A Model Toward Formalizing and Monitoring Compliance of Investment Funds Activities

Nicolas Sannier, Marcello Ceci,
Sallam Abualhaija, Domenico Bianculli
SnT - University of Luxembourg
Luxembourg
Email: firstname.lastname@uni.lu

Michael Halling
FDEF - University of Luxembourg
Luxembourg
Email: firstname.lastname@uni.lu

***Abstract*—Investment funds operate in a highly regulated sector and are thus required to act within strict constraints. Currently, compliance of fund activities is only checked through periodic inspections of the reports issued by the funds. This checking process is mostly performed manually and does not fully capture the data produced by the fund. Therefore, there is a need for providing automated solutions to support compliance monitoring of fund activities. We propose fulfilling this need using a model-driven approach. As a starting point, we create a conceptual model that formalizes the information types pertinent to compliance of fund activities. Building such a conceptual model requires analyzing multiple sources of compliance requirements, including both the applicable regulations and fund documents containing self-imposed requirements by the fund on itself. It also requires analyzing business data to understand the practice. This activity comes with various challenges mainly due to the discrepancies between the legal and business terminology. In this paper, we present our conceptual model and we further discuss the challenges we encountered during its creation. This model serves as an enabler for developing an automated support for monitoring fund activities.**

***Index Terms*—Conceptual Modeling, Legal Requirements, Regulatory Compliance, Fund Activities.**

I. Introduction

Compliance checking against regulations and technical standards is a widely-known research area in requirements engineering (RE) [1]–[6]. Regulatory compliance of systems deployed in financial markets (e.g., portfolio management software) is a novel, yet unexplored research area for requirements engineering. Financial investors, and in particular mutual funds[1] [7], not only represent an important part of financial markets, but they also play an important role in the economy as many pension systems, for example, rely on fund-based investment schemes [8].

Tracking fund activities and ensuring their compliance is a challenging task because of the traditional principal-agent conflicts in delegated portfolio management [9], where fund managers invest capital not on their own behalf but on their clients' behalf. This entails that funds must not only comply with the applicable regulations but also fulfill what they commit themselves to in their own investment policies published in their prospectuses and key information documents (KIDs). Such policies provide a contractual basis for their operations on behalf of their clients, and hence are an essential source for compliance requirements. From an RE standpoint, checking the compliance of fund activities requires *eliciting and reconciling the compliance requirements* from the applicable regulations and fund policies.

To illustrate the compliance checking process, let us consider the following simplified example. Let *fund F* be an Undertaking for Collective Investment in Transferable Securities (UCITS)[2] [10], further presented in Section II. UCITS funds are currently among the most popular investment schemes in Europe and are strictly regulated by a comprehensive legislative framework [8]. *Fund F* is therefore subject to the European Directive 2009/65/EC [11], hereafter "UCITS Directive". However, to ensure the compliance of *fund F*, requirements engineers must elicit and reconcile requirements from both the UCITS Directive as well as the investment policy of *fund F*, since the latter also contains compliance requirements that need to be met. An example from the former is Article 55, stating that *"A UCITS may acquire the units of UCITS or other collective investment undertakings [...], provided that no more than 10% of its assets are invested in units of a single UCITS or other collective investment undertaking. [...]"*. This requirement indicates the possibility for *fund F* to invest in other UCITS or other collective investment undertakings, but using for each investment no more than 10% of the assets of *fund F*. In its investment policy, *fund F* can define investment constraints on the instrument types or investment locations in which it intends to invest. For instance, in its KID, *fund F* can declare that "*The Sub-fund invests at least two-thirds of its assets in fixed income securities [...] of corporate issuers which are domiciled in Europe [...].*". Such a constraint indicates that *fund F* commits to invest a portion of its assets greater than or equal to $\frac{2}{3}$ in the instrument type "fixed income securities" from issuers that are domiciled in Europe.

The current practice of checking compliance of fund activities in regulated markets foresees periodical checks, by regularly reviewing snapshots of fund documentation and reports at

[1]A mutual fund is an investment option where money from many people is pooled together to buy a variety of stocks, bonds, or other securities

[2]As the name reveals, a UCITS is an undertaking for collective investment (or "investment fund") which invests in securities, e.g., stocks, bonds, short term treasury instruments and cash.

different timestamps. This practice has two limitations. First, it requires intensive manual involvement. Second, it does not identify (either at an online or offline stage [12]) deviations or breaches in the actual data produced by the funds that is not detailed in periodic reports. The need for providing automated assistance in monitoring fund activities has been acknowledged by Ceci et al. [13]. In this paper, we propose utilizing model-driven methods for eliciting and formalizing the compliance-relevant information from the applicable regulations and the fund policies. Specifically, we build a comprehensive conceptual model that characterizes the information types pertinent to fund compliance. Our work presents a step toward developing such an automated assistance.

To develop a model that is both comprehensive and applicable in practice, we analyze fund data in addition to the regulations and fund policies. The goal is to better understand the practice on how funds operate and what data they produce. This goal can be complex and challenging to achieve because of the significant gap between the terminology used in the documentation (regulations and fund policies) versus that used by the fund in practice.

Elicitation and specification of legal requirements has long been investigated in the RE literature. Most of the research focuses on security and privacy [14], [15] and personal data protection [6], [16]. Compared with existing work, we reconcile the diverse knowledge elicited from multiple sources of compliance requirements into a comprehensive conceptual model. Our model enables an end-to-end compliance checking of fund data both against regulations as well as against fund policies. We further shed light on the challenges for using modeling methods in complex domains such as the finance domain.

Contributions: The paper makes the following contributions:

(1) We develop a conceptual model made of 75 main information types, describing in a comprehensive manner the information types pertinent to compliance of fund activities, derived from multiple sources including regulations, policies, and available fund data. The model has been built through several iterations with the involvement of researchers having diverse expertise including software engineering, requirements engineering, legal informatics, and finance. This model represents a stepping stone toward devising an automated assistance for monitoring fund activities. Our work covers mainly the UCITS universe, the largest fund type in Europe.

