their construction and functioning covered by this Directive, if they satisfy the requirements of the latter. </sentence>
</point>
<point>
<num> 4. </num>
<sentence> Member States shall establish or appoint the authorities competent in matters concerning approval, and notify to the Commission such establishment or appointment in accordance with Article 43. </sentence>
</list>
<sentence> The notification act of the approval authorities shall include the name, the address, including electronic address, and their area of responsibility. </sentence>
<heading> Article 5. </heading>
<heading> Obligations of manufacturers. </heading>
<list depth="1" conjunction="">
<point>
<num> 1. </num>
<sentence> The manufacturer is responsible to the approval authority for all aspects of the approval process and for ensuring conformity of production, whether or not the manufacturer is directly involved in all stages of the construction of a vehicle, system, component or separate technical unit. </sentence>
</point>
<point>

<num> 2. </num>

<sentence> In the case of multi-stage type-approval, each manufacturer is responsible for the approval and conformity of production of the systems, components or separate technical units added at the stage of vehicle completion handled by him. </sentence>

<sentence> The manufacturer who modifies components or systems already approved at earlier stages shall be responsible for the approval and conformity of production of those components and systems. </sentence>

</point>

<point>

<num> 3. </num>

<sentence> For the purposes of this Directive, a manufacturer established outside the Community shall appoint a representative established in the Community to represent him before the approval authority. </sentence>

</list>

<heading> CHAPTER III. </heading>

<heading> EC TYPE-APPROVAL PROCEDURES. </heading>

<heading> Article 6. </heading>

<sentence> Procedures to be followed for the EC type-approval of vehicles. </sentence>

<list depth="1" conjunction="">

<point>

<num> 1. </num>

<list depth="2" conjunction="">

<point>

<num> (a) </num>

<sentence> <intro> The manufacturer may choose one of the following procedures: </intro> <content> step-by-step type-approval; </content> </sentence>

</point>

<point>

<num> (b) </num>

<sentence> <intro> The manufacturer may choose one of the following procedures: </intro> <content> single-step type-approval; </content> </sentence>

</point>

<point>

<num> (c) </num>

<sentence> <intro> The manufacturer may choose one of the following procedures: </intro> <content> mixed type-approval. </content> </sentence>

</list>

</point>

<point>

<num> 2. </num>

<sentence> An application for step-by-step type-approval shall consist of the information folder containing the information required under Annex III and shall be accompanied by the complete set of type-approval certificates required pursuant to each of the applicable regulatory acts listed in Annex IV or Annex XI. </sentence>

<sentence> In the case of the type-approval of a system or separate technical unit, pursuant to the applicable regulatory acts, the approval authority shall have access to the related information package until such time as the approval is either issued or refused. </sentence>
</point>
<point>

<num> 3. </num>

<sentence> An application for single-step type-approval shall consist of the information folder containing the relevant information required under Annex I, in relation to the regulatory acts specified in Annex IV or Annex XI and, where applicable, in Part II of Annex III.
</sentence>
</point>
<point>

<num> 4. </num>

<sentence> In the case of a mixed type-approval procedure, the approval authority may exempt a manufacturer from the obligation to produce one or more EC system type-approval certificates, provided that the information folder is supplemented by the particulars, specified in Annex I, required for the approval of those systems during the vehicle approval phase, in which case each of the EC type-approval certificates thus waived shall be replaced by a test report. </sentence>
</point>
<point>

<num> 5. </num>
<list depth="2" conjunction="">
<point>

<num> (a) </num>

<sentence> <intro> Without prejudice to paragraphs 2, 3 and 4, the following information shall be supplied for the purposes of multi-stage type-approval: </intro> <content> at the first stage, those parts of the information folder and the EC type-approval certificates required for a complete vehicle which are relevant to the state of completion of the base vehicle; </content> </sentence>
</point>
<point>

<num> (b) </num>

<sentence> <intro> Without prejudice to paragraphs 2, 3 and 4, the following information shall be supplied for the purposes of multi-stage type-approval: </intro> <content> at the second and subsequent

stages, those parts of the information folder and the EC type-approval certificates which are relevant to the current stage of construction, together with a copy of the EC type-approval certificate for the vehicle issued at the preceding stage of construction; </content> </sentence>

<sentence> in addition, the manufacturer shall supply full details of any changes or additions that he has made to the vehicle. </sentence>

</list>

<sentence> The information specified in points (a) and (b) may be supplied in accordance with the mixed type-approval procedure set out in paragraph 4. </sentence>

</point>

<point>

<num> 6. </num>

<sentence> The manufacturer shall submit the application to the approval authority. </sentence>

<sentence> Only one application may be submitted in respect of a particular type of vehicle and it may be submitted in only one Member State. </sentence>

<sentence> A separate application shall be submitted for each type to be approved. </sentence>

</point>

<point>

<num> 7. </num>

<sentence> The approval authority may, by reasoned request, call upon the manufacturer to supply any additional information needed to enable a decision to be taken on what tests are required or to facilitate the execution of those tests. </sentence>

</point>

<point>

<num> 8. </num>

<sentence> The manufacturer shall make available to the approval authority as many vehicles as are necessary to enable the type-approval procedure to be conducted satisfactorily. </sentence>

</list>

<heading> Article 7. </heading>

<sentence> Procedure to be followed for the EC type-approval of systems, components or separate technical units. </sentence>

<list depth="1" conjunction="">

<point>

<num> 1. </num>

<sentence> The manufacturer shall submit the application to the approval authority. </sentence>

<sentence> Only one application may be submitted in respect of a particular type of system, component or separate technical unit and it may be submitted in only one Member State. </sentence>

<sentence> A separate application shall be submitted for each type to be approved. </sentence>
</point>
<point>
<num> 2. </num>
<sentence> The application shall be accompanied by the information folder, the content of which is specified in the separate directives or regulations. </sentence>
</point>
<point>
<num> 3. </num>
<sentence> The approval authority may, by reasoned request, call upon the manufacturer to supply any additional information needed to enable a decision to be taken on what tests are required or to facilitate the execution of those tests. </sentence>
</point>
<point>
<num> 4. </num>
<sentence> The manufacturer shall make available to the approval authority as many vehicles, components or separate technical units as are required under the relevant separate directives or regulations for the performance of the required tests. </sentence>
</list>
<heading> CHAPTER IV. </heading>
<heading> CONDUCT OF EC TYPE-APPROVAL PROCEDURES. </heading>
<heading> Article 8. </heading>
<heading> General provisions. </heading>
<list depth="1" conjunction="">
<point>
<num> 1. </num>
<sentence> Member States may not grant any EC type-approval without first ensuring that the procedures referred to in Article 12 have been duly and satisfactorily implemented. </sentence>
</point>
<point>
<num> 2. </num>
<sentence> Member States shall grant EC type-approvals in accordance with Articles 9 and 10. </sentence>
</point>
<point>
<num> 3. </num>
<sentence> If a Member State finds that a type of vehicle, system, component or separate technical unit, albeit in conformity with the required provisions, presents a serious risk to road safety or seriously harms the environment or seriously harms public health, it may refuse

to grant EC type-approval. </sentence>

<sentence> In this case, it shall immediately send the other Member States and the Commission a detailed file explaining the reasons for its decision and setting out the evidence for its findings. </sentence>

</point>

<point>

<num> 4. </num>

<sentence> EC type-approval certificates shall be numbered in accordance with the method set out in Annex VII. </sentence>

</point>

<point>

<num> 5. </num>

<sentence> The approval authority shall, within 20 working days, send to the approval authorities of the other Member States a copy of the EC vehicle type-approval certificate, together with the attachments, for each type of vehicle which it has approved. </sentence>

<sentence> The hard copy may be replaced by an electronic file. </sentence>

</point>

<point>

<num> 6. </num>

<sentence> The approval authority shall inform without delay the approval authorities of the other Member States of its refusal or withdrawal of any vehicle approval, together with the reasons for its decision. </sentence>

</point>

<point>

<num> 7. </num>

<sentence> The approval authority shall send at three-monthly intervals to the approval authorities of the other Member States a list of the system, component or separate technical unit EC type-approvals it has granted, amended, refused to grant or withdrawn during the preceding period. </sentence>

<sentence> That list shall contain the particulars specified in Annex XIV. </sentence>

</point>

<point>

<num> 8. </num>

<sentence> If so requested by another Member State, the Member State which has granted an EC type-approval shall, within 20 working days of receiving that request, send a copy of the EC type-approval certificate in question, together with the attachments. </sentence>

