Defining a Model for Content Requirements from the Law: an Experience Report

Marcello Ceci
SnT, University of Luxembourg
Luxembourg
Email: marcello.ceci@uni.lu

Domenico Bianculli
SnT, University of Luxembourg
Luxembourg
Email: domenico.bianculli@uni.lu

Lionel C. Briand†
Lero SFI centre for Software Research
and University of Limerick, Ireland
School of EECS, University of Ottawa, Canada
Email: lionel.briand@lero.ie

***Abstract*—This paper reports on the experience of building a content model in collaboration with a national financial supervisory authority, with the goal of automating the compliance checking activity performed by the agents of the supervisory authority on fund documentation. The work is focused on modelling content requirements found in the law, i.e., deontic rules prescribing that some information is contained in an official document. For such requirements, the main modelling effort revolves around the required content and its information types. We therefore designed a process to build a content model, elaborating design criteria for the model which partly depend on the use case encompassing compliance checking. We built the content model through iterative interactions between a knowledge engineer and domain experts designed to ensure that the model is not limited to representing only the letter of the law, but rather represents the relevant distinctions in the practice of compliance checking. We drew lessons learned regarding the need for setting up classification criteria for information types and handling the trade-off between expressivity and maintainability of the model.**

***Index Terms*—Conceptual modeling, Methodologies and tools, Legal compliance, Regtech**

I. Introduction

Investment funds are required to issue documentation (e.g., prospectuses, Key Information Documents, management regulations) in order to inform the potential investors about the characteristics of the fund and the risks associated with investing in it. This documentation needs to comply with a number of legal requirements coming from different sources (e.g., laws, implementation acts, case law) and aimed at different goals (e.g., preventing money laundering, protecting investors, ensuring proper employee conduct). Automated compliance checking for those documents can bring great benefits, as this documentation plays an important part in ensuring the health and stability of the financial system, which is paramount to avoid a new financial crisis like that of 2008.

The research that encompasses the experience reported in this paper is part of a project in collaboration with a national financial supervisory authority, and has the goal of automating the activity of checking the compliance of fund documentation against the requirements introduced by relevant regulations.

Most of the requirements for legal documents are *content requirements*: in order to be compliant, a document must contain text that expresses the required information types. The core aspect of a content model, i.e., a machine-readable representation of content requirements, lies therefore in the identification of the information types that must be contained in the fund documentation. On that regard, discussions with agents of the national financial supervisor confirmed the hypothesis that compliance checking on the contents of fund documentation cannot be reduced to a checklist, where each requirement is matched to a span of text in the fund document that expresses the required information type, but it is rather a process that involves a more comprehensive evaluation of the local context (e.g., how the law is applied in that jurisdiction, industry or time period) and of how the regulated entity submitting the fund document fulfills the requirements [1]. While the law describes a desired reality using generic and abstract terminology, its application to the case in point — which may vary greatly — depends not only on the letter of the law (including implementing acts [2] such as technical specifications, case law, FAQs), but also on the specific situation being evaluated [3]. In the case of content requirements, while the law describes the desired information types with abstract terminology, fund documents are written taking into account additional factors such as industry standards and common practices, and employ a language aimed at clearly providing information, but also promoting investments and managing risk. This implies that, in order to support the application of legal requirements, a model cannot only represent the desired reality described in the law, but must rather adapt to the practice of fund documentation.

The main challenge in representing compliance with content requirements thus consists in capturing the criteria used by the agents of the national financial supervisor to classify the content of a prospectus as either matching or not matching required information. These criteria are not always directly, unambiguously, or logically derivable from the letter of the law. This challenge can be overcome by adapting the model to the practice of compliance checking as performed by the agents, asking questions such as *what do agents look at when they evaluate the compliance of a document?* and *how do agents combine the available information to decide whether to accept or reject a document?*

This paper reports on the experience of creating content requirements through the specification of a model of required information types. The model is conceived to support the

†Part of this work was done while the author was affiliated with SnT, University of Luxembourg.

compliance checking of fund-related documents against the requirements of Schedule A of the UCITS (Undertakings for Collective Investment in Transferable Securities) EU directive [4], providing the legal framework for the organization and management of investment funds. It is composed of three main modules: a simple model of fund documentation, a content model of information types required by Schedule A, and a model of fund-related entities that is used to express conditions and constraints. The classes of the model and the design criteria are presented in an online annex [5].

The paper presents the design criteria and the process adopted during the experience. The lessons learned revolve around the peculiarity of conceptual modeling in RegTech, a domain still lacking the theoretical foundations of an established academic discipline [6]. Specifically, such lessons concern the importance of defining **classification criteria** that reflect the freedom granted to the domain experts while interpreting legal requirements, and a domain-specific declination of the well-known **trade-off between expressivity and maintainability of a conceptual model**.

The paper is structured as follows: Section II introduces content requirements, as well as the context and the research methodology. Section III analyzes the state of the art. The next two sections cover the model itself: Section IV introduces the design criteria as impacted by the foreseen use cases, and Section V reports on the experience of building the content model for Article 69 of the EU UCITS directive. Section VI discusses the lessons learned from the experience, the differences with existing model development processes, and the threats to validity. Section VII introduces the resulting model and shows how it can be used to express legal requirements. Finally, Section VIII concludes the paper.

II. Background, Context, and Methodology

A. Content Requirements found in the Law

The concept of "compliance" usually means compliance with regulations or, more specifically, with the deontic rules contained therein. A *deontic* [7] (or regulative) *rule* is a rule that expresses a desired reality, implying one or more actions by an addressee. If the reality matches the desired reality we have *compliance*, otherwise we have a *breach*.

We define content requirements found in the law as *deontic rules requiring that some information* (e.g., the description of a process) *is contained within an official document* (i.e., a document that fulfills a legal function — such as a contract, a document submitted to an authority, or a publicly available document). For the addressee of the requirement, which is an author of the document, a content requirement imposes *an obligation to include that information in the document*.

The basic model for content requirements includes the concept of the regulated document and the concept of the required content (see Figure 1). Identifying content requirements in the law is quite straightforward: the required action is usually a verb that expresses containment (such as "contains", "includes") or conveyance of meaning (such as "indicates", "describes"). This is well exemplified in the legal article that

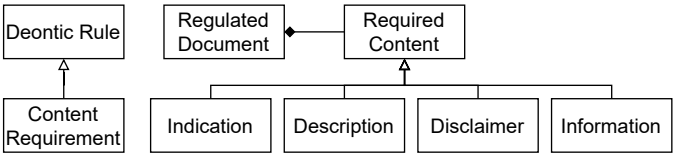

Fig. 1. Model of content requirements (left) and of the required state of affairs (right).

constitutes the core of our case study, Article 69 of the EU UCITS Directive: "The prospectus shall *contain* at least the information provided for in Schedule A of Annex I of this Law [...]". In this case, the concepts of regulated document and required content correspond to "prospectus" and "information provided for in Schedule A of Annex I of this Law", respectively. The basic model for the content requirement of Article 69 of the EU UCITS Directive would therefore be "the Prospectus *shall contain* Schedule A Information".

