

Toward Better Mapping between Regulations and Operations of Enterprises Using Vocabularies and Semantic Similarity

Sagar Sunkle, Deepali Kholkar and Vinay Kulkarni

Tata Research Development and Design Center
Tata Consultancy Services
54B, Industrial Estate, Hadapsar
Pune, 411013, India

sagar.sunkle@tcs.com, deepali.kholkar@tcs.com, vinay.vkulkarni@tcs.com

Abstract. Industry governance, risk, and compliance (GRC) solutions stand to gain from various analyses offered by formal compliance checking approaches. Such adoption is made difficult by the fact that most formal approaches assume that a mapping between concepts of regulations and models of operational specifics exists. Industry solutions offer tagging mechanisms to map regulations to operational specifics; however, they are mostly semi-formal in nature and tend to rely extensively on experts. We propose to use Semantics of Business Vocabularies and Rules along with similarity measures to create an explicit mapping between concepts of regulations and models of operational specifics of the enterprise. We believe that our work-in-progress takes a step toward adapting and leveraging formal compliance checking approaches in industry GRC solutions.

Keywords: Regulatory compliance, operations, business process models, SBVR, semantic similarity.

1 Introduction

With non-compliance being penalized severely in most countries and across various business domains [1], [2] effective and efficient resolution of regulatory compliance is high on priority for modern enterprises. Industry governance, risk, and compliance (GRC) solutions help enterprises in managing regulatory compliance; however, they mostly provide content management-based, document-driven and expert-dependent ways of managing regulatory compliance. They are usually semi-formal and are not as rigorous as formal approaches to compliance checking. Formal compliance checking offers several analysis benefits. It can enhance industry GRC solutions with such functionalities as formally finding out (non-)compliance to regulations [3], [4], [5], [6], [7], [8], [9] against document-based evidence as in industry GRC solutions, computable explanation of proofs of (non-)compliance [10], [11] against expert's judgement as in industry GRC, management of frequent changes in regulations [12], [13] against functional heat maps derived from experts' knowledge as in industry GRC, etc.

Each formal approach ideally requires to *relate* regulations to operational specifics of enterprises. The realms of regulations and enterprise operations are conceptually distinct and need to be reconciled in order to be related to each other. A terminological mapping would essentially tell *where* in the operational activities a rule from the regulation becomes applicable. Surprisingly, formal compliance checking approaches implicitly assume such mapping to exist without describing how to arrive at it as also indicated in [14], [15], and [16]. If some means were provided

whereby similarity between concepts from regulations and operational specifics could be formally established, then it would be easier to *relate* concepts from regulations with operational specifics and indicate *where* a rule from regulation becomes applicable. This would also make it easier to transfer results in formal compliance checking to practical usage.

We take a step in this direction by using Semantics of Business Vocabularies and Rules (SBVR) to model vocabularies of regulations and operational specifics of enterprises. We also detail our work in progress where we map the concepts from structured SBVR-based vocabularies of regulations and operational specifics using semantic similarity measures to find out, to which operational specifics do the regulations apply and need to be checked against.

The paper is arranged as follows. We review formal compliance checking approaches, along with industry GRC taxonomy tagging approaches in Section 2 to reveal if they support mapping of regulatory and operational concepts. We also review works that identify the need for mapping in Section 2. Section 3 outlines our approach for modeling the vocabularies using SBVR and mapping the concepts using SEMILAR similarity toolkit. In Section 4 we substantiate our approach with a case study. In Section 5 we discuss the future work and in Section 6 we conclude the paper.

2 Motivation and Related Work

Several formal compliance checking approaches have been presented in literature. These approaches treat business process (BP) models as the de-facto representation of operational specifics of enterprise and check BP models for compliance against regulations. In contrast, industrial governance, risk, and compliance (GRC) solutions tend to use *taxonomy tagging* mechanisms in their content management tools. Experts trained on GRC solution platforms tag regulations and use these to search and validate documents. Compared to academic compliance approaches, GRC solutions are document/artifact-oriented and rely on enterprise data to prove compliance to regulations [17].

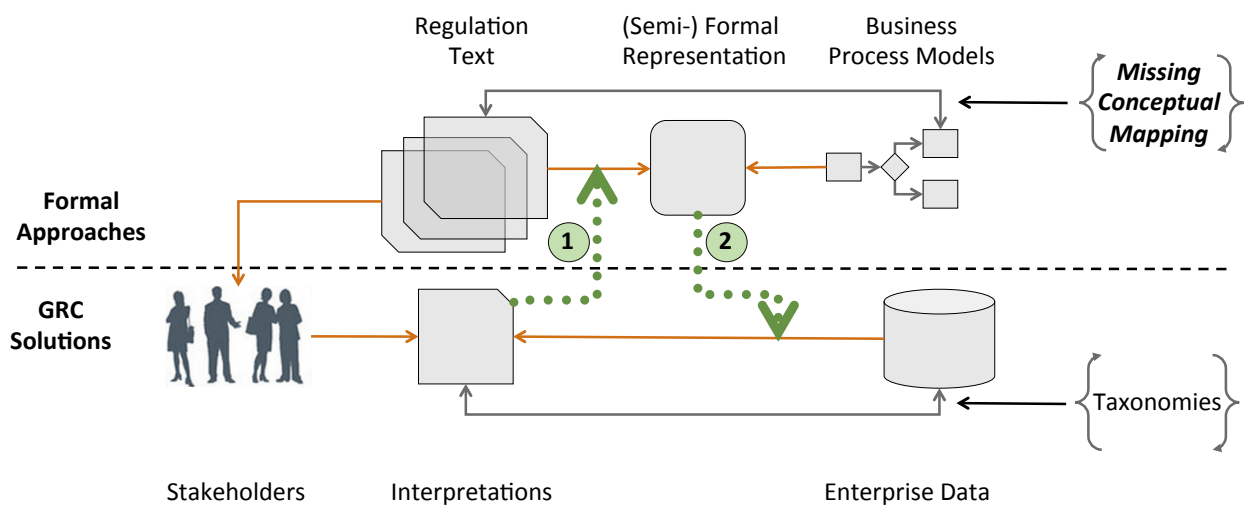


Figure 1. State of the art and practice in Compliance Management [18]

As shown above the horizontal dashed line in Figure 1, formal approaches represent both rules derived from the text of regulations and facts derived from business processes in a given formal language for compliance checking. To achieve this consistent representation of rules and facts, most formal approaches rely on implicit mapping of terminology from the two realms. On the other hand, steps followed in industrial GRC solutions are shown below the horizontal dashed line in Figure 1. Various stakeholders interpret regulations in the current context of enterprise and tag these interpretations to enterprise taxonomies. The tags indicate what enterprise data should be pulled and checked against interpreted regulations.

Two pointers are relevant in mapping regulations to operational specifics of the enterprise. ① in Figure 1 indicates that interpretations of regulation text by such stakeholders as enterprise legal advisors, compliance experts, CxO level business stakeholders, and operational managers, which are prevalent and even necessary in industry are ignored in formal approaches to a large extent [14]. Formal approaches often show direct translation of regulations to rules in the formal language used by the approach. ② in Figure 1 indicates that industrial GRC solutions hardly leverage formalisms available in research. Complexity of legal text of regulations and frequent amendments by regulatory bodies make it a demanding task to check and re-validate compliance. Formal methods to compliance checking can be very useful in industry solutions for this very reason.

In the both cases, a terminological mapping forms the first step of supporting enactment of regulations in enterprise operations. Next we present the related work in formal compliance checking and in industry GRC. Our specific aim in presenting this related work is to show how these approaches *map* concepts from regulations with concepts from models of enterprise operations, be they business process models or enterprise data/taxonomies.

2.1 Terminology Mapping in Formal Approaches

We consider five representative formal compliance checking approaches, namely, defeasible logic-based [9], [19], [20]; Petri-net based [11]; compliance rule graph-based [21], [8], [7]; extended business process modeling notation (BPMN) query and linear temporal logic (LTL)-based [3], [22], [23]; and Business Property Specification Language (BPSL) and LTL-based approaches [24], [25].

Table 1 illustrates these approaches in two columns. While the second column notes the formalism in that approach used for compliance checking, the first column shows how each approach maps labels/phrases from regulations to labels/phrases from approach-specific representation of business process models. In the following, we briefly elaborate the formal compliance checking approaches shown in Table 1 row by row concerning (a) compliance checking technique and (b) mapping between labels/phrases. As the mapping is of interest for this paper, we only briefly describe the compliance checking techniques, which the interested reader may understand from respective publications.

Defeasible Logic Approaches The first row from Table 1 shows defeasible logic-based approach for checking compliance of business process models against regulations [9].

