Toward Better Mapping between Regulations and Operational Details 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,deepali.kholkar,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. 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 enterprises. We believe that this proposal takes a step toward adapting and leveraging formal compliance checking approaches in industry GRC solutions.

Keywords: Regulatory Compliance, Operational Details, 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. While industry governance, risk, and compliance (GRC) solutions help enterprises in managing regulatory compliance, they are mostly document-oriented and are not as rigorous as formal approaches to compliance checking. Formal compliance checking can offer several analysis benefits to industry GRC solutions such as formally finding out (non-)compliance to regulations [3–9] 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. 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 [14–16].

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 propose to map the concepts from structured SBVR-based vocabularies of regulations and operational specifics using semantic similarity measures.

The paper is arranged as follows. In Section 2, we review several works in formal compliance checking with regards if and how they map concepts from regulations and operational specifics of enterprise and enlist our observations. Based on our observations, we propose the use of SBVR-based vocabularies and semantic similarity measures in Section 3 to map these conceptual realms. Section 4 concludes the paper.

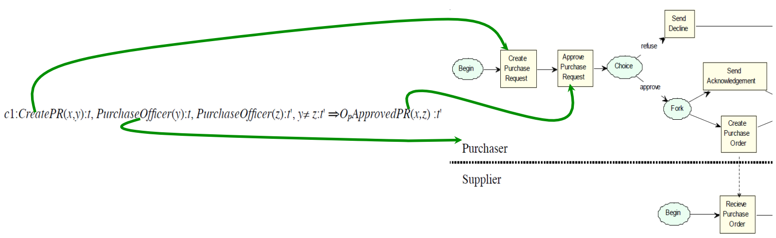
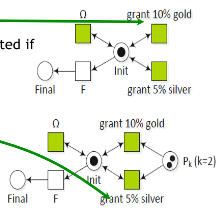
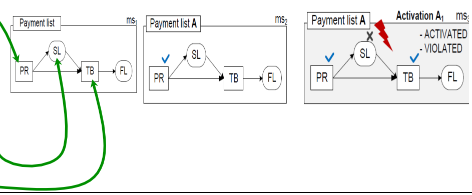
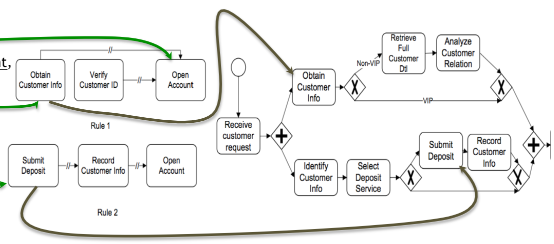
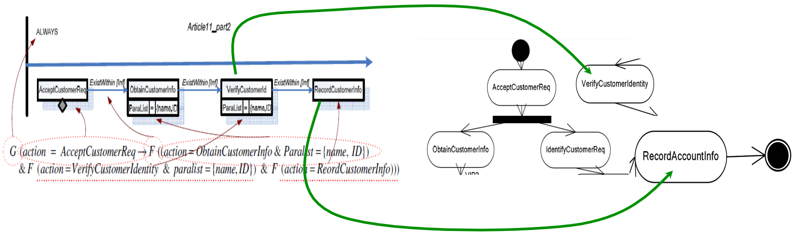
2 Related Work and Motivation

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. Our specific aim in presenting the related work is to show how these approaches *map* concepts from regulations with concepts from BP models. We consider five representative formal compliance checking approaches, namely defeasible logic-based [9, 17, 18], Petri-net based [11], compliance rule graph-based [7, 8, 19], extended BP modeling notation (BPMN) query and linear temporal logic (LTL)-based [3, 20, 21], and Business Property Specification Language (BPSL) and LTL-based approaches [22, 23]. Table 1 illustrates these approaches in two columns. First column shows how each approach maps labels/phrases from regulations to labels/phrases from approach-specific representation of BP models and second column notes formalism in that approach. In the following, we briefly elaborate the formal compliance checking approaches with regards mapping between labels/phrases row by row from Table 1.

