GRCTC WORKING PAPER

June 2015

KNOWLEDGE ELICITATION FOR REGULATORY COMPLIANCE IN THE FINANCIAL INDUSTRY

Tom Butler, Elie Abi-Lahoud, Leona O'Brien, Marcello Ceci, Bill Emerson, and Tabbasum Naz

Abstract

Financial institutions are struggling to address the volume, velocity, variety, and complexity of regulations. Failure to comply with existing regulations is costing the industry billions with 60% of profits being paid in fines and payments to customers since 2011. Such problems are caused by fundamental information system failures, be they people, process, or technology in origin. This design science research study addresses these problems by designing a standards-based methodology for interpreting regulatory texts. The methodology's constructs, model, methods and related guidelines form the architectural basis of an information technology artefact, or instantiation, called Ganesha. Based on the Object Management Group (OMG) standard - the Semantics of Business Vocabulary and Rules (SBVR), the methodology, guidelines, and related IT artefact enable practitioners to capture regulations in a Regulatory Natural Language (RNL), that is both human readable (in structured English) and machine readable (in XML and OWL). The methodology and guidelines have been evaluated through application in several regulatory domains and the results presented to major actors in the financial industry. Their instantiation in Ganesha therefore enables knowledge elicitation by legal and financial subject matter experts to facilitate knowledge sharing, training and decision making around regulatory compliance in financial services organisations.

Keywords: Design Science Research, SBVR, Semantic Technology, Regulations, Laws, Vocabulary.

KNOWLEDGE ELICITATION FOR REGULATORY COMPLIANCE IN THE FINANCIAL INDUSTRY

1 Introduction

Problems with, and deficiencies in, information systems (IS) being used in and across the financial industry contributed significantly to the organizational and institutional failures that occurred across this industry (Bamberger, 2010). The institutional response to the financial crisis in 2008 was industry reform and increased regulation on a global scale (Grant & Wilson, 2012). Kendall (2013) reports that "most of the largest banks understand that they not only have inadequate capabilities to address the regulations that have already been imposed, but that even more regulation is inevitable." The consequences of this are significant, as financial institutions have been paying over 60% of their profits in fines and payments to customers since 2011. In the US, for example, over $251 billion in fines were levied by US regulators since 2008; in the UK, £41 billion of fines and charges were incurred by banks. While the regulatory compliance failures of Global Systemically Important Banks (G-SIBs—Top tier banks) hit the headlines, those of Domestic Systemically Important Banks (D-SIBs—typically second tier institutions) and other financial institutions are no less significant. Regulatory compliance is top of the agenda across the industry. Take, for example, that JP Morgan Chase are spending $4 billion to "boost staff and fix compliance and risk-control problems"; in people terms, HSBC has 24,300 staff specialising in risk and compliance, almost 10% of its entire workforce. Hence while fines are a charge on profits, regulatory compliance is costing firms dearly, as they comprise anywhere from 5-20% of total operating costs. Such are the scale of the regulatory compliance challenges, banks such as HSBC, Standard Chartered etc. are withdrawing from smaller markets.

Our unpublished research indicates that IS support for regulatory compliance is patchy at best. Just 30% of US-based financial technology (FinTech) are focussed on compliance and risks, with support for key regulatory compliance processes the exception, rather than the norm. Hence, major G-SIBs are investing in more people than technologies due to the gap in the FinTech market. Evidence of the widespread failures to apply people, processes and IT comes from the Basel Committee on Banking Supervision viz. "One of the most significant lessons learned from the global financial crisis that began in 2007 was that banks' information technology (IT) and data architectures were inadequate to support the broad management of financial risks…This had severe consequences to the banks themselves and to the stability of the financial system as a whole." Researchers and practitioners argue that traditional information technologies are not up to the task of addressing what are significant regulatory compliance challenges facing the industry (cf. Bamberger 2010, Bennett 2011, 2013, Kendall, 2014). The Financial Industry GRC Technology Centre (GRCTC) was established in 2012, as an industry-oriented, government-funded research centre, to conduct R&D on the use of semantic technologies for regulatory compliance in the financial industry. One of its core objectives is to design and build semantic technologies to enable sense-making of complex regulations by actors in order to facilitate compliance in financial organisations.

This paper presents the findings of our design science research initiative on the development of a methodology (which includes the elaboration of concepts), guidelines, technical specification and IT artefact which enables legal and financial subject matter experts to express complex legislation, regulations and rules in a regulatory natural language (RNL). This RNL is expressed according to the Object Management Group's Semantics of Business Vocabulary and Rules (SBVR) standard, making it both human and machine readable and understandable. Specifically, we describe the Design Cycle of our research (Winter, 2008), which is tightly coupled with the Rigour Cycle of the design process (Hevner, 2007), underpinned as it is by adherence to an industry standard and the theory of

interpretation (described in Abi-Lahoud et al. 2014). Another output of the design cycle helps address the *vocabulary* and *translation problems* in the design of traditional IS (see Furnas et al., 1987; Bamberger, 2010), as it illustrates how legal, business and financial subject matter experts (SMEs) can capture domain knowledge of directives, regulations, rules, and regulatory guidance in a structured regulatory natural language. This SBVR-based RNL can also be transformed into the Resource Description Framework (RDF) or Web Ontology Language (OWL) for inclusion in a knowledge base. The design artefacts are also applicable to other domains, as indicated by insights obtained in the Rigour Cycle of our research.

The remainder of this paper is structured as follows. Section 2 first describes the motivation and context of this design research study and includes a short overview of design science research perspectives and their application in this study. Section 3 presents the findings of this design science research study, which involved two research phases, or iterative design (DR) processes, which produced several DR artefacts, culminating in the design of an IT artefact called Ganesha in Phase 2. The final section discusses the findings and offers concluding thoughts.

2 Motivation and Research Approach

The widespread failure of operational and regulatory compliance information systems across the financial industry contributed to the financial crisis in 2008 (Bamberger, 2010). However, there were also global failures in regulatory oversight (Campbell, 2011). Following the global financial crisis, the velocity, volume, and variety of regulations increased across all regulatory domains (Grant & Wilson, 2012). It is clear that big regulation has led to big fines, which are a charge on profits—typically 60% for larger banks. This in turn led to an industry-wide recruitment drive for regulatory compliance staff and an upsurge in demand for GRC and related information systems (KPMG, 2012). Such GRC IS are typically developed using traditional information technologies. Sheth (2005), however, argues that traditional GRC IS are incapable of delivering the type of risk and compliance functionality required by financial services organisations—the solution, he argues, lies in capturing the semantics of risk and regulatory compliance. Bennett (2011, 2013), Kendall (2014), and Atkin (2015) also identify problems with traditional approaches to GRC IS or data management in the financial industry and argue for solutions based on semantic technologies. Take, for example, Bamberger (2010, p. 706) that traditional IS are deficient due to "problems of translation… regarding the translation of both legal mandates and business understandings of risk into computer code and actionable controls." Hence, one of the key challenges facing financial services organisations and others servicing the sectors is in understanding the regulatory compliance implications of increasing volumes of local and global legislation, regulatory rules, and various forms of 'soft' law (Brummer, 2010). The *translation problem* arises due to primordial challenges involving the interpretation of regulatory texts, industry rules and standards, such as IFRS 9, by legal and financial subject matter experts (SMEs). This problem is magnified, however, when software engineers attempt to translate rules and compliance imperatives into IS, as the IS development process is plagued with poor communication and bias (Friedman & Nissenbaum, 1996).