(2) We discuss a set of pressing challenges concerning the creation of such a comprehensive model from diverse sources in the context of a complex domain like finance. We further provide recommendations on addressing the challenges in similar contexts.

Data Availability. We release the model and all other material relevant to our modeling activity in an online annex [17].

Structure. The remainder of the paper is organized as follows. Section II provides background information on fund activities and the context in which our work is set. Section III introduces our methodology for constructing the conceptual model. Section IV presents the model and how it can be potentially leveraged toward devising an automated compliance monitoring approach. The section further discusses the challenges emerged during its design. Section V discusses the threats to validity. Section VI positions our contribution against the related work. Section VII concludes the paper and outlines future work.

II. Background and Context

Our work focuses on *Undertakings for Collective Investments in Transferable Securities* funds (in short, UCITS) [11]. UCITS are economically highly relevant funds as well as popular investment schemes accounting for the majority of investments made in the European Union (EU) [8].

UCITS fund activities, for the purpose of our work, can be roughly summarized as follows. Once registered in a regulated market which is approved by the market regulator, a UCITS fund collects investments from investors via subscription of shares. The fund then invests the collected capital, on behalf of its clients (i.e., investors), according to the investment strategies outlined in the fund documents and in accordance with legal requirements. Investments options might vary according to the financial products targeted (e.g., equity versus fixed income fund), the risk appetite of its investors (e.g., aggressive versus defensive strategies), and the investment policy or style of the fund (e.g., value versus growth funds). Investors can redeem their shares daily, i.e., they can (ask to) get their investment back when they deem the moment to be appropriate, or at the time of the liquidation of the fund at the latest. The business model of the mutual fund industry relies on charging the investors fees on a regular basis (e.g., annually). These fees are usually expressed in percentage terms and are based on the assets under management. In addition to these fees, funds sometime charge additional costs that the investors have to pay when investing capital into or withdrawing capital from funds.

Several activities of UCITS funds are subject to compliance. For instance, the investments of the fund must be checked for compliance. Investments are typically recorded as they occur in dedicated transaction records, including details such as date, parties, types of instruments, and amounts. Other activities like the fees paid by investors, or the fund holdings composition or portfolio composition, i.e., the collection of the fund's financial investments (e.g., stocks, bonds, commodities, cash) are also subject to compliance. Driven by data availability, we focus in this paper on *fund holding composition*, a complex and constantly evolving activity due to several factors, e.g., risks, or market fluctuations [18]. Information on fund holding composition is usually updated on a daily basis and is freely available (although in aggregated form) through financial information providers such as Thomson Reuters[3] or MorningStar[4]. Other activities are typically not available outside the management company and are thus not usable in our context.

[3]https://www.reuters.com/markets/global-market-data/

[4]https://www.morningstar.com/

III. Eliciting Compliance Monitoring Information

In this section, we report on the qualitative study that we conducted to answer the following research question: **Which information types are pertinent to monitoring fund activities?**

We conducted the study over selected portions of the law as well as different sources of business data, including fund documents and financial data on holding composition. The study involved researchers with different areas of expertise, including requirements engineering, model-driven engineering, legal informatics, and finance. The outcome of the study was a conceptual model representing the different information regarding fund holding composition.

In the remainder of this section, we present the qualitative study design and illustrate the resulting outcomes and observations.

A. Qualitative Study Design

Study material. To cover both the regulatory and the business aspects of UCITS funds investments, we used the following source material for the study:

1) The EU Directive 2009/65/EC of the European Parliament and 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) — in short, Directive 2009/65/EC or the UCITS directive [11].
2) A proprietary dataset from a major financial information provider, which provides monthly aggregated information about fund holding compositions from September 2012 to June 2018. The dataset provides this information at a worldwide level, containing thereby financial transactions in major regulated markets, e.g., in the US as well as the EU.For our analysis, the last author of the paper (an expert in finance) selected the dataset that focuses on one of the largest worldwide UCITS funds which has 26 individual subfunds, and performs a substantial fraction of its investment globally (e.g., in the US). Although the data is related to only one fund, analyzing one of the largest funds enables us to obtain a large coverage of the domain.
3) A total set of 30 Key Information Documents (KIDs) from various UCITS funds collected during the period July–August 2023. While analyzing prospectuses would have been the straightforward choice in our context, we opted for analyzing KIDs. Despite being more complete, prospectuses are more complex to analyze since they tend to paraphrase the legal requirements in addition to describing each subfund's investment policies. KIDs of individual subfunds, on the other hand, are easier to analyze due to their specific format, and they contain only the subfund's specific investment policy.

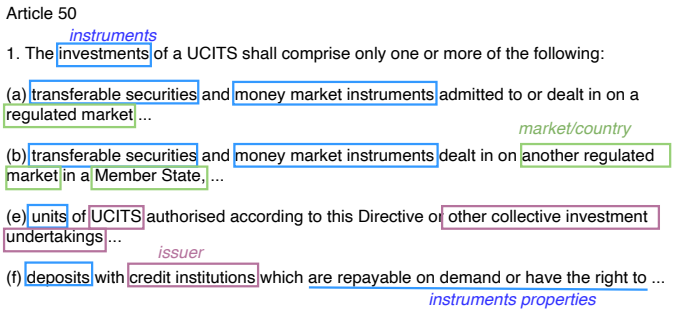

Fig. 1: Excerpt of the UCITS Directive

We remark that, even if the proprietary dataset and the KIDs included in our source material[5] refer to funds operated long after the issue date of the UCITS Directive, such a directive is still in force. We further remark that these timing considerations do not pose a threat to the development of the model. However, for the concrete monitoring of fund activities, funds data (information on trades and holdings and information on their investment policy) will have to be consistently aligned.