First row from Table 1 shows defeasible logic-based approach for checking compliance of BP models against regulations [9]. Regulations are modeled in Formal Contract Language (FCL) which is a combination of efficient non-monotonic defeasible logic and deontic logic of violations. 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 BP model respectively. Label *PurchaseOfficer* from FCL expression maps to *Purchaser* from BP 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 [18] and semantic annotations of BP models in [17], but a structured terminological mapping of concepts is yet not explored.

Second row from Table 1 shows an approach in which an event log describing the observed operational behavior is aligned with a Petri-net pattern that formalizes a regulation. From the regulation shown in the second row of Table 1, phrases *a discount of*

Table 1: Disparity between Labels in Formal Regulations and Operational Specifics

Mapping between Regulations and Operational Details	Compliance Rule/Operational Formalism
	Formal Contract Language based on Defeasible Logics; Operational Specifics as a Business Process Model
Regulation: “A discount of 10% is granted if the customer is a gold customer; 5% are granted if the customer is a silver customer.” 	Rules as Petri-net Patterns, Operations from the Event Log
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” 	Rules in terms of Compliance Rule Graph; Operations in terms of Events
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. 	Rules in Business Process Modeling Notation (BPMN) Query (BPMN-Q), later in Temporal Logic; operational specifics in terms of BPMN Models
	Rules in graphical Business Property Specification Language, later into Linear Temporal Logic; operational specifics in terms of BP models in the Business Process Execution Language, later into Pi calculus

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].

Third row from Table 1 shows an approach where events from operational event trace are checked against graph-based compliance rule language called compliance rule graph that formalizes a regulation. 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 [19] from which this example is taken or other publications from same authors [7, 8].

Fourth row from Table 1 shows an example from [3]. It uses BPMN-Q which is a visual language based on BPMN used to query BP 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 left. Interestingly, the concepts from BPMN-Q representation of the regulation match with the BP model shown by process graph on the right. This is to be expected since BPMN-Q visual queries are based on corresponding BP 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 [20, 21] similarly do not express the need for explicit mapping and presume that terminological mapping from regulation statements to BPMN-Q queries exists.

Finally, fifth row from Table 1 shows an example from [22]. BP 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. The example shows that BPSL formulation of labels *RecordCustomerInfo* and *VerifyCustomerId* map to BP 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 most formal compliance checking approaches assume that labels/phrases from regulation statements map to labels/phrases used in various regulation and BP specification languages. There are approaches that recognize the need for explicit mapping between the concepts such as [16] and use word databases which consider co-occurrent words and synonyms of an activity name from the BP models. However, this approach lacks formal compliance checking as in other approaches enumerated so far. Considerable research has been done on semantic similarity of texts outside the context of regulatory compliance. An approach in [24] uses information content of texts to yield similarity judgements that correlate more closely with human assessments than other measures. Since rules and activities are short length pieces of text, it is possible to use method described in [25], which combines corpus- and knowledge-based similarity measures targeted at matching short length pieces of texts more accurately.

Industry models of operational specifics, whether they are BPMN-based BP models or enterprise data, are generally extremely large. Texts of regulations such as Sarbanes-Oxley (SOX), Foreign Account Tax Compliance Act (FATCA), BASEL-III, Dodd-

Frank and various anti money laundering regulations like Know Your Customer (KYC), etc., are similarly large with several interdependent regulations. From semantic similarity point of view, specific and short length pieces of texts need to be matched. Without such an explicit terminological mapping between these two sets of concepts, it is difficult to practically apply formal compliance checking approaches. In the next section, we sketch our proposed approach in this direction.

3 Vocabularies and Semantic Similarity

Our approach for mapping concepts from regulations and operational specifics is illustrated in Figure 1. $Vocabulary_{Reg}$ and 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. The concepts from individual vocabularies $Vocabulary_{Reg}$ and Terminological $Dictionary_{Operations}$ are mapped using semantic similarity measures. By expressing these concepts with a predetermined set of synonyms for each pair of concepts from both $Vocabulary_{Reg}$ and Terminological $Dictionary_{Operations}$, it is possible to express compliance checking uniformly using a given formalism. We briefly describe next how SBVR can be used to model aforementioned vocabularies.