<sentence> The hard copy may be replaced by an electronic file. </sentence>

</list>

<heading> Article 9. </heading>
<heading> Specific provisions concerning vehicles. </heading>
<list depth="1" conjunction="">
<point>
<num> 1. </num>
<list depth="2" conjunction="">
<point>
<num> (a) </num>
<sentence> <intro> Member States shall grant an EC approval in respect of: </intro> <content> a type of vehicle which conforms to the particulars in the information folder and which meets the technical requirements specified by the relevant regulatory acts listed in Annex IV; </content> </sentence>
</point>
<point>
<num> (b) </num>
<sentence> <intro> Member States shall grant an EC approval in respect of: </intro> <content> a type of special-purpose vehicle which conforms to the particulars in the information folder and which meets the technical requirements specified by the relevant regulatory acts listed in Annex XI. </content> </sentence>
</list>
<sentence> The procedures set out in Annex V shall apply. </sentence>
</point>
<point>
<num> 2. </num>
<sentence> Member States shall grant a multi-stage type-approval in respect of a type of incomplete or completed vehicle which conforms to the particulars in the information folder and which meets the technical requirements specified by the relevant regulatory acts listed in Annex IV or Annex XI, having regard to the state of completion of the vehicle. </sentence>
<sentence> The multi-stage type-approval shall apply also to complete vehicles converted or modified by another manufacturer. </sentence>
<sentence> The procedures set out in Annex XVII shall apply. </sentence>
</point>
<point>
<num> 3. </num>
<list depth="2" conjunction="">
<point>
<num> (a) </num>
<sentence> <intro> In respect of each type of vehicle, the approval authority shall: </intro> <content> complete all the relevant sections

of the EC type-approval certificate, including the test results sheet appended thereto, in accordance with the model set out in Annex VIII; </content> </sentence>
</point>
<point>
<num> (b) </num>
<sentence> <intro> In respect of each type of vehicle, the approval authority shall: </intro> <content> compile or verify the index to the information package; </content> </sentence>
</point>
<point>
<num> (c) </num>
<sentence> <intro> In respect of each type of vehicle, the approval authority shall: </intro> <content> issue the completed certificate, together with its attachments, to the applicant without unjustified delay. </content> </sentence>
</list>
</point>
</act>
</akomaNtoso>

13

APPENDIX B: DEFINITION AND NORM EXTRACTION: DIRECTIVE D95/46/EC

This appendix contains the first extracted definitions and norms of the Directive. The complete file is available upon request.
indicates element found by pattern-matching.

<?xml version="1.0"?>
<Legislation ID="031995L0046-en">
<Norm ref="S_A1_B1_D">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="FullyCorrect"> protect the fundamental rights and freedoms of natural persons , and in particular their right to privacy with respect to the processing of personal data </Action>
<FullText> In accordance with this Directive , Member States shall protect the fundamental rights and freedoms of natural persons , and in particular their right to privacy with respect to the processing of personal data . </FullText>
</Norm>
<Norm ref="S_A1_B2_D">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> neither </Action>
<Action verdict="PartiallyCorrect"> restrict the free flow of personal data between for reasons connected with the protection afforded under S_paragraph_1 </Action>
<Action verdict="Missing"> nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1 </Action>
<FullText> Member States shall neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1 . </FullText>
</Norm>
<Norm ref="S_A2_B_La_R">
<NormType verdict="FullyCorrect"> Definition </NormType>
<Definiendum verdict="PartiallyCorrect" note="NormalisationError Apostrophe"> " personal data ' " </Definiendum>
<Definiens verdict="PartiallyCorrect" note="NormalisationErrorSubjected"> any information relating to an

identified or identifiable natural person ('data subjected ') </Definiens>
<Scope verdict="PartiallyCorrect"> For the purposes of S_this_Directive </Scope>
<FullText> For the purposes of this Directive , " personal data ' shall mean any information relating to an identified or identifiable natural person ('data subject ') ; </FullText>
</Norm>
<Norm ref="S_A2_B_La_R">
<NormType verdict="Wrong" note="Definition"> Unknown </NormType>
<Definiendum verdict="Missing"> an identifiable person </Definiendum>
<Definiens verdict="Missing"> one who can be identified , directly or indirectly , in particular by reference to an identification number or to one or more factors specific to his physical , physiological , mental , economic , cultural or social identity </Definiens>
<FullText> an identifiable person is one who can be identified , directly or indirectly , in particular by reference to an identification number or to one or more factors specific to his physical , physiological , mental , economic , cultural or social identity ; </FullText>
</Norm>
<Norm ref="S_A2_B_Lb_R">
<NormType verdict="FullyCorrect"> Definition </NormType>
<Definiendum verdict="PartiallyCorrect" note="NormalisationErrorApostrophe"> " processing of personal data ' ('processing ') " </Definiendum>
<Definiens verdict="FullyCorrect"> any operation or set of operations which is performed upon personal data , whether or not by automatic means , such as collection , recording , organization , storage , adaptation or alteration , retrieval , consultation , use , disclosure by transmission , dissemination or otherwise making available , alignment or combination , blocking , erasure or destruction </Definiens>
<Scope verdict="FullyCorrect"> For the purposes of S_this_Directive </Scope>
<FullText> For the purposes of this Directive , " processing of personal data ' ('processing ') shall mean any operation or set of operations which is performed upon personal data , whether or not by automatic means , such as collection , recording , organization , storage , adaptation or alteration , retrieval , consultation , use , disclosure by transmission , dissemination or otherwise making available , alignment or combination , blocking , erasure or destruction ; </FullText>
</Norm>
<Norm ref="S_A2_B_Lc_R">
<NormType verdict="FullyCorrect"> Definition </NormType>
<Definiendum verdict="PartiallyCorrect" note="NormalisationErrorApostrophe"> " personal data filing system ' ('filing system ') "

</Definiendum>
<Definiens verdict="FullyCorrect"> any structured set of personal data which are accessible according to specific criteria , whether centralized , decentralized or dispersed on a functional or geographical basis </Definiens>
<Scope verdict="FullyCorrect"> For the purposes of S_this_Directive </Scope>
<FullText> For the purposes of this Directive , " personal data filing system ' ('filing system ') shall mean any structured set of personal data which are accessible according to specific criteria , whether centralized , decentralized or dispersed on a functional or geographical basis ; </FullText>
</Norm>
<Norm ref="S_A2_B_Ld_R">
<NormType verdict="FullyCorrect"> Definition </NormType>
<Definiendum verdict="PartiallyCorrect" note="NormalisationError Apostrophe"> " controller ' " </Definiendum>
<Definiens verdict="FullyCorrect"> the natural or legal person , public authority , agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data </Definiens>
<Scope> For the purposes of S_this_Directive </Scope>
<FullText> For the purposes of this Directive , " controller ' shall mean the natural or legal person , public authority , agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data ; </FullText>
</Norm>
<Norm ref="S_A2_B_Ld_R">
<NormType verdict="Wrong" note="Power"> Permission </NormType>
<##Condition verdict="PartiallyCorrect"> where the purposes </##Condition>
<##Action verdict="Missing"> the controller or the specific criteria for his nomination may be designated by national or Community law </##Action>
<PassiveRole verdict="Missing"> the controller </PassiveRole>
<FullText> where the purposes and means of processing are determined by national or Community laws or regulations , the controller or the specific criteria for his nomination may be designated by national or Community law ; </FullText>
</Norm>
<Norm ref="S_A2_B_Ld_R">
<NormType verdict="Wrong" note="Power"> Permission </NormType>
<ActiveRole verdict="Wrong" note="PassiveRole"> the controller or the specific criteria for his nomination </ActiveRole>

<Action verdict="PartiallyCorrect"> are designated by national or Community law </Action>

<Condition verdict="FullyCorrect"> where the purposes and means of processing are determined by national or Community laws or regulations </Condition>

<FullText> where the purposes and means of processing are determined by national or Community laws or regulations , the controller or the specific criteria for his nomination may be designated by national or Community law ; </FullText>

</Norm>

<Norm ref="S_A2_B_Le_R">

<NormType verdict="FullyCorrect"> Definition </NormType>

<Definiendum verdict="PartiallyCorrect" note="NormalisationError Apostrophe"> " processor ' " </Definiendum>

<Definiens verdict="FullyCorrect"> a natural or legal person , public authority , agency or any other body which processes personal data on behalf of the controller </Definiens>