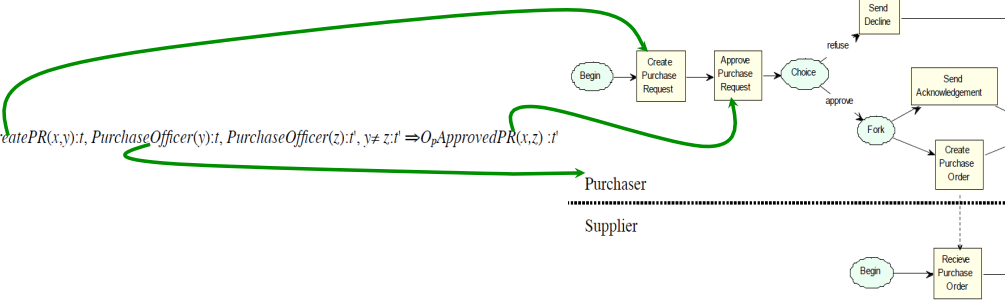
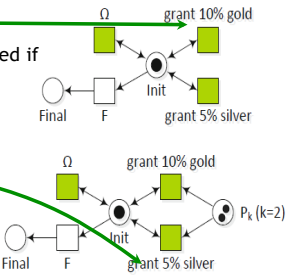
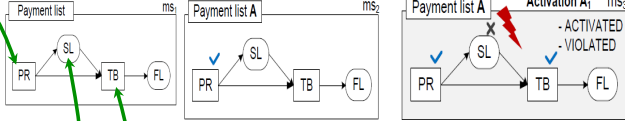
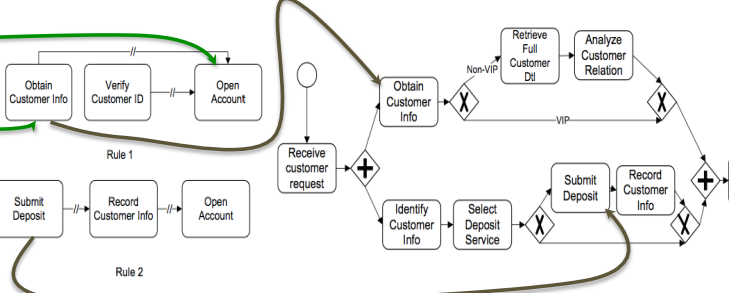
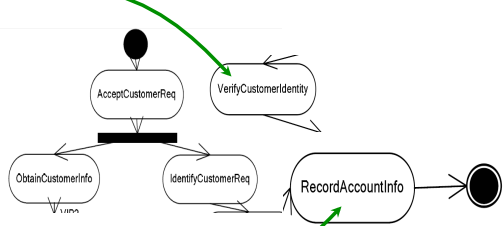
- **Compliance Checking:** Regulations are modeled in Formal Contract Language (FCL) which is a combination of efficient non-monotonic defeasible logic and deontic logic of violations [9].
- **Terminological Mapping:** First row shows a formulation of a regulation *the creation and approval of purchase requests must be undertaken by two separate purchase officers*. Labels *CreatePR* and *ApprovedPR* from FCL expression match with *Create Purchase Request* and *Approve Purchase Request* activities from business process model respectively. Label *Purchase-Officer* from FCL expression maps to *Purchaser* from business process model. It is evident that this mapping is presumed to exist implicitly in [9].

SBVR-based transformation of business rules to FCL expressions is suggested in [20] and semantic annotations of business process models in [19], but a structured terminological mapping of concepts is yet not explored.

Petri Net Approaches The second row from Table 1 shows a Petri net-based approach [11].

- **Compliance Checking:** An event log describing the observed operational behavior is aligned with a Petri-net pattern that formalizes a regulation.

Table 1. Disparity between labels in formal regulations and operational specifics [26]

Mapping between Regulations and Operational Details	Compliance Rule/Operational Formalism
<p>$c1: CreatePR(x,y):t, PurchaseOfficer(y):t, PurchaseOfficer(z):f, y \neq z: f \Rightarrow O_p ApprovedPR(x,z) :f$</p> 	<p>Formal Contract Language based on Defeasible Logics; Operational Specifics as a Business Process Model</p>
<p>Regulation: "A discount of 10% is granted if the customer is a gold customer; 5% are granted if the customer is a silver customer."</p>	
<p>Regulation: "For payment runs with amount beyond euro 10,000, the payment list has to be signed before being transferred to the bank and has to be led afterwards for later audits." + Event "payment list A is transferred to the bank"</p>	
<p>Rule 1: Before opening an account, customer information must be obtained and verified. Rule 2: Whenever a customer requests to open a deposit account, customer information must be recorded before opening the account.</p>	
<p>Article 11 part 2</p> <p>$G (action = AcceptCustomerReq \rightarrow F ((action = ObtainCustomerInfo \& Paralist = \{name, ID\}) \& F (action = VerifyCustomerIdentity \& paralist = \{name, ID\}) \& F (action = RecordCustomerInfo)))$</p>	

- **Terminological Mapping:** From the regulation shown in the second row of Table 1, phrases *a discount of 10% is granted if the customer is a gold customer* and *5% are granted if the customer is a silver customer* are mapped to phrases *grant 10% gold* and *grant 5% silver*. No explicit terminological mapping exists in this approach [11].

Compliance Rule Graph-based Approaches The third row from Table 1 shows an approach using specialized representation of regulations called compliance rule graph (CRG) [21], [8], [7].

- **Compliance Checking:** Events from operational event trace are checked against graph-based compliance rule language CRG that formalizes a regulation.
- **Terminological Mapping:** Phrases *payment runs*, *list has to be signed*, *transferred to the bank* from the regulation are presumed to match with similarly named events and are mapped to labels *PR*, *SL*, and *TB* respectively in the compliance rule graphs. No explicit terminological mapping has been suggested in [21], from which this example is taken, or other publications from same authors [8], [7].

BPMN-Q-based Approaches The fourth row from Table 1 shows an example from [3].

- **Compliance Checking:** This approach uses BPMN-Q. It is a visual language based on BPMN and used to query business process models by matching a process graph to a query graph. Visual queries labelled Rule 1 and Rule 2 in the middle indicate BPMN-Q queries adapted to expressing the regulation on the left.
- **Terminological Mapping:** Interestingly, the concepts from BPMN-Q representation of the regulation match with the business process model shown by process graph on the right. This is to be expected since BPMN-Q visual queries are based on corresponding business process models. Yet, translation of regulations to BPMN-Q queries does not preserve same concepts, for instance, phrase *customer information must be obtained* is mapped to phrase *Obtain Customer Info*. Other publications by the same authors [22], [23] similarly do not express the need for explicit mapping and presume that terminological mapping from regulation statements to BPMN-Q queries exists.

Pi Calculus-based Approaches Finally, the fifth row from Table 1 shows an example from [24] that uses Pi calculus for compliance checking.

- **Compliance Checking:** Business process models expressed in the Business Process Execution Language are transformed into Pi calculus and then into Finite State Machines. Compliance rules captured in the graphical BPSL are translated into LTL. This way, process models can be verified against these compliance rules by means of model checking technology.
- **Terminological Mapping:** The example shows that BPSL formulation of labels *RecordCustomerInfo* and *VerifyCustomerId* maps to business process labels *RecordAccountInfo* and *VerifyCustomerIdentity* respectively. This approach too does not consider an explicit terminological mapping and with several transformations between specifications, lack of explicit mapping is likely to be problematic.

Table 1 essentially shows that the most of formal compliance checking approaches *assume* that labels/phrases from regulation statements map to labels/phrases used in various regulation and business process specification languages without actually explicitly modeling them.

Next we review the related work about mapping between regulations and enterprise taxonomies in industry GRC solutions.

2.2 Terminology Mapping in Industry GRC

From industry GRC point of view, taxonomy is simply the collection of pre-defined tags, which are available for companies to “affix” to their financial data [27]. It contains facts that are defined by the elements in the taxonomy it refers to, together with their values and an explanation of the context, in which they are placed. Within a content management system, taxonomy refers to the hierarchical structure, into which content is authored, as well as the metadata elements and vocabularies created for (meta-)tagging content [28].

Industry GRC solutions offer scanning and classification facilities for enterprise content against customized check files based on regulatory compliance. Tags are specific to territories/geographies, timeframes, and business units [29]. Taxonomy tagging tools such as OpenCalais¹, Active Tags², and Compliance Guardian³, etc. employ the following kinds of techniques to build taxonomies:

- *Auto-populated taxonomies*, in which tags are extracted from and linked with content as the content is being added. This is done using natural language processing and machine learning algorithms.
- *User-defined taxonomies*, wherein users construct the classification of terms and then use them to tag content. In this approach, the taxonomies are carefully controlled by users.
- *Hybrid* approaches that support auto-populated as well as user-defined taxonomies and tagging. The auto-population feature performs verification on user-defined tags.

Industry GRC solutions may use global standards such as eXtensible Business Reporting Language (XBRL) [30]. XBRL is supposed to facilitate data tagging using XML for financial information. The tags could be drawn from pre-defined taxonomies. XBRL-based tagging can also be used to create audit trails by recording any change to a document.

It is interesting to note that whether industry GRC solutions enable tagging with proprietary or XBRL-based taxonomies, the initial tagging is manual or semi-automated with tools and largely expert-dependent. It is estimated that less than 50% of content is correctly indexed. In addition, the average cost of tagging an item ranges from \$ 4.00 to \$ 7.00. Studies regarding results of a comparison of XBRL filings for voluntary filing program of Securities and Exchange Commission have shown that the filings consisted of multiple labeling and classification errors [31]. Mostly these errors show up during initial filing, but it was argued that with experience, errors could be reduced in subsequent filings.

2.3 Approaches Using Mapping

In Sections 2.1 and 2.2, we reviewed related work in academic and industry GRC approaches in terms of support for mapping regulatory and enterprise operational concepts. We find that 1) academic approaches support formal compliance checking, but fall short in explicit modeling and mapping of concepts from regulations and operations of enterprise and 2) industry GRC approaches support taxonomy tagging to map relevant concepts from regulations and operations, but lack formal compliance checking. Next we look at those approaches, which recognize the need for explicit modeling and mapping of regulatory and operational concepts in the overall context of both academic and industry regulatory compliance and specific regulatory functionalities as enlisted below:

- *Mapping to Enact Compliance* Humberg et al. identify the need to map *situations*, i.e., one or more rules and part of the rule elements present in these rules, which may be involved persons (roles), objects or activities, to business processes in the context of CARiSMA framework [16].

