



The contribution of ontologies to the creation of knowledge bases for the management of legal compliance in occupational health and safety

Jonathan Vigneron, Franck Guarnieri, Jean-Marc Rallo

► To cite this version:

Jonathan Vigneron, Franck Guarnieri, Jean-Marc Rallo. The contribution of ontologies to the creation of knowledge bases for the management of legal compliance in occupational health and safety. 22nd European Safety and Reliability Conference - ESREL 2013, Sep 2013, Amsterdam, Netherlands. 6 p. - ISBN 9781138001237. hal-00870504

HAL Id: hal-00870504

<https://hal-mines-paristech.archives-ouvertes.fr/hal-00870504>

Submitted on 25 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The contribution of ontologies to the creation of knowledge bases for the management of legal compliance in occupational health and safety

J. Vigneron & F. Guarnieri

MINES Paristech, Center for Research on Risk and Crisis, Sophia Antipolis, France

JM. Rallo

Preventeo, Le Cannet, France

ABSTRACT: Occupational health and safety (OHS) is at the heart of corporate policies and the increasing trend towards litigation has contributed to the inflated number of regulatory texts published each year. OHS professionals are therefore faced with the need to process a growing amount of data in order to remain in compliance. They train themselves in law, draw upon internal expertise (the company's legal department) or external consultants (advisory and enforcement agencies) and equip themselves with information systems. These systems are mainly 'judicial monitoring' databases that contain regulatory texts and associated comments. For the most part, this data is complex to interpret and difficult to exploit given the growing number of texts, the expertise needed to understand them and the difficulty in being sure that they are relevant to the activities of the company. This need for clarity may be met by models and methods from artificial intelligence. Amongst these, we have selected the concepts of the ontology and the knowledge base. An ontology is a model of a particular domain of knowledge created for a specific purpose. Mainly used for the formalization of knowledge, it can also be used to define a common language between an information system and a community of actors (i.e. OHS professionals). A knowledge base is a set of knowledge that is specific to a given domain that can be exploited by a technology. The coupling between the ontology and the knowledge base makes it possible to automate certain tasks, such as monitoring, auditing, ensuring compliance, etc. This three-part article describes a methodology and an associated software tool that can ensure compliance. The first part identifies elements used in the definition and the proven benefits of ontologies and knowledge bases. The second describes the approach taken in the design of the ontology and its implementation as a rule base using a bespoke knowledge base editor. Finally, a concrete example of the formalization of regulatory knowledge (using the theme of harassment and violence at work) is presented to demonstrate the contributions and the limitations of the tool.

1 INTRODUCTION

In France, Occupational Health and Safety (OHS) is at the heart of corporate policy and an increasingly litigious society has contributed to a very significant increase in the number of regulatory texts published annually by the legislature. OHS professionals responsible for the implementation of OHS policy such as Health, Safety and Environment officers and safety managers are faced with this situation which they struggle to manage on a daily basis [1] [2] [3]. Nevertheless, they train themselves in labour law and draw upon internal (legal department) or external (regulatory authorities) expertise. They sometimes use information systems (databases, sophisticated software solutions, etc.) [4]. However, the number and nature of texts that are published on a regular basis makes it difficult to process all the new information and obligations, due to the degree of technical expertise that is necessary to understand

and implement texts that apply to the activities of the company.

This observation reflects a recognised organisational need for the popularization of legal knowledge in the field of OHS. This need can be met through the use of models and methods from artificial intelligence, from which the concepts of ontologies and knowledge bases were selected.

In this context, the Centre for Research on Risks and Crises (CRC) of MINES ParisTech, in partnership with the company PREVENTEO launched a research project that looked at the contribution of ontologies for data management and legal knowledge in the field of occupational health and safety.

