12-14-2016 12:00 AM

Regulatory Compliance-oriented Impediments and Associated Effort Estimation Metrics in Requirements Engineering for Contractual Systems Engineering Projects

Md Rashed I. Nekvi, *The University of Western Ontario*

Supervisor: Prof. Nazim H. Madhavji, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Computer Science

Abstract

Large-scale contractual systems engineering projects often need to comply with a myriad of government regulations and standards as part of contractual fulfillment. A key activity in the requirements engineering (RE) process for such a project is to elicit appropriate requirements from the regulations and standards that apply to the target system. However, there are impediments in achieving compliance due to such factors as: the voluminous contract and its high-level specifications, large number of regulatory documents, and multiple domains of the system. Little empirical research has been conducted on developing a shared understanding of the compliance-oriented complexities involved in such projects, and identifying and developing RE support (such as processes, tools, metrics, and methods) to improve overall performance for compliance projects. Through three studies on an industrial RE project, we investigated a number of issues in RE concerning compliance, leading to the following novel results:(i) a meta-model that captures artefacts-types and their compliance-oriented inter-relationships that exist in RE for contractual systems engineering projects; (ii) discovery of key impediments to requirements-compliance due to: (a) contractual complexities (e.g., regulatory requirements specified non-contiguously with non-regulatory requirements in the contract at the ratio of 1:19), (b) complexities in regulatory documents (e.g., over 300 regulatory documents being relevant to the subject system), and (c) large and complex system (e.g., 40% of the contractual regulatory requirements are cross-cutting); (iii) a method for deriving base metrics for estimating the effort needed to do compliance work during RE and demonstrate how a set of derived metrics can be used to create an effort estimation model for such work; (iv) a framework for structuring diverse regulatory documents and requirements for global product developments. These results lay a foundation in RE research on compliance issues with anticipation for its impact in real-world projects and in RE research.

Keywords

Systems engineering, requirements engineering, artefacts, impediments, case study, regulatory compliance, metrics, effort estimation.

Co-Authorship Statement

There are several co-authors of the published papers that are related to this dissertation written in an integrated-article format. The candidate of this dissertation is a primary author of all the thesis related published papers except one (appears in Chapter 5). The candidate planned, designed and conducted the studies, and analysed and interpreted results, and wrote and published papers (included as thesis chapters) under the close supervision by Prof. Nazim H. Madhavji. As my thesis supervisor, Prof. Madhavji's role has been active and involved in all phases of the research: from inception of research ideas to providing continuous guidance in planning, designing and conducting the studies, making decisions regarding scoping out scientific papers, and editing and revising those papers. He is co-author of all the published papers relevant to this dissertation (Chapters 2-5).

Besides, there are other researchers who co-authored in the papers appearing as thesis chapters (partly or as a whole). Dr. Remo Ferrari (a former Ph.D. student at UWO) and Dr. Brian Berenbach (a senior industrial researcher) are co-authors of the two short versions of papers on which later an extended version is produced (Chapter 2) and a journal paper is published (Chapter 3). Dr. Ferrari assisted in data analysis, study design, and editing the papers while Dr. Berenbach helped in obtaining data and also reviewed the papers. Ibtehal Noorwali (a current Ph.D. student at UWO) is a co-author of the short version of the paper for Chapter 4. She helped in bringing related literature and data analysis. In Chapter 5, the candidate assisted in designing a framework for structuring regulatory requirements, and analysed the necessary data to illustrate the framework and wrote the corresponding parts in the paper. Dr. Maria Spichkova (a lecturer at RMIT, Australia) is the primary author who wrote parts of the paper (Chapter 5), and Dr. Heinrich W. Schmidt (a faculty member at RMIT, Australia) is another co-author who reviewed the paper included in Chapter 5.

Acknowledgments

First and foremost, I would like to thank my parents for their relentless motivation, love and sacrifice that has been a great inspiration for me to continue my work.

Most importantly, I would like to sincerely thank my supervisor Prof. Nazim H. Madhavji for his constant guidance and tireless supervision. I am truly indebted for his strong motivation, valuable support, hard work and inspiration. Without his guidance and cooperation this work would have not been possible.

I gratefully acknowledge the support of all my former and present colleagues, members of the thesis advisory team, faculty members and staffs of the Department of Computer Science, UWO. In particular, I am grateful to Dr. Remo Ferrari for his invaluable suggestions in publishing several research papers and Ibtehal Noorwali for her timely cooperation in improving part of the thesis writing.

I would also like to thank Dr. Brian Berenbach for his help in obtaining data, and reviewing our work.

I am extremely thankful to my wife, Fabi, and our son Raatib for their immense patience and support during the long days and nights I spent doing this work.

I would like to thank all my friends and family members who were always beside me during my difficult times; special thanks to my local friends Dr. Ahmed Ishtiaque Amin Chowdhury, Dr. Rezwanul Quddus, Dr. Jahirul Mazumder, Dr. Bahlul Haider, Md Sayef Ishaque, and their respective family members.

Table of Contents

List of Tables

List of Figures

Glossary of Terms

Component requirements: Requirement s that apply to a particular component of a subsystem.

Contract: The term 'contract' refers to the legally binding agreement regarding system requirements, recorded in writing, between the customer organization and the development company.

Cross-cutting requirements: Requirements that apply to more than one subsystem or component.

HLRS: Stands for *high-level requirements sections* which refer to those sections of the contract document that contain customer requirements at a high-level.

IPRS: Stands for *implicit project requirements specification* which is a requirements artefact containing requirements that are not explicitly specified by the contract but are imposed by applicable standards, regulations or other external sources.

PERS: Stands for *project execution requirements specification* which is a requirements artefact containing the requirements related to certain processes or activities to be followed during development of the system.

PRS: Stands for *project requirements specification* which is a requirements artefact that contains project requirements to be used by those who work in the later phase of the development cycle such as designers, architects, developers, testers, etc.

Regulations: A regulation refers to executive order, announced by legislative authority (e.g., government, industry) having force of law, which may impose financial or criminal penalties to the respective person or organization.

Regulatory documents: Regulations and standards are commonly named as regulatory documents.

Regulatory requirements: Requirements that are expected to comply with certain requirements described in regulatory document.

Requirements artifacts: The documents (or their instantiation) used and created in requirements engineering.

RSRS: Stands for *referenced standards and regulations sections* which refer to those sections of the contract document that enlist the names of applicable standards and regulations for the system.

SR proxy: Stands for *standards and regulatory proxy* which is a relevant excerpt of the original regulation or standard as needed in a particular context.

Standards: A standard is an established norm or requirement. It is usually a formal document that establishes uniform engineering or technical criteria, methods, processes and practices. It has no legal power on its own.

Subsystem requirements: Requirements that apply to a particular subsystem of the system.

System requirements: Requirements that apply to the entire system.

RE compliance project: A RE project that mainly deals with regulatory requirements and their compliance to relevant regulatory documents.

Requirements-compliance: Demonstrating regulatory compliance of requirement at RE times.

Chapter 1

1. Introduction

During the last several decades, ensuring compliance of software-intensive systems has become an important issue in software engineering (SE). With the evolution of information technology, there has been a rapid increase of complex socio-technical systems (e.g., aviation, railway, telecommunications, banking, and healthcare) that combine systems, software, people, and numerous organisations (e.g., industry, regulatory authority, business, and government) (Perini et al., 2011). In order to ensure security, privacy, safety and other qualities that concern people using such socio-technical systems, governments around the world are increasingly enacting newer regulations (e.g., HIPAA[1] (1996) for protecting the privacy and security of medical information, Sarbanes-Oxley Act[2] (SOX) (2002) for protecting investors from accounting fraud by corporations, PIPEDA[3] (2000) for ensuring individual data privacy in commercial businesses, etc.) to specify system requirements that must be followed (Qureshi, 2007). Also, organisations include directives in their business contracts for certain engineering standards to be complied with by contracted systems (Berenbach et al., 2010).

Ensuring compliance is expensive work (Siena et al., 2009), e.g., the estimate of annual costs (2005) to ensure compliance with only Sarbanes-Oxley Act (SOX) by US organisations was $5.8 billion[4]. On the other hand, non-compliance to applicable regulations results in costly monetary penalties and even imprisonment (Otto and Anton, 2007). It also damages reputations of the organisations for developing non-compliant products or sys-

---

[1] https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/html/PLAW-104publ191.htm

[2] https://www.gpo.gov/fdsys/pkg/PLAW-107publ204/html/PLAW-107publ204.htm

[3] https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

[4] Online news published in DMReview.com, Nov 2004

tems that suffer from deficiencies in quality attributes such as safety, performance, security, and privacy. Thus, ensuring regulatory compliance becomes an integral part of software and systems engineering.

Requirements engineering (RE) plays a vital role in ensuring system compliance since a system's successful implementation depends largely on the success of the RE process (Damien et al., 2005). It is during this early phase of the development cycle where system requirements described in regulations and standards must be identified and elicited for implementation. However, eliciting system requirements from applicable regulations entails thorough analysis of regulatory texts by personnel who are acquainted with its domain. Yet regulatory texts have been found to have properties such as domain-specific contents, numerous cross-references (often cyclic), overlapping and conflicting codes to other texts, and abundant use of conditional terms (e.g., if, else, else if, except, unless, etc.) that significantly complicate the requirements elicitation task, often leading to an ambiguous understanding of requirements (Otto and Anton, 2007).

In this regard, several encoding and modelling techniques, including mark-up based representations (Kerrigan and Law, 2003), goal modeling (Massacci et al., 2005), and logical modelling (Antoniou et al., 1999), have been proposed to model regulatory text into some form of logical format as a means to facilitate the elicitation of requirements from regulations and validation of requirements for compliance. Breaux et al. (2006), among others, propose an elicitation technique for privacy requirements, which can extract '*rule statements*' (e.g., stakeholder rights, obligations, privileges) - indication of requirements - from regulatory text that are restated in restricted format using a 'semantic parameterization process' (Breaux and Anton, 2005). Several validation techniques are proposed for checking requirements for regulatory compliance, which include: (i) detection of prohibited rules included in requirements by checking requirements against a semantic representation of regulations based on a concept called '*case frame*' (i.e., verb and semantics of words that frequently co-occur with the verb) (Saeki and Kaiya, 2008); and (ii) production rule model (Maxwell and Anton, 2009) where regulatory rules are encoded using '*production rules*' (i.e., "if-then" structure) to be checked against input requirements for validating compliance.

An adaptability framework is proposed in (Maxwell et al., 2012) to deal with regulatory evolutions, which predicts changes in regulations based on three components, i.e., rationale behind changes (e.g., changes in another regulation, ambiguities and redundancy), taxonomy for types of changes (e.g., reorganization, introduced cross-references, deletion), and adaptability heuristics. Ghanavati et al., (2014) propose a goal-oriented method to handle issues with multiple regulations that partially overlap or even conflict; it compares and links between diversified goals of overlapping regulations and business organisations. In (Fernandez and Yiman, 2015), the authors describe an approach to collect analogous aspects (e.g., security mechanism to protect privacy) that may exist across multiple regulations in order to build reference architectures of analogies. Such reference architectures would enable repeated use by requirements analysts. To reduce dependence on manual analysis by domain experts, Sunkle et al., (2016) present a machine learning approach for identifying regulatory rules (i.e., indications for requirements) from regulations. It uses a semi-supervised technique that can learn from two given sets of regulatory sentences: rule sentences (regulatory rules) and non-rule sentences.

Nonetheless, these methods and techniques are still in the theoretical stage without having enough empirical validation in industry that would allow enterprises to start using them in practice. Consequently, a current industrial practice for compliance work in RE is significantly manual, ad-hoc (non-standardised), unpredictable and time-consuming.

More specifically, in contract-based, large-scale systems engineering projects that need to ensure compliance with a multitude of regulations and standards as part of contractual obligations, RE tasks such as identifying relevant regulatory documents, eliciting described requirements from them, managing changes of regulations during project lifecycle and system operation, and demonstrating contractual fulfillment regarding compliance becomes a monumental task (Berenbach et al., 2010). Further, such project work entails immense uncertainty and risk due to the lack of shared understanding of the complexity underlying the variety of artefacts used (e.g., diverse regulatory documents, contracts, project requirements, etc.) and the compliance-oriented inter-relationships existing among them in the projects. However, there is not much grounded theory on the different types of artefacts used, the characteristics of their inter-relationships, and the challenges to ensure regulatory compliance of requirements in a RE compliance project. This shortfall is deemed to cause variability and quality problems in RE projects. Another implication of this lack of understanding is that such large-scale projects become prone to underestimation of effort required to do compliance work in RE.

1.1 Research Problem

Enterprises that are facing compliance problems in RE need answers to two complementary questions: i) what are the complexity and impediments to achieving regulatory compliance of requirements in contractual systems engineering projects? and ii) how can the compliance-based project complexities in RE be estimated?

Academic research aiming to ensure RE compliance has two levels, where findings of the first level motivate and guide the second level. The first level encompasses research that seeks detailed shared understanding of the characteristics and associated challenges of the RE compliance project. Research at the next level comprises essential solution schemes (e.g., metrics, techniques, tool support, etc.) to approximate project complexity, which is guided by the body of knowledge gathered in the first level.

1.2 Research Contribution

In this thesis, we describe multiple empirical studies conducted in an industrial RE project that is part of a contract-based large-scale systems engineering project aiming to upgrade a railway system. The RE project is broadly characterised by: (i) a 1000+ page contract document describing approximately 12,000 customer (government) requirements; (ii) over 300 regulatory documents to which the target system must comply; and (iii) a multi-domain system consisting of seven major subs-systems with numerous components.

The goal of our study was set to explore the characteristics and challenges of RE in a contract-based, large-scale systems engineering project from a compliance viewpoint in order to form a foundation for developing corresponding RE methods, processes, metrics, models, and technological support. With regards to the goals, the contributions of this thesis are as follows:

a) Identification of the artefact-types (e.g., contract, set of regulatory documents, derived project requirements, and implicit project requirements (see Section 2.4.1)) and their types of inter-relationships, such as "*reference-to*" from contract to regulatory documents, "*impose*" from regulatory documents to implicit project requirements specifications, "*Is-derived-from*" from contract to project requirements specifications, that exist in large-scale contractual systems engineering projects (see Section 2.4.2). These artefact inter-relationships are also characterised in quantitative terms (see Section 2.4.3) (Nekvi et al., 2011).

b) Construction of a '*compliance meta-model for RE*' that depicts artefact inter-relationships and their characterisations based on empirical findings (see Section 2.4.4) (Nekvi and Madhavji, 2015). This meta-model would act as a domain guide for anyone interested in compliance-oriented RE projects.

c) Identification and analysis (both quantitative and qualitative) of a number of novel impediments (e.g., non-contiguity and abstractness of regulatory requirements in a contract; large set of regulatory documents and their voluminous contents and cross-references; and cross-cutting requirements) to ascertaining compliance of system requirements (see Section 3.6) through a case study of a contractual systems engineering projects (Nekvi et al., 2012; and Nekvi and Madhavji, 2015).

d) Development of a method to derive key metrics for effort estimation of requirements compliance work, see Section 3.4 (Nekvi and Madhavji, 2016). Using the method, we derive a number of key metrics such as: (i) size-metrics (e.g., percentage of regulatory requirements that are cross-cutting, number of sections having cross-references, and number of requirements having diverse references), (ii) project complexity metrics (i.e., number of sub-system teams and avg. number of components per subsystem), and (iii) metrics for process and product characteristics (e.g., average number of cross-references per segment of regulatory documents, average ratio of regulatory and non-regulatory requirements per page) (see Section 4.4). Then, we demonstrate how the derived metrics can be used in an algorithmic model-based effort estimation technique (see Section 4.5).

e) For managing the diversity of regulations and requirements of global products, we propose a framework for structuring the regulatory requirements for global products (see Section 5.4) (Spichkova et al., 2015).

f) We propose two emerging theories that generalise the findings of the thesis: (i) *characteristics of requirement artefacts and their inter-relationships existing in the contractual systems engineering projects underlie substantial impediments to doing compliance work in RE;* and *(ii) key metrics for an effort estimation model for compliance work of RE are derivable through analysis of impediments associated with compliance work of RE* (see Chapter 6).

1.3 Thesis Structure

The contributions of this thesis are organised into four core studies that are structured into four discrete chapters (i.e., Chapter 2-5) in the thesis, which are shown in Table 1-1. With reference to Table 1-1, the title of the core studies is shown in the middle column, the corresponding chapter numbers where these studies are described are shown in the first column, and the last column shows corresponding publications with publications year and venue.

The first two studies (presented in Chapter 2 and Chapter 3) are case studies performed on an industrial RE project (part of a contractual systems engineering project) aiming to ensure regulatory compliance of requirements of a railway system. The system contained both software and systems requirements.

In Chapter 2, the first study identifies and characterises (both qualitatively and quantitatively) types of artefacts (e.g., contract, regulations, standards, sub-system requirements, component requirements, and cross-cutting requirements) and their inter-relationships (e.g., contract reference to standards and regulations, contract derives regulatory requirements, standards and regulations impose implicit regulatory requirements, and project requirements complies with regulations) that exist in the project. Based on the empirical evidence, we develop a compliance meta-model that depicts such characterisations.

In Chapter 3, the second study further investigates into the project to identify and characterise a number of impediments (e.g., large number and size of regulatory documents, non-contiguity of regulatory requirements in the contract, abstractness of requirements in the contract, cross-references in regulatory requirements, cross-cutting nature of regulatory requirements) to regulatory compliance of requirements due to complexity of contract, large set of regulatory documents, and scale and complexity of the system.

Table 1-1 Thesis core

Chapter #	Study Title	Publications
2	Empirically Derived Compliance Meta-Model for System Requirements	RELAW (Nekvi et al., 2011)
3	Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects - A Case Study	REFSQ (Nekvi et al., 2012) ACM Trans. on MIS (Nekvi and Madhavji, 2015)
4	Metrics for Estimating the Effort Needed In Requirements Compliance Work	REFSQ (Nekvi et al., 2016)
5	Structuring Diverse Regulatory Requirements for Global Product Developments	RELAW (Spichkova et al., 2015)

In Chapter 4, based on scientific insights gained by the two studies and analysis of the impediments, we develop a method to derive key metrics to estimate the effort needed to perform compliance-related RE activities that are effort-critical (i.e., characterised by impediments). Using this method we derived the key metrics (e.g., number of requirements having diverse regulatory reference, average ratio of regulatory and non-regulatory requirements per page of contract, number of regulatory documents relevant to system) for effort estimation for the effort-critical RE activities identified in the studied project. We also explain how these metrics can be used in developing an algorithmic-based effort estimation model for compliance work in RE.

In Chapter 5, we propose a framework to structure and manage regulatory requirements and regulatory documents of products that operate in multiple jurisdictions (i.e., country). We illustrate the usability of this framework by examples collected from the project where we performed the earlier studies (Chapter 2-4).

In Chapter 6, following the above studies, we propose two descriptive emerging theories which are empirically grounded on the evidence from the earlier studies. The theories describe the relationships between: (i) the characterisations of artefacts and their inter-relationships and impediments to regulatory compliance of requirements, and (ii) the impediments and their associated metrics for estimating the effort needed in doing compliance work of RE for contractual systems engineering projects. These theories are validated against criteria for goodness of theory proposed by (Sjøberg et al. 2008) and constitute the core component of this thesis.

This thesis is structured in the integrated-article format[5]. With reference to this format, each study is presented in separate "middle" chapters, i.e., Chapter 2-5. Each of these "middle" chapters, thus, contains its own introduction, literature review, research procedure, results and discussions, and bibliography. In addition, Chapter 6 describes the emerging theories that are a generalisation and abstraction of the previous findings (Chapter 2-5). Finally, Chapter 7 concludes the thesis with directions for future work.

1.4 References

Antoniou, G., Billington, D., Governatori, G., and Maher, M. J., On the Modelling and Analysis of Regulations. Australian Conference on Information Systems, pp. 401–405, 1999.

[5] http://grad.uwo.ca/current_students/regulations/8.html

Berenbach, B., Lo, R. and Sherman, B., Contract-based Requirements Engineering. Int. Workshop on Requirements Engineering & Law, IEEE Comp. Soc., pp. 27-33, 2010.

Breaux, T. D., and Antón, A.I., Deriving Semantic Models from Privacy Policies. 6th IEEE Int. Works. on Policies for Distributed Systems and Networks (POLICY), pp. 67-76, 2005.

Breaux,T.D., Vail, M.W., and Antón,A.I., Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations.14th RE Conf., pp. 49-58, 2006.

Fernandez, E.B. and Yimam, D.,Towards Compliant Reference Architectures by Finding Analogies and Overlaps in Compliance Regulations. E-Business and Telecommunications (ICETE), SCITEPRESS, vol. 4, pp. 435-440, 2015.

Ghanavati,S., Rifaut, A., Dubois, E. and Amyot, D., Goal-oriented Compliance with Multiple Regulations.IEEE 22nd International RE Conference ,pp. 73-82, 2014.

Kerrigan, S., and Law, K. H., Logic-based regulation compliance-assistance. ACM 9th International Conference on Artificial Intelligence and Law, Scotland, pp. 126-135, 2003.

Massacci, F., Prest, M. and Zannone, N., Using a Security Requirements Engineering Methodology in Practice: the Compliance with the Italian Data Protection Legislation. Computer Standards & Interfaces, Vol. 27(5), pp.445-455, 2005.

Maxwell, J. C. and Antón, A. I., Checking Existing Requirements for Compliance with Law Using a Production Rule Model. Int. Work. on RELAW, IEEE Com. Soc., pp. 1-6, 2009.

Maxwell, J.C., Antón, A.I. and Swire, P.,Managing Changing Compliance Requirements by Predicting Regulatory Evolution. 20th IEEE Int. RE Conference, pp. 101-110, 2012.

Nekvi, M.R., Ferrari, R., Berenbach, B., Madhavji, N.H., Towards a Compliance Meta-model for System Requirements in Contractual Projects. 4th Int. Workshop on Requirements Engineering and Law (RELAW), IEEE Computer Society, CA, USA, pp. 74-77, 2011.

Nekvi, M.R., Madhavji, N.H., Ferrari, R., Berenbach, B., Impediments to Requirements-Compliance. Int.Conf. on REFSQ, Springer Berlin Heidelberg, pp. 30-36, 2012.

Nekvi, M.R.I., and Madhavji, N.H., Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects: A Case Study. ACM Transactions on Management Information Systems (TMIS), 5(3), p.15, 2015.

Nekvi, M.R.I., Noorwali, I. and Madhavji, N.H., Deriving Metrics for Estimating the Effort Needed in Requirements Compliance Work. Int. Working Conf. on REFSQ, Springer International Publishing, pp. 135-141, 2016.

Otto, P. N. and Antón, A. I., Addressing Legal Requirements in Requirements Engineering. Int. Requirements Engineering Conference, IEEE Comp. Society, CA, USA, pp.5-14, 2007.

Perini A, Qureshi N, Sabatucci L, Siena A, Susi A, Evolving Requirements in Socio-technical Systems: Concepts and Practice. Int. Con. on CM., Springer, pp. 440-447, 2011.

Qureshi, Z.H., A Review of Accident Modelling Approaches for Complex Socio-technical Systems.12th Aust. Conf. on Safety-Related Programmable Systems,, pp. 47 – 60, 2007.

Saeki, M., and Kaiya, H., Supporting the Elicitation of Requirements Compliant with Regulations. 20th Int. Conf. Advanced Information Systems Engineering (CAiSE), Springer-Verleg, Heidelberg, Germany, pp. 228 - 242, 2008.

Siena, A., Mylopoulos, J., Perini, A., Susi, A., The Nomos Framework: Modelling Requirements Compliant with Laws. Technical Report 0209-SMSP, FBK-Irst, 2009.

Sjøberg, D., Dyba, D., Anda, B. C., and Hannay. J., Building Theories in Software Engineering. Guide to Advanced Empirical Software Engineering, Shull, F., Singer,J., and Sjøberg, D.I.K, Eds., Springer, London, pp. 312–336, 2008.

Spichkova, M., Schmidt, H.W., Nekvi, M.R.I. and Madhavji, N.H., Structuring Diverse Regulatory Requirements for Global Product Development.IEEE International Workshop on Requirements Engineering and Law (RELAW), pp. 57-60, 2015.

Sunkle,S., Kholkar,D., Kulkarni, V., Informed Active Learning to Aid Domain Experts in Modeling Compliance. IEEE 20th Conf. on EDOC, pp. 1-10, 2016.

Chapter 2

2. Empirically Derived Compliance Meta-Model for System Requirements[1]

2.1 Introduction

A large system involving a number of engineering domains (e.g., civil, electrical, mechanical, rail, software, communications, etc.) is often required to be compliant with a myriad of engineering standards and government regulations. To support compliance, requirements imposed by standards and regulations need to be elicited and be incorporated into project requirements.

Non-compliance can result in non-standard products or those suffering from quality problems (e.g., security, safety, reliability, etc.). Non-compliance can also lead to violation of the law and contract which, in turn, can lead to financial penalties and criminal charges. For instance, a corporation that fails to comply with a Canadian Environment Act 1999 is charged, for a first offence, a fine of between $75,000 and $4,000,000 and, for a second or subsequent offence, the fine is doubled (Environment Canada, 1999)

To avoid such penalties, requirements analysts need to ensure that project requirements comply with the relevant standards and regulations. Manually, assurance of such compliance is a non-starter for large and complex projects. However, methods and tool support would not – by themselves – avoid non-compliance. There is thus a need for domain knowledge on artefacts-types and their inter-relationships so that humans who use the methods and tools can ensure compliance.

Many RE compliance projects try to ensure that they do not violate the contract (Berenbach et al., 2010), which specifies high-level customer requirements as well as the standards and regulations to which the target system must demonstrably conform. System

[1] A shorter version of this chapter was published in (Nekvi et al., 2011).

requirements for these projects must be derived from the contract as well as from the applicable standards and regulations. However, the regulatory documents, which cross-reference various standards and also among themselves (Kerrigan and Law, 2003) are also being referenced from within the contractual documents and system requirements specifications. Thus, in large system engineering projects, a highly complex inter-relationship network exists among the requirements and related artefacts. In such situations, a meta-model (a map of sorts) depicting the various types of artefacts and inter-relationships is invaluable because it can act as a domain guide in the RE process.

While a meta-model exists to support RE compliance for eliciting requirements from legal artefacts (Siena et al., 2009), there is a distinct lack of two key things in a systems engineering context where there are many regulatory codes and standards. One is a shared understanding of the types of artefacts (and their inter-relationships) involved in a RE compliance project and the other is a meta-model that holds all these type-level items together as a unified structural and semantic entity. Thus, our research goal is to characterise the numerous types of artefacts and inter-relationships and create a meta-model.

We had an opportunity to explore these two needs in a large-scale case study in industry involving a contractual document exceeding 1000 pages, approx. 300 engineering standards and approx. 30 regulatory documents. They highlight the synergy between traditional RE processes and compliance issues in systems engineering (specifically a railway system).

Two specific results emanate from this case study: (i) the different types of artefacts used and the characteristics of their inter-relationships, in a requirements engineering (RE) compliance project involving regulatory documents and standards and (ii) a meta-model for RE compliance based on these artefacts and inter-relationships. The meta-model is created from data gathered and observations made in the case study. This model can be useful in understanding and managing a requirements-compliance project. The findings and resultant meta-model have implications for: RE compliance support, requirements traceability tools, requirements elicitation, and further empirical work.

The rest of this chapter is organized as follow: Section 2.2 gives a brief overview of related work; Section 2.3 discusses the empirical study; Section 2.4 presents the study results and interpretations; Section 2.5 discussed the threats to validity of the results; Section 2.6 discusses the implications of this research; and finally, Section 2.7 closes this chapter with future work and conclusions.

2.2 Related Literature

There is a fundamental difference between the concepts in which regulations are expressed, and how requirements are defined (Siena et al., 2008). Regulations describe rights, obligations, privileges, and liabilities (Hohfeld, 1913), whereas requirements represent stakeholders' goals or needs. Legal texts tend to be well structured and are organized hierarchically (Otto and Anton, 2007). Regulatory text often has inter and intra-cross referencing between sections of regulatory document (Kerrigan and Law, 2003) and the presence of many conditional statements (i.e. containing conditional terms such as if, unless, except, only if) reduces the readability of the documents and often creates ambiguity. In addition, regulatory documents contain domain specific terms which are not necessarily used in the RE community (Kerrigan and Law, 2003). Furthermore, laws are dynamic in nature with frequent changes by amendments (Kiyavitskaya et al., 2008; Otto and Anton, 2007; and Penzenstadler and Leuser, 2008). Therefore, process support is necessary to manage requirements with the evolution of laws. New laws enforced by different regulative authority such as federal, provincial (or state) or local level may introduce contradictions, overlapping or duplication of contents (Kitchenham, 2004). Identification of an applicable portion from diverse set standards and regulations, and eliciting appropriate requirements from the tedious details of these documents offer unique challenges to requirement engineers (Kerrigan and Law, 2003).

A number of approaches have been proposed for analysing and modeling regulatory text (see (Otto and Anton, 2007) for a summary) which have been built upon a variety of legal text encoding techniques. A logical model provides unambiguous and strict format to content that enables machine certain operational facilities such as structured querying, searching of particular items of interest, categorization of contents into defined classification . However, the encoding of legal text is tedious and manually time consuming (Saeki and Kaiya, 2008); and incredibly difficult for large-scale project having fair number of regulatory documents to comply with.

Jureta, et al. (2010) presents a formal theory for regulatory compliance which describes: (i) what it means for requirements to be compliant, (ii) how analysts can verify that requirements are compliant, and (iii) testable hypotheses regarding the verifiability of compliance. Siena, et al. (2009) has developed a meta-model to derive law-compliant requirements that take into account both the law in a given domain and the stakeholders' goals. Islam et al. (2010) propose a framework to align existing requirements with changes in law. Breaux et al. (2006) propose a methodology (and supporting tool) to semi-automatically extract requirements from regulatory text.

Saeki and Kaiya (2008) proposed a compliance validation technique where the semantics of the regulations are represented using case frames and requirements statement are checked against the case frame for compliance. Maxwell and Anton (2009) propose regulation modeling technique using production rules where each rule is an if-then structure. The model takes existing requirements as input and outputs a set of validated requirements compliant with the regulations or any additional candidate requirements necessary to ensure compliance.

These research works focus on the problems rooted in the structure of legal and regulation texts, and their alignment with requirements. However, for a large RE project that has numerous standards and regulations to be complied with, an explicit understanding of the interconnections among the various types of artefacts is needed. This is where the RE compliance meta-model we have developed fits it.

The relationships among these artifacts (i.e. also called an artifacts model or a meta-model) convey information which is useful for planning the project, and also for defining RE processes, methods and tools to automate the compliance process (Berenbach et al., 2009).

2.3 The Empirical Study

In this section, we describe the empirical study. This includes: an overview of the project that was investigated; the research goals and questions; data collection and analysis procedures, and threats to the validity of the results.

2.3.1 Study context: rail infrastructure upgrade project

The studied case is a large RE project where the primary goal is the development of appropriate requirements to upgrade a rail corridor infrastructure system. According to the project's personnel, the project has been in existence since late 2008 and is expected to continue till the year 2014. The infrastructure of this rail corridor spans over a sizeable geographical area in North America, consisting of a complex network of tracks, passenger platforms and interlocking at major streets, approximately 200 signals and over 250 switch machines, and many kilometers of circuited track and associated infrastructure.

It is a multi-disciplinary, engineering-based domain consisting of rail, civil, electrical, mechanical, chemical, software, and communication. The overall system consists of several major subsystems (such as, network management subsystem, power supply subsystem, civil structures subsystem, signalling subsystem, and others), where each subsystem is composed of a number of components. For example, the major components within the signalling subsystem are Switch, Cables, Interlock, Circuit, and Relay. The case study focuses on requirements that deal with the whole system with no particular focus on any one subsystem. Also, the requirements deal with multi-disciplinary domains described above. It is not clear at this stage, what percentage of the total requirements are software-intensive; however, with over 12,000[2] requirements in all, one can easily imagine that even a meager 10% is over one thousand requirements.

[2] These 12,000 requirements are at this early stage of the project; these are anticipated to increase significantly through the project targeted to end in the year 2014.

2.3.1.1 Safety-critical domain

The system facilitates passenger and cargo transportation in a busy metropolitan area and is thus considered highly safety-critical. Any defect in the system leading to a system failure can potentially have devastating consequences (e.g., a train crash and its impact on the public). It is imperative that the RE project capture, document and trace critical requirements, and complies with the stipulated regulations and standards.

2.3.1.2 Role of contract

The RE project is bound by a legal contract, which is a primary source for the elicitation of the system requirements. The contract is a hierarchically structured document separated into ten broad divisions (e.g., mechanical, electrical, software, construction, metals, etc.). Each division is also hierarchically structured -- into sections, subsections, clauses, and sub-clauses, which we refer to as 'segments' in this study.

The contract is in excess of 1,000 pages. It describes the mentioned 12,000 requirements at a high-level. These are referred to as contractual requirements and encompass both functional and quality aspects. There are "other" requirements, referred to as project requirements, which are "derived" either from the contract or regulations and standards that the overall system is subjected to. Because the contractual requirements are at a high-level, they are generally not testable; whereas, the project requirements are meant to be testable. Also, an unspecified portion of the contractual requirements reference, explicitly, approximately 300 engineering standards and 30 government agency regulations (e.g., federal, provincial, city, etc.) against which the project must demonstrate compliance. This particular subset of requirements is referred to as regulatory requirements. An important reason to single out this subset of requirements is that, failure to comply against non-regulatory requirements can lead to issues with the customer; whereas, that against regulatory requirements can lead to legal issues. This case study, amongst other things, distinguishes between these two types of requirements.

The mentioned standards and regulations can be referenced from the contractual requirements in several ways: (i) specific or abstract part of the regulatory document (e.g., Part

11.5.1 of the standard AREMA or the entire AREMA standard, respectively); and (ii) specific or abstract part of the subject system (e.g., electrical cables or the Network Management system, respectively). This variety of ways of referencing adds to the complexity of the contractual requirements and their relationships with the standards and regulations. In particular, when either the regulatory document or the subject system is referenced in an abstract manner, there is a need to drill down the details of the abstract entities in order to make them specific so that compliance can be demonstrated. For example, "*Circuit breakers shall comply with AREMA requirements*" does not give any indication of which specific parts of the AREMA standard are to be complied with. This abstract referencing is particularly problematic because it requires a manual analysis of the (AREMA) standard, often involving domain experts and other stakeholders, to identify the applicable parts of the standard for this specific requirement. This problem is compounded by the fact that the size of each of the over 300 regulatory documents can range anywhere from approximately 50 to 1,200 pages! This is thus a monumental problem for the rail corridor infrastructure system project. In our case study, we dealt with the AREMA standard (approx. 1,200 pages) and the contract (over 1,000 pages) in order to create the meta-model (mentioned in the introduction section and described fully in Section 2.4.4).

2.3.1.3 Requirements engineering process

The main tasks of the RE project are to: (1) elicit and document detailed system requirements from the high-level contractual requirements; (2) elicit and document detailed requirements from the standards and regulations referenced in the contract; and, (3) create and maintain traceability links between specific segments of the contract to regulatory and standards documents, and to the elicited detailed system requirements.

The project used Rational DOORS[3] to capture the project requirements and to maintain traceability among the various software artefacts. This process was accomplished manually resulting in a time consuming and costly effort.

---

[3] Telelogic DOORS, http://www-01.ibm.com/software/awdtools/doors

The challenge of achieving compliance against the large set of voluminous standards and regulations significantly characterize the RE project while little is known from the current state of the practice or existing literature about it.

Multiple suppliers (company) were subcontracted with the actual contracted company for the development and delivery of the different subsystems and components of the system. Therefore, individual subsystem or component specification document development and their traces to the contract, corresponding standards and regulation for each supplier were another critical task. Since the high-level requirements in the contract are not categorized as the subsystem and component level; rather they are organized into engineering domains (e.g., electrical, mechanical, material, etc.).

2.3.2 Research questions

The goal of this study is to create a compliance meta-model for RE through empirical findings, which depicts the types of artefacts and their inter-relationships existing in RE projects. Based on the study goal, we have four pertinent research questions. The first three questions pertain to the "knowledge seeking" aspect of the study, where we characterize the RE compliance artefacts and their relationships. Q4 is related to the development of the meta-model.

Q1. What are the types of artefacts created and used in a RE compliance project?

In order to develop the RE compliance meta-model, we first needed to identify the types of artefacts used in a RE compliance project. To investigate this question, the collection of artefacts used in the RE compliance project (see Section 2.3.3) was analyzed to identify the different types of artefacts. The types of artefacts were determined by inspecting the purpose of a particular artefact (e.g., contract, subsystem requirement, regulation, etc.) in the RE project. Each artifact in the database was explicitly articulated with its purpose by the developers, reducing the possibility of researcher misinterpretation.

Q2.What are the inter-relationships that exist among the RE compliance artefacts?

This question is investigated by determining the inter-relationships between the artefacts identified from the investigation of Q1 (*types of artefacts)*. This involves determining: (1) whether a relationship exists between any two artefacts, and (2), if a relationship does exist, the *type* of this relationship (e.g., *cross-reference to, is-derived from*, *conform-to*, etc.). The types of relationships can be inferred directly from the artefacts themselves, where the relationships are explicitly mentioned (for example, a high-level contractual requirement explicitly references a particular standard), or the relationships are implicit, and must be inferred through researcher interpretation.

