# A Framework to Reason about the Legal Compliance of Security Standards*

Cesare Bartolini[1], Andra Giurgiu[1], Gabriele Lenzini[1], and Livio Robaldo[1]

University of Luxembourg,
Interdisciplinary Centre for Security, Reliability and Trust (SnT)
{*firstname.lastname*}@uni.lu

**Abstract.** Achieving compliance with legal regulations is no easy task. Normally, laws state general requirements but do not provide clear parameters to determine when such requirements are met. On a different level, industrial standards and best practices define specific objectives that can be certified by means of auditing procedures from qualified bodies. Implementing a standard does not *per se* guarantee legal compliance, with the rare exception when the standard is also endorsed by the law itself. But standards and laws in the same domain may have overlaps and correlations, so adopting the former may provide an argument to demonstrate that adequate measures were taken to achieve legal compliance. In this paper, we introduce a framework that, using state-of-the-art Natural Language Semantics techniques, helps process legal documents and standards to build a knowledge base to store their logic representations, and the correlations between them. The knowledge base will help legal experts assess what requirements of the law are met by the standard and, consequently, recognize what requirements still need to be implemented to fill the remaining gaps. An application of the framework is exemplified by comparing a provision of the European General Data Protection Regulation against the ISO/IEC 27001:2013 standard.

**Keywords:** Legal compliance; legal requirements; security standards; General Data Protection Regulation.

## 1 Introduction

Security standards have contributed to shaping the quality of services and have promoted a security-oriented culture by establishing best practices. Like standards in general, they can also create in favour of the implementing party a "*presumption of conformity with the specific legal provision they address*" [12].

In themselves, standards do not guarantee legal compliance. In order to have a direct effect on legal compliance, standards would need to be endorsed by governments as key elements in their framework for policies and regulations. This is

* Livio Robaldo has received funding from the EU's H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 661007. Cesare Bartolini has received funding from the EU's H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690974.

a solution that ISO and IEC are advocating in specific sectors, such as that Medical Device[1], but it would have severe limitations if applied at an international level. In the European Union, for instance, where there is the need to promote a single market while letting each country establish its own legal framework, supporting a standard in a legislative source (Regulation or Directive alike) would immediately exclude from conformity the countries where that standard is not commonly adopted, weakening their position in the single market [10]. This is why the European Union prefers to emit a legislation that defines an abstract level of safety/security of products and systems, and makes the adoption of standards to demonstrate compliance a voluntary choice[2].

That stated, standards nevertheless remain a viable way to support arguments for conformity with regulations. When widely adopted and subject to repeated audits by conformity-assessing bodies, a standard gives an organization that implements it an argument of compliance. Of course, such an argument is a presumption, giving a plaintiff the possibility to demonstrate that the organization failed to comply with the legal framework. Still, this offers an inversion of the burden of proof that the organization would not benefit from, if it simply adopted a personalized solution [12].

However, that presumption of compliance needs to be reassessed when new laws reshape the legal landscape. In this case, a company needs to know whether or not the standards it has adopted will preserve the company's presumption of conformity. Failing that, it may want to know what changes must be implemented to keep preserving that presumption. Waiting for the adoption of new standards would take a significant amount of time and may lead to interim problems. In the meanwhile, businesses face the risk of liability for not being compliant and may, on that ground, be sanctioned. A passive business strategy like this may have the benefit of being effortless but can backfire with high costs if an incident occurs.

A safer strategy would be to identify what provisions in a standard interpret or implement the provisions in the law. This allows to determine what gaps need to be filled in order to benefit from the presumption of compliance. Such *correlations* can express either a *formal compliance* based on the semantic of terms of the language in the standard and the law, or a *substantive compliance* based on decisions of courts and other competent authorities. This strategy has drawbacks as well: security standards, laws, and court decisions are written in natural language. Reading and analysing them by hand is inefficient and largely impractical. But this activity can be supported by computer processing.

We advocate such a pathway and propose a software framework that aids in determining the formal and substantive correlations between provisions in a standard and in a law. The framework's core is a logic-based methodology to represent, in a machine-processable format, $(a)$ the relevant syntactical concepts in the provisions, and $(b)$ the relevant correlations between them. In this paper,

[1] See for example `http://www.iso.org/sites/policy/sectorial_examples.html`.

[2] Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European Standardization, Article 2.1

we describe in detail this logic-based methodology, and exemplify how it works using an excerpt from the ISO/IEC 27001 security standard and a provision in the new European General Data Protection Regulation (GDPR)[3].