This article consists of three parts. It describes a methodology and associated software to ensure the management of legal compliance. The first part identifies the defining elements and the proven benefits of ontologies and knowledge bases. Then, the paper details the design process for the ontology and

its integration into a bespoke tool for the editing of knowledge bases. Finally, a concrete example of the formalization of regulatory knowledge is presented and its benefits and limitations are outlined.

2 DEFINING ELEMENTS

A rapid review of the current state of information systems for the management of OHS regulatory compliance shows that there are a range of technological solutions, from simple databases that provide legal texts and associated comments through to interactive online systems that offer, for example, regulatory monitoring facilities, a compliance audit process and management of action plans. Although it is clear that there have been technological advances in this area, management of legal texts and their content remains the key element. In France, the legislature, through the LEGIFRANCE database publishes on the Internet all the texts of the Official Journal of the French Republic. This website offers free access to many types of texts: laws that implement standards, acts resulting from French international commitments, jurisprudence and all official publications. To date, in the field of OHS this site includes more than 200,000 texts organised by themes. These texts are accessible using a search engine. This site is extremely important in order to follow new legal developments in OHS. However, access to the database is reserved to authorised users. Therefore, in order to open it up to the largest number of active OHS professionals, it appears that an ontology can provide real added value.

2.1 *Ontology: definitions*

The concept of ontology originated in artificial intelligence. The goal is to provide a shared language between an operator (here the OHS specialist), an expert in the domain (a lawyer) and computers and to enable the automation of certain tasks involving a cognitive representation of domain knowledge. The most cited definition of an ontology is that of Tom Gruber [5], "An ontology is an explicit specification of a conceptualization." It should be noted that many authors criticise the overly vague and generic nature of this characterisation. For this reason, our research uses the definition proposed by Studer et al. [6] which adds the notion of sharing, "An ontology is a formal, explicit specification of a shared conceptualization." This definition highlights four particularly important characteristics of ontologies:

- Formal: A standardised definition of the representation for use in a computer system.
- Explicit: Both concepts and related constraints are defined declaratively.

- Conceptualization: This reminds us that an ontology is only a representation of reality and an abstraction of the real world. Thus, there should be no ambiguity in the definition of terms.

- Sharing: This focuses on the fact that an ontology allows the sharing of knowledge in a given field.

The contributions that ontologies can provide are multiple. They aim to clarify the structure of knowledge by defining the relationships between domain concepts. They reduce terminological conceptual ambiguities through the definition of similar concepts in a given field. Ontologies concern the formalization and sharing of knowledge in order to make it easier to understand the vocabulary of a domain and the differences between terms.

An ontology is usually composed of different entities: "Concepts" are an abstraction of reality through a "class". "Relationships" reflect a link between two concepts. "Axioms" are information that enables the definition of a concept. Finally, "instances" are the "individuals" in the ontology, i.e. the members of different "classes".

2.2 *Method for ontology design*

Although the concept is to a certain extent mature in information systems, there is currently no textbook or manual for the design of ontologies. Thus, it is possible to list a dozen ways of creating and monitoring them. Depending on the degree of formalisation required, the operationalization of the ontology or even its internationalization, each approach has advantages and limitations that are often influenced by the objectives of the ontology to be created. The method used in our research is that described by Stanford University [7], which offers a structured seven-step approach:

1. Determine the domain and scope
2. Consider reusing existing ontologies
3. Define the key terms
4. Create the classes and the class hierarchy
5. Create the properties and attributes
6. Define the facets
7. Create instances

Similarly, there are tools available for the implementation of different methods. We selected the Protégé® tool developed by Stanford University. The choice of this tool was based on the size of the user community and the ability to adapt it depending on the formalization language. Software tools can sometimes implement proprietary languages, but the language that seemed most appropriate for our work was the Ontology Web Language (OWL) which, as its name suggests takes a web-based approach to knowledge formalization. This language is a mixture of Resource Description Format (RDF) and eXtensible Markup Language (XML), which are both widely used in web applications.

