UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DORGIVAL PEREIRA DA SILVA NETTO

Factors influencing companies for reducing ambiguity in legal requirements specification and achieving their compliance with data protection laws

Recife

2021

DORGIVAL PEREIRA DA SILVA NETTO

Factors influencing companies for reducing ambiguity in legal requirements specification and achieving their compliance with data protection laws

A Ph.D. Thesis presented to the Center of Informatics of Universidade Federal de Pernambuco in partial fulfillment of the requirements for the degree of Philosophy Doctor in Computer Science.

Main Area: Software Engineering and Programming Language

Advisor: Profª. Drª. Carla Taciana Lima Lourenço Silva Schuenemann

Recife

2021

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S586f Silva Netto, Dorgival Pereira da
Factors influencing companies for reducing ambiguity in legal requirements specification and achieving their compliance with data protection laws / Dorgival Pereira da Silva Netto. – 2021.
272 f.: il., fig., tab.

Orientadora: Carla Taciana Lima Lourenço Silva Schuenemann.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2021
.
Inclui referências e apêndices.

1. Engenharia de software. 2. Engenharia de requisitos. I. Schuenemann, Carla Taciana Lima Lourenço Silva (orientadora). II. Título.

005.1 CDD (23. ed.) UFPE - CCEN 2022-23

Dorgival Pereira da Silva Netto

“Factors influencing companies for reducing ambiguity in legal requirements specification and achieving their compliance with data protection laws”

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Engenharia de Software e Linguagens de Programação

Aprovado em: 29/10/2021.

_____

Orientadora: Profa. Dra. Carla Taciana Lima Lourenco Silva Schuenemann

BANCA EXAMINADORA

_____

Prof. Dr. Jaelson Freire Brelaz de Castro
Centro de Informática / UFPE

_____

Profa. Dra. Jéssyka Flavyanne Ferreira Vilela
Centro de Informática / UFPE

_____

Profa. Dra. Márcia Jacyntha Nunes Rodrigues Lucena
Departamento de Informática e Matemática Aplicada / UFRN

_____

Prof. Dr. Alberto Manuel Rodrigues da Silva
Instituto Superior Técnico / Universidade de Lisboa

_____

Prof. Dr. Marcos Kalinowski
Departamento de Informática / PUC/RJ

I dedicate this thesis to my family.

ACKNOWLEDGEMENTS

I thank God.

To my parents (Paulo and Paula), my wife (Mayara), and my son (Ravi) for all support and encouragement since the decision to do my doctorate and for helping me to overcome adversities. Thank you so much for helping me get here.

My advisor, Carla Silva, for the teachings and the partnership since the undergraduate research (2010), through the masters and doctorate.

To my friends who supported me in getting here. Anyone reading this thesis, please do not give up on your dreams, no matter how hard it is.

ABSTRACT

Software requirements are mainly specified using natural language, but it brings challenges as it is prone to produce ambiguous specifications. These challenges become bigger when dealing with software requirements that must comply with regulations, the so-called legal requirements. Ambiguous requirements specifications may cause the system not to satisfy the stakeholders' needs and not comply with the legislation. Existing Requirements Engineering approaches to addressing ambiguity and/or achieving legal compliance are not based on knowledge from empirical studies conducted in the software development industry. This thesis aims to overcome this limitation by providing a set of factors and guidelines that help reduce ambiguity in legal requirements specification and achieve specifications compliant with data protection laws. To achieve this objective, we initially carried out a broad study in the literature to characterize the landscape of legal requirements engineering concerning privacy and security. Then, we analyzed works that developed approaches to deal with ambiguity in the specification of legal requirements. We then investigated how the software development industry tackles the problem through an exploratory study based on semi-structured interviews with twenty-two professionals from public and private companies. Data collected from the interviews were analyzed using grounded theory techniques. We identified factors and outlined a theory explaining the relationships between them and how they reduce ambiguity in the specification of legal requirements and the compliance of such requirements with data privacy laws. To validate these factors, we conducted a self-administered online survey with professionals. Findings from the studies reveal that discussions among the team, customer, specialized support areas (Legal Sector, Ambiguity Analysis sector, Anonymization Sector), consulting experienced team members with domain knowledge reduce ambiguity and promote legal compliance in requirements specifications. The theory that emerged from the interviews explains a set of factors influencing the work practices used by public and private companies to deal with ambiguity in legal requirements specification and achieve their compliance with regulations. Researchers and practitioners can use these factors and guidelines to leverage the methods and tools they develop or use to support legal requirements specifications.

Keywords: requirements engineering; ambiguity; legal requirements; legal compliance; empirical study.

RESUMO

Requisitos de software são especificados principalmente utilizando linguagem natural, mas traz desafios, pois tende a produzir especificações ambíguas. Estes desafios tornam-se maiores quando lidam com requisitos de software que devem estar em conformidade com legislações, chamados de requisitos legais. Especificação de requisitos ambígua pode fazer com que o sistema não atenda aos desejos dos stakeholders e não esteja em conformidade com a legislação. As abordagens existentes da Engenharia de Requisitos têm como objetivo tratar a ambiguidade e/ou obter a conformidade legal não são baseadas no conhecimento que emergiu de estudos empíricos conduzidos na indústria de desenvolvimento de software. O objetivo desta tese é superar essa limitação fornecendo um conjunto de fatores e diretrizes que auxiliam na redução da ambiguidade na especificação de requisitos legais e na obtenção de especificações em conformidade com leis de proteção de dados. Para alcançar o objetivo, inicialmente realizamos um estudo amplo na literatura para caracterizar o panorama da engenharia de requisitos legais em relação à privacidade e segurança. Em seguida, realizamos uma análise de trabalhos que desenvolveram abordagens para lidar com ambiguidade na especificação de requisitos legais. Então, investigamos como o problema é tratado pela indústria de desenvolvimento de software através de um estudo exploratório baseado em entrevistas semiestruturadas com vinte e dois profissionais de empresas públicas e privadas. Os dados coletados a partir das entrevistas foram analisados utilizando técnicas de teoria fundamentada. Identificamos fatores e esboçamos uma teoria explicando o relacionamento entre esses fatores e como eles contribuem para a redução da ambiguidade na especificação de requisitos legais e a conformidade de tais requisitos com leis de proteção de dados. Para validar esses fatores, nós conduzimos um questionário online autoadministrado com profissionais. Os resultados dos estudos revelam que as discussões entre a equipe, o cliente e as áreas de suporte especializado (Setor Jurídico, setor de Análise de Ambiguidade, setor de Anonimização), a consulta a membros experientes da equipe com conhecimento do domínio reduzem a ambiguidade e favorecem a conformidade legal nas especificações de requisitos. A teoria que emergiu das entrevistas explica um conjunto de fatores influenciando as práticas de trabalho utilizadas por empresas públicas e privadas para lidar com a ambiguidade na especificação de requisitos legais e o alcance da conformidade com as legislações. Pesquisadores e profissionais podem utilizar estes fatores e as diretrizes para alavancar os métodos e ferramentas que desenvolvem ou utilizam para apoiar a especificação de requisitos legais.

Palavras-chaves: engenharia de requisitos; ambiguidade; requisitos legais; conformidade legal; estudo empírico.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ANPD	Autoridade Nacional de Proteção de Dados
BDD	Behavior Driven Development
DFD	Data Flow Diagram
DPIA	Data Protection Impact Assessment Report
DPO	Data Protection Officer
EU	European Union
GDPR	General Data Protection Regulation
ICO	Information Commissioner's Office
ICT	Information and Communication Technology
IT	Information technology
ITD	Information Technology Departments
K-W	Kruskal-Wallis
LGPD	Lei Geral de Proteção de Dados
M-W	Mann-Whitney
MVP	Minimum Viable Product
NHA	Brazilian National Health Agency
NL	Natural Language
PBD	Privacy by Design
PO	Product Owner
QA	Quality Assurance
RE	Requirements Engineering
RQ	Research Question
SCCs	Software Consultancy Companies
SDLC	Software Development Life Cycle
SHs	Software Houses

SLM	Systematic Literature Mapping
TDD	Test-driven Developmen
UK	United Kingdom

LIST OF SYMBOLS

ϕ	Greek letter Phi
$>$	Greater
$<$	Less than
α	Greek letter Alpha

CONTENTS

1 INTRODUCTION

This chapter presents, in Section 1.1, the motivation and justification for this thesis, highlighting the problems and research gaps related to ambiguity in legal requirements specification compliant with the laws. Section 1.2 presents the research questions and outlines the objectives. Section 1.3 presents the methodological classification. Section 1.4 shows research steps to answer research questions. Section 1.5 presents the summary of the publications. Furthermore, Section 1.6 presents the thesis structure.

1.1 MOTIVATION AND CONTEXT

Information technology (IT) companies must comply with many regulations, including privacy and data protection laws. The amount of data produced and stored in software systems and several incidents of data disclosure have made privacy has emerged as a critical concern. Because most companies use standard practices to collect, store and manage the user's personal information to deliver their services (GHARIB; MYLOPOULOS; GIORGINI, 2020). Software development organizations that process users' data must ensure compliance with data protection laws in all their software systems (CANEDO et al., 2020).

Several laws deal with privacy and protection of personal data as, for example, the legislation of the European Union (EU) named General Data Protection Regulation (GDPR), which entered into force on 24 May 2016 and applied since 25 May 2018, through the Regulation EU 2016/679 (REGULATION, 2016). In South America, several countries have Personal Data Protection Laws, as, among them, Argentina, which has the Personal Data Protection Law 25.326 (ARGENTINA, 2000) (in Spanish, Ley de Protección de Los Datos Personales or PDPA), in force since 1994. In Uruguay, the right to data protection is provided by Law 18.331, edited in 2008 (URUGUAY, REPúBLICA ORIENTAL DEL, 2008) (in Spanish, Protección de Datos Personales y Acción de "Habeas Data"). In Brazil, the General Law of Personal Data Protection, Law 13.709/2018, was approved on 14 August 2018 and came into force in August 2020 (BRASIL, 2018b).

GDPR aims to adapt data privacy laws across Europe to provide individuals more protection and control of their data. Therefore, all software systems that handle the personal data of European citizens must comply with the GPDR. (AYALA-RIVERA; PASQUALE, 2018). Failure to

comply with the GDPR may, in addition to other sanctions, impose fines of up to € 20 million or four percent of an organization's global turnover (REGULATION, 2016). Usman et al. (USMAN et al., 2020) affirm that inconsistent interpretations may have severe consequences. For example, varying interpretations between development organizations and regulating bodies may lead to rework, delays, financial and legal repercussions. This risk is exacerbated because compliance verification is often performed late in the software development process. Consequently, any issues discovered in the compliance check are costly to repair.

Therefore, IT companies must ensure that their business and system requirements are in legal compliance (GHANAVATI; AMYOT; RIFAUT, 2014), also called regulatory compliance. Regulatory compliance is the act of guaranteeing adherence of an organization, process, or (software) product to laws, guidelines, specifications, and regulations to avoid the risk of costly penalties, lost reputation, and brand damage resulting from non-compliance (AKHIGBE; AMYOT; RICHARDS, 2019). Compliance difficulties with privacy laws may lead to rework, delays, financial and legal repercussions (HADAR et al., 2018); (USMAN et al., 2020). Demonstrating compliance in a project cannot wait until the system is implemented; all development stages, including Requirements Engineering (Requirements Engineering (RE)), must play their part in ensuring compliance (NEKVI; MADHAVJI, 2014). Requirements Engineering is the initial phase of Software Engineering that aims to produce agreed requirements document specifying a system that satisfies stakeholder requirements (SOMMERVILLE; SAWYER, 1997).

Some works cite impediments or obstacles to achieve regulatory compliance: domain-specific terms (KERRIGAN; LAW, 2003), cross-reference among documents (MAXWELL et al., 2012; KERRIGAN; LAW, 2003; OTTO; ANTóN, 2007), conflicts among laws (OTTO; ANTóN, 2007; KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008; MAXWELL et al., 2012), changes in the law (OTTO; ANTóN, 2007; KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008). Ensuring that systems meet legal requirements and comply with the legislation has been treated as a challenging and essential problem by the RE community (KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008), (OTTO; ANTóN, 2007).

Kiyavitskaya et al. (KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008) mention that regulatory text often contains vague, ambiguous, and abstract terms or lacunae. Legal texts, in general, are inflexible, non-negotiable, incomplete, unclear, open to different interpretations, and can change with new legislation (RABINIA; GHANAVATI, 2017). In addition, legal texts are full of ambiguities, often planned ones, called intentional ambiguity (OTTO; ANTóN, 2007) (MASSEY et al., 2014). This type of ambiguity allows laws and regulations to avoid dependence on

technologies or practices that may change over time (BERRY; KAMSTIES, 2004) (OTTO, 2009).

The item (2) (a), paragraph §164.306 of the HIPAA (Health Insurance Portability and Accountability Act) (Centers for Medicare & Medicaid Services, 1996) presents an example of intentional ambiguity: *requires covered entities and business partners to "protect against any reasonably anticipated threats or risks to security or integrity of such information*." However, the legal text does not define what constitutes "reasonably" (OTTO; ANTóN, 2007). Ambiguities such as this can impair the understanding of the legal excerpt, which may lead to non-compliance of the system with regulations, resulting in legal and financial penalties.

However, when there is intentional ambiguity, analysts need to establish an interpretation of the law and maintain traceability with the section being interpreted (OTTO; ANTóN, 2007). The ability to maintain traceability from the source of the legal text to the relevant requirement should be addressed by any methodology that aims to support the requirements engineering, and compliance verification process (WEBEL; STEGLICH, 2017). Other ambiguities result from a lack of knowledge of the domain, which provides an essential justification for legal requirements or describes situations contrary (for example, conflicts) to legal requirements (BREAUX; ANTÓN, 2007). For Siena (SIENA, 2010), the complex structure of legal documents and their lack of standardization make their understanding even more complicated, requiring the need for human interaction, mainly to resolve or interpret ambiguous issues.

Wagner et al. (WAGNER et al., 2019) identified that the most systematic way to documentation requirements are: free-form textual domain/business process models, free-form textual structured requirements lists and use case models as text with constraints, structured requirements lists (documented textually with constraints), and free-form textual use case models (semi-formal). Software requirements specified in natural language bring challenges for RE as it is prone to produce ambiguous specifications.

The use of natural language to specify requirements has some benefits. Natural language is universal (can describe any requirement in any application domain), flexible (requirements can be characterized subjectively or in detail), and very widespread (anyone can read and write such requirements) (KAMSTIES; PEACH, 2000). However, the main disadvantage of natural language when used to specify requirements is the ambiguity inherent in the language (KAMSTIES; PEACH, 2000).

Ambiguity is not only a disadvantage of natural language, but it is also a characteristic (KAMSTIES; PEACH, 2000). When ambiguity is not perceived and identified in the requirements elicitation, it can lead to an incomplete, unmanaged, imprecise, or ambiguous requirement

specification, which may eventually lead to an increase in cost and time (FIRESMITH, 2007). In addition, ambiguous requirements generate confusion, wasted effort, and rework (SHAH; JINWALA, 2015). These challenges become bigger when dealing with software requirements that comply with regulations, the so-called legal requirements.

Many works propose approaches to help engineers address ambiguity in RE and align system requirements with legal constraints (BHATIA et al., 2016; GHANAVATI; AMYOT; RIFAUT, 2014; MASSEY; HOLTGREFE; GHANAVATI, 2017; OTTO, 2009; AYALA-RIVERA; PASQUALE, 2018). Other approaches aim to detect (KAMSTIES; PEACH, 2000) and classify ambiguity in requirements (MASSEY et al., 2014) and (MASSEY et al., 2015). Still, ambiguity remains a recurring problem. The works presented do not propose guidelines for reducing ambiguity in specifying legal requirements for software products.