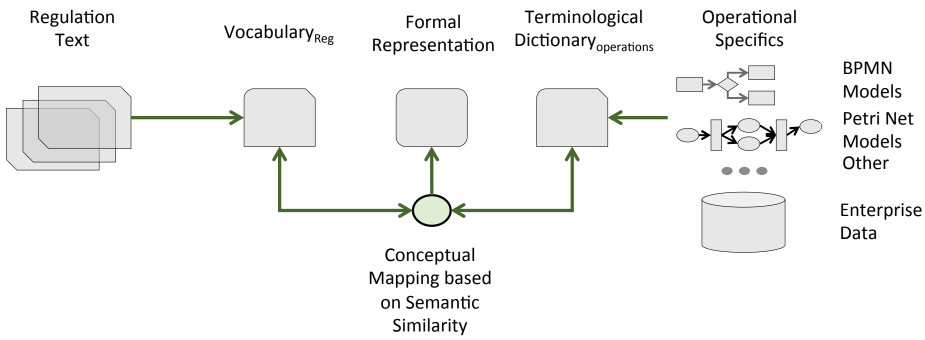


Fig. 1: Using Vocabularies and Semantic Similarity to Map Regulations and Operational Specifics

Modeling Concept Vocabularies SBVR vocabularies for regulations and operations can be defined in terms of four sections. 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 2.

Second, the body of concepts is modeled 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.

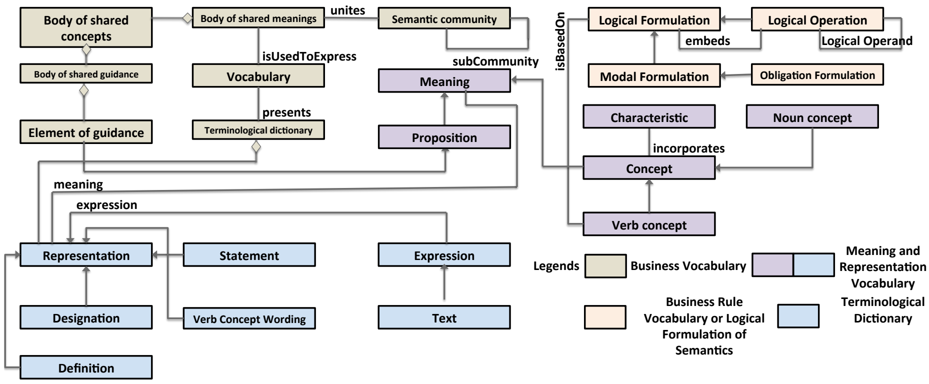


Fig. 2: SBVR Metamodel For Creating and Mapping Regulations and Operations Vocabularies

Binary verb concepts capture relations between two concepts. Characteristics are unary verb concepts. The SBVR metamodel for modeling regulation body of concepts are shown as *Meaning and Representation Vocabulary* in Figure 2.

Third, we build the body of guidance 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 2.

Fourth and lastly, we model the terminological dictionary that contains various representations used by a semantic community for its concepts and rules defined above. These consist of designations or alternate names for various concepts, definitions for concepts and natural language statements for policies stated in the regulation. We also use the terminological dictionary to capture the vocabulary used by the enterprise in its business processes. 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 2.

Mapping Concepts by Similarity Once the vocabularies of concepts from regulations and BP models are available, we can use similarity measures as in [16, 24, 25]. Note that the SBVR-based vocabularies provide a structured corpus where it is possible to encode domain knowledge including interpretations of regulations by various stakeholders [14]. The intended outcome of using vocabularies of regulations and BP models and similarity measures is that unlike approaches illustrated in Table 1, an explicit mapping between concepts of regulations and BP models will be achieved.