Q3. What are the key characteristics of the items in the artefacts that are associated with the artefacts inter-relationships?

Here, we probe deeper into the findings from Q2 where we provide a quantitative characterization of the identified artefact relationships. Basically, frequency counts of the occurrence of the artefact relationships are determined. This quantitative characterization is critical for understanding the strength of the relationship between any given pair of artefact types.

The findings of this question are also needed for the development of the meta-model, as they suggest the cardinality that should be applied to each type of artefact relationship.

Q4. What is a RE standards and regulatory compliance meta-model for system requirements?

This question is investigated by constructing a meta-model based on the findings from Q1-Q3. The meta-model is developed using recognized guidelines for meta-model construction (Berenbach et al., 2009). Further details regarding the meta-model construction are discussed in Section 2.3.4.

2.3.3 Data collection and analysis

We performed action research in our study (i.e., the researcher participates in the development project while investigating it). In particular, we were involved in the project activities such as (regulatory) requirements elicitation from the specified regulatory documents; interview sessions with domain experts to get identified the relevant segments of regulatory documents; training workshops on creating appropriate traces for regulatory requirements in DOORS, and RE project meetings to formulate action plan for RE compliance. Our participation in the project activities helped us to build solid understanding of the RE compliance activities; the artefacts-types and their role; the organization of file repository; and last but not the least, domain of the developing system.

We collected the RE project files (i.e., documents such as contract document, project requirements documents, standards, regulation, compliance process procedural documents, and so on) from a web-based file repository system used by the project. The collected documents represent the snapshot of the artefacts profile at a certain period. The set of documents comprise several thousand pages in total.

The collection of documents was the prime source of our study data. Based on domain knowledge we gathered during our active project participation, we qualitatively analyzed the contents of the documents. We also contacted to the technical manager of the project time to time to understand the project background, profile, and several fuzzy issues (e.g., how the contract was settled). Another Ph. D student of the University of Western Ontario from the field of Requirements Engineering was involved in the document analysis process.

In addition, we interviewed two different domains experts (senior designers of one subsystem) from same engineering domain, in three two-hour sessions, on determining the applicability of a standard document (of over 1000 pages) to the corresponding subsystem. In particular, we asked them to identify the list of segments of the document that are relevant to the subsystem. They also stated which parts of the subsystem are relevant for the applicable segments. The document was structured hierarchically from sections to subsections, sub-subsections and so on, the experts were asked to identify the lowest level of the document structure that applies to system.

2.3.4 Meta-model construction procedure

The rail corridor infrastructure project used a web-based file repository system that contained all the relevant documents including the contract, standards, regulations, requirements, etc. We analyzed these qualitatively with the objective to create a meta-model using the following steps:

Step 1 - Identify the types of artefacts (Q1):

We determined the types based on the distinctly different documents used in the project (e.g., requirements specifications, regulations, standards, contract, etc.).

Step 2 - Determine the relationships among the artefacts (Q2):

We manually checked each item in the artefacts (e.g., requirements, segments of standards, regulations, etc.) to determine if they are related or dependent on item(s) from other artefacts.

Step 3 - Characterise the items associated in the relationships (Q3):

We characterised the items in the artefacts that are associated with each other in a given relationship into various categories for the purpose of useful interpretations.

Step 4 - Create a meta-model for compliance (Q4):

We created a meta-model for compliance based on the conceptual guidelines provided in (Berenbach et al., 2009). This included the following steps:

Step 4.1 - Identification of the conceptual nodes of the meta-model:

Each of the artifacts used and created in the RE compliance project has specific purposes. The collections of artifacts with similar purpose are conceptualized as a single ‘node’ in the meta-model. Thus, a ‘contract’ node in the meta-model does not indicate the actual contract document of the project but refers to a conceptualization of a typical contract document.

Step 4.2 - Determination of relationships among artifacts-types:

Based on the relationships information found from the constructs of ‘Q2’, we determined the relationships among the artifacts.

Step 4.3 - Determination of the cardinality of the relationships:

We determined the cardinality of the relationships mostly based on the data captured through the construct of ‘Q3’ that represents the number of items associated in the relationships.

2.4 Results

In this section, we describe the results of the study. Types of artefacts used in RE projects are discussed in Section 2.4.1, the inter-relationships among the artefacts are discussed in Section 2.4.2, key characteristics of the items in the artefacts that are related each other in various relationships are discussed in Section 2.4.3, and Section 2.4.4 depicts the empirically derived compliance meta-model for RE.

2.4.1 Types of artefacts (Q1)

The contract and the regulatory documents are not homogeneous; they are a mosaic of different types of “objects” of interest in this project. For example, the contract has both regulatory and non-regulatory requirements; explicit and implicit requirements; cross-cutting requirements; project execution requirements; and so on. Likewise, the regulatory documents have specifications of regulatory codes; source of origin along the governmental jurisdictions; focus on specific technical domains (e.g., electrical, mechanical, and metals); etc. We refer to these “objects of interest” as different types of artefacts of the project. The 1st column of Table 2-1 describes the different types of artefacts that we identified in the project as relevant for understanding the “big picture” of compliance activities in RE. Regulatory compliance aspects are associated with each of these artefacts-types, which are given in the 2nd column of Table 2-1. For example, contract is the prime source of the relevant regulatory documents for the systems (see the 1st row of Table 2-1). In addition, there are high-level requirements in the contract that reference to specific regulatory codes or engineering standards to which the requirements have to comply with (see the 1st row of Table 2-1). Likewise, regulatory documents describe detail requirements for systems which are critical for system compliance (see the 2nd and 3rd row of Table 2-1).

In order to ascertain and demonstrate that the set of elicited project requirements (from the contract or regulatory documents) are in accordance with all applicable regulatory documents as well as with the contract, project staffs must know how the apparently scattered pieces of information (e.g., requirements, regulatory codes, and sections of a document) across the artefacts are inter-related. Hence, we determined the 'types' of the artefacts from the regulatory compliance perspective and identified their role to regulatory compliance (see the 2nd column of Table 2-1).

Table 2-1 Types of artefacts

Artefacts-types	Compliance Aspects
1. *Contract Document* – A contract is the legally binding agreement between the customer organisation and the development company. Conceptually, there are two parts of the contract that are distinguished in the manner in which they reference the standards and regulations, as follows: *i) High-level Requirements Sections (HLRS)* – HLRS explicitly specify high-level requirements for specific parts of the system. *ii) Referenced Standards and Regulations Sections (RSRS)* – RSRS has two sections such as: (a) 'referenced standards', and (b) 'referenced regulations' which provide lists of the standards and regulations without detailing requirements, respectively, to which the entire system is expected to comply with.	Contains high-level regulatory requirements 1(i) and 'Sections' listing set of regulatory documents applicable to system 1(ii).
2. *Engineering Standards Documents* – These are usually formal documents containing established norms, or requirements, which aim to establish uniform engineering or technical criteria, methods, and practices. It has no legal mandate by itself.	Describe regulatory requirements for systems
3. *Regulations Documents* – Legal documents that refer to executive order, announced by legislative authority having force of law. Legislative authorities can be National, State and City.	Describe regulatory requirements for systems
4. *Project Requirements Specification (PRS)* – It is an aggregation of multiple documents ((5) to (9) below) that contains testable system requirements to be used by developers.	Specify regulatory project requirements
5. *System Specification* – Requirements that apply to the entire system.	Specify system-level regulatory requirements
6. *Subsystem Specification* – Requirements that apply to a particular subsystem e.g., communication subsystem.	Specify subsystem-level regulatory requirements
7. *Component Specification* – Requirements that apply to a particular component within a subsystem e.g., switches.	Specify component-level regulatory requirements
8. *Cross-cutting Specification* – Requirements that apply to more than one subsystem and/or component.	Specify cross-cutting regulatory requirements
9. *Project Execution Requirements Specification (PERS)* – It contains project requirements that relate to processes or activities to be followed during the development of the system.	Specify PERS regulatory requirements
10. *Implicit Project Requirements Specification (IPRS)* – It contains requirements imposed by standards and regulations that are not explicitly mentioned in the contract. It is an aggregation of multiple documents ((5) to (9)).	Specify implicit regulatory requirements determined by domain experts.
11. *Standards and Regulations Proxy* (called *SR Proxy)* – It is a construct implemented in a RE tool that contains the segments of the standards and regulations that are applicable to the project. These tool "objects" represent regulatory segments and may be linked (e.g., for tracing purposes) to other objects in the tool (e.g., requirements and system elements).	Contains regulatory segments relevant to system.

2.4.2 Artefacts inter-relationships (Q2)

With reference to Table 2-2, we tabulate various types of inter-relationships between artefacts-pairs (e.g., *'Reference-to'*, *'Comply-to'*, and *'Impose'*) to define what they mean by the specific relationships-types. Such definitions help reducing any ambiguity to occur in the literature of RE. Two or more artefacts become related to each other by the relationships of their contents (i.e., units of items) which are typically requirements, sections of documents, regulatory codes, rules, test cases, traces, etc.

With reference to Table 2-3, requirements specified in the contract (i.e., RSRS – system-level requirements, and HLRS – high-level requirements) make a '*reference to*' certain standards and regulations for compliance. Regulations, in turn, '*cross-reference to*' among themselves and also other regulatory documents; see the 6th row of Table 2-3.

Table 2-2 Definitions of artefacts inter-relationships

'***Reference-to*' relationship** - Two artifacts '*A'* and *'B'* are said to be associated by a *'Reference-to'* relationship when one unit of item (i.e., a requirement) in artifact '*A'* includes a direct reference to an unit of item (i.e., section of a document) in artefact '*B'.* We will call '*A'* references to '*B'.*
'*Is-derived-from*' relationship - Two artifacts '*A'* and *'B'* are said to be associated by a *'Is-derived-from'* relationship when one (or more) low-level (testable) requirement(s) in artifact '*B'* is derived from a high-level requirement in artifact '*A'.* We call 'B' *is derived from 'A'.*
***'Impose'* relationship -** Two artifacts, '*A'* and *'B',* are said to be associated by a *'Impose'* relationship when a segment in artifact '*A'*, typically a standard or regulation, imposes one or more project requirement(s) in artifact '*B'* for a given target system.
Intra/Inter '*Cross-reference-to'* relationship - When one part of an artefact *'A'* references to other part of the artefact '*A*' then artefact *A* is said to associated by '*Intra-cross-reference*' relationship. However, if one part of an artefact *'A'* references to part of another artefact of the same type e.g., '*B'*, then the **t**wo artifacts (typically a standard or regulation) of the same type e.g., '*A' and 'B'* are said to be associated by a *' Inter Cross-reference-to'* relationship.
'*Conform- to*' relationship - Two artifacts, '*A*' and '*B*', are said to be associated by a '*Conform-to*' relationship when a project requirement '*a*' in artifact 'A' satisfies the requirements specified in segment '*b*' of a standard or regulation '*B*'. We will call requirement '*a*' conforms to segment '*b*' of '*B*'.
'*Proxies- to*' relationship - Artifact '*A*' is said to have a '*proxies-to*' relationship with artefact '*B*' if artefact *B* acts as a proxy for artefact *A*. Usually artefact A is a regulation or standard, and artefact B is a relevant excerpt of artefact A as needed in a particular context.

Derived project requirements (i.e., those requirements that are to be used by the developers) specified in the *Project Requirements Specifications* (PRS) document '*is derived from*' the contractual high-level requirements (HLRS) (see the 3rd row of Table 2-3). Further, the referenced standards and regulations '*impose*' implicit requirements (i.e., those not specified in the contract explicitly) that are discovered later as *Implicit Project Requirements Specification* (IPRS) during requirements elicitation (i.e., after the contract has been signed); see the 4th row of Table 2-3.

The relevant subsections of the standards and regulatory documents (i.e., those that are referenced directly from the contract and others indirectly referenced through the direct and indirect ones and are recognised by domain experts) are spread amongst over 300 voluminous standards and regulations documents. Due to the fact that only a small portion of the standards and regulations contain information relevant to the regulatory requirements at hand, a *Standards and Regulations Proxy* (*SR Proxy*) is used to maintain a copy of the relevant segments from all specified standards and regulations (see the 5th row of Table 2-3). The proxy then acts as a single, unified, information source against which the project requirements (PRS and IPRS) need to conform (see the bottom two rows of Table 2-3).

Table 2-3 Compliance-oriented inter-relationships among the artefacts

(The numbers that suffix the artefacts represent the approximate number of items associated in the relationship)

Artefact-type	Relationship-type	Artefact-type
Contract- Referenced Standards and Regulations Sections (RSRS) (300+)	Reference-to	Standards & Regulations (300+)
Contract- High-Level Requirements Sections (HLRS) (500+)	Reference-to	Standards & Regulations (150+)
Project Requirements Specification PRS (550+)	Is-derived-from	Contract- High-Level Requirements Sections (HLRS) (500+)
Standards (1) & Regulations	Impose	Implicit Project Requirements Specification (IPRS) (30+)
Standards & Regulations	Proxies-to	Standards and Regulations (SR) Proxy
Regulations	Cross-reference-to	Regulations
Project Requirements Specification PRS (550+)	Conforms-to	Standards and Regulations (SR) proxy (650+)
Implicit Project Requirements Specification (IPRS)	Conforms-to	Standards and Regulations (SR) proxy

2.4.3 Key characteristics of artefacts inter-relationships (Q3)

In this section, we describe the key characteristics of the items in the artefacts that are related in various relationships such as *'reference to*' in Section 2.4.3.1, '*is derived from*' in Section 2.4.3.2, *'impose'* in Section 2.4.3.3, and *'conform to*' in Section 2.4.3.4.

2.4.3.1 Characteristics of the items within 'reference-to' relationships

With reference to Table 2-4, regulatory documents are referenced from the contract document in two ways (see the 1st column of Table 2-4), through: (i) the RSRS sections (where names of the standards and regulations are enlisted without mentioning specific system requirements - mentioned earlier in Table 2-1 in Section 2.4.1); and (ii) the high-level requirements in the HLRS.

In the 'RSRS' of the contract (i), there are a total of 316 regulatory references (each of them is unique) to the same number (316) of unique regulatory documents that includes 287(approx. 90% of the total 316) engineering standards and 29 (approx. 10%) regulations from three levels of legislative authorities such as Federal, Provincial and City (shown in the 1st row of Table 2-4). On the other hand, 547 regulatory references made from the high-level requirements in the 'HLRS' (ii) are not unique; hence, the total number of unique regulatory documents referenced is only 154, including 143 (approx. 93% of the total 154) engineering standards and 11 (7% of the total) regulations (See the 2nd column of Table 2-4).

Table 2-4 Quantitative characteristics of referenced standards and regulations

# of Regulatory Ref. in Contract	**# of Unique Regulatory Documents referenced**					
Types of Sections (# of regulatory references)	**# of Regulations (R=F+P+C)**				**# of Standards (S)**	**Total (S+R)**
	Fed. (F)	***Prov. (P)***	***City (C)***	***Sum (F+P+C)***		
RSRS (316)	11	17	1	29 (9%)	287 (91%)	316 (100%)
HLRS (547)	5	6	0	11 (7%)	143 (93%)	154 (100%)
(RSRS without HLRS)	6	11	1	18 (62%)	144 (50%)	162 (51%)

Therefore, the percentage of standards referenced is significantly higher than the number of referenced regulations in the contract (91% vs. 9%). The reason for this substantial difference underlies in the multidisciplinary and safety-critical nature of railway systems (see Section 2.3.1.1) for which the system requirements have been specified in the contract. In such large, complex and safety-critical systems, there involves a variety of engineering domains (e.g., electrical, mechanical, IT, civil, power supply, etc.). For each of the domains, there are defined standards to be followed by systems as a way to ensure such system qualities as safety, reliability, interconnectivity, uniformity, etc.

With reference to the Table 2-4 (last row), there are significant *abstract* references to regulatory documents i.e., approx. 50% (144 out of 287) of the standards that are enlisted in the RSRS in the contract are never explicitly referenced through high-level requirements in the HLRS. Similarly, 62% (18 out of 29) of the regulations are not explicitly referenced from the HLRS in the contract. Altogether, the account of the regulatory documents having no such explicit reference from high-level requirements is over half of the total documents i.e., 51% (162 out of 316). Without having proper reference to regulatory documents, the requirements become abstract.

This figure (51%) of abstract requirements is quite substantial, as it suggests that in order to demonstrate compliance to these set of regulatory documents as part of contractual obligations, RE projects need not only to elicit applicable system requirements to ascertain compliance but also to provide reasonable arguments for why the rest of the requirements are not applicable and ignored in the elicitation process. When there is no explicit requirement mentioned with respect to these documents, then anything and everything from the documents can be relevant to the system. Without showing '*applicability matrix*' (i.e., that shows what sections of the documents are relevant or irrelevant for target system along with necessary argumentations behind the decisions) for all of the sections of these documents, compliance cannot be determined to concerned stakeholders such as customers, internal compliance auditors, external compliance auditors, etc.

We interviewed a senior RE project staff in order to determine the reason why more than half of the regulations and standards were referenced in the contract in an abstract manner. The response was that this is quite typical in a customer/bidder compliance based project, where it is considered impractical and time consuming for the customer to include detailed segments in the contract at such an early phase of the development process. The assumption is that the bidder (i.e., development organisation) must provide the necessary resources to elicit requirements from these voluminous documents.

RE challenges that are rooted in the regulatory texts such as cross-references, ambiguity, domain-specific terms, overlapping (Kitchenham, 2004; Kerrigan and Law, 2003) and (Otto and Anton, 2007) have already sought much attention in the RE community (previously discussed in Section 2-2). Compliance to standards has no direct legal obligation but it becomes binding when specified through contractual terms. Engineering standards belong to a wide range of technical domains, e.g., electrical, mechanical, IT, chemical, power, civil, etc. and they abound with technical and domain-specific terms. Without close supervision by acquainted domain experts at the process, identification of system requirements (for a given system) from the set of applicable standards is a non-starter task. Not only that, their large number (i.e., 287 in the studied project), wide technical diversities of the content, and complexity of the system (i.e., sub-systems, components, and interfaces) cause diversified domain experts from each technical domain (or each sub-system) to be involved in the requirements elicitation process. Employing experts is expensive and maintaining consistency across the artefacts produced in the process becomes challenging in the absence of established practices.

With reference to the 1st sub-column of the 2nd column of Table 2-4, the breakdown of the referenced regulations along three levels of legislative authorities (i.e., federal, provincial, city) shows that the target multi-domain transportation system (i.e., railway system) is governed by provincial laws slightly more (18 vs. 11) than the central federal regulations. On the contrary, only one city-level law was relevant for the system. However, quantitative characterisation of the referenced regulations would likely vary across countries in which the target system is to be operated based on respective administrative and legislative structure of the countries, and across domains of the systems.

2.4.3.2 Characteristics of items in 'is-derived-from' relationships

In Table 2-5, the number of project requirements that are derived from the contract is presented: the 1st column shows the categories (e.g., system-level, sub-system level, cross-cutting, etc.) of the derived requirements to be used by the project; the 2nd column shows the total number of corresponding requirements for each category; and finally, the last column shows the frequency of requirements that are regulatory.

With regards to Table 2-5 (last column), the highest two percentages of regulatory requirements belong to cross-cutting specifications (40%) and component specifications (37%) despite the fact that they specify only 27% (component specifications) and 16% (cross-cutting) of the total derived requirements (regulatory plus non-regulatory) in the PRS. On the contrary, the highest percentages of the total requirements (regulatory plus non-regulatory) belong to subsystem specifications (36%). There is only a very small percentage (2%) of regulatory requirements that belong to system-level specification.

Our further investigation into the cross-cutting specifications revealed that they mostly apply to those constraints or functionalities that tend to apply multiple types of components. For example,

'Electrical equipment shall be painted with ANSI 61 Gray to comply with CSA requirements'.

This is a typical example of a cross-cutting requirement that may belong to any electrical-based components in the system.

Table 2-5 Quantitative characteristics of the derived requirements in the PRS

Categories of PRS	Total # of Derived Requirements (%)	# Of Derived Regulatory Requirements (%)
System specification	1221 (10%)	12 (2%)
Subsystems (all) specifications	4304 (36%)	64 (11%)
Components (all) specifications	3189 (27%)	216 (**37%**)
Cross-cutting specification	1911 (16%)	240 (**40%**)
Project Execution Spec. (PERS)	1185 (10%)	62 (10%)
Total	11810 (100%)	594 (100%)

2.4.3.3 Indication for imposed regulatory requirements

Two domain experts have been interviewed to determine all the segments of a particular standard document (i.e., AREMA) that apply to one of the given sub-systems, i.e., switch clearing sub-system. The domain experts identified 13 sections out of the 21 sections of the document as relevant for the switch clearing sub-system, which is independent of the analysis of the project requirements in PRS referencing to different sections of this document. They also identified which of the sub-sections, and sub-sub-sections within the relevant section are specifically relevant. Technically, these identified segments contain at least one project requirement, which clearly hinted for new requirements as imposed by relevant regulatory documents apart from those explicitly specified in the contract. However, at the time of the project life time when we collected data, the actual elicitation of the requirements from the identified segments had not been completed. Therefore, intended characterisation of the imposed requirements is not possible.

2.4.3.4 Characteristics of the conformed regulatory codes

As said earlier in Section 2.4.2, a regulatory proxy (i.e., a program module implemented in a RE tool that contains the segments of the standards and regulations which are applicable to the project) stores relevant regulatory codes (i.e., segments of the regulations or standards describing constraint, rules, functionalities or qualities for systems) for the project derived either from the explicit reference of the contract (e.g., PRS) or determined by domain experts once analysis of the regulatory documents is done (e.g., IPRS). This proxy then acts as a unified information source against which the project requirements

(PRS and IPRS) need to conform[4]. Redundant proxy entries were avoided even if multiple requirements are found to reference to the same regulatory code; thus, each proxy entry was unique.

In Table 2-6, the quantitative characterisation of the regulatory codes (captured in SR Proxy) with respect to regulatory project requirements in the PRS is shown: types of PRS in the 1st column; number of regulatory requirements in each type of the PRS in the 2nd column; number of regulatory references within each type of PRS in the 3rd column; and the last column presents the numbers of regulatory codes in the SR Proxy complied by each type of PRS.

With respect to Table 2-6 (the 2nd and 3rd column), we observe that there are more regulatory references than the number of regulatory requirements for each type of PRS, e.g., 76 references vs. 64 requirements within the subsystem specification (76 > 64); 256 references vs. 216 requirements within the components specification (256 > 216); and 79 references vs. 62 requirements within the PERS (79 > 62). This is because there are a number of regulatory requirements in the PRS that have more than one regulatory reference. For example,

“Buried gas supply pipe shall utilize bolting materials that comply with ANSI B18.2.1 and ASTM A307, Grade B".

The above requirement is one example that has two regulatory references, i.e., (i) *ANSI B18.2.1* and (ii) *ASTM A307 Grade B*. The finding that the prevalent difference between the number regulatory references and the number o regulatory requirements within all types of PRS hints at the complexity involved in compliance analysis of the requirements, especially when the referenced codes from diverse authorities are found to be not aligned in their specified requirements or in conflict to each other (Otto and Anton, 2007).

---

[4] At the time of the project when we collected data, analysis of regulatory documents by domain experts for determining IPRS were not completed, thus there was no proxy entry with respect to IPRS.

Table 2-6 Quantitative characteristics of the conformed regulatory codes

Types of PRS	# of Regulatory Requirements in PRS	# of Regulatory References in PRS (Total)	# of Regulatory Codes in SR Proxy (Distinct)
System specification	12	14	09
Subsystem specification	64	76	42
Component specification	216	256	112
Cross cutting specification	240	259	102
Project Execution Spec. (PERS)	62	79	29

Another observation is that distinct numbers of regulatory codes (see column four in Table 2-6) in the SR proxy are less than the numbers of regulatory requirements (column two of Table 2-6) for each type of PRS, e.g., (i) nine distinct regulatory codes vs. twelve regulatory requirements within system specifications; (ii) 42 distinct regulatory code vs. 64 regulatory requirements within system specifications; and (iii) 102 distinct regulatory code vs. 240 regulatory requirements within system specifications.

This is because of the fact that more than one requirement often reference to the same regulatory code and the SR proxy does not maintain redundant entries. From technical point of view, each regulatory code is supposed to describe one requirement. However, requirements derived from contract often reference a higher level index in the document organisations (such as chapter number as a whole) that further includes a number of lower-level segments (such as sections, sub-sections within one chapter) each of which represent separate requirements for different parts of the systems. This finding also hints to complication in determining what specific part of the referenced segments is actually relevant for the requirement within which it is referenced.

2.4.4 Compliance meta-model (Q4)

Based on the artefact-types described in Section 2.4.1, and their inter-relationships and characteristics discussed, respectively, in Section 2.4.2 and Section 2.4.3, we created a compliance 'meta-model' for RE as shown in Figure 2-1 using the guidelines given in (Berenbach et al., 2009). In Figure 2-1, each rectangle represents a type of an artefact, denoted by a class node in UML. Between two artefact-types, only compliance-oriented inter-relationships (i.e., that needed for demonstrating requirements-compliance) are shown in the model. A relationship link between any two artefact-types denotes the relationship between the actual contents (e.g., contractual requirements, segments of a standard, project requirements, etc.) of those artefacts.

With reference to Figure 2-1, the contract has two conceptual parts, i.e., the HLRS (high-level requirements sections) and the RSRS (referenced standards and regulations section) from the viewpoint of regulatory-compliance. The HLRS is spread all over the contract, consisting of numerous sections containing high-level customer requirements some of which are regulatory. The RSRS refers to two dedicated sections (“Referenced Standards” and “Referenced Regulations”) of the contract, those that provide only the names of the standards or the regulations without detailing any requirement (functional or quality aspects) to which the target system is to demonstrate compliance.

In Figure 2-1, between the HLRS and Standards, and between the HLRS and Regulations, there is a '*Reference-to*' relationship because regulatory requirements contained in the HLRS refer to one or more standard(s) and regulation(s). Likewise, each entry (name of a standard or regulation) in the RSRS corresponds to exactly one regulatory document (i.e., standard or regulation). Thus, between RSRS and Standards (or Regulations) we have a '*Reference-to*' relationship.

There is an '*Is-derived-from*' relationship between the HLRS and PRS (*Project Requirements Specification*) because from each high-level requirement in HLRS one or more lower-level project requirements can be derived and specified in PRS.

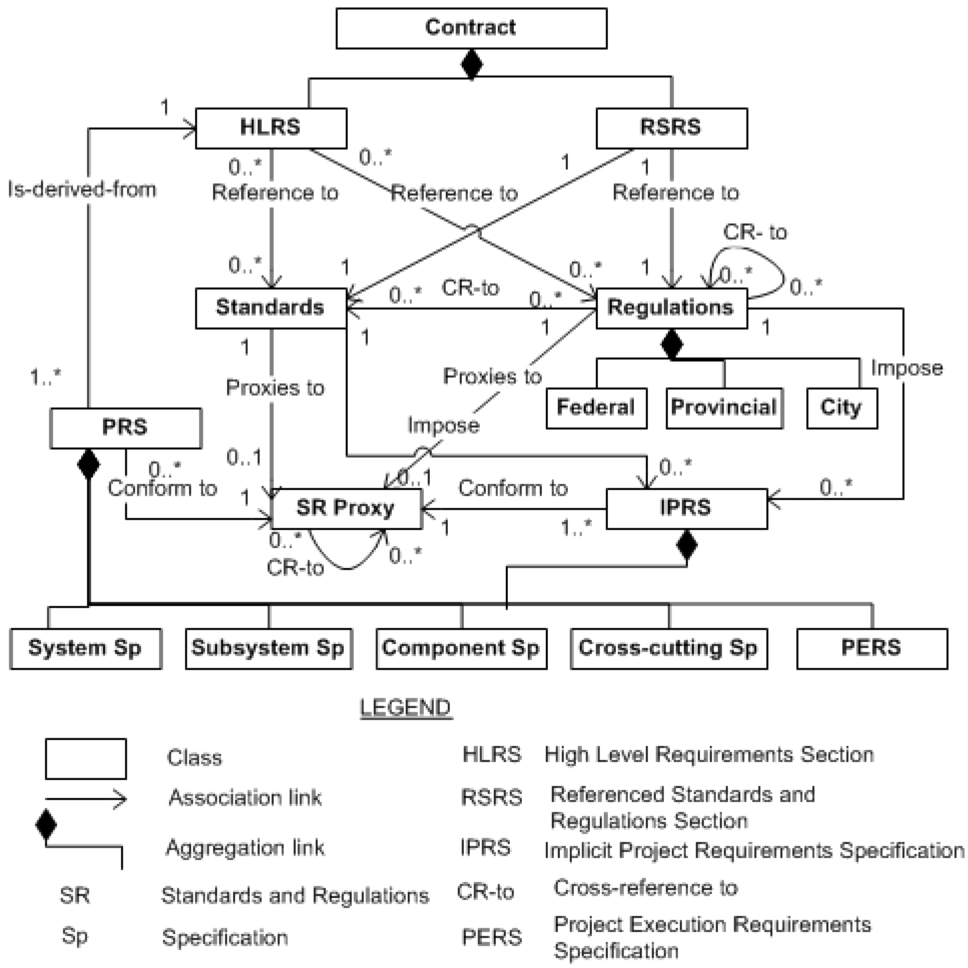


Figure 2-1 Compliance meta-model for RE

Also, a regulation can be at one of three levels: national, provincial and city, shown in Figure 2-1 as aggregation links between them and the node Regulations. Furthermore, Regulations refer to themselves and to other regulatory documents (i.e., standards or regulations), (see the '*CR-to*' (i.e., *Cross-reference-to*) relationship in Figure 2-1).

The Standards and the Regulations impose new requirements that are not specified in the contract (and hence called “implicit” requirements). These are elicited and specified in IPRS (*Implicit Project Requirements Specification*). Thus, the *'Impose'* relationship between the standards (or the regulations) and IPRS in Figure 2-1 illustrates that a standard (or regulation) may contain implicit requirements (not specified in the contract).

PRS and IPRS are both organised into five parts: system specifications, subsystem specification, component specification, PERS (*Project Execution Requirements Specifications*) and cross-cutting specifications; all these five parts thus have *aggregation* links to both PRS and to IPRS (See Figure 2-1).

For each relevant segment in the standards or regulations, there is a corresponding entry in the SR Proxy; that is why there is a '*Proxy-to*' relationship between standards (or regulations) and SR Proxy (See Figure 2-1). SR Proxy thus represents logical entries of the segments described in standards and regulations. This way, trace-links can be created and maintained in tool support (e.g., IBM DOORS) between regulatory requirements populated in the tool and segment-proxies.

Project and implicit requirements specified in PRS and IPRS, respectively, have '*Conform-to*' relationships with the regulatory entries (normative rules or policies) in the SR proxy because they are meant to conform to the corresponding regulatory entries in SR proxy. While every requirement in PRS is contractual (i.e., derived from the contract – HLRS), it is not necessarily regulatory. In contrast, each implicit requirement (described in a standard or a regulation but not in the contract) is regulatory. Thus, IPRS has a stronger relationship with SR proxy than does PRS. For this reason, for each entry in IPRS there is a corresponding entry in SR Proxy.

In order to create a tool-supported, document-based, mechanism (possibly using RE tools such as Rational DOORS) for: (i) demonstrating requirements-compliance, and (ii) downstream management of regulatory requirements, it is important to understand the artefacts and their inter-relationships as depicted in Figure 2-1. For example, the '*Impose*' relationship between regulations (or standards) and IPRS in Figure 2-1 indicates that there should be a trace between the precise part of a regulation (or standard) and the corresponding implicit requirement in IPRS. These artefact-artefact relationships thus help in developing and maintaining appropriate traces in the supporting tool. Without the artefact model, there would be no guiding principles (i.e., domain understanding) for creating and maintaining such links over the project and product life-times.

2.5 Threats to Validity

We describe the threats considered relevant to this case study: external validity in Section 2.5.1, construct validity in Section 2.5.2, and conclusion validity in Section 2.5.3.

2.5.1 External validity:

Here, we are concerned with generalisability of the results of this case study to other compliance projects (Johnson and Christensan, 2004). Clearly, this threat does exist if the results of this case study are used, verbatim, in other compliance projects without first ensuring equivalence of project attributes such as system domain, size, regulatory conditions, and so forth. Caution is thus warranted when reusing the results from this study. This does not mean that the described results are worthless. Rather, this case study forms an important, first, data-point on artefacts-types and their compliance-oriented inter-relationships existing in large-scale contractual systems engineering projects (and its graphical depiction i.e., a meta-model), and the threat calls for further case studies so that results can be categorised for increased generalisability.

2.5.2 Construct validity:

The primary issue here is whether the data captured or observed conforms to the theoretical constructs intended (Johnson and Christensan, 2004). The latter are the artefacts-types defined, types of inter-relationships existing among them, and a meta-model that depicts such entities, relationships and characterisations, presented in the four investigative questions: Q1-Q4 (see Section 2.3.2). The data gathered and analysed correspond to these questions, as will become evident in Section 2.3.3 and Section 2.3.4.

Of particular interest here is the artefact model depicted in Figure 2-1. The key entities and relationships capture the essence of requirements compliance. Thus, it is these key entities and relationships that act as constraints on the type of data that is gathered. For example, HLRS (high level requirements in the contract) make references to Standards and Regulations. Thus, given a particular high level requirement from the contract (see, for example, Section 2.4.3.1), the data gathering process would look for any references to Standards or Regulations that might be embedded in the high level requirement. Likewise, Standards and Regulations contain “implicit” requirements (not mentioned in the contract). Thus, identification of such requirements (by domain experts – see, for example, Section 2.4.3.3) would declare parts of standards and regulations as implicit. This ensured that the data gathered was in alignment with the meta-model (Figure 2-1). However, note that the entities and relationships in the meta-model were not all developed top-down (i.e., through interactions with project staff); some emerged through exploratory analysis of the regulatory and contractual documents. For example, regulations making reference to other regulations (shown as a loop over the node “Regulations” in Figure 2-1) and that there were three levels of government involved in the regulations (National, State and City) were discovered inductively during analysis of data. The key point, however, is that the resultant meta-model acts as a guide as to what elements and relationships in the project need to be examined during data analysis.

Also, several researchers and collaborators (authors and project staff) are involved in the discussions on the types of data captured and analysed. This threat is thus considered to be negligible in this case study.

2.5.3 Conclusion validity:

Here, we are concerned with the degree to which conclusions we make are based on the findings of the study (Johnson and Christensan, 2004). Among the accepted principles for improving conclusion validity (Trochim, 2006) that applies to our study is ensuring reliability of data gathering and analysis. In this regard, four researchers were involved in the study and two of them participated all through the process of data analysis. Thus, researcher biased is reduced. The conclusions in Section 2.6 are shown to be based on the results described in Section 2.4. Thus, this threat is considered to be negligible.

2.6 Implications

The implications of the findings and constructed meta-model center upon the areas of: enhancement of RE tools in the compliance-centric RE projects (see below in Section 2.6.1); and empirical study on architecture design (see below in Section 2.6.2).

2.6.1 Traceability model for RE tools

The artifacts inter-relationships presented in Table 2-3 (see Section 2.4.2) and in the meta-model (see Figure 2-1 and Section 2.4.3) suggest that compliance tracing during RE is a highly complex problem and, thus, tools could be augmented with tracing support for the artefacts and relationships identified in the meta-model. For example, a contract requirement referencing to a specific regulatory code (e.g., AREMA Part 3) needs to have traces from it to the specific portion of the regulatory code that is being incorporated in the RE tools and to the project requirements that has been derived from the contractual specification upon analysis of the referenced regulatory code. Further, if any consideration and rationale that is made at the time of specifying system requirements from the codes also needs to be captured and appropriately traced.

Existing RE Tools (e.g, Telelogic DOORS, XTie-RT (Cross Tie Requirements Tracer), IBM Rational RequisitePro, etc.) do provide indirect support for compliance artefacts through user-defined traceability models. However, the process for determining an appropriate RE compliance tracing model for a given project is currently ad-hoc. Furthermore, RE compliance literature lacks guidance on this issue. This, ad-hoc, custom model-building ultimately results in higher costs to development (Konrad and Degan, 2009).

Our meta-model thus provides a much-needed first-step for practitioners to develop custom tracing models. Essentially, specific tracing models can be instantiated from the meta-model depending on the compliance and RE-related artefacts present in a given project.

2.6.2 Empirical study on architectural design

Ferrari and Madhavji, (2008) had conducted a multiple-case study involving 16-teams to investigate requirements-oriented problems when architecting a banking system. They found that approx. 35% of all architecting problems were requirements-oriented. The most problematic areas included: quality satisfaction (22%), requirements understanding (18%), quality drivers (15%), and requirements abstraction (14%). In that study, however, no attempt was made to separate regulatory requirements and so it is not clear as to what extent the overall architectural problems are attributed to regulatory issues. Also, in the case study project described in this chapter, we did not have access to data on requirements problems experienced during system architecture development. Thus, there is no understanding at this point in time, as to the extent of regulatory issues in quality satisfaction problems during system architecting; likewise for other problems quantified above. This calls for an analogous empirical study observing architecting problems due to regulatory requirements.