3 AN ONTOLOGY COUPLED WITH AN AUDIT SYSTEM FOR REGULATORY COMPLIANCE

In addition to the organisation of knowledge, an ontology of OHS regulations is particularly useful when coupled with a tool for the editing of knowledge bases, which is itself linked to a tool for the management of regulatory compliance audits. The remainder of this article describes in detail the relationships between these three components that are integrated into a global information system for risk management (called PREVENTEO).

3.1 Overview of the OHS ontology

An ontology was created in the domain of French occupational health and safety regulations. It specifically addresses the obligations of “agents”, i.e. employers, business owners and more widely those holding delegated powers. Further research identified a higher-level ontology, i.e. a classification of high-level concepts, useful in defining the knowledge base. The Legal Knowledge Interchange Format (LKIF) ontology was selected. LKIF consists of 155 classes, 97 class properties, 266 annotations and it focuses on basic legal concepts. It offers no individuals or instances (a so-called “top-level” ontology). Efforts were focused on building a corpus of French regulations. Candidate terms were extracted using the Yatea method [8] together with numerous interviews with domain experts. The construction of the OHS ontology was carried out to a depth ranging from three to ten levels, which led to the definition of 200 terms organised into two broad categories: “Agent” and “Action”. Figures 1 and 2 illustrate two extracts from the ontology centred on these concepts. These two branches of the ontology are used mainly for the definition of regulatory requirements.