¹ See OpenCalais <http://new.opencalais.com/opencalais-api/>

² See Active Tags <http://www.wavetrend.net/activ-tags.php>

³ See Guardian <http://www.avepoint.com/products/compliance-management/>

They utilize word databases to achieve this mapping via simple similarity scoring between the concepts from two realms. Compliance checking is deferred to standard model checking.

Cheng et al. present techniques to map single enterprise/domain taxonomy to multiple regulations and from multiple taxonomies to a single regulation [32]. Regulations are compared based on conceptual information as well as domain knowledge through a combination of feature matching and mapped to taxonomies. They evaluate cosine similarity, Jaccard coefficient, and market-basket analysis for similarity measurement and find that cosine similarity offers acceptable precision and recall rates among the three.

- **Mapping to Explain Proofs** An approach in [19] annotates business process models with predicates from regulations and uses them in creating status reports of violations that are treated as explanations of proofs of (non-)compliance. Other compliance checking approaches that provide some forms of proofs using diagnostic information as in [11], [10], [33] do not explicitly model concepts from either regulations or operations.
- **Mapping to Reconcile Multiple Regulations** Legal statements from different regulations may enforce the same rules, contain overlaps, or even contradict each other. The approach presented in [34] integrates the Eunomos knowledge and document management system [35] with Legal-URN framework [36]. They identify relevant regulations by generating a list of the most similar pieces of legislation in Eunomos repository using Cosine Similarity. Then, the interaction between multiple legal statements is captured using the pairwise comparison algorithm of Legal-URN. Compliance checking is itself achieved by consistent representation of regulations and business process models in goal-oriented requirements language. Modern enterprises are often subject to new regulations from one or more governing bodies, when introducing new or existing products into a different jurisdiction, or when data is transferred across political borders. To address this problem, Gordon and Breau developed a framework called *requirements water marking* that business analysts can use to align and reconcile requirements from multiple jurisdictions (municipalities, provinces, nations) to produce a single high or low standard of care [37]. They evaluated similarity measurement techniques and found that cosine similarity measures to be ideal in comparing textual legal requirements. They do not discuss mapping concepts from regulations and operations.
- **Mapping to Change-enable Compliance** To map organizational processes with applicable regulatory guidelines, Sapkotaa et al. present RP-Match framework based on regulation process similarity computation leveraging organizational process ontology [38]. The representations of the processes, regulations, and design of the validation tasks need to be changed or updated in circumstances such as (1) change/update in the existing regulations or (2) the processes need to conform to regulations from other regulatory bodies or in other territories. In such cases, mapping of the *new* regulations with the processes and validation tasks constitutes an important step towards updating the affected processes and validation tasks. The change-enabled mapping identifies processes and validation tasks affected by a change in a regulation and generates a mapping table. A domain expert then verifies the recommendations.

The related work in this section shows that modeling and mapping regulatory and operational concepts are vital for enactment of compliance, explanation of proofs, and reconciliation of multiple regulations. They are also useful in change-enabling compliance management of an enterprise. We are working on creating an end-to-end compliance management solution, the implementation architecture of which is as illustrated in Figure 2.

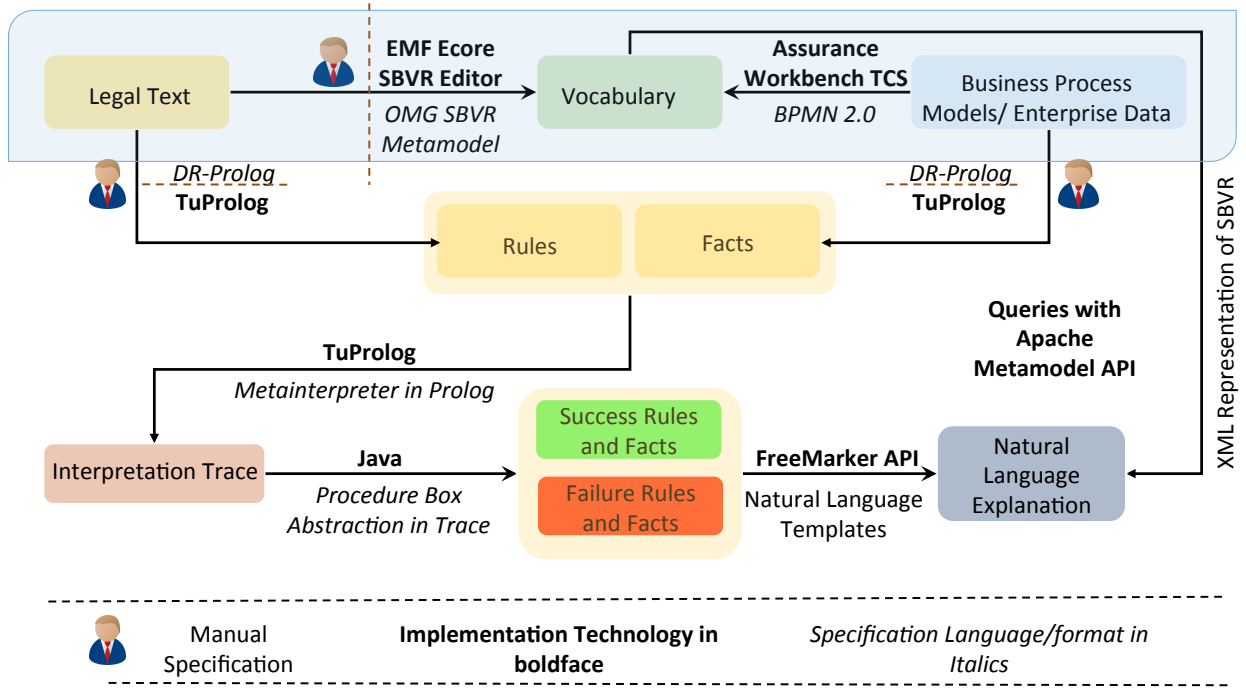


Figure 2. Implementation architecture

In the next section we describe only the vocabulary-based modeling and mapping shown at the top of Figure 2 with blue highlight. The interested reader is invited to read:

1. [18], in which we outlined how we use design science to understand and solve the semantic disparity problem and the early ideas in [26], of which this paper is an extension.
2. [39], in which we explain how rules and facts from the legal text and operations respectively are encoded as DR-Prolog rules and facts and how we obtain proofs of (non-)compliance and then query the respective vocabularies to generate natural language explanation.
3. [40], in which we explain how this implementation architecture is utilized to capture changes in governance, risk, and compliance when either regulations or operations or both change.

We use Semantics of Business Vocabulary and Rules (SBVR) to model vocabularies of regulatory and operational concepts as elaborated next.

3 Modeling and Mapping Vocabularies

Our approach for mapping concepts from regulations and operational specifics is illustrated in Figure 3. Vocabulary_{Reg} and $\text{Terminological Dictionary}_{Operations}$ indicate SBVR vocabularies of regulations and operational specifics respectively. Operational specifics may be present in any BP modeling form or as enterprise data/taxonomies. The concepts from individual vocabularies Vocabulary_{Reg} and $\text{Terminological Dictionary}_{Operations}$ are mapped using semantic similarity measures. By expressing these concepts with a pre-determined set of synonyms for each pair of concepts from both Vocabulary_{Reg} and $\text{Terminological Dictionary}_{Operations}$, it is possible to express compliance checking uniformly using a given formalism. Next we describe how SBVR can be used to model aforementioned vocabularies. In Section 3.2, we explain how we use semantic similarity measures to map the concepts from these vocabularies.

We imported elements shown in Figure 4 from the consumable XMI of SBVR meta-model available at OMG site⁴ into Eclipse Modeling Framework Ecore model. The BP model is created and traversed using an in-house tool that we describe in [41]. To interface with enterprise data, we are currently working with other in-house tools described in [42] and [43].

⁴ See under *Normative Machine Consumable Files* at <http://www.omg.org/spec/SBVR/20130601/SBVR-XMI-Metamodel.xml>

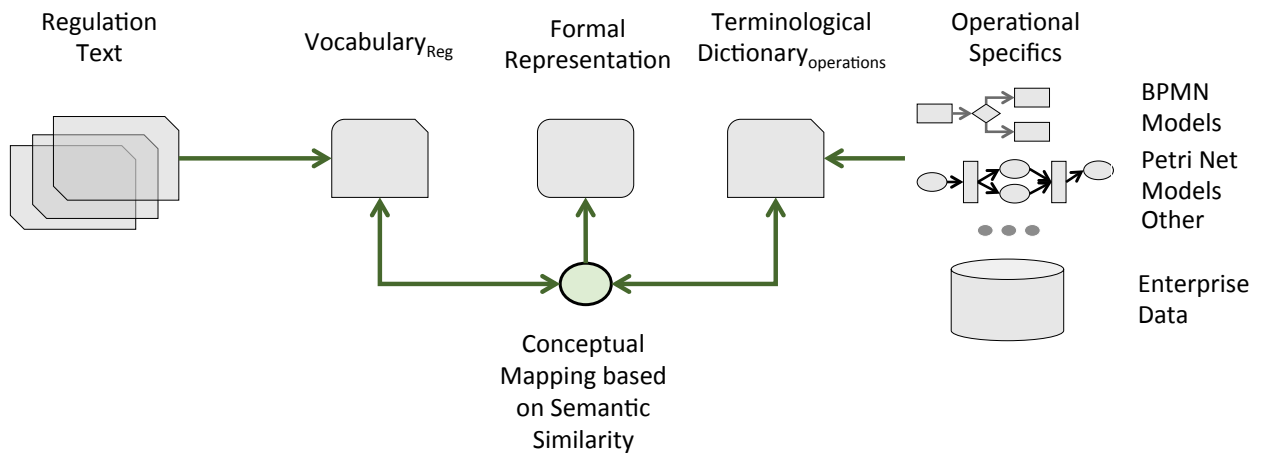


Figure 3. Using vocabularies and semantic similarity to map regulations and operational specifics [26]

In order to create vocabularies of regulatory and operational concepts, we first create a structured definition of concepts used in the problem domain using generalizations and specializations to create concept hierarchies based on SBVR metamodel shown in Figure 4.