<Scope verdict="FullyCorrect"> For the purposes of S_this_Directive </Scope>

<FullText> For the purposes of this Directive , " processor ' shall mean a natural or legal person , public authority , agency or any other body which processes personal data on behalf of the controller ; </FullText>

</Norm>

<Norm ref="S_A2_B_Lf_R">

<NormType verdict="FullyCorrect"> Definition </NormType>

<Definiendum verdict="PartiallyCorrect" note="NormalisationError Apostrophe"> " third party ' " </Definiendum>

<Definiens verdict="PartiallyCorrect" note="NormalisationErrorBecome"> any natural or legal person , public authority , agency or any other body other than the data subjected , the controller , the processor and the persons who , under the direct authority of the controller or the processor , are authorized to process the data </Definiens>

<Scope> For the purposes of S_this_Directive </Scope>

<FullText> For the purposes of this Directive , " third party ' shall mean any natural or legal person , public authority , agency or any other body other than the data subject , the controller , the processor and the persons who , under the direct authority of the controller or the processor , are authorized to process the data ; </FullText>

</Norm>

<Norm ref="S_A2_B_Lg_R">

<NormType verdict="FullyCorrect"> Definition </NormType>

<Definiendum verdict="PartiallyCorrect" note="NormalisationError Apostrophe"> " recipient ' " </Definiendum>

<Definiens verdict="FullyCorrect"> a natural or legal person ,

public authority , agency or any other body to whom data are disclosed , whether a third party or not </Definiens>
<Scope verdict="FullyCorrect"> For the purposes of S_this_Directive </Scope>
<FullText> For the purposes of this Directive , " recipient ' shall mean a natural or legal person , public authority , agency or any other body to whom data are disclosed , whether a third party or not ; </FullText>
</Norm>
<Norm ref="S_A2_B_Lg_R">
<NormType verdict="Wrong" note="Definition"> Obligation </NormType>
<Action verdict="Wrong"> however </Action>
<Definiendum verdict="Missing"> recipients </Definiendum>
<Excludes verdict="Missing"> authorities which may receive data in the framework of a particular inquiry </Excludes>
<FullText> however , authorities which may receive data in the framework of a particular inquiry shall not be regarded as recipients ; </FullText>
</Norm>
<Norm ref="S_A2_B_Lh_R">
<NormType verdict="FullyCorrect"> Definition </NormType>
<Definiendum verdict="PartiallyCorrect" note="NormalisationErrorApostrophe"> " the data subjected 's consent ' " </Definiendum>
<Definiens verdict="PartiallyCorrect" note="NormalisationErrorBecome"> any freely given specific and informed indication of his wishes by which the data subjected signifies his agreement to personal data relating to him being processed </Definiens>
<Scope> For the purposes of S_this_Directive </Scope>
<FullText> For the purposes of this Directive , " the data subject 's consent ' shall mean any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed . </FullText>
</Norm>
<Norm ref="S_A3_B1_D">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_This_Directive </Rule>
<Object verdict="FullyCorrect"> the processing of personal data , wholly or partly by automatic means , and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system </Object>
<FullText> This Directive shall apply to the processing of personal data wholly or partly by automatic means , and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system .

</FullText>
</Norm>
<Norm ref="S_A3_B2_D_B_B_H">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_This_Directive </Rule>
<ExcludesObject verdict="PartiallyCorrect"> the processing of personal data </ExcludesObject>
<ExcludesObject verdict="Wrong" note="NoAdditionalExcludesObject"> in the course of an activity which falls outside the scope of Community law , such as those provided for by S_Titles_V_and_VI_of_the_Treaty_on_European_Union and in any case to processing operations concerning public security , defence , State security (including the economic well-being of the State and the activities of the State in areas of criminal law, </ExcludesObject>
<##Condition verdict="PartiallyCorrect"> when the processing operation relates to State security matters) </##Condition>
<FullText> This Directive shall not apply to the processing of personal data , in the course of an activity which falls outside the scope of Community law , such as those provided for by Titles V and VI of the Treaty on European Union and in any case to processing operations concerning public security , defence , State security (including the economic well-being of the State when the processing operation relates to State security matters) and the activities of the State in areas of criminal law, </FullText>
</Norm>
<Norm ref="S_A3_B2_D_B_B_H">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_This_Directive </Rule>
<ExcludesObject verdict="PartiallyCorrect"> the processing of personal data </ExcludesObject>
<ExcludesObject verdict="Wrong"> a natural person in the course of a purely personal or household activity </ExcludesObject>
<FullText> This Directive shall not apply to the processing of personal data , by a natural person in the course of a purely personal or household activity . </FullText>
</Norm>
<Norm ref="S_A4_B1_D_B_B_La_R">
<NormType verdict="Wrong" note="Obligation"> Scope </NormType>
<Rule verdict="Wrong"> Each Member State </Rule>
<##Condition verdict="PartiallyCorrect"> where </##Condition>
<ActiveRole verdict="Missing"> Each Member State </ActiveRole>
<Action verdict="Missing"> apply the national provisions it adopts pursuant to this Directive to the processing of personal data

</Action>
<FullText> Each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data where , the processing is carried out in the context of the activities of an establishment of the controller on the territory of the Member State ; </FullText>
</Norm>
<Norm ref="S_A4_B1_D_B_B_La_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action "Wrong"> he </Action>
<Action verdict="PartiallyCorrect"> take t necessary measures </Action>
<Condition verdict="PartiallyCorrect"> wn t same controller is establisd on t territory of several Member States </Condition>
<Reason verdict="Wrong"> ensure that each of tse establishments complies with t obligations laid down by t national law applicable </Reason>
<FullText> when the same controller is established on the territory of several Member States , he must take the necessary measures to ensure that each of these establishments complies with the obligations laid down by the national law applicable ; </FullText>
</Norm>
<Norm ref="S_A4_B1_D_B_B_Lb_R">
<NormType verdict="Wrong" note="Obligation"> Scope </NormType>
<Rule verdict="Wrong"> Each Member State </Rule>
<##Condition verdict="Wrong"> where its national law applies by virtue of international public law </##Condition>
<ActiveRole verdict="Missing"> Each Member State </##Condition>
<Action verdict="Missing"> apply the national provisions it adopts pursuant to this Directive to the processing of personal data where , the controller is not established on the Member State 's territory , but in a place where its national law applies by virtue of international public law </Action>
<FullText> Each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data where , the controller is not established on the Member State 's territory , but in a place where its national law applies by virtue of international public law ; </FullText>
</Norm>
<Norm ref="S_A4_B1_D_B_B_Lc_R">
<NormType verdict="Wrong" note="Obligation"> Scope </NormType>
<Rule verdict="Wrong"> Each Member State </Rule>
<##Condition verdict="PartiallyCorrect"> where </##Condition>
<##Exception verdict="FullyCorrect"> unless such equipment is

used only for purposes of transit through the territory of the Community </##Exception>

<ActiveRole verdict="Missing"> Each Member State </ActiveRole>

<Action verdict="Missing"> apply the national provisions it adopts pursuant to this Directive to the processing of personal data </Action>

<FullText> Each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data where , the controller is not established on Community territory and , for purposes of processing personal data makes use of equipment , automated or otherwise , situated on the territory of the said Member State , unless such equipment is used only for purposes of transit through the territory of the Community . </FullText>
</Norm>
<Norm ref="S_A4_B1_D_B_B2_D">

<NormType verdict="FullyCorrect"> Obligation </NormType>

<ActiveRole verdict="FullyCorrect"> the controller </ActiveRole>

<PassiveRole verdict="FullyCorrect"> a representative established in the territory of that Member State </PassiveRole>

<Action verdict="PartiallyCorrect"> In the circumstances referred to in S_paragraph_1_(_c_) , designate without prejudice to legal actions which could be initiated against himself </Action>

<Condition verdict="Missing"> In the circumstances referred to in S_paragraph_1_(_c_) </Condition>

<FullText> In the circumstances referred to in paragraph 1 (c) , the controller must designate a representative established in the territory of that Member State , without prejudice to legal actions which could be initiated against the controller himself . </FullText>
</Norm>
<Norm ref="S_A5">

<NormType verdict="FullyCorrect"> Obligation </NormType>

<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>

<Action verdict="Wrong"> determine more precisely </Action>

<Action verdict="PartiallyCorrect"> determine the conditions under which the processing of personal data is lawful </Action>

<Timeframe verdict="Wrong"> within the limits of the provisions of S_this_Chapter </Timeframe>

