Designing Semantic Technologies for Regulatory Change Management in the Financial Industry

Research-in-Progress

Tom Butler
University College Cork
Cork, Ireland
tbutler@ucc.ie

Elie Abi-Lahoud
University College Cork
Cork, Ireland
e.abilahoud@ucc.ie

Angelina Espinoza
University College Cork
Cork, Ireland
angelina.espinoza.limon@gmail.com

Abstract

The volume, velocity, variety and complexity of regulations are causing major information-related problems for the financial industry. The industry is increasingly looking to semantic technologies to help solve the many problems it faces—all of which are, in one way or another, data-, information- and knowledge-related. This design research (DR) study applies several semantic technologies to create an innovative regulatory compliance change management system prototype (RCMS). This ontology-based information system enables ground-breaking semantic tagging of lengthy and complex regulatory and legal texts, which are then loaded into a triple store for semantic querying and analysis. The findings of this DR study illustrate how practitioners can easily and efficiently identify and extract compliance obligations and prohibitions in regulatory and legal texts. This will enable them to meet the significant information-based regulatory compliance challenges they face. Initial reviews from practitioners across the financial industry highlight the relevance of our research.

Keywords: Governance, Risk and Compliance, Financial Regulation, Regulatory Change Management, Semantic Technology, Knowledge Management, Ontology, Classification.
Introduction

The financial crisis in 2008 had disastrous consequences for the world economy (Campbell 2011). Regulators across the globe responded rapidly to this situation by instituting a raft of new regulations and strengthening existing regulations (Grant and Wilson 2012). A recent commentary in The Economist outlined the current challenges facing the financial industry: “Bankers had hoped that, after seven years of penance for their part in the financial crisis, the end of wrenching overhauls forced by fierce new regulations might be nigh. But to their dismay, the regulators’ zeal is undimmed. Far from giving banks reprieve, they are toughening up old rules and devising new ones, perhaps heralding a new wave of restructuring.” The scale of the regulatory compliance problem now facing the financial industry is evident in the US Banking Reform Dodd-Frank Act (2010). It stands at 2,319 pages, compared 66 pages in the Sarbanes-Oxley Act (2002). However, Dodd-Frank is being translated in over 400 regulatory rules which will fill over 40,000 pages of regulatory text. Currently, over 200 of these are being published across various Titles in the Code of Federal Regulation and the international financial industry is in considerable confusion as to what they mean (Acharya et al. 2010). While Dodd-Frank is large and onerous, so too are the number and complexity of regulatory rules are to be found across all regulatory regimes, including the European Union (Pagliari 2013). The inability of financial institutions to address the regulatory compliance problems is evidenced in the significant fines being imposed by regulators across the industry for operational risk events such as anti-money laundering (AML), product mis-selling, fraud events, and, more recently, the exchange-rates rigging scandal. Take, for example, that in the US since 2008 over $251 billion in fines were levied by US regulators; in the UK, £41 billion of fines and charges were incurred by banks in the same period.

All this poses significant challenges for governance, risk and compliance (GRC) for organizations in the financial industry and for GRC vendors. There is a dearth of IS research in this area. However, Kenneth Bamberger (2010) drew on IS perspectives in his analysis of the failures in GRC practice and related information systems leading up to the crisis. Bamberger (2010, p. 706) concluded that GRC-related IS failures were due to “problems of translation...of both legal mandates and business understandings of risk into computer code and actionable controls.” Business enterprises across industry sectors are failing to keep pace with regulatory compliance imperatives as they find it difficult “to establish what compliance means...and how to prove that they comply when asked for explanations of (non-)compliance” (Sunkle, Khokhar and Kulkarni 2015, p. 326) These problems are particularly acute in the financial industry, as Kendall (2013) points out in her presentation, Semantics in Finance: Addressing Looming Train Wreck in Risk Management, Regulatory Compliance and Reporting, to the Semantic Technology and Business Conference 2013. She argues that that current regulatory compliance approaches are ill-equipped to handle the challenges posed by the volume, velocity and variety of financial regulations.