Privacy Engineering is an emerging research area that focuses on designing, implementing, adapting, and evaluating theories, methods, techniques, and tools to capture and address privacy issues in developing products and services (GURSES; ALAMO, 2016). Privacy requirements engineering should be done with a broad and deep understanding of legislative requirements, standards, and policies, but requirements engineers usually do not have expertise in these areas (MEAD; MIYAZAKI; ZHAN, 2011). Many developers do not have sufficient knowledge and understanding about privacy, nor do they sufficiently know how to develop software with privacy (HADAR et al., 2018). Moreover, interpreting legal requirements is not a trivial task for individuals who lack legal expertise, as with most requirements engineers. As a result, requirements engineers are prone to make interpretations and inferences that are inconsistent with the law (OTTO; ANTóN, 2007) (HOSSEINI et al., 2021).

Moreover, many developers do not have sufficient knowledge and understanding about privacy, nor do they sufficiently know how to develop software with privacy (HADAR et al., 2018).Lawyers and engineers bring different, sometimes conflicting, perspectives to interpreting legal texts (SWIRE; ANTON, 2014). One reason is the difference between the software development process (whether based on plans or agile) and the one used for legal compliance. These gaps make specifying legal requirements a problematic and challenging activity for the requirements analyst. Among the implications of these insufficient guidelines, we can mention that organizational and system requirements can arise from different sources of law. Therefore, organizations must analyze all relevant legislation and prioritize them to identify the legal requirements to be met (KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008). Extracting requirements from the legal text and interpreting them is a complex and error-prone process (OTTO; ANTóN,

2007) (LE et al., 2019). Mapping legal obligations on software features are also not trivial (GJERMUNDRØD; DIONYSIOU; COSTA, 2016). To find requirements in legal texts, organizations must identify relevant parts of the information in these documents and understand the relationship between various pieces of information (KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008).

Due to these characteristics, requirements analysts find it difficult to capture the true meaning of legal statements, their implications, and possible consequences of incompatibility between regulations and system requirements (RABINIA; GHANAVATI, 2017).

Due to incompatibility between the legal concepts (as understood by a layman) and the technical state of the art, a literal implementation of the GDPR may decrease the attainable accurate security level, thus hurting privacy (KUTYLOWSKI; LAUKS-DUTKA; YUNG, 2020). Therefore, the participation of a lawyer or data protection law expert is required to support the identification, extraction, and specification of software requirements in compliance with legal text.

Besides, it is necessary to understand how professionals deal with this problem in practice. Therefore, conducting empirical studies exploring and discovering the current practices used in the industry to address ambiguity and legal compliance in the privacy requirements specification. Moreover, Gurses and Álamo (GURSES; ALAMO, 2016) state that empirical studies are needed to explore how different engineering contexts address privacy issues currently. A recent mapping study evidenced that current literature does not present the state of practice related to how the industry deals with ambiguity in legal requirements specification and legal compliance in developing software systems (NETTO; PEIXOTO; SILVA, 2019).

Based on this motivation, the following section describes the research question and the objectives of this thesis.

1.2 PROBLEM DEFINITION AND RESEARCH QUESTIONS

In this context, this thesis provides a set of factors and guidelines, obtained from industry practices, that influence the activity of the requirements analyst to elaborate a requirements specification with reduced ambiguity and compliant with legislation.

This work proposes to answer the following questions:

RQ1. What are the existing approaches to deal with the ambiguity in the specification of legal requirements and achieve compliance with data protection laws?

RQ2. How do organizations deal with ambiguity in legal requirements specifications and comply with data protection laws?

RQ3. What practices are defined in academia and industry to address ambiguity in legal requirements and specify legal compliance systems with data protection laws?

RQ4: How to produce a requirements specification with reduced ambiguity and legal compliance with data protection laws?

This thesis aims to define a theory based on constructs, propositions, and explanations and provide a set of factors and guidelines that help reduce ambiguity in legal requirements specification and achieve specifications compliant with data protection laws. To answer these research questions, we have defined the following specific objectives:

Understand which ones approaches to deal with the ambiguity in the specification of legal requirements and achieve compliance with data protection laws (to answer RQ1).

Present the current state of art on Requirements Engineering approaches to achieve unambiguous legal requirements specification and legal compliance;

Understand how do organizations deal with ambiguity in legal requirements specifications and comply with data protection laws (to answer RQ2).

Investigate in the industry how public and private companies treat legal compliance and address ambiguity in specifying legal requirements.

Define a theory for specifying legal requirements with reduced ambiguity and compliant with data protection laws (to answer RQ3).

Define a set of guidelines for software developers to specify legal requirements with reduced ambiguity and compliant with data protection laws (to answer RQ4).

1.3 RESEARCH METHODOLOGICAL CLASSIFICATION

We based on the methods and guidelines of Empirical Software Engineering (KITCHENHAM et al., 2002) that explore, describe, predict, and explains natural, social, or cognitive phenomena using scientific methods and the experience of evidence-based methods (SJØBERG et al., 2008). The Table 1 summarizes the methodological classification of the research.

Table 1 – Research methodological classification

Code	**Title**
Philosophical position	Constructivist
Research questions	Exploratory
Explanations	Inductive
Research method	Mixed-method
Data analysis	Sequential exploratory

Source: The author (2021).

This thesis adopts philosophical position **constructivist**. Constructivist concentrate less on verifying theories, and more on understanding how different people make sense of the world and assign meaning to actions (EASTERBROOK et al., 2007). Theories may emerge from this process, but they are always tied to the context being studied (EASTERBROOK et al., 2007). Therefore, we concentrate on building local theories that can explain the observed phenomenon, which is the treatment of ambiguity in the specification of legal requirements compliance with legislation.

The type of research questions that guide this study is primarily **exploratory**. Use exploratory questions in the early stages of research as we attempt to understand the phenomena and identify useful distinctions that clarify our understanding (EASTERBROOK et al., 2007). To elaborate the explanations about the studied phenomenon, we follow an **inductive** approach. In this case, the researcher aims to discover what is happening in a particular context, seek new insights, and generate ideas and hypotheses for future studies (RUNESON; HöST, 2009).

The choice of research methods depends upon the theoretical stance of the researcher, access to, and how closely the method aligns with the questions that have been posed (EASTERBROOK et al., 2007). **Mixed-method** research was used in order to investigate the phenomenon in the literature and industrial practice (through interviews and surveys) and build a description about it. Mixed method research employs data collection and analysis techniques associated with both quantitative and qualitative data (EASTERBROOK et al., 2007). Following the classification of Creswell (CRESWELL, 2013) the strategy used in this thesis is the **sequential exploratory**. We collected and analyzed qualitative data, followed by collecting and analyzing quantitative data and results to interpret qualitative findings. Qualitative researchers study phenomena in their natural settings, attempting to make sense of, or interpret, phenomena regarding the meanings people bring to them (MERRIAM; TISDELL, 2015).

AIM: Understand how ambiguity is treated in legal requirements specification and how to verify the legal compliance.

QUALITATIVE: data collected from interviews with practitioners.

QUANTITATIVE: qualitative data used to create a survey self-administered to practitioners and verify findings of qualitative data empirically.

SYNTHESIS: Sequential

This mix-method research investigated how ambiguity in legal requirements is handled in literature (snowballing) and industry through empirical studies (interviews and surveys). Based on the methodology here detailed, we obtained the results reported in Chapter 4, Chapter 5, Chapter 6 and Chapter 7. Section 4.2 shows the method for conducting the exploratory study through semi-structured interviews and threats to validity for the study. Section 5.2 presents the methodology for surveying professionals.

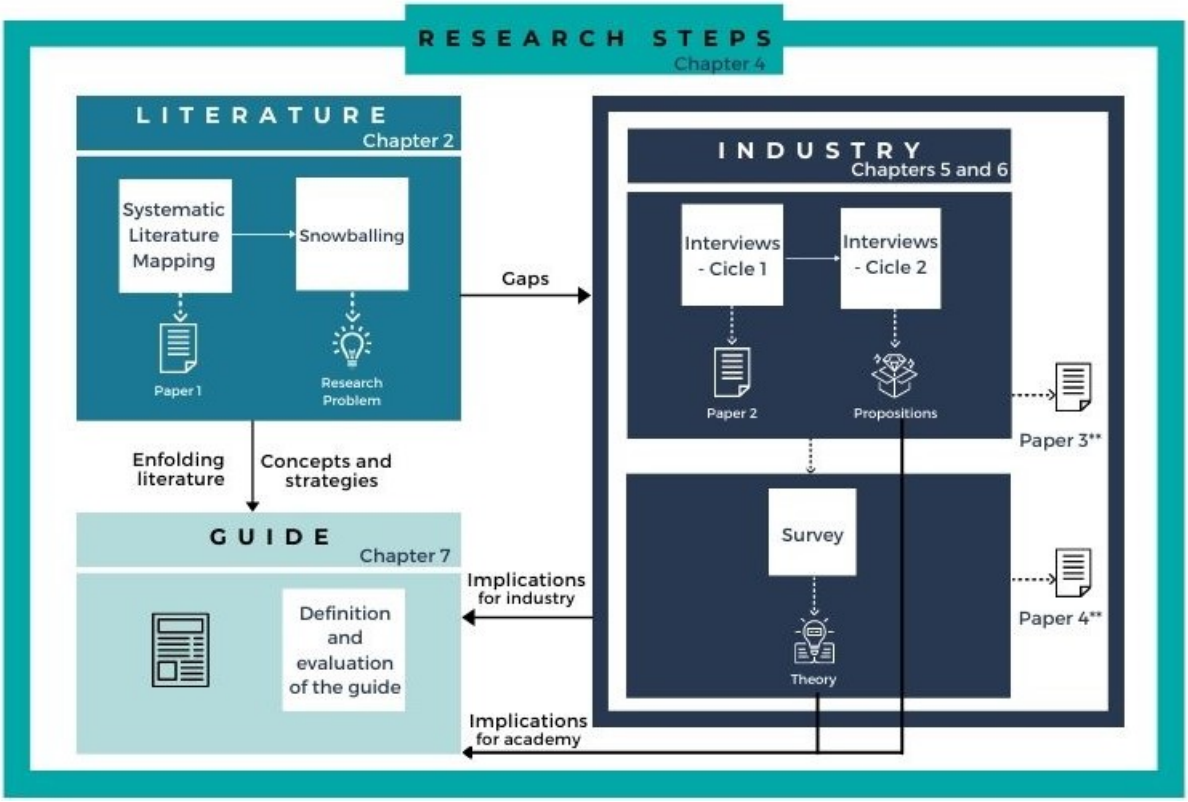
Based on our results, we developed an theory of how IT Companies address ambiguity resolution and compliance with Data Protection Laws in the requirements specification. Moreover, we analyze how the initiatives described can be used, based on the practices in literature and industry, and define a set of guidelines to specify legal requirements with reducing ambiguity and compliant with the laws. Requirements Analysts can adopt this guide proposed in this thesis and apply them in the context of compliance with legal requirements.

1.4 RESEARCH STEPS

Figure 1 presents the research steps to answer the research questions, presented in section 1.2.

To answer RQ1, we investigated the literature through a systematic literature mapping (SLM), shown in Figure 1. The SLM about security and privacy in requirements engineering followed the procedures indicated by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007). It covers the period from 2000 to 2016, and primary studies were selected, analyzed, and synthesized to understand the current state of approaches concerning privacy and security in the RE domain. The protocol, the execution procedures, and the analysis of the SLM results can be seen in Appendix A, and (NETTO; PEIXOTO; SILVA, 2019).

Figure 1 – Research steps protocol



Source: The author (2021).

The SLM pointed gaps related to ambiguity in legal requirements specification compliant with the law. Still, aiming to respond to RQ1, we conducted a more specific bibliographic study on ambiguity in legal requirements using the snowballing technique (WOHLIN, 2014) to identify research problems related to legal requirements, ambiguity, and legal compliance. The protocol, the execution procedures, and the analysis of the snowballing results can be seen in Chapter 2 and Appendix B. These two studies in the literature assisted us in delimiting the research problem of this thesis. In addition, we identified concepts and strategies related to ambiguity in the specification of legal requirements compliant with the legislation.

To answer RQ2, we conducted an exploratory study through semi-structured interviews with 22 professionals (developer, tester, lawyer, project managers, requirements analysts, and others) from eighteen Brazilian public (6) and private (12) IT companies. To understand how they deal with the inherent ambiguity in legal requirements specification and how they verify the legal compliance. This study was conducted through two cycles of interviews, shown in Figure 1. The first, with nine interviewees of three public and four private companies in Brazil (NETTO; SILVA; ARAÚJO, 2019). And the second cycle of interviews with thirteen interviewees

of three public and eight private companies in Brazil.

The interview data were analyzed using qualitative research coding techniques (MERRIAM; TISDELL, 2015) based on Grounded Theory (CORBIN; STRAUSS, 2014) techniques: open, axial, and selective. Kasurinen et al. (KASURINEN; TAIPALE; SMOLANDER, 2010) define the steps of grounded theory analysis: open (where categories and related codes are extracted from the data), axial (where identify connections between categories and codes), and selective (where the main category is identified and described). Planning consists of preparing, validating the interview script, and conducting a pilot interview to refine the script. After conducting the interviews, there was a transcription and, later, coding and analysis. Chapter 4 presents the protocol, the execution procedures, and the analysis of the results of these exploratory studies, Appendix F and G the interviews scripts and coding process [1].

Given state of the art (systematic literature review and snowballing) and the practice (twenty-two interviews), it was possible to identify a set of eight propositions influencing the specification of legal requirements with reduced ambiguity and compliance with data protection laws. Furthermore, we identify twenty-seven factors to promote or mitigate the categories from interview data.

The interview-based study pointed to propositions that need to be empirically validated. Then, we surveyed 39 IT professionals and legal experts through a self-administered online questionnaire to validate the strategies and proposals for reducing ambiguity in the specification of legal requirements compliant with the legislation, which emerged from the interviews.

This survey collected the software practitioners' perceptions regarding the factors and actions to achieve ambiguity reduction and legal compliance in the software requirements specification. Chapter 5 presents the protocol, the execution procedures, and the analysis of the survey results, and Appendix I the survey questions.

Based on survey data analysis, to answer RQ3, we identify practices defined in academia and industry to address ambiguity in legal requirements specification and legally compliant. Analyzing the results of the exploratory studies and in the investigations carried out with practitioners, we define a theory that helps the practitioners reduce ambiguity in legal requirements specification and achieve specifications compliant with data protection laws. A theory represents a set of constructs, general propositions, and, possibly, explanations of those propositions (SJØBERG et al., 2008). To our knowledge, this is the first study to bring in perspectives from state of the art and practice to elaborate a requirements specification with reduced ambiguity

[1] https://dorgivalnetto.github.io/journal2021/

and compliant with data protection laws.

Next, from these two studies with industry professionals, to answer RQ4, we have identified lessons learned on dealing with ambiguity in the specification of legal requirements compliant with legislation and define mitigation actions and a guide for professionals, in Chapter 7. A guide that helps Requirements Analysts to specify legal requirements with reduced ambiguity in compliance with the law.

1.5 THESIS-RELATED PUBLICATIONS

In this section, we list papers originated from this thesis.