3.1 SBVR for Regulations and Operations

We model the relations between concepts from declarative sentences that record business facts about these concepts. The constraints from restrictions mentioned on business facts form the rules. These steps lead to creation of a layered semantic model of regulations and operations, in a bottom-up manner, from concepts to fact types, to rules. Next we elaborate on specific sections of vocabularies to clarify how relevant aspects of regulations and operations are modeled:

1. **Modeling the Business Context:** First, vocabulary to capture the business context is created, consisting of the semantic community and sub-communities owning the regulation and to which the regulation applies. Each semantic community is unified by shared understanding of an area, i.e., body of shared meanings and a body of shared guidance containing business rules. These concepts are shown as *Business Vocabulary* in SBVR metamodel in Figure 4. The business domain is represented by a body of shared meanings comprising the concept model and rules that apply to these concepts. Each community defines a vocabulary that is used to designate concepts and rules defined in its body of meanings.
2. **Modeling the Meaning of Concepts:** We model the body of concepts by focusing on key terms in regulatory rules. Concepts referred in the rule are modeled as noun concepts. A general concept is defined for an entity that denotes a category. Specific details about an entity are captured as characteristics. Verb concepts capture behavior, in which noun concepts play a role. Binary verb concepts capture relations between two concepts. Characteristics are unary verb concepts. The SBVR metamodel for modeling regulation body of concepts is shown as *Meaning and Representation Vocabulary* in Figure 4. A representation represents a meaning and each meaning has a representation. Within a representation context, concepts characterize the domain of usage such that the expression of a representation has a unique meaning for a given speech community. Concepts can specialize other concepts, helping build hierarchies. Relations between concepts are captured as fact types, also known as verb concepts, in the form $\langle \text{role} \rangle \text{ verb } \langle \text{role} \rangle$, where each role stands for a noun concept that plays a specific role in this relation.
3. **Modeling the Rules:** The policies, rules, advices or guidelines are modeled based on logical formulations based on fact types from the body of concepts. We build the body of guidance

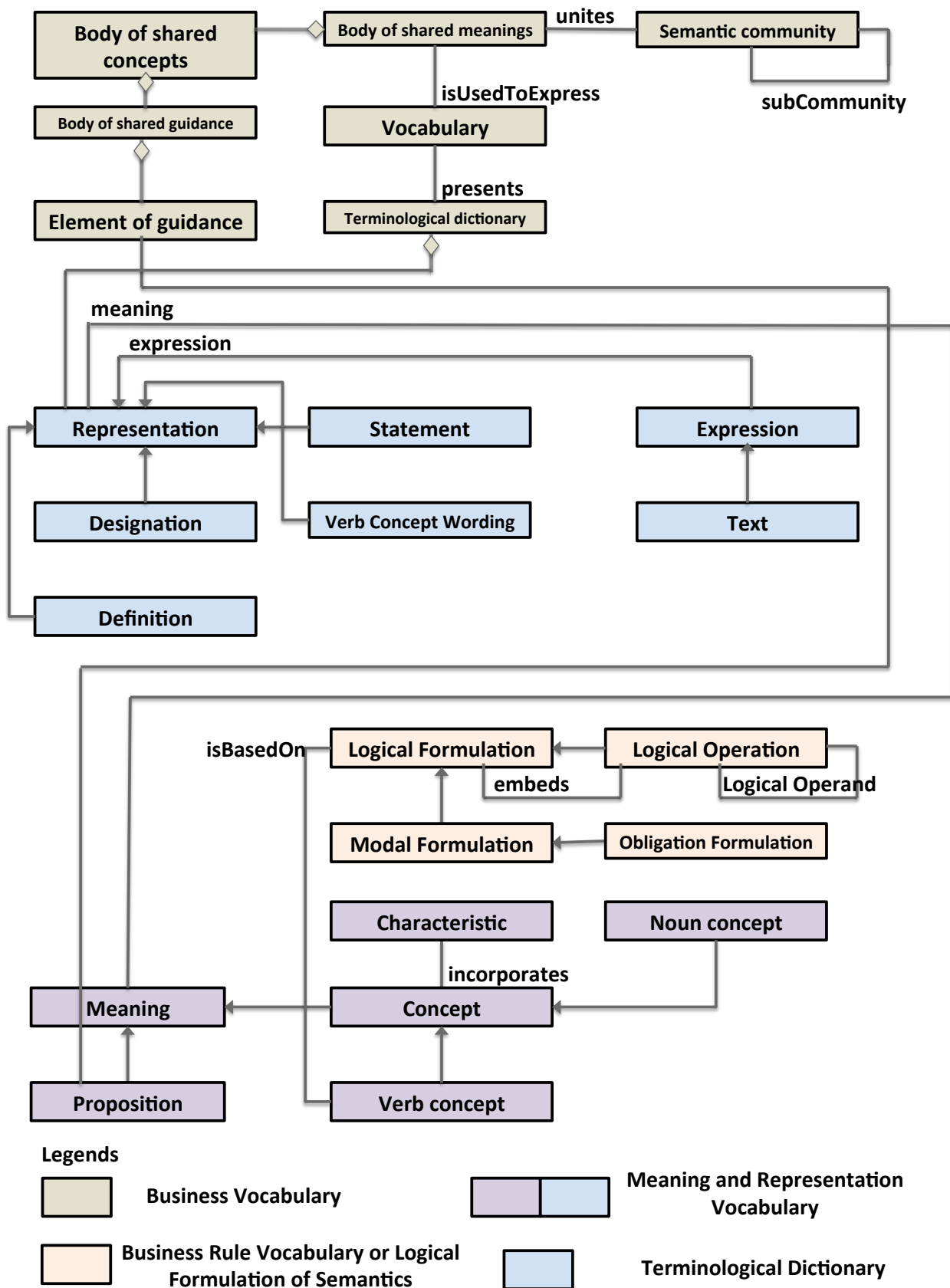


Figure 4. SBVR metamodel for creating and mapping regulations and operations vocabularies [40]

using policies laid down in the regulation. This includes logical formulation of each policy (an obligation formulation for obligatory rules) based on logical operations such as conjunctions, implications and negation. This is shown in *Business Rules Vocabulary* in Figure 4.

4. **Modeling the Terminological Variations:** The various forms of representations used by the communities for their vocabularies are modeled in terminological dictionary. These include designations or alternate names and additional information such as definitions and natural language statements for rules. We use the terminological dictionary to capture the vocabulary used by the enterprise in its operations. Depending on whether enterprise operations are represented as business process models or data/taxonomies, we extract concepts/phrases from these using proprietary tools [41]. Each activity in the process becomes a verb concept wording in the terminological dictionary. SBVR concepts for modeling terminological variations are shown as *Terminological Dictionary* in Figure 4.

3.2 Semantic Similarity

The problem of semantic similarity between two texts is defined as quantifying and identifying the presence of semantic relations between the two texts, e.g., to what extent each text has the same meaning as or is a paraphrase of the other text [44]. Semantic similarity can be measured between texts of any size such as word- to-word similarity, phrase-to-phrase similarity, sentence-to-sentence similarity, paragraph-to-paragraph similarity, or document-to-document similarity including mixed combinations of these.

Similarity measures could be broadly categorized as *geometric* measures, which enable to assess similarity between entities by considering them as points in a dimensionally organized metric space and *feature-based* measures, which utilized characteristics of the examined objects and assumes that similarity is a function of both common and distinctive features. Recent work also enables combining feature-based measures with information theoretic measures by including informativeness of concepts [45].

To apply similarity measures to regulatory and operational concepts, we utilize SEMILAR⁵ semantic similarity toolkit. We choose SEMILAR because it makes available extensive documentation and examples of usage for various similarity algorithms. It also enables implementing various semantic similarity approaches at different levels of text granularity with facilities to manually annotate texts with semantic similarity relations using semantic similarity annotation tool [44]. We particularly use the optimal matching measure, which provides an optimal solution for text-to-text similarity based on word-to-word similarity measures using ideas from optimal assignment problem. For further details on this measure, reader is requested to refer to [46].

Specializing Similarity Measure Usage

In applying similarity measures to find relatedness of regulatory and operational concepts, we are essentially automating the workflow followed by a domain expert. As described earlier in Sections 2.1 and 2.2, in both formal and industry GRC solutions, mapping is construed implicitly and explicitly respectively by domain experts working with those approaches. Ideally, regulation text gives hint about which operational process or a workflow at large is under consideration, which actors participate in enacting a regulation, and which structural or behavioral constraints does the regulation enforce. We utilize this way of arriving at the specific task description in a business process model or enterprise data to find applicability of a regulation as illustrated in Figure 5.