Methodology The study was aimed at creating a conceptual model, and was conducted in three iterations, one for analyzing each category of source material. During the first iteration, we focused on the legal aspects of UCITS funds portfolio composition in the UCITS Directive [11]. Specifically, we focused on Article 50 on authorized investments, which contains the list of instruments and the conditions for a fund to be able to invest. During the next two iterations, we analyzed data from the domain. As part of the second iteration, we analyzed the dataset, while in the third iteration, we analyzed the investment policies in the KIDs. In each iteration, we refined the conceptual model with complementary information.

Overall, during each iteration, the first author (with expertise in legal requirements engineering and model-driven engineering) and the second one (with expertise in legal analysis and legal informatics) elicited the relevant concepts from the material and prepared draft versions or revisions of the conceptual model through several bilateral sessions, cumulatively adding up to approximately 120 hours of work. The elements of interest and the model itself were further discussed and refined among the remaining authors during four collective workshops of 1.5 hours as well as an additional specific off-line validation session between the first and the last author, expert in finance, regarding the specializations of the concepts in the model.

B. Observations and outcomes.

First iteration: Analyzing the UCITS Directive. Fig. 1 shows an excerpt of Article 50 of the UCITS Directive, where we highlighted some important concepts regarding the fund

[5]We acknowledge that newer, specialized financial instruments could be referred to in data sources containing more recent information/documents, not included in our source material. Nevertheless, such instruments would still need to meet the UCITS Directive requirements.

holding composition. In this excerpt, the legislator considers as authorized investments the following financial instruments (highlighted in blue in the figure): investments in transferable securities, money market instruments, units of other UCITS, and units of other collective investment undertakings or deposits. These instruments may have some constraints on the markets where they were sold/bought (e.g., in regulated markets), as well as the countries in which they were sold (e.g., EU Member States). these constraints are highlighted in green in the example. Other constraints (in purple in the example) apply to the issuers of the instruments (UCITS, credit institutions, or generic "companies"). The legislator has also imposed additional constraints on the instruments (e.g., the possibility for deposits to be repayable on demand) and on the markets or issuers of the instrument (underlined in blue in the example).

Overall, during the first iteration on the conceptual model we articulated all the concepts retrieved in Article 50 around four main categories: (C1) the types of investments (financial instruments), the types of (C2) markets and (C3) countries in which these investments are performed, and (C4) information about the issuers of the shares owned by the UCITS. We noticed that the additional properties on instruments, issuers, or markets that were elicited from the regulations are not *necessarily observable*. Examples of such properties include transactions performed over the counter, authorized institutions, the level of supervision of the markets, and the open characteristics of markets. Such information might not appear in the real data produced by the fund; this may introduce challenges to the compliance monitoring process. Thus, assessing the observability of concepts was the main drive for conducting the second iteration, focusing on analyzing business data.

Second iteration: Analyzing the financial information provider dataset. This dataset consists of a table with 57 fields covering different information types (e.g., identifiers, countries where the shares are available for selling, type of the holding, issuing companies, amount of shares and their values) of the fund's individual investments at a certain timestamp. We disregarded all the information types that either have an empty value in the table or are related to specific properties that are not relevant in our context (e.g., the eligibility of the instrument to some US regulation is irrelevant since our study is grounded in the European regulatory framework).

The remaining relevant fields in the dataset include four main groups: (F1) fields related to the identification of the issuer in various marketplaces[6] (e.g., holding security name, holding ID, holding CUSIP, ISIN, SEDOL, and ticker); (F2) the (detailed) holding type, which describes the type of instrument; (F3) the ID of the country where the holding was bought. In addition, some fields such as (F4) the global sector ID and global industry ID are meant to provide information on the sector origins of the investments, which can be of interest in case the fund invests in specific industry sectors.

Regarding instruments, we noticed some discrepancies between the definitions of the authorized investments in the law, which are high-level, and those of the actual instruments mentioned in the dataset, which are business-oriented. Therefore, there is a need for mapping the instruments in the dataset and the information types derived from the law. To establish such a mapping, we extracted all the elements from the holding types and analyzed their definitions using public data such as Investopedia[7], or definitions available in financial websites.

Issuers, markets, and countries, which are key concepts elicited from the regulation (see C2–C4 mentioned above), are only barely described in the dataset. In particular, the dataset mentions the countries where the shares were available for sale and the origin country of the product. However, it does not mention the actual place (e.g., regulated marketplace, stock exchange, or over the counter) where those shares were bought by the fund. In few cases, the dataset mentions the names of the issuers without specifying its type. Further analysis of the CUSIP/ISIN numbers did not help infer additional relevant information about the issuers.

During this iteration, we extracted the content of all the fields pertaining to groups (F2), (F3), and (F4) and further refined the conceptual model by adding new subclasses, mostly on the instruments part of the model. (F1) is about identifiers, which represent additional properties in the model.

Third iteration: Analyzing UCITS funds investment policies. In the third iteration, we analyzed the investment policy provisions contained in KIDs (the other type of document containing requirements to monitor). KIDs are required to follow a mandatory structure: in our case, we are interested in the "Objectives" subsection of each KID, which describes the objectives of the fund and how it intends to achieve them through its investment policy. This description, which typically spans a couple of sentences or paragraphs, outlines the investment constraints exactly as they are stated in the law, or additional constraints specified by the sub-fund itself according to its strategy. For our study we elicited information types such as instrument type, industry sector, issuer, markets/country, as well as additional properties like investment limits, investment currency, and possible benchmark names.