In addition, we note that requirements *abstractness* (see Section 2.4.3.1) was cited as an impediment to achieving compliance, and as much as approx. 50% of all references from the contractual requirements were noted as abstract. In the multi-case banking study (Ferrari and Madhavji, 2008), requirements abstraction was also cited as a significant problem (14%) during architecture development. This would thus indicate that in the way contracts are drawn currently, there would likely be compliance-based difficulties during architecting due to requirements abstraction. The extent of the architecting problem or its impact on system compliance is however not currently known. While backtracking to requirements work may alleviate some of the uncertainty in architecting, it is clearly costly to backtrack and induces project delays, and there are no guarantees that defect slippage would not occur into downstream stages of development or, even worse, into the field. This thus calls for ways to reduce the approx. 50% abstract references in the contractual requirements.

2.7 Conclusions and Future work

We note the recent research progress on analyzing and modeling regulatory texts (such as Kerrigan and Law, 2003; Otto and Anton, 2007), techniques to automatically elicit requirements from regulations (Breaux et al, 2006), and the validation of requirements for compliance (Saeki and Kaiya, 2008). However, there is not much grounded theory on: (i) the different types of artefacts used, and the characteristics of their inter-relationships, in a requirements engineering (RE) compliance project involving regulatory documents and standards and (ii) a meta-model for RE compliance based on these artefacts and inter-relationships. This shortfall is deemed to cause variability and quality problems in RE projects.

In this study, we fill this need by describing emerging results pertaining to the described characterisation (see Table 4, Table 5, Table 6 and Section 2.4.3) of the artefacts inter-relationships. These results are from a large-scale industrial case study of a systems engineering RE compliance project involving a contractual document exceeding 1000 pages, approx. 300 engineering standards and approx. 30 regulatory documents.

Another contribution of this chapter is a meta-model for requirements-compliance (see Figure 2-1) that depicts key artefacts and relationships involved in a contractual requirements-compliance project. This model can be useful in understanding, conducting and managing a requirements-compliance project.

The characterisation and the meta-model are deemed to aid in RE compliance projects because the myriad types of artefacts and their inter-relationships are made explicit for analysts and stakeholders to consider. Also, the results lay a foundation upon which RE-oriented metrics, methods, processes and tools can possibly be developed. Our future work involves analysing case study data to identify artefacts and inter-relationships at a finer level of granularity than that described in this study.

2.8 References

Berenbach, B., Grusemann, D., Cleland-Huang, J. 2010. The Application of Just In Time Tracing to Regulatory Codes and Standards. In: 8th Conf. on SE Research. Hoboken, NJ.

Berenbach, B., Paulish, D., Kazmeier, J., Rudorfer, A. 2009. Software and Systems Requirements Engineering: In Practice. McGraw-Hill.

Breaux,T.D.,Vail,M.W.,&Antón,A.I. 2006. Towards Regulatory Compliance:Extracting Rights and Obligations to Align Requirements with Regulations.In:14th IEEE Int. RE Conf.,pp.49-58.

Environment Canada, http://www.ec.gc.ca.

Ferrari, R., and Madhvaji, N.H.,Architecting-Problems Rooted in Requirements. Information and Software Technology on Best Papers from REFSQ'05 and '06, Volume 50, Issue 1-2, pp. 53-66, 2008.

Hohfeld, W. N. (1913). Fundamental legal conception as Applied in Judicial reasoning. Yale Law Journal , 23 (1).

Islam, S., Mouratidis, H., Wagner, S. 2010. Towards a Framework to Elicit and Manage Security and Privacy Requirements from Laws and Regulations. In: 16th Int. Working Conf., REFSQ 2010. LNCS, vol. 6182, pp. 255-261.Springer, Heidelberg.

Johnson, R.B., and Christensan, L. Educational Research: Quantitative, Qualitative and Mixed Approaches, 2nd ed., Allyn & Bacon, 2004.

Jureta, I.J., Siena, A., Mylopoulos, J., Perini, A. 2010. Theory of Regulatory Compliance for Requirements Engineering. In: CoRR, vol. abs/1002.3711.

Kerrigan, S., & Law, K. H. (2003). Logic-based regulation compliance-assistance. 9th international conference on Artificial intelligence and law (pp. 126-135). Scotland, United Kingdom: ACM.

Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele University and NICTA.

Kiyavitskaya, N., Krausova, A., & Zannone, N. (2008). Why Eliciting and Managing Legal Requirements Is Hard. Requirements Engineering and Law (pp. 26-30). Barcelona, Spain: IEEE Computer Society.

Konrad, S., Degan, H. 2009. Lessons Learned from the Use of Artefact Models in Industrial Projects. In : 17th IEEE Int. RE Conf., pp 349-354. Atlanta, USA. IEEE Comp. Society.

Maxwell, J., & Antón, A. (2009). Validating Existing Requirements for Compliance with Law Using a Production Rule Model. IEEE Workshop on Requirements Engineering and the Law, (pp. 1-6). Atlanta, USA : IEEE Computer Society.

Nekvi, M.R., Ferrari, R., Berenbach, B., Madhavji, N.H., Towards a Compliance Meta-model for System Requirements in Contractual Projects. IEEE 4th International Workshop on Requirements Engineering and Law (RELAW), IEEE Computer Society, CA, USA, pp. 74-77, 2011.

Otto, P. N., & Antón, A. I. (2007). Addressing Legal Requirements in Requirements Engineering. (pp. 5-14). Los Alamitos, CA, USA: IEEE Computer Society.

Penzenstadler, B., & Leuser, J. (2008). Complying with Law for RE in the Automotive Domain. Requirements Engineering and Law (pp. 11-15). Barcelona, Spain: IEEE Computer Society.

Saeki, M., & Kaiya, H. (2008). Supporting the Elicitation of Requirements Compliant with Regulations. 20th international conference on Advanced Information Systems Engineering (CAiSE '08) (pp. 228 - 242). Montpellier, France: Springer-Verlag. 39

Siena, A., Mylopoulos, J., & Perini, A. (2008). From Laws to Requirements. Requirements Engineering and Law (pp. 6-10). Barcelona, Spain: IEEE Computer Society.

Siena, A., Perini, A., Susi, A., Mylopoulos, J. 2009. A Meta-model for Modelling Law-compliant Requirements. In: 2nd Int. Conf. on RE and Law, pp 45-51.Washington, DC, USA.

Trochim, W. M. K. 2006. Research Methods Knowledge Base. In http://www.socialresearchmethods.net /kb/design.php. Last accessed January 2009

Chapter 3

3. Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects - A Case Study[1]

3.1 Introduction

Compliance of software systems with acting government-regulations is of increasing interest in the requirements engineering (RE) community – see, e.g., (Otto and Anton, 2007; Nekvi et al., 2012). This is no more important than in large or complex software-intensive and critical systems, such as public transportation systems, banking systems, and health-care systems. From the business point of view, non-compliance can result in deficient or non-standard systems which, in turn, can lead to customer satisfaction issues. Also, non-compliance can lead to: violation of the law, penalties, and criminal charges (Otto and Anton, 2007). For example, failing to comply with the Canadian Environment Act 1999 (Environment Canada, 1999) can result in fine in the range of $75,000 to $4 million.

A large-scale systems engineering project is invariably bound by a contract between the customer and the supplier, where the contract specifies requirements while referring to a myriad of regulations and standards the target system must comply with [Berenbach et al. 2010]. For example, a rail-infrastructure upgrade project – the case of focus in this chapter -- involving various technical domains such as software, hardware, networks, communications, power, signalling and others -- would require the upgraded system to comply with regulations and standards to do with public safety, railway system, electrical devices, underlying operating system interfaces, etc. Clearly, demonstrating compliance in such a complex project cannot wait until the system has been implemented; all stages of

[1] A version of this chapter was published in (Nekvi and Madhavji, 2015) before a shorter version published in (Nekvi et al., 2012).

development, including requirements engineering (RE), must play their part in ensuring compliance.

It is to be understood that, ultimately, the term "compliance" means that the operational system must satisfy regulatory constraints. This is called "system compliance"[2]. Moreover, because a system's successful implementation depends significantly on RE (Damian et al., 2005), there is a need to demonstrate regulatory compliance at RE-time[3]. That is, it must be demonstrated at RE-time that all the relevant regulatory requirements from the contract, regulations, and standards have been elicited and traced to their sources and to the target system (as much as is known). Hereon, we call this requirements-compliance.

In a large-scale systems engineering project, however, the number and sizes of the various regulatory and contractual documents, and their inter-relationships, is mind-boggling. As will be shown later in the chapter, there can be hundreds of documents to contend with and many are thousands of pages long with countless cross-references, making requirements-compliance quite challenging.

The literature on compliance-challenges describes experiences with, and opinions about, ambiguity (Kiyavitskaya et al., 2008) and domain specific terms (Kerrigan and Law, 2003) in regulatory text; cross-referencing among regulatory documents (Kerrigan and Law, 2003; Otto and Anton, 2007; Maxwell et al., 2012); legislative conflicts (Otto and Anton, 2007; Kiyavitskaya et al., 2008 and Maxwell et al., 2012); the changing nature of the laws (Otto and Anton, 2007; Penzenstadler and Leuser, 2008 and Kiyavitskaya et al., 2008); complexity in a distributed environment (Penzenstadler and Leuser, 2008); and contractual specification practices (Berenbach et al., 2010). Typically, these are based on the analysis of one regulatory document, e.g., Federal Regulations (CFR 40) (Kerrigan and Law, 2003) and HIPAA (Breaux et al., 2006).

---

2 http://en.wikipedia.org/wiki/Intel_Cluster_Ready

3 RE-time is when bulk of the system's requirements are elicited and specified for downstream development. It is clear that, due to uncertainty or other reasons, requirements would also be elicited during system development.

In contrast, in this chapter, we describe a case study of a large-scale industrial RE upgrade-project; part of a contractual project for the rail infrastructure system. The scale of requirements-compliance in this project is characterised by: (i) a contract that is in excess of a thousand pages containing approx. 12,000 requirements, (ii) over 300 regulatory and standards documents to which the intended upgraded system must comply, and (iii) a multi-domain system consisting of seven major subs-systems with numerous components.

The novelty of this chapter is that it identifies new impediments to achieving requirements-compliance in a large-scale, contractual, systems engineering upgrade-project. An impediment is a hindrance or obstruction (in terms of effort required) in achieving requirements-compliance. Example findings of impediments are:

(a) Non-contiguity of regulatory requirements in the contract;

(b) Implicit regulatory requirements imposed by regulations and standards; and

(c) The cross-cutting characteristic of regulatory requirements.

The study's findings are both qualitative and quantitative and they add to the growing body of knowledge on the impediments to requirements-compliance. Clearly, any attempt at designing solutions[4] to overcome such impediments should not ignore their existence lest they be ineffective.

Another contribution of this research is a set of emergent metrics, identified during the analysis of the findings. These metrics attempt to describe the extent of the identified impediments to requirements-compliance. For example, the extent of regulatory (R) to non-regulatory (NR) requirements in the contract is formulated as ratio R: NR. Such a metric

---

[4] The reader is cautioned that there are no solutions described to the impediments in this chapter. The discovery of impediments is considered as a first important step in the quest for solution building and mobilises the research and tool-building community into an appropriate direction.

is anticipated to help in developing predictive theories concerning the compliance effort needed in new projects.

The rest of this chapter is organised as follows: Section 3.2 overviews related literature and analyses the research gap; Section 3.3 describes the rail infrastructure upgrade project; Section 3.4 describes the role of an artefact meta-model for an overall understanding of compliance process of RE; Section 3.5 describes the case study protocol; Section 3.6 describes the results of the case study; Section 3.7 reflects on related literature and discusses generalisability of the results; Section 3.8 describes emergent metrics on impediments; Section 3.9 discusses threats to validity; and Section 3.10 concludes the chapter and describes future work.

3.2 Related Literature

In this section, we first describe the literature on the challenges and solution approaches for analysing requirements, regulations, and policy documents for ensuring requirements-compliance (Section 3. 2.1). We then analyse this literature to highlight the "research gap" (Section 3.2.2).

3.2.1 Compliance-analysis of requirements, regulations, and policy documents

Here, we describe literature on challenges to requirements-compliance; logic-based approaches for modelling and analysing regulations; approaches for identifying and eliciting regulatory requirements; approaches for validating requirements for compliance; and analysis of requirements described in policy documents.

Compliance Challenges:

Recent literature suggests that understanding “regulations” pertaining to software-intensive systems (described in such “free-text” notions as rights, obligations, privileges, and liabilities (Siena et al., 2008)) can be difficult for requirements analysts and domain experts. For example, in (Kiyavitskaya et al., 2008) the authors mention that regulatory text often contains vague and abstract terms or missing text. Also, Kerrigan and Law

(2003) point out that regulatory documents contain domain-specific terms foreign to the RE community, and have inter- and intra- cross-references. Moreover, Kiyavitskaya et al. (2008) refer to multiple levels of laws from national and international legislative authorities. Furthermore, regulatory documents from different sources can be in conflict with one another (Otto and Anton, 2007; Kiyavitskaya et al., 2008). Penzenstadler and Leuser (2008) discuss requirements-compliance challenges in developing auto products, where multiple equipment manufacturers are involved in keeping up-to-date various documents, history, and regulatory references. Moreover, a change to a law may not be synchronously updated across all the relevant distributed documents. Berenbach et al. (2010) point out that: (i) regulations and standards are cited in the contract without specifying the associated system parts, and (ii) the relevant regulatory documents are often only cited abstractly as identifiers.

Modelling and Analysis of Regulations:

A number of logic-based approaches have been proposed for analysing and modelling regulatory text (e.g., by (Kerrigan and Law, 2003); (Antoniou et al., 1999) -- see (Otto and Anton, 2007) for a summary). These approaches incorporate a variety of encoding techniques (e.g., Mark-up based representation and Defeasible logic) to reduce ambiguities inherent in legal text. While the underlying logical model enables manually encoded text to be machine-processed, the manual encoding of voluminous text is arduous (Saeki and Kaiya, 2008). Likewise, Maxwell et al. (2012) propose a cross-reference taxonomy to guide requirements agents in determining conflicting requirements in different regulations, and provide guidelines for resolving conflicts.

Identification and Elicitation of Regulatory Requirements:

Siena et al. (2009) describe a meta-model that shows relationships among the key legal concepts. The meta-model links laws to stakeholders' strategic goals in a given domain, and guides the derivation of law-compliant requirements. Islam et al. (2010) propose a four-step framework to aid the elicitation and management of security and privacy requirements from relevant regulatory documents: (i) modelling a regulation in terms of goals, actors, task, etc., (ii) mapping the legal terms to security and privacy goals, (iii) elicitation of requirements, and (iv) refinement of the elicited requirements through analysis of security threats and privacy concerns.

Breaux et al.'s (2006) method for extracting rule statements from regulatory text was applied to subsections of HIPAA (a health Act) privacy rule, yielding 46 rights and 80 obligations. The method is characterised by its semantic parameterisation (Breaux and Anton, 2005) of regulatory text in a restricted form, which is then analysed to identify: keywords (such as may, might, must, has a right to, etc.) that indicate rule statements, and conditional keywords (such as if, unless, and except) that indicate constraints (e.g., preconditions and exceptions).

Validation and Checking of Requirements for Compliance:

Saeki and Kaiya (2008) propose a validation technique that detects any prohibited rule included in requirements. Semantics of the regulations are represented using "case frames" (i.e., verb and semantics of words that frequently co-occur with the verb), and requirements text is checked against the case frame for compliance.

Maxwell and Anton (2009) propose a regulation modelling technique using "if-then" (production) rules. The model takes existing requirements and encoded regulatory text as input, and validate that the input requirements are compliant.

Ramezani et al. (2012) propose a framework that can check, at run-time, whether a system's operation violates compliance rules. This framework uses business vocabulary to formulate regulatory requirements as rules in order to facilitate direct comparison of regulatory requirements and the on-going business process.

Ingolfo et al. (2011) propose an argumentation framework for establishing requirements compliance, that integrates with the i* framework (Yu, 1995) (for modelling requirements) and the Nomos framework (Siena, 2010) (for expressing legal concepts such as stakeholder rights, duty, and privileges) to generate compliant requirements. In this augmented framework, argumentation for/against a requirement with respect to a given fragment of law is iteratively generated to check for the requirement's compliance as long as the requirement is proved compliant. The set of arguments subsequently comprises the basis for evidence of compliance.

Regulators also require organisations to uphold their guarantees (declared in agreements) concerning the use of consumer data collected by their systems (Massey et al., 2013). Thus, proper implementation of internal policy requirements becomes obligatory for such organisations. Two approaches from the literature that deal with this issue are: CPR analysis and topic modelling (which is, in fact, a generic text mining technique (Steyvers and Griffiths, 2006)). CPR (Commitment, Privilege, and Right) analysis (Schmidt et al., 2012) enables requirements engineers to elicit regulatory requirements from privacy policies. Topic modelling has been shown to be effective (Massey et al., 2013) for identifying legal requirements in policy documents.

Analysis of Requirements in Policy Documents:

For multi-party organisational policies, some researchers are investigating into analysis techniques for relationship types (such as sharing information, policy conflicts, and relative stringency) encoded in policies and regulations. For example, Breaux and Rao (2013) have designed a formal language for representing policy statements. The formality aids during RE in: (a) detecting conflicts amongst policies, and (b) determining legitimacy of privacy data crossing application boundaries.

Also, Hassan and Logrippo (2013) propose an approach for flagging non-compliance of requirements extracted from organisational policies against relevant legal requirements. Requirements are modelled using Governance Analysis Model (GAM) and are then translated into those in a language called Governance Analysis Language (GAL). Using Governance Analysis Tool (GAT), the GAL specifications of both requirements types are checked for consistency and non-compliance.

Outcome-based regulations (focusing on the what, not how) are gradually gaining in importance (Yin et al., 2013). Unlike in prescriptive regulations, outcome-based regulations provide regulated parties with more solution options to choose from that best suits their contexts but, at the same time, raise technical issues concerning assessment of compliance. Performance modelling of legislations is claimed to be able to address this issue (Rashidi-Tabrizi et al., 2013).

3.2.2 Analysis

In contrast to the broad review of the RE literature in Section 3.2.1, in this section, we analyse only that literature that is at the core of our work on impediments so as to highlight the "research gap" in the RE field. Table 3-1 (column one) lists the impediments to achieving requirements-compliance, identified by other researchers (described in Section 3.2.1). Column two lists the source-context of the impediment; and

Table 3-1 Analysis of literature-based impediments

Impediment	Source Context	Paper	Type of Study
Vague, ambiguous, abstract terms	Legal text	(Kiyavitskaya et al., 2008)	Educated opinion
Domain-specific terms	Legal text	(Kerrigan and Law, 2003)	Educated opinion
Cross-referencing among documents	Legal text	(Kerrigan and Law, 2003)	Educated opinion
		(Otto and Anton, 2007)	Experience
		(Maxwell et al., 2012)	Case Study
Conflicts among Laws	Law	(Otto and Anton, 2007)	Experience
		(Kiyavitskaya et al., 2008)	Educated opinion
		(Maxwell et al., 2012)	Case Study
Changes in the law	Law	(Otto and Anton, 2007)	Experience
		(Penzenstadler et al,. 2008)	Experience
		(Kiyavitskaya et al., 2008)	Educated opinion
Complexity due to distributed environment	Project set-up	(Penzenstadler et al., 2008)	Experience
Abstract citation from contract	Contractual requirements	(Berenbach et al., 2010)	Experience

the last column indicates the study type. As evident, the impediments have originated largely from legal text and laws. Literature is thin on "solid" studies on impediments in a contractual, multi-domain (or inter-disciplinary) complex system involving a large set of standards and regulations – the focus of this chapter. Also, the studies cited in Table 3-1 are mainly of a small scale, leading to 'educated opinions' in some cases and 'experiences' (based on actual projects) in others.

In contrast to these small-scale studies, the impediments identified here are observations from: contractual requirements; complex, multi-domain systems; and a large set of regulatory documents. Also, the impediments are in qualitative and quantitative terms, reported later in Section 3.6. Further, in contrast to the solution papers cited in Section 3.2.1, this chapter focuses solely on discovering noteworthy impediments and not on creating solutions. Its value is in forming a strong foundation on which solution methods, tools and processes can be built.

3.3 Background: Rail Infrastructure Upgrade Project

In this section, we overview the system (Section 3.3.1); overview the RE project (Section 3.3.2); explain key aspects of the contract (Section 3.3.3); and describe the tasks and related issues involved in the compliance effort in the RE process (Section 3.3.4).

3.3.1 Description of the system

The rail infrastructure covers a large geographical area, consisting of a complex network of tracks, passenger platforms, inter-locking at major streets, several hundred signal bridges and switch machines, and many tens of miles of circuited track and associated infrastructure. The main hub of this system handles many millions of passengers a year. The key subsystems are:

Network Management: It provides central control, monitoring and management of the operations of all other subsystems. Designated user groups have pre-assigned access control from multiple sites. It has the largest volume of contractual (mainly software) requirements: over 30%. Two major components of network management are: (a) timetable management, which manages information about trains (ID, type, schedule, route, etc.) and facilitates analysis and detection of collision situations; and (b) rail-traffic control, which monitors and controls train traffic, detection and resolution of fault situations elsewhere in the system, report generation of failure or distressed cases, etc. For this purpose, it integrates various equipment and tools (e.g., database, interactive software, radios and telephones, closed circuit televisions, alarms, etc.). Also, it facilitates logging and backup of configuration and event data received over radios or telephones from other subsystems.

Signalling: This transmits periodic or incidental signals between Network Management and other equipment located elsewhere in the system, as a means of notifying the status of the equipment and their operations so that timely actions can be taken. The key components of this subsystem, each with its specific capability, include: interlocking control system, relays, track circuits, switches, cables, wires, etc. For example, interlocking ensures proper sequencing of train transportation over the rail tracks or crossings; track circuits are capable of detecting the absence of trains on the tracks; and relays enable controlling several circuits through one signal; etc.

Communication: This subsystem enables transmission of audio, video, and data across various communication systems such as network management, closed circuit televisions, telephones, clocks, and public address systems.

Switch Clearing Device: This ensures smooth operation of track switches during the winter by blowing off snow and ice from critical areas of the switches.

Power Supply: This subsystem is responsible for the distribution of electric power to all the devices in the entire system.

Civil Structures: This subsystem consists of the system's physical structures pertaining to civil construction, such as '*signal bridge'* (i.e., a special bridge located above railroads where rail-signalling equipment is installed) and communication backbone (needed by Communication subsystem to install its various apparatus).

Building Services: This subsystem provides general facilities needed in a building structure such as: air conditioning, fire alarms, heating, water, etc.

With passenger and cargo transportation in an urban area, the operational system is safety-critical from the point of view of derailment, crash, signalling failures, and the like. This aspect adds to the importance of compliance work throughout the project, including the RE process.

3.3.2 RE project overview

The studied case is a compliance aspect of an industrial-scale “RE project” the primary goal of which is to elicit requirements for upgrading the rail infrastructure system. The upgrading project duration is from late 2008 to 2014 and is multi-disciplinary, as illustrated by various subsystem descriptions in Section 3.3.1. The case study focuses on requirements for the whole system with no artificial separation between software only, or hardware only, requirements. The data for the study was gathered during the second half of 2009. Thereafter, we also had online communications and meetings with the project's staff.

3.3.3 The contract

The RE project is governed by over a 1000-page contract (between supplier and customer organisations) and is a primary source of the system’s requirements. It describes approx. 12,000 requirements at a high-level, referred to as *contractual* requirements; such a requirement can be regulatory or non-regulatory. For example, *“The transfer switch shall comply with Electrical Code.”* is a regulatory requirement; it refers to a regulatory document with which the transfer switch has to comply. A non-regulatory requirement does not refer to a regulation or standard.

Because contractual requirements are abstract, they are generally not testable. However, for driving development work, there is a need for testable requirements (a.k.a *project* requirements). These are derived from the contract and associated documents. Also, an unspecified portion of the contractual requirements refer explicitly to over 300 engineering “standards and regulation” documents (here-on *regulatory documents*) against which the project must demonstrate compliance. The requirements entrenched in the regulatory documents are called *regulatory requirements*. An important reason for singling out this subset of requirements is that failure to satisfy non-regulatory requirements can lead to customer issues, which may be negotiable whereas failure to satisfy regulatory requirements can lead to legal issues, which may not be negotiable and can incur penalties.

3.3.4 Requirements-compliance tasks

The main tasks involved in demonstrating compliance of the requirements are: (1) eliciting and documenting detailed project requirements from: (a) high-level contractual requirements and (b) relevant standards and regulations; and (2) creating and maintaining trace-links from specific segments of the contract to: (i) regulatory documents and (ii) the elicited project requirements.

A two-step process was used: a pilot project focusing on one subsystem and involving two RE agents and two domain experts, followed by fanning out of the process to other subsystems involving nine RE agents and eight domain experts. One agent has the task of managing numerous regulatory documents, including obtaining the documents, populating the document-base, monitoring changes, and bringing the changes to the attention of project staffs. The RE agents typically had two to five years of institutional experience in RE. For capturing the requirements and maintaining traces among the various RE artefacts (e.g., contract, project requirements specifications, etc.), the project used the Rational DOORS5 tool.

3.4 Role of artefact meta-model

Our role in the RE project was confined to the observational study (Runeson and Host, 2009) on impediments. As researchers, we had no influence over the direction of the project. In this role, it is difficult to see the "wood for the trees", especially in a, large, complex project, if our attention is lost in the details of specific documents. It was thus im-

[5]Telelogic DOORS, http://www-01.ibm.com/software/awdtools/doors.

portant to create a “big picture” (e.g., key artefacts and relationships) that can guide our investigation on determining impediments by understanding their characteristics (Berenbach et al., 2009, p.27). In Chapter 2, therefore, we depict such a “big picture” (called the artifact model (Berenbach et al., 2009)). This artefact meta-model is based on our observation of the types of artefacts in the project and the inter-relationships amongst the artefacts (see Chapter 2).

Why is it important to understand the different types of artefacts? Recall, our goal in this case study is to unravel the impediments to requirements-compliance work. It is, then, the characteristics of the different types of artefacts that will determine, in part, the challenges of compliance work in any given project. For example, the number of “regulatory” requirements specified in the 12,000-requirement contract is an indicator of the extent of impediments in the compliance work. That is, larger the set of contractual requirements, more (in general) the extent of impediments to achieving compliance.

It is important to note that regulatory requirements specified in the contract, alone, are not adequate for claiming compliance of the system’s requirements to regulations. These requirements often have tentacles in regulatory documents, which also need to be elicited as part of the compliance process in RE. The impediments to requirements compliance is dependent, in part, on the insidiousness with which the tentacles are spread amongst the numerous regulatory documents. Demonstrating compliance, thus, needs to show, for example, which regulatory requirements in the contract have tentacles in which regulatory documents and how the contractual and regulatory requirements are related to the “derived” project requirements (see Section 3.3.3 where derived requirements were first mentioned). Every significant requirements-compliance project needs to demonstrate this.

Also, the network of inter-linkages among the different types of artefacts (e.g., contract, regulations, standards, and project requirements - described in Chapter 2) is, clearly, an important factor in understanding the impediments to requirements-compliance work. For both these reasons, we set out to make the inter-relationships among the artefact-types (e.g., "reference to" relationship between contract and regulatory documents, "cross-reference to" among regulations, "is derived from" between project requirements and contract, etc.) explicit, which are described in the Table 2-3 of Chapter 2.

Based on the artefact-types and their inter-relationships, we created a graphical depiction of the artefact meta-model for requirements-compliance (Figure 2-1). As mentioned earlier, one purpose of this depiction was to make the domain of compliance, at least from this large systems engineering project, explicit for possible ease of understanding in future similar projects. Yet another important contribution of creation of the meta-model and its embedded artefacts-interrelationships is that it provides overall understanding of the complexities underlying the compliance process of such RE projects.

3.5 Case study

In this section, we describe the research goal and questions in Section 3.5.1; data gathering in Section 3.5.2, and data analysis in Section 3.5.3.

3.5.1 Research questions

The goal of this study was to explore different types of impediments to requirements-compliance in a large, multi-domain, systems engineering project. The rationale for this investigation is that, through our observations and interactions with the project personnel, we noted compliance-related impediments in this project (See Section 3.4) that seemed to be little understood in the RE practice and research communities. For any solution to effectively overcome such impediments, clearly, a sound understanding of the impediments is an important prerequisite. This study thus aimed at taking a step towards improved understanding of different types of impediments to requirements-compliance.

Using the gained project knowledge, including the creation of the artifact model (see Figure 2-1, Chapter 2), and following the guidelines on posing research questions (Creswell, 2003), led to the following research questions (Q1-Q4):

Q1. What are the impediments to requirements-compliance in identifying and accessing relevant regulations and standards?

This question is motivated by the apparent complexity in identifying the source of regulatory documents, and in obtaining these documents. Any large contractual project will likely have artefact-types such as those depicted in Figure 2-1 (Chapter 2).

Q2. What are the impediments to requirements-compliance due to the plethora of regulatory documents?

This question is motivated by the apparent complexity due to the size of, and interrelationships among, the identified standards and regulations. Figure 2-1 (Chapter 2) depicts the kind of interrelationships that can exist among the regulatory documents.

Q3. What are the impediments to requirements-compliance due to contractual complexity?

This question is motivated by the apparent complexity in the characteristics of the regulatory requirements specified in, or perhaps left implicit (but considered relevant) in, the contract. Figure 2-1 (Chapter 2) depicts the composition of the contract in terms of HLRS (high level requirements) and RSRS (referenced standards and regulations) as mentioned in Chapter 2. However, the referenced standards and regulations sections may also contain "implicit" requirements not mentioned in the contract.

Q4. What are the impediments to requirements-compliance due to a large-scale, multi-domain system?

This question is motivated by the apparent complexity in relating the regulatory requirements to the various parts of the system. Figure 2-1(Chapter 2) shows that the implicit requirements (IPRS) and project requirements (PRS) correspond to requirements at various levels of system abstraction (system to component levels); and they can be cross-cutting or be project execution requirements.

While the meta-model depicted in Figure 2-1 (Chapter 2) is helpful in understanding the "big picture" of requirements-compliance, the four questions posed above probe deeper into the big picture for a detailed understanding of the challenges to achieving requirements-compliance. The responses to these questions could thus aid in the design of solutions to deal with the impediments.

It is important to note that even though the identified research questions solidified through our increased understanding of the upgrade project, there were no readily available metrics for these questions (except for such rudimentary metrics such as "size" (page-length) of documents (in relation to Q2)) at the outset in this study. Reason is that the problem of impediments to requirements-compliance was not well-understood – neither in the scientific literature nor by the stakeholders including us. Only through exploration did meaningful metrics emerge for some of the questions. These, descriptive, metrics (as opposed to prescriptive ones as in the Goal-Question-Metric paradigm (Basili and Weiss, 1984)) are identified in Section 3.8.

3.5.2 Data gathering

We attended two 2-day workshops (conducted by staff mentioned in Section 3.3.4) where we learnt about, amongst other things: the system and the project; various types of regulatory documents; the role of RE agents and domain experts; the practice of, and challenges faced in, achieving compliance; repository organisation; and the tools used. This is a critical aspect of project understanding and forms a basis for data gathering. It also helped us to create the artefact model depicted in Figure 2-1 (Chapter 2). This sharpened our understanding about the key elements and relationships in the project, their bounds, and the research questions Q1-Q4 (described in Section 3.5.1).

We then gathered project artefacts such as: the contract, standards and regulations, system descriptions, etc. Thus, the gathered artefacts represent the snapshot at a particular point in time in the compliance project. During the study, we also had numerous clarification and other questions for project staff on specific matters. These interactions took place through emails and online meetings over a period of approx. 20 weeks. The notes taken during online meetings were documented in structured templates created for this purpose. Our understanding of the key issues from these interactions was shared with project staff for quality control and acceptance.

3.5.3 Data analysis

Driven by questions Q1-Q4 posed in Section 3.5.1, the regulatory requirements were analysed in conjunction with the gathered artefacts, and discussed with the project-staff, with the objective to determine a qualitative and quantitative understanding of the impediments to requirements-compliance. This analysis included issues such as the following (where relation to the research questions is identified):

- How does the organisation manage identification and accessing of the applicable standards and regulations (Q1);
- Complexity of the regulatory documents in terms of their numeracy, volume, relevance to systems, intra- and inter- cross-references, etc. (Q2);
- Contractual complexity in terms of: the contiguity of the regulatory requirements in the contract; cross-references (including their level of detail) to various standards and regulatory documents; extent to which the case study system is referred to in the contract, etc. (Q3);
- Complexity of the system in terms of sub-systems and components, inter-team communications, allocation of regulatory requirements to subsystems and components, cross-cutting requirements, etc. (Q4).

For analysis purposes, while project data was gathered at workshops and through online meetings and communications, bulk of the data source is project documents (e.g., regulatory and contractual documents). Thus, much of data analysis involves “content analysis” (Creswell, 2003):

i. *Content analysis as the research method:* Questions Q1-Q4 (in Section 3.5.1) are all “what” centred as opposed to “why”, and project data is mainly in narrative form that needs to be “analysed” (as opposed to “measured”) to answer the research questions.

ii. *Scope of material to be used in data analysis:* the “data space” is dependent on the question, available time and resources. For example, to characterise the size of documents (see later in Section 3.6.2.1), the full dataset was used; whereas, to characterise regulatory and non-regulatory contractual requirements (see later in Section 3.6.3.1), a random sample of 75 pages was used for line-by-line analysis from the 1000-page contract.

iii. *Unit of analysis:* Also called “recording unit”, a unit of analysis is the construct of interest in attempting to answer a research question. For research question Q1 (see Section 3.5.1), for example, our interest is in recognising “challenges” in identifying and accessing regulatory documents. From the data gathered, that segment of information that constitutes a challenge in identifying and accessing regulatory documents is our recording unit.

iv. *Coding categories*: We developed various non-overlapping categories, all directed to structure data in order to answer individual research questions. Examples include: contractual vs. project requirements; regulatory vs. non-regulatory requirement; abstract vs. concrete requirements; and explicit vs. implicit requirements. This is at the heart of the content analysis approach.

v. *Coding*: We manually analysed the documents to identify items of a particular category (e.g., requirements), and used the search feature of *MS Word* or *PDF* to find words of interest (e.g., "*comply with", "conform to",* or a name of a standard) for enumerating frequency or creating models.

vi. *Reliability*: Multiple researchers carried out content analysis to ensure reliability of recorded data, frequency counts and the like. We used spreadsheets to log counts and other attributes of data for further analysis.

vii. *Analysis and interpretation*: We tabulated and plotted appropriate charts, and interpreted information was shared with project staff for feedback.

The findings from this analysis are described in Section 3.6 and Section 3.8.

3.6 Results and Interpretation: Impediments to requirements-compliance

Below, we describe the key impediments to requirements-compliance identified in the case study. They are clustered as: identifying and accessing documents (Section 3.6.1); complexity of regulatory documents (Section 3.6.2); complexity of the contract (Section 3.6.3); and large scale of the project (Section 3.6.4). We also interpret the impediments here so as to give meaning to the results in situ.

3.6.1 Identifying and accessing regulatory documents (Q1)

One of the first steps in ensuring requirements compliance is to identify and gather the set of regulatory documents relevant for the system to-be-developed (or evolved). The complexity of this non-technical and foundational task is not to be underestimated for large, multi-domain, system such as the one we investigated.

For example, in our study, while the contract gives a list of approx. 300 applicable standards and 29 regulations, it is also open-ended: “The list is provided as a convenience only, and is not considered exhaustive.” An immediate implication of this is that the number of regulatory documents in the project scope is not clear and this adds to the uncertainty in the compliance task.

This uncertainty needs to be managed. The large number and variety of regulatory documents, as in the case studied, would suggest primacy of compliance across the entire system. Thus, in order to bring closure to the open-ended list of applicable regulatory documents, each part of the existing system needs to be analysed for applicability of any unlisted regulatory document deemed relevant.

A separate manager was thus responsible for obtaining all the necessary documents. Included in this responsibility is being vigilant about any updates to the regulatory documents and coordinating with requirements analysts with respect to the changes. Though tools are used for tracing regulatory requirements to the contract and to the regulatory documents, stakeholder experience suggests that this is not fool-proof and requirements often need to be re-reviewed when change-related problems are detected later in the RE and subsequent processes.

At the time of data collection in this study, 190 of the over 300 documents cited in the contract[6] had been populated in the document-base; the remaining ones were still at the acquisition stage. This shows that in large projects it is to be expected that the relevant set of documents would not be given by the customer to the project staff on a silver platter, necessitating the manager role for document management. The critical documents (such as the contract and certain standards), however, were already populated in the document-base and our study revolves around these documents.

Clearly, the task of identifying and accessing the set of applicable regulatory documents is deceptively risky. If not managed effectively, jitters are bound to be felt when conducting technical work and on project costs. A domain expert gave an illustrative example symptomatic of the kinds of hidden problems that can arise. At railway crossings on public roads, trains are bound by certain regulations in terms of maximum permissible elapsed time for stationary trains. The precise regulation on this matter took several weeks to identify and the matter passed hands across several agents in different departments of the authority. Management and administrative oversight can thus easily result in incorrect implementation of a regulation which, in turn, can lead to operational safety issues, not to mention escalating costs due to project delays.

In the case study project, three levels of government (national, provincial and city) imposed regulations on the system under development. Of the 29 regulations identified from

---

[6] The number 190 reflects project state at the time of the study was conducted

the contract, 17 belong to the national level, 11 to the state level, and one to the city level. This shows that compliance processes for large projects may have to deal with multiple levels of government.