The framework depends on two auxiliary functional blocks (see Section 2):

1. a *logic knowledge base* that stores the history of the machine-processable logic correlations, based on collaborative work; experts can not only use the knowledge base but also correct and extend it;
2. a *set of Natural Language Semantics (NLS) and Natural Language Processing (NLP) techniques* that allow a user to browse a XML representation of the documents, search and retrieve the words, terms, and phrases that he and others have found relevant for correlation.

The NLS and NLP techniques will help users show the established correlations within the knowledge base. Any expert user can contribute to reinforce, correct, justify, and expand the correlations. Due to differences in legal interpretation, it may happen that some of the correlations will be in contradiction. Initially, these will remain as contradictions within the knowledge base; over time however they will be overridden by interpretations from more authoritative sources (such as those given by courts of a higher instance).

Technically, the different correlations are expressed in a deontic and defeasible logic for legal semantics called Reified Input/Output Logic (see Section 3), while the *modus operandi* of the framework is similar to that of the collaborative tools for document translation “SDL Trados Studio”. It must be stressed that selecting the relevant terms and establishing the correlations is an activity that still requires human reading, processing and decision-making. The analysis of documents is therefore semi-automatic. It will be supported by the extensible knowledge base; the process of browsing, aided by the tools and techniques from the NLS and NLP, adds efficiency and precision.

The fast pace of technological innovation calls for new regulation. The European Union (EU)’s GDPR is the perfect example of a new legislation which is trying to address the challenges posed by new technologies. It will be applied from 25 May 2018[4], and it poses significant challenges for undertakings in terms of ensuring compliance with it, as it brings significant changes to the regulatory framework for personal data protection in Europe. Security standards, on the other hand, can be regarded as a building block [12] that helps data-processing organizations comply with the principle of accountability and with their obligations resulting from data protection laws. Considering the novelties of the GDPR, as well as the partial overlap between security and data protection, this article will illustrate a methodology based on the correlations between provisions of the GDPR and security standards, specifically ISO/IEC 27001.

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[4] GDPR, Article 99.2.

## 2 The framework at a glance

The framework we propose is schematically summarized in Figure 1. Users (who may be lawyers, regulators, auditors, or other legal experts), access a digital and annotated XML representation of the normative texts (both laws and standards). While browsing a document and selecting the relevant concepts—some which may have been previously annotated—NLP and NLS tools help traverse the rest of the documents, find related terms, and recall the correlations that already exist between them. Not all these correlations have the same degree of importance, as they come from different sources and may have different relevance in specific contexts. This subsidiary information will be stored in appropriate metadata. Potential contradictions and ambiguities among the different interpretations can be solved, as detailed in Section 4. Besides, the user can find new terms and define new correlations, or correct existing ones. The framework thus implements a collaborative, self-correcting, strategy to evaluate the stored correlations. The user's decisions are stored in the knowledge base for later use, after having been appropriately represented in a logic for legal semantics. The details of this transformation are also explained in Section 4.

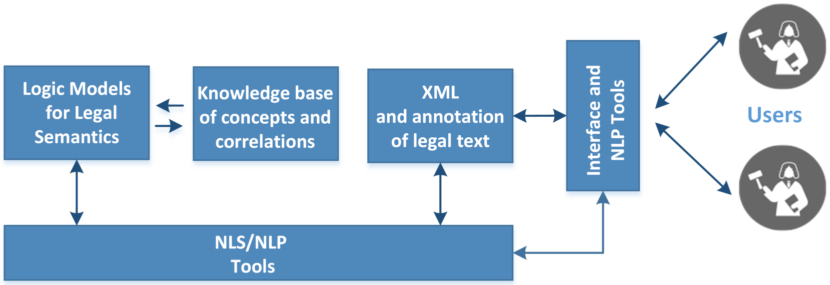
Logic Models for Legal Semantics
Knowledge base of concepts and correlations
XML and annotation of legal text
Interface and NLP Tools
Users
NLS/NLP Tools

**Fig. 1.** The framework at a glance.

Our framework offers a computer-aided methodology to anyone who intends to analyze standards to build an argument of compliance with respect to a specific piece of legal text. The framework, and in particular its knowledge base, does not pretend to be complete. Rather, it provides the expert user with the updated knowledge that helps him or her take autonomous and informed decisions, either when confirming the correlations the tool suggests or when choosing to define new correlations.

The knowledge base is designed to tolerate apparent inconsistencies of different interpretations of terms which may lead to different correlations and create conflicts. Conflicts are especially frequent in legal interpretation, but can be solved, at least in principle, considering that the interpretations of higher instances, such as the Court of Justice of the European Union (CJEU), prevail over others. In order to cope with interpretations of different legal weights, which may supersede one another, the logic formalism that the framework embeds is defeasible: correlations can be updated, modified, rewritten and weighted. If, despite the defeasible mechanism, conflicts do remain, the framework still embeds known strategies that help the user take a decision (see Section 4).