Overall, we analyzed a total of 190 sentences (across the 30 KIDs) containing relevant information. Regarding instrument types (C1), we identified 150 occurrences of various instruments, most of which being (variations of) the instruments that we previously identified in our model with additional business properties, e.g., bond ratings or grades when considering bonds, liquidity characteristics, returning interest rates, and incomes. We note that most of the additional properties mentioned in the KIDs have no correspondence in the dataset; we did not include this information in our model since they will be difficult or impossible to monitor. However, one notable information, not observed in the first two iterations but retrieved in KIDs, is related to the notion of "convertibles", which are securities or bonds that can be converted at a certain

[6]https://www.investopedia.com/terms/i/isin.asp

[7]https://www.investopedia.com/financial-term-dictionary-4769738

date into other types of assets. Convertibles were added to the model as an attribute of the instrument.

As for issuers (C4), most of the sentences (163/190) do not explicitly mention any kind of issuer, while the remaining 27 only mention generic types (e.g., companies, corporate, governmental entities) or complement information with general quantifiers (e.g., European companies, small/large companies). Consequently, we did not add new information types related to issuers to the conceptual model.

Regarding the market and country (categories (C2) and (C3) mentioned above), we note that 146/190 sentences do not contain any geographical information or notions of market. Out of the remaining 44 sentences, 38 of them provide some shallow information: 11 sentences mention "worldwide" while the remaining 27 mention continental regions or large regional areas (e.g., "the EU", "Asia excluding Japan", "emergent countries"). Moreover, only 6 sentences propose limitations related to the country-specific objectives of the fund.

We did not observe any description of industry sectors (F4 in the dataset resulting from the second iteration) in the analyzed policies. However, we are aware of some different KIDs (not considered in this study) that contain some remarks about the industry sectors; we speculate that indication of sectors would be retrieved for more specialized funds[8], different from those considered in our study (which have general orientations and are not targeting assets of specific sectors).

At the end of the third iteration, we only added to the model the information type corresponding to the notion of convertibles.

In summary, we created our conceptual model over three iterations related to different aspects of the domain modeling as well as requirements elicitation. The first iteration from the law yielded a model that is generic and universal (i.e., applicable to all possible UCITS funds) but not aligned with the concrete business practice and terminology. The second and third iterations complemented the initial model, resulting in a more specific and practical representation, integrating business language and specifying the information types that are introduced in the law from a practical standpoint.

IV. Modeling UCITS Funds Holding Composition

In this section, we first present the conceptual model and discuss its usage toward automated compliance monitoring. We then discuss the challenges we encountered while building the model.

A. The Conceptual Model and its Usage

The outputs and observations from each iteration described in Section III contributed to the design and adaptation of our final conceptual model, which captures the information types related to UCITS funds' holding composition and that are required for their compliance monitoring. Due to the size of the model, we show in Fig. 2 an excerpt of our conceptual model; the complete model is provided in an online annex [17].

[8]We did not analyze the characteristics of the funds when selecting the documents at the beginning of the analysis.

In total, the model consists of 75 concepts coming from the legal source (shown as white boxes in the figure) elicited during the first iteration and additions (shown as light blue boxes) derived from the other sources during the second and third iterations.

As discussed earlier, the conceptual model is centered around the following main concepts, which have been enriched with information from the UCITS directive, the fund data, and the investment policies:

- *Investment* is a wide category including 52 subcategories (not all of them are shown in Fig. 2) the most important being "financial instrument";
- *Issuer* identifies the company or institution that issues a financial instrument;
- *Country* identifies the location of the issuer or of a market;
- *Trading Venue* identifies the place where an instrument is traded.

Such a conceptual model can be leveraged in the various tasks required for developing automated compliance monitoring of UCITS fund activities.

(1) Extraction of monitoring information. This task aims at (automatically) extracting compliance-relevant information from fund documents which will then be used to monitor the fund activities. Our conceptual model is the basis for this activity as it lists all potential information types that are required to be extracted from the fund policies. To this end, we note that the goal of this activity is to identify the concrete values of these information types, e.g., the types of investments and their ratios that are defined in the fund policy. The initial annotations made on the 30 investment policies analyzed during our qualitative study enable the design of an automated information extraction solution. There is substantial research work in the RE literature on metadata extraction from legal texts [19], [20] or privacy policies [15]. One can apply similar approaches or utilize recent artificial intelligence (AI) technologies, notably the recent natural language processing technologies featured by zero-shot learning approaches [21] and large language models [22].

(2) Specification of compliance requirements. This task aims at specifying formal compliance requirements for monitoring fund activities. This activity can be performed through devising a specification language that (i) is understandable by domain experts who would define the compliance requirements, and (ii) can be used to formalize the requirements. To build such a language, our conceptual model in combination with the results from the first activity are essential to define compliance requirements from different sources, and develop an end-to-end solution that provides automated support for specifying formal requirements. This activity will highly benefit from the observation that emerged during our qualitative study regarding the inter-dependencies between the requirements originated from the law and those originated from the fund policies. Such a specification language can follow (tailored) patterns like the ones proposed by Rupp et al [23] and