1. NETTO, D.; Silva, C. . Addressing Ambiguity in Legal Requirements Engineering of Software Systems. In: VIII WORKSHOP DE TESES E DISSERTAÇÕES DO CBSOFT (WTD-SOFT 2018), 2018, São Carlos. Anais do 8º Workshop de Teses e Dissertações. São Carlos, 2018. p. 19-28.

2. NETTO, D.; MAIA, M. ; Silva, C. . Privacy and Security in Requirements Engineering: Results from a Systematic Literature Mapping. In: Workshop on Requirements Engineering, 2019, Recife. Proceedings 22nd Workshop on Requirements Engineering, 2019.

3. NETTO, D.; Silva, C. ; ARAUJO, J. . Identifying How the Brazilian Software Industry Specifies Legal Requirements. In: XXXIII Simpósio Brasileiro de Engenharia de Software, 2019, Salvador. Anais do X CONGRESSO BRASILEIRO DE SOFTWARE: TEORIA E PRÁTICA (CBSOFT), 2019. - **Best Paper Insightful Ideas and Emerging Results Track**.

4. *How Information Technology Companies Address Ambiguity Resolution and Compliance with Data Protection Laws in Requirements Specification?*" - Under evaluation by Requirements Engineering Journal.

5. *How Information Technology Companies Address Ambiguity Resolution and Compliance with Data Protection Laws in Requirements Specification - a survey-based study* Under evaluation by Journal of Systems and Software.

1.6 DOCUMENT OUTLINES

This thesis is structured as follows:

Chapter **2** presents the **Theoretical Background** regarding ambiguity in requirements engineering, legal compliance, ambiguity present in the legal text, and the specification of

legal requirements compliant with the legislation.

Chapter **3** presents the **Related works** to the theme.

Chapter **4** presents the findings from an exploratory study (**Interview-based study**) conducted through semi-structured interviews with employees of public and private companies.

Chapter **5** presents the findings from an online **Survey-based study** conducted through a self-administered questionnaire with IT professionals. We defined a set of propositions inferred from interview-based study and literature.

Chapter **6** presents the evaluation findings of the propositions and derive explanations, and the **Definition of Theory** based on a set of constructs, propositions, and explanations.

Chapter **7** describes the **Mitigation actions and Guidelines** for specifying legal requirements with reduced ambiguity in accordance with the legislation.

Chapter **8 Conclusion and future works**: presents the final considerations, including the contributions achieved and the indications of future work.

Finally, attached to this document, we have Appendix A - Systematic Literature Mapping protocol, Appendix B - Snowballing, Appendix D - Invitation Letter, Appendix E - Consent Term, Appendix F - Interview script, Appendix G - Interview script (specialized support), and I - Survey (Questionnaire).

2 THEORETICAL BACKGROUND

In this chapter, we briefly review the main themes that support this research. Section 2.1 briefly introduces Requirements Engineering (RE). Section 2.2 introduces Regulatory compliance in RE. Section 2.3 introduces the ambiguity in legal requirements specification.

2.1 REQUIREMENTS ENGINEERING

Requirements Engineering (RE) involves discovering, documenting, and maintaining a set of requirements for a computer-based system. Requirements Analysts should use systematic and repeatable techniques to ensure that the system requirement is complete, consistent, and relevant (SOMMERVILLE; SAWYER, 1997).

Discovering or requirements elicitation is the process of collecting system requirements from stakeholders (SOMMERVILLE; SAWYER, 1997). Stakeholders are individuals who use, request, and develop the system, who are interested in developing it, for example, analysts, customers, investors, and users (KOTONYA; SOMMERVILLE, 1998). Documenting or requirements specification is an official statement of the system requirements for customers, end-users, and software developers that will serve as a reference for other software engineering activities (SOMMERVILLE; SAWYER, 1997).

In a survey with participants from ten countries, Wagner et al. (WAGNER et al., 2019) identified that the interview as the most popular approach to capturing requirements, followed by a requirements workshop, prototyping, scenarios, and observation. Wagner et al. (WAGNER et al., 2019) also classified requirements specification or documentation techniques according to their formality: textual (natural language), semi-formal (for example, UML diagrams), and formal (with a mathematical or formal semantic basis). The most frequent documentation requirements are: free-form textual domain/business process models, free-form textual structured requirements lists and use case models as text with constraints, structured requirements lists (documented textually with constraints), and free-form textual use case models (semi-formal).

The following section presents the Legal compliance in Requirements Engineering, which is the act of ensuring the adherence of an organization, process, or software product to laws, guidelines, specifications, and regulations (AKHIGBE; AMYOT; RICHARDS, 2019).

2.2 LEGAL COMPLIANCE IN REQUIREMENTS ENGINEERING

Legal texts have some characteristics that make them unique compared to other sources of requirements: laws can complement, overlap or contradict one another; some areas of the law undergo a constant change; regulations can also refer to other sections of a specific regulation (OTTO; ANTóN, 2007). References to other sections within a given legal text or other parts of the law are called cross-references. These references force requirements engineers to spend additional time reading and understanding legal texts, even before they can begin to extract critical concepts (OTTO; ANTóN, 2007). Regulatory compliance is a source of non-functional requirements (GLINZ, 2007).

These features highlight the need for legal compliance to be adequately considered. Governatori (GOVERNATORI, 2010) defines legal compliance as the relationship between two particular specifications: the system to be developed and the legal specifications. According to Boella et al. (BOELLA et al., 2014), the legal requirements are the way found by the Requirements Engineering to make it possible that the relevant standards for the application domain are correctly codified in the Information Systems.

Compliance requirements typically aim at a broad range of stakeholders and use cases, and they are purposefully expressed in general terms, omitting implementation-specific details (USMAN et al., 2020).

From the perspective of information systems, organizations have to implement the regulatory requirements into their data processes to stay compliant. For organizations and companies to be compliant with regulations, they must ensure that appropriate technologies, privacy controls, and safeguards are implemented in their existing systems, and programs (TANKARD, 2016).

To demonstrate legal compliance, we have to ratify, through traceability, that all legal requirements have been elicited from the legal texts (NEKVI et al., 2012). Legal compliance in Requirements Engineering has been the subject of several studies (SIENA et al., 2013), (GHANAVATI; AMYOT; RIFAUT, 2014), (RABINIA; GHANAVATI, 2017), (ZENI et al., 2015).

According to Boella et al. (BOELLA et al., 2014), systems should allow a dialogue between legal and industry experts, allowing more informed judgments about legal requirements and relevant applications that transfer academic research to the legal sector.

Kiyavitskaya et al. (KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008) mention that regulatory text often contains vague, ambiguous, and abstract terms or lacunae. Legal texts are full of

ambiguities, often planned ones, called intentional ambiguity (OTTO; ANTóN, 2007)(MASSEY et al., 2014). This type of ambiguity allows laws and regulations to avoid dependence on technologies or practices that may change over time (BERRY; KAMSTIES, 2004) (OTTO, 2009).

Therefore, it is often the case that IT professionals or individuals responsible for ensuring legal compliance and accountability lack sufficient guidance to manage their legal obligations (BREAUX; ANTÓN; SPAFFORD, 2008). Lack of guidelines for IT professionals to understand what requirements must be operationalized and implemented in an organization's software system to support compliance (KOOPS; LEENES, 2014) (BREAUX; ANTÓN; SPAFFORD, 2008).

Failure to comply with privacy laws and regulations results in penalties imposed by data protection authorities. Some works cite impediments or obstacles to achieve regulatory compliance: domain-specific terms (KERRIGAN; LAW, 2003), cross-reference among documents (MAXWELL et al., 2012; KERRIGAN; LAW, 2003; OTTO; ANTóN, 2007), conflicts among laws (OTTO; ANTóN, 2007; KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008; MAXWELL et al., 2012), changes in the law (OTTO; ANTóN, 2007; KIYAVITSKAYA; KRAUSOVá; ZANNONE, 2008).

Privacy has emerged as a critical concern because most companies use common practice to collect, store, and manage the user's personal information to deliver their services (GHARIB; MYLOPOULOS; GIORGINI, 2020). Therefore, privacy requirements engineering should be done with a broad and deep understanding of legislative requirements, standards, and policies. However, requirements engineers usually do not have expertise in these areas (MEAD; MIYAZAKI; ZHAN, 2011).

Several laws deal with privacy and protection of personal data as, for example, the legislation of the European Union (EU) named General Data Protection Regulation (GDPR), which entered into force on 24 May 2016 and applied since 25 May 2018, through the Regulation EU 2016/679 (REGULATION, 2016). In South America, several countries have Personal Data Protection Laws, as among them, Argentina, which has the Personal Data Protection Law 25.326 (ARGENTINA, 2000) (in Spanish, Ley de Protección de Los Datos Personales or PDPA), in force since 1994. In Uruguay, the right to data protection is provided by Law 18.331, edited in 2008 (URUGUAY, REPúBLICA ORIENTAL DEL, 2008) (in Spanish, Protección de Datos Personales y Acción de "Habeas Data").

GDPR requires organizations and companies who process personal data to adjust and change their existing systems to meet the requirements that the GDPR puts forward. GDPR aims at all businesses and other entities situated outside of the EU to collect and process European citizens' data and comply with GDPR (TANKARD, 2016). Personal data refers to

the information managed by organizations and companies about an identified, or identifiable individual (BLIX; ELSHEKEIL; LAOYOOKHONG, 2017).

GDPR requires organizations and companies who process personal data to adjust and change their existing systems to meet the requirements that the GDPR puts forward. Under GDPR (REGULATION, 2016) principles for personal data processing: Lawfully (personal process data only when there is an appropriate legal basis or legislative measure under the GDPR, EU, or Member State law.

Complying with the GDPR will strongly affect small- and medium-sized enterprises (SME) that cannot necessarily afford juridical help to comply with the new rules of the GDPR (TIKKINEN-PIRI; ROHUNEN; MARKKULA, 2018)

The Brazilian General Law of Personal Data Protection or LGPD, Law 13.709/2018 (BRASIL, 2018b), came into force in August 2020, has ten chapters and 65 articles dealing with control measures, evidence of compliance, rights, and obligations concerning the processing of personal data. About its Art. 3, the law applies to any processing operation that aims to offer or provide goods or services or process data from individuals in the Brazilian territory.

The ten items of Art. 6 of the LGPD (BRASIL, 2018b) establish the guiding principles that must be observed as a minimum requirement for personal processing data and creating policies, services, or products.

> *Purpose: treatment for legitimate, specific and explicit purposes and informed to the holder;*
>
> *Adequacy: Compatibility of the treatment with the purposes informed to the holder;*
>
> *Need: check that all data is necessary and compatible with the purpose;*
>
> *Free access: guarantee, holders, free and easy consultation on the form of treatment;*
>
> *Data quality: guaranteeing the data subject's accuracy, clarity, relevance, and updating of data;*
>
> *Transparency: guaranteeing holders clear, accurate, and easily accessible information on the performance of the treatment;*
>
> *Security: use of technical, administrative, training and awareness measures to protect personal data;*

Prevention: adoption of technical and administrative measures to prevent the occurrence of damages due to the processing of personal data;

Non-discrimination: inability to perform treatment for discriminatory, illegal or abusive purposes;

Accountability: demonstration, by the agent, of the adoption of effective measures capable of proving the observance and compliance with the rules of protecting personal data and, even, of the effectiveness of these measures.

Noncompliance with the law incurs rough treatment and is prone to the following administrative sanctions applicable by the National Data Protection Authority (in Portuguese, Autoridade Nacional de Proteção de Dados - ANPD), provided for in Art. 52 (BRASIL, 2018b):warning; simple fine (2 % of the billing in the last financial year, limited to fifty million Brazilian Real, for infraction); daily fine; the publication of the infraction; blocking and/or elimination of the personal data to which the infraction refers; partial suspension of the operation of the database; suspension of the exercise of the treatment activity; total prohibition of the exercise of activities.

2.3 AMBIGUITY IN REQUIREMENTS ENGINEERING

Ambiguity is a natural language (Natural Language (NL)) attribute and a necessary feature that makes NL adaptable in various contexts (BANO, 2015). Gervasi and Zowghi (GERVASI; ZOWGHI, 2010) define ambiguity as a phenomenon whereby several different meanings can be assigned to the exact requirement (or, more generally, sets of requirements). Locating the source or root cause of ambiguity can be challenging.

Ambiguity has been a research topic in both linguistics and requirements engineering. Linguistic researchers have identified the various forms of ambiguity in language use and have classified textual ambiguity in various ways ((HOFFMANN et al., 2012); (MASSEY et al., 2014)). Casellas (CASELLAS, 2011) provides a comprehensive overview of the current literature on legal ontologies to represent and formalize legal knowledge.

According to Shah and Jinwala (SHAH; JINWALA, 2015), there are two significant categories of ambiguity: linguistic and Software Engineering. The first one can be noticed by any

reader who knows the language. The second depends on the domain involved and can only be perceived by readers who have sufficient knowledge of the domain.

Berry et al. (BERRY; KAMSTIES, 2004) present the types of linguistic ambiguity: lexical, syntactic, semantic, pragmatic and language error.

- **Lexical** ambiguity of a word arises when a word can have a few different meanings.
- **Syntactic** ambiguity arises when the sentence can be parsed more than one way, resulting in more than one grammatical structure where each provides different meanings.
- **Semantic** ambiguity arises when the predicate logic can lead to multiple interpretations, although there is no lexical and syntactic ambiguity.
- **Pragmatic** ambiguity is concerned with the relationship between the interpretations of a sentence and its context.
- Ambiguity **language error** is a grammatically incorrect construction that is understood and possibly in different ways due to the error.

Kamsties et al. (KAMSTIES et al., 2001) analyzed requirements documents and classified the ambiguity as lexical ambiguity, systematic polysemy, referential ambiguity, discourse ambiguity, and domain ambiguity. Kamsties and Paech (KAMSTIES; PEACH, 2000) present several ER-specific ambiguities (requirements document ambiguity, application domain, system domain, development domain):

- **requirements document ambiguity** occurs if a requirement allows several interpretations concerning what is known about other requirements in the requirements document;
- an **application domain ambiguity** occurs if a requirement allows for multiple interpretations of what is known about the application domain;
- a **system domain ambiguity** occurs if a requirement allows for multiple interpretations for what is known about the system domain;
- A **development domain ambiguity** occurs if a requirement allows for multiple interpretations concerning what is known about the development domain.

Breaux and Antón (BREAUX; ANTÓN, 2007) differentiated the intentional and unintentional ambiguity present in legal texts. Intentional ambiguity can be classified, according to Breaux and Antón (BREAUX; ANTÓN, 2007), in

Logic: when a single word can be interpreted in several ways when conjunctions are present;

Attributive: when a sentence within a sentence refers to another sentence, and it is not clear what this other sentence consists of;

Referential: when words like "this," "that," "they" are used to refer to other parts of the text. It is not always clear where the reference is.

Ambiguities are prevalent in laws and regulations (MASSEY et al., 2014). According to Massey et al. (MASSEY et al., 2015), many of the approaches developed to mitigate or disambiguate the requirements specifications are not appropriate to deal with legal ambiguities. They cannot quickly rewrite legal texts, and they must be clarified through interpretation rather than a reformulation if ambiguity appears in current law or regulation. Lawyers and engineers bring different, sometimes conflicting perspectives to interpreting legal texts (SWIRE; ANTON, 2014).