In the following, this paper will exemplify the core part of the methodology, which relates to building correlations.

## 3 Background

The background is structured along two main topics. 1) Natural Language Processing (NLP) [22,8] and Natural Language Semantics (NLS) [15,20]) in legal analysis, techniques which we invoke in the computer-assisted steps of our methodology. 2) Reified Input/Output Logic, the defeasible logic that we propose as the formal language to express correlations.

*Natural Languages Processing in Law.* The application of NLP and NLS to the legal domain is a research trend that has received a lot of attention and investments in recent years, as demonstrated by the several EU funded projects on the topic such as `Openlaws`[5], `Legivoc`[6], `EUCases`[7], `ProLeMAS`[8], `BO-ECLI`[9], and `MIREL`[10]. Modern NLP technologies [5] are able to classify and discover interlinks between legal documents thanks to parsers [2], statistical algorithms [7], and legal terminological databases or legal ontologies [23,7]. This is often done by transforming the legal documents into XML standards, such as *Akoma Ntoso*[11], where relevant information are tagged. An example of commercial legal document management system employing these technologies is *Eurocases*[12] which collects EU case law and uses NLP techniques to classify the documents on the basis of their topic [6].

Although these systems help navigate legislative documents and retrieve information, their overall usefulness is limited because they process words disregarding their possible different semantic interpretations. The latter would allow for legal reasoning, *e.g.*, correlating laws among them and determining whether they lead to inconsistencies. Semantic processing of documents like laws/regulations and security standards is what we are going to propose as a component of the methodology we present in Section 4.

*Reified Input/Output Logic.* This logic has been designed as an attempt to go further in state-of-the-art research in legal informatics, by investigating the *logical architecture* of the provisions, which are available in natural language only.

[5] `https://info.openlaws.com/openlaws-eu/`.

[6] `http://www.legivoc.eu`.

[7] `http://eucases.eu/start.html`.

[8] `http://www.liviorobaldo.com/Prolemas.htm`.

[9] `http://www.bo-ecli.eu/`.

[10] `http://www.mirelproject.eu/`.

[11] `http://www.akomantoso.org/`.

[12] `http://eurocases.eu/`.

Reified Input/Output logic is a logical framework [21] for representing the meaning of norms. Contrary to the great majority of its competitors [14,16,13], Reified Input/Output logic integrates modern insights from the NLS literature. Specifically, it merges Input/Output logic [17], a well-known formalism in Deontic Logic (*i.e.*, a logic that expresses concepts like permissions, obligations, prohibitions), with the first-order logic for NLS proposed by prof. J.R. Hobbs, which is grounded on the concept of *reification*. Reification is a concept originally introduced by the philosopher D. Davidson in [11]. It allows to move from standard notation in first-order logic such as "$(give\, a\, b\, c)$", asserting that "$a$" gives "$b$" to "$c$", to another notation in first-order logic "$(give\, e\, a\, b\, c)$", where "$e$" is the *reification* of the giving action. "$e$" is a first-order logic term denoting the giving event by "$a$" of "$b$" to "$c$". Thanks to reification, Hobbs's logic is able to express a wide range of phenomena in NLS such as named entities, quantification, anaphora, causality, modality, time, and more[13]. In particular, the simplified version of Reified Input/Output logic [21] that we use in our example is an extension of first-order logic that distinguishes three kinds of implication: "$\rightarrow$", "$\leadsto$", and "$\Rightarrow$".

The implication "$\rightarrow$" is the standard trust-value implication of first-order logic. The second, "$\leadsto$", is its *defeasible* version. Defeasible [19] here is to be understood in the sense that an implication "$\Phi \leadsto \Psi$" holds by default unless overridden by "stronger" implications. When instantiated properly, this notion of "stronger" implications resolves the potential contradictions emerging because of the non-monotonic nature of the defeasible reasoning. Reified Input/Output logic also includes other mechanisms to deal with unresolvable conflicts, such as the well-known conundrum about whether or not Nixon is a pacifist, considering that he is both a Quaker (therefore pro-peace) and a Republican (pro-war)[14]. A possible solution to deal with this type of conflicts is to leave the conflict open until more evidence will help the reasoner take a decision. This is, actually, what better fits a situation with conflicts due to multiple legal interpretations. The third implication, "$\Rightarrow$", is a deontic implication. Taken a formula $\Phi$, referring to a set of pre-conditions, and another formula $\Psi$, referring to a set of actions, the meaning of "$\Phi \Rightarrow \Psi$" is "given the pre-condition $\Phi$, actions $\Psi$ are obligatory", that is, the actions must be undertaken if the pre-conditions hold. Note that, according to the interpretation of "$\Rightarrow$", the formula "$\Phi \Rightarrow \Psi \wedge \Phi \wedge \neg\Psi$" is not inconsistent: it only means obligation $\Psi$ has been violated.