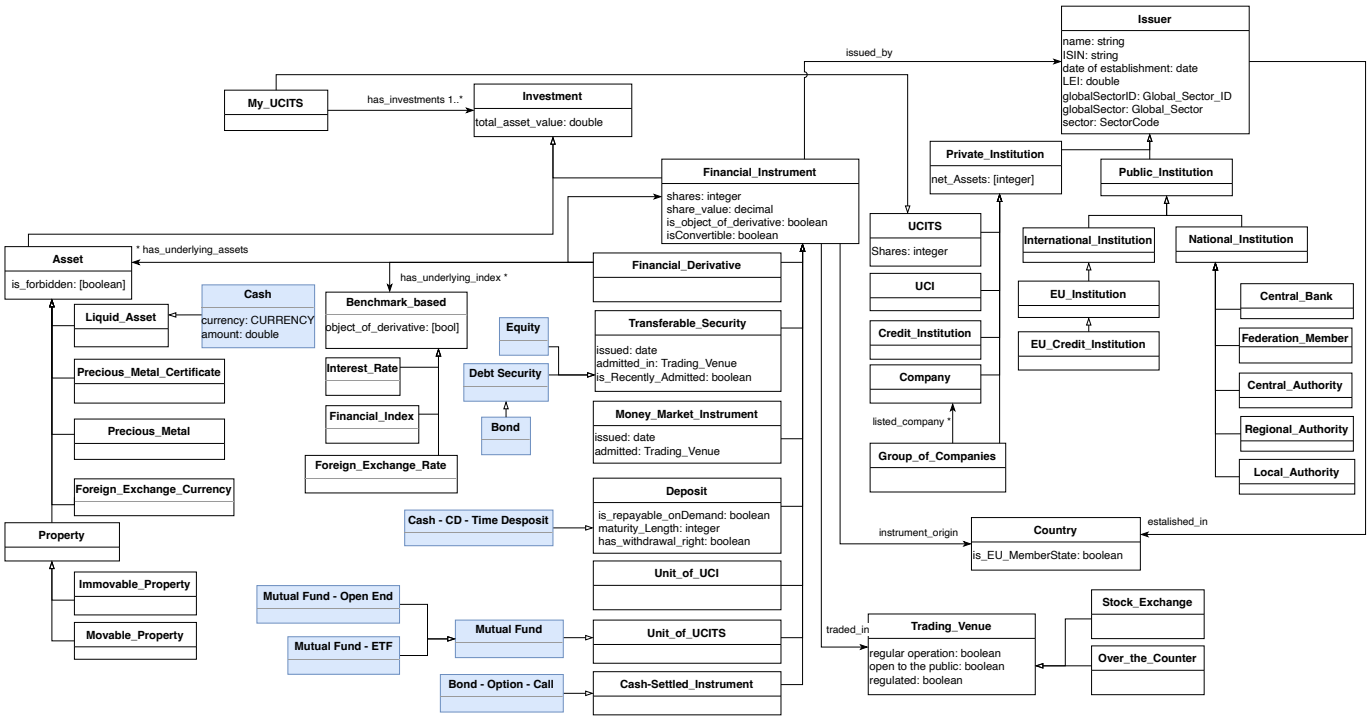

Fig. 2: Excerpt of the Conceptual Model for Fund Holding Composition

EARS [24] which were further used by Arora et al [25] and Veizaga et al [26]. Alternatively, one can investigate constrained natural languages such as that proposed by Konrad and Cheng [27]. In this case, the language should be adapted to fit in the context of fund composition. Another alternative is to develop a specification language similar to RELAX, proposed by Whittle et al [28] in the context of dynamic adaptive systems or a domain specific language such as CDSL proposed by Rabiser et al [29] for monitoring systems of systems.

(3) Automated Compliance Monitoring. By combining the outcomes of the previous tasks, it becomes possible to automatically perform compliance monitoring of fund activities on the actual data produced by the fund. This activity involves translating the monitoring requirements (formalized in the second task) into executable assertions that can be evaluated against monitoring data. Furthermore, monitoring data might potentially come from multiple providers and thus may also require to be reconciled before being checked.

The aforementioned tasks must account for challenges that might emerge from the analysis of fund documents. For instance, one needs to distinguish in the fund policies between the constraints that come from the law, and have been replicated in the investment policy, and those that are specific to the fund. Similarly, the developed automated approaches will be challenged by possible vague expressions found in the policies. For example, the fund objective stating that "the fund will mostly invest in the equity universe worldwide" contains the vague term "mostly" that is difficult to interpret, while the other terms ("worldwide", "equities") are processable. In such cases, integrating heuristics from the domain could be potentially beneficial in better interpreting vague terms. Otherwise, some requirement would be marked as not verifiable.

B. Challenges when Building the Model

During the construction of the model, we encountered the following challenges:

Dealing with tacit domain knowledge. The languages used in the law and in fund documents do not always align consistently. For example, the concepts of "investment", "asset", and "property" define (virtual) goods by giving them a role in relation to another entity, i.e., the owner. As such, these classes may overlap. For example, the class "Asset" overlaps with the class "Financial instrument", which by itself is just an instrument, but insofar as it is "owned" by a fund, it is considered an investment and also an asset of that fund.

Moreover, the terminology used in the law differs from that used both in KIDs and in the dataset. This is also related to the fact that funds, while being established in one country (in Europe for UCITS funds) operate worldwide and are thus subject to other regulatory requirements in addition to those issued by the EU Institutions and would use terminology appropriated to the places where they operate. For example, the term "mutual fund", which corresponds to a class of collective undertakings, is the US equivalent to the European "UCITS", but the two legal bases are slightly different. Therefore, there might be discrepancies in fund denomination and classifications while looking at the data; for example, "other undertaking for collective investments" might be used instead of "other UCITS", because funds can be categorized differently across different jurisdictions using different criteria.

Furthermore, in some cases, a one-to-one mapping of the terminology is not possible. For example, a financial instrument might have complex characteristics, making its classification not straightforward. Examples of such corner cases involve the maturity date of some investments, i.e., the period after which it is recommended to sell the investment, or the place it was traded (e.g., security market or money market), which can affect the classification of an instrument as "security" or "money market instrument". Instruments such as bonds could initially be considered as "money market instruments" but only when their maturity date is short (usually, a maturity of one year or less), or "securities" when their maturity/term date is longer.

We were able to tackle this challenge thanks to the interdisciplinary expertise of the research team, encompassing experts in requirements engineering, model-driven engineering, legal informatics, and finance. The team had regular meetings dedicated to reconciling the terminology among the different sources, incorporating the (tacit) domain knowledge provided by the subject-matter expert.

Generalizing from the EU specific context. Since we analyzed the EU legislation, the legal terminology was initially EU-oriented. In particular, the EU legislation considers its jurisdiction as its "local market" while considering other regions as "third-countries". However, as stated in Section II, funds considered as "UCITS" by EU legislation can invest worldwide and use terminology adapted to the markets where they operate (a circumstance which can also be observed in the dataset). Instead, the dataset provides data at a global level. Consequently, we had to refine the model to enable the possibility to account for regulations and terminology from other jurisdictions by removing EU-specific information, e.g., distinction between (EU member states) countries.