4 Conclusion

We illustrated several formal compliance checking approaches that assume a terminological mapping to exist between concepts of regulations and BP models when checking BP models for compliance against regulations. Structured vocabularies with semantic similarity between concepts would be needed when checking compliance of large BP models in industry against extensive regulations like BASEL-III, Dodd-Frank, FATCA, and various geography-specific KYC regulations. We plan to create an explicit mapping between regulations and BP models concept for which we briefly described how SBVR and semantic similarity measures can be used.

References

1. French Caldwell, J.A.W.: Magic quadrant for enterprise governance, risk and compliance platforms (Gartner) (2013)
2. English, S., Hammond, S.: Cost of compliance 2014 (Thomson Reuters Accelus) (2014)
3. Awad, A., Decker, G., Weske, M.: Efficient compliance checking using BPMN-Q and temporal logic. In Dumas, M., Reichert, M., Shan, M., eds.: Business Process Management, 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings. Volume 5240 of Lecture Notes in Computer Science., Springer (2008) 326–341
4. Becker, J., Ahrendt, C., Coners, A., Weiß, B., Winkelmann, A.: Business rule based extension of a semantic process modeling language for managing business process compliance in the financial sector. In Fähnrich, K., Franczyk, B., eds.: Informatik 2010. Volume 175 of LNI., GI (2010) 201–206
5. El Kharbili, M., Stein, S., Markovic, I., Pulvermüller, E.: Towards a framework for semantic business process compliance management. In: The Impact of Governance, Risk, and Compliance on Information Systems (GRCIS). Volume 339 of CEUR Workshop Proceedings., Montpellier, France (June 17 2008) 1–15
6. Hashmi, M., Governatori, G., Wynn, M.T.: Normative requirements for business process compliance. In Davis, J.G., Demirkan, H., Motahari-Nezhad, H.R., eds.: Service Research and Innovation. Volume 177 of Lecture Notes in Business Information Processing., Springer (2013) 100–116
7. Knuplesch, D., Reichert, M., Ly, L.T., Kumar, A., Rinderle-Ma, S.: Visual modeling of business process compliance rules with the support of multiple perspectives. In Ng, W., Storey, V.C., Trujillo, J., eds.: Conceptual Modeling - 32th International Conference, ER 2013, Hong-Kong, China, November 11-13, 2013. Proceedings. Volume 8217 of Lecture Notes in Computer Science., Springer (2013) 106–120
8. Ly, L.T., Knuplesch, D., Rinderle-Ma, S., Göser, K., Pfeifer, H., Reichert, M., Dadam, P.: Seaflows toolset - compliance verification made easy for process-aware information systems. In Soffer, P., Proper, E., eds.: Information Systems Evolution - CAiSE Forum 2010, Hammamet, Tunisia, June 7-9, 2010, Selected Extended Papers. Volume 72 of Lecture Notes in Business Information Processing., Springer (2010) 76–91
9. Sadiq, S.W., Governatori, G., Namiri, K.: Modeling control objectives for business process compliance. In Alonso, G., Dadam, P., Rosemann, M., eds.: Business Process Management, 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings. Volume 4714 of Lecture Notes in Computer Science., Springer (2007) 149–164
10. Antoniou, G., Bikakis, A., Dimaresis, N., Genetzakis, M., Georgalis, G., Governatori, G., Karouzaki, E., Kazepis, N., Kosmadakis, D., Kritsotakis, M., Lilis, G., Papadogiannakis, A., Pediaditis, P., Terzakis, C., Theodosaki, R., Zeginis, D.: Proof explanation for a nonmonotonic semantic web rules language. Data & Knowledge Engineering **64**(3) (2008) 662 – 687
11. Ramezani, E., Fahland, D., van der Aalst, W.M.P.: Where did I misbehave? diagnostic information in compliance checking. In Barros, A.P., Gal, A., Kindler, E., eds.: Business Process Management - 10th International Conference, BPM 2012, Tallinn, Estonia, September 3-6, 2012. Proceedings. Volume 7481 of Lecture Notes in Computer Science., Springer (2012) 262–278