According to the project staff, at each level of authority, the number of regulations that are relevant for the system depend on such factors as: (i) the types of domains applicable to the system; (ii) domain complexity and maturity (how long the domain is ingrained in the regulatory procedures); (iii) importance of the domains to the society (or specific sub-groups thereof) in terms of, for example, risks associated with hazards, security, and privacy; impact on the environment; treaties made with specific target groups (e.g., considering their rights, cultural needs and beliefs), etc., and (iv) size and complexity of the system. These factors also need to be analysed in the RE process in order to determine the scope of compliance issues in the project. To our understanding, there are no standardised methods for carrying out such analyses which, in a large project, can add to the complexity of identifying and accessing the relevant regulatory documents.

Note that among this huge set of standards, there may be a few that are extremely critical for, or pervasive in, the development project. This identification and prioritisation aspects are currently ad hoc. In the case study project, one particular standard stands out as prime above others in terms of the frequency of references made to it from the contract. This is the AREMA7 standard.

Whereas regulations are usually publicly available, the standards can be proprietary. Examples include such standards as ASHRAE Standard 62.1 (ventilation and air quality), ASTM B140/B140M (standards for zinc, copper, lead, alloy, etc.), CAN/CSA-A370-04 (Connectors for masonry), etc. Such standards can have multiple or concurrent versions which, in the context of a large set of standards can become problematic in terms of their identification and access.

[7] AREMA (American Railway Engineering and Maintenance of way Association) is a communication and signal manual of recommended practice of design, plan, instruction, information …: www.arema.org.

The gathered 190 documents as well as the requirements (both contractual and derived project requirements from the contract and regulatory documents) needed to be incorporated into the RE tool (DOORS) in order to create tool-supported traces among related items (such as: individual requirements, specific parts of regulatory documents, parts of the contract, parts of the target system, and others.). This was an arduous task because importing a document in its PDF or MS Word format would not enable tracing of the desired segments of that document to/fro the requirements and system parts. In particular, the tool had no feature to import a document such that its constituent segments are automatically recognised and represented internally as operable objects. Thus, each relevant fragment of the numerous documents had to be identified and copied manually, one by one, and was then incorporated in newly created "modules" (or proxies) in the tool. Six to nine people who were involved in this activity were specially trained over 3-days for this purpose. The entire importation process for documents soaked up a number of weeks.

3.6.2 Complexity of regulatory documents (Q2)

We characterise the regulatory documents in terms of their number and size (Section 3.6.2.1), relevant regulatory sections (Section 3.6.2.2), and cross-references among regulatory documents (Section 3.6.2.3), and associate these with impediments.

3.6.2.1 Large set of documents

In the project under study, as described in Section 3.3.3, over 300 distinct standards and regulations (each one a separate document) are referenced from numerous requirements in the contract. This is a large number of documents by any measure. It adds to the complexity of managing information sharing among the stakeholders in a large project. According to a project staff member, two particular challenges that impinge upon managerial responsibilities are: changes to the regulatory documents and visualization of the compliance project status:

i. *Changes*: Over the life of a project, managing and tracking the different versions of a large number of documents and which agents are in possession of which versions can become a challenging task. Versions can change in the midst of the project, and the impact of change on compliance can be non-trivial. Regulatory project requirements (as opposed to contractual requirements) may need changing. Affected traces from project requirements to the regulatory documents may need changing (including any indirect references to other regulatory documents). The set of implicit requirements may need changing; etc.

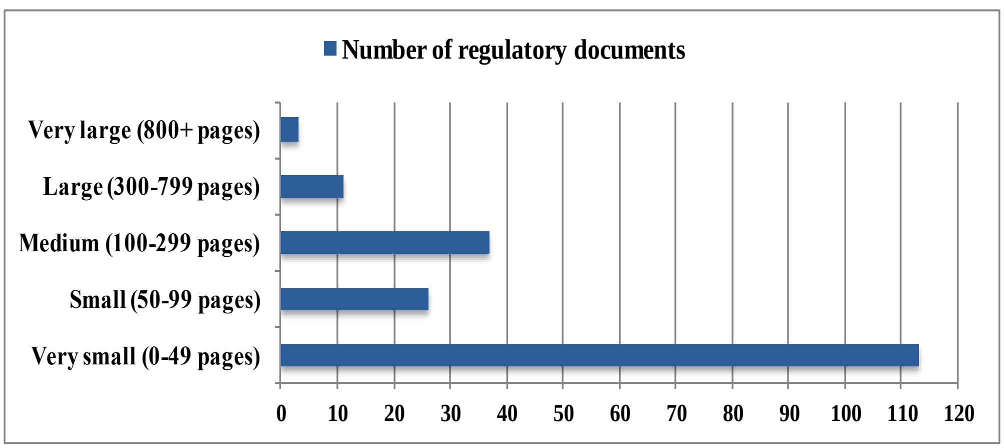
ii. *Visualisation of project status*: In a large compliance project with many regulatory documents, there are many different agents and domain experts involved in concurrently eliciting, elaborating and tracing regulatory requirements. With significant uncertainty in the extent of the linkages from a given contractual requirement to the relevant segments of the impending regulatory documents (including any indirect links) makes this task very people dependent and unpredictable. Collectively, visualising the completion status of the overall compliance project is at best only a guesstimate.

In the case study project, not only is the number of regulatory documents huge, the sizes of some of these documents are substantial too; see Figure 3-1. Examples of three of the largest documents include: IEEE Std. 1003.1 for IT--Portable O/S Interface (3,760 pages), CSA A23.1-09/A23.2-09 for Concrete materials and methods (573 pages), and AREMA for American railway standard (2,049 pages). Such sizes add to the impediments in the compliance project because: the larger the document, more complex it is to grasp in general, more time it takes to elicit requirements from them, not to mention the sheer human stamina required to sustain the process.

3.6.2.2 Identifying relevant regulatory sections

Furthermore, identifying the particular sections from a regulatory document, that are applicable to the system, can be arduous and error-prone in a large compliance project. Reasons not covered earlier include, e.g.: (i) diversity in the domain applications of the large number of documents (in the case study – over 300); and (ii) identifying the relevance of a particular section (from the various documents) to the appropriate parts of the large system – in fact, a many-to-many relationship.

In the case study, for example, a team of three people (a requirement engineer, a project supervisor and a domain expert) were assigned to determine the applicability of the sections of one of many standards to the Switch Clearing Subsystem, as part of a pilot project (mentioned in Section 3.3.4). Unexpectedly, the pilot process stalled deep into the process because of technical complexity in the document, which neither the domain expert nor the other two people could resolve, thus incurring significant delays. Subsequently, new domain experts from third-party vendors were brought in to help move the process forward. The entire pilot process took two months, yet this was still foundational work and not actual elicitation of requirements (for system implementation), which was yet to follow.



(190 available regulatory documents of the 316 referenced in the contract)

Figure 3-1 Size of regulatory documents

3.6.2.3 Cross-References

Concurring with the findings reported in (Kerrigan and Law, 2003; Otto and Anton, 2007 and Maxwell et al., 2012), of the existence of cross-references among regulatory documents, our analysis reveals example new, quantitative, insight into the extent of the problem. For example, the standard CGSB 1-GP-81 is referenced by the contractual requirement:

"Inside and outside surfaces of switchboards shall be painted with a high quality metal primer coat conforming to CGSB 1-GP-81, Type 1".

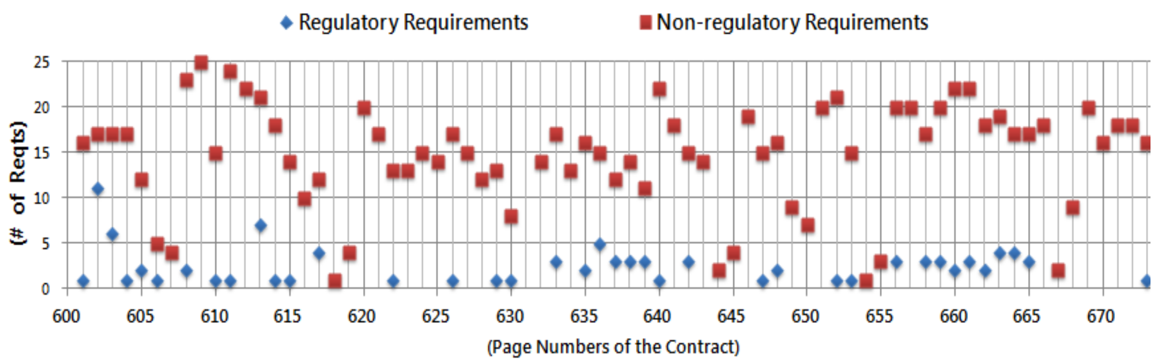
This standard, which is only a 6-page document, contains 59 cross-references in all (avg. of 10/page): 11 of which are intra-references to other paragraphs; and 48 are inter-references to 12 unique, external regulatory documents (e.g., CGSB 1-GP-70M, ASTM D1210, ASTM D2621 and nine others). Furthermore, the 12 externally referenced documents refer to yet other documents which, in turn, refer to yet others, and so on. Without appropriate visualization tools (Gotel et al., 2007), even if one is loaded with solid domain knowledge, these characteristics can easily add to the impediments in the compliance project.

3.6.3 Contractual Complexity (Q3)

This category of impediments belongs to the third question (Q3) on contractual complexity. We analyse the complexity from the points of view of: (i) the *spread* of regulatory and non-regulatory requirements in the contract (Section 3.6.3.1); (ii) the *diversity* of the references to the various regulatory documents from the contract (Section 3.6.3.2); (iii) the degree of *abstractness* of the regulatory requirements specified in the contract (Section 3.6.3.3); and (iv) the requirements that are not specified explicitly in the contract but are, nonetheless, relevant from regulatory standpoint (i.e., these requirements are *implicit*) (Section 3.6.3.4). Each of these issues is described in more detail in the sub-sections that follow.

3.6.3.1 Non-Contiguity of regulatory requirements

Across its 1000 pages and approx. 12,000 requirements (see Section 3.3.3), the regulatory requirements are specified non-contiguously in the contract (see Figure 3-2 for a random sample). Now, the contractual document is organized into ten domain-specific “divisions” (such as Electrical, Mechanical, Doors and windows, Metals, etc.) and so when identifying regulatory requirements from the contract for a particular sub-system (e.g., *power supply*), one needs to go through *all* the divisions of the contract (a thousand pages) carefully in order to identify the applicable regulatory requirements from the mixed set of requirements. There is no straightforward predictability as to when next to expect a regulatory requirement (as can be seen from Figure 3-2) – much less whether the next identified regulatory requirement is applicable to the *power supply* sub-system (which is only one of seven sub-systems). This non-contiguity in the contract makes the identification task manual, extremely slow, arduous and error-prone.



(The 75-page chunk of the contract is an extension of the randomly selected 30-page chunk (Nekvi et al., 2012) *(using the avarage of 10 random numbers between 1 and 1086 pages of the contract as a starting page. Ratio of regulatory to non-regulatory requirements in this chunk is 1:11)*

Figure 3-2 Non-contiguous requirements in the contract

Thus far, we have identified approximately 600 regulatory requirements in the contractual document, giving an overall ratio of regulatory to non-regulatory requirements as 1:19 in the entire contract. Note that this is almost twice as dilute as the ratio of the chunk of pages (1:11) in Figure 3-2, implying that identifying the regulatory requirements in the overall contractual document is considered more difficult than that shown in Figure 3-2.

Furthermore, “regulatory” requirements in the contract are not tagged explicitly as regulatory. They are rather mixed up with other general requirements in each page of the contract (see the sample ‘page# 635’ of the contract in Figure 3-3). This particular page contains 31 lines of text comprising 16 high-level requirements (R1-R16.) two of which are regulatory (R13 and R16). One can see that the two regulatory requirements can only be ascertained by parsing the entire page of text. This illustrates the tediousness and the care with which the work on compliance has to be carried out on each page of the contract.

The complexity of identifying regulatory requirements from the contract (culminating into cost, quality and time issues at the requirements stage) translates into difficulties in other project tasks, for example: (a) deriving *project* requirements (i.e., those that are actually used for system implementation) from the contract, ensuring accuracy, completeness, consistency, etc.; (b) creating traces for the derived requirements to/fro the sources in the contract; and (c) monitoring progress of the degree of requirements-compliance attained in other phases in the project life-cycle (e.g., development and testing, release configuration, and installation).

3.6.3.2 Diverse regulatory references

The described complexity is compounded when one considers the 300-odd regulatory documents (see Section 3.6.2) to be examined for regulatory requirements. For example, with reference to Figure 3-2, the following two regulatory requirements: (i) p. 622: “*All (switch clearing device) products shall comply with CSA B149*”, and (ii) p.629: “*Provide all materials and installation to ground the switch clearing devices housing including rods and conductors in accordance with Division 16 of AREMA*” – (which are from the same system component – “switch clearing device”) refer to two different standards

(CSA and AREMA), complicating the elicitation of requirements (because it may need different domain experts to comprehend the requirements for the same sub-system). In the case study, we see that, in total, 29 regulatory requirements belong to the “switch clearing sub-system” (See Table 3-2), which refer to 19 different regulatory documents for compliance. These regulatory documents are issued by six standards organizations (e.g., CSA, ANSI, AAR, etc.) and one legislative authority (i.e., a provincial government). Moreover, a single contractual requirement is often subjected to comply with diverse standards. For example, the requirement: “*Buried gas supply pipe shall utilize bolting materials that comply with ANSI B18.2.1 and ASTM A307 Grade B*” has to conform to both ANSI B18.2.1 and ASTM A307 Grade B. In the case study, our analysis of the requirements belonging to the “switch clearing sub-system” reveals that 34% (10 of 29) of the regulatory requirements need to comply with more than one standard. But the requirements from diverse regulatory documents may be in conflict with each other (Otto and Anton, 2007; Kiyavitskaya et al., 2008; Maxwell et al., 2012). This conflict must be resolved through analysis, negotiation, reviews and other processes, which adds to the impediments to requirements-compliance.

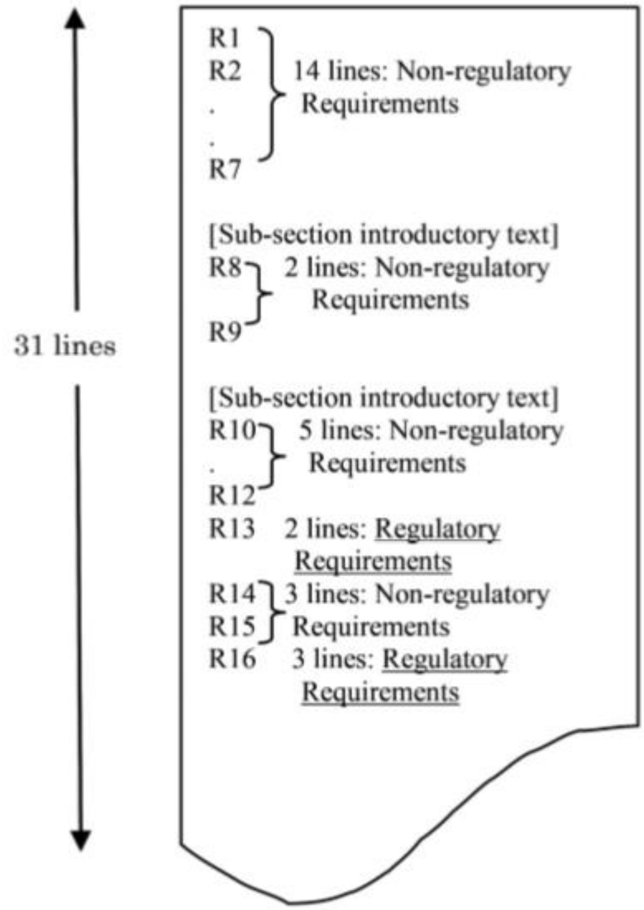


Figure 3-3 A sample contract, page 635

3.6.3.3 Abstractness

The regulatory requirements are mentioned in the contract in an abstract or specific manner, for example: (i) whether the requirement is about the whole system (e.g., *the* ***system*** *shall comply with the requirements of AREMA*) or a specific part of the system (e.g., *the* ***depth of buried gas supply pipe*** *shall be in accordance with CSA B149.1*); and (ii) whether the requirement cites the entire regulatory document (e.g., *the wayside track circuits shall be furnished in accordance with the* ***AREMA***) or a specific part of it (e.g., *nuts and washers shall be in conformance with the AREMA,* ***Part 14.1.11***). The variety of ways in which the regulatory requirements are stated in the contract adds to the impediments to compliance management.

In the case study, approx. 50% of all documents are referenced in their entirety from the contractual requirements. This is quite staggering because it suggests that in order to elicit concrete project requirements, approximately 50% of the regulatory documents must be analysed from cover to cover, requiring domain expertise in order to determine the relevant segments. Given the sizes and number of documents in the project (see Figure 3-1, for example), it is easy to see that achieving full compliance is extremely difficult, costly and time consuming.

Upon asking a senior RE project staff as to why half of the regulatory documents were referenced in their entirety in the contract, he responded that: "*it is considered impractical and time consuming for the customer to include detailed segments in the contract at such an early phase in a customer/bidder compliance-based project.*" It is expected that the bidder will provide the resources to elicit detailed requirements from these documents.

3.6.3.4 Implicit requirements

Not all requirements are mentioned explicitly in the contract, making it difficult to attain compliance. In the case study, there are requirements (called *implicit requirements)* in the 300-odd regulatory documents that are not referenced at all from the contract. According to a domain expert's analysis of the AREMA standard, the sections (or sub-sections, sub-sub-sections, etc.) that are applicable to the case study system were identified. These sections thus became candidate sources for eliciting implicit requirements. Table 3-2 shows the applicability of Section 15 of the AREMA standard to the Switch Clearing Subsystem as determined by the relevant domain expert. The last column of Table 3-2 indicates whether the corresponding section/sub-section/sub-sub-section is applicable ("Yes") or not ("No"). Applicable sections (unlike the non-applicable ones) are further analysed. For example, sub-section 15.3 is further analysed for applicability of its sub-sections (e.g., 15.3.1 ("Yes") and 15.3.10) ("No").

With reference to Table 3-2, it is interesting to note that none of the "Yes" items (i.e., those identified by the domain expert as relevant to the Switch Clearing sub-system) is referenced from the contract (which implies that these items are sources of *implicit* requirements). However, in order to ensure full compliance with the AREMA standard, *all* applicable requirements imposed by the standard need to be elicited from the standard, irrespective of whether these requirements are mentioned or not in the contract.

Table 3-2 Applicability of Section 15 of the AREMA standard to switch clearing subsystem by domain experts

Section/Sub-section/ Sub-sub-section		**Title**	**Applicability**
15[8]		Materials	Yes
15.1		Recommended Metallic Materials	Yes
	15.1.1	Criteria for Gray Iron Castings	No
	15.1.2	Criteria for Malleable Iron Castings	Yes
	15.1.4	Criteria for Various Types of Steel	Yes
	15.1.5	Non-Ferrous Metals and Alloys	Yes
15.2		Non-Metallic Materials	No
15.3		Coatings and Finishes	Yes
	15.3.1	Metallic Coating of Metals	Yes
	15.3.10	Signal Colors of Signal Glass	No
15.4		Recommended Oils, Greases	No
15.5		Identical Items	Yes
	15.5.1	Criteria for Oils and Greases	Yes

Table 3-3 shows the number of various types of segments (sections, sub-sections, etc.) in the standard AREMA that contain (by definition, regulatory) requirements For example, 10 sections containing requirements were identified by expert opinion (E) – see the “Sections” column in Table 3-3; 5 of these were referenced from the contract (C); and thus, 5 (50%) were not referenced from the contract (i.e., contain *implicit requirements*). Likewise, 74% of the Sub-sections and 59% of Sub-sub-sections contain implicit requirements. In the contract (approx. 12,000 requirements (mentioned in Section 3.3.3)),

[8] In the AREMA document, some sub-section numbers were missing, e.g., 15.1.3, 15.3.2-15.3.9, etc.

approx, 600 are regulatory. The implicit requirements[9] contained in the various types of segments in Table 3-3 add to the system's 600-odd regulatory requirements explicitly mentioned in the contract. Note that, this kind of analysis for deriving all the implicit requirements for the system needs to be done involving the 300-odd regulatory documents, which is truly a monumental task!

Table 3-3 Regulatory (AREMA) segments containing *implicit* requirements

(The accounted "Sub-sections" below are distinct segments in the standard and not a part of the accounted "Sections"; likewise for "Sub-sub-sections")

	Number and Types of Segments			
	Sections	***Sub-sections***	***Sub-sub-sections***	***All***
Expert opinion (E)	10	19	29	58
Referenced from the contract (C)	5	5	12	22
Implicit = (E - C)	5 (50%)	14 (74%)	17 (59%)	36 (62%)

3.6.4 Large, Complex System (Q4)

Here, we discuss the impediments due to the large, complex system. In particular, we describe: (i) how communications are affected by cross-cutting regulatory requirements (Section 3.6.4.1), (ii) the task of allocating the contractual requirements to specific parts of the system (Section 3.6.4.2), and (iii) the task of identifying requirements that are cross-cutting (Section 3.6.4.3).

3.6.4.1 Communications overhead

As mentioned in Section 3.3.1, the major sub-systems of the case study system are: civil structures, network management, communication, power supply, signalling, switch clearing device, and building services, consisting of thirty six components in all. For example,

---

[9] As of the time of this writing, the project staff had not elicited the implicit requirements from the identified 36 segments not referenced from the contract. However, because they have been identified as relevant by the experts, we thus know that there is at least one regulatory requirement per segment.

the signalling sub-system consists of the components: signals, switch, circuits, relay and six others. Table 3-4 shows the distribution of requirements across the sub-systems. By any measure, the system is a large, complex system.

In total, there are 1911 “cross-cutting” [10] requirements of which 240 (13%) are regulatory. This indicates that a significant amount of communication among diverse stakeholders, during RE (for elicitation, analysis, tracing, verification, management, etc.), is due to compliance-related matters. Further the 240 regulatory requirements cross-cut different sub-systems and components. This suggests a need for lateral and vertical compliance-related communications within the relevant team hierarchy.

Consider, for example, the following cross-cutting requirement: “Motors, electric heating, control and distribution devices and equipment shall operate satisfactorily at 60 Hz and DC voltage within normal operating limits established by ESA.” Agents dealing with different devices in different sub-systems need to communicate among themselves and have a shared understanding about ESA-established “normal operating limits” so that their respective devices can collaborate during system operation. Such interactions add to the impediments to requirements-compliance.

Table 3-4 Distribution of requirements (regulatory and non-regulatory)

<table>
<tr><th colspan="2">Requirements Type</th><th># of Requirements (all)</th><th># of Regulatory Requirements</th></tr>
<tr><td colspan="2">System Level</td><td>1221</td><td>12</td></tr>
<tr><td colspan="2">Project Execution</td><td>1185</td><td>62</td></tr>
<tr><td colspan="2">Cross-cutting</td><td>1911</td><td>240</td></tr>
<tr><td rowspan="7">Sub-systems</td><td>Switch clearing device</td><td>360</td><td>29</td></tr>
<tr><td>Building service</td><td>928</td><td>32</td></tr>
<tr><td>Civil structures</td><td>165</td><td>46</td></tr>
<tr><td>Communication</td><td>328</td><td>10</td></tr>
<tr><td>Network management</td><td>3799</td><td>6</td></tr>
<tr><td>Power supply</td><td>1146</td><td>97</td></tr>
<tr><td>Signalling</td><td>767</td><td>60</td></tr>
<tr><td colspan="2">Total</td><td>11,810</td><td>594</td></tr>
</table>

[10] A “cross-cutting” requirement is one that applies to multiple sub-systems. It can contribute significantly to the complexity of a system as exemplified by the task to locate related components across the system and reason about the system concern it represents (Georg et al., 2004).

3.6.4.2 Relevant contractual sections

With many diverse sub-systems (See Table 3-4), they were outsourced to third-party organisations. Thus, the primary organisation needed to develop separate sub-system specifications to be derived from the integrated contract.

However, the contract is written from the needs point-of-view of the customer, without *a priori* knowledge of how the desired system *will* be structured or which parts of the contract *will* be relevant to which sub-systems and components. Thus, when determining compliance-related requirements during system structuring, the analysts need to scan through the entire contractual document, in the case study a good 1000 pages. Table 3-5 shows, for each of the sub-systems, the number of relevant regulatory requirements and the number of different, associated, sub-sections from the contract.

For example, for the "Signalling" sub-system, there are 60 regulatory requirements specified in 11 sections. These sections are described in three "divisions" (akin to book chapters). The key impediment here is that there is no *a priori* knowledge of which divisions or sections are relevant to the "Signalling" sub-system, i.e., the requirements could have come from *anywhere* in the contract.

Furthermore, the 11 sections from the three divisions demand diverse expertise, as evidenced by the following division titles: *general requirements*, *special construction*, and *electrical;* and section titles: *standard signaling principles, entrance rack layout, signal wires and cables, vital microprocessor-based interlocking system, relays, track circuits, signals, switch machines, terminals, rail bonding, and signal system general requirements*. This problem is compounded by the fact that there are seven major sub-systems to elicit requirements for and demonstrate compliance.

3.6.4.3 Cross-cutting requirements

The communication issues mentioned earlier are not the only impediment due to cross-cutting requirements. Another impediment is in recognising that a regulatory requirement is in fact cross-cutting. This involves such analyses as: (i) whether multiple components or sub-systems of the large system are involved in the requirement and, if so, which ones, (ii) what their shared values and interactions (across system elements) are due to common properties, data communication and control, and (iii) what the characteristics are of the objects, transformations and constraints. All such analyses involves ploughing through the sea of text in numerous, sizeable, regulatory documents.

Table 3-5 Number of compliance-related sections in the contract

Sub-system name	# of Regulatory requirements	# of associated sections of the contract
Switch clearing device	29	3
Building service	32	8
Civil structures	46	4
Communication	10	5
Network management	6	1
Power supply	97	12
Signalling	60	11

In addition, the cross-cutting requirements need to be traced to the source documents as well as to the sub-system and component-level requirements at least for the purpose of demonstrating compliance at the requirements-stage. For systematic performance of this task in the case study project, the organisation uses a tracing model and defines tracing requirements, such as: (a) "*the contract links to standards and regulations*" (without giving any more details); (b) "*project requirements should be traced to relevant standards and regulations*"; and (c) "*project requirements should be traced to their contractual requirements*". However, the tracing model and the associated requirements, while cognitively helpful, only scratch the surface when manually dealing with the numerous impediments described thus far.

There is no easy respite even after the described arduousness of recognising and tracing cross-cutting requirements because, overall, there are 1911 (16%) of them in the case study system – quite a significant number. An interesting observation here is that *non*-regulatory cross-cutting requirements account for 15% of all non-regulatory requirements – see Table 3-4: (1911 - 240 = 1671) / (11,810 – 594 = 11,216); whereas, *regulatory* cross-cutting ones account for 40% of all regulatory requirements (240/594) – a factor of approx. three difference. One interpretation of this difference could be that though there are *non*-cross-cutting *regulatory* requirements (594 – 240 or 60%), there are bound to be a high percentage of cross-cutting ones (240 or 40%) because regulatory standards seem to have an inherent property of affecting multiple components or sub-systems (as in the cross-cutting requirement stated above in the section on *Communications*): *“Motors, electric heating, control … established by ESA.”* All in all, therefore, cross-cutting regulatory requirements add significantly to the impediments to requirements-compliance in the project.

3.7 Discussion

In this section, we first discuss the extent to which existing approaches address the type of impediments that we have identified in the case study. This is followed by discussion on some critical factors to be considered when attempting to generalise case study results from one environment to another.

3.7.1 Reflection on related literature

The state-of-the-art frameworks, techniques, and tools (discussed in Section 3.2.1) have created a significant impetus in regulatory requirements analysis research in the following areas:

(i) Modelling and analysis of regulations (e.g., see mark-up based representation (Kerrigan and Law, 2003); defeasible logic (Antoniou et al., 1999); and (Maxwell et al., 2012)), to support structured querying, searching, and classification of content of interest (e.g., detection of regulatory conflicts, inconsistencies, and ambiguities);

(ii) Support for identification and elicitation of regulatory requirements (e.g., see (Siena et al., 2009); (Islam et al., 2010); and (Breaux et al., 2006));

(iii) Support for validation of requirements for regulatory compliance (e.g., see (Saeki and Kaiya, 2008); (Maxwell and Anton, 2009); (Ramezani et al., 2012); (Ingolfo et al., 2011); and (Rashidi-Tabrizi et al., 2013));

(iv) Support for analysis of requirements in policy documents (e.g., see (Massey et al., 2013); (Schmidt et al., 2012); (Breaux and Rao, 2013); and (Hassan and Logrippo, 2013)).

Overall, the current approaches have paid particular attention to: (i) the complexity involved in understanding regulations and legal texts, and (ii) the relevance of the emerging technologies to systems or business. It may thus be possible that some of the proposed approaches can potentially deal with some of the impediments identified in the case study.

For example, the impediments to eliciting requirements from the voluminous regulatory documents (discussed in Section 3.6.2.1) may be mitigated by employing semi-automated techniques proposed in (Breaux et al., 2006). In addition, the cross-reference taxonomy (Maxwell et al., 2012) might prove to be useful for resolving inherent conflicts among cross-referenced regulations discussed in Section 3.6.2.3. Furthermore, regulations can be annotated with appropriate tags in a mark-up based representation (Kerrigan and Law, 2003)) to enable parsing them later to identify the regulatory segments of interest for a given system (i.e., the impediment discussed in Section 3.6.2.2).

In contrast, however, certain impediments discussed in this chapter have roots embedded in issues beyond regulations. For example, identifying and accessing relevant regulatory documents (see Section 3.6.1), is still heavily a human-centred activity. To our knowledge, currently, there are no known automated techniques that will identify the set of regulatory documents that are applicable to a given system.

Similarly, Section 3.6.3 (i.e., impediments due to contractual complexity) describes various impediments, such as: (i) non-contiguity of regulatory requirements in the contract (Section 3.6.3.1); (ii) diverse references in the requirements (Section 3.6.3.2); (iii) abstract requirements (Section 3.6.3.3); and (iv) regulatory requirements implicitly mentioned in the contract (Section 3.6.3.4). The current literature, however, does not seem to have approaches for dealing with these impediments.

Finally, impediments due to large, complex systems, discussed in Section 3.6.4, include: the need for lateral and vertical communications to resolve compliance issues in system design (Section 3.6.4.1); crosscutting requirements (Section 3.6.4.2); and relevant contractual sections (Section 3.6.4.3). To our knowledge, there are no tools or techniques currently that can handle these impediments.

3.7.2 Considerations for generalising the results

Conducting a case study in industry has its reward that the issues being examined are of factual concern and the findings are readily applicable in the context where unraveled. However, a significant challenge for industry is to be able to reuse lessons learnt, findings, and solutions from one context in another. Where results are portable, benefits accrue in terms of productivity gain, quality improvement, and cost reduction.

While Section 3.6 describes the impediment-findings from the case study, it is important to reflect upon these impediments from the point of view of their usability in another (target) environment. Specifically, are there any factors, assumptions or conditions under which these impediments are deemed to hold (or conjectured to not hold, with a sense of accompanying reasoning) in the source environment? Better to make such elements explicit so that when assessing for generalisability in the target environment these elements can be examined.

Smaling (2003) is a proponent of analogical reasoning for supporting case-to-case generalisation of results. In particular, he describes six quality criteria for analogical reasoning:

- *The relative degree of similarity between the source and target cases:* The more similar and less different the two cases are the better for generalisability of results.

- *The relevance for the conclusion:* The more the similarity (differences) between source and target cases is relevant for the study's conclusion, the more (less) plausible the analogical reasoning is.

- *Support by other, similar cases:* There are other cases similar to the source case where the results have been successfully used and the target case matches the source.

- *Support by means of variation:* Should there be larger differences between the non-significant characteristics of the source case (P) and other non-target cases (A, B, C, etc.) while there are also similarities between these cases that are significant for result generalisability then this condition makes it more plausible that the differences in the non-significant characteristics of the source case (P) and the target case (Q) do not matter and that the result will also be generalisable to Q.

- *The relative plausibility of the conclusion on its own:* The more context-free the result is, the more acceptable the analogical reasoning.

- *Empirical and theoretical support:* Where the knowledge about similarities and differences between source and target cases and their relevance is empirically and theoretically supported, the analogical reasoning is more likely to be plausible. Fewer the quality criteria that are supported, weaker the claim that the results of the source case would be generalisable to the target case.

Earlier, we mentioned the process-focused work by (Ghaisas et al., 2013) where they define "mechanisms" as sequences of actor stimuli and responses in an organisation. Their approach works on similarity of interactions patterns between source and target environments. In contrast, our artefact-focused perspective is complementary for determining critical generalisability factors.

For example, with reference to meta-model (shown in Figure 2-1 of Chapter 2), our case study project was a contractual project. The figure depicts how the contract is related to regulations and standards and, in turn, how regulations are related to the various administrative levels of government. Satisfying the contract was of paramount importance, lest there could be penalties particularly when different levels of government are involved. In non-contractual projects, the relationship between the supplier and the customer may not be as formal and thus the pressure to demonstrate requirements-compliance may not be as severe. In such projects, the impediments identified in our study may not carry as much weight, and so the user of our case study results would need to consider this factor in generalisability analysis.

Another factor that seems critical in our project is the large number of regulatory documents to manage. As mentioned in Section 3.3.4, a managerial role was assigned for this purpose. Without this role, the impediments to achieving compliance can only be expected to rise. In contrast, in another, much smaller project we were involved in (in the healthcare domain) (Yin et al., 2013), there was only one regulatory document to manage. Thus, both identification and accessibility were a non-issue.

The artefact meta-model also shows that the technology infrastructure demanded that *proxies* be created in the support tool (as described in Section 3.6.1) so that regulatory requirements in the contract can make reference to parts of regulatory documents and parts of the system. If, however, the target environment has fine-grain (e.g., object-oriented) infrastructure, proxies can possibly be avoided by creating links directly from an object to another, thereby reducing representation effort.

Another factor that seems to be related to impediments is the process used for building systems. An implicit assumption in the case study project is that the upgrade system would be developed piece-meal. However, in the target case, if the development process is, say, component-based, this could make a huge difference in the level of abstraction of the analysis carried out on regulatory and contractual requirements. Accordingly, the traces created would be at the component-level, minimising the tracing effort and the corresponding impediments to compliance.

As depicted in the artefact meta-model (Figure 2-1), the contract refers to standards and regulations. Data shows that (see Section 3.6.3.3) almost 50% of all documents are referenced by document names without giving any detail of the relevant contents. This is a major source of impediment, also shared by (Berenbach et al., 2010). Thus, if in the target environment the contract spells out the details of applicable regulatory content, then this could simplify eliciting system requirements and demonstrating compliance.

The above discussion, while not exhaustive, highlights some factors that need to be considered when attempting to generalise from a source case to a target case.

3.8 Metrics by-product

As mentioned in Section 3.5.1, while our observations lead to specifying investigative questions, Q1-Q4, there were no *a priori* metrics associated with these questions that we could use for data gathering. Metrics emerged, as a by-product, from our explorative analysis of the project's contract and regulatory documents. For example: *complexity in identifying and accessing regulatory documents*; *complexity of regulatory documents (from such points of view as number of regulatory documents to plough through, their size, their structure, need for different domain experts, intricacies of cross-references)*; or *complexity of the contract from a regulatory standpoint*. Knowing the complexity would throw some light on the effort needed to ensure compliance of requirements. In this section, we describe these metrics (shown in Table 3-6).

M4: Complexity of identifying and accessing regulatory documents.

Larger the number of technical domains to contend with in the project and more the legislative levels across which the regulatory documents originate from, the more complex would be to identify the applicable regulatory documents in the project. Larger the number of proprietary documents to contend with in the project, more complex would be to access these documents in the project. The degree of complexity is a function of the base metrics M1, M2 and M3.

M10: Degree of complexity of the regulatory documents.

This is a derived metric that is a function of several base metrics: M5 – M9 (described in Table 3-6). Higher the measures for the base metrics, higher would be the measure for the derived metric. Accordingly, more complex the regulatory documents for requirements elicitation and analysis.