The framework [21] further distinguishes between the formulæ belonging to the assertive contextual statements (ABox), *i.e.*, the formulæ denoted by the norms, from those belonging to the terminological declarative statements (TBox), *i.e.*, the definitions, axioms, and constraints on the predicates used in the ABox formulæ. Formulæ in the ABox are in the form "$\forall_{x_1} \ldots \forall x_n\, [\Phi(x_1 \ldots x_n) \Rightarrow \Psi(x_1 \ldots x_n)\,]$", where arguments "$\Phi(x_1 \ldots x_n)$" and "$\Psi(x_1 \ldots x_n)$" are conjunc-

[13] More information are available on line at the following URLs:
`www.isi.edu/~hobbs/csk.html`
`www.isi.edu/~hobbs/csknowledge-references/csknowledge-references.html`.

[14] See `plato.stanford.edu/entries/logic-nonmonotonic/#sec-2-2`.

tions in standard first-order logic. Formulæ in the TBox can be any formula in standard first-order logic augmented with the operator "$\leadsto$".

## 4 Methodology

Our methodology, after minor adaptations, can be applied to find matches between any laws and any standards. However, in order to explain it in a manner as precise and specific as possible, we will refer only to a security standard in ISO/IEC 27000 family, more specifically ISO/IEC 27001, and to the GDPR. The application example in Section 5 will also refer to these two documents.

### 4.1 Overview

The methodology we propose is highly interdisciplinary, in that it involves a strict interaction between law and computer science. When implemented, it will partition the provisions of the the law into three sets:

**set of covered provisions:** these are the provisions that have a full match with the provisions in the standard. A match, or correlation, is expressed in terms of a logic implication, one of the three implications that the Reified Input/Output logic admits (for an example seeSection 5);

**set of partially-covered provisions:** the provisions in this set have a partial match with provisions in the standard, that is, not all the requirements of the provision have corresponding elements in the standard. To guarantee compliance with these provisions, the (certified) compliance with the standard is helpful but not sufficient, and additional requirements must be addressed;

**set of uncovered provisions:** these are the provisions for which there does not appear to be any correspondence in the standard. Concerning compliance with these provision, a certified compliance with the standard does not provide any useful information.

The types of coverage stated above can be further divided into two different categories: *formal* and *substantive* correlations.

Formal correlations entail a mere textual overlap between concepts. For example, formal correlations would allow us to observe that both the GDPR and the ISO/IEC 27001 standard use the term "transfer" (*e.g.*, GDPR, Article 46 and ISO/IEC 27001, Article. 13.2.1, shown in Section 5).

On the other hand, substantive correlations are more complex and entail the analysis of the actual meaning of terms. To assert a correlation of this kind, requirements must be met in a concrete way. In other words, and following the previous example, to assert a correlation between a provision of the GDPR and one of the standard concerning transfer, it is necessary to verify the actual difference/similarity in meaning.

It should be pointed out that the correlations that define a match are *defeasible*, *i.e.*, they can be superseded by new correlations. A superseding correlation might be the consequence of a decision of a court or Data Protection Authority (DPA), or simply the evolution of the interpretation in doctrinal analysis. This way, the output of the methodology can be easily kept up to date by introducing the new correlations as they are developed by the legal actors.

### 4.2 Building Correlations

To build the correlations and define the types of coverage, the methodology follows three steps, which involve both a legal and a technical approach. The legal approach is focused on the interpretation of the provisions of laws (the GDPR in our example) and security standards, whereas the technical approach consists of modelling those provisions into an ontology, and expressing the interpretation by means of logical formulas. The sequence of steps is meant to be iteratively repeated. At each iteration, existing correlations between provisions in the law and provisions in the standard are presented to the user, but new correlations may be established and existing correlations may be updated. The knowledge base of correlations is thus used and updated at the same time.

The next paragraphs detail the steps.