In this basic model, the required content ("Schedule A Information") is classified into *information types*. Information types identify the contents expressed in documents by content requirements. At the highest level of abstraction, an information is a generic text, potentially covering several descriptions. It can be defined using one of the following forms:

- A single *indication* of a name, an amount, or another value (e.g., name of the director, date of establishment of the fund);
- A *disclaimer* with fixed content (e.g., disclaimer on availability of financial reports);
- A *description* expressed in a text span of varying length, from a few characters (e.g., the duration of the fund, the frequency of calculation of the issue price) to hundreds of words (e.g., the procedure for the liquidation of the fund, the method of calculation of the issue price);

The content required in the basic model for Article 69 ("Prospectus *contain* Schedule A Information") is generic *information*, that can be further specified through the analysis and interpretation of Schedule A, to derive a list of information "sub-types" (subconcepts of "Schedule A Information") including *indications*, *disclaimers*, and *descriptions*. Since information types identify semantic contents and not strings of text, it is possible that one text span in the documentation expresses more than one information type, or that spans expressing different information types overlap.

B. Context

The goal of the research behind this paper is to speed up the process of compliance checking of fund documentation by supervisory authorities. This process has become very cumbersome since the number of prospectuses increases constantly. However, such a process is not easily modifiable, since it pertains to the delicate mission of the supervisory authority to guarantee the credibility and stability of the financial sector.

The ultimate goal of the research is to capture the essence of the checking process in order to automate it. We aim for a manageable model that allows the automated compliance checking solution to provide outputs that are understandable

by humans (the domain experts) so that they can analyze these outputs and facilitate their application of the compliance checking process without altering it or reducing its precision.

C. Research Methodology

The process for developing the content model is an example of artifact engineered following the *design science* research methodology. More specifically, the paper follows the steps of the approach by Peffers et al. [8]: the *practical motivation* (1) of the research is illustrated in § I and the *problem definition* is further outlined in § III, based on published research. The *objectives* (2) are formulated in § II-B and serve as a basis for the *design and development* (3) presented in § IV and § V. Section VII *demonstrates* (4) the results of the process (i.e., the model). Section VI provides elements towards the *evaluation* (5) of the process.

III. State of the art

In the field of computer science, we have several examples of using models of the law to assess the compliance of business processes [2, 9]; the experience described in this paper stemmed from a different purpose, namely that of using models to asses the compliance of official documents.

Content requirements. Previous research has identified a category of legal requirements that aims at ensuring the *completeness* of documentation, e.g., in research related to GDPR [10, 11, 12], "completeness" in privacy policies is achieved if the document "contains descriptions which should be explained in privacy policies, such as how to deal with cookies". This is not to be confused with the completeness of requirements addressed in other work [13]: while the latter is a "requirement for the development of requirements", the completeness of documentation is a (set of) requirement(s) for the development of documentation fitting certain descriptions.

Modelling approaches. In the field of semantic technologies, Fawei et al. [14] present a methodology to develop criminal law ontologies and rules to support query answering. In the financial domain, the Financial Industry Business Ontology (FIBO [15]) is aimed at representing the relevant concepts in the financial industry, which can then be used as building blocks for expressing requirements or restrictions.

In AI and Law, the contributions are at a more theoretical level, and not specifically focused on content requirements. Examples of conceptualizations of requirements include models of business process compliance [16], compliance modeling frameworks [17], ontologies for the representation of legal requirements (the most relevant being LKIF [18]), and rule languages such as LegalRuleML [19], a markup language for legal rules based on RuleML.

In the field of Requirements Engineering, most contributions are focused on the creation of models from legal requirements [20], which however are very different from legal and regulatory texts. Among the contributions that focus on legal texts we have NomosT [21], which aims to automatically extract legal requirements.

In the field of data representation, Blums et al. [22] recently proposed the consolidation of existing economic exchange ontologies with the purpose of setting a standard for financial reporting: the same approach could be envisaged for fund documentation, where models such as the one presented in this paper could be used to clarify concepts and set a standard for the creation of documentation, which would in turn facilitate the automation of many tasks related to compliance checking.

IV. Building the Model: Design Criteria

In this section, we describe the criteria for a representation of a document's contents that enables the assessment of their compliance with content requirements. Since the level of granularity of the model (and thus the design criteria) depends on the use case, we first introduce the possible use cases involving the checking of content requirements.

A. Use Cases

As part of our collaboration with the financial supervisor, we identified two use cases related to compliance checking:

1) **Query-answering-assisted Compliance Checking (QCC)** consists in retrieving, from a document, the text blocks expressing the content relevant to a given requirement (e.g., as proposed in [23]). This use case is also called "legal information retrieval" in the literature [24]. For example, a query to retrieve the content relevant to the requirement of indicating the "duration of the fund" would return the following underlined sentence:

> *"The Fund is created for an unlimited period. The Management Company may [. . .] decide upon the liquidation of one or more Sub-Funds [. . .]".*

Note that in this use case the compliance checking is not fully automated, since the human expert still needs to make a decision by manually analyzing the retrieved text. The main limitation here is related to content requirements which are bound to conditions. For example, if a content requirement says that "if the fund contains compartments, it must indicate the procedure for their liquidation", query answering can allow the user to retrieve all occurrences of "indication of procedure for the liquidation". However, since the condition "the fund contains compartments" cannot be evaluated, the query result alone does not allow the user to assess which of those indications are legally required in the first place.

2) **Fully Automated Compliance Checking (ACC)** consists, as the name suggests, in automatically checking the compliance of a document with regard to content requirements (e.g., as proposed by Torre et al. [11]). In this use case, an automated procedure determines which content is *required* (checking any existing pre-conditions for the rules' applicability) and tries to locate text expressing this content in the document. If the required content is found, the document is marked as "compliant", otherwise it is marked as "non-compliant". Ideally, the outcome of the compliance check may go beyond the binary result "compliant/non-compliant", to assess different levels of compliance (e.g., the compliance

required for a regular company vs. the compliance required for a joint-stock company).

B. Structural Granularity

As explained in Section II-A, proper processing of content requirements calls for a detailed model of the required contents (information types). The level of detail of a model is called *granularity* in the field of conceptual modeling. In a model of content requirements from the law, the level of detail depends on the interpretation and application of the law. The different aspects of granularity have been described in the past [25]; in the case of a model for content requirements, granularity is meant as structural complexity regarding the number of model elements [25, p. 19], and their hierarchy.

When representing required information, structural granularity impacts two aspects of the model design: (1) how many **sub-types of information** to represent, and (2) whether to capture **both core and non-core information** (i.e., how to represent different degrees of relevance of a textual content to a content requirement).

Representation of information sub-types. Representing information sub-types means identifying the smaller pieces of information that compose an information type required by the law. An example is point 10 of the Schedule A in our case study, according to which prospectuses are required to describe "conditions and procedure for liquidation". Here, we could distinguish the sub-types "description of the conditions" and "description of the procedure", or represent the whole point 1.10 as a single class. Distinguishing between conditions and procedure enables a finer-grained compliance check but also increases the complexity of the model. Conversely, representing procedure and conditions as one single description will lead to a simpler model, easier to understand and maintain, but possibly too shallow to properly verify the completeness of the provided information.