The scale of these challenges is evidenced by the fact that JP Morgan Chase are spending over $4 billion to solve its regulatory compliance issues, while HSBC now has over 24,300 staff specializing in risk and compliance—this is almost 10% of its entire workforce. Thus, the Financial Times reports that the costs of regulatory compliance are causing concern in the industry². Kendall (2013) highlights two initiatives, based on the application of semantic technologies, being undertaken to help meet the challenges of regulatory compliance viz. “Ontologies for financial regulations (FIRO, from the University College Cork), financial contracts and transitive risk (FIBO, from the EDM Council), and aspects of XBRL are all active efforts to apply semantics to this highly complex and problematic domain. These efforts, within academic and standards bodies, and within and between individual institutions, are also highlighting some of the holes in current methods and technology required to provide much needed solutions.”

This Design Research study focuses on the development and application of FIRO—the Financial Industry Regulatory Ontology—which is part of a family of semantic technologies being developed by the FIRO research team to provide solutions to the aforementioned regulatory compliance problems. Following Winter (2008), we adopt a Design Research (DR) approach as design artefacts, such as FIRO, involve research on concepts and relationships, which are captured in a family of semantic models based on

---

extant theory and standards (Rigor Cycle, Hevner 2007). We also develop and apply novel methods in the construction and application of these concepts, relationships, axioms and related models (Design Cycle, Hevner 2007). Independent evidence of our transition through the Relevance Cycle (Hevner 2007) is given in Kendall's (2013) comments to the Semantic Technology and Business Conference 2013 cited above.

**What are Semantic Technologies?**

A Semantic technology is defined as a “software technology that allows the meaning of and associations between information to be known and processed at execution time”³. “Semantic technologies provide an infrastructure on top of which we can build intelligent applications for end-users” (Benjamins 2008, p. 76). Not only they constitute the foundation of the Semantic Web (cf. W3C 2008, 2012, 2014), Semantic technologies have been successfully applied in projects within business enterprises (cf. Declerck et al. 2007; Sheth 2005). This DR project applied W3C standard Semantic Web technologies, such as Web Ontology Language (OWL) (W3C 2012), Resource Description Framework (RDF) (w3c 2014), and SPARQL (W3C 2008), to develop our Proof of Concept. These standard technologies enabled the development the Financial Industry Regulatory Ontology during the Design Cycle by providing standardized languages that support our design requirements, which include portability and interoperability, robust semantics, and web integration. As a semantic technology FIRO is a representation that supports interoperability at both the syntactic and semantic levels. The OWL2 language has well-defined semantics that are grounded in first-order logic, set and model theory (W3C, 2012). This allows software programs to make inferences from declarative OWL expressions and assure that the subsequent interpretations are sound. An advantage for OWL over many other knowledge-based languages is that it has well-defined family members with appropriate levels of expressivity and complexity. As a semantic technology OWL might be applied for two purposes: (1) as a powerful semantic data modeling language and (2) to enable inference and reasoning over existing knowledge in a database in order to enrich it. RDF is a family of W3C specifications for the conceptual depiction and/or modeling of information, typically on the web, but it has widespread applications. The RDF data model is akin to entity–relationship (ERD) models or object-orient class diagrams. The RDF syntax is helps make statements about entities, concepts or web resources. These statements are in the form of triples, which are subject–predicate–object expressions, where the subject and object represents the concepts or resources, while the predicate describes relationships between the subject and the object. A triples store is a data store that persists RDF files. SPARQL is a recursive acronym for the SPARQL Protocol and RDF Query Language. Unlike SQL, which operates on relational data stores, SPARQL is a semantic query language for RDF triple stores. SPARQL Endpoints can also be used to query relational data stores. Putting it all together, an OWL semantic data model could enable SPARQL queries on relational data stores via SPARQL endpoints, or to query RDF triple stores. Thus OWL models enable the federation and integration of data for analysis.