<Condition verdict="Missing"> within the limits of the provisions of S_this_Chapter </Condition>

<FullText> Member States shall , within the limits of the provisions of this Chapter , determine more precisely the conditions under which the processing of personal data is lawful . </FullText>
</Norm>

<Norm ref="S_A6_B1_D_B_B_La_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect" note="NormalisationErrorBecome"> provide that personal data must are , processed fairly and lawfully </Action>
<FullText> Member States shall provide that personal data must be , processed fairly and lawfully ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_La_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="Wrong"> Member States </ActiveRole>
<Action verdict="FullyCorrect"> personal data , processed fairly and lawfully </Action>
<FullText> Member States shall provide that personal data must be , processed fairly and lawfully ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Lb_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect" note="NormalisationErrorBecome"> provide that personal data must are , collected for specified , explicit and legitimate purposes and not further processed in a way incompatible with those purposes </Action>
<FullText> Member States shall provide that personal data must be , collected for specified , explicit and legitimate purposes and not further processed in a way incompatible with those purposes . </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Lb_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action verdict="PartiallyCorrect"> personal data </Action>
<ActiveRole verdict="Missing"> Member States </ActiveRole>
<FullText> Member States shall provide that personal data must be , collected for specified , explicit and legitimate purposes and not further processed in a way incompatible with those purposes . </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Lb_R">
<NormType verdict="Wrong" note="Obligation"> LegalEffect </NormType>
<Situation verdict="Wrong"> Further processing of data for historical , statistical or scientific purposes </Situation>

<Result verdict="Wrong"> considered as incompatible </Result>
<##Condition verdict="FullyCorrect"> provided that Member States provide appropriate safeguards </##Condition>
<Action verdict="Missing"> Further processing of data for historical , statistical or scientific purposes shall not be considered as incompatible </Action>
<FullText> Further processing of data for historical , statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Lc_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect" note="NormalisationErrorBecome"> provide that personal data must are , adequate , relevant and not excessive in relation to the purposes for which they are collected and/or further processed </Action>
<FullText> Member States shall provide that personal data must be, adequate , relevant and not excessive in relation to the purposes for which they are collected and/or further processed ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Lc_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action verdict="PartiallyCorrect"> personal data </Action>
<ActiveRole verdict="Missing"> Member States </ActiveRole>
<FullText> Member States shall provide that personal data must be , adequate , relevant and not excessive in relation to the purposes for which they are collected and/or further processed ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Ld_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect" note="NormalisationErrorBecome"> provide that personal data must are , accurate and , , kept up to date </Action>
<##Condition verdict="FullyCorrect"> where necessary </##Condition>
<FullText> Member States shall provide that personal data must be , accurate and , where necessary , kept up to date ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Ld_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action verdict="PartiallyCorrect"> personal data </Action>
<##Condition verdict="PartiallyCorrect"> where necessary </##Condition>
<FullText> Member States shall provide that personal data must

be , accurate and , where necessary , kept up to date ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Ld_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action verdict="PartiallyCorrect"> every reasonable step </Action>
<FullText> every reasonable step must be taken to ensure that data which are inaccurate or incomplete , having regard to the purposes for which they became collected or for which they are further processed , are erased or rectified ; </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Le_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect" note="NormalisationErrorBecome"> provide that personal data must are , kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data became collected or for which they are further processed </Action>
<FullText> Member States shall provide that personal data must be , kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data was collected or for which they are further processed . </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Le_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action verdict="PartiallyCorrect"> personal data </Action>
<ActiveRole verdict="Wrong"> Member States </ActiveRole>
<FullText> Member States shall provide that personal data must be , kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data was collected or for which they are further processed . </FullText>
</Norm>
<Norm ref="S_A6_B1_D_B_B_Le_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> lay appropriate safeguards </Action>
<Reason verdict="Wrong"> for personal data stored for longer periods for historical , statistical or scientific use </Reason>
<FullText> Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical , statistical or scientific use . </FullText>
</Norm>

<Norm ref="S_A6_B1_D_B_B2_D">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<Action verdict="PartiallyCorrect"> It </Action>
<ActiveRole verdict="Missing"> the controller </ActiveRole>
<FullText> It shall be for the controller to ensure that paragraph 1 is complied with . </FullText>
</Norm>
<Norm ref="S_A7_B_La_R_O">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide that personal data may be processed only , the data subjected has unambiguously given his consent </Action>
<##Condition verdict="PartiallyCorrect"> if </##Condition>
<FullText> Member States shall provide that personal data may be processed only if , the data subject has unambiguously given his consent ; </FullText>
</Norm>
<Norm ref="S_A7_B_Lb_R_O">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide that personal data may be processed only , processing is necessary for the performance of a contract to which the data subjected is party or </Action>
<##Condition verdict="PartiallyCorrect"> if </##Condition>
<##Reason verdict="PartiallyCorrect" note="NormalisationErrorBecome"> in order to take steps at the request of the data subjected prior to entering into a contract </##Reason>
<FullText> Member States shall provide that personal data may be processed only if , processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract ; </FullText>
</Norm>
<Norm ref="S_A7_B_Lc_R_O">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide that personal data may be processed only , processing is necessary for compliance with a legal obligation to which the controller is subjected </Action>
<##Condition verdict="PartiallyCorrect"> if </##Condition>
<FullText> Member States shall provide that personal data may be processed only if , processing is necessary for compliance with a

legal obligation to which the controller is subject ; </FullText>
</Norm>
<Norm ref="S_A7_B_Ld_R_O">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide that personal data may be processed only , processing is necessary </Action>
<##Condition verdict="PartiallyCorrect"> if </##Condition>
<##Reason verdict="PartiallyCorrect" note="NormalisationErrorSubjected"> in order to protect the vital interests of the data subjected </##Reason> <FullText> Member States shall provide that personal data may be processed only if , processing is necessary in order to protect the vital interests of the data subject ; </FullText>
</Norm>
<Norm ref="S_A7_B_Le_R_O">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide that personal data may be processed only , processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed </Action>
<##Condition verdict="PartiallyCorrect"> if </##Condition>
<FullText> Member States shall provide that personal data may be processed only if , processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed ; </FullText>
</Norm>
<Norm ref="S_A7_B_Lf_R_O">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide that personal data may be processed only , processing is necessary for the purposes of the legitimate interests pursued by the third party or parties to whom the data are disclosed , and freedoms of the data subjected which require protection under S_Article_1_(_1_) </Action>
<##Condition verdict="PartiallyCorrect"> if </##Condition>
<##Exception verdict="FullyCorrect"> except where such interests are overridden by the interests for fundamental rights </##Exception>
<##PassiveRole verdict="PartiallyCorrect"> the controller </##PassiveRole>
<FullText> Member States shall provide that personal data may

be processed only if , processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed , except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1) . </FullText>
</Norm>
<Norm ref="S_A8_B1_D">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="PartiallyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> prohibit the processing of personal data revealing racial or ethnic origin , political opinions , religious or philosophical beliefs , trade-union membership , and the processing of data concerning health or sex life </Action>
<FullText> Member States shall prohibit the processing of personal data revealing racial or ethnic origin , political opinions , religious or philosophical beliefs , trade-union membership , and the processing of data concerning health or sex life . </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B_La_R_O">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_Paragraph_1 </Rule>
<ExcludesObject verdict="PartiallyCorrect"> the processing of those data </ExcludesObject>
<Exception verdict="PartiallyCorrect"> except the laws of the Member State provide that the prohibition referred to in S_paragraph_1 may not are lifted by the data subjected 's giving his consent </Exception>
<##Condition verdict="Wrong"> where </##Condition>
<FullText> Paragraph 1 shall not apply where , the data subject has given his explicit consent to the processing of those data , except where the laws of the Member State provide that the prohibition referred to in paragraph 1 may not are lifted by the data subject 's giving his consent ; </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B_Lb_R_O">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_Paragraph_1 </Rule>
<##Condition verdict="Wrong"> where </##Condition>
<Excludes verdict="Missing"> where , processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law </Excludes>
<Condition verdict="Missing"> in so far as it is authorized by national law providing for adequate safeguards </Condition>
<FullText> Paragraph 1 shall not apply where , processing is

necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards ; </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B_Lc_R_O">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_Paragraph_1 </Rule>
<##Condition verdict="PartiallyCorrect" note="NormalisationErrorSubjected"> where the data subjected is physically or legally incapable of giving his consent </##Condition>
<Excludes verdict="Missing"> where processing is necessary to protect the vital interests of the data subject or of another person
<FullText> Paragraph 1 shall not apply where , processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent ; </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B_Ld_R_O">
<NormType verdict="Wrong" note="Scope"> Unknown </NormType>
<Rule verdict="Missing"> S_Paragraph_1 </Rule>
<ExcludesObject verdict="Missing"> where , processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation , association or any other non-profit-seeking body with a political , philosophical , religious or trade-union aim </ExcludesObject>
<Condition verdict="Missing"> on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects </Condition>
<FullText> Paragraph 1 shall not apply where , processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation , association or any other non-profit-seeking body with a political , philosophical , religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects ; </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B_Le_R_O">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_Paragraph_1 </Rule>
<##Condition verdict="Wrong"> where </##Condition>
<ExcludesObject verdict="Missing"> where the processing relates

to data which are manifestly made public by the data subject or is necessary for the establishment , exercise or defence of legal claims </ExcludesObject>