On this regard, we note that the higher the level of granularity, the more concepts are represented and thus the more complex the model gets. This may have repercussions on certain applications that rely on the model. For example, some machine learning-based approaches for compliance checking might require a training set of documents annotated with the different information types, and the process of annotating such documents is more cumbersome and error-prone when adopting higher levels of granularity. Moreover, the more specific the information, the more its identification in the text may rely on legal interpretation, which in turn may lead to inconsistencies in the annotations. When representing content requirements, it is therefore paramount to strike a balance between the expressivity of the model and its complexity.

The use case has the following impact on the representation of information sub-types:

1) *For QCC*, it is important to attain at least the level of detail required to represent the relevance of the text strings to the requirements. For example, if a part of a description that is relevant to requirement *a* includes a smaller piece of information that is also relevant for requirement *b*, the latter should be represented as a distinct sub-type (and marked as relevant for both requirements *a* and *b*).

2) *For ACC*, a higher level of detail is required. It is in fact necessary to represent certain detailed aspects of the document's contents (e.g., the value of a measurement, or the country of establishment of an investment fund) in order to check the pre-conditions of content requirements.

Relevance of non-core information. The *core information* of a content requirement is defined as the minimal information required for a regulated document to be considered compliant with such a requirement. In contrast, *non-core information* may be useful or even necessary, but is not sufficient to determine compliance. For example, sentence (1) "the price is calculated every business day" is core information for the content requirement *description of the frequency of calculation*, whereas the definition of business day (e.g., (2) "a business day is defined as any working day") is non-core information for the same requirement. Capturing the quoted sentence (1) without (2) is enough to know that the information is there, even if we may not be able to understand the exact semantics of the description. In contrast, capturing (2) without (1) does not allow us to establish whether the information on the frequency of calculation is provided.

Including non-core information in the model allows users to represent different levels of relevance of a description to a required information type. This is paramount for conditional requirements and other complex compliance checking procedures that yield a more informative assessment than a simple binary "compliant/non compliant" answer (e.g., assessing compliance for different kinds of companies or assessing the penalty to be applied in case of breach). On the other hand, representing non-core information increases the number of classes of the model as well as its complexity, especially if we consider that the distinction between core and non-core information might depend on the case in point.

The use case has the following impact on the relevance of non-core information:

1) *For QCC*, it is not necessary to represent non-core information. It is however possible that certain non-core information helps evaluate a document content. For example, if a document uses uncommon terminology (e.g., "accountable day" instead of "business day") to describe the *frequency of the price publication* (core information), retrieving the definition of such an uncommon term (non-core information) could facilitate the proper evaluation of the compliance of this information. In these situations, non-core information should be included in the model, albeit clearly flagged so as not to influence any compliance assessment.

2) *For ACC*, representing non-core information helps better specify the pre-conditions of a content requirement. For example, if a fund has a characteristic (e.g., the presence of fund compartments) that is a pre-condition for a requirement (e.g., the requirement of indicating the procedure for the liquidation of those compartments), then the indication of that characteristic constitutes non-core information for the conditional content requirement.

V. Building the Model: Experience Report

In this section we describe our experience of building a model for content requirements in the financial domain, the design choices made while building the model in application of the criteria outlined in the previous section, and the validation performed on the model through interactions between a knowledge engineer and the domain experts.

A. Case Study

Our case study focuses on modeling the content requirements expressed in a national transposition† of article 69 of the EU UCITS Directive. A shortened version of this article was presented as example in Section II-A. The full article recites:

> "The prospectus shall *contain* at least the information provided for in Schedule A of Annex I of this Law in so far as such information does not already appear in the management regulations or instruments of incorporation annexed to the prospectus in accordance with Article 152(1)"†

This provision requires funds to include in their documentation (e.g., prospectus, management regulations, instruments of incorporation) all the information types that are listed in the annex of the law called "Schedule A". Schedule A is composed of two main types of content, structured in a list: while points 2 to 6 of the list are free-text paragraphs, point 1 (see Figure 2) is a table with 18 rows (numbered from 1.1 to 1.18) and three columns, for a total of 54 cells of which only 39 are filled. While the rows identify different types of information, the columns indicate the legal body to whom the required information must refer:

- Columns 1 and 3 correspond to two different types of investment funds, namely *mutual funds* and funds run by *investment companies*. Since a fund can only be of one of these two types, columns 1 and 3 are mutually exclusive in their application to any case in point.
- Column 2 refers to the management company. All mutual funds must appoint a management company, so the information types indicated in this column must always be included in the documentation of mutual funds. Investment companies instead have the option to entrust management and administrative functions either to a management company (so-called *externally managed companies*) or to an internal board (so-called *self-managed companies*): in the first case the information types of column 2 must be included in the documentation, while in the second case they are not required.

Each cell of the table of Schedule A (and each sentence, for points 2 to 6) corresponds to one or more information types. We remark that, in addition to the type of investment fund (mutual fund vs. fund of investment company) and adopted management scheme (self-managed vs. externally managed)

1. Information concerning the common fund	1. Information concerning the management company, including an indication whether the management company is established in a Member State other than the home Member State of the UCITS	1. Information concerning the investment company
1.1. Name	1.1. Name, corporate name, legal form, registered office and head office if different from registered office	1.1. Name, corporate name, legal form, registered office and head office if different from registered office
1.2. Date of establishment of the common fund. Indication of duration, if limited	1.2. Date of incorporation of the company. Indication of duration, if limited	1.2. Date of incorporation of the company. Indication of duration, if limited
1.3. In the case of common funds having different investment compartments, indication of the compartments	1.3. If the company manages other common funds, indication of those other funds	1.3. In the case of investment companies having different investment compartments, indication of the compartments

Fig. 2. Excerpt of Point 1 of Schedule A

— which determines the applicable columns in the table — it is also possible that a content requirement from Schedule A is *conditional*, as already discussed in Section IV-B. For example, point 1.3 of Schedule A requires to include the "indication of compartments", but only "in the case of common funds having different investment compartments". The condition might even be implicit. For example, point 1.8 for investment companies (column 3) requires to indicate the "names and positions in the company of the members of the administrative [...] body"; this requirement only applies to self-managed funds of investment companies, since externally-managed funds' administrative functions are performed by the management company. However, this circumstance is not explicitly stated in the Schedule A, being instead derived from the general law of investment funds [26, pp. 366 ff.].

B. Team and Settings

The work was done as part of a collaborative research project with a national financial supervisory authority, with the purpose of automating the compliance checking process performed by the agents of the authority on the fund documentation. The model was developed over 16 months, through meetings with the agents of the authority and exchange of clarification documents. The first phase was conducted in two rounds (two Q&A meetings with subsequent update of the model), corresponding to a total of 36 man-hours. The second phase required four rounds, each of which included one Q&A session, the creation of a clarification document, and the update of the model and of the classification criteria, resulting in a total of 176 man-hours. We used the Eclipse Modeling framework to create the model. The development of the model involved a team consisting of one legal knowledge engineer with background in legal informatics and four domain experts specialized in compliance of financial documentation, Such experts were all employees of the supervisory body; two of them had a background in finance, with respectively 4 and 10+ years of experience in the supervisory body, and two had a background in law, with 10+ and 20+ years of experience.