*Step 1: analysis of the provisions.* The provisions of the law and of the security standard are analyzed by legal experts. They provide a legal interpretation of the terms used in the provisions and compare them in search of semantic correlation. There is no need for this interpretation to be final—more interpretations can be added later, and old interpretations can be overridden by newer ones—but this start requires a significant manual activity. The analysis also builds a structured model of the provisions, using an annotation tool called LIME[15]. LIME does not modify the text of the law, but splits it into its components, *i.e.*, articles, paragraphs, and so on. This annotation converts the law and the security standard into the XML legal format `Akoma Ntoso`. The latter is a formalism that provides specific XML tags for linking the textual spans with separate semantic annotations. So, for example, relevant concepts can be embraced by `<concept eId=`$x$`>` ... `</concept>`, where "$x$" indexes an ontological concept (see next step).

To support the execution of this step (and especially its successive iterations), we link LIME to external NLP procedures so that it suggests (semi-automatically and while the legal expert browses the documents) previous translations and correlations on the basis of the ones currently stored in the knowledge base. Ultimately, it is the legal expert that must decide whether and how the correlations shown need to be overridden. In that case, together with the updating of the knowledge base, we can annotate the new correlations with the source which contributed to define it (*e.g.*, Court of Justice of the European Union, *Dapreco and Copreda Corp.*, C-XYZ/16). Correlations by more authoritative sources must override defeasible correlations by less authoritative ones, reshaping the partitions in the three coverage sets as described in Subsection 4.1.

[15] `http://lime.cirsfid.unibo.it/`.

*Step 2: creation of legal ontologies.* The legal interpretations are mapped onto legal ontologies of the law and the security standard. Legal ontologies [4] model the legal concepts, the parties and stakeholders affected by the law, the duties and rights of each stakeholder, and the sanctions and penalties for violating the duties. As per ontologies in general, legal ontologies must be expressed in a knowledge representation language. For this work, we chose the popular abstract language OWL [1]. It can be serialized using various XML notations, making it well fit for machine processing. For example, the OWL representation of the data protection ontology will contain concepts such as "controller", "data subject", "personal data", "processing" and so on.

Concerning the GDPR used in our example, a preliminary version of a legal ontology has been defined already [3]. Albeit partial and based on an older text of the GDPR, it was designed to express the duties of the controller. As such, it can be used to find the correspondences between the requirements expressed in the GDPR and in security standards, until an improved ontology is built.

*Step 3: generation of logic formulæ.* The third and final step of the methodology consists of generating the logical formulæ representing the set of provisions in the law and the set of provisions in the security standard, as well as the correlations between them. These formulæ are expressed in Reified Input/Output logic [21], the first-order language illustrated above in Section 3.

Associating textual provisions to logical formulæ amounts to converting ambiguous and vague terms into non-ambiguous items (predicates and terms). Also, as extensively explained above, these predicates, as well as the correlations connecting them, are subject to different legal interpretations. To this end, words in the provisions are represented via predicates reflecting their ambiguity/vagueness. For example, the word "appropriate", included in the sample GDPR provision used below in Section 5, will be represented via the homonym predicate "$\mathsf{appropriate}$". These "vague" predicates may be defined by adding correlations and further constraints (axioms). Those correlations will be *defeasible*, so that they can account for different legal interpretations.

## 5 Generation of Logic Formulæ: example

We exemplify step 3 of our methodology by using a provision from the GDPR and an article of the ISO/IEC 27001 security standard. Namely:

(*a*) GDPR, Article 46.1: *[. . . ] a controller or processor may transfer personal data to a third country or an international organization only if the controller or processor has provided appropriate safeguards [. . . ]*

(*b*) ISO 27001, Article. 13.2.1: *Formal transfer policies, procedures and controls shall be in place to protect the transfer of information through the use of all types of communication facilities.*

We underline that this example is based only on merely formal correlations between the two texts and relies on the formal conceptual identity of the term "transfer". In this paper we did not take into account the substantive interpretations, according to which the same term would be used with different meanings in the two contexts. The example has the sole purpose of exposing how formulæ are build; it does not mean to assess the substantive correlation between the two texts, which might lead to a different conclusion in terms of compliance.

The formalization of the two provisions in the simplified version of Reified Input/Output logic we use in this paper is rather straightforward. As explained above in Section 4.2, the formulæ will include predicates reflecting the vagueness of the terms occurring in the sentences. Thus, for instance, the main verb "transfer" is formalized into an homonym predicate "$\mathsf{transfer}$" and the adjective "appropriate" is formalized into an homonym predicate "$\mathsf{appropriate}$".