It was within this context that G-SIBs, D-SIBs, GRC and FinTech firms supported the institution of the Financial Industry GRC Technology Centre (GRCTC). As indicated, GRCTC is a government-funded research centre whose purpose is to investigate the design of semantic technologies for regulatory compliance and risk data aggregation in the financial industry. Industry members sit on the steering board and inform the Design Research programme. Thus, these organizations are also collaborating in the initiative as co-researchers. In the context of the current DR study, a multidisciplinary team of 2 IS researchers, 2 knowledge engineers, 2 software engineers, and 4 legal researchers from GRCTC undertook a programme of research to make financial laws and regulations human and machine readable. It is significant that this study was conducted with the initial participation of two of the primary authors of the OMG's SBVR standard. It is also significant that one of the research artefacts from the first phase of the study, the SBVR-based methodology for regulatory

interpretation, was applied as part of a Proof-of-Concept (PoC) on Regulation W in the US by the OMG's Financial Domain Task Force and the Enterprise Data Management Council (EDM Council). While Section 3 illustrates the process and product of the Design Cycle, the participation of the standards-making bodies and industry associations indicate the transition through both Rigour and Relevance Cycles, as indicated next.

2.1 A Design Science Research Approach

According to Winter (2008, p. 471) design research (DR) is aimed at "creating solutions to specific classes of relevant problems by using a rigorous construction and evaluation process" while "design science reflects the design research process and aims at creating standards for its rigour." This Design Research (DR) project accords with Winter's (2008) categorisation of this form of research. The design artefacts being produced in this study include (1) Constructs (i.e. regulatory concepts in a vocabulary); (2) Relationships between, and rules constituted by, these concepts; (3) Models represented in the Semantics of Business Vocabulary and Rules (SBVR); and (4) Methods (an approach to the construction of concepts, relationships and models). Hevner (2007) argues that design science research must incorporate: (1) a Design Cycle, which involves the core activities of developing and evaluating the design artefacts and research processes; (2) a Rigour Cycle connects the design cycle with a knowledge base of scientific theories, experience, expertise, and meta-artefacts; and (3) a Relevance Cycle that captures the interaction between the environment or context of the problem domain and the core design activities (cf. Hevner et al., 2004). These cycles were incorporated into our design science research.

In the first phase of the DR project described below, the Rigour Cycle was underpinned by our adoption of the SBVR standard from the Object Management Group (OMG) in the design of our methodology. Design Science (DS) theory was at one level, informed by Formalism (West, 2004) and the Bunge-Wand-Weber (BWW) Ontology (Wand & Weber, 1993, 1995, 2002), but chiefly by the concepts, tenets, and abstract models provided by the SBVR standard. The Design Cycle (Hevner, 2007; cf. Hevner et al., 2004) involved an iterative approach to the Build Design Artefact and Processes Phase, and Evaluation phase. The first Phase of the DR study was governed by a hermeneutic approach to developing both constituent Constructs and Models based on a pragmatic phenomenological hermeneutic theory of interpretation developed by team members (Abi-Lahoud et al. 2014: cf. Winograd & Flores, 1986; Fonseca & Martin, 2007). Thus, DS thinking informed the DR process and underpinned its rigour.

This Design Research study commenced in February 2013. The first phase of our study saw immersion in the philosophy, tenets and processes involved in the application of the SBVR standard and direct engagement with the designers of the SBVR standard. Following this, detailed requirements were gathered, scoping research conducted, and proofs of concept (PoCs) developed. These were component activities in Hevner's (2007) Rigour, Relevance and Design Cycles. While the focus was on the Design Cycle, the nature of the study saw natural movement into and between Rigour and Relevance Cycles due to its strong foundation on the OMG SBVR standard and engagement with, and feedback from, top-tier banks in the US and Europe, an international legal firm, the legal informatics research community in the US and Europe, and members of the EDM Council. In the second phase, the DR solution to the aforementioned problems was iteratively developed. The initial move through the Rigour Cycle saw the SBVR standard being adopted as the DSR Knowledge Base. In a movement through the Relevance and Design Cycles, the Evaluation phase was conducted internally with industry co-researchers, and externally with members of the Object Management Group, where the DR artefacts were presented to members of the Enterprise Data Management Council (EDM Council), Finance Domain Task Force (FDTF), Ontology PSIG and SBVR community, at OMG Technical Meetings in 2013 and 2014. In May 2015, the EDM Council CEO Mike Atkin publically acknowledged the contribution of the SBVR-based methodology to over 100 members from London City-based G-SIBs, UK D-SIBs, GRC vendors, including Chief Data Officers, Chief Information Officers and other practitioners.

3 Design, Rigour and Relevance Cycle Phase 1-2

Despite being one of the most IS-intensive industry sectors, the financial industry lags behind other business domains in the application of semantic technologies. However, all that is changing as the financial industry is, through the Enterprise Data Management Council developing a *common vocabulary* called the Financial Industry Business Ontology (FIBO) (Bennett, 2011, 2013; Atkins, 2015). We have previously adduced evidence to the effect that this DR initiative is unique in terms of process and product. However, we argue that it transcends—in terms of purpose, detail and rigour—extant approaches to capturing the semantics of the regulatory domain in financial services. Support for this contention comes from our previously published research (Abi-Lahoud et al. 2013, 2014), presentations to, and interactions with, academics and practitioners at the OMG and the International Conference on Legal Knowledge and Information Systems (Abi-Lahoud et al. 2013, 2014, O'Brien et al. 2014), and contemporaneous published work on SBVR (Ramakrishna & Paschke, 2014). The following section 3.1 presents the Knowledge Base that informed this study's Rigour Cycle: Sections 3.2 then delineates the DSR study in terms of the design artefacts

3.1 SBVR as a Knowledge Base for the Rigour Cycle

The Semantic Business Vocabulary and Rules (SBVR—Object Management Group, 2015) is an ISO TC 37-compliant Formal Terminological Dictionary (the SBVR Vocabulary) used to define a coherent set of interconnected business concepts. It is not merely a list of terms and definitions, however. It is much more as it enables the creation of a Rulebook that contains business policies, rules etc. that govern organisational activities and processes. Based on the principles of formal and first-order logic, it was developed to enable business practitioners to capture the unambiguous meaning of the concepts and entities that constitute a business domain and the rules that govern the relationships and behaviours of such entities. Thus, SBVR could, as its authors intended, be used to cross the chasm between business and software engineering domains by providing a common vocabulary and a formal, but business-oriented mechanism, for expressing business rules. SBVR is part of the Computation Independent Model of the OMG's Model Driven Architecture and supports the use of Structured English (SE) as a machine-independent natural language. Thus, it has a structured syntax for capturing business vocabularies and business rules by representing both structural and behavioural dimensions of business processes and activities in addition to policies and rules that govern actors' in organisational contexts. SBVR is being applied in research and practice as, for example: To extract business vocabularies from business process models (Njonko et al. 2014); As a controlled natural language for business rule specifications (Njonko et al. 2014; De Jesus & De Melo, 2014); and as a standard-based approach for knowledge transformation and knowledge representation of patent law (Ramakrishna & Paschke, 2014). Its use in the financial industry as a business natural language has been indicated by Chapin (2012), while our previous published work and presentations to OMG Technical Meetings indicate its potential as a regulatory natural language (Abi-Lahoud et al. 2013, 2014). Its use has been popularised among practitioners by Rob van Haarst (2013) in his instructional SBVR Made Easy: Business Vocabulary and Rules as a Critical Asset.