Figure 5 shows that with structured vocabularies by our side, it is possible to query $Vocabulary_{Reg}$ and $Terminological Dictionary_{Operations}$ so that space of applicable operational specifics is pruned by a particular process and within a process, by interactions between particular actors. Once these details are obtained from $Vocabulary_{Reg}$, we can separate the operational terms

⁵ See SEMILAR semantic similarity toolkit <http://www.semanticsimilarity.org/>

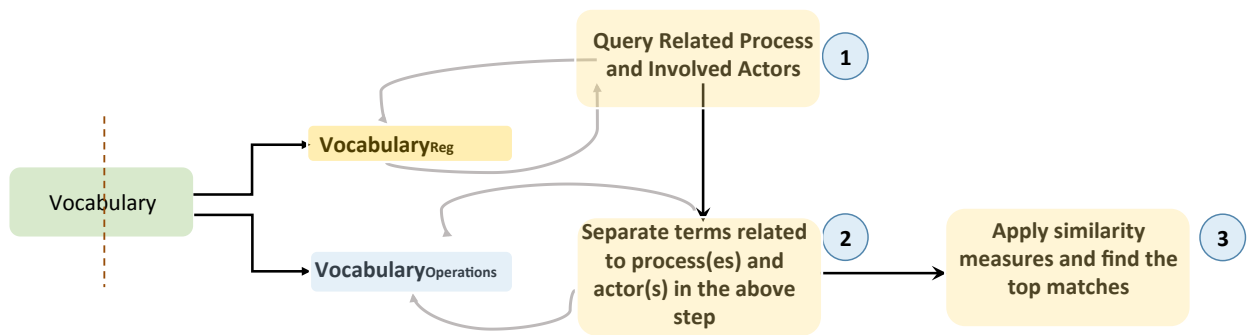


Figure 5. Steps in using semantic similarity measures

from Terminological Dictionary_{Operations}, which provide us with restricted set of texts to match from Vocabulary_{Reg} with Terminological Dictionary_{Operations}. We then use SEMILAR’s similarity measures, particularly *optimal matching* measure, to provide the domain expert with the top-K matches, amongst which she can choose the best match based on domain knowledge.

4 Case Study

Below we give some illustrative excerpts from RBI KYC regulation and then show, how we model both regulatory and operational concepts.

KYC Regulation text

§1.1 KYC Norms/ Anti-Money Laundering (AML) Measures/ Combating of Financing of Terrorism (CFT)

The objective of KYC/AML/CFT guidelines is to prevent banks from being used, intentionally or unintentionally, by criminal elements for money laundering or terrorist financing activities. KYC procedures also enable banks to know/understand their customers and their financial dealings better, which in turn help them manage their risks prudently.

§1.2 Definition of Customer

For the purpose of KYC Policy, a ‘Customer’ is defined as a person or entity that maintains an account and/or has a business relationship with the bank; one on whose behalf the account is maintained (i.e., the beneficial owner); beneficiaries of transactions conducted by professional intermediaries, and any person or entity connected with a financial transaction, which can pose significant reputational or other risks to the bank, say, a wire transfer or issue of a high value demand draft as a single transaction.

§2.2 KYC Policy

Banks should frame their KYC policies incorporating the following four key elements: Customer Acceptance Policy, Customer Identification Procedures, Monitoring of Transactions, and Risk Management.

§2.3 Customer Acceptance Policy

Banks should ensure that a) . . . ii) Parameters of risk perception are clearly defined to enable categorization of customers into low, medium and high risk. iii) Documentation requirements and other information to be collected in respect of different categories of customers depending on perceived risk iv) Not to open an account where the bank is unable to apply appropriate customer due diligence measures, i.e., bank is unable to verify the identity and

/or obtain documents required as per the risk categorization. . . . c) Risk categorization: Illustrative examples of low risk customers could be salaried employees, while examples of customers requiring higher due diligence include PEPs. Banks should apply enhanced due diligence measures based on the risk assessment.’

§2.4 Customer Identification Procedure

b) Salaried Employees: In case of salaried employees, it is clarified that with a view to containing the risk of fraud, banks should rely on certificate/letter of identity and/or address issued only from corporate and other entities of repute and should be aware of the competent authority designated by the concerned employer to issue such certificate/letter. Further, in addition to the certificate/letter issued by the employer, banks should insist on at least one of the officially valid documents as provided in the Prevention of Money Laundering Rules (viz. passport, driving licence, PAN Card, Voter’s Identity card, etc.) or utility bills for KYC purposes for opening bank accounts of salaried employees of corporate and other entities. f) Politically Exposed Persons (PEPs) resident outside India: Banks should verify the identity of the person and seek information about the sources of funds before accepting the PEP as a customer.

We use the following font styles of the SBVR Structured English to express our model: term font for noun concepts and roles; Name font for special concepts or names; *verb* font for fact types; and **keyword** font for other words in definitions and statements.

4.1 Modeling Vocabularies

In the following, we present how vocabularies of regulatory concepts, banking concepts, and policy statements are modeled.

Regulatory Concepts Vocabulary

Regulation *addresses* risk

Regulation *guides* policy

Policy *describes* Internal control

Internal control *contains* risk

Internal control *fulfills* policy

Process *has associated* risk

Process *contains* task

Actor *performs* task

KYC regulation (Concept Type: Regulation)

Risk type (Concept Type: Risk)

Money laundering (Concept Type: Risk type)

Financing terrorism (Concept Type: Risk type)

Customer acceptance policy (Concept Type: Policy)

Customer identification procedure policy (Concept Type: Policy)

Transaction monitoring policy (Concept Type: Policy)

Risk management policy (Concept Type: Policy)

Risk categorization (Concept Type: Internal controls)

High risk (Concept Type: Risk categorization)

Customer due diligence (Concept Type: Internal controls, Signifier: CDD)

Additional information collection (Concept Type: Customer due diligence)

Additional approval (Concept Type: Customer due diligence)

Documentation collection (Concept Type: Customer due diligence)
Risk profiling (Concept Type: Internal controls)
Account opening process (Concept Type: Process)
Bank official (Concept Type: Actor)
Senior manager (Concept Type: Actor)

KYC regulation *addresses risk of* Money laundering
KYC regulation *addresses risk of* Financing terrorism
KYC regulation *guides* Customer acceptance policy
Risk categorization *contains risk of* Money laundering
Risk categorization *fulfills* Customer acceptance policy
Customer due diligence *fulfills* Customer acceptance policy
Account opening process *has associated risk of* Money laundering
Account opening process *contains* Request documents from customer
Account opening process *contains* Customer submits documents
Account opening process *contains* Review documents
Bank official *performs* Review documents
Bank official *performs* Request documents from customer
Customer *performs* Customer submits documents

Banking Concepts Terminology

Customer (characteristics: isBeneficiary, nature of business, location, turnover, social status)
Client (Concept type: Customer)
Individual (Concept type: Customer)
Private Salaried Employee (Concept type: Individual)
Politically Exposed Person (Concept type: Individual)
Organization (Concept type: Customer)
Bank
Documentation

Employer certificate (Concept type: Documentation)
Documentation collection *requires* Documentation to be collected from customer
Customer *provides* documentation
Documentation collection *contains* Risk Of Fraud
Private salaried employee *provides* Employer certificate (specialization of above fact type)
Customer *maintains* Account
Customer *has business relationship with* Bank
Customer due diligence *depends on* Risk categorization
Politically exposed person *is categorized as* High risk (from §2.3 a) ii))
Additional Information *depends on* High risk (from §2.3 a))
Additional approval *depends on* High risk (from §2.5b)
BankOfficial *performs* Customer due diligence
SeniorManager *performs* Additional approval
Customer *passes* Customer due diligence
Bank *accepts* Customer

Policy Statements as Rules

Customer *maintains* Account or *has* Business relationship with Bank or *is* Beneficiary or *is* Beneficial owner or *is* entity associated with risky transaction

It is obligatory that (Customer provides documentation and BankOfficial performs Customer due diligence and Customer passes Customer due diligence) *implies* (Bank accepts Customer)

The obligation statement above is an example of customer acceptance rule, which is an operative business rule.

4.2 Mapping Vocabularies

In the following, we consider an Indian public sector bank, which must comply with RBI KYC regulations. Given that it would have set processes (although not necessarily business process models), we use the vocabularies to query and zoom in on applicable operational specifics as described in Section 3.2.

```
<SBVR.VocabularyforDescribingBusinessVocabularies:ComplianceModel>
<contains Id="RBI_reference">
  <presentsVocabulary Id="RBI_RegulationVocabulary"/>
  <expressesBodyOfMeanings Id="RBI_KYCRegulation"/>
  <includesBodyOfGuidance Id="RBI_KYCRules">
    <includesElementsOfGuidance Id="r3">
      <Id>r3</Id>
    </includesElementsOfGuidance>
  </includesBodyOfGuidance>
</contains>
<isMeantBy
  xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:obligationformulation">
  <antecedent xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:conjunction">
  <logicalOperand xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:atomicformulation">
    <Id>ind</Id>
    <isBasedOn>client_is_ind</isBasedOn>
  </logicalOperand>
  <logicalOperand xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:atomicformulation">
    <Id>pse</Id>
    <isBasedOn>client_is_pse</isBasedOn>
  </logicalOperand>
  <logicalOperand xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:atomicformulation">
    <Id>approvedCorporate</Id>
    <isBasedOn>approvedCorporate</isBasedOn>
  </logicalOperand>
  <logicalOperand xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:atomicformulation">
    <Id>acceptApprovedCorpCertificate</Id>
    <isBasedOn>acceptApprovedCorpCertificate</isBasedOn>
  </logicalOperand>
  </antecedent>
  <consequent xsi:type="SBVR.LogicalFormulationofSemanticsVocabulary:atomicformulation">
    <Id>open_account</Id>
    <isBasedOn>open_account</isBasedOn>
  </consequent>
</isMeantBy>
</includesElementsOfGuidance>
</includesBodyOfGuidance>
</contains>
</SBVR.VocabularyforDescribingBusinessVocabularies:ComplianceModel>
```

Figure 6. Private Salaried Employee Rule from §2.4 b in SBVR XML format

As an example, consider regulation §2.4 b) for salaried employees employed at private corporates described earlier in this section. Customer identification and acceptance policies such as §2.3 and §2.4 are indicated to relate to the bank’s account opening process (see the last statement in §2.4 b)). This information is captured in the Regulatory Concepts Vocabulary as Account opening process contains Review documents. Furthermore, the actors involved here are the bank (i.e., a bank official) and the salaried employee. The Banking Concepts Terminology shows the concept BankOfficial performs Customer due diligence and also Private Salaried Employee (Concept type: Individual).