Table 3-6 Emerging metrics

Ref.	Category	Metric	Metric ID	Q
Section 3.6.1	Identifying regulatory documents	Number of different domains involved in the target system	M1	Q1
		Number of legislative levels from where the documents originate.	M2	
Section 3.6.1	Accessing regulatory documents	Number of proprietary documents	M3	
Section 3.6.1	Identifying and accessing regulatory documents	Derived metric: complexity of identifying and accessing regulatory documents.	**M4**	
Section 3.6.2.1	Large set of regulatory documents	Number of regulatory documents applicable to the target system	M5	Q2
Section 3.6.2.1 (Fig 3-1)	Size	Average size of the applicable regulatory documents	M6	
Section 3.6.2.2	Identifying applicable sections	Number of domain experts needed to determine the applicability of one regulatory document to the entire system	M7	
Section 3.6.2.3	Cross-references	Average number of cross-references per page of a regulatory document	M8	
		Average number of cross-references involved per regulatory requirement	M9	
Section 3.6.2	Complexity in the Regulatory documents	Derived metric: Complexity of regulatory documents	**M10**	
Section 3.6.3.1 (Fig 3-2)	Non-contiguity	Ratio of regulatory to non-regulatory requirements in the contract	M11	Q3
Section 3.6.3.2	Diverse Reference	Average number of regulatory documents referenced from one contractual requirement	M12	
Section 3.6.3.3	Abstractness	Ratio of regulatory requirements in the contract referring to the whole system to that referring to specific parts of the system	M13	
		Ratio of regulatory requirements in the contract referring to entire regulatory documents to that referring to a specific part of those documents	M14	
Section 3.6.3.4 (Fig 3-3, Table 3-2, Table 3-3)	Implicit Requirements	I: % of regulatory requirements deemed implicit.	M15	
		II: % of segments of a standard deemed as containing implicit requirements.	M16	
Section 3.6.3	Contractual Complexity	Derived metric: Complexity in the contract	**M17**	
Section 3.6.4.1 (Table 3-4)	Communication	Number of planned subsystems	M18	Q4
		Avg. number of components in one subsystem	M19	
Section 3.6.4.2 (Table 3-5)	Relevant Contractual Sections	Avg. number of sections of the contract associated to one sub-system	M20	
Section 3.6.4.3 (Table 3-4)	Cross-cutting Requirements	% of the regulatory requirements that are cross-cutting requirements.	M21	
		Average number of subsystems or components applicable to one cross-cutting requirement.	M22	
Section 3.6.4	Large and Complex Systems	Derived metric: Complexity in large-scale systems	**M23**	

***M17: Degree of complexity of the contractual document*.**

This is a derived metric that is a function of several base metrics: M11 – M15 (described in Table 3-6). Higher the measures for the base metrics, higher would be the measure for the derived metric. Accordingly, more complex the contractual document for requirements elicitation and analysis. *M15: Implicit requirements - I*

Higher the percentage of the regulatory requirements that are implicit in the contract, more complex would be to elicit and analyse regulatory requirements.

M16: Implicit requirements - II

Higher the percentage of the segments (i.e., sections, sub-sections, sub-sub-sections, etc.) of a standard that are deemed to contain implicit requirements, more complex would be to elicit and analyse regulatory requirements.

M23: Degree of complexity due to large-scale systems

This is a derived metric that is a function of several base metrics: M18 – M22 (described in Table 3-6). Higher the measures for the base metrics, higher would be the measure for the derived metric. Accordingly, more complex it is to elicit and analyse regulatory requirements.

Analysis:

Metrics M4, M10, M17, and M23, are currently on the ordinal scale: Low, Medium and High, with the range of values for each rank currently subjective. While this is a start, helping to focus on certain critical aspects of requirements-compliance effort, it clearly calls for more studies along the lines of investigation taken in this study. This would help build a knowledge-base of effort expended in compliance projects of varying characteristics. In turn, this would help develop predictive theories (e.g., value-ranges for the ordinal scale) concerning the requirements-compliance effort needed in new projects (Humphrey, 2002).

3.9 Threats to Validity

We describe the threats considered relevant to this case study: external validity, construct validity, and conclusion validity. One more recognised threat in the literature is internal validity, which is concerned with causal effect of independent variables on the dependent variable in a controlled experiment. In particular, if the experiment design certifies that the effect on the dependent variable is only because of the manipulation of the independent variables and not due to other factors then the threat to internal validity is contained. Because in the case study we are not investigating causal effects, this particular threat does not apply.

External validity: Here, we are concerned with generalisability of the results of this case study to other compliance projects across populations of persons (population validity), settings (ecological validity), and time (temporal validity) (Creswell, 2003; Johnson and Christensen, 2007). The project under study involved professionals from industry. Thus, other similar populations would likely present little threat to the generalisability of the results from the case study.

Another dimension of interest is generalisability across different settings. Ghaisas et al. (2013) describe their approach to assessing generalisability of results across projects. Basically, they depict different actors and their interaction in an organisation – what they call “organisational project architecture”. Within this architecture, there are “mechanisms” (i.e., sequences of actor stimuli and responses). For generalisability of results from historical projects (the source) to other projects (the target), one needs to assess the source and target mechanisms and discern similarities and differences among them. Subjective decisions would then drive the extent to which results from the source project can be reused across target projects. Likewise, in Section 3.7, we discussed Smaling’s (2003) quality criteria to be considered during generalisability assessment.

The compliance project under study is relatively a small component – in terms of personnel engagement -- of the full upgrade project so there weren’t obvious patterns of agent organisation and interactions that are critical for the results to hold in another context.

Still, there was a managerial role for managing the numerous regulatory documents and the contract (see Section 3.3.4). Without such a role in another large project, gathering the relevant regulatory and contractual documents and managing the impact of changes to these during the project's lifetime could be a significant challenge. In turn, such a challenge could lead to steeper impediments to requirements compliance. Thus, generalisability of the case study results to an environment without assigned management role for documents can be under threat.

The third aspect of generalisability is temporal validity. Would the case study results be valid over time? A critical aspect in this analysis is to identify factors that could change over time and may have significant impact on the validity of the results of the case study. For example, regulations and standards can change over time. Thus, in the target projects, it is important to conduct change analysis of the volatile factors in order to assess whether the case study results described in this chapter are applicable to the target projects. The degree of threat to temporal validity of the case study results is thus context dependent.

Construct validity: The primary issue here is whether the data captured or observed conforms to the theoretical constructs intended (Johnson and Christensen, 2007). The latter are the impediment-types defined in the four investigative questions: Q1-Q4 (see Section 3.5.1). The data gathered, and analysed, correspond to these questions, as will become evident in Section 3.6.

A key point with respect to construct validity is that the interactions with project staff over several months generally involved at least two or three people in any given meeting. Thus, any question or domain understanding issues (mentioned in Section 3.5.2) were responded to openly by the project people involved and in consensus with the meeting participants. This also gives confidence on the relevance of the various constructs in this case study.

Yet another important point is that during the analysis of data, the emerging results (described in Section 3.6 were discussed between the two authors and two project staff members. Cross-checks were made between the emerging results and the research questions (Q1-Q4 in Section 3.5.1) for alignment. Any anomalies and issues were resolved with consensus. Thus, the threat to construct validity can be said to be contained in this case study.

Conclusion validity: Here, we are concerned with the degree to which conclusions we make are reasonable (Trochim, 2006) or are based on the findings of the study (Johnson and Christensen, 2007). Among the accepted principles for improving conclusion validity are ensuring reliability of data measurements (in quantitative studies) or data gathering (in qualitative studies) and soundness of data analysis. For reliability of data gathering, we based on the artefact meta-model (Figure 2-1), which depicts the key artefact types and relationships. This model has been instrumental in placing the research questions (Q1-Q4) into context (see Section 3.5.1), in gathering various documents and data contained within, and in data analysis. Also, for analysis, two researchers were at the core throughout the process, thereby mitigating researcher bias. In addition, the findings of this study are shown to be traceable to the objectives of the study. Further, should any assumptions underlying the case study not be valid then the validity of the results and, consequently, of the conclusion drawn, would be threatened. Section 3.10 discusses these issues as they pertain to the case study described in this chapter.

3.10 Conclusions and Future Work

Demonstrating compliance of a system's *requirements* with government regulations raises a number of challenges. Previous work, for example, has cited such challenges as terminology issues in regulatory text (Kiyavitskaya et al., 2008); cross-referencing (Kerrigan and Law, 2003; Otto and Anton, 2007; Maxwell et al., 2012); conflicts (Kiyavitskaya et al., 2008; Otto and Anton, 2007; Maxwell et al., 2012); managing change (Penzenstadler and Leuser, 2008; Otto and Anton, 2007; Kiyavitskaya et al., 2008); and others.

Complementing previous research, in this chapter we describe a number of hitherto uncovered impediments to achieving requirements-compliance in a large-scale, requirements engineering project that aims to upgrade a railway infrastructure system. These impediments were observed in a case study comprising the following four investigative questions:

Q1. What are the impediments to requirements-compliance in identifying and accessing relevant regulations and standards?

In large systems engineering compliance projects, identifying and accessing the relevant set of regulatory documents is problematic (see Section 3.6.1). Contributory factors include: domain-types applicable to the project, domain complexity and maturity, societal importance of the domain, system complexity, availability of regulatory documents, levels of regulative authorities to contend with, and clarity in contractual specifications.

Q2. What are the impediments to requirements-compliance due to the plethora of regulatory documents?

In the case study project, there were over 300 regulatory documents, some quite sizeable (see Figure 3-1). There can be the many-to-many relationships between the sections of numerous regulatory documents and the many parts of the system (see Section 3.6.2.2). Our investigation also yielded quantitative insight into the density of cross-references across a sample standard, i.e., 10 cross-references per page (see Section 3.6.2.3), which add to the requirements elicitation challenges.

Q3. What are the impediments to requirements-compliance due to contractual complexity?

Section 3.6.3 describes several impediments due to contractual complexity: non-contiguity of regulatory requirements – e.g., approx. 1:20 ratio between regulatory and non-regulatory requirements in the contract (see Section 3.6.3.1); diverse references from one regulatory requirement to several different regulatory documents – e.g., approx. 35% of regulatory requirements refer to more than one regulatory document (see Section 3.6.3.2); approx. 50% of all documents are referenced in their entirety from the contractual requirements (see Section 3.6.3.3); and approx. 60% of the segments of a sample standard contain *implicit* requirements (see Table 3-3).

Q4. What are the impediments to requirements-compliance due to a large-scale, multi-domain system?

Our investigation suggests that several impediments exist due to system complexity. For example, with seven major sub-systems and numerous components and approx. 1900 “cross-cutting” requirements, this suggests a significant amount of communication across the various development groups (see Section 3.6.4.1 and Table 3-4). Also, each of the sub-systems is associated to a number of overlapping segments distributed in the contract (see Table 3-5). Moreover, Section 3.6.4.3 describes the impediment of recognising “cross-cutting” requirements in the contract.

An overall conclusion from all these findings is that there are a number of significant impediments to requirements-compliance in a large scale systems engineering project. The validity of this conclusion rests on the traceable links to the sources of the individual findings, through the data analysis and gathering procedures, to the specific constructs linked to the four research questions posed above. The motivation for each of these questions can be found in Section 3.5.1 where the questions are first posed.

Threats to external, construct and conclusion validity are described in Section 3.9. Of particular interest is the discussion on analogical factors a user should consider in attempting to utilise the findings of the case study in a target environment (see Section 3.7.2).

This chapters' key contribution is that it uncovers numerous new impediments to requirements-compliance, through qualitative and quantitative analysis. This lays a significant foundational piece upon which new solutions can be created to address these impediments.

Yet another contribution is a by-product of the case study; some emergent metrics concerning requirements-compliance effort (see Table 3-6). These metrics, along with historical data on requirements-compliance effort, can be helpful in estimating efforts in new projects.

An important question left open for future work is the changes of regulatory documents and their impact on requirements-compliance work. In our project, we didn't have any opportunity to gather such data. Still, in a large compliance project, this question can be challenging to address. Reasons include: numerous independent documents that can evolve due to external change factors; involvement of numerous agents who are seemingly acting independently and concurrently on compliance tasks, making project monitoring a challenging task, let alone logging trustworthy change and impact data. Clearly, such challenges call for more effort to be put into empirical studies so as to build theoretical foundations for any solutions that are created.

Acknowledgements

We sincerely thank the reviewers for their excellent comments that have truly helped us to improve the chapter. This work was, in part, supported by Natural Science and Engineering Research Council (NSERC) of Canada.

3.11 References

Antoniou, G., Billington, D., Governatori, G. and Maher, M. J. 1999. On the Modeling and Analysis of Regulations. In Australian Conf. Information Sys., Wellington, New Zealand, 20-29.

Basili, V. R. and Weiss, D. M. 1984. A Methodology for Collecting Valid Software Engineering Data. In *IEEE Transactions on Software Engineering*, Vol. SE-10, No 6. IEEE Press, NJ, USA, 728-739. DOI: 10.1109/TSE.1984.5010301.

Berenbach, B., Lo, R. and Sherman, B. 2010. Contract-based Requirements Engineering. In Int. Workshop on Requirements Engineering & Law (RELAW'10). IEEE Computer Society, Los Alamitos, CA, USA, 27-33. DOI: 10.1109/RELAW.2010.5625354.

Berenbach, B., Paulish, D. J., Kazmeier, J. and Rudorfer, A. 2009. In Software & systems requirements engineering: in practice (1st Ed.). McGraw-Hill, New York, NY, USA.

Breaux, T. D. and Antón, A. I. 2005. Deriving Semantic Models from Privacy Policies. In Proc. of the 6th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'05). IEEE Computer Society, Washington, DC, USA, 67 - 76. DOI: 10.1109/POLICY.2005.12.

Breaux, T. D. and Rao, A. 2013. Formal Analysis of Privacy Requirements Specifications for Multi-tier Applications. In 21st IEEE Int. Requirements Engineering Conference(RE'13). IEEE Computer Society, CA, USA, 14-23. doi: 10.1109/RE.2013.6636701.

Breaux, T. D., Vail, M. W. and Antón, A. I. 2006. Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations. In IEEE Int. Requirements Engineering Conf.(RE'06). IEEE Com. Soc., Los Alamitos, CA, USA, 49-58. DOI: 10.1109/RE.2006.68.

Creswell, J. W. 2003. In Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, CA: Sage Publications.

Damian, D., Chisan, J., Vaidyanathasamy, L. and Pal, Y. 2005. Requirements Engineering and Downstream Software Development: Findings from a Case Study. In Empirical Software Engineering, Vol. 10 (3). Kluwer Academic Publishers, 255 – 283. DOI: 10.1007/s10664-005-1288-4.

Environment Canada. 1999. Canadian Environment Protection Act (CEPA 99). In www.ec.gov.ca, date of last access: April 01, 2013.

Ferrari, R., and Madhvaji, N. H. 2008. Architecting-Problems Rooted in Requirements. In Information and Software Technology, Vol. 50 (1-2). Butterworth-Heinemann, Newton, MA, USA, 53-66. DOI: 10.1016/j.infsof.2007.10.012

Georg, G., Reddy, R. and France, R. 2004. Specifying Cross-cutting Requiremens Concern. In Proc. Int. Conf. UML & Applications (UML'04). LNCS, Vol. 3273, Spinger-verlag, London, 113-127.

Ghaisas, S., Rose, P., Daneva, M., Sikkel, K. and Wieringa, R. J. 2013. Generalizing by Similarity: Lessons Learnt from Industrial Case Studies. In Proc. 1st Int, Workshop on Conducting Empirical Studies in Industry (CESI'13). IEEE Comp. Society. CA, USA, 37-42. doi: 10.1109/CESI.2013.6618468

Gotel, O. C., Marchese, F. T. and Morris, S. J. 2007. On requirements visualization. In 2nd Int. Workshop on Requirements Engineering Visualization (REV'07). IEEE Computer Society, Los Alamitos, CA, USA, 11-11. DOI: http://doi.ieeecomputersociety.org/10.1109/REV.2007.4.

Hassan, W. and Logrippo, L. 2013. Towards a Process for Legally Compliant Software. In 6th Int. Workshop on Requirements Engineering and Law (RELAW'13). IEEE Explore Digital Library.

Humphrey, W. S. 2002. In Personal Software Process (PSP). John Wiley & Sons, Inc. DOI: 10.1002/0471028959.sof238.

Ingolfo, S., Siena, A. and Mylopoulos, J. 2011. Establishing regulatory compliance for software requirements. In Conceptual Modeling (ER'11). LNCS, Vol. 6998,Springer-Berlin, Heidelberg, 47-61.DOI:10.1007/978-3-642-24606-7_5.

Islam, S., Mouratidis, H. and Wagner, S. 2010. Towards a Framework to Elicit and Manage Security and Privacy Requirements from Laws and Regulations. In Int. Working Conf. on Requirements Engineering: Foundation for Software Quality (REFSQ'10). LNCS, Vol. 6182. Springer-Verleg, Heidelberg, Germany, 255-261. DOI: 10.1007/978-3-642-14192-8_23.

Johnson, B. and Christensen, L. 2007. In Educational Research: Quantitative, Qualitative, and Mixed Approaches (3rd Ed.). SAGE Publications, Incorporated, California, USA.

Kerrigan, S. and Law, K. H. 2003. Logic-based Regulation Compliance-Assistance. In Int. Conf. on Artificial Intelligence and Law (ICAIL'03). ACM Press, New York, NY, 126-135. DOI:http://doi.acm.org/10.1145/1047788.1047820.

Kiyavitskaya, N., Krausová, A. and Zannone, N. 2008. Why Eliciting and Managing Legal Requirements is Hard. In Int. Workshop on Requirements Engineering & Law (RELAW' 08). IEEE Computer Society, Los Alamitos, CA, USA, 26-30. DOI: http://doi.ieeecomputersociety.org/ 10.1109/RELAW.2008.3

Massey, A. K., Eisenstein, J., Ant´on, A. I., and Swire, P. P. 2013. Automated Text Mining for Requirements Analysis of Policy Documents. In 21st IEEE Int. Requirements Engineering Conference(RE'13). IEEE Computer Society, CA, USA, 4-13. doi: 10.1109/RE.2013.6636700.

Maxwell, J. C. and Antón, A. I. 2009. Checking Existing Requirements for Compliance with Law Using a Production Rule Model. In the 5th IEEE International Workshop on Requirements Engineering and Law (RELAW'09). IEEE Computer Society, Los

Alamitos, CA, USA,1-6. DOI: http:// doi.ieeecomputersociety.org /10.1109/ RELAW.2009.3.
Maxwell, J. C., Anton, A. I., Swire, P., Riaz, M. and McCraw, C. M. 2012. A Legal Cross references Taxonomy for Reseanong about Compliance Requirements. In Requirements Engineering Journal, Vol. 17, No 2. Springer-Verlag New York Incorporate, Secaucus, NJ, USA, 99-115. DOI: 10.1007/s00766-012-0152-5.
Nekvi, M. R. I., Madhavji, N. H., Ferrari, R. and Berenbach, B. 2012. Impediments to requirements-compliance. In Proc. of 18th Int. Conf. on Requirements Engineering: foundation for software quality (REFSQ'12). Springer-Verleg, Heidelberg, Germany, 30-36. DOI: 10.1007/978-3-642-28714-5_3.
Nekvi, M.R.I., and Madhavji, N.H., Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects: A Case Study. ACM Transactions on Management Information Systems (TMIS), 5(3), p.15, 2015.
Otto, P. N. and Antón, A. I. 2007. Addressing Legal Requirements in Requirements Engineering. In IEEE Int. Requirements Engineering Conference (RE'07). IEEE Computer Society, Los Alamitos, CA, USA, 5-14. DOI: 10.1109/ RE.2007.65.
Penzenstadler, B. and Leuser, J. 2008. Complying with Law for RE in the Automotive Domain. In Int. Workshop on Requriements Engineering & Law (RELAW'08). IEEE Computer Society, Los Alamitos, CA, USA, 11-15. DOI: http://doi.ieeecomputersociety.org/10.1109/RELAW.2008.3.
Ramezani, E., Fahland, D., Werf, J. M. and Mattheis, P. 2012. Separating Compliance Management and Business Process Management, In Business Process Managements Workshops. LNBIP, Vol. 100, Springer-Verleg, Heidelberg, 459-464. DOI: 10.1007/978-3-642-28115-0_43.
Rashidi-Tabrizi, R., Mussbacher, G. and Amyot, D. 2013. Transforming Regulations into Performance Models in the Context of Reasoning for Outcome-Based Compliance. In 6th Int. Workshop on Requirements Engineering and Law (RELAW'13). IEEE Explore Digital Library.
Runeson, P. and Höst, M. 2009. Guidelines for conducting and reporting case study research in software engineering. In Empirical Software Engineering Journal, Vol. 14, No 2. Springer, United States, 131-164.DOI:10.1007/s10664-008-9102-8.
Saeki, M. and Kaiya, H. 2008. Supporting the Elicitation of Requirements Compliant with Regulations. In Proc. Of 20th Int. Conf. Advanced Information Systems Engineering (CAiSE '08). Springer-Verleg, Heidelberg, Germany, 228 - 242. DOI: 10.1007/978-3-540-69534-9_18.
Schmidt, J. Y., Anton, A. I. and Earp, J. B. 2012. Assessing Identification of Compliance Requirements from Privacy Policies. In 5th Int. Workshop on Requirements Engineering and Law (RELAW'12). IEEE Computer Society, Los Alamitos, CA, USA, 52-61. DOI: http://doi.ieeecomputersociety.org/10.1109/RELAW.2012.6347806.
Siena, A. 2010. Engineering Law-compliant Requirements. The Nomos Framework. Ph. D thesis. University of Trento, Italy.
Siena, A., Mylopoulos, J., Perini, A. and Susi, A. 2008. From Laws to Requirements. In Int. Workshop on Requirements Engineering and Law (RELAW'08). IEEE Computer Society, Los Alamitos, CA, USA , 6-10. DOI: http://doi.ieeecomputersociety.org/10.1109/RELAW.2008.6.
Siena, A., Perini, A., Susi, A. and Mylopoulos, J. 2009. A Meta-model for Modelling Law-compliant Requirements," In 5th IEEE International Workshop on Requirements Engineering and Law (RELAW'09) . IEEE Computer Society, Los Alamitos, CA, USA, 45-51. DOI: http://doi. ieeecomputersociety.org/10.1109/ RELAW.2009.1.

Smaling, A. 2003. Inductive, Analogical, and Communicative Generalization. In International Journal of Qualitative Methods, Vol. 2, No 1, 1-31.

Steyvers, M. and Griffiths, T. L. 2006. Probabilistic Topic Models. In Latent Semantic Analysis: A Road to Meaning. T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, Eds. Laurence Erlbaum.

Trochim, W. M. K. 2006. Research Methods Knowledge Base. In http://www.socialresearchmethods.net /kb/design.php. Last accessed January 2009.

Yin, Q., Madhavji, N. H. and Pattani, M. 2013. An Approach for Ensuring Regulatory Compliance of Process Outcomes. In 6th International Workshop on Requirements Engineering and Law (RELAW'13). IEEE Explore Digital Library.

Yu, E. S. K. 1995. Modelling Strategic Relationships for Process Reengineering. Ph.D. thesis. University of Toronto, Canada.

Chapter 4

4. Metrics for Estimating the Effort Needed In Requirements Compliance Work[1]

4.1 Introduction

Escalation of cost estimate is rampant in large systems engineering projects. Through an empirical investigation on historical records of 258 transportations projects, Flyvbjerg et al., (2003) showed that 90% of the projects overran cost estimate, where the percentage of average cost escalation is the highest (45% of the predicted cost) in the rail infrastructure projects among others (e.g., aviation and road transport). Importantly, effort-related cost is one of the most dominant factors in the total project cost (Somerville, 2006). Poor estimation of project effort and cost is identified as one of the key factors to cost escalation in large projects while the other factors include complexities in engineering, changes of schedule and project scope, ambiguous provisions of contracts (Shane et al., 2009).

Similarly, making acceptable effort estimates of ensuring regulatory compliance of requirements in large systems engineering projects has eluded project management. In practice, the process of ascertaining regulatory compliance of requirements at the phase of requirements engineering (RE) can be extremely difficult and arduous (Nekvi and Madhavji, 2015) because of: unbounded cross-references within and across documents, non-contiguity of regulatory requirements, abstract requirements, multi-domain complexity, implicit regulatory requirements, and others. There is supporting evidence for some of the causes: e.g., cross-references in (Breaux and Gordon, 2013); non-contiguity of regulatory requirements in the health act HIPAA (Breaux and Anton, 2008) and across multiple jurisdictions (Granavati et al., 2014); and detection of relevant regulatory codes (Cleland-Huang et al., 2010).

---

[1] A shorter version of this chapter was published in (Nekvi et al., 2016)

This situation raises several uncertainties, for example: whether all the regulatory requirements have been elicited or identified from complex documents; whether changes in regulations have been accounted for; and whether the effort estimation of requirements compliance work is realistic. A noteworthy complaint from industry is that underestimation of effort is uncontrolled, with consequences on cost overrun, project delay, quality problems, and customer dissatisfaction (Personal Communication, 2012).

The literature abounds with traditional approaches for effort estimation, for example: COCOMO (Boehm, 1981), neural networks (Wittig and Finnie, 1994), regression and decision trees (Srinivasan and Fisher, 1995), case-based reasoning (Mukhopadhyay et al., 1992) and estimation by analogies (Shepperd and Schofield, 1997), etc. These are based on such metrics as Lines of Code (LOC), Function Points (FP) (Albrecht and Gaffney, 1983), Object Point (OP) (Kauffman and Kumar, 1993), and Use Case Point (UCP) (Karner, 1999) to estimate the size of software systems. Obviously, these metrics are meant for general software development effort as a whole and they do not provide separate estimations for RE work, especially for compliance-related complexity in RE (Nekvi and Madhavji, 2015).

In requirements engineering (RE), little research on effort estimation of RE work has been conducted so far. For example, Hoffman et al. (2001) suggest an estimate of 16% of overall project effort for RE work. Further, *Seilevel* develops a tool for estimating RE effort, which also considers a rough estimate of 15% of overall project effort for RE (Beatty, 2012). However, effort estimation for compliance work in RE is still in its infancy where required metrics are almost non-existence let alone be there any model for estimating the effort.

Compared to the cited related work on effort estimation, our work is fundamentally different in two ways. One, since “compliance” work at RE-time, as described earlier, has particular characteristics (e.g., analysing a large set of legal documents) that are quite different from those of a “standard” elicitation process (e.g., interviews, focus groups, prototyping, etc.). Thus, any effort estimation method aimed at requirements compliance work needs to take this into account. Two, here we are concerned with systems engineering projects, which have not only software elements but non-software elements too (e.g., devices, appliances, computing hardware, non-computing hardware, etc.).

In this chapter, we describe a novel method for deriving key metrics for estimating the effort needed in requirements compliance work. The proposed method is developed from the analysis of the impediments (i.e., those compliance-related RE activities that are considered challenging) identified in (Nekvi and Madhavji, 2015). Consequently, following this method we derive key metrics for each of the effort-critical activities (i.e., impediments). These metrics are of fundamental importance for creating an effort estimation model, which we demonstrate by aligning these metrics with an algorithmic-based effort estimation model used for overall software development projects (Boehm, 1981). The actual construction of an estimation model requires adequate historical data from similar projects that are not known to exist at this time anywhere. Therefore, the model itself is outside the scope of this chapter. We anticipate that the proposed method can possibly be applied in some form to other projects for deriving their own metrics tailored to the specifics of the projects.

The rest of the chapter is organised as follows: Section 4.2 discusses the related literature and background of the project that is investigated, Section 4.3 describes our approach for deriving effort-estimation metrics, Section 4.4 presents the set of derived metrics, Section 4.5 illustrates the use of metrics in creating an effort estimation model, Section 4.6 gives a summary analysis of the results, and Section 4.7 wraps up the chapter with future work and conclusions.

4.2 Background

In this section, we describe the background of the work presented in this chapter. First, we describe the existing literature related to software metrics and effort estimation techniques in SE and RE in section 4.2.1, and second, in section 4.2.2, we sketch the background of the project that we investigated for deriving metrics towards constructing an effort estimation model for compliance work in RE.

4.2.1 Related literature

Software metrics generally have two main focuses: (i) to measure quality of software work product (such as requirements specifications, designs, codes, etc.), resources (e.g., personnel, teams, and tools) and SE processes (e.g., testing, coding, and architecting); and (ii) to predict or estimate the required effort and cost for future projects (Fenton and Neil, 2000). This chapter concerns the latter (ii). Below, we describe the literature on: metrics for traditional effort estimation techniques in software engineering and effort estimation approaches in RE.

Metrics for traditional effort estimation techniques in software engineering:

The literature is abundant with algorithmic-based effort estimating models that use a variety of metrics for estimating efforts. Examples of such models include: (i) Cost Construction Model (COCOMO) proposed by (Boehm, 1981), (ii) Software Lifecycle Model (SLIM) proposed by (Putnam, 1978), (iii) COCOMO II (Boehm et al., 2000), and others.

Estimate of software size is the prime component of algorithmic-based effort estimation models for software development. However, use of particular metrics for software sizing varies across the models. For example, Boehm, (1981) developed one of the earliest algorithmic models 'COCOMO 81' which uses *Lines of Code* (LOC) for sizing a software. In SLIM (Putnam, 1978), effort is estimated based on size, time, and productivity metrics, where size is estimated in *Effective Source Lines of Code* (ESLOC). COCOMO II (Boehm et al., 2000) is an advanced version of COCOMO 81 that applies to the sequential development process (i.e., waterfall process). Therefore, COCOMO II has been gradually improved to incorporate capabilities to estimate effort in the latest software development processes such as rapid software developments (Beck, 2000), reengineering, reuse-based developments (Griss and Wosser, 1995), and object-oriented software developments (Wolfgang, 1994). COCOMO II has several internal sub-models such as *application compositions* (i.e., that uses CASE tools for rapid developments), *early design* (to be used prior to architectural decisions), and *post-architectural model* (to be used after basic software design has been completed). The size metrics used in the sub-models are:

(a) object points (Kauffman and Kumar, 1993), i.e., count of screens, reports, and program modules, for *application compositions*, (b) function points (Albrecht and Gaffney, 1983) for *early design*, and (c) both lines of codes and function points in *post-architectural model* (Boehm et al., 2000). Function point is a measurement of software functionality used as a basis for sizing software. It differs from the concept of using physical size (i.e., lines of code) that cannot provide a consistent measure of productivity of personnel for diverse programming languages. Karner, (1999) proposes a Use Case Point (UCP) model that predicts effort based on analysis of use case diagrams, where count of use case, actors in the use cases and interactions were used as base metrics. These algorithmic-based effort estimation models also include other metrics such as for project complexity, and product, and process characteristics along with software size.

In the *estimation by analogy* technique (Mukhopadhyay, 1992; Shepperd and Schofield, 1997), features and data of the current project are extracted by experts so that they can be compared with previously completed projects to determine their similarity. Based on regression analysis on completed similar projects whose actual effort and cost is known, the estimation is made for current project. No fixed metrics are defined for this technique.

Neural network (Wittig and Finnie, 1994) is another kind of model where an estimation algorithm is constructed by learning from training data collected from historical projects. It can always adjust the parameter values of the algorithm upon new data fed into it. It can work on whatever metrics are defined appropriate for current context.

There are other expert-based estimation techniques that rely on expert opinions other than metrics-based algorithms, e.g., *expert judgment* (Hughes, 1997). This technique is handy in situations when necessary empirical data is not available for constructing a predictive effort model. However, accuracy of this technique depends solely on the opinions of experts whose years of experience may not necessarily be sufficiently competent.

Effort estimation approaches in RE:

In requirements engineering (RE), research on metrics has been conducted in two areas: i) utilising requirements to estimate entire software development effort (Verlaine et al., 2014); and ii) approaches to estimate the RE effort only (Goldsmith, 2010). The latter (ii) is of concern here since the former (i) one does not estimate RE effort rather uses characteristics of requirements as a basis for estimating the size of overall software and systems. For example, the model proposed by (Verlaine et al., 2014) analyses requirements specifications to estimate the complexity of service-oriented software in terms of design and structural complexity (i.e.., count and weights of inputs, outputs, storage, and interfaces), computational complexity, and conceptual complexity (e.g., capabilities of personnel to understand system requirements and domains). With respect to estimating RE effort (ii), Hoffman et al. (2001) mention average effort in RE at 16% of the overall project, while the most successful projects expend RE effort as high as 28%. However, fixed estimates (16%) of RE effort provided by (Hoffman et al., 2001) can only serve as a rough guideline but can't guarantee accuracy for diversified projects in the real world. Further, Seilevel's approach for estimating RE effort is based on three primary estimates: (i) 15% of overall work effort estimation; (ii) 6:1 developer to Business Analysts (BA) ratio; and (iii) bottom-up estimation derived from breakdown of RE activities and their associated historical effort (Beatty, 2012). Historical effort data from similar projects are basic ingredients of this approach. However, this approach also does not consider the scale and complexity of compliance-related activities in RE such as analysing large set of regulatory documents, analysis of a voluminous contract describing regulatory requirements in abstract, non-contiguous and implicit manners, and analysis of regulatory requirements that cross-cuts multiple components in large systems, which are described in (Berenbach et al., 2010; Nekvi and Madhavji, 2015).

4.2.2 Overview of Project

We describe the RE-part of a rail upgrade infrastructure project (Nekvi and Madhavji, 2015) from which we derived metrics for estimating the effort for carrying out requirements compliance work. The RE project had a 1000-page contract that describes approximately 12,000 requirements referred to as contractual requirements. Approximately 6% of the contractual requirements refer to a variety of 'regulations and engineering standards' (i.e., regulatory documents) with which they need to comply. The total number of regulatory documents referenced from the contract is in excess of 300. The size of approx. 25% of the documents is over 100 pages; a few amongst them are much larger (over 2000 pages).

The RE process had to identify regulatory requirements from the contract. Since the contractual requirements are specified at a high-level (i.e., not testable), the requirements for the project had to be derived from the contract (and regulatory documents) and categorised (e.g., subsystem, component, and cross-cutting). Also, note that regulatory documents often contained requirements that are characterised by numerous cross-references, ambiguities, conflicts, domain-specific terms, etc. (Nekvi and Madhavji, 2015).

Further, the elicitation of regulatory requirements involved frequent aid from legal and domain experts. Once elicited, the regulatory requirements were logged in a requirements management tool and appropriate tracing links were generated.

However, time and effort spent on analysing these documents are typified by numerous impediments (Nekvi and Madhavji, 2015). Thus, in such projects, there is a significant amount of uncertainty as to when the task of compliance analysis would actually be completed. This situation was a strong motivator to define appropriate metrics in order to reduce estimation variability and, hence, improve such project variables as resource allocation, time to completion and requirements (and hence system) quality.

4.3 A Method for Deriving Effort-estimation Metrics

The method we present in this section to derive the core set of metrics for RE work on regulatory compliance is based on three investigative questions. The general idea behind the method is to first determine the “scope of effort-consuming items” that is involved in the compliance work. Based on this scope, we want to identify the key characteristics that make them effort-critical. Finally, we want to define metrics that correspond to these effort-critical characteristics. The metrics then become the foundation of an effort-estimation model for compliance work in RE. These three investigative questions are:

Q1. What are the effort-critical activities and artefacts in the compliance work in RE?

Effort-critical activities and artefacts are those that are considered to take an inordinate amount of person-hours to accomplish the goals of those activities and artefacts. Since our objective is to determine metrics to estimate the effort needed for compliance tasks, it is important to identify the activities that contribute significantly to this effort so that they are not ignored in the overall effort estimation.

Q2. Which characteristics of the activities and artefacts identified in Q1 are primarily responsible for making them effort-critical?

Effort-critical activities and artefacts have complicated characteristics (e.g., non-contiguity of regulatory requirements in the contract, abstractness of requirements in the contract, and cross-references in regulatory documents) that are root causes for imposing impediments (and thus adding extra effort) in doing RE work. This question investigates characteristics of the artefacts and activities (Q1) that are effortful to analyse

Q3. What are the metrics that can be used to measure the effort-criticality level of the characteristics identified in Q2?

It is important to know the relative contribution of each of the effortful characteristics (of the artefacts and activities) to effort so that overall effort can be summed up. Such a figure for overall estimated effort guides critical project activities such as project resourcing, budgeting, and scheduling. This question (Q3) probes into metrics that can correspond to the effort-criticality level of the characteristics (Q2) of the activities and artefacts.

Below, we discuss techniques to derive effort-critical activities and artefacts (Q1) in Section 4.3.1; their characteristics (Q2) in Section 4.3.2; and the effort-estimation metrics (Q3) in Section 4.3.3.

4.3.1 Identification of effort-critical activities and artefacts (Q1)

In compliance work, the complexity of effort-critical activities typically originates from certain types of artefacts and associated activities. In our study (Nekvi and Madhavji, 2015), we obtained information about compliance work and its associated impediments: (i) through a couple of workshops; (ii) by gathering and analysing project artefacts such as contract, regulatory documents, and system descriptions; and (iii) by interacting with project staff. By analysing the gathered information, we determined the number of task (see the 1st column in Table 4-1) carried out by the analysts that addressed compliance-based issues (e.g., implicit regulatory requirements, diverse regulatory references in documents, and abstract requirements). It is this type of task that was not accounted for at the outset in a requirement engineering process and thus was a factor in under estimating the overall effort. Clearly, metrics (described in Section 4.3.2 and Section 4.3.3) need to be associated with this type of task for estimating compliance effort.

We grouped the identified tasks into clusters of artefacts and activities (see column two in Table 4-1) according to the guideline, i.e., logically grouping of related activities, (PMI, 2006) defined by the standard of *Project Managements Institute* for *Work Breakdown Structure (WBS)* that is widely used in effort-estimation (Trendowicz and Jeffery, 2014). These clusters represent project specific variables such as the contract, regulatory documents, and system structure. With respect to Table 4-1, the first column shows the list of effort-critical activities and artefacts, and the second column shows the clusters for the activities in column one. Each activity is numbered by a unique identifier so that they can be referenced from the textual descriptions.