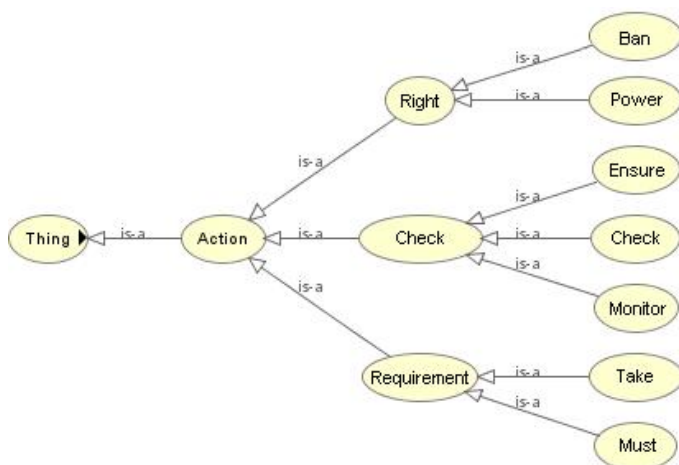


Figure 1: Overview of an extract of the ontology – Action

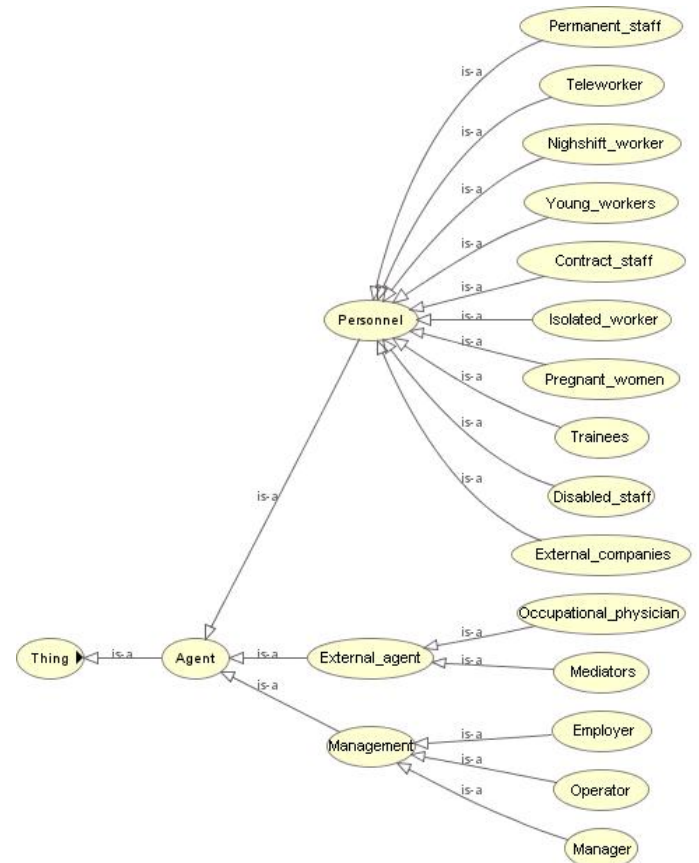


Figure 2: Overview of an extract of the ontology – Agent

3.2 The Knowledge Base editor: Cogniteo

To exploit the OHS ontology, a tool for editing the knowledge base was designed for the formalization and use of regulatory audit questionnaires: the Cogniteo tool.

Cogniteo is an online tool. It guides the user through several steps for the creation of online questionnaires:

- The first step is to organise a corpus of regulatory texts in a database. Each text is defined by a set of descriptors (title, date, subject, etc.).
- Each text is then divided into a series of articles; each article corresponds to a key point of the regulatory text (step 2). Articles are in turn stored and described in a database.
- Articles are “reduced” (step 3) to an elementary “atom” of knowledge of the requirement (the requirement is represented as a verb complemented by a comment).
- Requirements are used to construct audit questionnaires.

Cogniteo is particularly intuitive to use. It provides a set of functions to organise questionnaires into themes and sub-themes and make it easy to review the order of questions. Updates are possible. Each question is accompanied by “its” response and additional help in understanding it.

3.3 A prototype tool to automatically identify regulatory requirements

The coupling of the ontology and the knowledge base editor (Cogniteo) is achieved using a prototype tool for the automatic identification of regulatory requirements found in texts. The prototype is designed to improve both the productivity and quality of the process of identifying regulatory requirements, which is particularly time-consuming.

Initially, the tool “cuts” the text into sentences; it then runs various queries on the OHS ontology in order to recover useful terms considered to be generators of requirements (Ontology: Agent) and checks for the presence of terms that are related to obligations (Ontology: Action). For each sentence of the text, if an “Agent/Action” pair is found, the sentence is detected and stored in the database. Once the automatic reading of the text is complete, the list of proposed requirements is notified to the user. Candidate terms are classified into three categories: Management, Staff or Action. The same is true for the requirements that have been identified.

When this operation is complete, the list created by Cogniteo is examined by a legal expert who confirms, clarifies and completes the proposed suggestions.

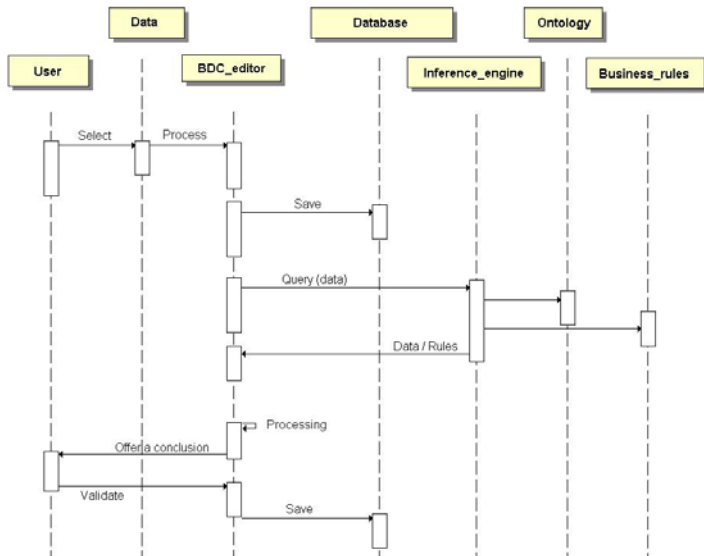


Figure 3: The process of creating a semi-automated questionnaire

4 EXPERIMENTS, RESULTS, DISCUSSION

To validate the coupling between the ontology and the knowledge base editor, an experiment was conducted. It was based on a simple principle; a legal text (“Violence and Harassment in the Workplace”) [9] was “simultaneously” scanned by a legal expert and the prototype tool for the automatic identification of regulatory requirements.

