

# Towards Legal Knowledge Management Systems for Regulatory Compliance

Guido Boella<sup>1</sup>, Joris Hulstijn<sup>2</sup>, Llio Humphreys<sup>1,3</sup>, Marijn Janssen<sup>2</sup>, and  
Leendert van der Torre<sup>3</sup>

<sup>1</sup> Università di Torino, Italy

`guido@di.unito.it`, `llobh@gmail.com`

<sup>2</sup> Delft University of Technology, Netherlands

`j.hulstijn@tudelft.nl`, `M.F.W.H.A.Janssen@tudelft.nl`

<sup>3</sup> University of Luxembourg, Luxembourg

`leon.vandertorre@uni.lu`

**Abstract.** Maintaining regulatory compliance is an increasing concern. Legal Knowledge Management systems could support the work of compliance managers. However, there are challenges to overcome, of interpreting legal knowledge and mapping it onto business processes. In this discussion paper we determine requirements for Legal Knowledge Management systems that can handle these challenges. We argue that an approach to Legal Knowledge Management should consist of four pillars: (1) a legal ontology to express the legal concepts that are central to the interpretation of rules and regulations in practice, (2) natural language processing techniques, to semi-automatically populate the ontology, (3) a systematic method for mapping the legal concepts onto the actual data and business processes of a company, and (4) a method to construct an audit trail as part of the process design, providing legally acceptable evidence of compliance. The challenges are illustrated by an example of MIFID, a set of European regulations in the financial domain.

**Keywords:** regulatory compliance, legal ontology, evidence

## 1 Introduction

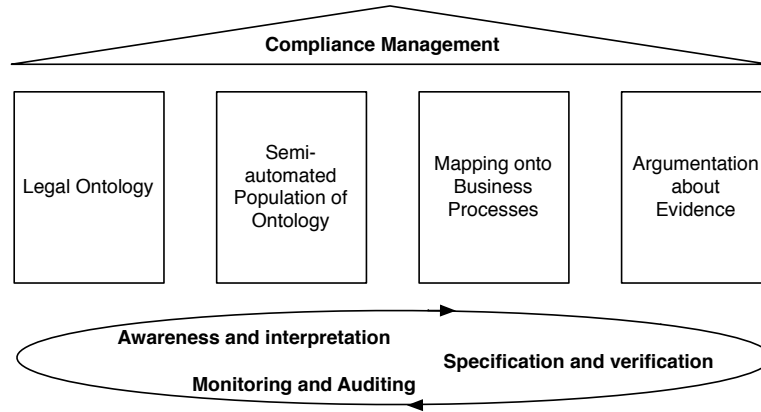
Every organization has to deal with operational risks: the risks arising from the execution of a company's primary business functions. Operational risks include monetary loss, fraud, physical or environmental risks, risks related to human resources, and so forth. Risk Management departments typically collect and assess data for each risk in order to make management decisions, increasingly also using ICT support. For one type of risk there is a lack of adequate ICT support: regulatory compliance. With the financial crisis the costs associated with managing regulatory compliance are becoming an important concern [12]. Large enterprises such as banks and insurance companies are increasingly being subjected to regulations. The risks of failure to comply with regulations are increasing in volume and complexity.

There is increasing awareness in the compliance world that information systems can play a larger role in improving compliance. Most attention has been devoted to the prevention side. Focus on prevention leads to an approach called *compliance by design* [16]: business processes and supporting information systems are designed in such a way that conducting undesirable behavior is made difficult or impossible. Process and system specifications are verified to conform to the stated control objectives (conformance testing) [22]. However, compliance is not just about specification and verification of processes, beforehand. It also involves monitoring and auditing the system after it has been implemented, in order to detect potential violations and provide evidence of compliance. Examples of developments in this direction are continuous auditing or continuous control monitoring [3, 18]. Moreover, the organization must know which rules and regulations apply to the business processes of the company. This requires awareness of new laws and regulations and the legal debate about its interpretations, to keep an up-to-date repository of legal prescriptions relevant to the operation. So in fact compliance management involves a whole *compliance life-cycle*, containing activities for regulatory awareness and interpretation (adoption), specification and verification (prevention) and monitoring and auditing operating effectiveness (detection). See also the COMPAS project [13].