Trade-offs between completeness of the model and observability. While building the conceptual model and analyzing business data, we mentioned additional properties that could be attached to instruments, issuers, or markets. Such additional properties, which express essential characteristics, might however not always be observable or verifiable in the data. For instance, "recently issued" transferable securities are from issuers whose shares should be fully registered within the year. Such information might not be visible in the data. Similarly, some investment funds tend to remain vague when describing their investment policies, as described previously. Observability [30] and vagueness issues, as emerged during our analysis, can impact the specification of complete and precise compliance requirements. When building the model, we decided to favor the completeness of the model, by including these properties independently from their degree of observability.

V. Threats to Validity

Internal validity. The main threat to internal validity is related to the interpretation of the source materials performed by the authors to create the conceptual model. This threat is mitigated since our work in this paper, including the creation of the conceptual model, is interdisciplinary. Specifically, the conceptual model was created in an iterative manner by three different researchers who have diverse experience, with more than ten years experience in RE and modeling, background in legal informatics, and expertise in finance. During each iteration, the researchers had several online sessions to discuss their interpretations and define the domain-specific terminology. Our conceptual model is further made publicly available and is thus open to scrutiny. Another threat is related to the completeness of our conceptual model. This threat is reduced since the study material consisted of documents from both the regulatory and business sides, complemented with a dataset from a major financial information provider. Further studies with a broader spectrum of regulations and fund documents are nonetheless beneficial for improving the validation of our observations and the resulting model.

External validity. The main threat to external validity is related to the generalizability of our observations since we conducted our study on a specific case study targeting UCITS funds. We note that extending the analysis and the model to other types of funds, e.g., alternative investment funds, would require analyzing, using a similar methodology, different legislative corpora, possibly from various jurisdictions that would need to be aligned [31]. Such an extension would lead to expanding the knowledge that we have presented in our model, while not invalidating the current content or the observations regarding the challenges that we encountered. We further note that our work is representative since UCITS funds cover a substantial fraction of the European regulated markets activities [8]. That said, further studies on the application of our proposed model are necessary to address this threat and provide more confidence in the generalizability of our observations.

VI. Related Work

Domain modeling is a well-known activity in (model-based) software engineering. Using model-based approaches is typically beneficial for addressing complex domains with sensitive, safety, or regulatory compliance-related concerns. Conceptual models and metamodels have been proposed in the literature to describe (large) domains such as the general data protection regulation (GDPR) [32], safety standards [33], [34], or smaller domains such as a subset of Luxembourg's tax law [35] and the domain model for satellite systems [36].

Approaches for automating the construction of domain models have been also proposed using various technologies, including NLP [37], [38], machine learning [39], and more recently large language models [40]. Despite providing promising results, automated domain modeling has limitations regarding the completeness and complexity of the resulting models, as well as the need for human analysts to guide or correct these models. In addition, such automated approaches have been tested in a specific context on limited material, e.g., requirements specifications or problem descriptions. Not only are these documents simpler and more structured than those from the finance domain, but they also do not capture

the complexity of utilizing various sources (e.g., regulations and fund documents) to build a consistent and homogeneous model. Given the complexity of financial regulations and the discrepancies in terminology used in investment policies, we opted for manually building our proposed model.

Requirements monitoring has been extensively studied for quite some time by the RE (and more largely the software engineering) community [41], [42], focusing primarily on the verification of functional and non-functional requirements of the monitored systems, including events or data coming from other systems in the case of systems of systems [43]. In contrast, in this work we investigate building a system-to-be to serve as a compliance checker aiming at specifying compliance requirements and monitoring data or events that are external to the system. In relation with monitoring as discussed in the business and finance domain, existing work focuses on monitoring global enterprise organizations and processes [44]. The approaches in this research direction follow state-of-the-art business methodologies, such as GRC (governance, risk, and compliance) and further rely on standards such as the Semantics of Business Vocabulary and Business Rules (SBVR) [45] for formalizing the domain knowledge or documentations. Regarding compliance of data produced by funds, the work by Roychoudhury et al. [46], [47] is the closest to our context and research problem. The authors investigated the use of controlled natural language (SBVR' Structured English) for formalizing legal provisions into SBVR models. Specifically, they use machine learning and NLP in order to extract vocabulary elements and facts, as well as support the semi-automated specifications of regulatory rules and the translation of the domain into formal logic representation (e.g., DROOLS or Prolog). Compliance checking is then performed by checking facts against the generated rules. Their work is based on the Money Market Statistical Reporting (MMSR) regulation [48], which regulates the daily reporting of financial transactions produced by financial institutions on the money markets. The authors further analyzed transaction logs based on rules on the log entries fields. In contrast to their work, we created our model by analyzing multiple sources of requirements from which our compliance requirements are also originated. Instead of utilizing general-purpose structured English, we plan to develop a domain-specific language that is more tailored to our specification needs. In addition, we aim to specify requirements whose evaluation would be more dynamic, since investment policies vary from one fund to another and fund composition can vary due to the simple fluctuation of the markets.

VII. Conclusion

In this paper, we proposed a conceptual model created from diverse sources including applicable regulations, fund policies, and fund data. We further discussed the challenges emerged during building the model, such as the terminology gap between the different sources and discrepancies between different legal frameworks added to variations in business practices. We also discussed tasks that we envision for developing automated compliance monitoring by drawing on our conceptual model.

In the future, we plan to devise an end-to-end automated approach for monitoring compliance of fund activities by conducting the following tasks: (1) automated extraction of monitoring information from fund policies according to our conceptual model; (2) specification of formal compliance requirements that should be monitored; (3) integrating these approaches into an automated compliance monitoring solution. Completing such tasks will allow us to validate (and refine) our model in concrete settings.