An SBVR ISO TC 37-compliant formal terminological dictionary and rulebook—i.e. business vocabulary and rules—is comprised of the following:

- *Noun Concepts*—these correspond to things or entities that populate a business (or regulatory) domain
- *Verb Concepts*—these capture the relationships between entities expressed as Noun Concepts
- *Definitional Rules*—these constrain the relationships between noun concepts and are expressed as Verb Concepts
- *Behavioural rules*—these govern business activities in which Noun Concepts play roles

Thus, SBVR contains the *noun concepts*, *verb concepts*, *definitional rules* and *behavioural rules* for a specific business, business domain, and regulatory-business domain such as, for example, Anti-Money-Laundering (AML) or MiFID 2, the Revised Markets in Financial Instruments Directive.

It is interesting to note that SBVR is itself defined according to the SBVR specification; significantly, any domain-specific terminological dictionary and related rulebook is an extension of SBVR. However, SBVR does not include behavioural rules, as such. SBVR is simply a terminological dictionary that defines what SBVR is, including constituent concepts such as what behavioural rules are and how to specify them.

According to the SBVR specification, a terminological dictionary and associated rulebook should be complete and consistent:

- Each *noun concept* should be defined explicitly or adopted from an authoritative source—alternatively it may be acknowledged as implicitly understood in that it is the everyday natural language meaning of the term;
- Only defined *noun concepts* may play constituent roles in *verb concepts*;
- *Rules* should be built from defined *verb concepts* only and articulated using a defined set of structure elements such as an *obligation*, *necessity*, *if … then*, *that*, *at least* and so on.

SBVR does not have a normative syntax; any syntax that has adequate expressive power is acceptable. SBVR is specified in SBVR Structured English (SE), a simplified version of natural English. SBVR-SE is probably the most widely used syntax for domain-specific SBVR models. Thus, the conceptual model is separated from the external representation, and any (suitable simplified) natural language may be used. Other dimensions to the application of our DSR Knowledge Base and the movement through the Rigour Cycle are integrated into the following sections which describe Phase 1 and Phase 2 of our DSR study.

3.2 Iterating Though the Design Cycle: Phase 1 Methodology and Guidelines

In January 2013, when the first phase of the present DR study commenced, the team's 4 legal researchers began their training in the Semantics of Business Vocabulary and Rules (SBVR) specification from the Object Management Group (OMG). This culminated in a 2-day workshop provided by two of the chief authors of the OMG specification in February 2013. They subsequently collaborated with the team in the design of the first PoC for U.S. Anti-Money Laundering regulations drawn from the Bank Secrecy Act and associated Federal Register Final Rules. The then objective of expressing U.S. regulations thus was to transform them into a business natural language, such that: (1) Non-lawyers from the financial industry could read and understand the rules; (2) Provide an audit trail for the formation of organizational governance policies; (3) Enable software or knowledge engineers to transform from SBVR to OWL vocabularies and other formalisms, without encountering *translation problems*; and (4) Contribute to the development of the EDM Council's Financial Industry Business Ontology (FIBO). It was from this experience which emerged the Regulatory Interpretation Methodology now described.

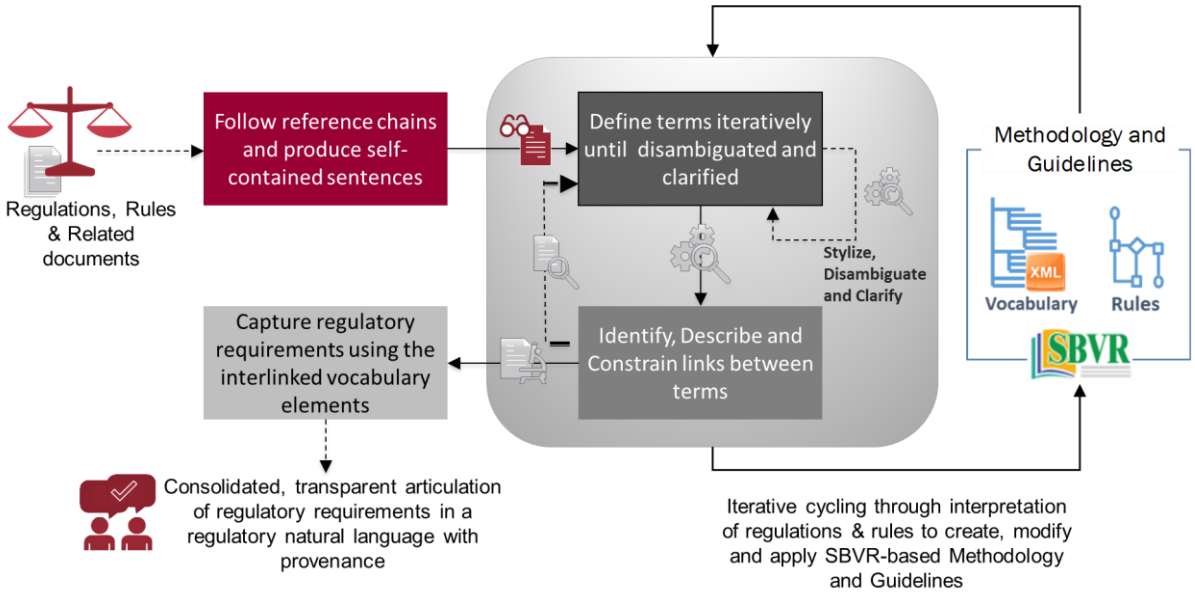

Figure 1: Iterative Design Cycle phase 1 - Regulatory Interpretation Methodology and Guidelines

The focus of the initial transition, which in this case, is the U.S. Anti-Money Laundering (AML) regulations – Title 31of the Code of Federal Regulations (CFR) Chapter X. The latter is over 100,000 words of source text before following external references to related texts in the United States Code (USC). The U.S. CFR is an annual codification of U.S. Federal Register Final Rules. Generally, Federal Register Final Rules are self-contained. Thus, a final rule adds to or amends sections of the CFR related to one or more regulated domains, for example AML sub-domains such as Money Services Businesses, Reports of Foreign Financial Accounts, etc. Beginning with such foundational texts presents legal SMEs with the opportunity to limit the scope of the *knowledge representation* using SBVR—that is, in the creation of a regulatory natural language. This commences with the identification of key sections in the legal or regulatory text. Figure 1 presents a high level view of the iterative design process which led to the development of the Methodology, Guidelines and Technical Specification which we use to create our Regulatory Natural Language. Significantly, it also underpinned the design of our DSR IT artefact – Ganesha.

We now describe the iterative design process that underpinned the development of our SBVR-based Methodology and Guidelines for legal and financial SMEs to help them identify, interpret, stylise, disambiguate, clarify the meaning of a regulation and how it is then represented as a regulatory natural language expressed in the form of SBVR-compliant vocabulary and rules. The SBVR specification enables such SMEs to structure natural text around concepts specified in the SBVR Metamodel, which may then be expressed as an XML grammar (XML-Schema). However, the SBVR meta-model is just that—it may be, as we have illustrated, modified to represent domain-specific knowledge. As previously indicated, the two main categories of concepts in the SBVR Metamodel are *Noun Concepts* and *Verb Concepts.*

- *Noun Concepts* are a groups of entities in the regulatory or business domain of discourse. For example: regulator, financial institution, etc. *Individual Noun Concepts* are a particular type of *Noun Concept* representing actual entities or individuals. For example, Securities and Exchange Commission (Regulator), RegulationW (Regulation), Wells Fargo Bank (Regulated Entity), and so on.
- *Verb Concepts* (or *Fact Types*) capture the relationships between *Noun Concepts*. For example, the *Verb Concep*t ‘money services business submits suspicious activity’ captures the

submission relationship between the *noun concepts* money services business and a suspicious activity report.