Note that the sentences may refer to *tacit knowledge*, *i.e.*, they may contain ellipses. For instance, the GDPR uses the term "data subject" to generically refer to individuals to which the personal data pertains. However, in many GDPR provisions, such as the one shown above, it is not specified that the "personal data" are the "personal data of a data subject". Tacit knowledge ought to be restored in the formula; thus we will represent "personal data" as "$(\mathsf{personalData}\, y) \wedge (\mathsf{of}\, y\, z) \wedge (\mathsf{dataSubject}\, z)$", where "$y$" is a (generic) unit of personal data and "$z$" is the (generic) data subject that owns "$y$".

Elliptical/tacit knowledge is only one of the issues we must deal with when translating natural language into logical formulæ. In this paper, we cannot illustrate all major problems in NLS and how it is possible to address them in a reification-based framework such as the one of prof. Hobbs.

The GDPR provision in $(a)$ is formalized as follows:

$$\begin{aligned}
&\forall_e \forall_x \forall_y \forall_z \forall_w \forall_k [\\
&\quad ((\mathsf{dataController}\, x) \wedge (\mathsf{personalData}\, y) \wedge (\mathsf{of}\, y\, z) \wedge (\mathsf{dataSubject}\, z) \wedge\\
&\quad (\mathsf{thirdCountryOrIntOrg}\, w\, z) \wedge (\mathsf{transfer}\, e\, x\, y\, w) \wedge (\mathsf{instrOf}\, e\, k))\\
&\quad\ \Rightarrow (\mathsf{appropriate}\, k)]
\end{aligned} \tag{1}$$

In (1), "$e$" is a first-order variable referring to the event of transfer of "$y$" (patient) performed by "$x$" (agent) to "$w$" (receiver). "$x$" has to be a data controller[16] while "$y$" refers to personal data of a data subject "$z$". "$k$" is the instrument of the action "$e$", *i.e.*, the communication facility through which data are transferred. The predicate "$\mathsf{thirdCountryOrIntOrg}$" is a convenient predicate to state that its first argument is either an international organization or a country different from the one where the data subject lives. This is enforced by adding the following definition of "$\mathsf{thirdCountryOrIntOrg}$" to the TBox:

$$\begin{aligned}
&\forall_w \forall_y \forall_c [(\mathsf{thirdCountryOrIntOrg}\, w\, z) \leftrightarrow\\
&\quad ((\mathsf{IntOrg}\, w) \vee ((\mathsf{country}\, w) \wedge (\mathsf{diffFrom}\, w\, c) \wedge (\mathsf{countryOf}\, c\, z)))]
\end{aligned} \tag{2}$$

[16] Or a data processor, but in this example we assume for simplicity that the provision only refers to data controllers.

The crucial predicate in (1) is the predicate "$\mathsf{appropriate}$", which may be true or false with respect to a communication facility "$k$".

The TBox of the ontology will include defeasible axioms that define when a communication facility is "appropriate". For instance, an undertaking certified to ISO 27001 has adopted a formal transfer policy according to which data transfers must take place via email with electronic signature or registered (hard) mail. Assuming that such data transfers are to be considered appropriate, the following two (defeasible) axioms are added to the knowledge base.

$$\begin{aligned}&\forall_x[\,(\mathsf{emailWithES}\,x)\ \rightsquigarrow\ (\mathsf{appropriate}\,x)\,],\\&\forall_x[\,(\mathsf{regMail}\,x)\ \rightsquigarrow\ (\mathsf{appropriate}\,x)\,]\end{aligned}\tag{3}$$

In case a judicial authority decides, for example, that emails are no longer appropriate means, an additional higher-priority axiom will be added to the TBox in order to override the first axiom in (3). However, axioms may be associated to time stamps (although this is not shown in (3), so that the new axiom will override the old one only for transfer actions performed after a certain date, *i.e.*, the one from which the new law enters into force. On the other hand, the old axiom will still assert that email with electronic signature is an "appropriate" communication facility for all transfer actions performed before that date.

Formula (4) models the ISO/IEC 27001 provision in ($b$). This provision refers to any kind of information, not only personal data. It does not specify the agent or the receiver of the transfer actions, so those are assumed to be any entity. The formal transfer policies, procedures and controls to protect the transfer of information are simply formalized in terms of a predicate *secure*, which is true for any communication facility that is deemed to be "secure".

$$\begin{aligned}&\forall_e\forall_x\forall_y\forall_w\forall_k[\\&\quad(\,(\mathsf{transfer}\,e\,x\,y\,w)\ \wedge\ (\mathsf{information}\,y)\ \wedge\ (\mathsf{instrOf}\,e\,k)\,)\\&\quad\Rightarrow\ (\mathsf{secure}\,k)\,]\end{aligned}\tag{4}$$

Since personal data are a particular subclass of information, the TBox includes the following axiom:

$$\forall_y[\,(\mathsf{personalData}\,y)\ \rightarrow\ (\mathsf{information}\,y)\,]\tag{5}$$

On the other hand, since in (4) the agent and the receiver may be any entity, the formula may be obviously evaluated also when these are data controllers and third countries or international organizations.