12. Gómez-Sebastià, I., Álvarez-Napagao, S., Vázquez-Salceda, J., Felipe, L.O.: Towards runtime support for norm change from a monitoring perspective. In Ossowski, S., Toni, F., Vouros, G.A., eds.: Proceedings of AT 2012, Dubrovnik, Croatia, October 15-16, 2012. Volume 918 of CEUR Workshop Proceedings., CEUR-WS.org (2012) 71–85
13. Dastani, M., Meyer, J.C., Tinnemeier, N.A.M.: Programming norm change. Journal of Applied Non-Classical Logics **22**(1-2) (2012) 151–180
14. Boella, G., Janssen, M., Hulstijn, J., Humphreys, L., van der Torre, L.: Managing legal interpretation in regulatory compliance. In Francesconi, E., Verheij, B., eds.: International Conference on Artificial Intelligence and Law, ICAIL '13, Rome, Italy, June 10-14, 2013, ACM (2013) 23–32
15. Becker, J., Delfmann, P., Eggert, M., Schwittay, S.: Generalizability and applicability of modelbased business process compliance-checking approaches — a state-of-the-art analysis and research roadmap. BuR — Business Research **5**(2) (2012) 221–247 Publication status: Published.
16. Humberg, T., Wessel, C., Poggenpohl, D., Wenzel, S., Ruhroth, T., Jürjens, J.: Ontology-based analysis of compliance and regulatory requirements of business processes. In Desprez, F., Ferguson, D., Hadar, E., Leymann, F., Jarke, M., Helfert, M., eds.: CLOSER 2013 - Proceedings of the 3rd International Conference on Cloud Computing and Services Science, Aachen, Germany, 8-10 May, 2013, SciTePress (2013) 553–561
17. Governatori, G., Hoffmann, J., Sadiq, S., Weber, I.: Detecting regulatory compliance for business process models through semantic annotations. In Ardagna, D., Mecella, M., Yang, J., eds.: Business Process Management Workshops. Volume 17 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg (2009) 5–17
18. Kamada, A., Governatori, G., Sadiq, S.: Transformation of sbvr compliant business rules to executable fcl rules. In: RuleML 2010: 4th International Web Rule Symposium. Number 6403, Springer (2010) 153–161
19. Ly, L.T., Rinderle-Ma, S., Knuplesch, D., Dadam, P.: Monitoring business process compliance using compliance rule graphs. In Meersman, R., Dillon, T.S., Herrero, P., Kumar, A., Reichert, M., Qing, L., Ooi, B.C., Damiani, E., Schmidt, D.C., White, J., Hauswirth, M., Hitzler, P., Mohania, M.K., eds.: On the Move to Meaningful Internet Systems: OTM 2011. Volume 7044 of Lecture Notes in Computer Science., Springer (2011) 82–99
20. Awad, A., Polyvyanyy, A., Weske, M.: 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) 85–94
21. Awad, A., Weidlich, M., Weske, M.: Visually specifying compliance rules and explaining their violations for business processes. J. Vis. Lang. Comput. **22**(1) (2011) 30–55
22. Liu, Y., Müller, S., Xu, K.: A static compliance-checking framework for business process models. IBM Syst. J. **46**(2) (April 2007) 335–361
23. Müller, S., Supatgiat, C.: A quantitative optimization model for dynamic risk-based compliance management. IBM Journal of Research and Development **51**(3/4) (2007) 295–308
24. Seco, N., Veale, T., Hayes, J.: An intrinsic information content metric for semantic similarity in wordnet. In de Mántaras, R.L., Saitta, L., eds.: Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, Valencia, Spain, August 22-27, 2004, IOS Press (2004) 1089–1090
25. Mihalcea, R., Corley, C., Strapparava, C.: Corpus-based and knowledge-based measures of text semantic similarity. In: Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA, AAAI Press (2006) 775–780