C. The Process of Building the Content Model

The process for building the content model of the information types required by Article 69 is illustrated in Figure 3: steps and activities are enclosed in squared boxes, and decision

†Article 151(1) of Luxembourg's the Law of 17 December 2010 relating to undertakings for collective investment

†Loi du 17 décembre 2010 concernant les organismes de placement collectif (Mémorial A n. 239 de 2010)

points are enclosed in diamond boxes. The process involves two actors: a knowledge engineer (KE) and several domain experts (DEs). In our case these are the first author and the agents of the financial supervisory authority, respectively. Figure 3 indicates the actor who performs or contributes to each step or activity. The process can be divided in two parts: the construction of the initial taxonomy (top part of the process model in Figure 3) and the validation and refinement of the content model (bottom part). The initial input of the process consists of the text of the law for which compliance checking needs to be automated. The output is a content model that can be used in downstream tasks (e.g., classification of text blocks relevant to a certain concept of the model).

Construction of the initial taxonomy from the letter of the law. The process begins with the KE extracting an initial taxonomy of information types from the regulation, with a rigorous approach strictly adherent to the legal text: each distinct class is derived only from the letter of the law. For example, for point 1.12 of Schedule A "Procedures and conditions of issue and/or sale of units", the KE initially identified four information types: "procedure of issue of units", "procedure of sale of units", "conditions of issue of units", and "conditions of sale of units". This taxonomy is then reviewed through one or more Q&A sessions (**QA1** in Figure 3), resulting in the following indications for its refinement:

- If the representation is too shallow for the practice (**DP1**), e.g., because several sub-types of such content may be present at the same time in the regulated document, new information types are created (**MR1**). For example, the indication of 1.1 "Fund Name" was distinguished into "Mutual Fund Name" (for mutual funds) and "Investment Company Name" (for funds run by investment companies).
- If, on the contrary, a distinction from the letter of the law is not meaningful or useful in practice (**DP2**), it is removed (**MR2**). For example, regarding 1.12 ("Procedures and conditions of issue and/or sale of units"), the KE asked the DEs whether "issue" and "sale" correspond to different activities of an investment fund. The feedback provided by DEs highlighted that "issue" and "sale" identify the same activity, and therefore for 1.12 the initial list of four information types was reduced down to two: "procedure of issue/sale of units" and "conditions of issue/sale of units".
- Successively, in order to improve the structure of the model and make it more easy to navigate, generalizations are created whenever possible (**DP3**, **MR3**). For example, during our experience several information types pertaining to rows 1.10, 1.12 and 1.13 were generalized as "Information on Units".

The review described above is repeated until, in **DP4**, the taxonomy is considered complete (i.e., no foreseeable situations are overlooked) and consistent (i.e., no foreseeable situations lead to ambiguous or wrong classifications), at which point it becomes the initial version of the content model. We remark that in **DP4** the model is verified against the letter of the law, i.e., against the semantics expressed by the regulations.

Validation and refinement of the content model. The initial version of the content model is validated through instantiation, i.e., by reviewing sample annotations of the information types in regulated documents. Inconsistencies in annotations are noted and give rise to critical questions leading to the refinement of the model, with the final goal of adapting it to the practice of compliance checking. For this activity we followed an approach where initially the KE annotates a few sample documents and the DEs review such annotations. This was made possible by the KE being a legal expert. This strategy puts the KE in an active role in the development of the model (e.g., suggesting labels or distinctions), thus saving the DEs' time and favouring a more consistent approach to the model construction. On the other hand, taking away the initiative from the DEs poses the risk of introducing criteria and distinctions for the model that are not familiar to them. In the first iteration of the validation, the KE also established a first version of the *classification criteria* (see Section V-D below), which were not shared with the DEs to avoid bias. In a second phase, the roles were inverted: the DEs annotated the sample documents and the KE reviewed the annotations highlighting any inconsistencies or discrepancies with the model. In this phase, the classification criteria were shared with the DEs, for two purposes: to measure their impact on the annotations (especially on their consistency), and to put the DEs in charge of the accuracy and consistency of the model, thus preparing them to autonomously maintain (and further refine) it. However, as we will see in Section VI-A, sharing the criteria poses the risk of influencing the very process that we are trying to automate.

In case the review highlights inconsistencies among the instantiations (**DP5**), or discrepancies between the instantiations and the model (**DP6**), a Q&A session (**QA2**) is performed and the results are recorded in a clarification document. Through such document, the KE can determine whether two information types are ambiguous when instantiated (**DP7**): in that case, if the distinction is not useful for compliance checking (**DP8**), the distinction is removed and the model is modified accordingly (**MR4**). For example, it emerged that the distinction between the two information types of 1.12 was often unclear in the text, and that the distinction did not help the DEs in their compliance check: following clarification between the KE and the DEs, the two information types of 1.12 were merged together into a single information type ("procedures and conditions for issue/sale of units"). The same happened for 1.10 ("procedures and conditions of liquidation", see Section IV-B). On the contrary, if in DP8 it emerges that the distinction is useful for compliance checking, this implies that the model lacks the detail needed to represent certain information sub-types, in which case a new information type is created (**MR5**) and the criteria for classification are updated accordingly. In our experience this was not needed (possibly because the initial taxonomy was detailed enough).

As the model gets refined, ambiguous distinctions in the information types are expected to decrease, with changes focusing instead on new types of content that were not found