At a practical level, Davies (2008) argues that a semantic technology can associate unstructured information, such as text in documents, with domain ontologies, thereby enabling information extraction into a knowledge base. He (ibid. p. 77) argues that “This makes possible more intelligent information-access facilities by annotating documents (and parts of documents) with semantic meta-information—information, formally expressed, which tells the machine what the document or subdocument is about. This allows more sophisticated analysis of documents. For example, named entity recognition is a language-processing technique that can identify particular locations, organizations, or people mentioned in textual documents with ontological descriptions of those entities (a technique known as semantic annotation... This offers numerous advantages).” There is a nascent body of research in legal informatics on the application of semantic technologies to make regulatory and legal texts machine readable. Diverse approaches are being employed, such as representing texts as linked documents, semantically tagging text, or translating texts to formal representations for inferencing and reasoning (Wyner et al. 2012). There is a notable absence of success in making regulatory and legal texts machine readable, however (cf. Wyner et al. 2012; Lesmo et al. 2013). Researchers have also proposed concepts and models of regulatory

---

texts and created conceptual models of laws and regulations (cf. Zeni et al. 2013; Zeni and Mich 2014). However, there is little evidence that such models or underlying concepts have been subjected to evaluation (Design Cycle) or been subjected to field test (Relevance Cycle), although the work of Zeni et al. (2013) looks promising. More recently, researchers in India are using semantic vocabularies in an ongoing design science research (DSR) initiative to map regulations and compliance imperatives to business processes and products (Sunthek, Kholkar and Kulkarni 2015). While initial results look promising, there are significant obstacles to overcome.

The Governance, Risk and Compliance (GRC) Technology Centre (TC) was founded by the Irish Government at the behest of the financial industry to conduct R&D on the use of semantic technologies for governance risk and compliance in the financial industry. The current DR initiative was sponsored by a number of Globally Systemically Important Banks (G-SIBs) and a leading GRC vendor to the financial industry who identified operational risk and its sub-domain of Anti-Money Laundering (AML) risk as areas that required urgent research attention. They argued that existing business intelligence and knowledge management systems (KMS) were not addressing their concerns and they therefore sought an innovative solution based on semantic technologies (cf. Declerck et al. 2007; Sheth 2005), which they argued overcame the limitations of traditional IS for GRC. The current DR study, however, breaks new ground, while also building on extant state-of-the-art research and practice, to propose a novel approach to achieve its research objective, as explained below.

Research Objective

The objective of this study is to conduct Design Research (DR) on the application and development of semantic technologies for regulatory compliance change management (RCM) in the Financial Industry. The paper describes how regulatory ontologies are being developed at the GRCTC and embedded in a prototype to enable regulatory texts to be queried in order to help GRC executives and legal practitioners answer questions such as ‘What are the various restrictions in an individual instrument of legislation or a regulatory rule?’ The prototype also to provides practitioners with the capability to query legislation and regulatory texts and identify obligations, derogations, exemptions, exclusions, etc. It is planned to develop regulatory compliance change management systems based on the outputs of this research in order to informate the development of governance policies, risk management strategies and compliance reporting in financial organizations.

The remainder of this paper is structured as follows. The next section describes our design research approach. The third section describes the research-in-progress on our nascent regulatory compliance change management system (RCMS). The final section offers conclusions and describes on-going R&D towards the completion of this project.

Design Science Research Approach to Developing and Applying Semantic Technologies

In positioning our research we look to Winter (2008, p. 471), who states that design research (DR) is aimed at “creating solutions to specific classes of relevant problems by using a rigorous construction and evaluation process.” Winter (ibid.) indicates: “design science reflects the design research process and aims at creating standards for its rigour.” We therefore classify our research-in-progress project as Design Research (DR) as it concurs with Winter’s (2008) conception of this type of research. The design artefacts being produced in this study include: (a) Constructs (i.e. concepts in an ontology); (b) Relationships between, and axioms that govern, these constructs; (b) Models (in Web Ontology Language (OWL2) represented in Protégé); and (d) Methods (an approach to the construction of concepts, relationships, axioms and models). According to Hevner (2007) design science research should include: (a) a Design Cycle, which involves the essential activities of developing and evaluating the design artefacts and research processes; (b) a Rigor Cycle, which connects the design cycle with a knowledge base of scientific theories, experience & expertise, and meta-artefacts; and (c) a Relevance Cycle, which incorporates interactions between the environment of the problem domain and the core design activities (cf. Hevner et al. 2004). Each of these cycles was incorporated into our design science research.