Typically, an SBVR knowledge base or semantics repository has two parts: a Vocabulary and a Rulebook. An SBVR Vocabulary is a Terminological Dictionary where entries are termed *Noun Concepts,* while relationships are captured using *Verb Concepts*. It also contains *Definitional Rules* which act to constrain these relationships using *alethic modalities* (modifiers such that '*it is necessary that'*, etc.) and related *advices of possibility*. An SBVR Rulebook is a set of guidance statements containing *Behavioural Rules* in the form of *deontic modalities* (e.g. it is obligatory that, must, etc,) and *advices of permission.*

To be compliant, an SBVR instance should be complete and consistent. This is determined by three basic principles: (1) *Noun Concepts* should be explicitly defined from the text, from other authoritative sources, or recognized as implicitly-understood by the SMEs; (2) Previously defined *Noun Concepts* are used to construct *Verb Concepts*; while (3) *Verb Concepts* are the basic ingredients in *Definitional Rules* and *Behavioural Rules*. It must be noted that SBVR does not have a normative syntax; however, the OMG specification includes SBVR Structured English (SBVR-SE), which is a relaxed convention based on natural language. SBVR-SE uses *text styles* to visually identify elements from the SBVR Metamodel. In the following example we adopt a typical style to express examples in SBVR for monochrome publication. As with software programming tools and integrated development environments, metamodel elements are stylised using different colours and italics. In the basic approach, Noun Concepts are underlined with a single line. Individual Noun Concepts are doubled underlined. The ***verb part of speech*** of a *Verb Concept* is in italic-bold font face. **Keywords** are in a bold font face and they are usually the natural language representation of logical operators, logical quantifiers, modal operators and selected articles and determiners.

Applying the SBVR Metamodel provides legal and/or financial SMEs with the ability to apply their experiential and technical knowledge to transform a complex legal rule into a human, and ultimately machine readable vocabulary. Hence, in reading the text, legal SMEs focus on identifying *modalities*; that is, the *obligations*, *prohibitions*, etc., specified or referenced by the text. Following this, we propose the following activities: (1) The *stylizing activity* consists of indicating which element in the SBVR metamodel a term (or set of terms) corresponds to while undertaking a linguistic analysis of the text. This is achieved by applying the appropriate SBVR-SE font styles to entities as *Noun Concepts* and *Verb concepts*; (2) The *disambiguation activity* consists of consolidating and identifying the explicit meaning of the text; and (3) The *clarification activity* which identifies the implicit meaning of the text in order to formulate guidance when the regulatory intent is not clear (subsequent to each *disambiguation activity*). Each of these activities demand legal experiential and technical knowledge. The Methodology and Guidelines aid these activities.

3.2.1 Use Case based on MiFID 2

In order to validate the extended SBVR Metamodel, legal researchers conducted a short experiment transforming a legal text into a RNL by following the processes set out in the methodology and guidelines. To ensure its applicability across different financial regulatory domains, a decision was taken to move away from the anti-money laundering domain which informed the original development of the methodology and guidelines. The researchers were directed towards MiFID 2 by a tier-1 bank and a leading international legal firm due to its impending enactment that will have a direct effect on all international financial entities trading into and within the EU.

The test involved two legal researchers, working independently on the transformation from the regulatory text to the metamodel, and one knowledge engineer using the metamodel to build the XML-enriched version of the law. At the end of the test, the two resulting transformations were compared to measure the consistency of the outputs.

TRANSPARENCY FOR TRADING VENUES

CHAPTER 1

Transparency for equity instruments

Article 3

Pre-trade transparency requirements for trading venues in respect of shares, depositary receipts, ETFs, certificates and other similar financial instruments

1. Market operators and investment firms operating a trading venue shall make public current bid and offer prices and the depth of trading interests at those prices which are advertised through their systems for shares, depositary receipts, ETFs, certificates and other similar financial instruments traded on a trading venue. That requirement shall also apply to actionable indication of interests. Market operators and investment firms operating a trading venue shall make that information available to the public on a continuous basis during normal trading hours.

Figure 2: Excerpt of a section from a relevant MiFIR Article

The preliminary phase was implemented by identifying the themes contained within the MiFID domain (any legal or regulatory text concerned with MiFID); these themes were then assigned to the articles of MiFID 2 and MiFIR. The identification of themes in the regulatory text allows for a more dynamic (machine assisted) knowledge exploration and retrieval. While the assignment of themes is not a hard task by itself, the consistency of the resulting classifications is of primary importance towards an effective high-level, generic and jurisdictionally independent categorization, applicable across an entire domain. The resulting classifications were different, yet consistent with each other so as to allow a common assignment of articles within the identified themes, with the only issue coming to the fore when choosing the legal perspective to adopt for the classification (e.g. legal subject involved, regulated activity, the addressee of the law).

After choosing the section to transform, the legal researchers proceeded to stylise, disambiguate and clarify the regulatory text by performing the following activities:

- divide the section into segments;
- rewrite segments in plain English, producing self-contained sentences and fleshing out references to external documents or to other parts of the same document;
- identify the modality of the rule (obligation/prohibition/permission);
- identify and define terms and relations between terms.

The resulting transformations carried out by each of the legal researchers differed in number of rules and vocabulary, yet this is not a weakness of the methodology or of the guidelines: it proves in fact that they allow for different approaches, not constraining the legal expert into a limited language.

Here is an example of a rule resulting from the transformation of a regulatory text (Article 3 of MiFIR, see Figure 2 for the original text fragment):

> **It is obligatory that each** market operator **and each** investment firm **that** *operates* **a** trading venue *makes public* current bid price **and** current offer price **and** depth of trading interest *for* shares **and** depositary receipts **and** exchange-traded fund **and** certificates **and other** financial instruments *traded on* **that** trading venue *of* **the** market operator **or that** trading venue *of* **the** investment firm

Each of the noun concepts and verb concepts of the rule are described in the vocabulary part of the SBVR-compliant RNL transformation. For example, this is the entry for “market operator”:

market operator

Definition:	is a person or persons who manages and/or operates the business of a regulated market and may be the regulated market itself.
Concept Type:	general noun concept
General Concept:	person
Source:	Article 4(1)(18) **of** Directive 2014/65/EU
Necessity:	**it is necessary that each** person **that** ***manages*** **the** business of a regulated market **or** ***operates*** **the** business of a regulated market **or** ***manages and operates*** **the** business of a regulated market **or** ***is*** **a** regulated market ***counts as*** market operator ***in*** Directive 2014/65/EU and Regulation (EU) No 600/2014