Knowledge management systems could be developed to support the work of compliance managers in each of these activities. However, compliance officers are trained as legal scholars and are often reluctant to use ICT support. We believe current tool support is inadequate, because of specific challenges in processing and interpreting legal knowledge and applying it to actual business processes. Partly based on interviews with compliance officers, we identified four challenges.

- (1) Existing compliance support systems that model the law tend to extract norms and terminology directly from the letter of legislative sources, without taking the *legal interpretation process* into account, e.g. [19, 7].
- (2) State of the art legal knowledge management systems rely almost entirely on manual population of specialist ontologies, which is not only inefficient but impractical. It is difficult to find knowledge engineers with the right interdisciplinary competences, i.e., expertise in both formal ontologies and law [12]. In addition, a large enterprise will have hundreds of processes each with hundreds of activities and states. This leads to a *resource bottleneck*.
- (3) It is difficult to find connections between legal concepts and norms expressed in legal ontologies, and business processes, because the two ‘worlds’ do not share a *common terminology*. Norms are generic, and need to be adapted to a specific situation to be implemented [2, 9].
- (4) Current approaches to compliance management largely neglect the importance of *auditing and evidence*. Regulators will want to see proof that the process as designed was in indeed performed according to specification.

In this discussion paper we will derive requirements for an approach to legal knowledge management systems that can face these challenges. The research is part of a project called LeKMaS (Legal Knowledge Management Systems). We argue that an approach to Legal Knowledge Management should consist of four



**Fig. 1.** Overview of the four pillars and the activities in a compliance life-cycle

pillars, each addressing one of the challenges: (1) a legal ontology expressing the legal concepts that are central to the rules and regulations to serve as an intermediary between legal text and interpretation, (2) natural language processing techniques to semi-automatically populate the ontology, (3) a systematic method of adapting legal prescriptions to the context and mapping the legal concepts onto the actual data and business processes, (4) an analysis method to design an audit trail, providing legally acceptable evidence of compliance. The general objective is to develop an integrated approach to compliance, ensuring synergy between law, business processes and auditing. Figure 1 illustrates these four pillars, and the activities of the compliance life-cycle.

The remainder of this paper is structured as follows. In Section 2 – 6 we discuss the challenges and requirements for an approach to face them. The paper ends with conclusions and recommendations for further research.

## 2 Example: MIFID regulations

By way of illustration, please consider the Markets in Financial Instruments Directive 2004/39/EC (MIFID). MIFID is a EU directive for harmonizing legislation in member states regarding investment services. The directive aims to increase competition and provide consumer protection in investment services. The directive covers most tradable financial products, like brokerage, advice, dealing, portfolio management, underwriting etc. MIFID requires financial institutions to archive and provide frequent reports about transactions. These archiving and reporting obligations make the example particularly interesting for investigating the possibilities of compliance tool support. Moreover, in response to the financial crisis, MIFID has been evaluated and the Commission has issued proposals to adapt and extend the regulations (MIFID II, MIFIR, EMIR). These adaptations will again provide an interesting test case for tool support, because we can follow the adoption process.

Particularly interesting is the so called best execution obligation:.

- (33) It is necessary to impose an effective best execution obligation to ensure that investment firms execute client orders on terms that are most favourable to the client. This obligation should apply to the firm which owes contractual or agency obligations to the client (Directive 2004/39/EC).

Now we will illustrate the challenges in the context of MIFID.

Challenge 1. As you can imagine, the question what ‘counts as’ most favorable, is open for interpretation and debate. Typically, different companies will adopt different policies in dealing with the best execution obligation. Branch organizations and expert groups have organized hearings, written guidelines and asked for explanation (e.g. MIFID Connect). The Commission subsequently provided an explanatory letter, detailing issues like scope and providing indicative examples of interpretation (working document ESC-07-2007). Regulators will sometimes respond to adoption behavior by issuing sanctions, further demarcating the boundary between what is acceptable and what not. To harmonize regulators’ responses, also CESR (now ESMA), the association of financial regulators, has issued guidelines.