We query the XML format of stored SBVR vocabularies using Apache Metamodel⁶, which provides SQL like query API to query XML data. After querying the Vocabulary_{Reg} for concepts related to regulatory rule about private salaried employee (as in §2.4 b)), we use the Terminological Dictionary_{Operations} to query specific activities referred to in the regulations. A generic rule, such as the one shown above in Policy Statements, applies to all employees with employee type specific variations. The specific rule for private salaried employee is shown as an excerpt from XML format of the vocabulary in Figure 6.

Note specifically the <isBasedOn> elements, which refer to conditions in the rule related to specific activities such as document collection from customer by a bank official indicated by Documentation collection requires Documentation to be collected from customer and Customer provides documentation, which captures document submission by the customer. Given that the bank has several processes, we already have vocabulary concepts that can help us get to the applicable process and specific activities where, for instance, the rule for private salaried employee needs to be enacted. This is shown in Figure 7.

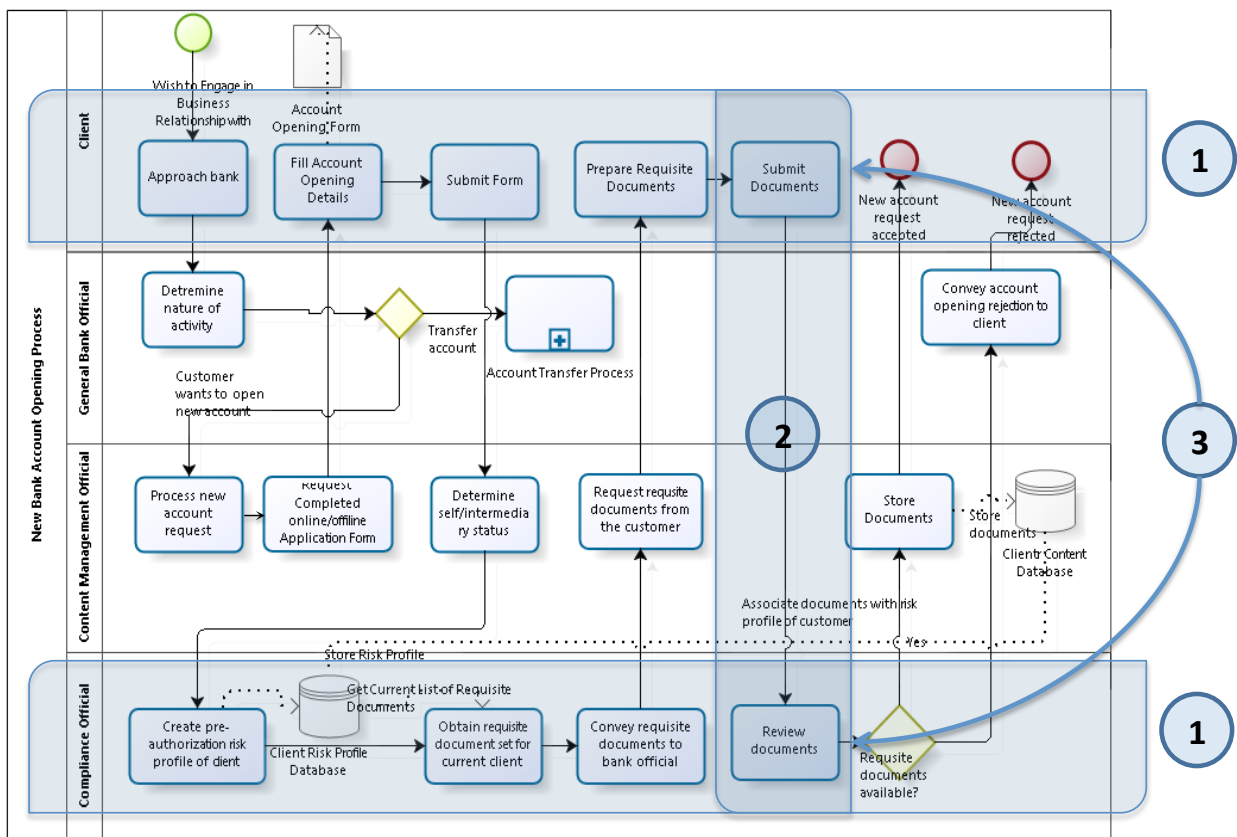


Figure 7. Bank Account Opening process

The steps of specialized semantic similarity usage from Figure 5 are shown as overlays ①, ②, and ③ in Figure 7.

For the rule formulation shown in Figure 6, we find two tasks in the business process, namely, *submit documents* and *review documents* by ①, ②. Once we apply semantic similarity measures between the sets of terms *provides* and *performs* related to the Customer and BankOfficial, and terms *submit* and *review* related to Client and Compliance Official and get satisfactory similarity threshold, the domain expert is pointed to these specific tasks in the business process where she can choose to attach the rule to *review documents* task.

⁶ See <http://metamodel.apache.org/>

Finding the applicability of regulatory rules to operational specifics is, thus, made explicit in our approach with vocabularies of regulatory and operational concepts that we modeled and mapped. We have not yet experimented with Latent Semantic Analysis and Latent Dirichlet Allocation methods offered in SEMILAR, but with RBI KYC corpus it might be possible to utilize these methods with greater accuracy.

From compliance checking point of view, if trace-based compliance checking mechanism is used as in any of the compliance checking approaches reviewed earlier in Section 2.1 or if compliance was checked against enterprise data signifying, which documents were submitted by the customer as in [47], it becomes possible to formally establish whether or not this rule was complied with.

5 Discussion and Future Work

Our overall motivation is to enable better compliance management by bringing together best of both academic and industry GRC features. In doing so, we do not intend to replace either of them, rather we envision a concomitant use of our framework with existing approach at a given enterprise.

Our approach can be used in conjunction with content management systems used in GRC frameworks, to bring formalism into the current compliance process and reduce burden on experts. Available content management tools can be used for population of the semantic models of regulation and enterprise, proposed by our approach, from natural language text documents and other enterprise information sources and to map these models.

Automated compliance checking can cut cost and time for checking compliance and bring in accuracy to current industry approaches. As described in Section 2, we utilize formalism from academic approaches toward formal compliance checking, proof explanation, and change management, the details of which can be found in [18], [26], [39], and [40].

We envision the future work, thus, along the following lines:

- We plan to utilize existing GRC services such as OpenCalais as well as techniques from academia such as [48] and NL2SBVR⁷ to extract vocabularies from natural language which we currently carry out manually.
- We plan to combine taxonomy tagging in current GRC approaches with our approach as suggested above.
- By leveraging our work in enterprise intentional models [49], [50] and vocabularies of concepts, thereof, we plan to extend explanations of proofs of (non-)compliance with risk categorization and corresponding business reasons; a functionality that current GRC frameworks provide by leveraging tagging to create risk-adjusted decision report based on input from domain experts.

We are currently working on capturing all of customer categories in RBI KYC and later plan to evaluate our approach for compliance to FATCA⁸ and BASEL III⁹ for Indian Banks.

6 Conclusion

We presented an exhaustive review of both academic and industry regulatory compliance approaches. Formal compliance checking approaches mostly assume a terminological mapping to exist between concepts of regulations and operations while industry approaches rely on tagging mechanisms and tools for the same. Explicitly modeling and mapping vocabularies of regulatory and operational concepts in the context of compliance frameworks eases the burden for the domain expert to find applicability of specific regulation in operational details. We demonstrated how we

⁷ See NL2SBVR Project <http://www.cs.bham.ac.uk/~bxb/NL2OCLviaSBVR/NL2SBVR.html>

⁸ See FATCA <http://www.irs.gov/Businesses/Corporations/Foreign-Account-Tax-Compliance-Act-FATCA>

⁹ See BASEL III <http://www.bis.org/bcbs/basel3.htm>

use SBVR and SEMILAR in conjunction to model and map the vocabularies. Results from work in progress seem promising, and with compliance regime looming over enterprises, considerable further research possibilities exist.