Berry et al. (BERRY; KAMSTIES, 2004) present a set of words frequently used in legal texts and requirements with ambiguity, imprecision, or uncertainty. Therefore, they must be avoided or used with care. Tjong (TJONG, 2008) states that any requirements specification with the words all, and, or, and/or, but, unless, if, only, also, it, they, are potentially ambiguous. Massey et al. (MASSEY et al., 2014) developed a taxonomy with six types of ambiguity to identify and classify ambiguous terms.

- **Lexical ambiguity** occurs when a word or phrase has multiple valid meanings. Consider "*Melissa walked to the bank*." This could mean that Melissa walked to a financial institution or she walked to the edge of a river.

- **Syntactic ambiguity** occurs when a sequence of words has multiple valid grammatical parsings. Also: "Quickly read and discuss this paragraph."

- **Semantic ambiguity** occurs when a sentence has more than one interpretation based entirely on the surrounding context. Consider "Fred and Ethel are married." and "Fred kissed his wife, and so did Bob." Further context is needed to determine if Fred and Ethel are married to each other or separately. Nor do we know if Fred has cause to be annoyed.

- **Vagueness** occurs when a term or statement admits border line cases or relative interpretation. Consider: "Fred is tall." If Fred was a North American male and 5'2" tall, then

the claim is not true. If Fred was 7'0" tall, then the claim is supported. Somewhere in between lie heights that reasonable people might disagree as to constituting "tall."

- **Incompleteness** occurs when a statement fails to provide enough information to have a single clear interpretation. Similarly, "Combine flour, eggs, and salt to make fresh pasta." omits some necessary information such as quantity of materials and techniques to be employed.

- **Referential ambiguity** occurs when a word or phrase in a sentence cannot be said to have a clear reference. Consider: "The boy told his father about the damage. He was very upset." The pronoun 'he' could refer to either the boy or the father.

Ambiguity in requirements specification can lead to different stakeholders, including software designers, regulators, and users, to interpret system behavior and functionality (BHATIA et al., 2016).

Berry, Kamsties, and Krieger (BERRY; KAMSTIES; KRIEGER, 2003) define two strategies to minimize ambiguity. The first is to establish a context, as the language is always interpreted about a context; if the context is not explicit and agreed upon by all stakeholders in the elicitation session, misinterpretations can occur. Another strategy is that the requirements engineer should paraphrase what he understood from the specifications of the customers and users in their own words so that the customers and users themselves can identify the ambiguity.

Reidenberg et al. (REIDENBERG et al., 2016) investigate a specific type of ambiguity, vagueness, in website privacy policies. To do this, they developed a theory of vague and ambiguous terms. Table 2 shows a set of keywords that have imprecision (vagueness), identified and classified (condition, generalization, modality, numerical quantifier) based on the application of the taxonomy to privacy policies.

Berry et al. (BERRY; KAMSTIES, 2004) defined words that are ambiguous and commonly used and should be avoided or used carefully in a legal requirement or declaration: *and, or, any, include, after, before, next, previous, minimum, maximum*. The authors also cite some vague adjectives and adverbs: *acceptable, accurate, appropriate, easy, efficient, essential, immediately, periodically, sufficient, user-friendly*. Verbs and nouns that can be vague: *support, handle, process, reject, use, etc., the user*. Some examples of words expressing uncertainty: *can, may, probably, possibly, usually*. A few more terms from work by Berry et al. (BERRY; KAMSTIES, 2004) that can introduce ambiguity are *all, each, every, even, this, otherwise, not,*

Table 2 – Vague terms identified in Reidenberg's work et al. (2016)

Category	**Keywords and phrases**
Condition	Depending, necessary, appropriate, inappropriate, as needed, as applicable, otherwise reasonably, sometimes, from time to time
Generalization	Generally, mostly, widely, general, commonly, usually, normally, typically, largely, often, primarily, among other things
Modality	May, might, can, could, would, likely, possible, possibly
Numerical quantifier	Anyone, certain, everyone, numerous, some, most, few, much, many, various, including but not limited to

Source: Reidenberg et al. (2016).

because, the, different from, a, any, each, one, some, many, few, by, until, only, also, other. Wiegers and Beatty (WIEGERS; BEATTY, 2013) present some ambiguous terms and cite some ways to remove ambiguity (Table 3).

Reidenberg et al. (REIDENBERG et al., 2016) suggest principles to improve imprecision: avoid the ambiguous terms in Table 2 that are problematic. Moreover, increase imprecision, specifically generalizations and modal verbs. Use a glossary of key terms to standardize terminologies with many software developers (website developer, mobile app developer, database administrator, back-end) — flag when the meanings of terms change in a policy. Berry et al. (BERRY; KAMSTIES; KRIEGER, 2003) present strategies that can be used to deal with ambiguity in requirements specification: increasing the accuracy of natural language. Provide more background information to allow the reader to resolve ambiguities themselves.

According to Berry et al. (BERRY; KAMSTIES; KRIEGER, 2003), glossaries, style guides, standard sentences, and controlled languages increase natural language accuracy and decrease ambiguity. The glossary helps to avoid lexical ambiguity; style guides assist authors in writing requirements. A standard sentence assists in the articulation of requirements. Controlled language increases the understanding of any technical documentation, reducing the ambiguity inherent in natural language through restricted grammar and a fixed vocabulary. Kamsties and Paech (KAMSTIES et al., 2001) describe how to detect ambiguities in natural language requirements using a checklist.

According to Berry et al. (BERRY; KAMSTIES; KRIEGER, 2003), any strategies help provide more context information and consequently disambiguate a requirement:

A comment can be to provide the rationale for a requirement.

Table 3 – Ambiguous terms to avoid in requirements

Ambiguous terms	**Ways to improve them**
acceptable, adequate	Define what constitutes acceptability and how the system might judge it
and/or	Specify if you mean "and", "or" or "any combination of", so the reader doesn't have to guess
as much as practicable	Don't leave it to developers to determine what is workable.
at least, at a minimum, not more than, not to exceed	Specify the maximum and minimum acceptable values
best, greatest, most	Define the desired level of achievement and the minimum level desired
between, from X to Y	Define whether boundary points are included in the range
depends on	Describe the nature of the addiction
efficient	Define how efficiently the system uses resources, how quickly it performs specific operations, or how quickly users can perform certain tasks with the system
fast, quick, rapid	Specify the minimum acceptable time that the system takes some action
flexible, versatile	Describe the ways in which the system should be able to adapt to changing operating conditions, platforms or business needs
i.e., e.g.	Use words in your native language, don't mess with Latin abbreviations
improved, better, faster, superior, higher quality	Quantifying how much better or faster constitutes adequate improvements in a specific functional area or quality aspect
including, including but not limited to, and so on, etc., such as, for instance	List all possible values or functions, not just examples, or direct the reader to the full list location.
in most cases, generally, usually, almost always	Clarify when the stated condition or scenario does not apply and what happens to them. Describe how the user or system can differentiate one case from another.
match, equals, agree, the same	Define whether a text comparison is case sensitive and whether the phrase means "contains" "starts with" or is "exact".
maximize, minimize, optimize	Enter the minimum and maximum values of some parameters
normally, ideally	Identify abnormal or non-ideal conditions and describe how the system should behave in these situations
optionally	Clarify whether this is a developer, system, or user choice

Source: Wiegers and Beatty (2013)

Table 4 – Ambiguous terms to avoid in requirements (continued)

Ambiguous terms	**Ways to improve them**
probably, ought to, should	Will you go or not?
reasonable, when, necessary, where appropriate, if possible, as applicable	Explain how the developer or user can make this judgment
robust	Define how the system handles exceptions and responds to unexpected operating conditions
seamless, transparent, graceful	Translate user expectations into specific observable product characteristics
several, some, many, few, multiple, numerous	Indicate how much or provide the minimum and maximum limits of an interval
shouldn't, won't	Try to state the requirements as positive, describing what the system will do.
sufficient	Specify how much of something is enough
support, enable	Define exactly what functions the system will perform that "supports" some capability

Source: Wiegers and Beatty (2013)

The rationale describes why a requirement is necessary.

Fit criteria describe a condition that the software product must meet to satisfy a requirement, providing contextual information and leaving less room for interpretation.

An inverse requirement describes functionality that the software product does not perform, helping to reduce pragmatic ambiguity, generality, and imprecision

Information traceability also helps to disambiguate a requirement because it helps identify related requirements that provide contextual information

2.4 CHAPTER SUMMARY

This chapter presented the baseline of this work: Requirements Engineering, Legal Compliance in Requirements Engineering, and Ambiguity in Requirements Engineering.

We described how the ambiguity in the specification of legal requirements had been addressed. Next, we presented the difficulties faced by the Requirements Engineer when faced with legal requirements related to personal data protection laws.

In the next chapter, we present the Related Works. We intended to find similar works dealing with ambiguity in legal requirements, legal compliance, and empirical studies in legal requirements.

3 RELATED WORKS

3.1 AMBIGUITY IN LEGAL REQUIREMENTS

Many approaches deal with developing techniques to identify or eliminate ambiguity in the specification of legal software requirements and ensure legal compliance.

Ambiguities are prevalent in laws and regulations (MASSEY et al., 2014). According to Massey et al. (MASSEY et al., 2015), many of the approaches developed to mitigate or disambiguate the requirements specifications are not appropriate to deal with legal ambiguities. Legal texts cannot be easily rewritten; if ambiguity appears in current law or regulation, it must be clarified through interpretation rather than a reformulation. Lawyers and engineers bring different, sometimes conflicting perspectives to interpreting legal texts (SWIRE; ANTON, 2014).

Otto and Antòn (OTTO; ANTóN, 2007) surveyed 38 articles of research in regulatory modeling, extracting critical concepts from legal texts, and compliance monitoring. The works were grouped into nine categories: symbolic logic, knowledge representation, deontic logic, defeasible logic, first-order temporal logic, access control, markup-based representations, goal modeling, reusable requirements catalogs. Also, in this work, they identified critical elements that systems to support the analysis of legal texts for requirements specification and compliance monitoring must have: identification of relevant legislation, classification of legislation with metadata, prioritization of legislation and exceptions, management of the evolution of legislation and the law, traceability between references and requirements, data dictionary and glossary, to ensure consistency, semi-automated navigation, and search, annotation of legal declarations, queries comparing legal concepts and compliance.

Ghanavati et al. (GHANAVATI; AMYOT; RIFAUT, 2014) proposed the Legal-GRL, a goal-oriented legal requirement modeling approach that shows traceability links between the system requirements model and the legal requirements model. The approach has a qualitative and quantitative analysis of the degree of legal compliance for a system's requirements model. Rabinia and Ghanavati (RABINIA; GHANAVATI, 2017) present an extension of the Legal-GRL to model legal requirements in first-order logic and extract requirements from legal documents. However, both approaches do not address the identification, classification, and resolution of ambiguity in legal requirements aiming at the legal compliance of systems.

Fernandes et al. (FERNANDES; SILVA; GONÇALVES, 2018), presents a methodology for prepar-

ing a reusable catalog of personal data protection requirements. The proposed approach deals with linguistic patterns and styles for specifying personal data protection requirements extracted from legal documents for designing and implementing information systems capable of communicating with those who process individuals' data. The use of standards is one of the strategies to reduce ambiguity, but the work is focused to the GDPR and does not consider the specification of legal requirements. Caramujo et al. (CARAMUJO et al., 2019) developed RSL-IL4Privacy that is a domain-specific language for the specification of privacy policies from the web and mobile applications.

Massey et al. (MASSEY et al., 2011) carried out an experimental assessment with students to show that students with a background in software engineering are not prepared to make legal compliance decisions. Legal requirements metrics are helpful to assist analysts when they have to make legal compliance decisions. Massey et al. (MASSEY et al., 2014) define a ambiguity' taxonomy with six types: lexical, syntactic, semantic, imprecision, incompleteness, and referential ambiguity. In this study, the correctness of the identified types of ambiguity is assessed against a HIPAA paragraph. In a second work, Massey et al. (MASSEY et al., 2015) carry out a case study using the ambiguity' taxonomy and compare the types identified by the participants for the same pieces of legislation in the first work (MASSEY et al., 2014).

3.2 LEGAL COMPLIANCE

Hoffmann et al. (2012) proposed software legal requirements standards (extracted from recurrent legal requirements), described in a technical language that guarantees that even if the requirements analyst does not have the legal knowledge, he can use them. Requirements standards will be used to produce catalogs of legal requirements.

The identification and extraction of software requirements from legal texts is a recurrent problem in requirements engineering. Boella et al. (BOELLA et al., 2014) contrasted the methodologies of extracting legal requirements in the RE community (software analysts) from legal practice (legal experts). They propose that systems enable dialogue between legal and industry experts, increase awareness of the cultural gap between experts, and facilitate clear communication in the legal RE process. The analysis of what legal provisions mean for legal requirements engineering must be carried out by legal experts who must investigate legislation, the interaction between legislative provisions and case law, legal doctrine, and regulatory conversations. Legal experts need to explain the values and assumptions behind norms, given

in the legal community but not well-known in the industry. Legal experts should be aware that in RE, all requirements must be made explicit. It is the job of the legal expert to supply any' missing' rule that is commonly understood but not articulated.

Ayala-Rivera and Pasquale (AYALA-RIVERA; PASQUALE, 2018) have defined a systematic approach called GuideMe that helps organizations understand the data protection obligations imposed by the GDPR and identify measures to ensure compliance. However, the approach supports practitioners in eliciting solution requirements from the GDPR legal obligations, not addressing the ambiguity in the specification of legal requirements.

3.3 EMPIRICAL STUDIES IN LEGAL REQUIREMENTS

Usman et al. (USMAN et al., 2020) performed an industrial case study with seven participants and identified several challenges related to the compliance requirements grouped into three categories: requirements specification-related challenges, process-related challenges, and resource-related challenges. In the first category, all participants highlighted the challenge of interpreting the compliance requirements in the context of their product, differences in understanding the compliance requirements, and abstractness of the compliance requirements. They identified the interpretation of the compliance requirements as the most challenging aspect of the compliance work. In process-related challenges, the participants highlighted the need to improve the alignment between different compliance activities, have a consistent process for all compliance requirements, and balance compliance and business requirements. In the third challenge, all participants highlighted the lack of available resources and time for handling the compliance requirements, lack of awareness among developers about compliance requirements, and lack of awareness among developers about design rules.

Some surveys investigate requirements engineering problems in the industry. For example, Beecham et al. (BEECHAM; HALL; RAINER, 2003) analyze in twelve software organizations that most RE problems are organizational rather than technical; and Kalinowski et al. (KALINOWSKI et al., 2016) investigate incomplete/hidden requirements in a survey with 14 companies from Austria and 74 from Brazil. However, it is out of scope to investigate ambiguity or legal compliance.

Difficulty in communication between legal professionals and IT professionals is one of the problems mentioned by respondents in our study. Nevertheless, the work by Boella et al. (BOELLA et al., 2014) does not present strategies to address problems related to compliance

with the legal requirements specification, and they investigate aspects related to methodologies that support organizations achieving compliance with the GDPR.

In a study made by Billgren and Ekman (EKMAN; BILLGREN, 2017), the authors interviewed employees (in positions such as security management, product management, and project management) working with and in organizations trying to comply with the GDPR about the challenges imposed by GDPR on organizations. Their findings showed that many companies had trouble interpreting the regulation, the actual meaning behind the articles and recitals, and translating it into implementable requirements (EKMAN; BILLGREN, 2017).In addition, they found seven data process-related challenges organizations face as they adjust to the GDPR: Interpretation of Regulations, Ad-hoc and Generic Solutions, Resource Allocation, Organizational Compliance, Continuous Compliance, Documentation, Monitoring, Legacy Systems, and Competing Compliance Measures. In Billgren and Ekman's work (EKMAN; BILLGREN, 2017) they cite challenges related to legal compliance, such as *how to interpret regulatory documents into implementable requirements*, but do not propose strategies to mitigate them. We surveyed professionals with the most diverse roles and companies that operate in different sectors and considered the ambiguity in the legal requirements specification.