All of this means, that a compliance officer in an investment firm must not only consider the text of the original EU directive, but also of the national act implementing the directive, of the various guidelines and best practices, of internal compliance policies and of recent jurisdiction. Clearly, there is no single ‘legal text’ from which a rule can be derived.

Challenge 2. Concerning the resource bottleneck, it is clear that MIFID archiving and reporting obligations have a huge impact on IT systems. Detailed client profiles must be archived (see 4 below). How this is done, depends on the existing application landscape. A person deciding on implementing MIFID regulations, must therefore know the legal requirements, the business processes and the applications supporting them. Such people are rare. Moreover, it is likely that a large part of the processes in an investment firm are affected. The scale of the problem is huge. Tool support is mandatory.

Challenge 3. Regarding the mapping to business processes, different companies have made different choices in how they have organized their trading processes. For example, some companies trade by providing price quotes beforehand. They would argue that clients are capable of comparing quotes and selecting the best execution. But there are other influences too, such as marketing, legacy software, and the risk appetite of the firm. All of these aspects determine how a legal prescription must be handled in a specific business process. A representation of a business process as a series of boxes and arrows (as in conformance testing) is insufficient.

Challenge 4. The outcome of what counts as ‘most favorable’ depends on the individual needs of a client. Therefore, MIFID requires investment firms to record detailed client profiles (see e.g. article 51(3) of the MiFID execution guideline). Such profiles must provide evidence, later, that indeed the institution had taken “all reasonable steps to obtain the best possible result” That means, probably, that also alternatives to the transaction that was offered, must be

recorded. Depending on the interpretation of the legal concepts, more or less evidence must be collected as part of the ‘audit trail’ to establish compliance later.

Summarizing, the MIFID regulations and particularly clause 33 show that indeed these challenges are very real, and make it hard for legal knowledge management tools to adequately support a compliance officer, in deciding how to adapt business processes and make them compliant. At various points, decisions must be made among alternatives. The currently dominant translation model is inadequate for supporting such decision making.

We will now briefly highlight how we hope to overcome these challenges, deriving requirements for a possible solution.

### 3 Legal Ontology

We believe that previous work on the representation of legal content rests on false assumptions (e.g. projects like Compas, Regnet, GRC GRID). Many technology driven approaches assume a one-to-one mapping between legislative text and rules [19, 7]. In compliance however, rules are generally the result of legal interpretation involving insights from different sources.

Ontologies may act as an intermediary between users and the source legislative text, helping to locate and explain the law. Ontologies enable sharing and integration of data from different domains and data sources. In particular, differences can be due to homonyms (i.e., nodes with identical concept-names, but possibly different semantics) and synonyms (concepts with different names but same semantics) [11]. For this reason, there is also the need of an alignment among distinct domains like the legal domain and the domain of business processes, which are designed by persons with different competences.

There is some progress in legal document management systems that model the complexity of the law, so that the system can support the work of lawyers and legal scholars [1, 4]. Such systems can also be applied to compliance [6].

### 4 Populating the Ontology

Even if an acceptable formalization of law is devised, there remains the problem of the resource bottleneck. Doing it entirely by hand is unfeasible; they are not enough knowledge engineers with the required inter-disciplinary competences.

Semi-automated ontology extraction and mapping can play a critical role in aligning norms with the processes they govern. The problem of semi-automatically constructing taxonomies or ontologies from texts can be faced at different levels of granularity, starting from the discovery of domain-specific dictionaries to more complex tasks aimed at organizing the terms in some structural form [8]. This work becomes important when there is a lack of resources or dictionaries that can serve for automatic matching between processes and regulations. In general, common practices used to make probabilistic inferences about the terms contained in a corpus leverage on their frequency and their context, defined as the set of terms that often co-occur with them [14].