References

- [1] J. A. W. French Caldwell, “Magic quadrant for enterprise governance, risk and compliance platforms,” Gartner, 2013.
- [2] S. English and S. Hammond, “Cost of compliance 2014,” Thomson Reuters Accelus, 2014.
- [3] A. Awad, G. Decker, and M. Weske, “Efficient compliance checking using BPMN-Q and temporal logic,” in *Business Process Management, 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings*, ser. Lecture Notes in Computer Science, M. Dumas, M. Reichert, and M. Shan, Eds., vol. 5240. Springer, 2008, pp. 326–341. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-85758-7_24
- [4] J. Becker, C. Ahrendt, A. Coners, B. Weiß, and A. Winkelmann, “Business rule based extension of a semantic process modeling language for managing business process compliance in the financial sector,” in *Informatik 2010: Service Science - Neue Perspektiven für die Informatik, Beiträge der 40. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Band 1, 27.09. - 1.10.2010, Leipzig*, ser. LNI, K. Fähnrich and B. Franczyk, Eds., vol. 175. GI, 2010, pp. 201–206. [Online]. Available: <http://subs.emis.de/LNI/Proceedings/Proceedings175/article5838.html>
- [5] M. El Kharbili, S. Stein, I. Markovic, and E. Pulvermüller, “Towards a framework for semantic business process compliance management,” in *The Impact of Governance, Risk, and Compliance on Information Systems (GRCIS)*, ser. CEUR Workshop Proceedings, vol. 339, Montpellier, France, June 17 2008, pp. 1–15.
- [6] M. Hashmi, G. Governatori, and M. T. Wynn, “Normative requirements for business process compliance,” in *Service Research and Innovation - Third Australian Symposium, ASSRI 2013, Sydney, NSW, Australia, November 27-29, 2013, Revised Selected Papers*, ser. Lecture Notes in Business Information Processing, J. G. Davis, H. Demirkan, and H. R. Motahari-Nezhad, Eds., vol. 177. Springer, 2013, pp. 100–116. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-07950-9_8
- [7] D. Knuplesch, M. Reichert, L. T. Ly, A. Kumar, and S. Rinderle-Ma, “Visual modeling of business process compliance rules with the support of multiple perspectives,” in *Conceptual Modeling - 32th International Conference, ER 2013, Hong-Kong, China, November 11-13, 2013. Proceedings*, ser. Lecture Notes in Computer Science, W. Ng, V. C. Storey, and J. Trujillo, Eds., vol. 8217. Springer, 2013, pp. 106–120. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-41924-9_10
- [8] L. T. Ly, D. Knuplesch, S. Rinderle-Ma, K. Göser, H. Pfeifer, M. Reichert, and P. Dadam, “Seaflows toolset - compliance verification made easy for process-aware information systems,” in *Information Systems Evolution - CAiSE Forum 2010, Hammamet, Tunisia, June 7-9, 2010, Selected Extended Papers*, ser. Lecture Notes in Business Information Processing, P. Soffer and E. Proper, Eds., vol. 72. Springer, 2010, pp. 76–91. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-17722-4_6
- [9] S. W. Sadiq, G. Governatori, and K. Namiri, “Modeling control objectives for business process compliance,” in *5th International Conference, BPM, Proceedings, 2007*, pp. 149–164. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-75183-0_12
- [10] G. Antoniou, A. Bikakis, N. Dimareisis, M. Genetzakis, G. Georgalis, G. Governatori, E. Karouzaki, N. Kazepis, D. Kosmadakis, M. Kritsotakis, G. Lilis, A. Papadogiannakis, P. Padiaditis, C. Terzakis, R. Theodosaki, and D. Zeginis, “Proof explanation for a nonmonotonic semantic web rules language,” *Data & Knowledge Engineering*, vol. 64, no. 3, pp. 662 – 687, 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.datak.2007.10.006>

- [11] E. Ramezani, D. Fahland, and W. M. P. van der Aalst, “Where did I misbehave? diagnostic information in compliance checking,” in *Business Process Management - 10th International Conference, BPM 2012, Tallinn, Estonia, September 3-6, 2012. Proceedings*, ser. Lecture Notes in Computer Science, A. P. Barros, A. Gal, and E. Kindler, Eds., vol. 7481. Springer, 2012, pp. 262–278. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-32885-5_21
- [12] I. Gómez-Sebastià, S. Álvarez-Napagao, J. Vázquez-Salceda, and L. O. Felipe, “Towards runtime support for norm change from a monitoring perspective,” in *Proceedings of the First International Conference on Agreement Technologies, AT 2012, Dubrovnik, Croatia, October 15-16, 2012*, ser. CEUR Workshop Proceedings, S. Ossowski, F. Toni, and G. A. Vouros, Eds., vol. 918. CEUR-WS.org, 2012, pp. 71–85. [Online]. Available: <http://ceur-ws.org/Vol-918/111110071.pdf>
- [13] M. Dastani, J. C. Meyer, and N. A. M. Tinnemeier, “Programming norm change,” *Journal of Applied Non-Classical Logics*, vol. 22, no. 1-2, pp. 151–180, 2012. [Online]. Available: <http://dx.doi.org/10.1080/11663081.2012.682784>
- [14] G. Boella, M. Janssen, J. Hulstijn, L. Humphreys, and L. van der Torre, “Managing legal interpretation in regulatory compliance,” in *International Conference on Artificial Intelligence and Law, ICAIL '13, Rome, Italy, June 10-14, 2013*, E. Francesconi and B. Verheij, Eds. ACM, 2013, pp. 23–32. [Online]. Available: <http://doi.acm.org/10.1145/2514601.2514605>
- [15] J. Becker, P. Delfmann, M. Eggert, and S. Schwittay, “Generalizability and applicability of modelbased business process compliance-checking approaches — a state-of-the-art analysis and research roadmap,” *BuR — Business Research*, vol. 5, no. 2, pp. 221–247, 2012, publication status: Published. [Online]. Available: <http://dx.doi.org/10.1007/BF03342739>
- [16] T. Humberg, C. Wessel, D. Poggenpohl, S. Wenzel, T. Ruhroth, and J. Jürjens, “Ontology-based analysis of compliance and regulatory requirements of business processes,” in *CLOSER 2013 - Proceedings of the 3rd International Conference on Cloud Computing and Services Science, Aachen, Germany, 8-10 May, 2013*, F. Desprez, D. Ferguson, E. Hadar, F. Leymann, M. Jarke, and M. Helfert, Eds. SciTePress, 2013, pp. 553–561.
- [17] N. Racz, E. Weippl, and A. Seufert, “Governance, risk & compliance (GRC) software - an exploratory study of software vendor and market research perspectives,” in *Proceedings of the 2011 44th Hawaii International Conference on System Sciences*, ser. HICSS '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 1–10. [Online]. Available: <http://dx.doi.org/10.1109/HICSS.2011.215>
- [18] S. Sunkle, D. Kholkar, and V. Kulkarni, “Solving semantic disparity and explanation problems in regulatory compliance- A research-in-progress report with design science research perspective,” in *Enterprise, Business-Process and Information Systems Modeling - 16th International Conference, BPMDS 2015, Held at CAiSE 2015, Stockholm, Sweden, June 8-9, 2015, Proceedings*, ser. Lecture Notes in Business Information Processing, K. Gaaloul, R. Schmidt, S. Nurcan, S. Guerreiro, and Q. Ma, Eds., vol. 214. Springer, 2015, pp. 326–341. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-19237-6_21
- [19] G. Governatori, J. Hoffmann, S. Sadiq, and I. Weber, “Detecting regulatory compliance for business process models through semantic annotations,” in *Business Process Management Workshops*, ser. Lecture Notes in Business Information Processing, D. Ardagna, M. Mecella, and J. Yang, Eds. Springer Berlin Heidelberg, 2009, vol. 17, pp. 5–17. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-00328-8_2
- [20] A. Kamada, G. Governatori, and S. Sadiq, “Transformation of sbvr compliant business rules to executable fcl rules,” in *RuleML 2010: 4th International Web Rule Symposium*, no. 6403. Springer, 2010, pp. 153–161. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-16289-3_14
- [21] L. T. Ly, S. Rinderle-Ma, D. Knuplesch, and P. Dadam, “Monitoring business process compliance using compliance rule graphs,” in *OTM 2011, Proceedings, Part I*, 2011, pp. 82–99. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-25109-2_7