This formalization is also represented in machine-readable format (XML), in the following fashion:

```
<nounConcept nounConceptID="N1.1" status="underConstruction">
        <name>market operator</name>
        <definition>is a person or persons who manages and/or operates the business of a
                regulated market and may be the regulated market itself.</definition>
        <usedInRule>R3.1.1</usedInRule>
        <usedInRule>R3.1.2</usedInRule>
        <usedInRule>R3.1.3</usedInRule>
        <usedInRule>R3.1.4</usedInRule>
        <usedInRule>R3.2.1</usedInRule>
        <usedInRule>R3.3.1</usedInRule>
        <conceptType>
                <type>GeneralNounConcept</type>
        </conceptType>
        <generalConcept>
                <name>person</name>
        </generalConcept>
        <source>Article 4(1)(18) of Directive 2014/65/EU </source>
        <definitionContext> Directive 2014/65/EU; Regulation 600/2014 </definitionContext>
        <synonym/>
        <dictionaryBasis/>
        <linkToNecessityStatement>RR4.18</linkToNecessityStatement>
        <entryLevelMeta>
                <subjectMatterExpert>Leona O'Brien</subjectMatterExpert>
                <dateInterpreted>2015-03-05</dateInterpreted>
        </entryLevelMeta>
</nounConcept>
```

The iterative design process in this phase primarily involved the creation of SBVR-based MS Word Templates to support the stylizing and disambiguation activities and to underpin our nascent SBVR-based semantic repository, and the XML version of the repository was written by the knowledge engineer. Such (mostly) manual process confirmed to be positively time-consuming and highly complex (being difficult for the legal expert to keep track of the many implicit links established between the rules and the vocabulary entries), further underlining the need for a specific tool assisting the legal experts in the transformation. The artefacts we produced using anti-money laundering legislation and MiFID 2 (and its associated regulation MiFIR) contribute towards this goal by helping software engineering researchers design DSR IT, which is explained in the following section.

3.3 *Ganesha*: Phase 2 DSR IT Artefact

Given the scale of regulations and rules being produced globally, we realised that MS Word Templates and hypertext links would not scale up to the task. Thus, once R&D funding had been approved and allocated, 2 software engineers were recruited with an aggressive 6 month schedule of deliverables. The initial stage of Phase 2 had a newly appointed Principal Investigator collaborate with the research fellow, a post-doctoral researcher and a legal researcher to undertake the initial iterations using an Agile software development approach. One of the software engineers was from India—in

brainstorming the design and purpose of the proposed IT artefact it was christened Ganesha—the 'God of wisdom, knowledge and new beginnings'. The first task in the design of the DSR IT Artefact was to design and develop an XML schema to tag the SBVR Metamodel. This not insignificant task is now outlined.

Figure 3 illustrates the form and function of the Ganesha application. Legal documents in PDF or XML texts are imported into Ganesha by legal Subject Matter Experts. These are then presented for interpretation and the iterative process of *stylising*, *disambiguating* and *clarifying* begins. The system will highlight candidate terms—both *noun concepts* and *verb concepts*—if they have already been defined. Where multiple definitions from different jurisdictions are present in the knowledge base, the SME will select the most relevant. Ganesha will, however, maintain and present links to alternative definitions and the source texts. All actions of the SME will be recorded in the form of MetaData. This will permit versioning and document management. Thus, the SME will be enabled to create regulatory vocabularies in the form of noun and verb concepts. These will then form the building blocks of the rules, which will be drafted interactively by SMEs. Ganesha's SBVR Logic Layer will enforce compliance with the SBVR standard, as extended by this DR study. As straightforward as this sounds, the devil of detail is hidden in the complexity of the underlying SBVR-based XML schema (presented in the following sub-section).

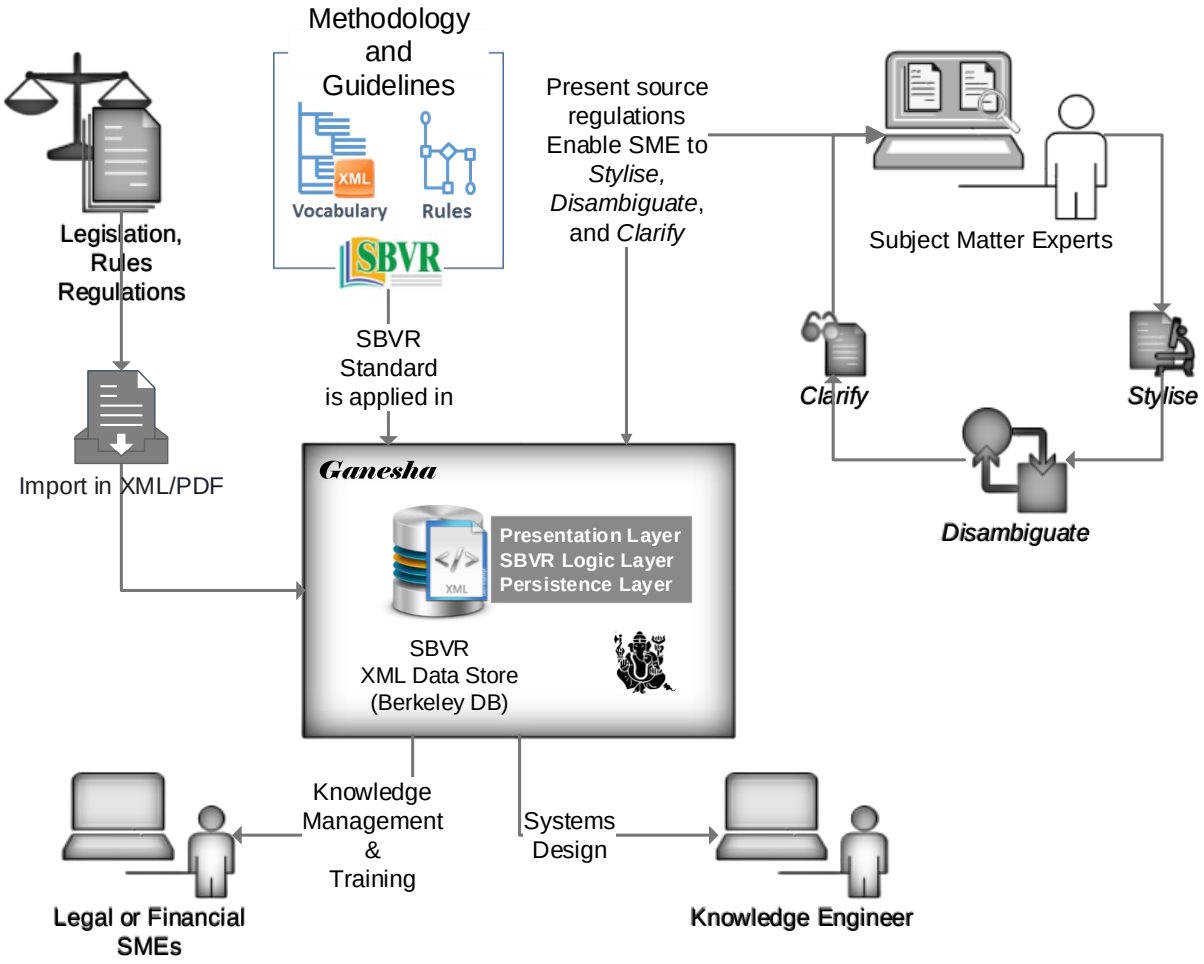

Figure 3: Ganesha – Form and Function

What the legal SME will create is a cumulative vocabulary, based on regulations, but also incorporating terms from the financial industry. Initially, the meanings and constructions on financial concepts imposed by the regulators will be included; however, these will be linked to the nascent

vocabulary being produced by the EDM Council's Financial Industry Business Ontology (FIBO), which is expressed in the web ontology language (OWL). Our interactions with the OMG indicate that OWL to SBVR transformations are possible (and vice-versa), thus linking and integration with industry standard vocabularies and taxonomies are possible. To this end, it is intended to exploit such links to enrich and widen the scope of the vocabularies. Likewise, we have investigated, at the behest of an international consulting firm, the possibility of capturing the semantics and rules in International Financial Reporting Standard (IFRS) on Financial Instruments or IFRS 9 and linking to the IFRS Taxonomy. Our scoping research indicated that this is possible. Indeed, having both regulatory and standards-based (which have a regulatory dimension) compliance imperatives in a single knowledge base. Feedback from industry indicates that this is immensely attractive and valuable proposition, for reasons will be presently adduced.