Canedo et al. (CANEDO et al., 2020) performed a systematic literature review to identify work-related to privacy requirements and what methodologies and techniques are used to specify them. They did not find industry reports using the methodologies and techniques found in the literature and studies reporting the benefits of their practical application.

Canedo et al. (CANEDO et al., 2020) surveyed practitioners in the software development industry to identify what is the perception of these professionals regarding software privacy, privacy requirements, LGPD, and how the organizations in which they work are handling the need to develop LGPD-compliant systems. Practitioners stated that they lack the knowledge necessary to implement privacy principles and LGPD guidelines. Furthermore, organizations need to disclose their privacy solutions based on their organizational policies, and organizations must provide their professionals with specific training related to data privacy laws.

Sirur, Nurse e Webb (SIRUR; NURSE; WEBB, 2018) performed an in-depth understanding of the challenges and concerns faced by organizations when interacting with the regulations and the processes used by organizations when implementing GDPR. They performed twelve semi-structured interviews with respondents (Senior Security Executive, Privacy and Data Protection Consultancy, software developer) from various backgrounds working in different areas (education, government, telecommunications). This study found challenges in translating

GDPR into a technically implementable format. The most significant non-technical issue for most respondents of smaller organizations was deciphering the expectations of GDPR (or the corresponding state-issued guidance); considerable effort had to be expended to understand what was expected of their organizations to comply. Some respondents discussed GDPR's generality, which is not itself to concrete technical requirements. Understanding the semantics and meaning behind the words of GDPR was not as simple. Despite their thoughts on the clarity of GDPR's intentions, the more technically focused respondents also expressed that without a legal professional of some kind, the average engineer would struggle to utilize the regulations directly.

The work by Canedo et al. (CANEDO et al., 2020) and Sirur, Nurse and Webb (SIRUR; NURSE; WEBB, 2018) identify professionals' perceptions of how companies are developing systems in compliance with personal data protection laws (LGPD and GDPR, respectively). However, these work do not propose strategies to mitigate the problems, nor are they concerned with the ambiguity in the legal requirements specification.

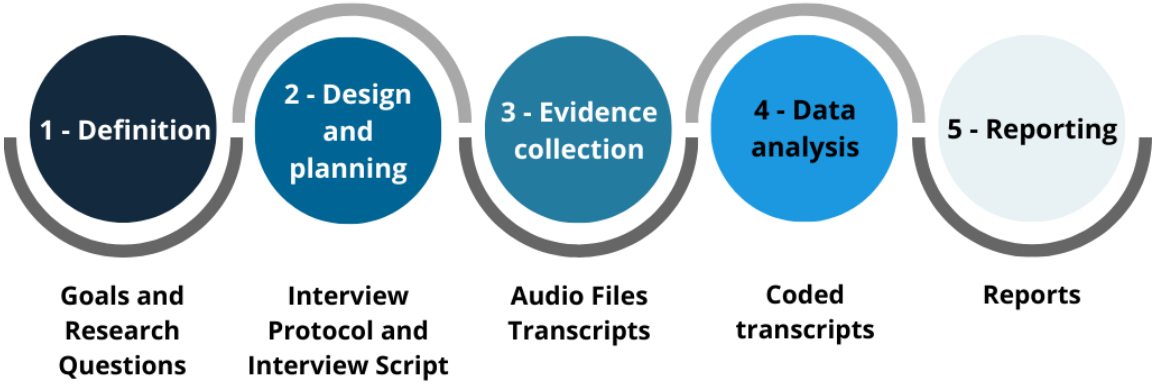
3.4 CHAPTER SUMMARY

Ambiguity in legal and software requirements specification is a well-known problem in academic and industry communities. Privacy is a matter that deserves attention from everyone within the company because it is a point of vulnerability in the actions that it performs in the company. Thus, this chapter presented research that deals with the inherent ambiguity in legal requirements specification and how they perform legal compliance verification.

4 INTERVIEW-BASED STUDY

This section presents the design and planning of an interview-based study involving professionals from different companies and people who support the software development process (legal sector, project managers, analysts, among others). Understand how the software development industry (public and private companies) deals with the ambiguity inherent in specifying legal requirements and achieving legal compliance. Figure 2 shows five phases of the research method, based em Kitchenham et al. (KITCHENHAM et al., 2002):

Figure 2 – Overview of the research method



Source: The author (2021).

1. Definition of goals and research questions.
2. Design and planning of interview protocol, including interview script.
3. Evidence collection between performing interviews.
4. Data analysis (record audio files transcripts, coding, interpreting).
5. Reporting.

4.1 RESEARCH GOAL (DEFINITION)

This interview-based study aims to understand better how companies (public and private) deal with the inherent ambiguity in the legal requirements specification and how they verify legal compliance. Does the following second research question guide this study: *How do*

organizations deal with ambiguity in legal requirements specification and achieve law compliance?, according to the practitioner's perspective. We investigate how the companies perform legal requirements elicitation (Research Question (RQ)1) and specification (RQ3) because the ambiguity (RQ2) in legal requirements can lead to an ambiguous requirements specification, which may eventually entail non-compliance with data protection laws (RQ4).

RQ1. How do companies perform legal requirements elicitation?

RQ2. How do companies deal with inherent legal requirements ambiguity?

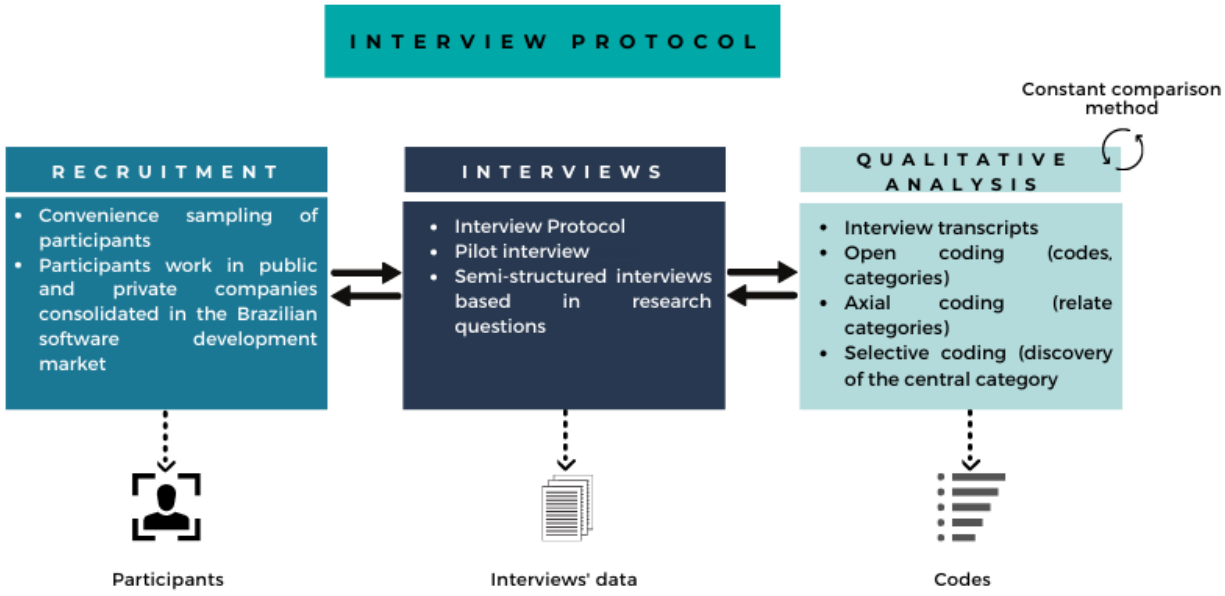
RQ3. How do companies perform legal requirements specification?

RQ4. How do companies verify Legal Compliance?

4.2 STUDY DESIGN AND PLANNING

Figure 3 presents the steps performed in the exploratory study with data collection through identification.

Figure 3 – Interview protocol



Source: The author (2021).

We used semi-structured interviews to help ensure that standard information on predetermined areas is collected but allow the interviewer to go deeper when required (ROBSON,

2002). Interviews allow a better understanding of the questions and explain the aspect under study. Besides, interviews allow discussions and clarifications when gathering the data, making it possible to investigate and compensate for differences in understanding and terminology (PALOMARES; QUER; FRANCH, 2017). It is essential to consider that requirements practices and requirements-related concepts can differ from project to project. Then, the interviews were transcribed, codified, and analyzed.

Determining population is a crucial element for qualitative researches. The target population was a global community of software professionals with experience in legal requirements projects. We adopted a non-probabilistic sampling method (KITCHENHAM; PFLEEGER, 2008) due to the difficulty of identifying and approaching a large number of professionals to form different organizations. Then, we use convenience sampling (KITCHENHAM; PFLEEGER, 2008), participants are selected based on the researcher's accessibility.

4.2.1 Sample

The sample of participants is employees of public and private companies consolidated in the Brazilian software development market. In this study, public companies are governmental organizations in charge of developing software and providing IT services for public administration companies. These companies typically have a robust legal department composed of legal experts involved in understanding the rules they have to comply with when regulations like LGPD (BRASIL, 2018b), General Data Protection Regulation (GDPR) (REGULATION, 2016), or others come into force.

To increase data diversity, we looked for companies with different characteristics regarding size (large and small, having respectively more and fewer than 50 workers), sector (public or private), and domain (see Table 2). We use the domain classification of software companies defined by Palomares et al. (PALOMARES; QUER; FRANCH, 2017):

1. *Software Consultancy Companies (Software Consultancy Companies (SCCs))* that performed software development tasks for different Clients as their primary business;

2. *IT Departments (Information Technology Departments (ITD))* that usually performed or outsourced some software development tasks for covering the organization's internal demands;

3. *Software Houses (Software Houses (SHs))* that develop and commercialize specific proprietary solutions.

About 70 IT professionals of different positions (developers, requirements analysts, project managers, legal specialists, and others) were invited by email (available in Appendix D). From these, 22 agreed to be interviewed voluntarily. Those who agreed were required to sign a Consent Term (available in Appendix E), which guaranteed confidentiality of the data, anonymity, and the right to withdraw from the research at the moment. The objective was to interview at least two participants from the same Company. However, in the end, we had one public Company (C01) with three subjects, and two private (C06 e C10) with two subjects, and the other companies with only one.

We aimed for good coverage of age, background, education, years of employment, and participation in different organizations to ensure a potentially fertile sample. The average participants' experience is 14.7 years, with values ranging from 3 to 31 years (see Table 2). Of the 22 participants, thirteen have more than ten years of experience.

Gathered data were anonymized, transcripts, and analyzed using open, axial, and selective coding techniques from qualitative research (MERRIAM; TISDELL, 2015). The constant comparison method (SEAMAN, 2008) was used to code, categorize and synthesize it. The codes were grouped into categories representing factors that explain how the companies deal with ambiguity in privacy requirements specification or factors that explain how the companies ensure legal compliance. Relationships among factors represent propositions. Propositions are causal relationships among concepts that explain a phenomenon (PANDIT, 1996). As a result, some factors have been discovered to create two stories representing the ambiguity treatment in the legal requirements specification. Second, they achieved legal compliance in developing software systems in Brazilian IT companies. A story is simply a descriptive narrative about the phenomenon of study (PANDIT, 1996).

Table 5 – Characterization of the Participants

S ID[1]	**IT Exp. (yrs)**[2]	**Position**	**C ID**[3]	**Publ. or Priv.**[4]	**C Domain**[5]	**Nat. or Mult.**[6]	**C Age (yrs)**7	**Emp.**[8]	**Legal Dep.**[9]
S1	31	IT Director	C01	Public	SCC ITD	Nat.	50+	1000-5000	Yes
S2	10	Project Mgr.	C01	Public	SCC ITD	Nat.	50+	1000-5000	Yes
S3	5	Programmer Analyst	C02	Private	ITD	Mult.	20-30	6000+	Yes
S4	18	Senior Mgr.	C03	Public	ITD	Nat.	20-30	51-200	Yes
S5	22	Program Analyst	C04	Public	SCC	Nat.	10-19	11-50	Yes
S6	3	Deployment Analyst	C05	Private	SCC	Nat.	10-19	51-200	Yes
S7	9	Requirements Analyst	C06	Private	SCC	Nat.	20-30	501-1000	Yes
S8	3	Requirements Analyst	C07	Private	SCC	Nat.	0-5	11-50	No
S9	15	Project Mgr.	C01	Public	SCC ITD	Nat.	50+	1000-5000	Yes
S10	7	Product Owner	C08	Private	SH	Nat.	10-19	11-50	No
S11	8	Requirements Analyst	C09	Public	ITD	Nat.	50+	501-1000	Yes
S12	9	Privacy and Security Mgr.	C10	Private	SCC SH	Nat.	5-10	51-200	Yes
S13	5	Anonymization Mgr.	C10	Private	SCC SH	Nat.	5-10	51-200	Yes
S14	18	IT Lawyer	C11	Private	SCC	Nat.	20-30	51-200	Yes
S15	17	Project Mgr.	C12	Private	SH	Nat.	10-19	300+	Yes
S16	23	Project Mgr.	C13	Private	SH ITD	Nat.	20-30	51-200	No
S17	20	Senior Developer	C14	Private	SH	Nat.	15	201-300	No
S18	14	Quality Analyst	C15	Public	SCC ITD	Nat.	50+	1000-5000	Yes
S19	8	Security Consultant	C16	Private	SCC	Nat.	0-5	1-10	Yes
S20	15	Project Mgr.	C17	Private	SCC	Nat.	20-30	300+	Yes

Source: The author (2021).

Table 6 – Characterization of the Participants (continued)

S ID[1]	IT Exp. (yrs)[2]	Position	C ID[3]	Publ. or Priv.[4]	C Domain[5]	Nat. or Mult.[6]	C Age (yrs)[7]	Emp.[8]	Legal Dep.[9]
S21	30	IT Mgr.	C06	Private	SCC	Nat.	20-30	501-1000	Yes
S22	18	IT Lawyer	C18	Public	Education	Nat.	20-30	501-1000	Yes

[1] Interviewed ID; [2] Years IT Experience in Industry; [3] Company ID; [4] Public or Private Company; [5] Company Domain; [6] National or Multinational Company; [7] Company Age; [8] Number of Employees; [9] Legal Department in place

Source: The author (2021).

4.3 DATA COLLECTION

The second stage of the interview-based study protocol, shown in Figure 3, is to perform semi-structured interviews. We are using an interview script specifically designed and composed of 15 open-ended questions (interview script for IT professionals available in Appendix F, interview script for other professionals available in Appendix G), following Strandberg's (STRANDBERG, 2019) guidelines for ethical interviews. The questions were presented in a funnel format, with more general questions presented initially (i.e., interviewee education, background, professional experience, Company characterization, which are relatively simple to answer). The preliminary questions were followed by the more specific questions related to the context of this work (RUNESON et al., 2012). Two researchers with more than 15 years of experience in the Requirements Engineering field analyzed the interview script. A pilot interview was performed with a senior IT professional from a public Company, validated the interview script, and improved it. Then, the answers from the pilot interview were discarded.

Before the interview, each participant received by email the Consent Term (available in (NETTO; SILVA, 2021)), which explained the overall objective and importance of the research, guaranteed data confidentiality, participation anonymity, and the right to withdraw from the research at any moment.