- [22] A. Awad, A. Polyvyanyy, and M. Weske, "Semantic querying of business process models," in *12th International IEEE Enterprise Distributed Object Computing Conference, ECOC 2008, 15-19 September 2008, Munich, Germany*. IEEE Computer Society, 2008, pp. 85–94. [Online]. Available: <http://dx.doi.org/10.1109/EDOC.2008.11>
- [23] A. Awad, M. Weidlich, and M. Weske, "Visually specifying compliance rules and explaining their violations for business processes," *J. Vis. Lang. Comput.*, vol. 22, no. 1, pp. 30–55, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.jvlc.2010.11.002>
- [24] Y. Liu, S. Müller, and K. Xu, "A static compliance-checking framework for business process models," *IBM Systems Journal*, vol. 46, no. 2, pp. 335–362, 2007. [Online]. Available: <http://dx.doi.org/10.1147/sj.462.0335>
- [25] S. Muller and C. Supatgiat, "A quantitative optimization model for dynamic risk-based compliance management," *IBM Journal of Research and Development*, vol. 51, no. 3.4, pp. 295–307, May 2007.
- [26] S. Sunkle, D. Kholkar, and V. Kulkarni, "Toward better mapping between regulations and operational details of enterprises using vocabularies and semantic similarity," in *Proceedings of the CAiSE 2015 Forum at the 27th International Conference on Advanced Information Systems Engineering (CAiSE 2015), Stockholm, Sweden, June 10th, 2015.*, ser. CEUR Workshop Proceedings, J. Grabis and K. Sandkuhl, Eds., vol. 1367. CEUR-WS.org, 2015, pp. 229–236. [Online]. Available: <http://ceur-ws.org/Vol-1367/paper-30.pdf>
- [27] AvePoint. (2014, Oct.) AvePoint compliance guardian product brochure. [Online]. Available: http://www.avepoint.com/assets/pdf/Compliance_Guardian_product_brochure.pdf
- [28] C. Searching. (2013) Compound term processing- delivering the ability to identify and correctly weight multi-word concepts in unstructured text. [Online]. Available: http://www.conceptsearching.com/wp/wp-content/uploads/collateral/CS_Compound_Term_Processing_Solution.pdf
- [29] M. B. Pohlman, *Oracle identity management: governance, risk, and compliance architecture, third edition*, 3rd ed. Boston, MA, USA: Auerbach Publications, 2008.
- [30] H. Du, M. A. Vasarhelyi, and X. Zheng, "XBRL mandate: Thousands of filing errors and so what?" *Journal of Information Systems*, vol. 27, no. 1, pp. 61–78, 2013. [Online]. Available: <http://dx.doi.org/10.2308/isys-50399>
- [31] J. W. Bartley, Y. A. Chen, and E. Z. Taylor, "A Comparison of XBRL Filings to Corporate 10-Ks - Evidence from the Voluntary Filing Program," *Social Science Research Network*, Feb 2010. [Online]. Available: <http://ssrn.com/abstract=1397658>
- [32] C. P. Cheng, G. T. Lau, K. H. Law, J. Pan, and A. Jones, "Improving access to and understanding of regulations through taxonomies," *Government Information Quarterly*, vol. 26, no. 2, pp. 238 – 245, 2009, building the Next-Generation Digital Government Infrastructures. [Online]. Available: <http://dx.doi.org/10.1016/j.giq.2008.12.008>
- [33] A. Awad, M. Weidlich, and M. Weske, "Specification, verification and explanation of violation for data aware compliance rules," in *Service-Oriented Computing, 7th International Joint Conference, ICSOC-ServiceWave 2009, Stockholm, Sweden, November 24-27, 2009. Proceedings*, ser. Lecture Notes in Computer Science, L. Baresi, C. Chi, and J. Suzuki, Eds., vol. 5900, 2009, pp. 500–515. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-10383-4_37
- [34] S. Ghanavati, L. Humphreys, G. Boella, L. Di Caro, L. Robaldo, and L. van der Torre, "Compliance with multiple regulations," in *Conceptual Modeling*, ser. Lecture Notes in Computer Science, E. Yu, G. Dobbie, M. Jarke, and S. Purao, Eds. Springer International Publishing, 2014, vol. 8824, pp. 415–422. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-12206-9_35

- [35] G. Boella, L. Humphreys, M. Martin, P. Rossi, and L. W. N. van der Torre, “Eunomos, a legal document and knowledge management system to build legal services,” in *AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents - International Workshop AICOL-III, Held as Part of the 25th IVR Congress, Frankfurt am Main, Germany, August 15-16, 2011. Revised Selected Papers*, ser. Lecture Notes in Computer Science, M. Palmirani, U. Pagallo, P. Casanovas, and G. Sartor, Eds., vol. 7639. Springer, 2011, pp. 131–146. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-35731-2_9
- [36] G. Boella, S. C. Tosatto, S. Ghanavati, J. Hulstijn, L. Humphreys, R. Muthuri, A. Rifaut, and L. W. N. van der Torre, “Integrating legal-urn and eunomos: Towards a comprehensive compliance management solution,” in *AI Approaches to the Complexity of Legal Systems - AICOL 2013 International Workshops, AICOL-IV@IVR, Belo Horizonte, Brazil, July 21-27, 2013 and AICOL-V@SINTELNET-JURIX, Bologna, Italy, December 11, 2013, Revised Selected Papers*, ser. Lecture Notes in Computer Science, P. Casanovas, U. Pagallo, M. Palmirani, and G. Sartor, Eds., vol. 8929. Springer, 2013, pp. 130–144. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-45960-7_10
- [37] D. G. Gordon and T. D. Breaux, “Reconciling multi-jurisdictional requirements: An case study in requirements water marking,” in *Proceedings of the 20th IEEE Requirements Engineering Conference*, ser. RE '12, 2012, pp. 91–100.
- [38] K. Sapkota, A. Aldea, M. Younas, D. A. Duce, and R. Bañares-Alcántara, “RP-Match: A framework for automatic mapping of regulations with organizational processes,” in *IEEE 10th International Conference on e-Business Engineering, ICEBE 2013, Coventry, United Kingdom, September 11-13, 2013*. IEEE Computer Society, 2013, pp. 257–264. [Online]. Available: <http://dx.doi.org/10.1109/ICEBE.2013.39>
- [39] S. Sunkle, D. Kholkar, and V. Kulkarni, “Explanation of proofs of regulatory (non-)compliance using semantic vocabularies,” in *Rule Technologies: Foundations, Tools, and Applications - 9th International Symposium, RuleML 2015, Berlin, Germany, August 2-5, 2015, Proceedings*, ser. Lecture Notes in Computer Science, N. Bassiliades, G. Gottlob, F. Sadri, A. Paschke, and D. Roman, Eds., vol. 9202. Springer, 2015, pp. 388–403. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-21542-6_25
- [40] S. Sunkle, D. Kholkar, and V. Kulkarni, “Model-driven regulatory compliance: A case study of “know your customer” regulations,” in *Model-Driven Engineering Languages and Systems - 18th International Conference, MODELS 2015, held at Ottawa, Canada, Accepted, 2015*. [Online]. Available: <http://dx.doi.org/10.1109/MODELS.2015.7338275>
- [41] D. Kholkar, P. Yelure, H. Tiwari, A. Deshpande, and A. Shetye, “Experience with industrial adoption of business process models for user acceptance testing,” in *ECMFA*, ser. Lecture Notes in Computer Science, P. V. Gorp, T. Ritter, and L. M. Rose, Eds., vol. 7949. Springer, 2013, pp. 192–206. [Online]. Available: <http://dx.doi.org/10.1007/978-3-642-39013-5>
- [42] S. Shah and S. Reddy, “Context aware ontology based information extraction,” in *Proceedings of the 18th International Conference on Management of Data, COMAD 2012, 2012, Pune, India*, C. Sahasrabudhe, A. E. Abbadi, K. Murthy, and A. Bhattacharya, Eds. Computer Society of India, 2012, pp. 32–43. [Online]. Available: <http://comad.in/comad2012/pdf/shah.pdf>
- [43] R. R. Yeddula, P. Das, and S. Reddy, “A model-driven approach to enterprise data migration,” in *Advanced Information Systems Engineering - 27th International Conference, CAiSE 2015, Stockholm, Sweden, June 8-12, 2015, Proceedings*, ser. Lecture Notes in Computer Science, J. Zdravkovic, M. Kirikova, and P. Johannesson, Eds., vol. 9097. Springer, 2015, pp. 230–243. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-19069-3_15
- [44] V. Rus, M. C. Lintean, R. Banjade, N. B. Niraula, and D. Stefanescu, “SEMILAR: the semantic similarity toolkit,” in *51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, Proceedings of the Conference System Demonstrations, 4-9 August 2013, Sofia*,

- Bulgaria*. The Association for Computer Linguistics, 2013, pp. 163–168. [Online]. Available: <http://aclweb.org/anthology/P/P13/P13-4028.pdf>
- [45] G. Pirró and J. Euzenat, “A feature and information theoretic framework for semantic similarity and relatedness,” in *The Semantic Web – ISWC 2010*, ser. Lecture Notes in Computer Science, P. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Pan, I. Horrocks, and B. Glimm, Eds. Springer Berlin Heidelberg, 2010, vol. 6496, pp. 615–630. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-17746-0_39
- [46] V. Rus, M. Lintean, C. Moldovan, W. Baggett, N. Niraula, and B. Morgan, “The similar corpus: A resource to foster the qualitative understanding of semantic similarity of texts,” in *Semantic Relations II: Enhancing Resources and Applications, in the proceedings of The 8th Language Resources and Evaluation Conference (LREC 2012) at Istanbul, Turkey*. Springer Berlin Heidelberg, May 2012.
- [47] A. Bikakis, C. Papatheodorou, and G. Antoniou, “The DR-Prolog tool suite for defeasible reasoning and proof explanation in the semantic web,” in *Artificial Intelligence: Theories, Models and Applications, 5th Hellenic Conference on AI, SETN 2008, Syros, Greece, October 2-4, 2008. Proceedings*, ser. Lecture Notes in Computer Science, J. Darzentas, G. A. Vouros, S. Vosinakis, and A. Arnellos, Eds., vol. 5138. Springer, 2008, pp. 345–351. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-87881-0_31
- [48] S. Ramzan, I. S. Bajwa, I. U. Haq, and M. A. Naeem, “A model transformation from NL to SBVR,” in *Ninth International Conference on Digital Information Management, ICDIM 2014, Phitsanulok, Thailand, September 29 - Oct. 1, 2014*. IEEE, 2014, pp. 220–225. [Online]. Available: <http://dx.doi.org/10.1109/ICDIM.2014.6991430>
- [49] S. Sunkle, V. Kulkarni, and S. Roychoudhury, “Analyzing Enterprise Models Using Enterprise Architecture-Based Ontology,” in *MoDELS*, ser. Lecture Notes in Computer Science, A. Moreira, B. Schätz, J. Gray, A. Vallecillo, and P. J. Clarke, Eds., vol. 8107. Springer, 2013, pp. 622–638. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-41533-3_38
- [50] S. Sunkle, S. Roychoudhury, and V. Kulkarni, “Using Intentional and System Dynamics Modeling to Address WHYs in Enterprise Architecture,” in *ICSOFIT*, J. Cordeiro, D. A. Marca, and M. van Sinderen, Eds. SciTePress, 2013, pp. 24–31.