<FullText> Paragraph 1 shall not apply where , the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment , exercise or defence of legal claims . </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B3_D">
<NormType verdict="FullyCorrect"> Scope </NormType>
<Rule verdict="FullyCorrect"> S_Paragraph_1 </Rule>
<##Condition verdict="Wrong" note="ExcludesObject"> where processing of the data is required for the purposes of preventive medicine </##Condition>
<##Condition verdict="Wrong" note="ExcludesObject"> where those data are processed by a health professional subjected under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subjected to an equivalent obligation of secrecy </##Condition>
<FullText> Paragraph 1 shall not apply where processing of the data is required for the purposes of preventive medicine , medical diagnosis , the provision of care or treatment or the management of health-care services , and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy . </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B4_D">
<NormType verdict="Wrong" note="Power"> Unknown </NormType>
<ActiveRole verdict="Missing"> Member States </ActiveRole>
<Reason verdict="Missing"> for reasons of substantial public interest </Reason>
<Action verdict="Missing"> lay down exemptions in addition to those laid down in paragraph 2 either by national law or by decision of the supervisory authority </Action>
<FullText> Subject to the provision of suitable safeguards , Member States may , for reasons of substantial public interest , lay down exemptions in addition to those laid down in paragraph 2 either by national law or by decision of the supervisory authority . </FullText>
</Norm>
<Norm ref="S_A8_B2_D_B_B5_D">
<NormType verdict="FullyCorrect"> Permission </NormType>
<Action verdict="PartiallyCorrect"> are carried out only under the control of official authority , or , subjected to derogations which

may be granted by the Member State under national provisions providing suitable specific safeguards </Action>

<##Condition verdict="FullyCorrect"> if suitable specific safeguards are provided under national law </##Condition>

<FullText> Processing of data relating to offences , criminal convictions or security measures may be carried out only under the control of official authority , or if suitable specific safeguards are provided under national law , subject to derogations which may be granted by the Member State under national provisions providing suitable specific safeguards . </FullText>

</Norm>

<Norm ref="S_A8_B2_D_B_B5_D">

<NormType verdict="FullyCorrect"> Permission </NormType>

<Action verdict="PartiallyCorrect"> are kept only under the control of official authority </Action>

<FullText> A complete register of criminal convictions may be kept only under the control of official authority . </FullText>

</Norm>

<Norm ref="S_A8_B2_D_B_B5_D">

<NormType verdict="FullyCorrect"> Permission </NormType>

<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>

<Action verdict="PartiallyCorrect" note="NormalisationErrorBecome"> provide that data relating to administrative sanctions or judgements in civil cases shall also be processed under the control of official authority </Action>

<FullText> Member States may provide that data relating to administrative sanctions or judgements in civil cases shall also be processed under the control of official authority . </FullText>

</Norm>

<Norm ref="S_A8_B2_D_B_B5_D">

<NormType verdict="FullyCorrect"> Obligation </NormType>

<ActiveRole verdict="Missing"> Member States </ActiveRole>

<Action verdict="PartiallyCorrect"> data relating to administrative sanctions or judgements in civil cases </Action>

<FullText> Member States may provide that data relating to administrative sanctions or judgements in civil cases shall also be processed under the control of official authority . </FullText>

</Norm>

<Norm ref="S_A8_B2_D_B_B6_D">

<NormType verdict="FullyCorrect"> Obligation </NormType>

<Action verdict="PartiallyCorrect"> in S_paragraphs_4_and_5 </Action>

<ActiveRole verdict="Missing"> the Commission </ActiveRole>

<FullText> paragraph 1 provided for in paragraphs 4 and 5 shall be notified to the Commission . </FullText>

</Norm>
<Norm ref="S_A8_B2_D_B_B7_D">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="FullyCorrect"> determine the conditions under which a national identification number or any other identifier of general application may be processed </Action>
<FullText> Member States shall determine the conditions under which a national identification number or any other identifier of general application may be processed . </FullText>
</Norm>
<Norm ref="S_A9">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<Action verdict="PartiallyCorrect"> provide for exemptions or derogations from the provisions of S_this_Chapter , S_Chapter_IV and S_Chapter_VI </Action>
<Action verdict="Wrong"> provide for the processing of personal data carried out solely for journalistic purposes or the purpose of artistic or literary expression only </Action>
<##Condition verdict="FullyCorrect"> if they are necessary to reconcile the right to privacy with the rules governing freedom of expression </##Condition>
<FullText> Member States shall provide for exemptions or derogations from the provisions of this Chapter , Chapter IV and Chapter VI for the processing of personal data carried out solely for journalistic purposes or the purpose of artistic or literary expression only if they are necessary to reconcile the right to privacy with the rules governing freedom of expression . </FullText>
</Norm>
<Norm ref="S_A10_B_La_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<PassiveRole verdict="PartiallyCorrect"> that the controller or his representative must provide a data subjected from whom data relating to himself are collected with at least the following information , and of his representative , </PassiveRole>
<##Action verdict="PartiallyCorrect"> provide ; </##Action>
<##Exception verdict="PartiallyCorrect"> except where he already has it : the identity of the controller </##Exception>
<##Condition verdict="FullyCorrect"> if any </##Condition>
<FullText> Member States shall provide that the controller or his representative must provide a data subject from whom data relating

to himself are collected with at least the following information , except where he already has it : the identity of the controller and of his representative , if any ; </FullText>
</Norm>
<Norm ref="S_A10_B_La_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="PartiallyCorrect"> the controller or his representative </ActiveRole>
<Action verdict="Wrong"> provide a data subjected from whom </Action>
<Action verdict="PartiallyCorrect"> provide data relating to himself are collected with at least the following information , and of his representative , </Action>
<##Exception verdict="PartiallyCorrect"> except where he already has it : the identity of the controller </##Exception>
<##Condition verdict="FullyCorrect"> if any </##Condition>
<FullText> Member States shall provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information , except where he already has it : the identity of the controller and of his representative , if any ; </FullText>
</Norm>
<Norm ref="S_A10_B_Lb_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>
<PassiveRole verdict="PartiallyCorrect"> that the controller or his representative must provide a data subjected from whom data relating to himself are collected with at least the following information , </PassiveRole>
<##Action verdict="PartiallyCorrect"> provide ; </##Action>
<##Exception verdict="PartiallyCorrect"> except where he already has it : the purposes of the processing for which the data are intended </##Exception>
<FullText> Member States shall provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information , except where he already has it : the purposes of the processing for which the data are intended ; </FullText>
</Norm>
<Norm ref="S_A10_B_Lb_R">
<NormType verdict="FullyCorrect"> Obligation </NormType>
<ActiveRole verdict="FullyCorrect"> the controller or his representative </ActiveRole>
<Action verdict="Wrong"> provide a data subjected from whom </Action>

<Action verdict="PartiallyCorrect"> provide data relating to himself are collected with at least the following information , </Action>

<##Exception verdict="PartiallyCorrect"> except where he already has it : the purposes of the processing for which the data are intended </##Exception>

<FullText> Member States shall provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information , except where he already has it : the purposes of the processing for which the data are intended ; </FullText>

</Norm>

<Norm ref="S_A10_B_Lc_R_B_B_H">

<NormType verdict="FullyCorrect"> Obligation </NormType>

<ActiveRole verdict="FullyCorrect"> Member States </ActiveRole>

<PassiveRole verdict="PartiallyCorrect"> that the controller or his representative must provide a data subjected from whom data relating to himself are collected with the following information , except he already has it : any further information such as the recipients or categories of recipients of the data </PassiveRole>

<##Action verdict="PartiallyCorrect"> provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information , , : in so far as such further information is necessary , having regard to the specific circumstances in which the data are collected , to guarantee fair processing in respect of the data subjected </##Action>

<##Exception verdict="PartiallyCorrect"> except where he already has it : any further information such as the recipients or categories of recipients of the data </##Exception>

<Condition verdict="Missing"> in so far as such further information is necessary </Condition>

<Reason verdict="Missing"> to guarantee fair processing in respect of the data subject </Reason>

<FullText> Member States shall provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information , except where he already has it : any further information such as the recipients or categories of recipients of the data , : in so far as such further information is necessary , having regard to the specific circumstances in which the data are collected , to guarantee fair processing in respect of the data subject . </FullText>

</Norm>

</Legislation>

14

APPENDIX C: GOLD MAPPING OF NORMS TO RECITALS: DIRECTIVE D95/46/EC

This appendix contains the mapping of recitals to the first articles of this Directive. The complete file is available upon request.