Further improvements on pure statistical methods rely on the use of natural language processing (NLP) technologies that can filter out data noise while maintaining high levels of accuracy. [20] defined a system for extracting definitional sentences (i.e., sentences that contain at least one hyponymy-hyperonym relationship), using them to build an IS-A graph (subsumption graph). This mined knowledge can be pruned later to form a taxonomy [21]. Such techniques have a high precision but a significantly low recall. This is what NLP approaches usually achieve in terms of accuracy levels, as opposed to statistical methods.

## 5 Mapping onto Business Processes

Compliance management involves not only legal knowledge but also knowledge concerning the business processes, and we cannot assume that the two worlds share the same terminology. While every jurisdiction has its own terminology, the same can be said for organizations. Simple string matching between an ontology concept with the regulation texts does not provide accurate results. We need to enrich an ontology concept with additional information from two different sources: the structural position and context within the ontology or, eventually, a set of concept-associated documents that help depict its semantics [17]. These two types of information are the basis of existing algorithms for *ontology alignment*, namely structurally-informed or extensionally-informed [11]. The first method takes into account the structural neighbourhood of a concept in the ontology to make more fine-grained the matching with the parts of text contained in the regulations. The clear advantage of using this technique is that it can be done in an automatic fashion without any external resource, though it still suffers from possible divergences in the used terminology. On the contrary, relying on a first, approximate, automatic classification of regulations into the ontology, already usually provides an initial overview of the concepts that can be used later for matching two data sources [10].

Therefore these techniques should be build into tools, that may assist human experts to verify and improve the automatically derived ontologies, before they can be mapped onto information required for making actual decisions in business processes. Regulations tend to have a stifling effect, limiting future innovative power. There has been progress in using techniques for balancing compliance objectives with meta qualities like flexibility and agility [15].

## 6 Establishing evidence

How can we generate evidence that the business processes are indeed compliant? Conceptually, the method is based on the realization that what ‘counts as’ evidence of some conduct or behavior being compliant or not, is essentially a social or institutional matter; compare the notion of constitutive norms [23]. Some person in some role is authorized to make a decision. This requires extending the legal ontology with the notion of roles and with the notion of a *constitutive norm* representing the ‘counts as’ relation [5].

Objectives of a process are stated in terms of desired *outcomes*, for example, a reduced financial risk for society. What is measured however is *output*, for instance, risk assessment reports. The art of process design is to find performance measures (outputs), which accurately approximate the actual effects (outcomes). Note moreover that deciding the amount of evidence to be collected, is often a trade-off between collecting too much, which will make the system slow or will make it impossible to find needed evidence later, and collecting too little, which make it impossible to use the evidence later in legal proceedings. Such trade offs usually point into the direction of a risk analysis, or alternatively, to a form of argumentation [9]. Only evidence which is deemed both necessary and sufficient for demonstrating compliance objectives, needs to be collected.

The method will at least address the following aspects. *Stakeholder analysis*: Who are the relevant stakeholders and what are their interests, in relation to the regulations? *Control direction*: Parties may have an interest in adapting the evidence. They may for instance want to overstate (e.g. revenue) or understate (e.g. liabilities). This determines which kinds of evidence need to be recorded. *Representation*: Finally, we need a way of expressing the evidence in an accessible format. The format should accommodate the decision boundaries according to the legal ontology. Not only the evidence itself, but also meta-data needs to be represented, like the date and time of collection, the responsible person, etc.

## 7 Conclusions

The costs of compliance are increasing. In particular in the financial industry, compliance officers are struggling to keep up with new rules and regulations. Information systems could help. However, there are challenges to be overcome.

Using the MIFID directive as example, we have discussed these challenges and derived requirements for legal Knowledge Management Systems (LekMAS).

To deal with the challenges, we argue that the LekMAS approach should consist of four pillars: (1) the use of a legal ontology, to represent and reason with legal concepts, and serve as an intermediary stage between the legal sources and the adopted prescription, (2) a semi-automatic method for populating the ontology, to overcome the resource bottleneck, (3) a method for mapping legal concepts and prescription onto business processes, taking the economic value into account and (4) a method for collecting reliable evidence of the business processes, and for auditing that they are actually compliant.

Their key idea is to view compliance management as a series of decisions, that need to be supported, rather than as a translation exercise.