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 grant agreement, the author has applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

References

[1] T. D. Breaux, M. W. Vail, and A. I. Antón, "Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations," in *RE'06, 14th IEEE International Requirements Engineering Conference*, pp. 46–55, 2006.

[2] S. Ghanavati, D. Amyot, and L. Peyton, "Compliance analysis based on a goal-oriented requirement language evaluation methodology," in *17th IEEE International Requirements Engineering Conference*, 2009.

[3] S. Ghanavati, A. Rifaut, E. Dubois, and D. Amyot, "Goal-oriented compliance with multiple regulations," in *22nd IEEE International Requirements Engineering Conference*, 2014.

[4] N. Zeni, N. Kiyavitskaya, L. Mich, J. R. Cordy, and J. Mylopoulos, "Gaiust: supporting the extraction of rights and obligations for regulatory compliance," *Requirements Engineering*, vol. 20, 2015.

[5] O. Akhigbe, D. Amyot, and G. Richards, "A systematic literature mapping of goal and non-goal modelling methods for legal and regulatory compliance," *Requirements Engineering*, vol. 24, no. 4, 2019.

[6] O. Amaral, S. Abualhaija, M. Sabetzadeh, and L. C. Briand, "A model-based conceptualization of requirements for compliance checking of data processing against GDPR," in *29th IEEE International Requirements Engineering Conference Workshops, RE 2021 Workshops*, pp. 16–20, 2021.

[7] Investopedia, "Mutual Fund - Definition," 04 2024. https://www.investopedia.com/terms/m/mutualfund.asp.

[8] The European Commission, "Undertakings for collective investment in transferable securities – amended Directive (UCITS V): Frequently asked questions." https://ec.europa.eu/commission/presscorner/detail/en/MEMO_14_298, 04 2014.

[9] M. C. Jensen and W. H. Meckling, "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics*, vol. 3, no. 4, pp. 305–360, 1976.

[10] Association of the Luxembourg Fund Industry (ALFI), "UCITS - Definition," 04 2024. https://www.alfi.lu/en-gb/understandinginvesting/post/what-is-a-ucits.

[11] The European Parliament and the Council of the European Union, "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)." https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009L0065-20140917, 07 2009.

[12] Y. Falcone, S. Krstic, G. Reger, and D. Traytel, "A taxonomy for classifying runtime verification tools," *Int. J. Softw. Tools Technol. Transf.*, vol. 23, no. 2, pp. 255–284, 2021.

[13] M. Ceci, N. Sannier, S. Abualhaija, D. Shin, D. Bianculli, and M. Halling, "Toward automated compliance checking of fund activities using runtime verification techniques," in *IEEE/ACM International Workshop on Software Engineering Challenges in Financial Firms (FinanSE), Lisbon, Portgual, April 16, 2024*, 2024.

[14] T. Breaux and A. Antón, "Analyzing regulatory rules for privacy and security requirements," *IEEE Transactions on Software Engineering*, vol. 34, no. 1, pp. 5–20, 2008.

[15] M. B. Hosseini, T. D. Breaux, R. Slavin, J. Niu, and X. Wang, "Analyzing privacy policies through syntax-driven semantic analysis of information types," *Inf. Softw. Technol.*, vol. 138, p. 106608, 2021.

[16] 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 *RE'20, 28th IEEE International Requirements Engineering Conference*, pp. 136–146, 2020.

[17] N. Sannier, M. Ceci, S. Abualhaija, D. Bianculli, and M. Halling, *"Online Annex (online)"*, 2024. Available at https://figshare.com/s/889e554cdd58675bd550, February 2024.

[18] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.

[19] A. Sleimi, N. Sannier, M. Sabetzadeh, L. C. Briand, M. Ceci, and J. Dann, "An automated framework for the extraction of semantic legal metadata from legal texts," *Empir. Softw. Eng.*, vol. 26, no. 3, p. 43, 2021.

[20] N. Zeni, L. Mich, and J. Mylopoulos, "Gaiust 2.0: Evolution of a framework for annotating legal documents," in *Metadata and Semantics Research - 10th International Conference, MTSR 2016, Göttingen, Germany, November 22-25, 2016, Proceedings*, pp. 43–54, 2016.

[21] W. Alhoshan, A. Ferrari, and L. Zhao, "Zero-shot learning for requirements classification: An exploratory study," *Inf. Softw. Technol.*, vol. 159, p. 107202, 2023.

[22] 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.

[23] K. Pohl and C. Rupp, *Requirements Engineering Fundamentals - A Study Guide for the Certified Professional for Requirements Engineering Exam: Foundation Level - IREB compliant*. rockynook, 2011.

[24] A. Mavin, P. Wilkinson, A. Harwood, and M. Novak, "Easy approach to requirements syntax (EARS)," in *RE'09, 17th IEEE International Requirements Engineering Conference*, pp. 317–322, 2009.

[25] C. Arora, M. Sabetzadeh, L. C. Briand, and F. Zimmer, "Requirement boilerplates: Transition from manually-enforced to automatically-verifiable natural language patterns," in *4th IEEE International Workshop on Requirements Patterns, RePa 2014, Karlskrona, Sweden, August 26, 2014*, pp. 1–8, IEEE Computer Society, 2014.

[26] A. Veizaga, M. Alférez, D. Torre, M. Sabetzadeh, and L. C. Briand, "On systematically building a controlled natural language for functional requirements," *Empir. Softw. Eng.*, vol. 26, no. 4, p. 79, 2021.

[27] S. Konrad and B. H. C. Cheng, "Real-time specification patterns," in *27th International Conference on Software Engineering (ICSE 2005), 15-21 May 2005, St. Louis, Missouri, USA*, pp. 372–381, ACM, 2005.