ARTICLE	**RECITAL(S)**
Article 1 Object of the Directive 1. In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data. 2. Member States shall neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1.	(1) Whereas the objectives of the Community, as laid down in the Treaty, as amended by the Treaty on European Union, include creating an ever closer union among the peoples of Europe, fostering closer relations between the States belonging to the Community, ensuring economic and social progress by common action to eliminate the barriers which divide Europe, encouraging the constant improvement of the living conditions of its peoples, preserving and strengthening peace and liberty and promoting democracy on the basis of the fundamental rights recognized in the constitution and laws of the Member States and in the European Convention for the Protection of Human Rights and Fundamental Freedoms. (2) Whereas data-processing systems are designed to serve man. Whereas they must, whatever the nationality or residence of natural persons, respect their fundamental rights and freedoms, notably the right to privacy, and contribute to economic and social progress, trade expansion and the well-being of individuals.

ARTICLE	RECITAL(S)
	(3) Whereas the establishment and functioning of an internal market in which, in accordance with Article 7a of the Treaty, the free movement of goods, persons, services and capital is ensured require not only that personal data should be able to flow freely from one Member State to another, but also that the fundamental rights of individuals should be safeguarded.
Article 1 Object of the Directive 1. In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data. 2. Member States shall neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1.	(7) Whereas the difference in levels of protection of the rights and freedoms of individuals, notably the right to privacy, with regard to the processing of personal data afforded in the Member States may prevent the transmission of such data from the territory of one Member State to that of another Member State. Whereas this difference may therefore constitute an obstacle to the pursuit of a number of economic activities at Community level, distort competition and impede authorities in the discharge of their responsibilities under Community law. Whereas this difference in levels of protection is due to the existence of a wide variety of national laws, regulations and administrative provisions.

ARTICLE	RECITAL(S)
	(8) Whereas, in order to remove the obstacles to flows of personal data, the level of protection of the rights and freedoms of individuals with regard to the processing of such data must be equivalent in all Member States. Whereas this objective is vital to the internal market but cannot be achieved by the Member States alone, especially in view of the scale of the divergences which currently exist between the relevant laws in the Member States and the need to coordinate the laws of the Member States so as to ensure that the cross-border flow of personal data is regulated in a consistent manner that is in keeping with the objective of the internal market as provided for in Article 7a of the Treaty. Whereas Community action to approximate those laws is therefore needed.
Article 1 Object of the Directive 1. In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data. 2. Member States shall neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1.	(9) Whereas, given the equivalent protection resulting from the approximation of national laws, the Member States will no longer be able to inhibit the free movement between them of personal data on grounds relating to protection of the rights and freedoms of individuals, and in particular the right to privacy. Whereas Member States will be left a margin for manoeuvre, which may, in the context of implementation of the Directive, also be exercised by the business and social partners. Whereas Member States will therefore be able to specify in their national law the general conditions governing the lawfulness of data processing. Whereas in doing so the Member States shall strive to improve the protection currently provided by their legislation. Whereas, within the limits of this margin for manoeuvre and in accordance with Community law, disparities could arise in the implementation of the Directive, and this could have an effect on the movement of data within a Member State as well as within the Community.

ARTICLE	RECITAL(S)
	(10) Whereas the object of the national laws on the processing of personal data is to protect fundamental rights and freedoms, notably the right to privacy, which is recognized both in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms and in the general principles of Community law. Whereas, for that reason, the approximation of those laws must not result in any lessening of the protection they afford but must, on the contrary, seek to ensure a high level of protection in the Community.
Article 3 Scope 1. This Directive shall apply to the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system. 2. This Directive shall not apply to the processing of personal data: - in the course of an activity which falls outside the scope of Community law, such as those provided for by Titles V and VI of the Treaty on European Union and in any case to processing operations concerning public security, defence, State security (including the economic well-being of the State when the processing operation relates to State security matters) and the activities of the State in areas of criminal law, - by a natural person in the course of a purely personal or household activity.	(12) Whereas the protection principles must apply to all processing of personal data by any person whose activities are governed by Community law. Whereas there should be excluded the processing of data carried out by a natural person in the exercise of activities which are exclusively personal or domestic, such as correspondence and the holding of records of addresses. (13) Whereas the acitivities referred to in Titles V and VI of the Treaty on European Union regarding public safety, defence, State security or the acitivities of the State in the area of criminal laws fall outside the scope of Community law, without prejudice to the obligations incumbent upon Member States under Article 56 (2), Article 57 or Article 100a of the Treaty establishing the European Community. Whereas the processing of personal data that is necessary to safeguard the economic well-being of the State does not fall within the scope of this Directive where such processing relates to State security matters.

ARTICLE	RECITAL(S)
	(14) Whereas, given the importance of the developments under way, in the framework of the information society, of the techniques used to capture, transmit, manipulate, record, store or communicate sound and image data relating to natural persons, this Directive should be applicable to processing involving such data. (15) Whereas the processing of such data is covered by this Directive only if it is automated or if the data processed are contained or are intended to be contained in a filing system structured according to specific criteria relating to individuals, so as to permit easy access to the personal data in question.

ARTICLE	RECITAL(S)
Article 3 Scope 1. This Directive shall apply to the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system. 2. This Directive shall not apply to the processing of personal data: - in the course of an activity which falls outside the scope of Community law, such as those provided for by Titles V and VI of the Treaty on European Union and in any case to processing operations concerning public security, defence, State security (including the economic well-being of the State when the processing operation relates to State security matters) and the activities of the State in areas of criminal law, - by a natural person in the course of a purely personal or household activity.	(16) Whereas the processing of sound and image data, such as in cases of video surveillance, does not come within the scope of this Directive if it is carried out for the purposes of public security, defence, national security or in the course of State activities relating to the area of criminal law or of other activities which do not come within the scope of Community law. (21) Whereas this Directive is without prejudice to the rules of territoriality applicable in criminal matters. (27) Whereas the protection of individuals must apply as much to automatic processing of data as to manual processing. Whereas the scope of this protection must not in effect depend on the techniques used, otherwise this would create a serious risk of circumvention. Whereas, nonetheless, as regards manual processing, this Directive covers only filing systems, not unstructured files. Whereas, in particular, the content of a filing system must be structured according to specific criteria relating to individuals allowing easy access to the personal data. Whereas, in line with the definition in Article 2 (c), the different criteria for determining the constituents of a structured set of personal data, and the different criteria governing access to such a set, may be laid down by each Member State. Whereas files or sets of files as well as their cover pages, which are not structured according to specific criteria, shall under no circumstances fall within the scope of this Directive.

ARTICLE	RECITAL(S)
Article 4 National law applicable 1. Each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data where: (a) the processing is carried out in the context of the activities of an establishment of the controller on the territory of the Member State; when the same controller is established on the territory of several Member States, he must take the necessary measures to ensure that each of these establishments complies with the obligations laid down by the national law applicable; (b) the controller is not established on the Member State's territory, but in a place where its national law applies by virtue of international public law; (c) the controller is not established on Community territory and, for purposes of processing personal data makes use of equipment, automated or otherwise, situated on the territory of the said Member State, unless such equipment is used only for purposes of transit through the territory of the Community. 2. In the circumstances referred to in paragraph 1 (c), the controller must designate a representative established in the territory of that Member State, without prejudice to legal actions which could be initiated against the controller himself.	(18) Whereas, in order to ensure that individuals are not deprived of the protection to which they are entitled under this Directive, any processing of personal data in the Community must be carried out in accordance with the law of one of the Member States. Whereas, in this connection, processing carried out under the responsibility of a controller who is established in a Member State should be governed by the law of that State. (19) Whereas establishment on the territory of a Member State implies the effective and real exercise of activity through stable arrangements. Whereas the legal form of such an establishment, whether simply branch or a subsidiary with a legal personality, is not the determining factor in this respect. Whereas, when a single controller is established on the territory of several Member States, particularly by means of subsidiaries, he must ensure, in order to avoid any circumvention of national rules, that each of the establishments fulfils the obligations imposed by the national law applicable to its activities. (20) Whereas the fact that the processing of data is carried out by a person established in a third country must not stand in the way of the protection of individuals provided for in this Directive. Whereas in these cases, the processing should be governed by the law of the Member State in which the means used are located, and there should be guarantees to ensure that the rights and obligations provided for in this Directive are respected in practice.