In our DR project described below, the Rigor Cycle was underpinned by Design Science (DS) theory based on Formalism (West 2009), which adhered to the Bunge-Wand-Weber (BWW) Ontology (Wand
Weber 1993, 1995, 2002), knowledge engineering principles, and in particular the formalisms underpinning the application of the Web Ontology Language (OWL) (W3C 2012). We also align our DR with standards published by the Object Management Group (OMG), particularly the Semantic of Business Vocabulary and Business Rules (SBVR) standard (OMG 2008) and the OMG and Enterprise Data Management (EDM) Council’s Financial Industry Business Ontology (FIBO) standard (Bennett 2011, 2013). The Semantics of Business Vocabulary and business Rules (SBVR) is an Object Management Group (OMG) specification for a Business Natural Language that is grounded in ISO Common Logic. SBVR structures natural text around concepts from the SBVR Metamodel which is expressed as an XML grammar (XML-Schema) (W3C, 2008). The Financial Industry Business Ontology (FIBO) is a family of financial domain ontologies sponsored by the Enterprise Data Management (EDM) Council. The development of FIBO is a collaborative initiative led by industry members of the EDM Council in collaboration with the Object Management Group (OMG).

The Relevance Cycle in this project includes regular feedback and demonstrations to GRC executives, lawyers, and GRC application vendors, as well as progress reports to the OMG’s Financial Domain Task Force, which includes on members of the OMG Ontology Special Interest Group, the SBVR Task Force and subject matter experts from the financial industry globally. The relevance of our theoretically informed DR project is therefore ascertained. We next outline our Design Cycle Activities.

**Design Cycle: Towards Effective Regulatory Compliance Change Management Systems**

The purpose of this design research project is to leverage semantic technologies to assist Subject Matter Experts (SMEs), be they lawyers or GRC officers or banking executives, in making sense of the wide and complex spectrum of legal documents, regulatory texts, and other rule sources in order to perform better regulatory change management.
<table>
<thead>
<tr>
<th>Research Phases</th>
<th>Activities and Description</th>
</tr>
</thead>
</table>
| 1. **Ontology Engineering** | a) Relevant Terms and Relationships Identification  
- Subject Matter Experts create a regulatory vocabulary and capture regulatory intent in a rulebook using the Object Management Group’s (OMG) Semantics of Business Vocabulary and Business Rules (SBVR) specification (OMG 2008).  
- The regulatory document structure is also analyzed in this step based on the European Union’s Akoma Ntoso Standard (UNDESA 2014).  

b) Ontology Design  
- Knowledge Engineers create a family of formal ontologies (from 1.a above). Design decisions such as mapping OWL axioms derived from the SBVR rules, are then created in the corresponding FIRO ontology. FIRO-H contains high-level concepts in the regulatory domain, such as prohibitions, obligations, derogations, etc. FIRO-S is an ontology that captures the semantics and structural description of legislative and regulative texts according to the Akoma Ntoso standard (UNDESA 2014). FIRO-AML captures the semantics and axioms for the Anti-Money Laundering (AML) domain.  
- The ontology to support a specific domain in financial industry is also created in this task, FIRO-Op (Operational or Purpose Specific, by merging all FIRO ontologies: FIRO-H, FIRO-S and FIRO-AML. Currently, FIRO-RCM is an operational ontology for Regulatory Change Management.  |
| 2. **Training**     | a) Annotation  
- In this Phase, SMEs identify, as a parameter for the prototype, the smallest unit of text that will be classified. This unit of text can be a section, a subsection or a paragraph.  
- SMEs manually annotate a limited set of regulations. SMEs identify and highlight in a regulation concepts from the Ontology, FIRO-RCM in this instance, which include domain concepts (such as Customer Due diligence, Monitoring, Training, etc.) and structural concepts (such as sections, etc.).  |
| 3. **Classification** | a) Relevant Instance Data Extraction  
- A chain of automatic classifiers are used to attach to semantic labels to units of text. One classifier suggests domain tags including AML concepts, while another suggests modalities tags including Obligations, Prohibitions, etc.  