Table 4-1 Effort Critical Activities and Artefacts

Effort-critical Activities and Artefacts (Activity ID)	**Clusters of Activities and Artefacts (Cluster ID)**
Identifying standards and regulations applicable to the target system (AC 01)	Obtaining relevant regulatory documents
Collecting identified regulatory documents and incorporating them as objects in RE tools (AC 02)	
Identifying regulatory requirements from contract (AC 03)	Analysing contractual complexities
Identifying abstract specification and determining their proper regulatory reference (AC 04)	
Identifying those contractual requirements having diverse regulatory reference and resolving any conflict among the referenced regulatory specifications (AC 05)	
Eliciting regulatory requirements that are implicitly mentioned in the contract (AC 06)	
Monitoring and managing changes made in regulations (AC 07)	Analysing complexities in regulatory documents
Identifying which requirements of the regulatory documents are relevant to the system and where in the system they apply (AC 08)	
Following all cross-referenced segments to understand the requirements correctly by their semantics (AC 09)	
Inter subsystem/component team communication to resolve compliance issues resulting from cross-cutting requirements (AC 10)	Analysing the aspects of large and complex system
Preparing separate subsystem and component specifications for third party developers (AC 11)	
Identifying requirements that cross-cut multiple subsystem and/or components (AC 12)	

Note, however, that RE activities and artefacts used can vary across projects. Thus, in the manner described in this section, one must consider project-specific variables.

4.3.2 Characteristics of effort-critical activities and artefacts (Q2)

In this step, we analysed each type of the artefacts and the associated activities in order to identify their underlying effort-critical characteristics (referred to as impediments in the RE process). It is important to identify these (effort-critical) characteristics because without knowing them it is not possible to determine the compliance workload which, in turn, is needed to estimate the compliance effort.

The criteria we used to identify the effort-critical characteristics include such aspects as: volume and complexity of associated artefacts, ad-hoc practices, lack of tool support, need for domain-expertise (usually external to project), and need for inter-team communication for cross-functional issues. Although there are no established criteria for determining effort-critical characteristics specifically for compliance work of RE, impeding factors for various RE activities and practices are described dispersedly in the RE literature. For example, (i) Ramesh, (1998) identified the impeding characteristics for traceability practices, which includes ad-hoc practices, use of external staff, incompatible tools, and needs for standards compliance; (ii) communication among diversified teams in distributed projects is regarded as a problem (Damian and Zowghi, 2003; and Al-Rawas and Easterbrook, 1996), (iii) size and complexity of regulatory documents is widely accepted as challenges to RE (Otto and Anton, 2007; Nekvi and Madhavji, 2015; and Breaux and Gordon, 2013); and others.

Using these criteria, we derived the effort-critical characteristics (impediments) as listed in column two of Table 4-2. The column three of Table 4-2 gives rationale of why the impediments (Table 4-2, column two) fit the effort-critical criteria mentioned above.

Table 4-2 Derived effortful activities and associating characteristics

Activity Cluster (ID)	Effortful Characteristics (Impediments)	Effort-Critical Aspect
Obtaining relevant regulatory documents	**Diversity of regulatory documents** in terms of engineering domains and regulatory authorities	It requires analysing an unbounded set of engineering standards and laws from diverse authorities for determining their relevance to system.
	Available format for regulatory text	Incorporation of regulatory codes available in hard copy or PDF as "objects" within RE tools is a manual and tedious.
Analysing contractual complexities	**Non-contiguity** of regulatory requirements in the contract	Regulatory requirements mixing in non-contiguous manner in the voluminous contract.
	Diverse regulatory references in contractual requirement	All referenced documents (codes) must be analysed for resolving possible conflicts and to define concrete requirements.
	Abstractness of contractual requirements	Those regulatory documents abstractly (without proper index) referenced need thorough analysis by domain experts from all the subsystems covered by

		the abstract requirements.
	<u>Implicit regulatory references</u> in contract	Eliciting implicit requirements from indirectly referenced documents needs help of domain-experts.
Analysing complexities in regulatory documents	**<u>Large number</u>** of relevant regulatory documents	Monitoring and managing legal changes made by external authorities (e.g., government officials) require dedicated role and technique.
	<u>Multi-Domain contents</u> in regulatory documents	Separate domain experts are required to analyse regulatory documents from various domain.
	Frequent **<u>cross-references</u>** within the regulatory text	All cross-referenced segments needs to be followed and understood correctly by their semantics, possibly with the help of legal experts.
Analysing the aspects of large and complex system	**<u>Vertical and lateral communications</u>** among the sub-teams	Inter subsystem/component team communication is required to resolve regulatory-related issues resulting from cross-cutting requirements.
	<u>Non-aligned contents</u> of contractual chapters with respect to system organisation	Contractual requirements need restructuring in order to generate subsystem or component specifications to be delivered to third party developers responsible for subsystem or component delivery.
	<u>Cross-cutting</u> requirements that apply to multiple components or sub-systems	"40% of the cross-cutting requirements are regulatory"(Nekvi and Madhavji, 2015) indicates substantive compliance effort.

4.3.3 Deriving metrics (Q3)

Below, we describe three analytical steps for deriving appropriate metrics for a given characteristic identified in Q2 (see Column two of Table 4-2 and Section 4.3.2):

<u>Step (i) Identify the type of items affected by the given characteristic:</u>

In this step, there is a need to identify the type of items to which the given characteristic (e.g., *cross-references*) belongs.

Example *item-types* are:

- Project requirements,
- Contractual requirements,
- Sections of a regulatory document, and

- System organisation into sub-systems and components.

Step (ii) Metrics concerning the breadth of impact:

In this step, we assess the *extent* to which the given characteristic (e.g., *cross-references*) exists in the item-type identified in step (i) (e.g., *sections of a regulatory document*).

Example metric is: *percentage of the sections of a regulatory document containing cross-references.*

Step (iii): Metrics concerning the depth of impact:

In this step, we assess the how *deeply* the *item-type* (identified in Step (i)) is affected by the given *characteristic*. In other words, it is the *intensity* with which the characteristic (e.g., *cross-references*) has an impact on an individual item (e.g., *a section of a regulatory document*).

Example metric is: *average number of cross-references per section of a regulatory document.*

4.4 Set of Derived Metrics

In this section, we present and discuss the set of metrics that have been derived using the method described in Section 4.3. The set of metrics presented here is similar but more refined than those metrics presented in Chapter 3 (Table 3-6, Section 3.8). In both cases, the metrics are produced from the same project data and importantly, they are founded partially on analysis of the impediments (Nekvi and Madhavji, 2015) previously identified.

However, metrics presented in Chapter 3 were not derived but emerged as a by-product of cursory analysis of the impediments causing effort-intensive activities in the project. Hence, they lacked the necessary vision of how they can be used in an actual effort-estimation model. Nevertheless, emergence of those metrics was critical for deriving revised metrics since it opened up scope for further analysis. On the contrary, metrics presented here are derived using a methodological process (discussed in Section 4.3) that guides the deriving process into three methodological steps, i.e., (i) identifying effort-critical activities (Section 4.3.1); (ii) identifying characteristics of effort-critical activities (Section 4.3.2); and (iii) defining appropriate metrics for each characteristic that relate to extent and depth of impact by the characteristics of artefacts and activities (Section 4.3.2). Such approach yields metrics that are linked to effort critical activities and their associative metrics; hence, are readily usable in an effort estimation model.

Below, for each cluster of activities and artefacts (See Table 4-1), we derive their corresponding set of metrics using the method discussed in Section 4.3. Since each cluster of effort-critical activities is centered on its unique set of artefacts, their corresponding metrics are clustered similarly and discussed altogether.

4.4.1 Metrics for activities associated with obtaining relevant regulatory documents

Identifying regulatory documents applicable to multi-domain systems is challenging since numerous engineering standards (e.g., ANSI B16.5, CSA C22.1, and AREMA) from various agencies define system requirements to follow. Further, there could be multiple regulatory authorities (i.e., states) that mandate certain regulations to be followed by systems operating within their legislative boundary. This is especially true for global products. To begin with, count of the engineering domains involved and count of the regulatory authorities relevant to the system to be operated can provide an initial rough estimate of the extent of effort needed to identify the regulatory documents relevant for the target system. This is because there will likely be more regulatory documents relevant for the system as more engineering domains and regulatory authorities are involved in the system. So the breadth of impact of the diversity of regulatory documents on the aspects of documents collection is estimated based on the number of engineering domains and regulatory authorities involved (see the 3rd column of the 1st row in Table 4-3). The average number of regulatory documents per engineering domain (or regulatory authority) indicates the depth of impact (see the 4th column of the 1st row in Table 4-3).

After collecting the documents, incorporating them into RE tools in order to enable automatic processing can become manual work if the formats of the documents are not compatible with the tools used. We define the sets of metrics for such characteristics (i.e., format of document) such as: number (#) of documents available in hard copy, PDF or other incompatible format for breadth of impact (see the 3rd column of the 2nd row in Table 4-3), and average size of the documents for depth of impact (see the 4th column of the 2nd row in Table 4-3).

Table 4-3 Metrics - Obtaining relevant regulatory documents

Effort-critical Characteristics (Activity ID)	Affected Item-types	Metrics for Breadth of Impact	Metrics for Depth of Impact
Diversity of regulatory documents (AC 01)	System domains	# of engineering domains and regulatory authority involved in system	Avg. # of documents per engineering domain
Format of regulatory documents (AC 02)	Regulatory documents	# of documents available in hard copy or PDF	Avg. size of the incompatible documents

4.4.2 Metrics for activities associated with contractual complexity

The set of metrics for RE activities associated with *contractual complexity* is presented in Table 4-4. In Table 4-4, the 1st column shows the effort-critical characteristics (e.g., non-contiguity, diverse regulatory references, and abstractness of requirements in contract), the 2nd column shows the item-types (e.g., contractual requirements or contract document as a whole) affected by the corresponding characteristics. The metrics for the breadth and depth of impact by the corresponding characteristics over the item-types are provided, respectively, in the column three and the column four. With reference to the 2nd column of Table 4-4, the item-type affected by all such activities under the cluster of *contractual complexity* is contract documents and its requirements. However, the *extent* to which this item-type (contract) is affected (i.e., breadth of impact) by different characteristics (e.g., non-contiguity, diverse regulatory references, and abstractness of requirements in the contract) is not always the same. For instance, metrics for *breadth of impact* by the characteristics "non-contiguity" and "diverse regulatory references" are respectively: '*number (#) of pages having regulatory requirements*' and '*number (#) of requirements having di-*

verse references' (see the first two rows in column three of Table 4-4). Likewise, we can observe the difference in the metrics for *depth of impact* across different characteristics (see the 4th column of Table 4-4). For example, metrics for *breadth of impact* for the characteristic "non-contiguity" is *"Avg. ratio of regulatory and non-regulatory requirements per page"* whereas metrics for the characteristic "implicit requirements" is *"Avg. number (#) of implicit requirements per section."*

Table 4-4 Metrics - Contractual Complexity

Effort-critical Characteristics (Activity ID)	Affected Item-types	Metrics for Breadth of Impact	Metrics for Depth of Impact
Non-contiguity (AC 03)	Contract document	# of pages having regulatory requirements	Avg. ratio of regulatory and non-regulatory requirements per page
Diverse Regulatory references (AC 04)	Contractual regulatory requirements	# of requirements having diverse regulatory references	Avg. # of regulatory references per requirement of this kind[2]
Abstractness (AC 05)	Contractual regulatory requirements	# of requirements that are abstract to some degree	Avg. level of abstraction[3] per abstract requirement
Implicit requirements (AC 06)	Contract document	# of sections having implicit requirements	Avg.# of implicit requirements per section

4.4.3 Metrics for activities associated with complexity in regulatory documents

Table 4-5 shows those metrics that are associated with activities clustered in *complexity in regulatory documents*. Regulatory text (or pages or sections) is the sole item-type that is affected by the characteristics present in the set of activities under the cluster of *complexity in regulatory documents* (see the 2nd column of Table 4-5). Metrics derived for assessing the *breadth of impact* account for the number of regulatory documents (or total sections therein) that are particularly affected by given characteristics (see the 3rd column

[2] further techniques can use such metrics as avg. number of sections or pages to be reviewed per reference

[3] The level of abstraction can be quantified as: System level - 3, sub-system level-2, component level-1.

of Table 4-5). The last column of Table 4-5 shows the metrics for *depth of impact*, which dictate the average degree of complexity associated with each unit of the affected item-type.

It is interesting to note that multiplying the metrics values concerning breadth of impact by the metrics values of depth of impact would result in total volume of impact; this would indicate the estimate of effort required for the corresponding activity. For instance, with reference to the characteristic of multi-domain content (the 2nd row of Table 4-5), if there are *x* number of documents containing multi-domain contents and on average, *y* domain experts are required per document for eliciting requirements, then x * y experts would be required for analysing those regulatory documents for determining their segment-wise relevance to the system and for eliciting appropriate requirements (in collaboration with requirements engineers).

Table 4-5 Metrics - Complexity in Regulatory Documents

Effort-critical Characteristics (Activity ID)	**Affected Item-types**	**Metrics for Breadth of Impact**	**Metrics for Depth of Impact**
Large number & voluminous Size (AC 07)	Regulatory text (or pages or sections)	# of regulatory documents relevant to system	Avg. size of regulatory documents
Multi-domain contents (AC 08)	Regulatory text (or pages or sections)	# of documents having multi-domain content	Avg. # of domain experts required per document
Cross-references (AC 08)	Regulatory text (or pages or sections)	# of sections having cross-references	Avg. # of cross-references per section

4.4.4 Metrics for activities associated with a large and complex system

The set of metrics for the activities associated with a *large, complex system* is presented in Table 4-6. Although the effort-critical characteristics underlying the activities of the cluster of a large, complex system relate to the complex system organisation (i.e., several sub-systems, components and their interfaces), the affected item-types are either the contract or aspects of the communication between sub-teams working for various subsystems and components (see the 2nd column of Table 4-6). Metrics for breadth and depth of impact for these activities are presented in the 3rd and 4th column of Table 4-6 respectively. We describe their use in the creation of an effort estimation model in Section 4.5.

Table 4-6 Metrics - Large and Complex System

Effort-critical Characteristics (Activity ID)	**Affected Item-types**	**Metrics for Breadth of Impact**	**Metrics for Depth of Impact**
Vertical & lateral communications (AC 10)	communications in RE	(i) # of team at sub-system level ; (ii) Avg. # of components per sub-system	Avg. # of vertical levels involved in the project, e.g., component team to subsystem team to system team makes three levels.
Non-aligned content (AC 11)	Contract	# of chapters in the contract	Avg.# of sub-system requirements per chapter
Cross-cutting requirements (AC 12)	Contract	# of cross-cutting requirements in contract	Avg. # of components per cross-cutting requirements

4.5 Use of Metrics in Effort-Estimation Model

In this section, we illustrate how the metrics derived in Section 4.4 (See Table 4-3, Table 4-4, Table 4-5, and Table 4-6) can contribute to creating an effort-estimation model for compliance work in RE. In particular, we focus on where in the overall design of an effort-estimation model the metrics would fit.

Effort-estimation of software development is a challenging process (Sommerville, 2006) that needs to consider myriad of factors such as types of workload (e.g., design, coding, testing, etc.), product (or system) characteristics (e.g., embedded, multi-domain, safe-critical, and legacy), development process (e.g., agile, waterfall, incremental, etc.), productivity of personnel doing the work (e.g., experience, training, and technical skills), process characteristics (e.g., degree of reuse, degree of integration of components, tools capability, and schedule), etc. Identifying and quantifying these factors that influence required effort of a project is fundamental for creating a model for effort estimation. Appropriately defined metrics enable project managements to measure the impact of the corresponding factors to project effort.

Below, in Section 4.5.1, we discuss algorithmic modelling (Boehm, 1981) being the probable estimation technique using our derived metrics, show the links between the derived metrics and bottom-up approaches for effort-estimation in Section 4.5.2, assign the metrics into appropriate parameters of an algorithmic model in Section 4.5.3.

4.5.1 Recognising appropriate estimation technique

It is important to recognise the particular estimation technique among the variety of established techniques (described in Section 4.2 - such as algorithmic modelling (used in COCOMO 81 (Boehm, 1981) and SLIM model (Putnam, 1978)), expert judgments (Hughes 1996), estimation by analogy (Shepperd and Schofield, 1997)), for which the derived metrics would have direct use.

Expert judgment involves consultation among a group of experts who provide estimations from their experience (Hughes, 1996). The experts do not necessarily use any explicit metrics in the decision making; their judgment process is subjective and does not follow any standard pattern. However, communication and coordination among the experts to reach a satisfactory conclusion is provided by the use of the Delphi technique (Dalkey and Helmer, 1963).

In the 'estimation by analogy' (Shepperd and Schofield, 1997), designated experts use their own experiences to extract important project characteristics to determine their similarity with previous projects for which historical records are kept. Then, based on historical records from such analogous projects, they estimate effort for a new project. This technique also does not consider using any explicit metric.

In contrast, metrics are the building blocks for the algorithmic modelling technique (i.e., *"a model is developed using historical information that relates some software metric - usually its size - to project cost, then an estimate is made of that metric and the model predicts the effort required"* (Sommerville, 2006, p 643)). This kind of model predicts effort (or cost) based on a set of parameter values representing project and product characteristics. Several metrics determine the parameters of the model. Although primarily used for software development cost estimation (e.g., Boehm, 1981), algorithmic modeling can be useful for other areas in SE such as estimates for risk assessment based on alternative strategies, to make decision on outsourcing and reuse (Boehm et al., 2000). Therefore, algorithmic modelling provides a generic computational model that can be applied to other domains as well provided that appropriate metrics for the domains are defined. The metrics we derived (described in Section 4.4) are directly associated with the effort-critical activities and artefacts for compliance work in RE. So they are considered to be particularly constructive to an algorithmic modelling technique. We adopt this widely used algorithmic modelling technique as a basis for illustrating the usages of the derived metrics.

Algorithmic Modelling Technique:

The standard formula for algorithmic modelling takes the following form (Boehm, 1981):

$$\textbf{Effort} = \textbf{A} * \textbf{Size}^{\textbf{B}} * \textbf{M} \qquad (4.1)$$

In the above formula, parameter *A* is a constant that represents organisational practices and type of systems to be developed. Typically, a different fixed value is assigned for this parameter depending on the complexity of the project, system domains, and organisational capability to handle such complexity. In the simplest form (e.g., COCOMO 81(Boehm, 1981)), three levels (e.g., simple, moderate, and embedded) of such complexity are determined where each level has its fixed value for parameter A (i.e., Simple - 2.4, Moderate - 3.0, and Embedded - 3.6). Complexity of a project is to be determined by experts based on experience.

Size corresponds to estimates of the amount of associated work (for software development, typical metrics for size are lines of codes, function points, and object points). In an algorithmic model, size is the main driver of effort.

The parameter B is the exponent value (typically ranges between the values of 1 and 1.5 in software development estimation) used to reflect the non-linear nature of the increase of project cost (and effort) by the increase of project size in software development. This is due to exponential increase of communication overhead, configuration management, and system integration by the increase of project size (Sommerville, 2006). Similar to project complexity metrics (i.e., parameter A), the exponent value (B) is fixed for the three types of projects (i.e., Simple - 1.05, Moderate - 1.12, Embedded - 1.20) in COCOMO 81 (Boehm, 1981).

Lastly, the parameter M is a multiplier whose value depends on other development process and product attributes (such as: degree of reuse, platform difficulty, experiences of personnel, and degree of support facility such as tools and training). In COCOMO II (Boehm et al., 2000), parameter M is a composite value that is based on seven process and product attributes, where value of the attributes is estimated using a six-point scale (i.e., 6 corresponds to very high and 1 corresponds to very low).

The resultant effort is typically measured in person-months (PM). It is the amount of work to be performed by one person in one month.

It is worthwhile to mention that values of the parameters of the above formula are determined based on historical project data (e.g., average effort required per unit of workload) and current project characteristics (e.g., size of the proposed software). In the core, this formula estimates effort as a function of the project size and complexity, and capability of personnel, where the size metric is the main factor.

Effort Estimation Models for Compliance Work in RE:

In order to develop an effort estimation model for compliance work in RE, our goal would be to answer the following question:

Question: How do we determine the values of the parameters used in the above formula for compliance work in RE?

The very first challenge of this attempt is determining the *size* of compliance work in RE. Whereas, Lines of Codes (LOC), Function Points (FP) or Objects Points (OP) are widely used metrics for size in overall software development, there are no a priori metrics for *size* in compliance work of RE. Furthermore, metrics for the project complexity (parameters *A*), exponent for size metric (*B*) and product and process characteristics (parameter *M*) in RE projects are not established yet. Not only is the set of metrics for various model parameters known but also there is no known record of effort from past projects against such metrics. Given the set of such metrics along with a historical dataset from similar contextual projects, established curve-fitting techniques such as regression analysis (Srinivasan and Fisher, 1995) and neural network (Wittig and Finnie, 1994) could be applied to produce realistic value for the model parameters (i.e., *A*, *B*, and *M*), which is out of scope of this chapter. Nevertheless, one of the very first steps towards this goal is to determine the metrics and/or the set of project and process attributes (outlined above in 'algorithmic modelling technique' sub-section) to be used in the metrics.

4.5.2 Bottom-up approaches

There are two approaches for effort estimation: (i) top-down and (ii) bottom-up. The top-down approaches estimate a summary effort for a total project (i.e., includes all project activities) directly and such a summary estimate is then broken down into proportionate efforts for individual activities (Trendowicz and Jeffery, 2014). For example, in the estimation by analogy technique (Shepperd and Schofield, 1997), a new project as a whole is compared with historical projects of similar complexity to predict overall effort for it. The top-down approach is functional when a collection of historical project data is available and particularly when enough information regarding details of each project activities and tasks for the new project is not well known.

In contrast, bottom-up approaches divide project work into individual bottom-level activities and estimate the efforts for completing each activity (called bottom-up estimates) (Trendowicz and Jeffery, 2014). Finally, the total project effort is estimated by aggregating all bottom-up estimates made for each activity. Obviously, availability of enough supported information regarding bottom-level activities is a precondition to this approach.

In general, bottom-up approaches lead to better estimates than top-down approaches (Trendowicz and Jeffery, 2014). This is because of the following reasons:

(i) It is easier for experts to estimate a smaller piece of project activity than an entire project, thus the margin of errors is reduced;

(ii) In practice, it is likely that bottom-up estimates would be a mixture of under and over the actual efforts; however, such errors of opposite direction (under and over) will cancel each other's effect and will result in a lesser error than top-down estimates.

(iii) High-uncertainty and complex tasks are often underestimated or forgotten when the project as a whole is considered for estimation, leading to underestimation of total effort.

Further, in the context of expert-based effort estimation (e.g., (Hughes, 1996)), Jørgensen, (2004) suggests to use bottom-up approaches if bottom-level activities are known and there is a lack of necessary information about similar historical projects to conduct a top-down approach.

Historical records for a similar type of project from which we derived the metrics (see Table 4-7) are currently not adequate. Moreover, there were numerous impediments (Nekvi and Madhavji, 2015) in the studied project such as non-contiguity of regulatory requirements in the contract, regulatory requirements were specified in a number of abstract forms, regulatory documents contain substantial implicit requirements for the system, etc. Such impediments were not completely anticipated until later in the RE process and intuitively, they contributed to underestimation of the required effort. Most importantly, the metrics are directly derived from individual effort-critical activities (bottom-level activities), which make them suitable for those bottom-level activities. Thus, we select a bottom-up approach of effort estimation.

Despite providing more reliable estimates, one drawback of the bottom-up approach is that the more granular the estimates are in terms of project activities and tasks, the more time-consuming the estimation process becomes and the more need of expertise is required by experts doing the bottom-level estimates. In this regard, a set of a priori metrics (i.e. unit of measurements) for various bottom-level effort-consuming project activities can be a useful basis for experts to make reliable estimates. This is another aspect where the derived metrics have practical use.

4.5.3 Assigning metrics to model parameters

In bottom-up approaches, efforts are estimated separately for each bottom-level project activity for which adequate historical effort data is available. Overall effort is then estimated by summing up all activity-wise efforts. With regards to Table 4-1, as many as 12 RE activities associated with compliance work are identified as being effort-critical. Their corresponding metrics are derived which are presented in the four tables (i.e., Table 4-3, Table 4-4, Table 4-5, and Table 4-6) representing the four clusters of the activities (e.g., contractual complexity (Table 4-4), complexities in regulatory documents (Table 4-5), and large and complex system (Table 4-6). We assign these metrics to appropriate parameters, i.e., input, size, project characteristics (A), exponent (B) of size, product and process characteristics (M), of the algorithmic model (shown in equation 4.1 in Section 4.5.1) below.

Inputs of the model: Every algorithmic-based estimation technique requires *inputs* as a basis for making the estimation of effort (e.g., *requirements document* is a typical input for software development effort estimation) (Somerville, 2006). With regards to the activities and artefacts defined under the cluster of *contractual complexity* (see Table 4-1), the item-type (i.e., artefacts-type) that is predominant in the RE analysis is the contract document (see the 2nd column of Table 4-4). Hence, the contract becomes the *input* for the effort estimation model for these clustered activities (see the 2nd column along the 3rd to 6th row in Table 4-7). Likewise, regulatory documents become the input for those activities associated with regulatory analysis (see the 2nd column along the 7th to the 9th row, in Table 4-7). In Table 4-7, the 2nd column of the 10th and 11th row shows that both the contract and system description documents serve as inputs for the activities under *large, complex systems* cluster.

Size of the Workload: Size of workload is the dominant driver of effort required to complete the workload. Sizing is a process that estimates the probable size of a workload while effort estimation predicts the effort needed to complete the workload. The productivity of the persons doing the work relates the size of work and the required effort. In software development, the most widely used metrics for software sizing are *Lines of Code* (LOC), *Function Point* (FP) and *Object Point* (OP).

On the other hand, compliance-based works in RE produces heterogenic work products such as: *applicability matrixes* (that depict the relationships between regulatory segments and corresponding components of a system), project requirements, and trace links among the contract, project requirements, and corresponding regulatory codes. These work products are outcome of variety of compliance-related project activity carrying out towards ensuring compliance of requirements. It includes analysis of contract, analysis of regulatory documents, analysis of project requirements, and analysis of the system. Therefore, technically there is a need for defining separate units of measurement (i.e., metrics) for estimating the size for corresponding activities so that each type of compliance activities involved in the RE projects has associated metrics.

Generally, the size of contract documents containing a number of regulatory requirements indirectly indicates the extent of RE activities associating with the contract. More specifically, those parts of the contract specifying regulatory aspects (e.g., regulatory requirements and references to relevant regulatory documents) would best approximate the size of the compliance-related workload in RE projects. These parts can be denoted by *regulatory parts* of the contract.

The size of the *regulatory parts* of the contract can be measured by either of the two metrics (as indicated by the 3rd column of Table 4-4): (a) number of regulatory requirements in the contract (i.e., number of contractual regulatory requirements), and (b) number of pages (or segments) in the contract which comprise its regulatory parts. However, for a specific activity, instead of using entire regulatory parts of the documents only those portions of regulatory parts that are affected by corresponding effort-critical characteristics can be considered for the size metrics. Such a step lets us exclude from calculations the extraneous regulatory parts from consideration and thus provides more realistic estimation of size. Let us assume that the variable *s* denotes the total size of the contract, n is the number of excluded parts and y is the average size of a part to be excluded. Then, the effort-prone size of the (contract) document is equal to (x - n*y).

Therefore, the size metric is likely to vary on the particular activity for which effort is to be estimated (see the 4th Column in Table 4-7). For example, the size metrics for the activities "*identifying abstract specification and determine their proper regulatory reference* (AC 04)" and "*identifying contractual requirements having diverse regulatory reference and resolve any conflict among the referenced regulatory specifications"* (AC 05) (shown in Table 4-1), respectively, are: '*number of requirements having diverse regulatory references'* and *'number of regulatory requirements that are abstract to some degree*' (see the column four of the 4th and 5th row in Table 4-7). It is likely that the number of regulatory requirements having diverse references is not same as the number of abstract requirements since there is no causal relationship between these two characteristics.

On the other hand, *'number of pages having regulatory requirements'* is the size metric for the activity "identifying regulatory requirements from contract (AC 03)" (see Table 4-7), which accounts for pages instead of requirements as the unit of measurements.

For the activities clustered under *complexity in regulatory documents* (e.g., "following all cross-referenced segments to understand the requirements correctly by their semantics" - shown in Table 4-1), it is essential to analyse the entire document. Size metrics for these activities thus corresponds to the size of the regulatory documents (in terms of page numbers) relevant for the system (see the 4th column along the 7th to 9th row in Table 4-7).

Product and Process Characteristics (M): Factors that contribute to product and process characteristics (M) are subjective, i.e., determination of factors and their evaluation criteria likely to vary from person to person. This parameter is a multiplier to the model, which is usually comprised of a range of process and product characteristics (such as: degree of reuse, platform difficulty, experiences of personnel, and degree of support facility such as tools and training) that have an influence on effort. The decision on considering the characteristics depends on the particular contexts of projects (e.g., types of application, system domains, developing organisations, and concerned regulatory body). Our analyses on the metrics for 'depth of impact' (see the 4th column in Table 4-4, Table 4-5, Table 4-6) suggest that they relate to the degree of complexity for each unit of size. For example, with respect to the 2nd row of Table 4-5 (metrics for complexity in regulatory documents) defining metrics for the characteristic "multi-domain content", the size metrics is "# of documents having multi-domain content" where the number of documents having this characteristic estimates the size of the workload pertaining in dealing with the issue of multi-domain content of regulatory documents. This issue is translated by the considerations that acquainted domain expert is likely to be employed to understand the content. Therefore, the more documents of such kind a project has to deal, the more domain experts will be required. Hence, with the increase of domain experts the required effort in doing such work will also be increased. Now, metrics for 'depth of impact' for this characteristic is defined as "*average number of domain experts required per (regulatory) document*" that expresses further complexity involved per *document* (i.e., unit of size in this case). Nekvi and Madhavji, (2015) showed that multiple experts may be needed to understand requirements of one regulatory documents due to their diversified technical contents. If average number of domain experts required per document increases, then it would require increased effort. Therefore, we suggest using metrics of this kind (depth of impact) to be one of the factors that can be used to construct parameter M. In Table 4-7, the last column summarizes all such metrics that can contribute to the corresponding parameter M.

Project Complexity (A): This parameter is used to represent the complexity of systems and organisational complexity. In its simplest form, discrete constants are assigned to this parameter depending on the project complexity. For example, in the COCOMO 81 (Boehm, 1981) model, the assigned constants for three types of projects are:

- Simple (*"well understood applications developed by small teams"*) - 2.4
- Moderate ("*more complex projects where team members have limited experience of related systems*") - 3.0

- Embedded (*"Complex projects where software is part of a strongly coupled complex of hardware, software, regulations, and operational procedures"*) - 3.6

Similarly, RE work is also influenced by project complexity, especially in multi-domain, compliance-oriented, large, and complex systems (Nekvi and Madhavji, 2015). At least, communication overhead among various sub-teams working for different sub-systems and components in RE is increased by the increase of project scale and complexity. Communications is a critical issue in large projects since increased communication overhead due to the increase of team members to catch up the schedule actually incurs more efforts (Brooks, 1995). Further, diverse culture, nationality, and knowledge of the personnel working in several teams impede coordination among them (Damian and Zowghi, 2003).

Table 4-6 (in Section 4.4.4) presents metrics for effort-critical activities (e.g., Inter sub-system/component team communication to resolve compliance issues resulting from cross-cutting requirements, prepare separate component specifications for third party developers) dealing with the issues (e.g., need for inter-team communications and cross-cutting aspects of requirements) mainly resulting from scale and complexity of systems.

We propose that the two derived metrics (i.e., (i) number (#) of teams at sub-system level, and (ii) average number (#) of components per sub-system) representing the breadth of impact by the characteristic '*vertical and lateral communications*' (1^{st} row of Table 4-6) to be used for constructing the project complexity (A) parameter (see the 3^{rd} column of Table 4-7). This is because they represent the scale of the system by its number of sub-systems and components.

Exponent (B) for Size Metric: This parameter reflects non-linearity of effort with respect to project size. As yet, we do not have a metric defined for this exponent. The value of this parameter is to be obtained through historical data set (outside the scope of this thesis) that would relate how the effective size of the workload is increased by the increase in project size.

Table 4-7 Parameters of Estimation Model

Activity ID	Input	Project Complexity(A)	Size	Exponent (B)	Process & product characteristics (M)
AC 01	System domains information	(i) # of sub-system wise project teams; (ii) Avg. # of components per subsystem	# of engineering domains and regulatory authority involved to system	(To be determined from analysis of historical data set)	Avg. # of documents per engineering domain
AC 02	Regulatory documents		# of documents available in hard copy or PDF		Avg. size of the incompatible documents
AC 03	Contract		# of contract pages having regulatory requirements		Avg. ratio of regulatory and non-regulatory requirements per page
AC 04	Contract		# of requirements in the contract having diverse regulatory references		Avg. # of references per requirement of this kind
AC 05	Contract		# of regulatory requirements in the contract that are abstract		Avg. level of abstractions per abstract requirement
AC 06	Contract		# of sections in the contract having implicit requirements		Avg.# of implicit requirements per section
AC 07	Regulatory Documents & Contract		# of regulatory documents relevant to system		Avg. size of regulatory documents
AC 08	Regulatory documents		# of documents having multi-domain content		Avg. # of domain experts required per document
AC 09	Regulatory documents		# of sections having cross-references		Avg. # of cross-references per section
AC 11	System descriptions & contract		# of chapters in the contract		Avg.# of sub-system requirements per chapter
AC 12	System descriptions & contract		# of cross-cutting requirements		Avg. # of components per cross-cutting requirements
AC 10	This activity (communication) is actually part of other activities rather than being on its own as a separate activity. Therefore, metrics associated with this activity becomes parameters (i.e. One factor for coefficient A) for the model for other activities where communication overhead is accounted.				

4.6 Discussion

In Section 4.3, we propose a three-step method for deriving metrics for complexities associated with regulatory compliance-oriented RE activities and artefacts. This method probes into identifying (effort-critical) characteristics (e.g., frequent cross-references in regulatory documents, abstractness of regulatory requirements in the contract, diverse regulatory references in requirements - see Table 4-2) from each effort-critical artefact and activity (see Table 4-1). Consequently, applying this method to one RE project (part of a contractual systems engineering project), we derive a set of metrics (see Table 4-3, Table 4-4, Table 4-5, and Table 4-6 described in Section 4.4) for each of these effort-critical activities and artefacts.

One apparent drawback regarding completeness of the metrics is that they are derived from one case study project, thus, are not generic enough for diverse projects in industry. With respect to completeness of metrics, it is not impossible to derive new metrics other than the ones described here by applying the method (shown in Section 4.3) to more projects. Hence, metrics derived in this chapter provide an important step towards this achievement.

Although deriving metrics from only one case study has generalisability threats, it is important to mention here that there is indeed literature support for certain impediments upon which our metrics are based. For example: "non-contiguity of regulatory requirements" (Breaux and Anton, 2008; Ghanavati et al., 2014); "cross-references" in regulatory documents (Breaux and Gordon, 2013); and some others.

Another technical issue that pertains to some of the metrics is how does one know a priori the metric value (e.g., percentage of abstract requirements, ratio of regulatory and non-regulatory requirements in the contract, average number of cross-references per section of a regulatory document) before actually doing the work? In this regard, there are at least two approaches to address this issue: (i) to maintain a record of average effort for such metrics from historical projects; or (ii) to use a random sample from the corresponding artefacts to obtain the value of the metrics of interest, e.g., to determine the percentage of abstract requirements in the contract, one can consider analysing some sample pages of the contract instead of analysing entire thousand-page long contract.

Further, we have demonstrated that the set of metrics (derived in Section 4.4 using our three-step method described in Section 4.3) were constructive towards creating an algorithmic-based effort estimation model (Boehm, 1981) (see Section 4.5.1). The algorithmic model is compared with other established effort estimation techniques (such as estimation by analogy (Shepperd and Schofield, 1997) and expert judgment (Hughes, 1996)) that are not based on explicit use of metrics.

Since the derived metrics are directly associated with complexities associated with various compliance-related RE artefacts and activities (shown in Table 4-1), it is very possible to estimate effort per activity given that productivity of doing the corresponding activities are known. Such activity-wise estimates of effort across all activities can be summed up to obtain the total estimate. This is why bottom-up estimation strategies (Trendowicz and Jeffery, 2014) over top-down strategies have appeared as more appropriate (see Section 4.5.2). By using a bottom-up strategy, the planned model becomes more inclusive, customised, and usable for diverse projects since one can add new metrics and select among the metrics (presented in Table 4-7) depending on the complexities involved in the particular project.

In Section 4.5.3, we illustrated the fitness of the derived metrics (e.g., # of contract pages having regulatory requirements, # of requirements in the contract having diverse regulatory references, average number of cross-references per page, etc.) to appropriate parameters of an established effort estimation model (Boehm, 1981) by assigning the metrics to model parameters such as size, project characteristics as multipliers, an exponent of size, and product and process characteristics as multipliers (see Table 4-7).

In an actual bottom-up effort estimation model, effort will be estimated by aggregating the effort estimate for each activity. Thus far in this chapter, we define metrics that capture factors contributing to effort such as size, project complexity, and product and process characteristics for each activity. However, the productivity of personnel doing those activities is an essential element to effort estimation model that embeds as coefficient value in the model. Obtaining accurate coefficients needs access to historical records from similar projects which are rare. Metrics presented in this chapter will guide researchers and industry to determine what effort-related data they need to collect. This is another implication of the metrics.