ARTICLE	RECITAL(S)
Article 5 Member States shall, within the limits of the provisions of this Chapter, determine more precisely the conditions under which the processing of personal data is lawful.	(9) Whereas, given the equivalent protection resulting from the approximation of national laws, the Member States will no longer be able to inhibit the free movement between them of personal data on grounds relating to protection of the rights and freedoms of individuals, and in particular the right to privacy. Whereas Member States will be left a margin for manoeuvre, which may, in the context of implementation of the Directive, also be exercised by the business and social partners. Whereas Member States will therefore be able to specify in their national law the general conditions governing the lawfulness of data processing. Whereas in doing so the Member States shall strive to improve the protection currently provided by their legislation. Whereas, within the limits of this margin for manoeuvre and in accordance with Community law, disparities could arise in the implementation of the Directive, and this could have an effect on the movement of data within a Member State as well as within the Community. (22) Whereas Member States shall more precisely define in the laws they enact or when bringing into force the measures taken under this Directive the general circumstances in which processing is lawful. Whereas in particular Article 5, in conjunction with Articles 7 and 8, allows Member States, independently of general rules, to provide for special processing conditions for specific sectors and for the various categories of data covered by Article 8.

ARTICLE	RECITAL(S)
Article 5 Member States shall, within the limits of the provisions of this Chapter, determine more precisely the conditions under which the processing of personal data is lawful.	(23) Whereas Member States are empowered to ensure the implementation of the protection of individuals both by means of a general law on the protection of individuals as regards the processing of personal data and by sectorial laws such as those relating, for example, to statistical institutes.
Article 6 1. Member States shall provide that personal data must be: (a) processed fairly and lawfully; (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards; (c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed; (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified; (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use. 2. It shall be for the controller to ensure that paragraph 1 is complied with.	(22) Whereas Member States shall more precisely define in the laws they enact or when bringing into force the measures taken under this Directive the general circumstances in which processing is lawful. Whereas in particular Article 5, in conjunction with Articles 7 and 8, allows Member States, independently of general rules, to provide for special processing conditions for specific sectors and for the various categories of data covered by Article 8. (26) Whereas the principles of protection must apply to any information concerning an identified or identifiable person. Whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person. Whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable. Whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible.

ARTICLE	RECITAL(S)
Article 6 1. Member States shall provide that personal data must be: (a) processed fairly and lawfully; (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards; (c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed; (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified; (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use. 2. It shall be for the controller to ensure that paragraph 1 is complied with.	(28) Whereas any processing of personal data must be lawful and fair to the individuals concerned. Whereas, in particular, the data must be adequate, relevant and not excessive in relation to the purposes for which they are processed. Whereas such purposes must be explicit and legitimate and must be determined at the time of collection of the data. Whereas the purposes of processing further to collection shall not be incompatible with the purposes as they were originally specified. (29) Whereas the further processing of personal data for historical, statistical or scientific purposes is not generally to be considered incompatible with the purposes for which the data have previously been collected provided that Member States furnish suitable safeguards. Whereas these safeguards must in particular rule out the use of the data in support of measures or decisions regarding any particular individual.

ARTICLE	RECITAL(S)
Article 7 Member States shall provide that personal data may be processed only if: (a) the data subject has unambiguously given his consent; or (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or (c) processing is necessary for compliance with a legal obligation to which the controller is subject; or (d) processing is necessary in order to protect the vital interests of the data subject; or (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed; or (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1).	(22) Whereas Member States shall more precisely define in the laws they enact or when bringing into force the measures taken under this Directive the general circumstances in which processing is lawful. Whereas in particular Article 5, in conjunction with Articles 7 and 8, allows Member States, independently of general rules, to provide for special processing conditions for specific sectors and for the various categories of data covered by Article 8. (30) Whereas, in order to be lawful, the processing of personal data must in addition be carried out with the consent of the data subject or be necessary for the conclusion or performance of a contract binding on the data subject, or as a legal requirement, or for the performance of a task carried out in the public interest or in the exercise of official authority, or in the legitimate interests of a natural or legal person, provided that the interests or the rights and freedoms of the data subject are not overriding. Whereas, in particular, in order to maintain a balance between the interests involved while guaranteeing effective competition, Member States may determine the circumstances in which personal data may be used or disclosed to a third party in the context of the legitimate ordinary business activities of companies and other bodies. Whereas Member States may similarly specify the conditions under which personal data may be disclosed to a third party for the purposes of marketing whether carried out commercially or by a charitable organization or by any other association or foundation, of a political nature for example, subject to the provisions allowing a data subject to object to the processing of data regarding him, at no cost and without having to state his reasons. (31) Whereas the processing of personal data must equally be regarded as lawful where it is carried out in order to protect an interest which is essential for the data subject's life.

ARTICLE	RECITAL(S)
Article 7 Member States shall provide that personal data may be processed only if: (a) the data subject has unambiguously given his consent; or (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or (c) processing is necessary for compliance with a legal obligation to which the controller is subject; or (d) processing is necessary in order to protect the vital interests of the data subject; or (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed; or (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1).	(32) Whereas it is for national legislation to determine whether the controller performing a task carried out in the public interest or in the exercise of official authority should be a public administration or another natural or legal person governed by public law, or by private law such as a professional association.

ARTICLE	RECITAL(S)
Article 8 The processing of special categories of data 1. Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life. 2. Paragraph 1 shall not apply where: (a) the data subject has given his explicit consent to the processing of those data, except where the laws of the Member State provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject's giving his consent; or (b) processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or (c) processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or (d) processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or (e) the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.	(22) Whereas Member States shall more precisely define in the laws they enact or when bringing into force the measures taken under this Directive the general circumstances in which processing is lawful. Whereas in particular Article 5, in conjunction with Articles 7 and 8, allows Member States, independently of general rules, to provide for special processing conditions for specific sectors and for the various categories of data covered by Article 8. (33) Whereas data which are capable by their nature of infringing fundamental freedoms or privacy should not be processed unless the data subject gives his explicit consent. Whereas, however, derogations from this prohibition must be explicitly provided for in respect of specific needs, in particular where the processing of these data is carried out for certain health-related purposes by persons subject to a legal obligation of professional secrecy or in the course of legitimate activities by certain associations or foundations the purpose of which is to permit the exercise of fundamental freedoms. (34) Whereas Member States must also be authorized, when justified by grounds of important public interest, to derogate from the prohibition on processing sensitive categories of data where important reasons of public interest so justify in areas such as public health and social protection - especially in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system - scientific research and government statistics. Whereas it is incumbent on them, however, to provide specific and suitable safeguards so as to protect the fundamental rights and the privacy of individuals. (35) Whereas, moreover, the processing of personal data by official authorities for achieving aims, laid down in constitutional law or international public law, of officially recognized religious associations is carried out on important grounds of public interest.

ARTICLE	RECITAL(S)
3. Paragraph 1 shall not apply where processing of the data is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy. 4. Subject to the provision of suitable safeguards, Member States may, for reasons of substantial public interest, lay down exemptions in addition to those laid down in paragraph 2 either by national law or by decision of the supervisory authority. 5. Processing of data relating to offences, criminal convictions or security measures may be carried out only under the control of official authority, or if suitable specific safeguards are provided under national law, subject to derogations which may be granted by the Member State under national provisions providing suitable specific safeguards. However, a complete register of criminal convictions may be kept only under the control of official authority. Member States may provide that data relating to administrative sanctions or judgements in civil cases shall also be processed under the control of official authority. 6. Derogations from paragraph 1 provided for in paragraphs 4 and 5 shall be notified to the Commission. 7. Member States shall determine the conditions under which a national identification number or any other identifier of general application may be processed.	(36) Whereas where, in the course of electoral activities, the operation of the democratic system requires in certain Member States that political parties compile data on people's political opinion, the processing of such data may be permitted for reasons of important public interest, provided that appropriate safeguards are established.