Use cases for Ganesha have been provided by top-tier banks, an international legal firm focusing on the financial industry and FinTech companies. An illustrative small example from our MiFID 2 proof of concept is presented in section 3.2.1. While we have articulated the Rolls Royce solution above, the current needs of banks and financial institutions are much more fundamental. There are currently no solutions available on the market that enable legal and financial subject matter experts capture, store and transfer/share knowledge of regulatory compliance imperatives according to an international standard and map these onto business vocabularies and rules to enhance governance, risk management and regulatory compliance. Indeed, it is our experience, and that of regulators, that financial institutions are deficient in the basic capabilities of capturing business vocabularies and rules. Thus, Ganesha provides the capabilities to present to legal or financial SMEs disambiguated and clarified vocabularies and rules in a regulatory natural language that is complete, logical and free of legalese, complexity, and ambiguity. Our methodology and guidelines have proven that this is possible—Ganesha provides the functionality to automate and informate this. In addition, while legal opinion or guidance on regulatory compliance is typically presented in Word or PDF format, such documents cannot be queried in a meaningful way. Likewise, GRC vendors typically use SQL-based repositories which do not support querying of unstructured data such as text. X-query is to an XML database such as Berkeley, what SQL is to a relational database such as Oracle. An XQuery expression can query many XML documents simultaneously and return more valuable information with greater precision and granularity than mere word searches or SQL queries (cf. McCreary & Kelly, 2013).

The sophisticated tagging, in a regulatory natural language, of SBVR-compliant XML documents in Ganesha provides powerful capabilities to extract, transform and load regulatory compliance information into a user interface for the purpose of knowledge sharing or training inexperienced personnel. In both the desktop and web-based (see below) versions, all vocabulary and rule elements with be presented in hypertext links rendering them highly navigable with the ability to explore, drill down and follow links to related content, up to and including, the original source regulations and rules, and their chapters, articles, sections and paragraphs. Furthermore, our intention to add functionality to transform from SBVR-XML to OWL provides the ability for semantic queries and inferencing of the knowledge base. Enroute to this will be the provision of graphing functionality for visualisation. Thus, vocabularies and rules will be mapped and multiple relationships highly visible.

Another use case harks back to the original reason for the institution of SBVR by the OMG to help solve both the vocabulary and translation problems identified earlier. With one notable exception, tool support for SBVR is absent from the OMG's catalogue. Hence, one of the use cases identified by a top-tier bank is to have domain specific SBVR-compliant tool to enable software engineers understand better the business and regulatory vocabulary and rules such that they can be embedded in applications and databases. Similarly, IT governance can be enabled by identifying, such as in the above paragraph of MiFID 2, the mandated information provision functionality of investment banking systems. Thus, such systems can help address operational risk where IT is concerned.

3.3.1 A Foundational SBVR-based XML Schema

This section therefore begins by introducing and explaining the XML schema that underpins the architecture of the DSR IT Artefact codenamed Ganesha. The schema provides details about regulatory document structures, elements, and attributes. Design guides and conceptual assumptions that were made in the development of the schemas are discussed, as well as an overview of the XML document using this schema. The objective of this phase of the DSR project is:

- To design solutions that capture regulatory compliance vocabularies and rules.
- To design an application to help create a semantics repository in the domain of regulatory compliance based on the Semantics of Business Vocabulary and Rules (SBVR) specification.
- The schema is used as a foundational model for Ganesha. This model is based on SBVR and is utilised to generate an SBVR-compliant XML document store. SBVR structures natural text around concepts from the SBVR Metamodel which is expressed as an XML grammar (XML-Schema).

As indicated, SBVR-compliant Microsoft Word Templates were used by legal SMEs to create RNL representations of regulatory rules. However, among other things, Word documents are not machine readable and it is not possible to directly validate the correctness of knowledge representations.

The design of Ganesha's XML schema makes it easier to enforce content structure and to validate the correctness of data. The XML schema describes the data facets (restrictions on data) and data patterns (data formats), and ensures a mutual understanding of the content between legal SME and software engineers. Based on this schema, SBVR-based XML documents can be generated, stored in an XML-based document store, and subsequently queried to identify specific information about regulations and rules. To the best of our knowledge, this is the first XML schema in the domain of regulatory compliance based on the SBVR-XML Schema. The schema follows the SBVR model in that it structures text around a vocabulary and a rulebook. We extended this with a meta-model to capture the rules, noun concepts, verb concepts and meta-data in the regulatory compliance domain.

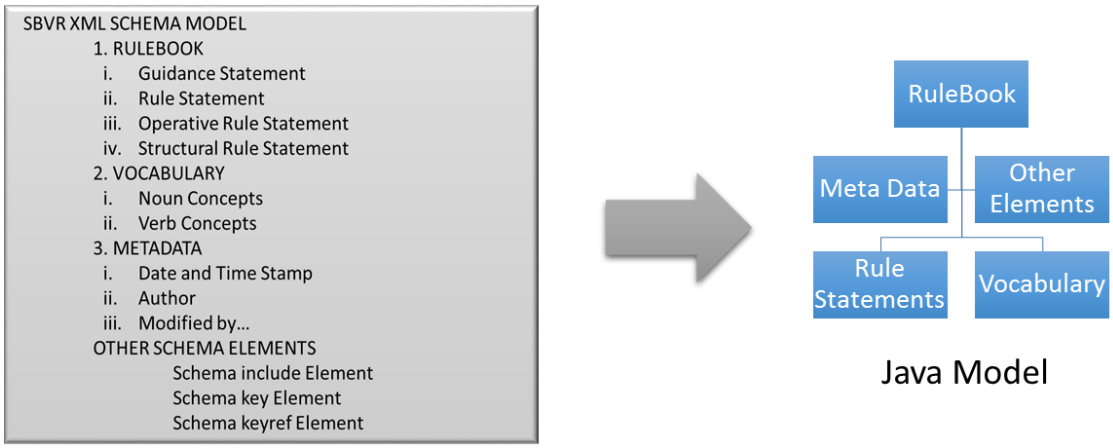

Figure 4: JAXB Architecture

The XSD (XML Schema Definition) for regulatory compliance has three main elements: RuleBook, a Vocabulary (Noun Concepts and Verb Concepts), MetaData and Other Data elements. This schema plays a key role in the architecture of the design science research IT artefact, which is next described.

3.3.2 IT Artefact Detailed Design

In the design of the IT artefact, the following challenges had to be met. The XML schema contains a rich set of connected concepts, and extends the standard SBVR schema. This gives a comprehensive base for modelling regulatory concepts, but is challenging from a software engineering point of view in terms of complexity. In addition, the XML schema is a much richer model than that possible with a UML Class Diagram or an Entity-Relationship model, and this has to be reflected in the code to support the model

Given the novelty of the approach and the demand for a usable IT artefact within an aggressive timeframe, the architecture of the system would need to be flexible and designed for possible changes in requirements. To this end, it was decided that the best way to leverage the richness of the XML schema, was to use Java Architecture for XML Binding (JAXB) to reflect the XML schema in a comprehensive java model – this is using the Domain Model software design pattern. (See Figure 4.) Any changes to the XML schema, can then be reflected with ease in the java model. The JAXB technology generates a set of tagged java classes which can be persisted to XML and marshalled back from an XML file or database with minimum effort.