So far, the example serves the purpose of illustrating how formal correlations work. To take it one step further we would have to assume a substantive correlation between the terms in the two norms. This would mean that *transfer* is used in the two text with the same meaning. Under this hypothesis, if there is a legal interpretation stating that whenever a means of communication is "secure" with respect to the ISO/IEC 27001 standard, then it is "appropriate" for transferring personal data with respect to the GDPR, such an interpretation can be expressed as shown in (6).

$$\forall_k [\, (\mathsf{secure}\, k) \;\rightsquigarrow\; (\mathsf{appropriate}\, k) \,] \tag{6}$$

Based on such an interpretation, it would then be possible to automatically infer that, whenever the ISO/IEC 27001 standard is respected by an agent "$x$", so is (part of) the corresponding GDPR's provision.

Again, the correlation formalized in axiom (6) is defeasible. In case a court or a data protection authority later decides that not all secure communication facilities are appropriate from the point of view of provision ($a$), it is possible to add further higher-priority axioms to the knowledge base in order to override (6), thus keeping track of that specific legal interpretation.

## 6 Discussion and Conclusion

This paper introduces a semi-automated methodology to reuse existing security standards to help ensure legal compliance. By following the illustrated methodology, one can build a machine-processable *knowledge base* of logic formulæ that model relevant concepts from the text of a law and a security standard, together with their possible correlations. Once the knowledge base has been populated with a sufficient number of correlations, since standards can be certified by auditors, an enterprise that implements the standard can have an argument for compliance, at least with the provisions to which the standard correlates.

The knowledge base will thus allow to detect the legal provisions covered by the standard, and can then be used to assess to what extent the enterprise that has implemented the standard(s) complies with the law.

Although this paper focuses on the methodology only, several technical challenges related to building and updating the knowledge base are raised.

First off, the translations from natural language to logical formulæ must be uniform for excerpts of text that are similar to each other. To achieve this, we must overcome the limitation of a manual translation, which would be time-consuming and error-prone. For this reason, our work must rely on current NLP technologies. However, even at the best of their performances, NLP algorithms are still unable to automatically carry out the translation with a reasonable level of accuracy, so we advocate a *semi-automatic* translation of the provisions.

Similar approaches are applied to translations in general. Here, translators are helped by collaborative tools such as the "SDL Trados Studio"[17], which suggests, via pattern-recognition text-similarity NLP techniques [18,9], how to translate a sentence on the basis of the translations of similar sentences that the translators have previously stored in the tool. Using a feature used in tools

[17] http://www.translationzone.com/products/trados-studio/.

to learn foreign languages such as Duolingo[18], some translations acquire more importance the more they are chosen by highly-qualified end users.

Inspired by that approach, we will extend the LIME text editor to assist the manual translation of provisions into formulæ. For each provision, the editor will display the translations of similar provisions found via NLP procedures applied to the provisions already stored in the knowledge base.

Secondly, the knowledge base must be consistent, *i.e.*, without contradictions. To check for consistency, we plan to store formulæ in an XML-based data model, and employ/extend reasoners[19], to monitor the consistency of the knowledge base, whenever new formulæ are added to it.

The methodology herein is currently a work in progress and not fully implemented yet. The complete methodology requires the definition of a detailed taxonomy of concepts extracted from the law and the security standards.

In the future, we envision significant developments of the research presented in this paper. The knowledge base will be populated with *legal* interpretations that will be translated into defeasible formulæ. To this end, and along the GDPR example, web crawlers can be used to parse documents from relevant portals, *e.g.*, those of national DPAs in the EU, in order to check (via NLP) whether new interpretations of the GDPR provisions are available. The knowledge engineers can then update the knowledge base, overriding some defeasible implications occurring therein. This task can be more easily achieved if the the methodology and the techniques we have illustrated here are embedded in a web service. By accessing the service, users, such as legal experts, judicial courts and the like, will benefit from the collaborative expertise brought by the knowledge base; at the same time, they can help build and update the substantive correlations. In addition to the technical skills needed to manage the knowledge base, this step would greatly benefit from a close interaction with legal authorities. We envision that in a small country like Luxembourg that could be more easily achieved.