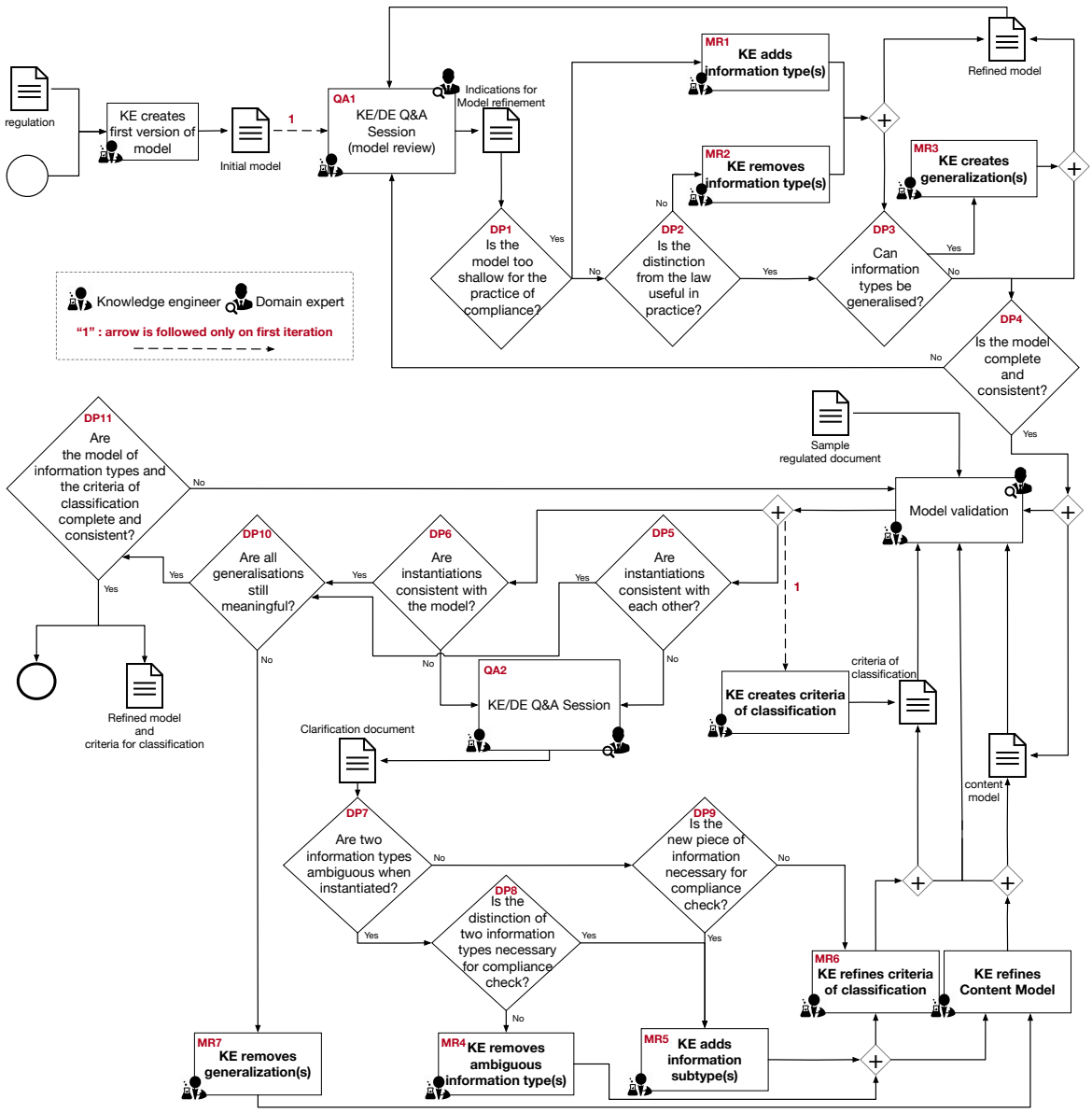

Fig. 3. The process of constructing the Content Model

in previous instantiations. In these cases, if the new piece of information is relevant for compliance checking (**DP9**), it is added as a new information sub-type (**MR5**), otherwise it is added to the classification criteria of an existing information type (**MR6**). For example, "minimum subscription" was identified as information in 1.12, but not relevant enough to become an information sub-type, and therefore it was added to the classification criteria for that information type.

After all the issues that emerged during instantiation have been addressed, the generalizations are checked again (**DP10**), to determine if they have become irrelevant (e.g., because all the information types in the generalization have been removed), in which case the generalization is removed (**MR7**).

The validation is repeated using new sample documents, until the model is considered to be complete and consistent enough for practical application (**DP11**), at which point it becomes a *definitive version*, together with the criteria of classification. We remark that in **DP11** the model is verified against concrete examples of financial documents.

D. The Activities for the Refinement of the Content Model

The feedback on the content model was gathered both through discussion on abstract categories (asking questions such as "*do you want to distinguish conditions of liquidation and winding-up procedure?*") and through model instantiation (i.e., the review of sample annotations). We noticed that the review of sample annotations was often contradicting what was decided in abstract terms and thus, following our approach prioritizing the practice of compliance checking, we favoured the feedback given on the case in point. During this process we performed two main activities: (1) the refinement of the

information types, and (2) the refinement of the criteria for the classification.

Refinement of the information types. Adapting the model in light of the feedback received from the DEs implies adding or removing information types, distinguishing core and non-core information, and creating generalizations. We ***added or removed information types*** in two circumstances:

In the model review phase, if an information type needed to be differentiated into two sub-types to support the practice of compliance checking, we added those sub-types in the model (activity **MR1**). For example, we distinguished between the concepts of "fund name" and "company name". If instead the distinction in the letter of the law was not reflected in a clear difference in practice, and resulted in two (or more) logically undistinguishable information types, we did not maintain it (activity **MR2**). An example is the distinction between "issue of units" and "sale of units", which in practice identify the same activity. To make these decisions we gathered the opinion of the DEs, asking questions such as "*are both fund name and company name likely to appear in the same document?", "should we distinguish the calculation of the issue price into the criteria for calculation and the frequency of calculation?"*, and *"should we distinguish asset enumeration within the issue price calculation?*". Overall, we tried whenever possible to maintain the granularity expressed by the letter of the law: one concept (as expressed in a sentence or in a cell of the table) corresponds to one information type.

In the validation phase, we removed ambiguous distinctions expressed by the letter of the law that are not strictly necessary in the practice of checking (activity **MR4**). We considered ambiguous any information type for which the DEs cannot achieve a consistent instantiation (i.e., those for which the DEs cannot agree on consistent criteria for the classification/annotation). For example, we initially distinguished an information type "information on liquidation" into sub-types "conditions" and "procedure", but the DEs were inconsistent in annotating a block of text (related to "notification of liquidation") as one or the other sub-type, and therefore we removed the two sub-types. In these cases, to avoid inconsistencies and the overburdening of the DEs, it is usually preferable not to represent such ambiguous and irrelevant distinctions, leaving it to the DEs to manually evaluate the completeness of the information provided by looking at a larger text span (in our example, a text span including both "conditions" and "procedure" for liquidation). If the ambiguous distinction is instead considered decisive for assessing compliance, we maintain the distinction and identify a third sub-type of information that includes the ambiguous content (activity **MR5**). In our example, we would need to create a third sub-type for information on liquidation, called "notification of liquidation". A new information sub-type is added also in case of a new piece of information that is considered decisive for assessing compliance (decision point **DP9**). As said before, in our case study the initial taxonomy was detailed enough that further distinctions among information types proved unnecessary in the validation phase.

Distinguishing between core and non-core information enables the identification of different degrees of relevance of an information type towards compliance checking (see Section IV-B). For each information type, feedback from the DEs allowed the KE to detect which information pieces are non-core. The feedback was gathered by asking questions such as "*is this text decisive for your assessment on compliance?*", or conversely "*is this text providing additional information?", "is this text relevant to determine whether other requirements apply?*". To limit the complexity of the content model, non-core information was not represented in the model but rather included in the classification criteria (see below). If a distinction is not necessary for QCC, but it is necessary for ACC (e.g., for checking pre-conditions), this distinction can be excluded from the version of the model aimed at QCC, and introduced only in the model aimed at ACC (see Section V-E).

Creating generalizations for information types that concern the same topic is an activity that was performed at the end of each round of review of the model (activity **MR3**). If different information types concerned the same fund-related entity or activity (e.g. issue of units, price of units), the KE created a generalization that included them. To make this decision, the KE asked questions such as "*do you think that 'price of units' is the common topic among the following information types?*". The main purpose was to make the model more manageable. Note that in some cases this activity resulted in one generalization including information types from multiple sentences or cells of the table. During the validation phase, the KE removed any generalization that was rendered meaningless (activity **MR7**). This can happen e.g. because the distinction between the generalized information types is ambiguous. In our case study, however, this circumstance did not arise.

Refining the classification criteria. Having clear criteria for what is included in — and excluded from — an information type is very important to review and validate the content model built from the letter of the law. In our experience, it was impossible to set precise classification criteria during the first phase of model construction, because most of the criteria were not expressed in legal or regulatory texts, but rather emerged from the practice of compliance checking. In this regard, instantiating the model by annotating sample documents and reviewing those annotations proved to be more efficient than discussing design decisions at an abstract level with the DEs.