Conducted the interviews using a remote call, such as phone call, Skype, Google Meet, or Hangouts and recorded all interviews with each interviewee's permission. They occurred between November 2018 and October 2019. Each interview lasted an average of 47 min and, altogether, it resulted in 17h and 23 min of audio time. Collected data were discussed

between the authors, and inconsistencies were addressed in discussions and/or complementary explanations provided by the participants. When doubts arose regarding answers, we contacted the respondent again to clarify these doubts.

The first cycle of Interviews was carried out from November 2018 to March 2019 with nine interviews (NETTO; SILVA; ARAÚJO, 2019). The second cycle of interviews was conducted from October 2019 to March 2020 with thirteen interviews (NETTO; SILVA; ARAÚJO, 2021). Despite an interval of more than six months between the interview cycles, and the topic of privacy and data protection and the General Data Protection Law (LGPD) have been widely discussed in society and academia, we did not notice any divergences in the outcomes.

4.4 QUALITATIVE DATA ANALYSIS

We followed the guidelines provided by Strauss and Corbin (CORBIN; STRAUSS, 2014) to categorize and synthesize data to identify the practices adopted to reduce ambiguity in legal requirements specification and the practices adopted to achieve legal compliance. The first author transcribed each audio interview, all interviews were anonymized, and QDA Miner[1] was used to support the analysis process. The manual transcription process was valuable for the research because it allowed a more accurate interpretation of the data. The interview transcripts lasted an average of 2h and 12 min, and, altogether, it lasted 48h 40 min.

Portions of the text were labeled using codes (we started using open coding, then axial coding, and, finally, selective coding) (CORBIN; STRAUSS, 2014). Coding gives a label (representing a specific theme, area, construct) to essential portions of the interview transcript. One code is usually assigned to many pieces of text, and one piece of text can be assigned more than one code (RUNESON; HöST, 2009). At the beginning of the analysis, we used open coding to identify relevant portions of each interview transcript and created the nodes from the interviewees' excerpts.

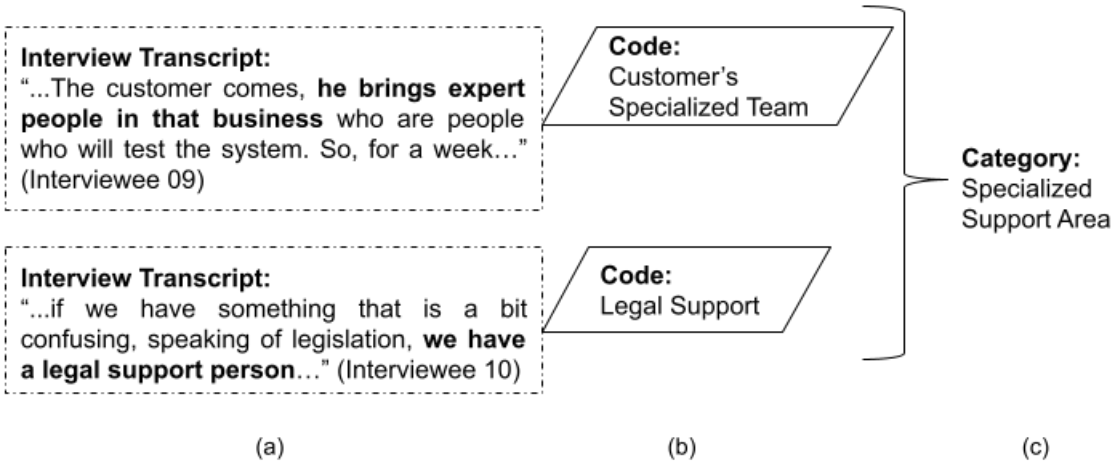
According to Strauss and Corbin (CORBIN; STRAUSS, 2014), axial coding is needed to investigate the relationships between concepts and categories that have been developed in the open coding process and grouped in similar codes. Vollstedt (VOLLSTEDT; REZAT, 2019) states that after having broken up the data in open coding, they are joined together in a new way in axial coding links are worked out between a category and its subcategories. The focus of axial coding is on a category (the phenomenon) concerning the following aspects (VOLLSTEDT;

[1] https://bit.ly/2gGLnTP

REZAT, 2019).

Figure 4 illustrates the category creation process. We started from the interview transcripts by marking and coding relevant parts of the text (Figure 4 (a)), that gave rise to the codes (Figure 4 (b)). Then, we compared all codes built in the first step and grouped the codes referring to the same concept into a category (Figure 4 (c)). To perform this step, we analyzed the coded interview transcripts again to ensure that the similar codes referred to the same concept.

Figure 4 – (a) Open coding of interview transcripts; (b) codes; (c) Category building from the codes



Source: The author (2021).

For example, analyzing the excerpt from interview 09, "...The Customer comes, he brings expert people in that business who are people who will test the system...", we assigned the code **Customer's Specialized Team**. In the excerpt from interview 10, "...if we have something a bit confusing, speaking of legislation, we have a legal support person...", we assigned the code **Legal Support**. These codes were constantly compared to codes in the same interview and other interviews. We perceived the similarity between these codes (Customer's Specialized Team and Legal Support) and merged them to create the "Specialized Support Area" category in axial coding.

In the coding stage, two authors participated, and the third author solved the conflicts. The two researchers discussed each code and refined the coding structure to reduce overlaps following this analysis. Each transcript was analyzed again to ensure that all relevant detail was captured and correctly codified. The categories represent how the industry addresses ambiguity in privacy requirements specification and achieves legal compliance.

We also found some secondary factors (represented as a statement with an arrow to a

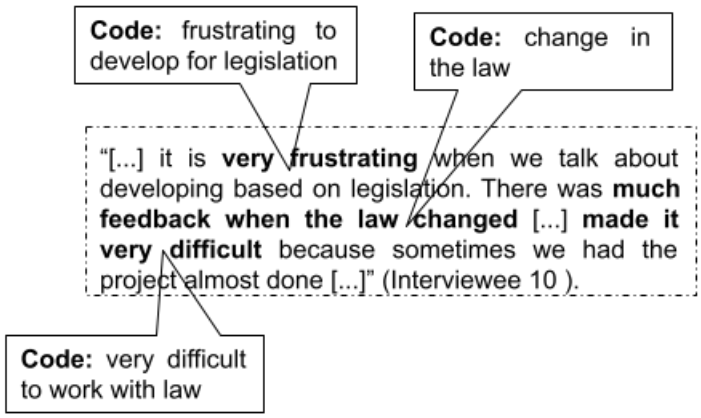
category) which can influence positively (+), i.e., corroborate, or negatively (-), i.e., oppose the factors. Each factor's contribution (positive or negative) over the categories was derived from interpreting the interviews and data analysis performed by the authors. Identifying these factors contributes to elaborating effective methods for reducing ambiguity in privacy requirements specification and achieving legal compliance with regulations.

Below we have an excerpt from an interview interpreted as a negative influence (we highlighted, in bold, the terms that caught our attention when we were coding).

> "[...] it is **very frustrating** when we talk about developing based on legislation. There was **much feedback when the law changed** [...] **made it very difficult** because sometimes we had the project almost done and **had to go back and adapt to attend law** "(Interviewee 10).

The analysis of this excerpt with the other interviews, using the constant comparison method, can create a factor that negatively impacts a category. In this case, Figure 5 shows the process for identifying factor negative **(2) Constant changes in the law make legal compliance difficult**. Similarly, a factor with positive influence appears.

Figure 5 – Process for identifying a factor with negative influence



Source: The author (2021).

Lastly, in selective coding, these codes, from axial coding, were related in categories. The goal of selective coding is to integrate the different categories that have been developed, elaborated, and mutually related during axial coding into one cohesive theory (VOLLSTEDT; REZAT, 2019). Thus, selective coding chooses the core category and relates it with the other categories from axial coding. The core category described "the central phenomenon around which all the other categories are integrated" (CORBIN; STRAUSS, 2014).

The core is the category that accounts for a large portion of the variation in a pattern of behavior and is considered the central theme or main concern or problem for the participants (GLASER, 1978). Some criteria for choosing a category as the core: it must be central and related to several other categories and their properties; it must frequently reoccur in the data; it relates meaningfully and quickly with other categories (GLASER, 1978). Having detected the core category, the researcher knows the central phenomenon of his/her research and can finally answer the research question. We found the two categories that passed all the criteria for the core: "**Specialized Support Area**," and a "**Communication between Development Team Members**." Figure 6 shows the core categories and their relationship.

The "**Specialized Support Area**" assists both in **Requirements Elicitation** and **Requirements Specification**. The company has professionals with specialized skills that support projects, **Reducing Ambiguity** and helps with **Achieving Legal Compliance**.

"**Communication between Development Team Members**" assists both in **Requirements Elicitation** and **Requirements Specification**, and helps in **Reducing Ambiguity**. "**Communication between Development Team Members**" and the team's expertise for **Working with Data Protection Regulations** assist **Achieving Legal Compliance**. A **Requirements Specification** with **Reducing Ambiguity**, in turn, favors **Achieving Legal Compliance**.

4.5 THREATS TO VALIDITY

We classify the threats according to the categories defined by Runeson et al. (RUNESON; HÖST, 2009). For the **Construct Validity**, we established rigorous planning and protocols for data collection and analysis, as suggested by Runeson et al. (RUNESON; HÖST, 2009). Additionally, we carefully designed the interview script and performed a pilot interview with a public company professional with extensive software development experience, whose answers were later discarded. Besides, it is essential to mention that we carried out this study with Brazilian software companies, where Portuguese is the first language. Thus, translated the participants' quotes were reproduced in this document.

Also, the Consent Term informs the subjects that participation should be voluntary and withdrawal is possible at any time. The data they provided is confidential, anonymized, and aggregated with the other interviews, i.e., they will not be analyzed individually, not allowing the identification of research participants so that the subjects could freely share their practical

experiences.

Concerning **Internal validity**, we used maximum variation sampling to have good coverage, background, and different views regarding the requirements elicitation and specification activities for legal requirements and legal compliance. The interviewees were utterly free to present a project, and the interviewer did not influence this decision. Finally, the presented projects were of different sizes and types, and the interviewees had different backgrounds (see Table 2). All the interviews were analyzed independently to avoid researchers' bias in the coding process. Following this individual analysis, the two researchers discussed each code and its content and refined the coding structure to reduce overlaps. Moreover, they refined aspects to identify and eliminate any individual biases. Furthermore, we check the categories from the data gathered to confirm that none of them refuted any conclusions.

External validity Due to interviews carried out in Brazilian software companies with different interviewers' backgrounds, company size, and characteristics, interviewed could have used certain practices in a specific company and not in others. Therefore, it will not be easy to replicate the study. Considering the number of respondents (with different expertise) and companies that operate in diversified domains, the study results can be generalized with a certain degree of confidence.

Reliability In qualitative research, the data analysis consists of interpretation and coding of excerpts from the interviews. Therefore, the codes would probably be partly different with a different set of interviewees. To increase the reliability, we realized a review of the findings by another researcher, maximum variation sampling to have good coverage (ROBSON, 2002), and get different views regarding the ambiguity in privacy requirements specification and legal compliance. We use a non-probabilistic sample to increase the potential for generalization of the findings and reduce bias in selection. Conducted the interviews at different companies, and each interview happened in only one work session, thus avoiding bias through subjects discussing the interview amongst themselves.

When conducting sufficient interviews, one criterion is "saturation," i.e., when no new information or viewpoint is gained from new subjects (CORBIN; STRAUSS, 2014). Dey (DEY, 2003) states that instead of theoretical saturation, it is better to guarantee that categories are consistently built from the data. Therefore, saturation is closely related to the notion of theoretical sampling — the idea that sampling is guided by the necessary similarities and contrasts required by the emerging theory (DEY, 1999). At least two researchers analyze all the interviews and the coding process in this research. This study interviewed 22 participants with

different backgrounds, viewpoints, and perceptions about reducing ambiguity in legal requirements specification compliant with the data protection law. Therefore, although theoretical saturation was not tested, we are confident that we achieve the theoretical sufficiency to apply the findings to other situations.

To support replication and validation by independent researchers, we are making available the protocol and all coding results to answer the research questions (available in Appendix D). The following section presents the results and analysis for the four research questions.

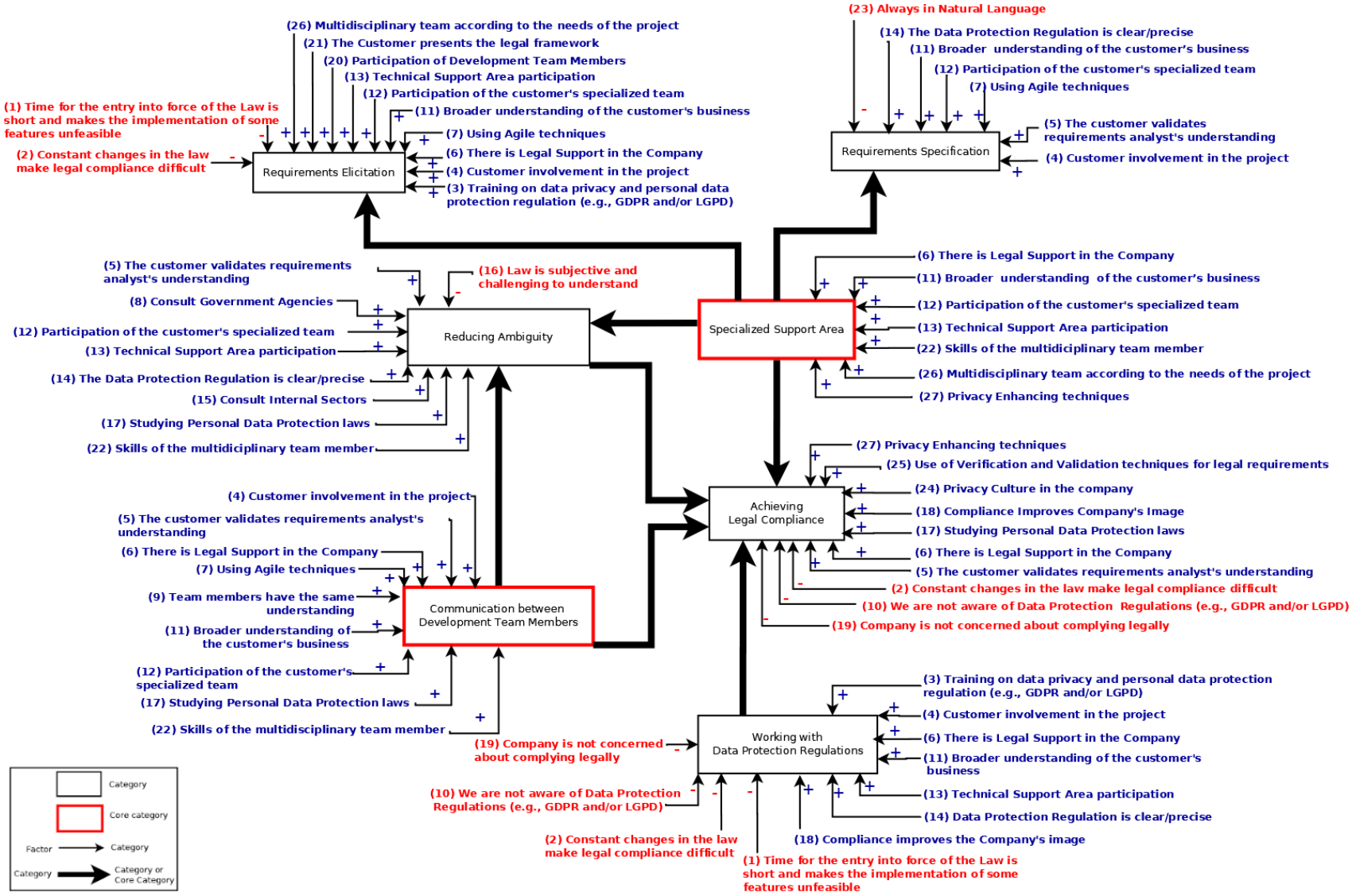
4.6 RESULTS AND ANALYSIS

We have followed a narrative style integrating quasi-quotes from the interviewees in the general explanations to present the results. These quasi-quotes mean syntactical adaptations of the sentences to make them fit the story (e.g., including the missing context in the sentence, aligning verb tenses, and others) (GUTIÉRREZ; BONACHE; QUER, 2020).These quasi-quotes include the identifier of the subjects between curly braces.