b) Knowledge Base Population  
- When the automatic classification is complete, the knowledge base is automatically populated following the FIRO-RCM model.  |
| 4. **Verification**  | a) Competence Questions Definition  
- The first validation step is checking the consistency of the ontologies (FIRO). Then SMEs define Competency Questions (CQs) to test the prototype.  

b) Testing the knowledge base  
- Based on the CQs, Knowledge Engineers define relevant queries using SPARQL (W3C 2008).  |
| 5. **Application**   | - User Interfaces Development  
- SMEs validate if the ontology is properly answering the CQs. Prototype screens and Excel spreadsheets are used to present results (cf. Figure 3).  |

**Table 1. Completed Design Cycle Phases**

A secondary goal is to help financial institutions draft more effective governance policies, do enhanced risk management, and relevant compliance reporting. More precisely, we are combining several techniques in a system that provides the capability to answer such important but technically elusive questions such as:
• What are the compliance imperatives (obligations, prohibitions etc.) in a regulation or rule and where do they appear?

• How can semantic technologies support regulatory change management?

Figure 1 illustrates the phases of our Design Cycle methodology which have been executed in full through several iterations during 2015. The figure illustrates several innovative techniques developed by GRCTC researchers, and related semantic technologies they designed. These are combined for the purpose of developing a working prototype of a Regulatory Compliance Change Management System or RCMS. First, during the Ontology Engineering phase, Subject Matter Experts (SMEs) create a regulatory vocabulary and a reuse it to capture the regulatory intent in a rulebook. The produced output is then used by Knowledge Engineers to create a family of formal ontologies called FIRO. Second, SMEs manually annotate, using FIRO, a set of documents used as training for the automatic classification algorithms. Third, several classification algorithms are executed in order to populate the ontology, or in other words, tag the regulatory text with concepts from the ontology. Fourth, SMEs communicate to the Knowledge Engineers a series of queries. Finally, the queries are answered using content from the resulting knowledge base that is made accessible via a SPARQL endpoint to facilitate queries. Details of each phase are described in Table 1 above, with a list of activities and descriptions identifying the technology and/or the formalisms underpinning it.

![Figure 2. FIRO-Tagged excerpt from UK AML 2007 unstructured text](image)

**RCMS Prototype**

Having iterated through the Design Cycle phase of the Regulatory Compliance Change Management system (RCMS) to develop the prototype, we entered the DR Relevance Cycle where the prototype application was demonstrated to executives from the financial industry and technology sectors. To achieve this, we first illustrate the target text—The 2007 UK Money Laundering Regulation. Here in figure 2 we see an excerpt—Section 15 (tagged in FIRO-S) which has been tagged as an Obligation (FIRO-H), and which covers Anti-Money Laundering concepts such as Record-keeping, Customer Due Diligence and Ongoing Monitoring. Figure 3 illustrates the application query interface. In this case several pre-
structured queries are presented. These are submitted to the SPARQL (the OWL/RDF query language) endpoint which then returns the results. The queries have several parameters, such as to list Obligations with certain parameters attached or Obligations that relate to the AML concept of Customer Due Diligence, Monitoring, Reporting, and so on. The location of the Obligation is returned by default (e.g. Section, Sub-section, Page etc.). However, the actual text as in Section 15 above may also be presented. Figure 3 also illustrates the result of the query in an Excel spreadsheet for further analysis. Here all the AML categories which carry an Obligation are presented, as is their Section and Sub-section. The text describing the obligation is next displayed.