4.1 Regulatory context

Before describing the initial results, we will outline the experimental domain. The French regulatory framework contains a number of definitions of occupational violence or harassment. Tensions, bullying, insults and assaults are clearly defined and can be directed at either an individual or a group. According to one survey [10], one in six workers claim to be under strain. These sources of tension have important implications for both the employee and the company. The effects on health, such as stress and fatigue, rapidly become apparent. This leads to reduced levels of performance and disinvestment, which has negative impacts that include a poor working atmosphere, difficulties in recruitment and absenteeism. It is therefore necessary, for both social and economic reasons, to identify any behaviour that could be termed occupational harassment or violence. As such, the Agreement of 26 March 2010 on occupational harassment and violence establishes the regulatory foundation. This first test uses the automatic detection method to identify regulatory requirements in the domain.

4.2 Initial results

The “Violence and Harassment in the Workplace” text corresponds to articles 3–5 of Agreement No. 2010-01. It consists of nine pages.

After reading, the legal expert identified a total of twenty-six requirements for the text, while the software prototype counted twenty requirements. The legal expert completed the task in 50 minutes. It took five seconds for the prototype to deliver the result. It should however be noted that that expert provided an accurately formulated requirement, while the prototype provided an extract “candidate” phrase that obviously must be looked at again and clarified by the lawyer in order to eventually become a requirement that can be inserted into the database.

Part of the ontology produced feeds the proposed definitions of French law on occupational harassment and violence. This section of French law falls into the category of psychosocial risks (PSR)

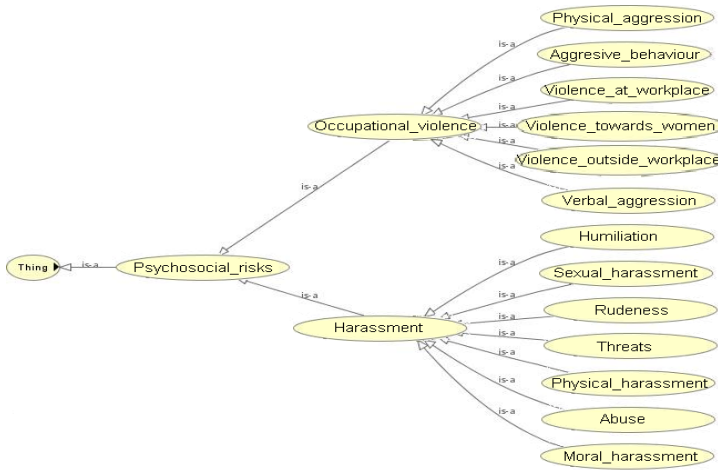


Figure 4: An extract from the ontology – PSR

Among the six requirements that the system did not find, five concerned mandatory obligations. Specifically, the tool was not able to identify requirements in the following extract, “To this end, it is important to identify, as appropriate, the phenomena of occupational harassment or violence in order to measure its extent, to understand the circumstances and to seek appropriate preventive measures”. This was because the concepts of “Agent” and “Action” (necessary for the definition of a regulatory requirement) do not appear.

The final undetected requirement refers a concept that was overlooked in the ontology, namely “internal regulation”.

Once the automatic processing was finished, the expert completed and reformulated the identified information in order to establish a list of regulatory requirements. These requirements contribute to the development of the audit questionnaire. They are classified according to prevention lever, and linked to the relevant company department and other related mandatory documents.

The audit questionnaire is based on two types of questions: simple (Yes / No / Proof required) and a more complex multiple-choice format that can check a list of obligations related to the same question.