After the first round of validation, the KE drafted a first version of the classification criteria on the basis of the instantiations and validated it through questions to the DEs, such as "*do you always include the 'notification of liquidation' in the 'procedure of liquidation'?*", "*do you always exclude the 'asset valuation' from the 'method of price calculation'?*", and "*do you consider 'portfolio composition' as 'asset valuation'?*". At first the criteria were not shared with the DEs, to avoid influencing the instantiation process. After a few rounds of annotation and validation, the criteria were made explicit. This was done in order to facilitate the construction of the training set by the agents themselves. As already noted, this introduces the risk of influencing the way in which the information is

processed (see Section VI-A), but it is also a necessary step to increase consistency among instantiations, thus enabling proper validation of the model.

Whenever an inconsistency among instantiations emerged and it was not due to ambiguous information types, we identified new classification criteria for an existing information type (activity **MR6**) or we created new information sub-types (activity **MR5**). The choice depends on **DP9**, and also on considerations regarding *the trade-off between the expressivity of the model and its maintainability*. Representing all relevant pieces of information as information sub-types will in fact result in a complex model that is difficult to navigate and expensive to instantiate. On the other hand, a shallow model is less intuitive to instantiate as the DEs would often need to check the criteria for classification. A shallow model is also less expressive, as it does not allow the user to automatically check compliance with content requirements beyond the level of the information type. For example, it would be impossible to determine whether information on "minimum subscription" is missing in the fund documentation (see Section V).

E. Building the Model of Fund-related Entities

As said before, in order to enable ACC, a model must express the concepts that are needed to specify the preconditions and the restrictions posed by requirements. For Schedule A, we built a *model of fund-related entities*, in the following way:

1) During the analysis of Schedule A, the KEs created a list of concepts and relations that played a role in the specification of compliance criteria (especially of their conditions, for example, the concept of "UCITS compartment" in order to verify the precondition of point 1.3);

2) These concepts and relations were then arranged in a model, relying on the expertise in banking and company law of the legal expert in the research team;

3) This initial model was refined through Q&A documents similar to the ones used for content models, asking questions such as: "*Is it possible for a self-managed SICAV to appoint third parties?*" and "*Are the administrative bodies of externally managed SICAVS and FCPs the same of their appointed management company?*".

Since the model of fund-related entities is a domain model and neither a content model nor a formal ontology, we did not follow strict formal criteria for its construction. Still, while building the model we only included concepts that correspond to actual entities in the real world (people, institutions, shares), excluding immaterial concepts (e.g., "majority", "liquidation").

VI. Discussion

A. Lessons Learned

From designing the process, we learned the following lessons related to the additional challenges posed by a model for content requirements:

Trade-off between expressivity and maintainability. The first lesson regards the question of how far should the modelling effort go in order to represent the minutiae of legal and domain knowledge. Whenever an inconsistency among instantiations emerged, and it was not due to ambiguous information types, we identified new classification criteria for an existing information type (activity **MR6**) or we created new information sub-types (activity **MR5**). The choice depends on **DP9**, but we also learned that it depends on considerations regarding the foreseen use case and the resulting trade-off between the expressivity (i.e., the level of detail to which data can be modeled [27]) of the model and its maintainability, i.e., the effort required for building and updating it. Representing all relevant pieces of information as information sub-types will result in a complex model, expensive and possibly confusing to instantiate. On the other hand, a shallow model is both difficult to interpret (the DEs would often have to check the classification criteria), and less expressive (since ACC cannot be performed beyond the level of information types). The authors find that an ideal trade-off consists in (1) creating new information types for distinctions that are common in the practice of compliance checking, and (2) updating compliance criteria when the distinctions are (a) rarely relevant in practice, (b) have likely no practical impact on the compliance assessment, or (c) are normally assigned to experts who would be aware of these specificities.

Importance of establishing classification criteria and associated risks. The second lesson concerns the classification criteria, which we have introduced as a way to ensure consistent instantiations of the model without increasing the complexity of the model. Because they are not embedded into the model but rather described in the guidelines for the use of such model, classification criteria guide the domain experts in their annotations without binding them. They should in fact be seen as suggestions or arguments towards a certain classification of content, in a perspective of compliance checking as an evaluation whose ultimate criteria are always up to debate. However, if seen as strict rules, bespoke criteria of classification pose the risk of constraining the agents' autonomy in their compliance checking activity. Indeed, agents enjoy a certain degree of discretionary autonomy in interpreting the regulations, and reducing (or removing altogether) this autonomy would impact the checking process itself. Technologies in fact are not merely tools for implementing goals: they shape the meaning of those goals themselves, creating "a world view that alters the perceptions of the decision-makers they inform" [1]. This stays an open dilemma as we learned that it is impossible to formalize the checking process without influencing it, making it more structured, consistent, and ultimately predictable.

These lessons are inspired by the goal of bringing closer together the research in the representation of legal requirements and the practice of compliance checking, which — as it is now clear — cannot be reduced to a checklist. The model can be scaled by applying the same process described in Section V to other sections of the law. This needs a KE to interact with DEs to check, discuss, and adapt both the model and the classification criteria. The entire process can be very expensive and time-consuming and requires a strong commitment of resources. A possible solution could consist

in the use of Large Language Models [28, 29] to speed up parts of the process, especially in the first phase of creation from the letter of the law (DPs 1-3); however, this is outside the scope of this work. Another approach would be to create protocols to allow the DEs to autonomously build and refine the model and the criteria for classification.

B. Differences with existing model development processes

The development process of our model differs from the processes followed in existing work in the following ways:

1) In the domain of RE, Zeni et al. [21] focused on the creation of models of the law that only take into account the legal text as a source. Instead, in this work we acknowledge the existence of two layers: the *provision* and the *norm* [30]. The norm is the result of the legal interpretation; its granularity can be different from the one provided in the letter of the law. There is in fact a discrepancy between the letter of the law and the practice of compliance checking [2]. In the case of financial supervision, the relevant interpretation is that in place in the supervisory body's office, as applied by its agents. For this reason, in our process we first extract the legal terms from the law and then expand them through questions to the experts. In this way, we do represent the law not only at a high level (as done by Zeni et al. [21]) but also in the practice of compliance checking.

2) Amaral et al. [31] asked legal experts to perform the validation of the model representing the relevant concepts of the law. In this work, we involved as domain experts in the model development process agents from the supervisory body, who are the people performing the actual compliance checking. The former can be seen as a form of *doctrinal* contribution, since the legal experts provided an interpretation of the law in an abstract situation. The latter can be seen as a form of *jurisprudential* contribution, since the domain experts provided their expertise in the context of a uniform course of decision by the supervisory body [32, p. 272].

3) Soltana et al. [33] followed an approach similar to ours; however, they described the process only briefly. In this work, we present the process in details and draw lessons from it.

4) At the same time, we limit the complexity of the model compared to ontological models [30, 14] (which are aimed at semantic web and have interoperability/reusability as primary goal) and other solutions in the field of AI and Law [34]. In this way, we ensure that the model can be perused both by IT experts and by domain experts. We therefore create the concept of content requirement: instead of modelling an entire rule, we focus the modelling effort on the information types.

The main advantage of our approach is the relevance of the knowledge represented therein, since it supports the actual compliance checking process. Using only the information derived from the law would not be sufficient, since the interpretation of the language of the law is what matters [2]. The main disadvantage is the need to define some complex criteria of classification. Defining such criteria may bring some discrepancies between the letter of the law and the practice of compliance checking. This may affect the compliance process and disrupt the perception of the process by the people involved [1].