ARTICLE	RECITAL(S)
Article 9 Processing of personal data and freedom of expression Member States shall provide for exemptions or derogations from the provisions of this Chapter, Chapter IV and Chapter VI for the processing of personal data carried out solely for journalistic purposes or the purpose of artistic or literary expression only if they are necessary to reconcile the right to privacy with the rules governing freedom of expression.	(17) Whereas, as far as the processing of sound and image data carried out for purposes of journalism or the purposes of literary or artistic expression is concerned, in particular in the audiovisual field, the principles of the Directive are to apply in a restricted manner according to the provisions laid down in Article 9. (37) Whereas the processing of personal data for purposes of journalism or for purposes of literary of artistic expression, in particular in the audiovisual field, should qualify for exemption from the requirements of certain provisions of this Directive in so far as this is necessary to reconcile the fundamental rights of individuals with freedom of information and notably the right to receive and impart information, as guaranteed in particular in Article 10 of the European Convention for the Protection of Human Rights and Fundamental Freedoms. Whereas Member States should therefore lay down exemptions and derogations necessary for the purpose of balance between fundamental rights as regards general measures on the legitimacy of data processing, measures on the transfer of data to third countries and the power of the supervisory authority. Whereas this should not, however, lead Member States to lay down exemptions from the measures to ensure security of processing. Whereas at least the supervisory authority responsible for this sector should also be provided with certain ex-post powers, e.g. to publish a regular report or to refer matters to the judicial authorities.

ARTICLE	RECITAL(S)
Article 10 Information in cases of collection of data from the data subject Member States shall provide that the controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information, except where he already has it: (a) the identity of the controller and of his representative, if any; (b) the purposes of the processing for which the data are intended; (c) any further information such as - the recipients or categories of recipients of the data, - whether replies to the questions are obligatory or voluntary, as well as the possible consequences of failure to reply, - the existence of the right of access to and the right to rectify the data concerning him in so far as such further information is necessary, having regard to the specific circumstances in which the data are collected, to guarantee fair processing in respect of the data subject.	(38) Whereas, if the processing of data is to be fair, the data subject must be in a position to learn of the existence of a processing operation and, where data are collected from him, must be given accurate and full information, bearing in mind the circumstances of the collection.

ARTICLE	RECITAL(S)
Article 11 Information where the data have not been obtained from the data subject 1. Where the data have not been obtained from the data subject, Member States shall provide that the controller or his representative must at the time of undertaking the recording of personal data or if a disclosure to a third party is envisaged, no later than the time when the data are first disclosed provide the data subject with at least the following information, except where he already has it: (a) the identity of the controller and of his representative, if any; (b) the purposes of the processing; (c) any further information such as - the categories of data concerned, - the recipients or categories of recipients, - the existence of the right of access to and the right to rectify the data concerning him in so far as such further information is necessary, having regard to the specific circumstances in which the data are processed, to guarantee fair processing in respect of the data subject. 2. Paragraph 1 shall not apply where, in particular for processing for statistical purposes or for the purposes of historical or scientific research, the provision of such information proves impossible or would involve a disproportionate effort or if recording or disclosure is expressly laid down by law. In these cases Member States shall provide appropriate safeguards.	(39) Whereas certain processing operations involve data which the controller has not collected directly from the data subject. Whereas, furthermore, data can be legitimately disclosed to a third party, even if the disclosure was not anticipated at the time the data were collected from the data subject. Whereas, in all these cases, the data subject should be informed when the data are recorded or at the latest when the data are first disclosed to a third party. (40) Whereas, however, it is not necessary to impose this obligation of the data subject already has the information. Whereas, moreover, there will be no such obligation if the recording or disclosure are expressly provided for by law or if the provision of information to the data subject proves impossible or would involve disproportionate efforts, which could be the case where processing is for historical, statistical or scientific purposes. Whereas, in this regard, the number of data subjects, the age of the data, and any compensatory measures adopted may be taken into consideration.

ARTICLE	RECITAL(S)
Article 12 Right of access Member States shall guarantee every data subject the right to obtain from the controller: (a) without constraint at reasonable intervals and without excessive delay or expense: - confirmation as to whether or not data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed, - communication to him in an intelligible form of the data undergoing processing and of any available information as to their source, - knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in Article 15 (1); (b) as appropriate the rectification, erasure or blocking of data the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data; (c) notification to third parties to whom the data have been disclosed of any rectification, erasure or blocking carried out in compliance with (b), unless this proves impossible or involves a disproportionate effort.	(41) Whereas any person must be able to exercise the right of access to data relating to him which are being processed, in order to verify in particular the accuracy of the data and the lawfulness of the processing. Whereas, for the same reasons, every data subject must also have the right to know the logic involved in the automatic processing of data concerning him, at least in the case of the automated decisions referred to in Article 15 (1). Whereas this right must not adversely affect trade secrets or intellectual property and in particular the copyright protecting the software. Whereas these considerations must not, however, result in the data subject being refused all information.

ARTICLE	RECITAL(S)
Article 13 Exemptions and restrictions 1. Member States may adopt legislative measures to restrict the scope of the obligations and rights provided for in Articles 6 (1), 10, 11 (1), 12 and 21 when such a restriction constitutes a necessary measures to safeguard: (a) national security; (b) defence; (c) public security; (d) the prevention, investigation, detection and prosecution of criminal offences, or of breaches of ethics for regulated professions; (e) an important economic or financial interest of a Member State or of the European Union, including monetary, budgetary and taxation matters; (f) a monitoring, inspection or regulatory function connected, even occasionally, with the exercise of official authority in cases referred to in (c), (d) and (e); (g) the protection of the data subject or of the rights and freedoms of others. 2. Subject to adequate legal safeguards, in particular that the data are not used for taking measures or decisions regarding any particular individual, Member States may, where there is clearly no risk of breaching the privacy of the data subject, restrict by a legislative measure the rights provided for in Article 12 when data are processed solely for purposes of scientific research or are kept in personal form for a period which does not exceed the period necessary for the sole purpose of creating statistics.	(42) Whereas Member States may, in the interest of the data subject or so as to protect the rights and freedoms of others, restrict rights of access and information. Whereas they may, for example, specify that access to medical data may be obtained only through a health professional. (43) Whereas restrictions on the rights of access and information and on certain obligations of the controller may similarly be imposed by Member States in so far as they are necessary to safeguard, for example, national security, defence, public safety, or important economic or financial interests of a Member State or the Union, as well as criminal investigations and prosecutions and action in respect of breaches of ethics in the regulated professions. Whereas the list of exceptions and limitations should include the tasks of monitoring, inspection or regulation necessary in the three last-mentioned areas concerning public security, economic or financial interests and crime prevention. Whereas the listing of tasks in these three areas does not affect the legitimacy of exceptions or restrictions for reasons of State security or defence. (44) Whereas Member States may also be led, by virtue of the provisions of Community law, to derogate from the provisions of this Directive concerning the right of access, the obligation to inform individuals, and the quality of data, in order to secure certain of the purposes referred to above.

ARTICLE	RECITAL(S)
Article 14 The data subject's right to object Member States shall grant the data subject the right: (a) at least in the cases referred to in Article 7 (e) and (f), to object at any time on compelling legitimate grounds relating to his particular situation to the processing of data relating to him, save where otherwise provided by national legislation. Where there is a justified objection, the processing instigated by the controller may no longer involve those data; (b) to object, on request and free of charge, to the processing of personal data relating to him which the controller anticipates being processed for the purposes of direct marketing, or to be informed before personal data are disclosed for the first time to third parties or used on their behalf for the purposes of direct marketing, and to be expressly offered the right to object free of charge to such disclosures or uses. Member States shall take the necessary measures to ensure that data subjects are aware of the existence of the right referred to in the first subparagraph of (b).	(45) Whereas, in cases where data might lawfully be processed on grounds of public interest, official authority or the legitimate interests of a natural or legal person, any data subject should nevertheless be entitled, on legitimate and compelling grounds relating to his particular situation, to object to the processing of any data relating to himself. Whereas Member States may nevertheless lay down national provisions to the contrary.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

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

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of March 22, 2017 (`classicthesis` Final Version).