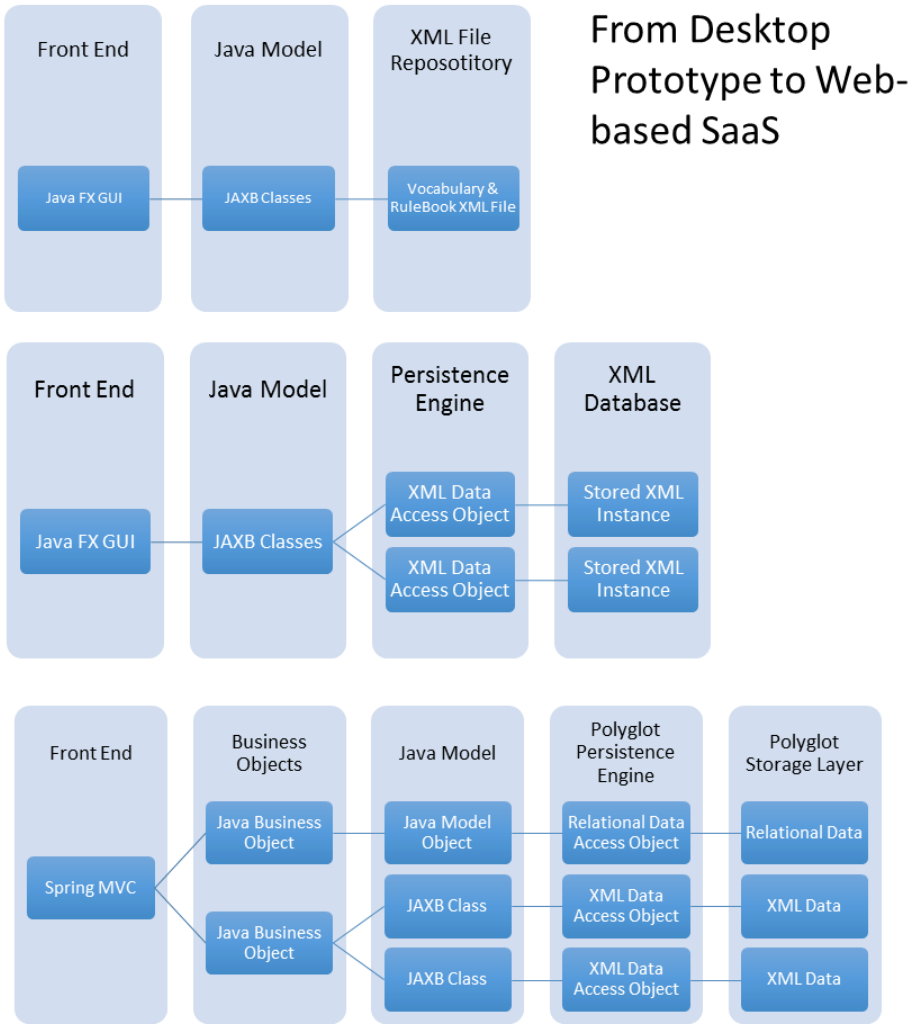

Figure 5: Iterative IT Artefact Detailed Architectural Design

To cater for the constraints outlined above, the software is being constructed in three phases – desktop, database enabled and web enabled Ganesha versions. Figure 5 illustrates an approach to IT artefact design that it both iterative and incremental. The following figure illustrates the incremental evolution of Ganesha from the desktop version, to database-enabled version to a multi-tier web version to be deployed as SaaS during 2015. The technical design of the IT artefact is based on established architecture and design patterns (cf. Knoernschild 2012; McCreary & Kelly, 2013). The desktop version was designed to be full featured in order to transition through the Relevance Cycle and maximise the fit with end-user needs in capturing end-user requirements. The database version ensured the application would scale to enterprise-level requirements. Finally, the web-based version was originally designed as Software-as-a-Service multi-instance application, while a multi-tenancy version is in the planning stage.

4 Conclusions

This design research (DR) study makes several valuable contributions to research and practice. Its pragmatic, practitioner-oriented approach presents an approach to DR that applies the philosophy (formal, first-order, deontic and alethic logic) and metamodel of the SBVR standard in order to underpin its rigour. This is further enhanced by the researchers' direct engagement with the standard's creators in the Object Management Group. The teams' network of research relationships with leading academic researchers in the area of legal informatics also places a check on the rigour cycle activities. In an industry that is seeking to increasingly standardise both process and product, including creating a common language (Bennett 2011, 2012), looking to a theoretically-informed standard to inform rigour also makes the outputs of such DR more relevant. Engaging with the broader community of practitioners through membership and engagement with an industry association such as the financial industry's Enterprise Data Management Council, also enhanced the relevance of the DR study, in that the researcher team were regularly engaging with tier-1 and tier-2 retail and investment banks. Also, specific research engagements had the research team work with FinTech companies. This approach ensured that the Design Cycle was tightly coupled with both Rigour and Relevance Cycles, as there was a natural movement into and between Rigour and Design, and Design and Relevance Cycles. While the broader engagements with practitioners were important, it was the specific engagement with and feedback from one top-tier bank, an international legal firm, and the legal informatics research community in the US and Europe that shaped the outputs of the first phase which produced design artefacts such as the regulatory interpretation methodology and guidelines. It must be noted that the methodology and guidelines were designed in an iterative fashion and applied across jurisdictions (US and Europe) and problem domains (AML, MiFID, and IFRS 9).

In the second phase, the DR solution – Ganesha – was also iteratively developed. Again, standards-based design patterns and architectures were adopted. However, innovation and contribution in the design of IT artefact was in the translation of the methodology and guidelines into the SBVR Logic Layer and the design of the complex SBVR-XML Data Schema. Here, the design of the application was informed by the use cases outlined in the previous section, thus ensuring its relevance. The design also factored in the need to scale the application and deliver it as Software-as-a-Service. Thus, in a global regulatory environment, where, for example, domestic regulations are now international in scope (Brummer, 2010), an IS such as Ganesha can be used to create a common understanding of regulatory rules from London to New York across the same financial institution. It is cautionary and incredible to believe that this is not currently possible.

References

ABI-LAHOUD E. BUTLER T and O'BRIEN L (2014) A Hermeneutic Approach to Solving the Translation Problem in Designing Ontologies 22nd. European Conference on Information Systems (ECIS), Tel Aviv, Israel, 2014.

ABI-LAHOUD E BUTLER T CHAPIN D and HALL J (2013) Interpreting Regulations with SBVR The 7th International Web Rule Symposium: Research Based and Industry Focused.

ATKIN M (2015) Unraveling the Data Mandate. EDM Council presentation to DAMA, New York, March 19, 2015, doi: http://dama-ny.com/meetinginfo.php?id=70&ts=1424461139.

BENNETT M (2011) Semantics standardization for financial industry integration. In Collaboration Technologies and Systems (CTS), IEEE, 23-27 May 2011, pp 439-445.

BENNETT M (2013) The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation* 14(3-4) 3-4.

BRUMMER C (2010) How International Financial Law Works (and How It Doesn't). *Geo. LJ* 99, 257–327.