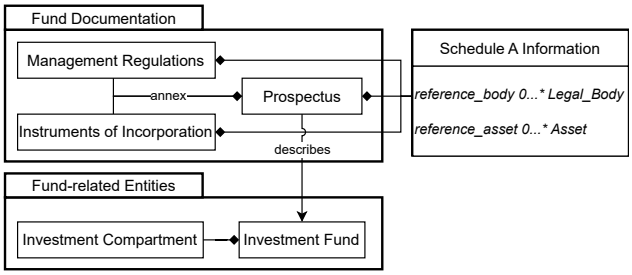

Fig. 4. Packages in the model of Article 69 of the EU UCITS Directive

C. Threats to Validity

The main threat to the validity of our process for model construction is the possibility that the experts do not fully provide their expertise either because they are unable to formalize it, or because they misunderstand the logical implications of the modeling choices. To mitigate this, we designed the activities (document exchange, QA sessions, ...) in a way that maximizes the possibility to elicit clear design choices while at the same time avoiding ambiguities (e.g., the use of direct questions such as the examples presented in the paper).

VII. The Model and its Use to Support Compliance Checking

A. The Consolidated Version of the Model

At the end of the process described above we reached a consolidated version of the model, which is composed of three packages: the Model of Fund Documentation, the Model of Schedule A Information, the Model of Fund-Related Entities. The Model of Fund Documentation is composed of three classes. The Model of Schedule A Information has a total of 28 main classes which, when adding the subclasses and the (meaningful) combinations of attribute values, allows one to identify 152 information types. The model of fund-related entities is composed of 25 classes. An overview of the packages is provided in Figure 4, while a detailed presentation of the model, together with the design criteria, is contained in the online appendix [5]. As already noted in Section II-A, the basic model for content requirements includes the concept of the "regulated document" (Fund Documentation) and the concept of the "required content" (Schedule A Information) that should be contained in the document for it to be considered complete (and, from a legal point of view, compliant). The general structure of the content model consists of the concept of a document having a containment relation with information types, which are in turn organized in a taxonomy. This structure is dictated by the type of task (supporting compliance with requirements of contents) as seen in similar work [11]. The complete model is composed of three packages:

1) **The model of fund documentation:** The model of fund documentation is a simple model representing the documents

issued by investment funds. The model is also used as a bridge between the other two models. Note that the model (entirely represented in Figure 4) is composed of three entities (*Prospectus*, *Management Regulations*, and *Instruments of Incorporation*) because the complete requirement of Article 69 of the EU UCITS Directive (see Section V-A) indicates all three documents as possible containers of the required information.

2) **The content model of Schedule A information:** It contains the information types that allow checking the compliance of the text contained in fund documentation.

3) **The domain model of fund-related entities:** The purpose of the model of fund-related entities is to represent the entities that are relevant to the domain of UCITS Funds. This enables the specification of application conditions, for those requirements that are conditional on the presence of some circumstances in the UCITS fund that is publishing the prospectus.

B. Using the Model to Support Compliance Checking

To explain how the model can be used to express compliance criteria, let us take the example of point 1.3 of Schedule A: according to the regulation, "if a fund has different compartments, the prospectus (or annexed management regulations or instruments of incorporation) must contain an indication of those compartments". The subset of the model that is relevant for this example is shown in Figure 5.

A simple representation of this requirement is possible by searching for the presence of the information type "Indication of Compartments", but this would not allow for ACC, as the lack of indication of compartments does not, by itself, imply a breach: in order to assess that, it is necessary to know whether the UCITS has compartments in the first place. In situations such as this, when it is necessary to verify preconditions, the model of fund-related entities comes into play: by combining it with the other two models, it is possible to express the compliance requirement of 1.3 as follows: "if a UCITS fund X has at least one Compartment, the Prospectus that presents X must contain the Indication of Compartments".

The representation could be even more complex, taking into account the quantifiers of the legal requirement: "*for each* compartment Y of the UCITS fund X, the prospectus that presents X would have to include the indication of the compartment Y". In order to support such a representation, our model should be modified by adding an association (*indicates*) between the "indication of compartment" in the content model and the "Compartment" in the model of fund-related entities.

The complexity of the model is justified by the complexity of the subject matter. An example of such complexity is point 1.18: "Information concerning the manner, amount and calculation of remuneration payable by the common fund to the management company, the depositary or third parties [...]" The information to be provided involves three aspects (manner, amount and calculation) and three destinataries (management company, depositary, and third parties) thus resulting in nine

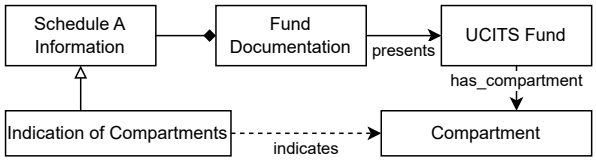

Fig. 5. Subset of model relevant for the representation of the requirement from point 1.3 of Schedule A.

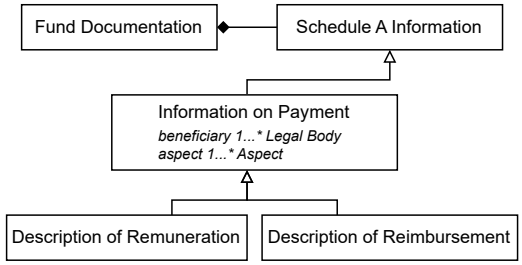

Fig. 6. Subset of model relevant for the representation of the requirement from point 1.18 of Schedule A.

possible information types, all nine of which must be accounted for in the prospectus. In the model, the aspects and destinataries are represented via attributes of the information type "Information on Payments", a super-class of "Description of Reimbursement". The model for this second example is shown in Figure 6.

VIII. Conclusions

In this paper we reported on the experience of building a model to support compliance checking of content requirements from regulations. We provided design criteria to extract and refine these requirements from the letter of law, and described a process to build and validate the model. We then presented the resulting artifacts and showed how they can be used to express content requirements.

The lessons learned provide the basis for a definition of content modeling and compliance checking in RegTech, contributing to the foundations of this young, not yet academically established discipline [6]. In that perspective, it would be ideal to further develop these lessons and this approach for the elaboration of a common ontology and interchange language, bridging financial regulatory knowledge and content models.

Ongoing work includes the validation of the model in the field, through its integration in tools for automatic extraction of information types from fund documents [35] and for model-driven compliance checking of fund documents. As part of future work, we plan to define a DSL, based on the proposed model, for expressing content requirements from the law, to enable automated compliance checking on fund documentation for any content requirement.

Acknowledgment

This research was funded in whole, or in part, by the Luxembourg National Research Fund (FNR), grant reference NCER22/IS/16570468/NCER-FT. For the purpose of open access, and in fulfillment of the obligations arising from the

Lionel Briand was partially supported by the Science Foundation Ireland grant 13/RC/2094-2, and the Canada Research Chair and Discovery Grant programs of the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] K. A. Bamberger, "Technologies of compliance: Risk and regulation in a digital age," *Texas Law Review*, vol. 88, p. 669, 2009.