Feedback from the financial industry and technology sectors on the first phase prototype, as illustrated in Figures 2 and 3, was extremely positive. With minimum re-training the application was applied to the US Bank Secrecy Act, 31 CFR B Chapter X, which covers AML. We were more than pleased with the query output, which contained, as early subject matter experts conducted tests revealed, a high percentage of accurate results in the form “units of text” tagged with appropriate concepts from FIRO-RCM. This bodes well for the future uptake of our research by industry, which we expect in the first quarter of 2016.

Conclusion and Future Work

This research-in-progress paper describes our Design Research (DR) on the key components of a regulatory compliance change management system (RCMS) for the financial industry. The semantic technologies—the FIRO family of OWL models—which form the core components of our nascent RCMS, and which are the products of our DR, can help reduce this burden by enabling financial organizations to query complex regulatory texts, as unstructured data. As indicated, the core design feature of these semantic technologies involves the implementation of an approach to ontology-enabled semantic tagging of regulatory documents and texts by identifying the embedded obligations, prohibitions, derogations, etc. This helps reduce the burden of identifying and understanding the compliance implications of complex regulations. This is a non-trivial problem for which viable, cost-effective solutions do not currently exist.

One of the major contributions of our Design Research is undoubtedly in the innovative processes described in our Design Cycle and captured in Figure 1 and briefly described in Table 1. This is, we believe,
an exemplar of how to go about developing IT artefacts based on semantic technologies for the purpose of creating knowledge bases to classify, store and query unstructured data. Likewise, the design or articulation of the Financial Industry Regulatory Ontology is, we believe, equally unique in the field of legal informatics. The overall Design Research approach taken in this study ensured that the Design Cycle was tightly integrated with both Rigour and Relevance Cycles. In terms of the Rigour Cycle, the research at the GRC Technology Centre is the subject of external, independent, peer academic review, to ensure coherence with extant theory and research. Likewise, research presentations made at Object Management Group Technical Meetings ensure that compliance with industry standards are adhered to. In terms of the Relevance Cycle, regular presentations to the industry members of the Technology Centre, and the broader financial industry ecosystem in the US and Europe, helped ensure the relevance of the semantic technology-based RCMS being developed.

The practical outcomes of our design research initiative are extremely positive. In the first half of 2015, our ongoing work addressed the challenge of improving and calibrating the RCMS underlying ontologies and classification algorithms in terms of their scope, accuracy and granularity. In this we are achieving a significant measure of success. To provide automated support for the improvement process, we are designing and developing a set of user interfaces for interactive data curation by legal and financial SMEs which can then be used for training the classification algorithms. The RCMS prototype for the Anti-Money Laundering domain will be completed by December 2015. Once processed a yet-to-be developed software algorithm will be used to compare existing and new regulations to identify changes to existing, and new, compliance imperatives. This comparison is enabled and driven by the semantic tags that were attached to the original regulatory text following an implementation of our DR approach illustrated in Figure 1 and Table 1.

**Future Work**

We are in the process of extending the findings of our DR to the parent domain of operational risk in early 2016. Here, we are developing a Financial Industry Regulatory Ontology (FIRO) for Operation Risk (OR) called FIRO-OR, expressed in OWL2, using WebProtégé (Tudorache et al. 2015). We intend to use FIRO-OR to extend our nascent RCMS to a critical area of interest for both regulators and organizations in the financial industry. In terms of such future applications, our semantic technologies can be calibrated, through the application of domain-specific taxonomies and ontologies, to query (as opposed to a simple word search) unstructured texts for specific categories of risk data. This DR study is a vital first step in providing a solution to the problem of querying unstructured data for risk management and regulatory compliance reporting.

**Acknowledgement**

This research was by Enterprise Ireland and the IDA Technology Centre Grant number TC-2012-0009.
References

W3C. 2012. “OWL 2 web ontology language”.