The model for public and private companies presented in Figure 6 show categories (represented by rectangles) and explains the factors that affect positively [+], i.e., corroborate, or negatively [-], i.e., oppose the factors, how Brazilian IT companies address ambiguity in privacy requirements specification and how they achieve legal compliance. To facilitate their location and reference in Fig. 6, they are bold and numbered. The larger arrows connecting only categories represent that the related categories can influence each other. The rectangle with the red border highlights the core categories. We found the two categories that passed all the criteria for the core were "**Specialized Support Area**" and a "**Communication between Development Team Members**."

The "**Specialized Support Area**" assists both in **Requirements Elicitation** and **Requirements Specification**. The Company has professionals with skills specialized that support projects, **Reducing Ambiguity** and helps with **Achieving Legal Compliance**. "**Communication between Development Team Members**" assists both in **Requirements Elicitation** and **Requirements Specification**, and helps in **Reducing Ambiguity**. "**Communication between Development Team Members**" and the team's expertise for **Working with Data Protection Regulations** assist **Achieving Legal Compliance**. A **Requirements Specification** with **Reducing Ambiguity**, in turn, favors **Achieving Legal Compliance**.

Figure 6 – Core categories and its Relationships



Source: The author (2021).

We also found some secondary factors (represented as a statement with an arrow to a category) which can influence positively (+), i.e., corroborate, or negatively (-), i.e., oppose. Figures 7 and 8 show a set of factors influencing positively (represented in blue) and negatively (represented in red) the categories that emerged from interview data. They are bold and numbered to facilitate their location and reference in the text.These factors appear for each interviewee's skills, capabilities, and experiences to handle ambiguity in privacy requirements specification compliant with the law. We categorize these factors in project-related (see Figure 7), dealing with the culture organizational, personal experience, and technical expertise (see Figure 8).

Figure 7 – Project-related Factors that influence categories

Source: The author (2021).

The following subsection presents in Table 2 Public and Private Companies' characterization.

4.6.1 Characterization of Public and Private Companies

Brazilian IT public companies have certain peculiarities because their units are spread over several cities or states. These companies serve public customers, have their own data center infrastructure, use a consolidated development process based on SCRUM to develop new products, or carry out major evolutionary maintenance, which has a durability of more than three months. Own separate development units and technology standards to ensure software

Figure 8 – Organizational, Personal, and Technical Factors that are influencing categories

Source: The author (2021).

developed at any development unit has the same quality. Eight respondents from six different companies are from Public Companies (see Table 2).

In this context, in general, all projects are aligned with legislation, provisional measure [2], or the law. Thus the company's software development processes are published for society as a whole, and there is a repository internally with standards, laws, articles, or any artifact related to customer demand. Articles 23 to 27 of the LGPD (BRASIL, 2018b) deal with the processing of personal data by the Public Authorities.

IT Private companies have certain peculiarities. These companies have several customers (tax, health, legal, education, among others domains), its development process is based on Agile Methods. All companies adhere to Agile Scrum for large-scope projects or new systems development. Correction demands projects adhere to the Waterfall method. Fourteen

[2] A provisional measure is a legal act in Brazil through which the President of Brazil can enact laws without approval by the National Congress. There are two requirements for a provisional measure to be used: urgency and relevance of the matter regulated.

respondents from twelve different companies are from Private Companies (see Table 2).

According to the interviewees, legal requirements are those that the legislation discusses and states how to treat information and the functional requirements that the system must-have. There is no specific guideline for treating legal requirements; they use general guidelines about the software development process.

The following subsections present the synthesis of the respondent's answers to the four research questions, shown in Section 4.2) in the context of public and private software development companies. Presents in the following sections the factors that emerged from the data from relationships between categories.

4.6.2 How do Companies perform Legal Requirements Elicitation?

The software development project starts when the client sends a demand to the Company. Then, the Business Analyst raises the Client's needs at a very high level. Many respondents reported that the source of legal requirements is legislation, manuals, and regulations that determine the mechanics of operation, interpretation of specific legislation for each domain, and interaction with the Customer.

The Requirements elicitation session occurs, in all interviewed companies, through interviews with Customers to identify needs and understand the Company or department's routines that the software product to be developed will be executed. Wagner et al. (WAGNER et al., 2019) indicate the interview as the most popular approach to capture requirements.

In addition to interviews, Company C03 carries out some ethnographic activities: " *One has to go to the user's environment, spending a day with him doing it, basically trying to see what he does and how he does it is an observation*" [Interviewee 04]. The creation of prototypes is a strategy widely used, and its validation by the Customer, as stated by Interviewee 15," *That outline was concrete for the Client. In the end, we managed to have a very high level of assertiveness*".

We cannot affirm that using these techniques positively or negatively influences legal requirements' elicitation, as we do not verify if the techniques are used correctly. We only present the techniques mentioned by the interviewees.

Some companies use other techniques, such as workshops, document analysis, meeting records, and as stated by Interviewee 18: "*We set up a template to facilitate elicitation with some key questions, so let us assume that it could guide the requirements survey*."

Most interviewees' companies work **[7] Using agile techniques**. Roles on the agile team (Product Owner (Product Owner (PO)), Development Team, Team Lead) plan the sprints for each set of items in the Product Backlog that will be developed. Therefore, closer contact with the Client (**[4] Customer involvement in the project**) is necessary because **[21] The Customer presents the legal framework** related to its area of activity (laws, norms, standards, among others), and there is the breakdown of the requirements for the sprint.

The elicitation of legal requirements has its particularities to software requirements. It demands knowledge of those involved in the project concerning the laws that regulate the software's domain to be developed. Thus, Company must provide **[3] Training on data privacy and personal data protection regulations (e.g., GDPR and/or LGPD)**, to stimulate the requirements analyst, business analyst, or other participants in the elicitation session about privacy and personal data protection (through lectures, events, courses, among others). **[3] Training on data privacy and personal data protection regulation (e.g., GDPR and/or LGPD)** is also a personal factor, as the initiative to take courses related to privacy and data protection may come from the employee. Training on data privacy or personal data protection is essential to assist the Customer identify the software's legal requirements.

Another factor that positively influences the **Requirements Elicitation** category has **[26] Multidisciplinary team according to the project's needs** with members with **[11] Broader understanding of the Customer's business**.The requirements elicitation and specification steps make it easy when the domain is known.

Conversely, an unknown customer domain negatively influences the requirements elicitation category, and it can cause interpretation difficulties. However, this negative factor can be resolved if there is the support of team members from the specialized support areas in the requirements elicitation sections (**[6] There is legal support in the Company**, **[12] Participation of the Customer's specialized team**, and **[13] Technical Support Area participation**), which positively influence the **Requirements Elicitation** category.

Analyzing these excerpts from the interviews, we identified that no support from the Specialized Legal Area at the company could negatively influence eliciting legal requirements. As such, **[6] There is a Legal Support in the company** helps understand and interpret specific legal terms, reducing ambiguous software requirements specifications.

The **[20] Participation of Development Team Members** in the requirements elicitation session positively influences the category **Requirements Elicitation** because they give examples and, based on practical cases, raise business rules and the first needs. "*Thus, we al-*

ready have a good outline of our backlog to be estimated and placed in the project structure", as Interviewee 09 cites.

[4] Customer involvement in the project is a facilitator and success criterion for carrying out **Requirements Elicitation** of legal requirements for employees **Working with Data Protection Regulations** in regulated project environments.

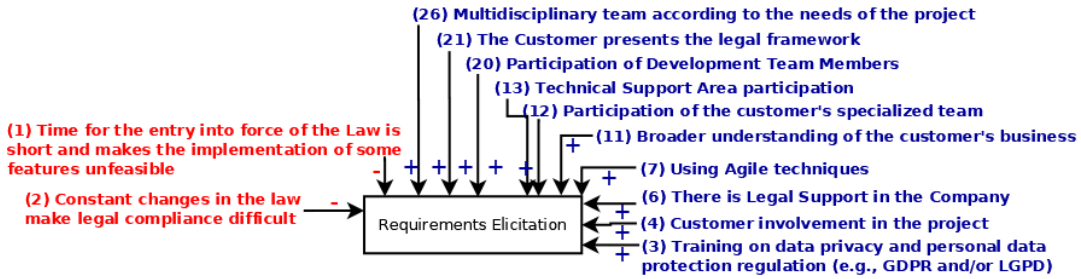
The factor **[1] the time to entry into force of the law is short and makes the implementation of some features unfeasible** influences the **Requirements Elicitation** negatively. Interviewee 04 states *"sometimes the law has a concise date to come into force, some requirements should be in the software do not go because of the development time"*.

Another factor that negatively influences the elicitation of requirements is the **[2] Constant changes in the law make legal compliance difficult**. According to Interviewee 06 *"if I develop software for six months, this law can change. Can happen. We read, scrutinized, detailed, studied a lot, understood several things, and finished the development. The legislation is changing*".

Consequently, to avoid interpretation problems and noncompliance, they try to be as conservative as possible or as conservative as possible in interpretation, not to make the development of the product unfeasible.

Figure 9 shows the factors that influence the Requirements Elicitation category.

Figure 9 – Factors influencing the Requirements Elicitation Categories



Source: The author (2021).

4.6.2.1 *Specialized Support Area category*

The Legal Requirements have a particularity related to the requirement of in-depth knowledge about the domain. All interviewed report that Clients allocate a specialist or a specialized team to accompany the project (that we call the **[12] Participation of the Customer's specialized team**). The team members are Legal Experts, Business Analysts, Domain Experts,

Auditors, and other stakeholders. It functions as the facilitator in understanding the Client's domain, mediating the discussion with the Business Analyst and Requirements Analyst. The Customer's specialized team present the legal basis, assisting in interpreting the legal privacy requirements, monitoring the software development life cycle as a whole, conducting the legal compliance analysis, and certifying that the software complied with the legislation and was delivered as specified. Some companies reaffirmed the importance of **[12] Participation of the Customer's specialized team**:

> "*The customer's team has to be responsible for giving us regulations and standards that regulate his business area, the data flow that they will use. Furthermore, after that, we go through an analysis stage; the development team sits down and takes the books on GDPR, takes the internal notes... people who have been working longer, and starts the discussion*" [Interviewee 03].

The first thing to meet a legal requirement is to map the Data Flow Diagram (DFD) to understand what will happen to people's data. Detailed mapping of this data is required to identify which sectors can access this data. Data flow mapping was vital in any compliance attempt. As states Sirur, Nurse e Webb (SIRUR; NURSE; WEBB, 2018) without understanding where their data was transmitted and stored, organizations felt they could not hope to have enough control over their data to protect it. While this was feasible for more essential or more data protection-focused companies, this was a highly challenging task due to the overhead involved in mapping out complex webs of data networks.

> *"To analyze that a given data passes through several sectors of the Company. In this process, we need to integrate Legal, Information Technology, and Governance. The legal to give the legal interpretation, the governance to explain what the internal procedures (business rules) are and, then the IT translates into the computational language*" [Interviewee 07].

As public companies deal with the product of the same Customers and with similar businesses, they acquire knowledge and **[11] Broader understanding of the Customer's business**. Respondents 09 states: "*with our practical expertise, the development teams become a little bit specialized in that business because the context, the very words he uses, the more particular business needs are also in our domain. We understand a lot of that business*".

The formation of **[26] Multidisciplinary team according to the project's needs** is crucial for eliciting legal requirements. The **[22] Skills of the multidisciplinary team member**, diverse knowledge, and versatile characteristics (specify, code, testing, database, and user interface skills), or as Interviewee 05 quoted, "*wearing different hats at certain times*." The roles participating are the most diverse: Client Project Management, Technical Leader (also named, Focal Point), Service Management Analyst (from the perspective of the development company is the person who best knows that service), Product Owner (PO), Scrum Master, developers, analysts (requirements, integration, and testing), software designers, and architects. In addition to these roles, Business Analysts specialized in the field (tax, health, Legal, education, among others).

In the software development Company, the **[13] Technical Support Area participation**, which is not part of the development team but provides consultancy (software architects, information security, project managers, lawyers, tests, among others). Its role is to help the team overcome a complex problem and transfer their skills to one or more developers on the team.

One Public Company has specialists identifying and treating ambiguity, and one Private Company has specialists in data anonymity. Support from technical experts has a compliance characteristic regarding team members' work, project adherence, and execution against established company standards, identifies vulnerabilities, points that may violate the law, protects sensitive personal data, or any other identified risk.

In all Public Companies **[6] There is Legal Support in the company** that guides the project involved regarding the developed software's security and legal privacy requirements. However, this support is not within the Development Team, as stated by Interviewee 18 "*We only adapt to the new laws, but there is no legal person in the development*." The Business Analyst, along with the Legal Support Area, will make a deep understanding of this legislation and support the Requirements Analyst and Development Team in discussion with the customer about how these legal requirements translate into technical solutions. Together, they will understand the law and extract the details that need to be in the system, both functional software and work process requirements.

> "*We take this law, do the first job of reading, and, together with the Client, we identify in the system what these points are, which must be changed and, must be changed in this way, by our law's understanding*" [Interviewee 05].

We identified two scenarios in these private companies related to factor **[6] There is Legal Support in the Company**. The legal sector is part of the software project team [Interviewee 12] or working together only in specific situations [Interviewee 10]. One of the companies has a legal person dedicated to privacy, bringing the Development Team and the Legal Sector closer. We believe that this is the ideal composition of a Team when it comes to Legal Requirements.

> *"Today, there is a person in the Legal sector dedicated to privacy. We hope that he will read and understand the law, explain and pass on this knowledge to people who need to act on the adequacy part, and that act as a consultant to help us get into compliance"* [Interviewee 12].

> *"We have a person of Legal support, a lawyer, that is when it is very critical, and there is a very high risk of us doing something, not in compliance. We trigger it much more as support"* [Interviewee 10].

Moreover, the second scenario in private companies, that there is no support from the Company's Specialized Legal Area [Interviewees 07 and 16].

> "*There is no independent sector for legal software compliance, and there is a legal sector responsible for the compliance. This means that the focus is not 100% on software compliance because of several other day-to-day demands*" [Interviewee 07].