[2] H. A. López, S. Debois, T. Slaats, and T. T. Hildebrandt, "Business process compliance using reference models of law," in *Fundamental Approaches to Software Engineering*. Cham: Springer International Publishing, 2020, pp. 378–399.

[3] H. M. Hart and A. M. Sacks, *The legal process: basic problems in the making and application of law*. Foundation Press, 1994.

[4] European Parliament and Council, "Directive 2009/65/EC of the European Parliament and of the Council of 13 July 2009 on the coordination of laws, regulations and administrative provisions relating to undertakings for collective investment in transferable securities (UCITS)," 2009. [Online]. Available: http://data.europa.eu/eli/dir/2009/65/oj

[5] M. Ceci, D. Bianculli, and L. Briand, "Defining a model for content requirements from the law: an experience report - online appendix," https://hdl.handle.net/10993/60775, 2024.

[6] E. Johansson, K. Sutinen, J. Lassila, V. Lang, M. Martikainen, and O. Lehner, "Regtech — a necessary tool to keep up with compliance and regulatory changes," *ACRN Journal of Finance and Risk Perspectives, Special Issue Digital Accounting*, vol. 8, pp. 71–85, 2019.

[7] G. H. von Wright, "Deontic logic," *Mind*, vol. 60, no. 237, pp. 1–15, 1951. [Online]. Available: http://www.jstor.org/stable/2251395

[8] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.

[9] A. Awad, M. Weidlich, and M. Weske, "Visually specifying compliance rules and explaining their violations for business processes," *Journal of Visual Languages & Computing*, vol. 22, no. 1, pp. 30–55, 2011, special Issue on Visual Languages and Logic. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1045926X10000716

[10] W. Tesfay, P. Hofmann, T. Nakamura, S. Kiyomoto, and J. Serna, "PrivacyGuide: towards an implementation of the EU GDPR on internet privacy policy evaluation," in *Proc. 4th ACM Int. Workshop on Security and Privacy Analytics*, 2018, pp. 15–21.

[11] D. Torre, S. Abualhaija, M. Sabetzadeh, L. C. Briand, K. Baetens, P. Goes, and S. Forastier, "An AI-assisted Approach for Checking the Completeness of Privacy Policies Against GDPR," in *28th RE Conf.*, 2020.

[12] N. Guntamukkala, R. Dara, and G. Grewal, "A machine-learning based approach for measuring the completeness of online privacy policies," in *14th ICMLA Conference*, 2015, pp. 289–294.

[13] B. Mar, "Requirements for development of software requirements," *INCOSE International Symposium*, pp. 34–39, 1994.

[14] B. Fawei, J. Z. Pan, M. Kollingbaum, and A. Z. Wyner, "A methodology for a criminal law and procedure ontology for legal question answering," in *Semantic Technology*, R. Ichise, F. Lecue, T. Kawamura, D. Zhao, S. Muggleton, and K. Kozaki, Eds. Springer, 2018.

[15] G. G. Petrova, A. F. Tuzovsky, and N. V. Aksenova, "Application of the financial industry business ontology (FIBO) for development of a financial organization ontology," *Journal of Physics: Conference Series*, vol. 803, p. 012116, 2017.

[16] G. Governatori and A. Rotolo, "A conceptually rich model of business process compliance," in *Proc. 7th Asia-Pacific Conf. on Conceptual Modelling*, 2010, Conference Proceedings, pp. 3–12.

[17] M. Hashmi and G. Governatori, "Norms modeling constructs of business process compliance management frameworks: a conceptual evaluation," *Artificial Intelligence and Law*, vol. 26, pp. 251–305, 2017.

[18] R. Hoekstra, J. Breuker, M. Di Bello, and A. Boer, "The LKIF core ontology of basic legal concepts," *International Journal of High Performance Computing Applications - IJHPCA*, vol. 1, pp. 43–63, 2007.

[19] M. Palmirani, G. Governatori, A. Rotolo, S. Tabet, H. Boley, and A. Paschke, "LegalRuleML: XML-Based Rules and Norms," in *Rule-Based Modeling and Computing on the Semantic Web*. Springer, 2011, pp. 298–312.

[20] C. Arora, M. Sabetzadeh, L. C. Briand, and F. Zimmer, "Extracting domain models from natural-language requirements: approach and industrial evaluation," in *Proceedings of MODELS'16*, 2016, pp. 250–260.

[21] N. Zeni, E. A. Seid, P. Engiel, S. Ingolfo, and J. Mylopoulos, "Building Large Models of Law with NómosT," in *Conceptual Modeling*. Cham: Springer International Publishing, 2016, pp. 233–247.

[22] I. Blums and H. Weigand, "Consolidating economic exchange ontologies for financial reporting standard setting," *Data & Knowledge Engineering*, vol. 145, p. 102148, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169023X23000083

[23] S. Abualhaija, C. Arora, A. Sleimi, and L. C. Briand, "Automated question answering for improved understanding of compliance requirements: A multi-document

study," in *2022 IEEE 30th International Requirements Engineering Conference (RE)*, 2022, pp. 39–50.

[24] E. Francesconi and G. Governatori, "Patterns for legal compliance checking in a decidable framework of linked open data," *Artificial Intelligence and Law*, 2022.

[25] J. F. Maier, C. M. Eckert, and P. John Clarkson, "Model granularity in engineering design – concepts and framework," *Design Science*, vol. 3, p. e1, 2017.

[26] N. Moloney, *EU securities and financial markets regulation*. Oxford University Press, 2014.

[27] K. Sengupta and P. Hitzler, *Web Ontology Language (OWL)*, 01 2014, pp. 2374–2378.

[28] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[29] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, B. Yin, and X. Hu, "Harnessing the power of llms in practice: A survey on chatgpt and beyond," *ACM Trans. Knowl. Discov. Data*, feb 2024, just Accepted. [Online]. Available: https://doi.org/10.1145/3649506

[30] E. Francesconi and G. Governatori, "Patterns for legal compliance checking in a decidable framework of linked open data," *Artificial Intelligence and Law*, vol. 31, no. 3, pp. 445–464, Sep. 2023.

[31] O. Amaral, M. I. Azeem, S. Abualhaija, and L. C. Briand, "Nlp-based automated compliance checking of data processing agreements against gdpr," 2022.

[32] R. Pound, "Sources and forms of law," *Notre Dame Law.*, vol. 21, p. 247, 1945.

[33] G. Soltana, E. Fourneret, M. Adedjouma, M. Sabetzadeh, and L. Briand, "Using uml for modeling procedural legal rules: Approach and a study of luxembourg's tax law," in *Model-Driven Engineering Languages and Systems*, J. Dingel, W. Schulte, I. Ramos, S. Abrahão, and E. Insfran, Eds. Cham: Springer International Publishing, 2014, pp. 450–466.

[34] L. Robaldo, F. Pacenza, J. Zangari, R. Calegari, F. Calimeri, and G. Siragusa, "Efficient compliance checking of RDF data," *J. Log. Comput.*, vol. 33, no. 8, pp. 1753–1776, 2023. [Online]. Available: https://doi.org/10.1093/logcom/exad034

[35] X. Li, D. Bianculli, and L. C. Briand, "An ai-based approach for tracing content requirements in financial documents," *CoRR*, vol. abs/2110.14960, 2021. [Online]. Available: https://arxiv.org/abs/2110.14960