## References

1. Gianmaria Ajani, Leonardo Lesmo, Guido Boella, Alessandro Mazzei, and Piercarlo Rossi. Terminological and ontological analysis of european directives: multilingualism in law. In *Proceedings of ICAIL 2007*, pages 43–48. ACM, 2007.
2. H. Aldewereld, F. Dignum, A.Garcia Camino, P. Noriega, J.A. Rodriguez-aguilar, and C. Sierra. Operationalisation of norms for usage in electronic institutions. In *Proceedings COIN'06*, volume LNC1 4386, pages 163–176, 2007.

3. M. Alles, G. Brennan, A. Kogan, and M.A. Vasarhelyi. Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at siemens. *Accounting Information Systems*, 7:137–161, 2006.
4. G. Boella, L. Humphreys, M. Martin, Rossi P., and L. I. Springer. van der Torre. Eunomos: a legal document and knowledge management system to build legal services. In Springer, editor, *In Proceedings of AICOL 2011*), Berlin, 2011.
5. G. Boella and L. van der Torre. Regulative and constitutive norms in normative multiagent systems. In *Proceedings of KR 2004*, pages 255–265. AAAI Press, 2004.
6. Guido Boella, Marco Martin, Piercarlo Rossi, Leendert van der Torre, and Andrea Violato. Eunomos, a legal document and knowledge management system for regulatory compliance. In *Proceedings of ITAIS 2011*, Berlin, 2012. Springer.
7. T. D. Breaux.: *Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems*. PhD thesis, North Carolina State University, Raleigh (NC), 2009.
8. P. Buitelaar, P. Cimiano, and B. Magnini, editors. *Ontology Learning from Text: Methods, Evaluation and Applications*, volume 123 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, Amsterdam, 2005.
9. B. Burgemeestre, J. Hulstijn, and Y.-H. Tan. Value-based argumentation for justifying compliance. *Artificial Intelligence and Law*, 19(2-3):149–186, 2011.
10. K. S. Candan, L. Di Caro, and M. L. Sapino. Creating tag hierarchies for effective navigation in social media. In *Proceeding of SSM 2008*, pages 75–82, 2008.
11. K. S. Candan, Mario Cataldi, M.L. Sapino, and C. Schifanella. Structure- and extension-informed taxonomy alignment. In *Proc. of ODBIS*, pages 1–8, 2008.
12. S. English and S. Hammond. Cost of compliance survey 2012. Technical report, Thompson Reuters, 2012.
13. Patrícia Silveira et al. Aiding compliance governance in service-based business processes. In *Handbook of Research on Service-Oriented Systems and Non-Functional Properties*, chapter 22, pages 524–548. IGI Global, 2012.
14. B. Fortuna, M. Grobelnik, and D. Mladenčić. Semi-automatic data-driven ontology construction system. 2006.
15. Y. Gong and M. Janssen. From policy implementation to business process management: Principles for creating flexibility and agility. *Government Information Quarterly*, 29(1):S61 – S71, 2012.
16. G. Governatori and S. Sadiq. *The journey to business process compliance*, pages 426–445. IGI Global, 2009.
17. J.W. Kim and K.S. Candan. CP/CV: concept similarity mining without frequency information from domain describing taxonomies. In *Proceedings of CIKM '06*, pages 483–492. ACM, 2006.
18. J.R. Kuhn and S.G. Sutton. Continuous auditing in erp system environments: The current state and future directions. *J. of Information Systems*, 24(1):91–11, 2010.
19. G. Lau. *A Comparative Analysis Framework for Semi-Structured Documents, with Applications to Government Regulations*. PhD thesis, Stanford, 2004.
20. R. Navigli and P. Velardi. Learning word-class lattices for definition and hypernym extraction. In *Proceedings of ACL'10*, pages 1318–1327, 2010.
21. R. Navigli, P. Velardi, and S. Faralli. A graph-based algorithm for inducing lexical taxonomies from scratch. 2011.
22. A Rozinat and W.M.P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.
23. J. R. Searle. *The Construction of Social Reality*. The Free Press, 1995.