4.7 Conclusions

Compliance work on requirements can be difficult and arduous because of: unbounded cross-references within and across documents, ambiguity in the content, abstractness of the requirements, multi-domain complexity, levels of jurisdictions to contend with, and others (Nekvi and Madhavji, 2015). This situation makes the task of estimating the effort needed for requirements compliance work particularly challenging. Traditional effort estimation techniques (see Section 4.2), normally used for estimating development effort, are not suited to estimating requirements compliance work that involves characteristics such as those described in Table 4-2.

Section 4.3 proposes a new method for deriving effort-estimation metrics for conducting requirements compliance work in a large systems engineering project. Further, by using this method we derived a number of metrics that relate to key parameters of an algorithmic-based effort estimation model for compliance work of RE (see Section 4.5), including; (i) size metrics such as number of pages (of contract) having regulatory requirements, number of requirements having diverse regulatory references, number of regulatory documents relevant to system, number of sections (of regulatory documents) having cross-references, etc.), (ii) project complexity metrics (i.e., number of sub-system teams, and average number of components per subsystem), and (iii) product and process characteristics metrics such as average ratio of regulatory and non-regulatory requirements per page (of contract), average number of implicit requirements per section of a regulatory document, average size of the regulatory documents, average number of domain experts required per document, etc (see Table 4-7 in Chapter 4).

On the basis of these findings, we conclude that effort estimation metrics are derivable using our proposed method and the derived metrics can be used in the development of an actual effort estimation model for requirements-compliance. Thus, from the above, the key contributions of this chapter are: (i) a method for deriving key effort metrics and (ii) a set of metrics for estimating the effort needed in requirements-compliance.

While the preliminary method and the derived metrics that are described in this chapter is encouraging, much work still remains to be done; for example, solidifying the metrics (shown in Table 4-7) through more empirical studies, constructing and validating the effort estimation model, and transferring the model for productive use in industry.

4.8 References

Albrecht, A., and Gaffney, J.E., Software Function, Source Lines of Code and Development Effort Prediciton: A software Science Validation. IEEE Trans. On Software Engineering, Vol. 9(6), pp. 639–648, 1983.

Al-Rawas, A. and Easterbrook, S., Communication problems in requirements engineering: a field study, NASA Technical Report, 1996.

Beatty, J., Requirements Estimation for Business Analysts. WWW.requirements.seilevel.com/requirements-estimation-for-business-analysts-free-estimation-tool-download, 2012.

Beck, K., Extreme programming explained: embrace change. addison-wesley professional, 2000.

Berenbach, B., Lo, R., and Sherman, B., Contract-based Requirements Engineering. Int. Workshop on Requirements Engineering & Law, IEEE Computer Society, Los Alamitos, CA, USA, pp. 27-33, 2010.

Boehm, B. W., Clark, Horowitz, Brown, Reifer, Chulani, Madachy, R., and Steece, B.: Software Cost Estimation with COCOMO II with Cdrom (1st ed.), Prentice Hall PTR, Upper Saddle River, NJ, USA, 2000.

Boehm, B. W., Software Engineering Economics. Prentice Hall, 1981.

Breaux, T.D. and Anton, A.I., Analyzing Regulatory Rules for Privacy and Security Requirements. IEEE Trans. on Software Engineering. Vol. 34(1), pp. 5-20, 2008.

Breaux,T.D. and Gordon,D.G., Regulatory Requirements Traceability and Analysis Using Semi-Formal Specifications. 19th Int.Work.Conf. on REFSQ, pp. 141-157, 2013.

Brooks, F.P., The Mythical Man-Month, Addison-Wesley Longman Publishing Co. Inc., MA, USA, 1995.

Cleland-Huang, J., Czauderna, A., Gibiec, M., Emenecker, J., A machine learning approach for tracing regulatory codes to product specific requirements. 32nd ACM/IEEE Int. Conf. on Software Engineering (ICSI), pp. 155-164, 2010.

Dalkey, N. and Helmer, O., An Experimental Application of the Delphi Method to the Use of Experts, Management Science, Vol. 9, pp.458-467, 1963.

Damian, D. E., and Zowghi, D., RE challenges in Multi-site Software Development Organisations. Requirements engineering, Vol 8.3, pp.149-160, 2003.

Fenton, N. E., and Neil, M.,Software metrics: roadmap. Proceedings of the conference on the Future of Software Engineering, pp. 357-370, 2000.

Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L., How Common and How Large Are Cost Overruns in Transport Infrastructure Projects? Transport Reviews, 23, 71–88, 2003.

Ghanavati, S., Rifaut, A., Dubois, E., Amyot, D., Goal-oriented Compliance with Multiple Regulations. IEEE 14th Requirements Engineering Conference, pp. 73-82, 2014.

Goldsmith,R.F., Reliably Estimating the Soft. Requirements Effort.Tech Target, 2010.

Griss, M.L. and Wosser, M., Making Reuse Work at Hewlett-Packard. IEEE Software, Vol 12(1), pp. 105-107, 1995.

Hoffman, H.F., and Lehner, F., Requirements Engineering as a Success Factor in Software Projects. IEEE Software, pp. 58-66, 2001.

Hughes, R. T., Expert judgment as an estimating method, Information and Software Technology, Vol. 38, pp. 67-75, 1996.

Jørgensen, M., Top-down and Bottom-up Expert Estimation of Software Development Effort.Information and Software Technology, vol. 46, no. 1, pp. 3–16, 2004.

Karner, G., Resource Estimation for Objectory Projects, Objective Systems SF AB, 17, 1993

Kauffman, R. and Kumar, R., Modeling Estimation Expertise in Object Based ICASE Environments, Stern School of Business Report, New York University, January,1993.

Mukhopadhyay, T., Vicinanza, S.S., and Prietula, M. J., Examining the Feasibility of a Case-Based Reasoning Model for Software Effort Estimation. MIS Quarterly, Vol. 16(2), pp. 155-171, 1992.

Nekvi, M.R.I., Noorwali, I. and Madhavji, N.H., Deriving Metrics for Estimating the Effort Needed in Requirements Compliance Work. International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ), Springer International Publishing, pp. 135-141, 2016.

Nekvi, M.R.I, Personal communication with personnnel from industry, March 2012.

Nekvi, M.R.I. and Madhavji, N.H., Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects: A Case Study. ACM Trans. on Management Information Systems, Vol. 5(3), pp. 15:1-15:35, 2015.

Project Management Institute, Practice Standard for Works Breakdown Structures, 2nd Edition. Project Management Institute, Inc., Pennsylvania, USA, 2006.

Putnam, L.H., A General Empirical Solution to the Macro Software Sizing and Estimating Problem. IEEE Transactions on Software Engineering, Vol. SE-4, pp. 345-361, 1978.

Ramesh, B., Factors influencing requirements traceability practice.Communications of the ACM, 41(12), pp.37-44, 1998.

Shane, J. S., Molenaar, K. R., Anderson, S., and Schexnayder, C., Construction Project Cost Escalation Factors. Journal of Manage. in Engineering, Vol. 25(4), pp. 221-229, 2009.

Shepperd, M. and Schofield, C., Estimating Software Project Effort Using Analogies. IEEE Trans. On Software Engineering. vol 23(12), pp. 736-743, 1997.

Sommerville, I.: Software Engineering (7th ed.), Pearson Education, 2006.

Srinivasan, K. and Fisher, D., Machine Learning Approaches to Estimating Development Effort. IEEE Transaction on Software Engineering. vol 21(2), pp. 126-137, 1995.

Trendowicz, A. and Jeffery, R., Software Project Effort Estimation: Foundations and Best Practice Guidelines for Successs, Springer, 2014.

Wolfgang, P., Design Patterns for Object-oriented Software Development. Reading, Mass.: Addison-Wesley, 1994.

Verlaine, B., Jureta, I. J., Faulkner, S., A Requirements-Based Model for Effort Estimation in Service-Oriented Systems. ISOC Workshops, LNCS, pp. 82-94, 2014.

Wittig, G.E. and Finnie, G.R., Using Artificial Neural Networks and Function Points to Estimate 4GL Software Development Effort. Australasian Journal of Information Systems, Vol. 1(2), pp.87-94, 1994.

Chapter 5

5. Structuring Diverse Regulatory Requirements for Global Product Developments[1]

5.1 Introduction

Globalisation of system development not only offers great opportunities but also results in new challenges due to the diversity of requirements in different contexts (such as countries, organisations and situations), in particular dealing with regulatory requirements. Some obvious reasons for requirements diversity for a given system in different contexts are historical, cultural, natural, and economical. However, less visible reasons are due to diversity in applicable standards, regulations and laws in different jurisdictions. The challenges in the global context have some resemblance with those of product customisation and the development of product lines (Alves et al., 2010) as well as requirements variability across product lines (Buhne et al., 2005). However, the regulatory domain provides a specific kind of non-functional requirements that necessitates the analysis of diversity and variability not only at the level of requirements for concrete products, but also at the level of regulations. The requirements engineering (RE) task is challenging even in the case of a local (non-distributed) development of a system for deployment in a single country or organisation, where there is an acknowledged set of standards and regulations. Clearly, in the multi-context case of global development, the challenges are considerably more serious and, thus, there is a more dire need for systematic and scalable ways to deal with these challenges.

In this chapter, we present a novel framework for requirements structuring as well as traceability and change analysis in the global context. In this framework, we start the analysis at the level of regulations, with the aim to provide a formal basis for structured analysis of legal requirements for software and systems meant for multiple jurisdictions.

[1] This version was published in (Spichkova et al, 2015) where the candidate was a co-author.

We illustrate the proposed ideas through example data from an industrial RE project (Nekvi and Madhavji, 2015), where one of key activities was to ensure regulatory compliance of the project's requirements. Since this chapter does not present empirical findings from a case study, we only extracted examples from an actual regulatory document[2] and artificially created their variants for the purpose of illustration of the proposed framework.

This chapter is structured as follows: Section 5.2 provides the background literature; Section 5.3 describes the key feature of the framework for structuring regulatory requirements in global product development; Section 5.4 illustrates the framework using three examples; and Section 5.5 concludes this chapter.

5.2 Related Literature

There are proposed approaches that check requirements for compliance ((Breaux et al., 2008); (Maxwell and Anton, 2009)) or check the compliance of the outcomes of business processes against outcome focused regulations (Yin et al., 2013). We complement these prior efforts, with our focus on creating a framework for methodical structuring of regulatory requirements for geographically distributed system development. The proposed approach is anticipated to help analyse the relations among requirements and to trace changes to requirements in a global context.

From non-functional requirements (NFR) viewpoint, the paper (Glinz, 2007) surveyed the existing definitions of non-functional requirements (NFR), highlights and discusses the problems with the current definitions, and contributes concepts for overcoming these problems. There are also several methodologies for developing software systems from requirements, e.g., (Spichkova et al., 2012). In this chapter, we focus on regulatory, non-functional requirements (NFRs) taking into account human factors in requirement model-

[2] Federal Communications Commission (FCC) 47 CFR Part 15; American Railway Engineering and Maintenance-of-way Association (AREMA), Part 10.3.20; CGSB (Canadian General Standard Board)-GP-71

ling (Spichkova, 2012). A survey of efforts to support the analysis of legal texts in the context of software engineering is presented in (Otto and Anton, 2007). The authors discuss the role of law in requirements and identify key elements that a system could support in the analysis of regulatory texts for requirements specification, system design, and compliance monitoring. In (Nekvi et al., 2011), we identify key artefacts, relationships and challenges in the demonstration of compliance of systems requirements against engineering standards and government regulations. This is foundational in the creation of a compliance meta-model depicted in (Nekvi and Madhavji, 2015). In (Kiyavitskaya et al., 2008), the authors describe the problem of designing regulation-compliant systems and, in particular, the challenges in eliciting and managing legal requirements. The paper (Breaux et al., 2006) reports on an industry case study in which product requirements were specified to comply with the U.S. federal laws. In (Maxwell and Anton, 2009), the authors describe a case study on the evaluation of the iTrust Medical Records System requirements for compliance with the U.S. Health Insurance Portability and Accountability Act (HIPAA). The paper (Siena et al., 2009) presents guiding rules and a framework for deriving compliant-by-construction requirements, focusing on the U.S. federal laws. An approach to mapping legal and regulatory requirements for electronic health records onto concrete security controls is presented in (Breaux et al., 2013). While there are numerous interesting issues on compliance such as recognition of natural language specification, model creation from such specification, and automated analysis of regulatory statements, which are not in the scope of this chapter.

5.3 Regulatory Requirements Dependency in Global Developments

For global and remote development (Spichkova et al., 2013), we need to deal with diversity in regulations in different contexts since this affects the specifics of the requirements implemented in the deployed systems in different jurisdictions. Suppose a product P has to be developed for application in N countries $C_1 \dots C_N$ with the corresponding sets of regulations $RegulC_1, RegulC_{2, \dots}, RegulC_N$. These sets could have a (non-empty) joint subset $Regul$ of the regulations that are equal for all $C_1 \dots C_N$. We denote for each $RegulC_j$ the corresponding compliment to $Regul$ - i.e. the country-specific subset - by the notation

$RegulCj' = RegulC_j \backslash Regul$. For simplicity we avoid discussion on the side-effects of changing a particular regulatory requirement.

We denote the set of requirements to the product *P* valid for the country C_j by R^{Cj}, where *R* denotes the set of requirements to the product *P*. The complete set *R* of requirements to the product *P* is then defined as a union of the sets R^{Cj}, $j \in \{1... N\}$. The sets of requirements might be different for each country, i.e. R^{Ci} is not necessary equal to R^{Cj} for the case $i \neq j$; $i, j \in \{1... N\}$. We divide the set *R* of requirements to the product *P* into three (disjoint) subsets:

- The set *RFN* denotes the functional and non-functional requirements that are independent from the regulations. We assume that the requirements of this category are the same for any of $C_1... C_N$.
- The set *RLcommon* denotes the requirements based or depending on the regulations that are common across all countries of interest i.e., $C_1... C_N$.
- The sets $RLspecific^{Cj}$ denote the requirements based or depending on the regulations that are specific to a country.

By RL^{Cj} we denote the set of requirements for the product *P*, which are valid for the country C_j, i.e., the union of *RLcommon* and $RLspecific^{Cj}$. Figure 5-1 presents the requirements dependencies for the product *P* that should be developed for the countries $C_1... C_N$. Consider the set RL^{Cj}_{ALL} to be the set of all legal requirements of the country *Cj*; we can then see the set RL^{Cj}_{P} as a "*projection*" of RL^{Cj}_{ALL} on the product *P*.

On this basis, we can adapt and extend the definitions for product lines (Pohl et al., 2005) to specify the variability aspects of the regulations and concrete requirements. We define a *variation point* in regulations as a part of a regulation that can vary within several jurisdictions (i.e., within several countries, provinces, states, or organisations). We say that a regulation *x* is a *variant* of a regulation *y*, if these regulations are equal modulo (a number of) variation point(s).

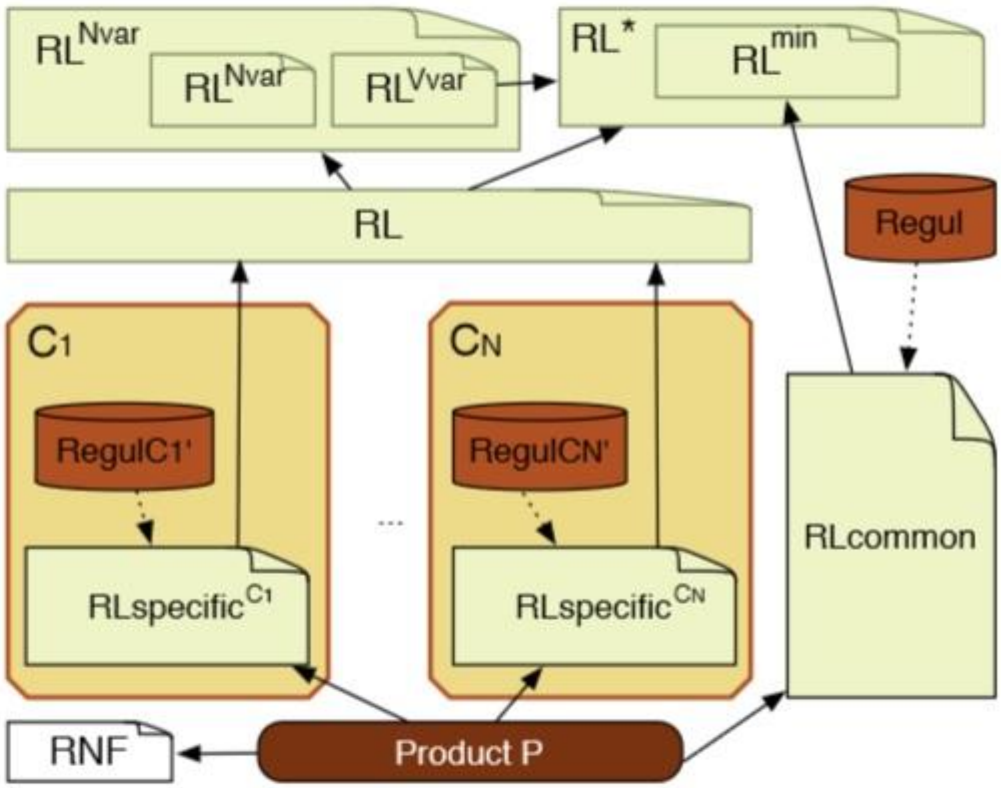


Figure 3 Requirements dependency for product *P*

As all requirements from the sets *RLcommon*, $RLspecific^{C}_{1}$, ..., $RLspecific^{C}_{N}$ (let us call them *contract requirements*) are based or depending on the regulations, we suggest for better traceability to specify them following one of the predefined formats, e.g., *"X shall be compliant with the limits set in the applicable regulations.* $[rg^{C}_{1} ... rg^{C}_{N}]$*"*, where *X* denotes component or its feature, the list $[rg^{C}_{1} ... rg^{C}_{N}]$ represents the variation point, and each rg^{C}_{i} denotes the name of the corresponding standard of the country C_i with reference to concrete regulations (regulatory requirements). Thus, regulatory requirements become parameters of the contract requirements. While instantiating the parameter with a concrete regulation, we obtain the corresponding contract requirement.

Example 1: Let us assume the case when a product is built for three different countries C_1, C_2, and C_3. Each of these countries has their own regulations describing allowed limit for emission energy produced by Communication Backbone Network (CBN) equipments used by the product. The decision of frequency bands for used by Short Range Devices (SRD) is a national matter, and the assignment of frequency spectrum is not consistent worldwide.

Not only the spectrum assignment but also the operational limitations (e.g., emission energy) of the emitting devices (e.g., LAN, alarms, and cell phones) can be diverse across countries or regions. This diversity reflects different degrees of tolerance to risk associated with health hazard caused by radio emission. For example, laws in some countries are relatively more lenient with respect to restricting the use of electrical devices, so as to stimulate economic growth; this is in contrast to the precautionary approach to health safety (Mazar, 2009).

An example of a *contractual* requirement is:

req1: *The CBN equipment (such as 'Ethernet network') shall be compliant with emitted field strength limits set forth in the applicable regulations* [$rg1^{C1}$, $rg1^{C2}$, $rg1^{C3}$].

The corresponding concrete requirement for each country C_i would be $req1[rg1^{Ci}]$, where $rg1^{Ci}$ denotes the corresponding regulatory requirement, e.g., let us consider $rg1^{C1}$ as *Federal Communications Commission Standards, FCC 47 CFR Part 15*, where $rg1^{C2}$ and $rg1^{C3}$ will be artificially created for illustration purposes (cf. Table 5-1). Thus, $rg1^{Ci} \in Regul^{Ci}$; for $i \in \{1,2,3\}$, which also means that $req1[rg1^{Ci}]$ should belong to the set $RLspecific^{\mathrm{Ci}}$.

Table 5-1 Regulations applicable to *req1* and *req2*

$rg1^{C1}$	The emissions from an intentional radiator operating over 960 MHz shall not exceed the field strength 500 µV/meter when measuring at 3 meter distant
$rg1^{C2}$	The emissions from an intentional radiator operating between 2GHz-3 GHz shall not exceed the field strength 500 µV/meter when measuring at 3 meter distant
$rg1^{C3}$	The emissions from an intentional radiator operating between 2400-2483.5 Hz shall have power range 10 to 1000 microwatt in EIRP measurement
$rg2^{C1}$ $rg2^{C2}$ $rg2^{C3}$	The primer shall dry tack-free in not more than 2h and shall dry hard in not more than 10h. The primer film shall withstand against scratch test under a load of 900 g; and the primer shall show no cracking when subjected to the bending test at 23◦ using 6.4 mm diameter mandrel.

Example 2: Let us continue the *Example 1* by discussing the following contract requirement:

req2: Inside and outside surfaces of switchboards shall be painted with a high quality metal primer coat conforming to applicable standards or regulations [$rg2^{C1}$, $rg2^{C2}$, $rg2^{C3}$].

Let $rg2^{C1}$ is defined to be the standard *Canadian General Standard Board GP-71*, where $rg2^{C2}$ and $rg2^{C3}$ are the same as $rg2^{C1}$ (cf. Table 5-1), i.e., $rg2^{C1}, rg2^{C2}, rg2^{C3} \in Regul$.

5.4 Framework in Use

In this section, we first define the various concepts concerning the inter-relationships amongst the requirements and regulations from various countries, which are ultimately used in the proposed framework for structuring them.

First, it is important to identify the relations amongst the subsets $Regul^{C1}... Regul^{CN}$ as well as amongst their corresponding requirements subsets of $RL^{C1}... RL^{CN}$ since it helps to trace the *Regulatory* requirements' changes more efficiently.

Further, some requirements in $RL^{Cj;}$ for j= {1...N} can be stronger versions of other requirements from this set. For example, if $req1 \in RL^{Ci}$ and $req2 \in RL^{Ci}$ are not equal but *req1* is a refinement of *req2*, we call *req2* is a weaker version of *req1*. The same principle applies for regulations.

Thus, we need to have a structured specification of legal requirements on both level, i.e., contract requirements and regulation. In this regard, we build phenomenon such as *strongest set* and *requirements variants* to be implemented as features of the requirements structuring framework.

We define the following notations for them:

- *Regul* $*$ denotes the strongest set of legal requirements for the product *P*.
- $Regul^{var}$ denotes the set of variants of the requirements. It has two disjoint subsets:
  - $Regul^{Nvar}$ denotes the set of variants, which cannot be compared in the sense that one of the variants is stronger than the other(s). This could be the case when the two regulations differs in the description value (e.g., material) that should be used for the products.
  - $Regul^{Vvar}$ denotes the set of variants, which can be compared (variants are numeric) to identify the strongest version.

Then we construct the sets $RL*$, RL^{var}, and RL^{min}. Here, RL^{min} denotes the set of contract requirements that should be fulfilled within all countries C_1... C_N. It is defined on the basis of the set *RLcommon* (i.e. to be contained in *Regul*), s.t. we do not analyse the sets $RLspecific^{Cj}$, for $1 \leq j \leq N$ to construct the set RL^{min}. While identifying RL^{min}, we will analyse which components of a target product can be built once for one country and then can be reused for the entire product family across all countries of interest such as $P^{C1},...,P^{CN}$. This allows us to have more efficient process for the global software and systems development. It also helps in efficient tracing of requirements changes that might come from the changes in the regulations. For example, changes in RegulC1′ imply changes in $RLspecific^{C1}$ only since only the *C1*-specific part is affected, where any changes in *Regul* might influence global changes. While identifying $RL*$ we will obtain the global view on the products’ requirements, which is not overloaded with the variants of the similar requirements but are just weaker versions of other.

Example 3: Let $rg3^{C1}$ is defined to be the standard *American Railway Engineering and Maintenance-of-way Association (AREMA), Part 10.3.20*, where $rg3^{C2}$ and $rg3^{C3}$ are artificially created on its basis (cf. Table 5-2).

Table 5-2 Regulations applicable to req3

$rg3^{C1}$	Jacketing shall be a durable properly vulcanised black Neoprene; with average thickness of not less than 15 mils, and with min. thickness at any point shall not be 90% of average thickness
$rg3^{C2}$	Jacketing shall be a durable properly vulcanised Chlorinated Polyethylene; with average thickness of not less than 20 mils, and with minimum thickness at any point shall not be 90% of average thickness.
$rg3^{C3}$	Jacketing shall be a durable properly vulcanised Chlorosulfonated Polyethylene; with average thickness of not less than 15 mils, and with minimum thickness at any point shall not be 85% of average thickness.

In case of Example 1, $rg1^{C1}$, $rg1^{C2}$, and $rg1^{C3}$ were already elementary, because all the characteristics were either ordinal rather than nominal (i.e., frequency band) or expressed in diverse means (i.e., emission strength limit expressed interleaving manner among electric field strength and EIRP) which are not comparable.

However for Example 3, we need to simplify the regulations first. In this case, each regulatory requirement consists of three elementary parts (we denote them by necessary indexes - see Table 5-3). Without necessary simplification, we obtain that $rg3^{Ci} \in RegulCi'$ for each i $\in\{1,2,3\}$,. This also implies that the corresponding contract requirement *req3* [$rg3^{Ci}$] would belong to the set $RLspecific^{Ci}$. One might have considered them as variants of each other but this would provide an insufficient basis for further analysis. However, simplification of $rg3^{Ci}$ allows piecing them into a set of variants of each other such as $rg3^{Ci}$j, if j $\in$ number of variants. We cannot compare them to assert that some of them are weaker or stronger than other. Thus, these regulatory requirements should be added to the set $Regul^{Nvar}$.

Table 5-3 Elementary Regulatory Requirements from rg3

$rg3^{C1}1$	Jacketing shall be a durable properly vulcanised black Neoprene
$rg3^{C1}2$	The average jacketing thickness should be not less than 15 mils
$rg3^{C1}3$	The minimum jacketing thickness at any point shall not be 90% of average thickness
$rg3^{C2}1$	Jacketing shall be a durable properly vulcanised Chlorinated Polyethylene
$rg3^{C2}2$	The average jacketing thickness should be not less than 15 mils
$rg3^{C2}3$	The minimum jacketing thickness at any point shall not be 90% of average thickness
$rg3^{C3}1$	Jacketing shall be a durable properly vulcanised Chlorosulfonated Polyethylene
$rg3^{C3}2$	The average jacketing thickness should be not less than 15 mils
$rg3^{C3}3$	The minimum jacketing thickness at any point shall not be 85% of average thickness

The regulations $rg3^{Ci}2$, for i ∈ {1, 2, 3} are equal (see the 2nd, the 5th and the 8th row of column two in Table 5-3), which means that they should be added to the set $Regul^{min}$ and therefore also to the set $Regul*$.

However the regulations $rg3^{Ci}1$, for i ∈ {1, 2, 3}, they are variants of each other (see the 1st, the 4th and the 7th row of column two in Table 5-3). Since the variants (e.g., black Neoprene, Chlorosulfonated Polyethylene, and Chlorinated Polyethylene) are descriptive rather than being numeric, they will be placed to the set $Regul^{Nvar}$. But the variants (e.g., 90% average thickness, 85% average thickness) for $rg3^{Ci}3$, for i ∈ {1, 2, 3} are numeric (see the 3rd, the 6th and the 9th row of column two in Table 5-3), they will be placed to the set $Regul^{Vvar}$. The strongest version will be both $rg3^{C1}3$ and $rg3^{C2}3$ than $rg3^{C3}3$ (90% > 85%) and should be added to the set $RL*$.

5.5 Conclusions

In this chapter, we presented our ongoing work on requirements specification and analysis in a global context. Requirements for a system meant for use in multiple jurisdictions (e.g., different countries, organisations, or situations) can differ according to the needs of the particular environments. Organising the overlapping sets of requirements across the diverse environments, including variant and change management, can be challenging. This is compounded when we consider diverse regulations, in different jurisdictions, with which deployed system must comply in specific environments. We have developed a framework to deal with this diversity in a systematic manner, avoiding contradictions and non-compliance. In this framework we start the analysis at the level of regulations. This provides a basis for structured analysis of legal requirements for the system to be built for multiple jurisdictions. The chapter describes, in formal terms, the framework and an illustrative example of the use of the framework.

Organising regulations and requirements into logical structures is anticipated to be relatively straightforward in terms of effort, cost and skills required for small-to-medium sized systems in a given locality. However, when dealing with large, complex systems to be deployed in diverse jurisdictions where regulatory compliance is critical, organisation of regulations along with systems regulatory requirements and change management can be challenging. This chapter shows a step in the direction towards a comprehensive solution to tackle this problem. Our future work includes further investigation of the framework and actual variant requirements from distributed locations or contexts and to provide tool support for the analysis with the framework.

5.6 References

Alves, V. N., Alves, N. C., and Valenca, G., Requirements Engineering for Software Product Lines: A Systematic Literature Review.Information and Software Technology, Vol. 52(8), pp. 806 – 820, 2010.

Breaux, T., Anton, A., Boucher, K., and Dorfman, M., Legal Requirements, Compliance and Practice: An Industry Case Study in Accessibility. International Conf. on Requirements Engineering, pp. 43–52, 2008.

Breaux, T., Gordon, D., Papanikolaou, N., and Pearson, S.,Mapping Legal Requirements to IT Controls. 6th international workshop on Requirements Engineering and Law, pp. 11–20, 2013.

Breaux, T., Vail, M., and Anton, A., Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations. International Conf. on Requirements Engineering, pp. 49–58, 2006.

Buhne,S., Lauenroth, K., and Pohl, K., Modelling Requirements Variability Across Product Lines. International Conf. on Requirements Engineering, pp. 41–50, 2005.

Glinz, M., On Non-Functional Requirements. International Conf. on Requirements Engineering, IEEE Computer Society, pp. 21–26, 2007.

Kiyavitskaya, N., Krausova, A., and Zannone, N., Why Eliciting and Managing Legal Requirements Is Hard.1st Workshop on Requirements Engineering and Law, pp. 26–30, 2008.

Maxwell, J., and Anton, A., Checking Existing Requirements for Compliance with Law Using a Production Rule Model.2nd Workshop on Requirements Engineering and Law, pp. 1–6, 2009.

Mazar,H., A Global Survey and Comparison of Different Regulatory Approaches to Non-ionizing Radhaz and Spurious Emissions.IEEE international conference on COMCAS, pp. 1–6, 2009.

Nekvi, M. R. I., and Madhavji, N. H., Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects: A Case Study. ACM Trans. on Management Inf. Syst., Vol. 5, no. 3, pp. 15:1–15:35, 2015.

Nekvi,M.R.I., Ferrari, R., Berenbach, B., and Madhavji, N., Towards a Compliance Meta-model for System Requirements in Contractual Projects.4th Workshop on Requirements Engineering and Law, pp. 74–77, 2011.

Otto,P., and Anton, A., Addressing Legal Requirements in Requirements Engineering. International Conf. on Requirements Engineering, pp. 5– 14, 2007.

Pohl, K., B¨ockle, G., and Linden, F. J. V. D., Software Product Line Engineering: Foundations, Principles and Techniques. Springer, 2005.

Siena, A., Perini, A.,Susi, A., and Mylopoulos, J., A Meta-model for Modelling Law-Compliant Requirements.2ndWorkshop on Requirements Engineering and Law, pp. 45–51, 2009.

Spichkova, M., Human Factors of Formal Methods. IADIS Interfaces and Human Computer Interaction, IADIS Press, pp. 307–310, 2012.

Spichkova,M., H¨olzl, F., and Trachtenherz, D., Verified System Development with the Autofocus Tool Chain. Workshop on Formal Methods in the Development of Software, EPTCS, pp. 17–24, 2012.

Spichkova,M., Schmidt, H., and Peake, I., From Abstract Modelling to Remote Cyber-physical Integration or Interoperability Testing. Improving systems and software engineering conference, 2013.

Spichkova, M., Schmidt, H.W., Nekvi, M.R.I. and Madhavji, N.H., Structuring Diverse Regulatory Requirements for Global Product Development.IEEE International Workshop on Requirements Engineering and Law (RELAW), pp. 57-60, 2015.

Yin, Q., Madhavji, N., and Pattani, M., Eros: An Approach for Ensuring Regulatory Compliance of Process Outcomes.7th Workshop on Requirements Engineering and Law, IEEE Computer Society, pp. 21–24, 2013.

Chapter 6

6. Emerging Theory[1]

6.1 Introduction

Existing literature on regulatory compliance in RE focuses mainly on developing technological solutions (e.g., (i) modelling and analysis of regulatory text (Kerrigan and Law, 2003; Antoniou et al., 1999), (ii) techniques and framework (Islam et al., 2010; Breaux et al., 2006) to elicit regulatory requirements from applicable regulations and standards for compliance, and (iii) techniques (Ramezani et al. 2012; Ingolfo et al., 2011; Saeki and Kaiya, 2008) to validate requirements for regulatory compliance. However, there is not much "empirically grounded theory" on the complexities arising from the characteristics of a large set of artefacts and their inter-relationships that exist in compliance work in RE of contract-based large systems engineering projects. For example, questions such as the following do not have responses grounded in empirical theory today:

Q1: What are the types of inter-relationships among RE artefacts existing in compliance work in RE?

Q2: Which characteristics of RE artefacts and their inter-relationships complicate the tasks of ensuring regulatory compliance of requirements?

Q3: What are the impediments to ascertaining regulatory compliance of requirements?

Q4: What are the effort-critical activities and artefacts in compliance work in RE?

Q5: How do industries measure the complexities associated with effort-critical activities and artefacts involved in compliance work in RE of contractual systems engineering projects?

---

[1] This chapter is not published as an independent research paper but is written as part of thesis composition.

The main goal of this dissertation was to investigate the characteristics of compliance work in RE of large contractual systems engineering projects. Therefore, we performed multiple industrial-scale studies on identifying and characterising: (i) the artefacts used in compliance and their inter-relationships (see Chapter 2), and (ii) the impediments existing in compliance projects (see Chapter 3). Consequently, based upon the study results (e.g., artefacts types and inter-relationships, and impediments) we derive a number of RE metrics towards creating an effort estimation model for compliance work of RE (see Chapter 4) and develop a preliminary framework for structuring regulatory requirements in global product development settings (see Chapter 5).

Based on the findings of the described studies, we propose an emerging descriptive theory that inter-relates characteristics of numerous artefacts, impediments to ensuring regulatory compliance of requirements, and emerging metrics towards an effort estimation model for compliance work in RE of large contractual systems engineering projects. The theory and its associated propositions are inferred directly from observational data following the hypothetico-inductive model[2] (Sjoberg et al., 2008). The emerging theory and propositions are described in Section 6.2. The propositions are evaluated using the criteria for measuring the goodness of a theory in Software Engineering (Sjøberg et al., 2008) in Section 6.3. The implications of this theory are provided in Section 6.4. Lastly, in Section 6.5, we conclude this chapter.

6.2 The Emerging Theory

In this section, we first describe background information about how the theory is formulated. Then, we explain the theoretical propositions as they are derived from observations from the studies.

Sjøberg et al., (2008) describes three levels of sophistications for theory propositions, i.e.

[2] It is a bottom-up approach of building theories that starts with specific observations or real examples and progresses analytically to broader generalizations, and theories based on those observed cases.

i. Level 1 propositions are minor working relationships that are concrete and based directly on observations,

ii. Level 2 are theories of the middle range that involve abstraction of possibly many Level 1 theories but are still closely linked to observations, and

iii. Level 3, all-embracing theories that seek to explain an aspect of Software Engineering.

The preliminary theory propositions are given in Table 6-1. The different levels of abstractions denoted in the theory propositions are structured hierarchically in the table. For example, Level 1 propositions are derived from direct empirical findings of the thesis (described in Chapter 2, Chapter 3 and Chapter 4) or direct observations (described in Chapter 5). The propositions of Level 2 represent higher level abstractions formulated on the more concrete propositions denoted in Level 1. Since Level 1 propositions are based on concrete evidence, they are directly testable (i.e., such evidence can be checked in other projects). They also formed the basis for testing the propositions at Level 2. With reference to Table 6-1, we formulate 12 propositions at level 1 and two level 2 propositions. However, we do not have any theory at the level 3 since they are typically comprised of meta-analysis of related findings from multiple studies as the field becomes mature (Sjøberg et al., 2008).

In the beginning of Table 6-1 are labels of the different studies (i.e., [S1], [S2], [S3], and [S4]) which form the context of the propositions stated in this thesis. Then, Level 1 propositions are stated with reference to the studies of the thesis from where they are derived. From these level 1 propositions, more generalised propositions are provided, which are listed under level 2 denotations. Both level 1 and level 2 propositions are labeled using unique identifiers (described in the table) to reference them from the textual descriptions of the theories.