## References

1. Antoniou, G., van Harmelen, F.: Web Ontology Language: OWL. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, chap. 4, pp. 67–92. International Handbooks on Information Systems, Springer Berlin Heidelberg (2004)
2. Arora, C., Sabetzadeh, M., Briand, L.C., Zimmer, F.: Automated checking of conformance to requirements templates using natural language processing. IEEE Transactions on Software Engineering 41(10), 944–968 (2015)
3. Bartolini, C., Muthuri, R., Santos, C.: Using ontologies to model data protection requirements in workflows. In: Proceedings of the Ninth International Workshop on Juris-informatics (JURISIN). pp. 27–40 (November 2015), extended version to be published in LNAI book.
4. Benjamins, R., Selic, B., Gangemi, A. (eds.): Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications, Lecture Notes in Artificial Intelligence, vol. 3369. Springer-Verlag Berlin Heidelberg (2005)

[18] `https://www.duolingo.com/`.

[19] `https://www.w3.org/2001/sw/wiki/OWL/Implementations`.

5. Boella, G., Di Caro, L., Humphreys, L., Robaldo, L., Rossi, R., van der Torre, L.: Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. Artificial Intelligence and Law to appear (2016)
6. Boella, G., Di Caro, L., Graziadei, M., Cupi, L., Salaroglio, C.E., Humphreys, L., Konstantinov, H., Marko, K., Robaldo, L., Ruffini, C., Simov, K., Violato, A., Stroetmann, V.: Linking legal open data: Breaking the accessibility and language barrier in european legislation and case law. In: Proc. of the 15th International Conference on Artificial Intelligence and Law. ACM, New York, USA (2015)
7. Boella, G., Di Caro, L., Rispoli, D., Robaldo, L.: A system for classifying multi-label text into eurovoc. In: Proceedings of the 14th International Conference on Artificial Intelligence and Law. ICAIL '13, ACM, New York, USA (2013)
8. Boella, G., Di Caro, L., Robaldo, L.: Semantic Relation Extraction from Legislative Text Using Generalized Syntactic Dependencies and Support Vector Machines, pp. 218–225. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
9. Boella, G., Di Caro, L., Ruggeri, A., Robaldo, L.: Learning from syntax generalizations for automatic semantic annotation. The Journal of Intelligent Information Systems 43(2), 231–246 (2014)
10. Chalmers, D., Davies, G., G.Monti: European Union Law: Cases and Materials. Cambridge University Press (2008)
11. Davidson, D.: The logical form of action sentences. In: Rescher, N. (ed.) The Logic of Decision and Action. Univ. of Pittsburgh Press (1967)
12. De Hert, P., Papakonstantinou, V., Kamara, I.: The cloud computing standard ISO/IEC 27018 through the lens of the EU legislation on data protection. Computer Law & Security Review 32(1), 16–30 (February 2016)
13. Governatori, G., Olivieri, F., Rotolo, A., Scannapieco, S.: Computing strong and weak permissions in defeasible logic. Journal of Philosophical Logic 42(6), 799–829 (2013), `http://dx.doi.org/10.1007/s10992-013-9295-1`
14. Hansen, J.: Prioritized conditional imperatives: problems and a new proposal. Autonomous Agents and Multi-Agent Systems 17(1), 11–35 (2008)
15. Hobbs, J.R.: Deep lexical semantics. In: Proc. of the 9th International Conference on Intelligent Text Processing and Computational Linguistics (2008)
16. Horty, J.: Reasons as Defaults. Oxford University Press (2012)
17. Makinson, D., van der Torre, L.W.N.: Input/output logics. Journal of Philosophical Logic 29(4), 383–408 (2000)
18. Mihalcea, R., Corley, C., Strapparava, C.: Corpus-based and knowledge-based measures of text semantic similarity. In: Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1. AAAI Press (2006)
19. Reiter, R.: A logic for default reasoning. Artificial Intelligence 13, 81–132 (1980)
20. Robaldo, L.: Interpretation and inference with maximal referential terms. The Journal of Computer and System Sciences 76(5), 373–388 (2010)
21. Robaldo, L., Humphreys, L., Sun, L., Cupi, L., Santos, C., Muthuri, R.: Combining input/output logic and reification for representing real-world obligations. In: Post-proceedings of the 9th International Workshop on Juris-informatics. LNCS (2016)
22. Robaldo, L., Caselli, T., Russo, I., Grella, M.: From italian text to timeml document via dependency parsing. In: Proc. of 12th International Conference 'Computational Linguistics and Intelligent Text Processing'. (2011)
23. Vibert, H., Jouvelot, P., Pin, B.: Legivoc - connectings laws in a changing world. Journal of Open Access to Law 1(1), 165–174 (2013)