[28] J. Whittle, P. Sawyer, N. Bencomo, B. H. C. Cheng, and J. Bruel, "RELAX: a language to address uncertainty in self-adaptive systems requirement," *Requir. Eng.*, vol. 15, no. 2, pp. 177–196, 2010.

[29] R. Rabiser, J. Thanhofer-Pilisch, M. Vierhauser, P. Grünbacher, and A. Egyed, "Developing and evolving a dsl-based approach for runtime monitoring of systems of systems," *Autom. Softw. Eng.*, vol. 25, no. 4, pp. 875–915, 2018.

[30] I. Gorton, L. Fong-Jones, and A. Larsson, "Observability q&a," *IEEE Softw.*, vol. 41, no. 1, pp. 50–54, 2024.

[31] D. G. Gordon and T. D. Breaux, "Reconciling multi-jurisdictional legal requirements: A case study in requirements water marking," in *2012 20th IEEE International Requirements Engineering Conference (RE), Chicago, IL, USA, September 24-28, 2012*, pp. 91–100, 2012.

[32] D. Torre, M. Alférez, G. Soltana, M. Sabetzadeh, and L. C. Briand, "Modeling data protection and privacy: application and experience with GDPR," *Softw. Syst. Model.*, vol. 20, no. 6, pp. 2071–2087, 2021.

[33] R. K. Panesar-Walawege, M. Sabetzadeh, and L. C. Briand, "Supporting the verification of compliance to safety standards via model-driven engineering: Approach, tool-support and empirical validation," *Inf. Softw. Technol.*, vol. 55, no. 5, pp. 836–864, 2013.

[34] N. Sannier and B. Baudry, "INCREMENT: A mixed MDE-IR approach for regulatory requirements modeling and analysis," in *Requirements Engineering: Foundation for Software Quality - 20th International Working Conference, REFSQ 2014, Essen, Germany, April 7-10, 2014. Proceedings*, pp. 135–151, 2014.

[35] G. Soltana, E. Fourneret, M. Adedjouma, M. Sabetzadeh, and L. C. Briand, "Using UML for modeling procedural legal rules: Approach and a study of luxembourg's tax law," in *Model-Driven Engineering Languages and Systems - 17th International Conference, MODELS 2014, Valencia, Spain, September 28 - October 3, 2014. Proceedings*, pp. 450–466, 2014.

[36] C. Arora, M. Sabetzadeh, L. C. Briand, and F. Zimmer, "Extracting domain models from natural-language requirements: approach and industrial evaluation," in *Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems, Saint-Malo, France, October 2-7, 2016* (B. Baudry and B. Combemale, eds.), pp. 250–260, ACM, 2016.

[37] C. Arora, M. Sabetzadeh, S. Nejati, and L. C. Briand, "An active learning approach for improving the accuracy of automated domain model extraction," *ACM Trans. Softw. Eng. Methodol.*, vol. 28, no. 1, pp. 4:1–4:34, 2019.

[38] L. Burgueño, R. Clarisó, S. Gérard, S. Li, and J. Cabot, "An nlp-based architecture for the autocompletion of partial domain models," in *Advanced Information Systems Engineering - 33rd International Conference, CAiSE 2021, Melbourne, VIC, Australia, June 28 - July 2, 2021, Proceedings*, vol. 12751 of *Lecture Notes in Computer Science*, pp. 91–106, Springer, 2021.

[39] R. Saini, G. Mussbacher, J. L. C. Guo, and J. Kienzle, "Automated, interactive, and traceable domain modelling empowered by artificial intelligence," *Softw. Syst. Model.*, vol. 21, no. 3, pp. 1015–1045, 2022.

[40] K. Chen, Y. Yang, B. Chen, J. A. H. López, G. Mussbacher, and D. Varró, "Automated domain modeling with large language models: A comparative study," in *26th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, MODELS 2023, Västerås, Sweden, October 1-6, 2023*, pp. 162–172, IEEE, 2023.

[41] W. N. Robinson, "A roadmap for comprehensive requirements modeling," *Computer*, vol. 43, no. 5, pp. 64–72, 2010.

[42] M. Vierhauser, R. Rabiser, and P. Grünbacher, "Requirements monitoring frameworks: A systematic review," *Inf. Softw. Technol.*, vol. 80, pp. 89–109, 2016.

[43] M. Vierhauser, R. Rabiser, P. Grünbacher, and B. Aumayr, "A requirements monitoring model for systems of systems," in *23rd IEEE International Requirements Engineering Conference, RE 2015, Ottawa, ON, Canada, August 24-28, 2015*, pp. 96–105, 2015.

[44] C. Fischer-Pauzenberger and W. S. A. Schwaiger, "Ontorea© accounting and finance model: Hedge portfolio representation of derivatives," in *PoEM'18, The Practice of Enterprise Modeling - 11th IFIP WG 8.1. Working Conference*, pp. 372–382, 2018.

[45] The Object Management Group, "Semantics of business vocabulary and business rules (sbvr)," 2006.

[46] S. Roychoudhury, S. Sunkle, D. Kholkar, and V. Kulkarni, "From natural language to SBVR model authoring using structured english for compliance checking," in *EDOC'17, 21st IEEE International Enterprise Distributed Object Computing Conference*, pp. 73–78, 2017.

[47] S. Roychoudhury, S. Sunkle, N. Choudhary, D. Kholkar, and V. Kulkarni, "A case study on modeling and validating financial regulations using (semi-) automated compliance framework," in *PoEM'18, The Practice of Enterprise Modeling - 11th IFIP WG 8.1. Working Conference*, vol. 335 of *Lecture Notes in Business Information Processing*, pp. 288–302, Springer, 2018.

[48] European Central Bank (ECB), "Regulation (EU) No 1333/2014 of the European Central Bank of 26 November 2014 concerning statistics on the money markets (ECB/2014/48)," 11 2014. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014R1333.