Table 6-1 Emerging Theory

[S1] (Chapter 2): Empirically Derived Compliance Meta-Model for System Requirements
[S2] (Chapter 3): Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects - A Case Study
[S3] (Chapter 4): Metrics for Estimating the Effort Needed in Requirements Compliance Work
[S4] (Chapter 5): Structuring Diverse Regulatory Requirements for Global Product Developments

P.L.I represents a unique proposition identifier, where L = level number, and I= Proposition ID at that level
P.I represents a unique proposition identifier at level 2, where I= Proposition ID

Level 1	Level 2
P.1.1. Numerous voluminous requirements artefacts such as contract, a set of standards, a set of regulations, various levels of requirements documents (e.g., sub-system level, component level, and cross-cutting) exist for regulatory analysis of requirements in large scale contractual systems engineering projects [S1, S2]	P1. Characteristics of requirements artefacts and their inter-relationships existing in contractual systems engineering projects underlie substantial impediments to doing compliance work in RE
P.1.2. Large number and voluminous contents of regulatory documents relevant to target system complicate the RE task of: (i) managing requirements changes with the change of regulations and standards; and (ii) visualisation of project status. [S2]	
P.1.3. Diversity of technical domains (e.g., electrical, IT, power system, and mechanical) associated with relevant regulatory documents impedes the RE task of identifying which parts are applicable for target system and where in the system do they apply. [S2]	
P.1.4. Inter and intra cross-references of the regulatory documents are so many (e.g., 10 cross-references per page in one sample standard *CGSB 1-GP-81*) that they complicate the elicitation of requirements from their associated textual description. [S2]	
P.1.5. Non-contiguity of regulatory requirements in the contract complicates and impedes the RE task of identifying them and deriving project requirements from contract, creating requirements traces, and monitoring progress of compliance work in RE. [S2]	
P.1.6. Diverse regulatory references present in so many of the contractual requirements (e.g., 34% of the requirements in Switch Clearing Sub-System) that deriving project requirements from the complementary and often contradictory referenced documents becomes a cautious and time consuming RE task entailing negotiation, review, and analysis. [S2]	
P.1.7. Abstractness of regulatory requirements specified in contract are so many in number (i.e., 50% of the regulatory documents are referenced without proper citation) complicates the RE task of deriving concrete project requirements. [S2]	
P.1.8. Significant portion of regulatory documents (62% of the segments of AREMA standard) containing implicit requirements extensively overloads the RE agents job to elicit appropriate re-	

quirements for system. [S2]	
P.1.9. Cross-cutting requirements comprising of significant percentage of regulatory requirements (i.e., 40%) indicates a great deal of communication overhead among sub-teams working for various sub-systems and components. [S2]	
P.2.1. Size metrics, and process and product complexity related metrics for an effort estimation model for compliance work in RE are derivable from impediments associated with contractual complexities [S3, S4]	P2. Key metrics for an effort estimation model for compliance work of RE are derivable through analysis of impediments associated with compliance work of RE
P.2.2. Size metrics, and process and product complexity related metrics for an effort estimation model for compliance work in RE are derivable from impediments associated with complexities in regulatory documents [S3, S4]	
P.2.3. Size metrics, and process and product complexity related metrics for an effort estimation model for compliance work in RE are derivable from impediments associated with large and complex systems [S3, S4]	

Below, we describe the level 1 propositions (see the 1st column of Table 6-1) on which each of the two level 2 emerging theories (i.e., P1 and P2 - see the 2nd column of Table 6-1) are derived.

P1. ***Characteristics of requirements artefacts and their inter-relationships existing in contractual systems engineering projects underlie substantial impediments to doing compliance work in RE.***

In study [S1], the set of artefacts that were identified to be used in compliance work in RE included a contract document comprising customer requirements, set of regulations and standards applicable to target system, project requirements specifications classified into sub-system specifications, components specifications, cross-cutting specifications, etc., and implicit project requirements specifications comprising requirements imposed by external sources such as regulations and standards (Section 2.4.1 of Chapter 2). Therefore we imply proposition P.1.1. We also identified how these artefacts are inter-linked to each other by compliance-oriented inter-relationships, such as: contract references to regulations, regulations cross-references to themselves, regulations impose new requirements, contract derives project requirements, project requirements complies with regulations (Section 2.4.2 of Chapter 2). Such inter-relationships are testament for their significance in compliance work in RE.

In the study [S2], we found that there were approx. 300 standards and regulations identified as relevant for the system (Section 3.6.2.1 of Chapter 3). Managing changes of requirements due to change of regulatory documents of such substantial number indicate higher degree of impediments. Not only the number of the documents was substantial but also sizes of some of the documents were so massive (i.e., ranging from over 100 pages to thousand pages long) to for RE agents to elicit requirements from without losing stamina (Figure 3-1 of Chapter 3). We thus imply proposition P.1.2.

We found from study [S2] that the regulatory documents had come from various technical domains such as IT, electrical, mechanical, and chemical in the system engineering project. Reviewing one of such documents for determining its applicability to system had been stalled when more than one domain experts felt unfamiliar with the technical contents of the document (Section 3.6.2.2 of Chapter 3). This implies proposition P.1.3.

The regulatory documents were abundant with inter and intra cross-references as discovered by our study [S2]. Analysis to the standard CGSB 1-GP-81 showed that it had an average of 10 cross-references per page (Section 3.6.2.3 of Chapter 3). Among them were 12 external documents (e.g., CGSB 1-GP-70M, ASTM D1210, ASTM D2621) that must also be analysed (Section 3.6.2.3 of Chapter 3), thus implying another impediments - as asserted by proposition P.1.4.

With reference to study [S2], regulatory requirements in the contract are specified non-contiguously and mixed up with non-regulatory requirements with 1:11 ratio in the studied 75 pages chunk (Fig 3-2 in Chapter 3) of the contract. Proposition P.1.5 is asserted because such non-contiguity complicates project tasks, e.g., identifying regulatory requirements, deriving project requirements from them, creating traces for the requirements, etc (Section 3.6.3.1 of Chapter 3).

In study [S2], we observed from analysis of regulatory requirements of one subsystem (i.e., Switch Clearing Sub-System) that 34% (10 out of 29) of them contained diverse regulatory references (Section 3.6.3.2 of Chapter 3). Not only that but also a total 19 different regulatory documents were specified from the 29 requirements of this sub-system. Since diverse regulatory documents can be in conflict with each other (Kiyavitskaya et al., 2008; and Maxwell et al., 2012), this must be resolved through negotiation, reviews, analysis and other techniques. Thus we assert proposition P.1.6.

In study [S2], we found that regulatory documents in the contract were specified in a variety of abstract forms, e.g., without proper reference to specific part of the regulatory documents and without mentioning the specific part of system that has to comply against cited documents. Approx. 50% of the regulatory documents were mentioned as a whole, which means that they have to be analysed cover to cover for eliciting requirements (Section 3.6.3.3 of Chapter 3). Therefore, proposition P.1.7 is asserted.

Proposition P.1.8 is implied from the fact that approx. 62% of all segments of the standard AREMA contained implicit requirements for the system as we found in study [S2] (Table 3-3 of Chapter 3). Considering 300-odd regulatory documents, and their sizes (shown in Figure 3-1), identifying and eliciting those implicit requirements becomes a monumental task (Section 3.6.3.4 of Chapter 3).

In study [S2], we found that cross-cutting requirements comprised of 40% of the regulatory requirements (Table 3-4 and Section 3.6.4.3 of Chapter 3). The degree of communication overhead among sub-teams working for various sub-systems and components involved in analysing these requirements imply proposition P.1.9.

P2. ***Key metrics for an effort estimation model for compliance work of RE are derivable through analysis of impediments associated with compliance work of RE.***

In study [S2], we identified impediments associated with contractual complexities were non-contiguity of regulatory requirements in the contract, diverse regulatory references from contractual requirements, abstractness of contractual regulatory requirements, and implicit regulatory requirements with respect to contractual specifications (Section 3.6.3

of Chapter 3). In Section 4.3 of Chapter 4 [S3], we defined a three-step method for deriving metrics associated with effort critical activities and artefacts (essentially impediments to compliance work in RE). Table 4-4 of Chapter 4 shows effort estimation metrics associated with contractual complexities (e.g., "Avg. ratio of regulatory and non-regulatory requirements per page" for the characteristics 'non-contiguity of regulatory requirements') that were derived using that method. In Section 4.5, use of these derived metrics into an effort estimation model (adopting algorithmic modelling technique) is explained and the correspondence between the metrics and model parameters, especially the parameters of 'size' and process and product characteristics, for each of the effort critical activities pertaining to contractual complexity is shown in Table 4-7. This implies to proposition P.2.1.

In study [S2], we also identified impediments associated with complexities in regulatory documents that include: large number and size of regulatory documents, numerous cross-references, and multi-domain contents (Section 3.6.2 of Chapter 3). In study [S3], a three-step method for deriving metrics associated with effort critical activities and artefacts (essentially impediments to compliance work in RE) is defined and explained in Section 4.3 of Chapter 4. Table 4-5 of Chapter 4 shows effort estimation metrics associated with complexities in regulatory documents (e.g., "Avg. # of cross-references per section" for the characteristics 'cross-references in regulatory documents') that were derived using that method. In Section 4.5, use of these derived metrics into an effort estimation model is explained and the correspondence between the metrics and model parameters, especially the parameters of 'size' and process and product characteristics, for each of the effort critical activities pertaining to complexities in regulatory documents is shown in Table 4-7. This implies to proposition P.2.2.

Similarly in study [S2], impediments associated with the aspect of large and complex system are identified such as vertical and lateral communications and cross-cutting requirements (Section 3.6.4 of Chapter 3). In Study [S3], using the three-step method (Section 4.3 of Chapter 4), we derived metrics associated with system complexity, which are shown in Table 4-6. In Section 4.5, use of these derived metrics into an effort estimation model is explained and the correspondence between the metrics and model parameters, especially the parameters of 'size' and process and product characteristics, for each of the effort critical activities pertaining to system complexity is shown in Table 4-7. This implies to proposition P.2.3.

6.3 Evaluation of the Emerging Theory

For evaluation the goodness of a theory, we need to establish criteria by which it is to be evaluated. Sjøberg et al. (2008) also describes some criteria for this, which are similar to the criteria listed by (Boehm and Jain, 2006). These criteria are adapted for empirically-based theory in SE, which were traditionally being used in other disciplines such as Business Management (Bacharach, 1989), Psychology (Haig, 2005), and Sociology (Cohen, 1989). The criteria provided by (Sjøberg et al., 2008) are one of the most up-to-date for evaluating theories in software engineering. Ralph, (2015) also provides guidance for empirically evaluating "process" theories (i.e., theories that describes or predict changes of entities such as artefacts, tests, teams, and organisations) in software engineering. However, the theories we describe (see Section 6.2) are not process theories describing changes; they describe condition and characteristics of products and process. Therefore, we also adopt the criteria provided in (Sjøberg et al., 2008) for its applicability to our empirical research in the field of RE. Several other researchers from SE used the criteria by (Sjøberg et al., 2008) for evaluating the empirically-based theory, e.g., Ferrari (2010) in the area of interaction between RE and software architecture (SA) and Murtaza (2011) in software testing.

We first list the criteria below and then we describe the degree of support (i.e., low, medium, or high) along each criterion for the emerging theory.

1. Empirical support - The degree to which a theory is supported by empirical studies that confirm its validity.

2. Utility - The degree to which a theory supports the relevant areas of the software industry.

3. Generality - The breadth of the scope of a theory and the degree to which the theory is independent of specific settings.

4. Parsimony - The degree to which a theory is economically constructed with a minimum of concepts and propositions.

5. Testability - The degree to which a theory can be empirically refuted.

6. Explanatory power - The degree to which a theory accounts for and predicts all known observations within its scope, is simple in that it has few ad hoc assumptions, and relates to that which is already well understood.

Below, we describe the goodness of the emerging theory based on the criteria given above. The evaluation is based on our subjective judgment that is expressed in nominal ranks ranging from "high" to "low" to represent the two opposite extreme evaluation where the rank "moderate" lies somewhere in the middle between the extremes ("high" and "low").

- **Empirical Support:**

  Empirical support of the propositions asserted at level 1 in Table 6-1 are moderate since their basic findings (e.g., characterisations of artefacts and inter-relatioships, effort critical activities, emerging metrics, and impediments to compliance work in RE) are directly grounded on empirical evidence from the studies [S1],[S2], and [S4]. Still, the effects of such characterisations of artefacts and inter-relationships on RE compliance work (in the form of impediments) are inferred mostly on the basis of logical argumentations, observations, or communications with project staffs. Yet another limitation to the propositions is that they are supported by only one case (i.e., project), thus more empirical evidence from similar studies on other projects are required to strengthen the propositions. On the other hand, the propositions at level 2 in Table 6-1 are not directly derived from empirical findings but are abstractions of the lower-level propositions. Still they have secondary empirical support since they are actually generalised from empirically-grounded level 1 propositions and are not just conjecture-based deduction. The empirical support for these propositions is thus low to moderate.

- **Utility**:

  The emerging theory would be useful in RE for software and systems engineering projects. The propositions concerning impediments (e.g., P.1.2 to P.1.9) can be used in the RE decision making by higher level managements regarding the process, roles, and tools to use depending on the types of artefacts existing in the projects. For example, the theory indirectly suggests using domain experts and RE tools capable of managing large documents (i.e., storing large, voluminous documents and their versions, enabling tracing among their contents, and maintain changes) in order to reduce complexity of compliance work in RE. Further, the propositions on emerging metrics (e.g., P.2.1-3) would aid management to better estimate the effort required in compliance-centric work that is currently predominantly underestimated. In Section 6.4, we describe several uses of the theory both in industrial practice and academic research. The utility of the emerging theory is considered high.

- **Generality**:

  The empirical evidence of the theory is generalised from analysis of project artefacts such as set of regulations, standards, contract, and project requirements at various levels (e.g., sub-system, component, and cross-cutting). Whereas the contract and project requirements are specific to a particular project but the standards and regulations are originated either nationally (e.g., Canadian Electrical Code (CEC), Canada Standards Association (CSA), and American Railway Standards (AREMA)) or internationally (e.g., IEEE standards). Thus the characteristics of these regulatory documents and their impact on other projects aiming to develop or upgrade similar systems would be generic irrespective of systems and project settings. Since the empirical evidence pertaining to aspects of contract document and complex organisation of systems (i.e., organised into various sub-systems and components) are derived from real world "industrial" settings as opposed to "lab" settings, this increases the generality of the findings. However, the evidence came from studies of a single project (i.e., upgrading a railway system) that has unique system requirements particular to features of railway and its infrastructure, which limits the generalisability of the findings. Therefore, the generality of the empirical findings and the derived theory based on the findings are considered to be moderate.

- **Testability**:

  Each proposition of the emerging theory is expressed in non-ambiguous and consistent manner so that other studies in similar settings can be reliably designed and performed to confirm or refute the stated propositions. Moreover, one can derive specific hypothesis from the propositions. One example hypothesis from propositions "P.1.5" could be derived as "H1: *non-contiguity of regulatory requirements in the contract impede the task of deriving of project requirements from contract*". The dependency between the propositions at the same level is nearly zero so they can be tested independently to each other. For example, the two propositions concerning 'diverse regulatory references from requirements' (P.1.6) and 'abstractness of requirements' (P.1.7) can be tested of their own without requiring the result from others because they describe distinct phenomenon of the requirements. Further industrial case studies in other systems engineering projects are likely to be the most appropriate study design to validate the propositions. Therefore the testability of the theory is regarded as high.

- **Explanatory Power:**

  Human and ogranisational factors play a great role in SE, which are often difficult to explain limiting the explanatory power of SE theories. With contrast to this, our propositions are based on findings of the aspects of artefacts characteristics that are essentially non-human and non-organisational. Still, humans are involved in the process of dealing with the characteristics existing in the project, so as the organisational practices on the way customer requirements are being drafted in a contract. This theory explains the characteristics of artefacts used in compliance work in RE and how such characteristics translated into a variety of challenges to RE activities. It also provides a ground for deriving key metrics towards developing an effort estimation model for compliance work in RE. However, it does not explain how exactly an effort estimation model would be developed using the metrics or how the impediments affect the project's process in terms of time, cost and quality. So, we think that further complementary studies can expand the explanatory power of the theory. Explanatory power of this theory is thus considered as low to moderate.

- **Parsimony**:

  We have attempted to use constructs (such as metrics, impediments (challenges), and effort) for the emerging theory as clearly and precisely as possible. So the constructs used in the theory are fairly understood in the field of RE. Thus we think that parsimony is high for the theory.

We provide a summary of evaluation of the theory in Table 6-2 where the 1st column lists the criteria and last column shows their corresponding assessments. It is evident from Table 6-2 that the evaluation of the theory is satisfactory on some criteria (i.e., utility, testability, and parsimony). However, there is opportunity to improve the assessment on generality, explanatory power and empirical power, which can be achieved through further investigation on similar projects.

Table 6-2 Evaluation of the theory

Criteria	Assessment
Empirical power	Moderate
Utility	High
Generality	Moderate
Testability	High
Explanatory power	Low to moderate
Parsimony	High

6.4 Implications

The emerging theory has several implications both in industrial practice and academic research. We briefly explain them as follows:

- **Industrial Practice:**

  - Understanding the types of impediments would help higher managements of RE projects for better preparing in terms of selecting types and number of roles (i.e., domain experts), planning for project activities specific to potential impediments, and determining requirements for RE tools to be used in project at hand.

  - Theory related to the impediments due to contractual complexities (e.g., non-contiguity of regulatory requirements and abstractness of regulatory requirements) might make the managements rethink on the ways they specify regulatory requirements and draft contract document so that over-all effort for contractual analysis during RE is reduced.

  - Project managements can derive their own project specific metrics which could provide early estimates of effort needed to do compliance work in RE.

  - Effort estimated from the metrics can be used in project planning and scheduling.

- **Academic Research:**

  - Researchers can further derive metrics from the effort critical artefacts and activities from diverse projects. They can also check the validity of the metrics by case studies in actual projects.

- Researchers can collect effort related data against the derived metrics to be used as coefficient value for an effort estimation model for compliance work in RE.
- Researchers can validate this theory in further case studies by deriving specific hypothesis from the theory, which would strengthen the foundations of the theory.
- Researchers can explore the impediments to derive requirements (e.g., automatic traceability, capability of semantic analysis of regulatory text, and visualisations of monitoring progress) for RE tools to be used in such projects.
- Researchers can probe into developing new techniques that can deal the stated impediments by analysing the underlying characteristics (e.g., cross-references, abstractness, non-contiguity of regulatory requirements) of the artefacts.

6.5 Conclusions

Ensuring regulatory compliance of requirements for contractual systems engineering projects is challenging due to complexities of contractual specifications and numerous regulatory documents involved in the projects (Nekvi and Madhavji, 2015; and Berenbach, 2010). Especially, effort required to do such work is grossly underestimated (Nekvi et al., 2016) as the traditional effort estimation tools built for whole software development work do not account for the complexities involved in compliance work in RE. In this chapter, we propose two emerging theories concerning complexities of compliance work in RE and associated metrics for effort (see Section 6.2). The emerging theories are stated as:

(i) *Characteristics of requirements artefacts and their inter-relationships existing in contractual systems engineering projects underlie substantial impediments to doing compliance work in RE;* and

(ii) Key metrics for an effort estimation model for compliance work of RE are derivable through analysis of impediments associated with compliance.

This emerging theories are formulated based on empirical evidence from multiple studies performed in an industrial systems engineering project using a theory building model (i.e., *hypothetico-inductive model*) described by (Sjøberg et al., 2008). The supporting empirical studies are described in Chapter 2, Chapter 3, and Chapter 4. The goodness of the theories are evaluated based on six criteria (e.g., explanatory power, generality, testability, and utility) also proposed by (Sjøberg et al., 2008).The evaluation of the theories are fairly good (see Table 6-2 and Section 6.3) even if they are derived from a single project. It is obvious that only more empirical investigations (in industry or lab settings) can solidify the theories by collecting further supportive, deep, or complementary evidence.

6.6 References

Antoniou, G., Billington, D., Governatori, G., and Maher, M. J., On the Modelling and Analysis of Regulations. Australian Conference on Information Systems, pp. 401–405, 1999.

Bacharach, S.B., Organizational Theories: Some Criteria for Evaluation.The Academy of Management Review, vol. 14(4), pp. 496–515, 1989.

Berenbach,B., Lo, R. and Sherman, B., Contract-based Requirements Engineering. Int. Workshop on Requirements Engineering & Law (RELAW), IEEE Computer Society, CA, USA, pp. 27-33, 2010.

Boehm,B.; and Jain, A., An Initial Theory of Value-based Software Engineering. Value-Based Software Engineering, 1st Edition. Biffl, S.; Aurum, A., Boehm, B., Erdogmus, H., & Grünbacher G., Eds., Springer, LNCS, Germany, pp. 15-33, 2005.

Breaux, T. D., Vail, M. W., and Antón, A. I., Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations. 14th IEEE International Requirements Engineering Conference, Minnesota, USA, pp. 49-58, 2006.

Cohen, B., Developing Sociological Knowledge: Theory and Method. 2nd ed., Belmont, CA: Wadsworth Publishing, 1989.

Ferrari, R., An emerging theory on the interaction between requirements engineering and systems architecting based on a suite of exploratory empirical studies, Ph.D. thesis, University of Western Ontario, September 2010.

Haig, B.D., An Abductive Theory of Scientific Method.Psychological Methods, Vol.10, No. 4, pp. 371–388, 2005.

Ingolfo, S., Siena, A. and Mylopoulos, J., Establishing regulatory compliance for software requirements.Conceptual Modeling (ER'11). LNCS, Vol. 6998,Springer-Berlin, Heidelberg, pp. 47-61, 2011.

Islam, S., Mouratidis, H., Wagner, S., Towards a Framework to Elicit and Manage Security and Privacy Requirements from Laws and Regulations. 16th Int. Working Conf. on REFSQ, LNCS, vol. 6182, Springer, Heidelberg, pp. 255-261, 2010.

Kerrigan, S., and Law, K. H., Logic-based regulation compliance-assistance. ACM 9th International Conference on Artificial Intelligence and Law, Scotland, pp. 126-135, 2003.

Kiyavitskaya, N., Krausova, A., & Zannone, N., Why Eliciting and Managing Legal Requirements Is Hard. Int. Workshop on Requirements Engineering and Law, Barcelona, IEEE Computer Society, pp. 26-30, 2008.

Maxwell, J. C. and Antón, A. I., Checking Existing Requirements for Compliance with Law Using a Production Rule Model. Int. Work. on RELAW, IEEE Com. Soc., pp. 1-6, 2009.

Murtaza, S. S., Finding Faulty Functions from the Traces of Field Failures, Ph.D. thesis, University of Western Ontario, April 2011.

Nekvi, M.R.I., and Madhavji, N.H., Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects: A Case Study. ACM Transactions on Management Information Systems (TMIS), 5(3), p.15, 2015.

Nekvi, M.R.I., Noorwali, I. and Madhavji, N.H., Deriving Metrics for Estimating the Effort Needed in Requirements Compliance Work. International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ), Springer International Publishing, pp. 135-141, 2016.

Ralph, P., Developing and evaluating software engineering process theories. 37th International Conference on Software Engineering, volume 1, IEEE Press, pp. 20-31, 2015.

Ramezani, E., Fahland, D., Werf, J. M. and Mattheis, P., Separating Compliance Management and Business Process Management. Business Process Managements Workshops. LNBIP, Vol. 100, Springer-Verleg, Heidelberg, 2012.

Saeki, M., and Kaiya, H., Supporting the Elicitation of Requirements Compliant with Regulations. 20th Int. Conf. Advanced Information Systems Engineering (CAiSE), Springer-Verleg, Heidelberg, Germany, pp. 228 - 242, 2008.

Sjøberg, D., Dyba, D., Anda, B. C., and Hannay. J., Building Theories in Software Engineering. Guide to Advanced Empirical Software Engineering, Shull, F., Singer,J., and Sjøberg, D.I.K, Eds., Springer, London, pp. 312–336, 2008.

Chapter 6

7. Conclusions and Future Work

In this section, we conclude the thesis in Section 7.1 and outline the directions for future work in Section 7.2.

7.1 Conclusions

Ensuring system compliance to applicable regulations and engineering standards is widely considered as a challenging job due, in part, to the myriad of complexities involved in the elicitation and analysis (Otto and Anton, 2007; Kerrigan, 2003; and Breaux et al., 2006) of the system's requirements from regulations . There is a good deal of research on the encoding techniques to model regulatory text (see (Otto and Anton, 2007) for summary) in order to facilitate requirements elicitation; techniques to elicit regulatory requirements (Breaux et al, 2006; Islam et al., 2010; Siena, 2010; Sunkle et al., 2016); validation and checking of compliance of requirements (Saeki and Kaiya, 2008; Ramezani et al., 2012; Maxwell and Anton, 2009; and Ingolfo et al., 2011); frameworks to deal regulatory evolution in RE (Maxwell et al, 2012); methods to handle issues (e.g., overlapping) in multiple regulations (Ghanavati et al, 2014; Fernandez and Yimam, 2015); and others. However, our analysis of the current RE literature suggests that it still lacks knowledge of the characteristics and challenges of RE projects for large-scale, contract-based, multi-domain systems engineering projects.

To help ameliorate this situation, we conducted three studies on an industrial-scale RE project that sought to ensure regulatory compliance of requirements for a contractual systems engineering project aiming to upgrade a railway system. The RE project is characterised by an over thousand page contract, over 300 standards and regulations, and project requirements specified at various levels of the system such as system level, subsystem level, component level, cross-cutting, and project execution requirements. Project requirements falls in both categories of system and software.

From the first study, we determined the types of the artefacts (e.g., contract, regulations, standards, and project requirements at sub-system, components, and cross-cutting levels) (shown in Table 2-1 of Chapter 2) and their types of inter-relationships (shown in Table 2-3 of Chapter 2) and quantitative characterisations of the inter-relationships (shown in Section 2.4.3) that are existing in RE projects for contractual systems engineering projects. We conclude from these findings that artefacts used in RE process of contractual systems engineering projects have numerous types of inter-relationships for the matter of regulatory compliance of requirements.

This thesis also describes the resultant compliance meta-model that depicts the inter-relationships among the artefacts used in RE compliance (see Figure 2-1 of Chapter 2). The meta-model acts as a domain guide for technical and managerial agents in determining project scope and making project decisions. This meta-model was instrumental in identifying the various impediments (see Section 3.4) for which we later derived effort-related metrics (see Sections 4.3).

From the second study, we further analysed the project data and identified a number of unique impediments (e.g., non-contiguity, abstractness, implicitness, cross-cutting nature, diverse regulatory references of the regulatory requirements) to ascertaining compliance of the requirements (see Section 6.4 of Chapter 3). This thesis concludes that there are a number of significant impediments to regulatory compliance of requirements due to scale and complexity of the contract, regulatory documents, and systems.

These impediments are analysed both in quantitative and qualitative terms to reflect their impact on compliance-related activities in RE (shown in Table 3-3, Table 3-4, Table 3-5, Figure 3-1 and Section 6.4). From our study:

(i) There were over 300 regulatory documents, some quite sizeable (i.e., several hundred pages) (see Figure 3-1) leads to the conclusion that in large systems engineering government project, hundreds of standards and regulations (with hundreds of thousands pages) need to be analysed to ensure compliance;

(ii) Average ratio between regulatory and non-regulatory requirements was 1:20 in the entire document, and 1: 11 in the sample 75 pages chunk used in the study (shown in Section 3.6.3.1 and Figure 3-2) - we conclude from this that regulatory requirements may be specified non-contiguously in the contract, makes RE compliance harder;

(iii) Approx. 50% of all documents were referenced in their entirety from the contractual requirements (see Section 3.6.3.3) - we conclude from this that standards and regulations are mentioned in the contractual requirements as abstract or partial references. This then requires that the analysts need to plough through voluminous or numerous regulatory documents to identify and specify precisely the relevant requirements; and

(iv) Approx. 60% of the segments of a sample standard contained *implicit* requirements (see Table 3-3). We conclude from this that standards may impose a substantial number of new requirements in addition to contractual requirements.

In the third study, we develop a method for deriving metrics to estimate the effort needed to compliance work in RE (see Section 4.3). Further, by using this method we derived a number of key metrics towards algorithmic-based effort estimation model for compliance work of RE (see Table 4-7). We therefore conclude that effort estimation metrics are derivable using our proposed method, which can be constructive to develop an actual effort estimation model for compliance work of RE.

In Chapter 5, we propose a framework for structuring the diverse regulatory requirements and regulatory documents maintained in global products development settings. We illustrate this framework by using examples collected from the industrial project that we have studied (see Section 5.4). This framework is anticipated to increase efficiency for change managements of regulatory requirements.

In Chapter 6, we propose the following two emerging theories that generalise the findings of the thesis:

(i) *Characteristics of requirements artefacts and their inter-relationships existing in contractual systems engineering projects underlie substantial impediments to doing compliance work in RE;* and

(ii) *Key metrics for an effort estimation model for compliance work of RE are derivable through analysis of impediments associated with compliance work of RE.*

These theories are evaluated against criteria for goodness of theory proposed by (Sjøberg et al. 2008). The evaluation is fairly satisfactory but support from more studies is needed to strengthen them further.

As with all case studies, the findings of this thesis have potential threats for generalisability (as discussed in Sections 2.5, 3.7, and 3.9) despite the fact these findings have resulted from an industrial setting (as opposed to a "lab" setting). The findings are novel and also have promising implications to industry and academic research. However, a family of empirical studies needs to be conducted by researchers from the RE community to establish a solid ground for the findings so that these can be reliably used in wider practice (Kitchenham et al, 2004). In this regard, this thesis lays a foundational step towards building grounded theory in the domain of regulatory compliance in RE.

7.2 Future Work

The suite of empirical studies and the emerging theories presented in this dissertation provide important but initial body of knowledge on characteristics and challenges of RE in contractual systems engineering projects from regulatory compliance viewpoint. Since these studies were exploratory, they basically opened up new avenue of scientific knowledge rather than confirming any hypothesis. Thus, directions for future work include replications of these studies in other project contexts and settings. Such studies can test the emerging theory in a variety of domains (e.g., financial, health-care, and communication), and can also go further to include new research questions asking why some phenomena occur (e.g., large percentage (40%) of the regulatory requirements specified in the contract are likely to be cross-cutting - see Section 3.6.2.3). Such research questions include: (i) How likely it is that all the component requirements associated with cross-cutting requirements are not recognised during RE, resulting missing requirements? (ii) What are the architectural problems attributed to regulatory requirements? (iii) Which structuring of contract enables increased ease of processing during RE? Etc.

There are also important future directions for technological research (e.g., developing new RE techniques, methods, and processes) with respect to the identified impediments. It can be researched to identify (and model) common semantics and syntax of: (a) regulatory requirements in the contract and the project requirements specifications, and (b) regulatory rules and constraints mandated in standards and regulations so that parsing tools can be employed to automatically relate corresponding entities in the documents. Technological solutions can be directed towards designing technique and tools for monitoring and visualising the progress of compliance work with respect to various project dimensions such as sub-systems and components (e.g., extent of completed components specifications), contractual analysis (e.g., how much of the contract is analysed for compliance work), and regulatory analysis (e.g., how many documents have been analysed by experts for eliciting requirements).

There is still much work remains towards constructing an effort estimation model for compliance work in RE. This includes solidifying the metrics we derived in Chapter 4 by providing supporting evidence from other studies and identifying more metrics (especially for project complexity (e.g., distributed system), and product and process characteristics (e.g., degree of reuse, platform difficulty, experiences of personnel, etc.) through conducting further empirical studies in different contexts. Future research should seek to collect effort-related data against derived metrics. Once adequate historical data is collected, trustworthy effort estimation model can be constructed. While it is tempting to "guesstimate" effort data for various metrics in the absence of historical data, it is not clear at this stage which metrics could be considered for gathering such preliminary effort data from the analysts and domain experts. Needless to say, regardless of the method used for creating the effort estimation model, it would need to be validated through industrial use. Though this is arduous work and long-term research, I do believe that it is worthwhile in order to reduce the error range of the estimates from the actual, giving an improved capability for resource and time estimation and the quality outcome of a system's compliance requirements.

7.3 References

Breaux, T. D., Vail, M. W., and Antón, A. I., Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations. 14th IEEE International Requirements Engineering Conference, Minnesota, USA, pp. 49-58, 2006.

Fernandez, E.B. and Yimam, D.,Towards Compliant Reference Architectures by Finding Analogies and Overlaps in Compliance Regulations. E-Business and Telecommunications (ICETE), SCITEPRESS, vol. 4, pp. 435-440, 2015.

Ghanavati,S., Rifaut, A., Dubois, E. and Amyot, D., Goal-oriented Compliance with Multiple Regulations.IEEE 22nd International RE Conference ,pp. 73-82, 2014.

Ingolfo, S., Siena, A. and Mylopoulos, J., Establishing regulatory compliance for software requirements.Conceptual Modeling (ER'11). LNCS, Vol. 6998,Springer-Berlin, Heidelberg, pp. 47-61, 2011.

Islam, S., Mouratidis, H., Wagner, S., Towards a Framework to Elicit and Manage Security and Privacy Requirements from Laws and Regulations. 16th Int. Working Conf. on REFSQ, LNCS, vol. 6182, Springer, Heidelberg, pp. 255-261, 2010.

Kerrigan, S., and Law, K. H., Logic-based regulation compliance-assistance. ACM 9th International Conference on Artificial Intelligence and Law, Scotland, pp. 126-135, 2003.

Kitchenham, B., A., Dyb !a, T., Jøorgensen, M., Evidence-based software engineering. 26th International Conference on Software Engineering., Washington, USA, pp. 273–281, Washington, USA, 2004.

Maxwell, J. C. and Antón, A. I., Checking Existing Requirements for Compliance with Law Using a Production Rule Model. 5th IEEE International Workshop on Requirements Engineering and Law. IEEE Computer Society, Los Alamitos, CA, USA,pp.1-6, 2009.

Maxwell, J.C., Antón, A.I. and Swire, P.,Managing changing compliance requirements by predicting regulatory evolution. 20th IEEE Int. RE Conference, pp. 101-110, 2012.

Otto, P. N. and Antón, A. I., Addressing Legal Requirements in Requirements Engineering.IEEE International Requirements Engineering Conference (RE). IEEE Computer Society, CA, USA, pp.5-14, 2007.

Ramezani, E., Fahland, D., Werf, J. M. and Mattheis, P., Separating Compliance Management and Business Process Management. Business Process Managements Workshops. LNBIP, Vol. 100, Springer-Verleg, Heidelberg, 2012.

Saeki, M., and Kaiya, H., Supporting the Elicitation of Requirements Compliant with Regulations. 20th Int. Conf. Advanced Information Systems Engineering (CAiSE), Springer-Verleg, Heidelberg, Germany, pp. 228 - 242, 2008.

Siena, A. 2010. Engineering Law-compliant Requirements. The Nomos Framework. Ph. D thesis. University of Trento, Italy.

Sjøberg, D., Dyba, D., Anda, B. C., and Hannay. J., Building Theories in Software Engineering. Guide to Advanced Empirical Software Engineering, Shull, F., Singer,J., and Sjøberg, D.I.K, Eds., Springer, London, pp. 312–336, 2008.

Sjøberg, D., Dyba, D., Anda, B. C., and Hannay. J., Building Theories in Software Engineering. Guide to Advanced Empirical Software Engineering, Shull, F., Singer,J., and Sjøberg, D.I.K, Eds., Springer, London, pp. 312–336, 2008.

Sunkle,S., Kholkar,D., Kulkarni, V., Informed Active Learning to Aid Domain Experts in Modeling Compliance. IEEE 20th Conf. on EDOC, pp. 1-10, 2016.

Curriculum Vitae

Name:	Md Rashed Iqbal Nekvi
Post-secondary Education and Degrees:	University of Western Ontario London, Ontario, Canada 2010 – 2016 Ph.D. (Computer Science)
	The University of Western Ontario London, Ontario, Canada 2008–2010 M.Sc. (Computer Science)
	Islamic University of Technology (IUT) Gazipur, Dhaka, Bangladesh 2000-2003 B.Sc. (Computer Science and Information Technology)
Honours and Awards:	Organization of the Islamic Conference (OIC) Scholarship Islamic University of Technology, Bangladesh 2000-2003
	Western Graduate Research Scholarship University of Western Ontario 2008-2014
	Govt. of Bangladesh Scholarship Ministry of Education Board, Bangladesh 2000-2003
	Best Presenter at UWORCS Dept. of Computer Science 2012
Related Work Experience	Teaching Assistant University of Western Ontario 2008-2014
	Graduate Research Assistant University of Western Ontario 2008-2016
	Lecturer Int. Islamic University Chittagong, Dhaka, Bangladesh 2004-2008

Software Programmer
Grameen Software Limited, Dhaka, Bangladesh
Jan 2004 - Aug 2004

Publications

1. Nekvi, M.R., Ferrari, R., Berenbach, B., Madhavji, N.H., Towards a Compliance Meta-model for System Requirements in Contractual Projects. IEEE 4th International Workshop on Requirements Engineering and Law (RELAW), IEEE Computer Society, CA, USA, pp. 74-77, 2011.
2. Nekvi, M.R., Madhavji, N.H., Ferrari, R., Berenbach, B., Impediments to Requirements-Compliance. International Working Conference on Requirements Engineering: Foundation for Software Quality, Springer Berlin Heidelberg, pp. 30-36, 2012.
3. Nekvi, M.R.I., and Madhavji, N.H., Impediments to Regulatory Compliance of Requirements in Contractual Systems Engineering Projects: A Case Study. ACM Transactions on Management Information Systems (TMIS), 5(3), p.15, 2015.
4. Nekvi, M.R.I., Noorwali, I. and Madhavji, N.H., Deriving Metrics for Estimating the Effort Needed in Requirements Compliance Work. International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ), Springer International Publishing, pp. 135-141, 2016.
5. Spichkova, M., Schmidt, H.W., Nekvi, M.R.I. and Madhavji, N.H., Structuring Diverse Regulatory Requirements for Global Product Development.IEEE International Workshop on Requirements Engineering and Law (RELAW), pp. 57-60, 2015.