<p>The employer, working with staff or their representatives, investigates instances of occupational harassment or violence as soon as such situations are reported.</p>	<p><input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> PR</p>
<p>The investigation focuses on the following elements of the working environment :</p>	<p><input type="checkbox"/> Individual behavior <input type="checkbox"/> Management style <input type="checkbox"/> Relations with clients <input type="checkbox"/> How the business operates</p>

Figure 5: An extract from the assessment questionnaire

4.3 Areas for improvement

The initial results of the experiment were encouraging, but they are still far from providing real assistance to lawyers responsible for the design of audit questionnaires.

There seem to be two main areas for improvement.

The first aims to improve and expand the domain covered by the ontology. The ontology developed so far is limited to a subdomain of OHS, in this case, the management of regulatory compliance. This subdomain defines the set of mandatory actions to be implemented by “agents”. The incompleteness of the ontology naturally conceals “candidate” requirements that could be linked to other domains. It would be useful to offer a broader scope of knowledge, and process different domains in the ontology such as chemical risk and the strenuousness of work for example.

The second area of improvement relates to the performance of the prototype tool for the automatic recognition of requirements. Beyond the quality of the ontology, some requirements could be made more accurate through better analysis. The prototype is restricted to offering a candidate term; consequently some requirements may be “forgotten” given the formalism of an extract: tables and figures cannot currently be processed by the prototype. Although figures or images require complex processing that is difficult to implement in the context of this work, it is possible to implement some features to manage tables. Texts are inserted into the database in HTML format, and work is in progress to modify the proposed cutting method in order to take into account this format.

5 CONCLUSION

Our work is still at the prototype stage, but initial results are encouraging. The automatic identification software still needs a lot of improvements. This work is closely related to the enrichment of the OHS ontology. There are also plans to expand the use of ontology concept to other areas of risk management. Thus, work conducted in the field of industrial safety [11] could be put to good use. As for the Cogniteo knowledge base editor, it is used on a daily basis by lawyers and HSE professionals working at PREVENTEO to update and design audit questionnaires. In the domain of OHS, it is now possible to manage more 4,000 regulatory requirements and almost two hundred more are added every month.

REFERENCES

[1] Audiffren T, Rallo JM, Guarnieri F, Martin C. A quantitative analysis of health, safety and environment policy in France, ESREL Conference 2013, Amsterdam, Netherlands, September 2013.

[2] Guarnieri F, Miotti H, Martin C, Besnard D, Rallo JM. Occupational safety and health in France: Practitioners and policy, Edited by the AFNOR Group, November 2010.

[3] Audiffren T, Rallo, JM, Guarnieri F. The contribution of case law to compliance management in Occupational Health and Safety (OHS) in France, ESREL Conference 2012, Helsinki, Finland, June 2012.

[4] Cambon J, Guarnieri F, Groeneweg J. Towards a new tool for measuring Safety Management Systems performance, Proceedings of the 2nd Symposium on Resilience Engineering, 2006, Edited by Presses des MINES (France).

[5] Gruber T. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.

[6] Studer R, Benjamins VR, Fensel D. Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering*, 25 (1–2) (March 1998): 161–197.

[7] Noy NF, McGuinness DL. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

[8] Aubin S, Hamon T. Improving term extraction with terminological resources, *Advances in Natural Language Processing -Lecture Notes in Computer Science Volume 4139*, 2006, pp 380-387

[9] Accord n° 2010-01 du 26 mars 2010 sur le harcèlement et la violence au travail [Law 2010-01 of 26 March 2010 on harassment and violence at work]

[10] Sumer 2003 – Medical risk monitoring (regular survey initiated by the Ministry of Labour and performed using data collected by occupational physicians)

[11] Bourreau L, Audiffren T, Rallo JM, Guarnieri F. Industrial maintenance in the gas sector: the contribution of knowledge bases to compliance evaluation, ESREL Conference 2012, Helsinki, Finland, June 2012.