BUTLER T and ABI-LAHOUD E (2014) Applying Semantic Technologies for Risk Data Aggregation *Consortium for System Risk Analytics (CSRA) Meeting, MIT Sloan Center for Finance and Policy, Cambridge, MA, December, 2014.*

CASELLAS N (2011) Law, Governance and Technology Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge. Springer.

CHAPIN D (2012) Why SBVR? Towards a Business Natural Language (BNL) for Financial Service. Panel Presentation, Demystifying Financial Services Semantics Conference, New York,13 March 2012, doi: http://www.omg.org/news/meetings/FS-CONF/Presentations/1015_Chapin.pdf.

DE JESUS JS and DE MELO ACV (2014) Business Rules: From SBVR to Information Systems. *Business Process Management Workshops*, 489-503.

FONSECA FT (2007) The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology* 58(6), 786-793.

FONSECA FT and MARTIN JE (2005) Toward an alternative notion of information systems ontologies: Information engineering as a hermeneutic enterprise. *Journal of the American Society for Information Science and Technology* 56(1), 46-57.

FRIEDMAN B and NISSENBAUM H (1996) Bias in computer systems. *ACM Transactions on Information Systems (TOIS)* 14(3), 330-347.

FURNAS GW, LANDAUER TK, GOMEZ LM and DUMAIS ST (1987) The vocabulary problem in human-system communication. *Communications of the ACM* 30(11), 964-971.

GRANT W and WILSON GK, Eds (2012) The Consequences of the Global Financial Crisis: The Rhetoric of Reform and Regulation. OUP Oxford.

GREGOR S and HEVNER AR (2013) Positioning and Presenting Design Science Research for Maximum Impact. *Management Information Systems Quarterly* 37(2), 337–355.

HEVNER AR (2007) The three cycle view of design science research. *Scandinavian journal of information systems* 19(2), 87-92.

HEVNER AR, MARCH ST and PARK J (2004) Design Science in Information Systems Research, *MIS Quarterly* 28(1), 75–105.

KENDALL E (2013) Semantics in Finance: Addressing Looming Train Wreck in Risk Management, Regulatory Compliance and Reporting. In Semantic Technology and Business Conference, Oct 2-3, 2013, doi: http://semtechbiznyc2013.semanticweb.com/sessionPop.cfm?confid=76&proposalid=5402.

KNOERNSCHILD K (2012) Java Application Architecture: Modularity Patterns with Examples Using OSGi. Prentice Hall.

KPMG (2012) The Convergence Evolution: Global survey into the integration of governance, risk and compliance, doi: http://www.kpmg.com/ES/es/ActualidadyNovedades/ArticulosyPublicaciones/Documents/The-Convergence-Evolution.pdf

MCCREARY D and KELLY A (2013) Making Sense of NoSQL. Manning Publications, Greenwich, Connecticut.

NIEHAVES B (2007) On epistemological diversity in design science: New vistas for a design-oriented IS research? In Proceedings of the Twenty Eighth International Conference on Information Systems, ICIS 2007, Montreal.

NJONKO PBF, CARDEY S, GREENFIELD P and EL ABED W (2014) RuleCNL: A Controlled Natural Language for Business Rule Specifications. In *Controlled Natural Language*, pp 66-77. Springer International Publishing.

OBJECT MANAGEMENT GROUP (2015) Semantics of Business Vocabulary and Business Rules V1.3 (SBVR V1.3), doi: http://www.omg.org/spec/SBVR/1.3/.

RAMAKRISHNA S and PASCHKE A (2014) A Process for Knowledge Transformation and Knowledge Representation of Patent Law. In *Rules on the Web. From Theory to Applications*, pp 311-328, Springer International Publishing.

SHETH A (2005) Enterprise Applications of Semantic Web: The Sweet Spot of Risk and Compliance. In IFIP International Conference on Industrial Applications of Semantic Web (IASW2005), Jyväskylä, Finland, August 25-27, 2005, doi: http://www.cs.jyu.fi/ai/OntoGroup/IASW-2005/

SKERSYS T, KAPOCIUS K, BUTLERIS R and DANIKAUSKAS T (2014) Extracting business vocabularies from business process models: SBVR and BPNM standards-based approach. *Computer Science and Information Systems* 11(4), 1515-1535.

SMITH B (2003) Ontology. In The Blackwell Guide to the Philosophy of Computing and Information (FLORIDI L, Ed), pp. 155-166, Blackwell, Malden, Massachusetts.

TUDORACHE T, NYYLAS C, NOY NF and MUSEN MA (2013) WebProtege: a collaborative ontology editor and knowledge acquisition tool for the Web. *Semantic Web* 4(1), 89-99.

VAN HAARST R (2013) SBVR Made Easy: Business Vocabulary and Rules as a Critical Asset, Conceptualheaven.

VAN KRALINGEN R (1997) A conceptual frame-based ontology for the law. In Proceedings of the First International Workshop on Legal Ontologies (VISSER P and WINKELS R, Eds) pp 6-17, University of Melbourne.

WAND Y and WEBER R (1993) On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal* 3(4), 217-237.

WAND Y and WEBER R (1995) On the deep structure of information systems. *Information Systems Journal* 5(3), 203-223.

WAND Y and WEBER R (2002) Research commentary: information systems and conceptual modeling—a research agenda. *Information Systems Research* 13(4), 363-376.

WEST D (2009) Object thinking. O'Reilly Media, Inc.

WINOGRAD T and FLORES F (1986) Understanding Computers and Cognition: A New Foundation for Design. Ablex Publishing Corporation, Norwood, NJ.

WINTER R (2008) Design science research in Europe. *European Journal of Information Systems* 17(5), 470-475.

ZENI N, KIYAVITSKAYA N, MICH L, CORDY JR and MYLOPOULOS J (2013) GaiusT: supporting the extraction of rights and obligations for regulatory compliance. Requirements Engineering, 1-22, doi: 10.1007/s00766-013-0181-8.

About the Authors

Tom Butler is Technology Centre Principal Investigator of Ireland's Financial Services Governance Risk and Compliance Technology Centre (GRCTC). The GRCTC conducts research on the design, development and implementation of semantic technologies for GRC in the financial industry. Tom has 112 publications since joining academia in 1998.

Elie Abi-Lahoud is a Research Fellow of the GRCTC. Elie plays a key role in engaging the research team with academic and industry partners. Elie is currently leading multiple projects with thought leaders in the Financial Services industry. He published over 20 peer-reviewed papers and advises companies on Semantic Technology.

Leona O'Brien is a legal researcher at the GRCTC. Leona plays a key role in several projects and engaging with academic and industry partners. A former financial services practitioner and lawyer, she has published several peer-reviewed papers at conferences and journals.

Bill Emerson is Principal Investigator of the Ganesha project at the GRCTC. A former Systems Architect in the Financial Industry, he has an in-depth knowledge of software engineering practice and finance. Bill has published peer-reviewed papers at conferences and leading journals.

Marcello Ceci is a postdoctoral researcher at the GRCTC. He provides the team with expertise in Legal Theory, Legal Informatics, Legal Knowledge Modelling, Regulatory Natural Languages and Operational Risk Ontology Modelling for the development of regulatory compliance information systems. Marcello has published peer-reviewed papers at leading conferences.

Tabbasum Naz is a postdoctoral researcher at the GRCTC. She provides the team with expertise on knowledge engineering, the Semantic of Business Vocabulary and Rules, software development, and Operational Risk Ontology Modelling for the development of regulatory compliance information systems. Tabbasum has published several peer-reviewed papers at conferences and journals.