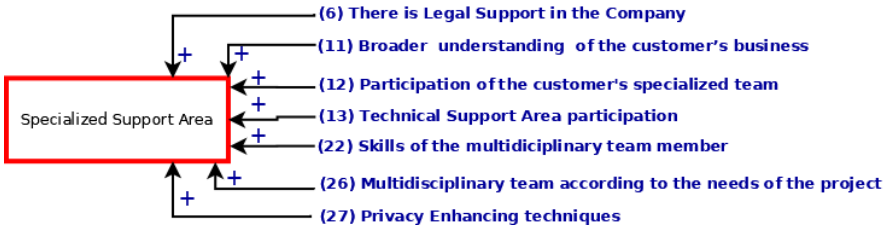
> "*Today, the Legal department is more focused on another activity line. It turns out that people in the IT technical area have to interpret everything we already seek legal support, but it is never satisfactory*" [Interviewee 16].

The company of respondents 12 and 13 has an Anonymization sector. This sector is responsible for ensuring the user's privacy, and one of the ways to achieve it is to have the data anonymized. LGPD (BRASIL, 2018b) article 5th indent III defines *anonymized data* as those related to the holder that cannot be identified or by whoever performed the collection or any other person, considering reasonable and available use of technical means at the time of treatment. Some **[27] Privacy-enhancing techniques** mentioned by the interviewees are encryption, hash, differential privacy, among others.

> *We are trying to anonymize all products, and we are doing "privacy weeks," which is, like, every quarter, we take a week for everyone to focus on privacy, try to find privacy solutions for the company* [Interviewee 12].

Having a **[11] Broader understanding of the Customer's business**, the **[12] Participation of the Customer's specialized team** and **[13] Technical Support Area participation** (Anonymization Team, Compliance Area, and has a Legal Person dedicated to privacy) positively influences the Specialized Support Area category (see Figure 10).

Figure 10 – Factors influencing the Specialized Support Area category



Source: The author (2021).

*4.6.2.2 **Communication between Development Team Members category***

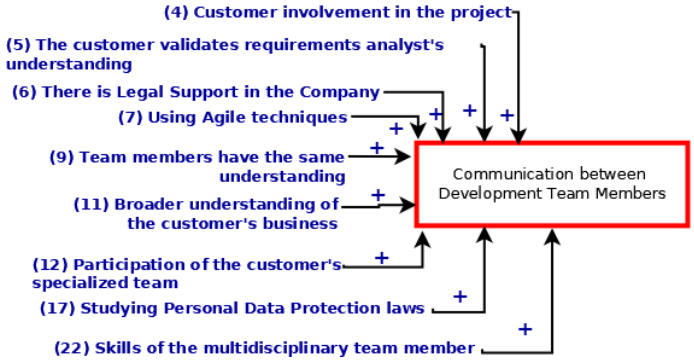
Communication between all project members is one of the most cited factors to understand ambiguous legal requirements and better specify legal requirements (see Figure 11). The **[12] Participation of the Customer's Specialized Team** or **(5) The customer validates the requirements analyst's understanding**. Another factor that positively influences **Communication between Development Team Members** category are **[11] Broader understanding of the Customer's business**, and **[22] Skills of each multidisciplinary team member**.

Nevertheless, the Requirements Analyst and Team member first meet to discuss and develop a shared understanding inside the team. It is essential that **[9] Team members have the same understanding**. If it is impossible to understand within the team or between teams working on similar needs to mature the understanding, but if the decision still seems ambiguous, the Project Manager or Business Analyst will contact the Customer (**[4] Customer involvement in the project**), or Customer's Specialized Team (**[12] Participation of Customer's specialized team**) or **(6) There is Legal Support in the Company** to clarify the doubt. Nevertheless, present the problem and some suggestions for a solution, as the customer expects this position from the company.

> *We try to talk to the other teams. Because they have a better understanding than the other, they try to mature understanding between us* [Interviewee 02].

The agile process itself predicts that, throughout the sprints, backlog refinement meetings, called grooming, will be held to get into the higher level of detail of what will be in the next sprint. The most valid is the team's understanding of what we want to do [Interviewee 05].

Figure 11 – Factors influencing the Communication between Development Team Members category



Source: The author (2021).

The fact that the software development process, in almost all companies interviewed, is based on Agile Methods (**[7] Using Agile techniques**). They make weekly meetings, and they support the Company's legal area. *"With each sprint demonstration, the team has a vision of what they did not fully understand or misunderstood. It is not something that prolongs and reaches the end of the development process*" [Interviewee 05]. Nevertheless, agile methods also make Customers more open to participating in the process. [**5] The Customer validates the requirement analyst understanding**. Suppose the Customer does not validate what was delivered because it is not what he expected; until the resolved ambiguity it does not execute, the feature returns, budgeted again and entered as an activity in the next sprint.

The origin was a very high-level Legal requirement, and I think there was a natural maturation of the understanding on the part of the Client and, as at the time, we did not follow the agile methods, as he matured the idea, we did not interact much with him. We gave him the solution, and the solution was not adequate [Interviewee 15].

During the development, if any doubts arise because the information is ambiguous or written so that it does not complete the possible flows that functionality, the first thing to do is talk to an experienced team member to understand better. Usually, some people have contacted the Customer during the survey of the product's first view within the team. These people remain as a specific reference to answer questions about that business. This strategy is

similar to that presented by Kamsties and Berry (BERRY; KAMSTIES, 2004) “is communicating an interpretation back to the author of the requirements, after which she can easily point out misinterpretations".

> *So if the developer is going to implement something and generated some doubt, he starts from an artifact there, from a rule written, suddenly from a mind map that was raised with the Client if he has a doubt, the first thing is to contact with the team member who created that artifact he is taking as a base to build* [Interviewee 09].

Usually, when Team Project members need some understanding from the Legal Text, the people involved read the law itself and talk with the customer. There are a deepening understanding and **[17] Studying personal data protection laws**. The best way to remedy ambiguity is to communicate with someone knowledgeable in the business, between team members, a team that works on similar demands, or the Client himself. **[12] Participation of the Customer's specialized team** positively influences the reduction of ambiguity in specifying requirements and achieving legal compliance.

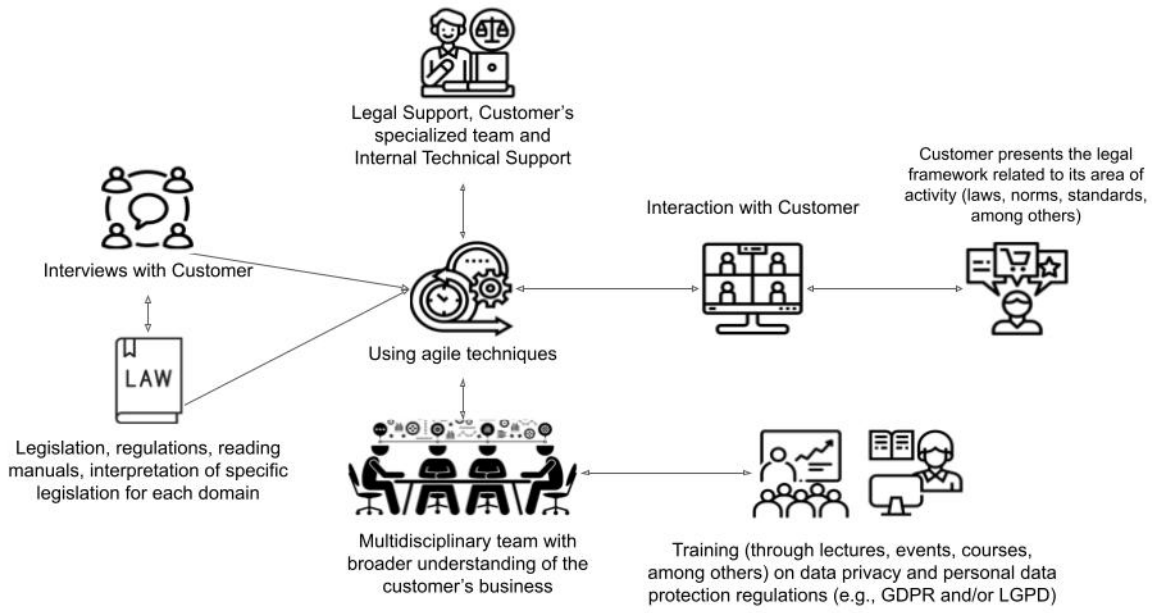
> *“Moreover, ask him about some points of the law that are ambiguous, or difficult to understand, or know in the IT area*" [Interviewee 18].

> *“The best way for us to overcome this problem of knowing if it is aligned with the Client's needs is to have the most frequent contact with him*" [Interviewee 05].

We can summarize the process Requirements Elicitation in the following steps, presented in Figure 12. The main techniques used are the interview with the Client and the analysis of legislation to identify legal requirements. An essential factor that appears is communication. Therefore, agile techniques (as Sprint Planning Meeting, Daily Meetings, or others) favor communication between stakeholders in the requirements elicitation process. The specialized support area, be it legal support or the Customer's specialized team or a technical support area (which are people who are not part of the development team. For example, an information security specialist, a software architect, among others), support the process as consultants to clarify any doubts that may arise during the elicitation of requirements. Teams must be multidisciplinary and possess knowledge of the Client's domain; the interviewees presented this as a successful strategy to elicit legal requirements. If necessary, the Company or the employee can seek training on privacy and data protection laws to improve their skills. Frequent

contact with the Customer is essential. Therefore, frequent meetings must be held to clarify the understanding of the requirements, and any possible ambiguities can be reduced and resolved as soon as possible. It is essential because the Client presents the regulations related to his business domain. Then, based on these steps, we identify how companies carry out the elicitation of legal requirements.

Figure 12 – How do Companies perform Legal Requirements Elicitation?



Source: The author (2021).

The following subsection presents the synthesis of the respondent's answers to the second research questions, shown in Section 4.2), in the context of public and private software development companies.

4.6.3 How do Companies deal with inherent Legal Requirements ambiguity?

Respondents cite that when **[16] Law is subjective and challenging to understand**, the procedures for reducing ambiguity in legal requirements vary among Companies: consult the customer, **[8] Consult Government Agencies** and **[15] Consult Internal Sectors**, case law, request the **[12] Participation of the Customer's Specialized Team** to try to resolve the identified ambiguity, and cases of penalties for companies that failed to comply with the GDPR, for example. Others bring together the **[13] Technical Support Area participation**

or experienced team members to reach a consensus on understanding the identified ambiguity ([22] Skills of the multidisciplinary team member).

In two Public Companies and one Private Company, three areas discuss the interpretation of a legal requirement (IT, Business/Governance, and Legal). The Legal (which interprets the law) and Business/Governance (which explains what the procedures are and how to operate them) need to reach an understanding and, therefore, present to the IT sector to "translate" the legal requirement into computational language. The different points of view are presented, how this legal requirement will impact each area, a risk assessment and analysis of the state of the art techniques is carried out. The articulation of actions will give more protection to the customer product development.

The **Reducing Ambiguity** category presents considerable divergences between public and private companies when it comes to LGPD, for example. Analyzing the excerpts of Interviewee 22, from a Public Company, mentions that *"the law is extremely generalist, it does not objectively say what the requirements are"*. We analyzed excerpts similar to this one and categorized them in the factor **[16] Law is subjective and challenging to understand**. Already, Respondent 21, an employee of a private company, cites *"LGPD is very didactic. I think of it as a step-by-step*". Similar excerpts were categorized under the factor **[14] Data protection regulation is clear/precise**. Such divergences are due to several factors such as a background in the law, position at the company, experience in projects related to legal requirements and the interpretation of the law may be different for each individual. Therefore, we have two factors that influence the Reducing Ambiguity category: one positively **[14]**, and the other negatively **[16]**.

When Team members are unaware of the law, they need **[17] Studying personal data protection laws**, making on a Workshop, and sometimes specific training to have the same understanding.

> "*The Customer presents the Laws and standards that regulate your business area. We read the laws, do the workshop, sometimes, specific training. Sometimes, we have to specialize in legal language to collect laws and resolutions for that particular business.* " [Interviewee 04].

When the law comes from a Federal Agency, for example, which is much larger than the Company, the customer does not feel comfortable giving an understanding of the law, companies **[8] Consult Government Agencies** and **[15] Consult Internal Sectors**.

> *It was full of ambiguity and left much to the managers to decide there as claimed that it was very ambiguous, that was leaving much to the people to decide, then the Ministry of Planning went there and rewrote, put a new Normative Instruction where it was much more rigid, objective, descriptive and left no space for ambiguity* [Interviewee 04].

LGPD (BRASIL, 2018b), in its chapter X Section I, defines the Brazilian National Data Protection Authority (in Portuguese, Autoridade Nacional de Proteção de Dados - Autoridade Nacional de Proteção de Dados (ANPD)), which until the date of realization of this interviews is not acting. One interviewee pointed out that he could consult the ANPD to obtain guidance for the interpretation of ambiguity. The ANPD will function as an Information Commissioner's Office (Information Commissioner's Office (ICO)) [3], for example, is the United Kingdom (UK)'s independent authority, set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

> *Another way we would do that we do not do today is if the ANPD had been consolidated. So, if it already existed, we could consult it to remove this doubt about the law's ambiguity, and, at this moment, we would have a greater certainty that we are doing it right* [Interviewee 12].

> *Opened the Brazilian National Health Agency (NHA) to clarify these doubts. The channel with Brazilian National Health Agency (NHA) is complicated, so we started to contact the operators, look for what the NR (Normative Resolutions) said, start to discuss with the operational members* [Interviewee 16].

The public Company C01 has a team for Ambiguity Analysis. That is not involved in demand, does not interact with the Customer, does not engage with the context. This team analyzes the description of the requirements produced to identify if it is objective, precise, and does not give rise to double interpretation. If it is not satisfactory, ask the responsible for making the necessary corrections. Thus, software documentation produces better quality and decreases rework and risks not correctly meeting what the Customer requests- contributing to strengthening the Company's image (i.e., gain credibility with your Customers, competitive advantage over the competition, attracting investors, and investments).

The **[13] Technical Support Area participation** of the Company in the development process influences the reduction of ambiguity in the specification of requirements and the

[3] https://ico.org.uk/

achievement of legal compliance positively, as stated by Interviewee 01: "*we have an area not linked to my Department, which guides us in Data Protection*." Interviewee 18, from a public company, cites: " have the support of a Legal Sector within the Company, but the rules demanded the project have already passed the Legal Sector, so we make it fit when it comes to us".

In private companies, respondents cite that support for the technical area for reducing ambiguity happens in several ways during the software development process:

> *"Although I work very hard in this field, I have people on my team responsible for making all this interpretation and translating what the legislation requires into a requirement for developers to code"* [Interviewee 20].
>
> *There is a second team, besides mine, specialized in scoring the security of applications within the company* [Interviewee 03].
>
> *Currently, we have this anonymization team and the Legal team. So, together we define these privacy parameters* [Interviewee 12].

The Legal Sector members act as consultants. So whenever there is a failure to understand a particular item, and it has a legal or operational nature, usually hold a meeting. This legal sector positions itself about ambiguity or difficulty and presents its understanding.

Interviewee 21 cited cases where the interpretation of ambiguous sections of the legislation caused sensitive personal data from users to be collected and sent to partners for marketing purposes. These facts were analyzed by the Legal sector, which suggested changes in collecting and sharing with partners based on LGPD. Furthermore, notified to the partner Company to send its commitment to safeguarding the data, which is the motivation for use and all the guiding principles that must be observed as a minimum requirement for an excellent personal data processing activity, as established in I to the X items of LGPD Section 6 (BRASIL, 2018b).

> *We try to understand the principles and look at the studies already done in the law. Based on these studies, and it is also on the GDPR case law, for example, in Europe, the law is running. We know that if you treat the data in this way, it is wrong because the law was applied there, and that company was fined* [Interviewee 12].

Respondents 14 and 19 act as consultants in IT projects supporting the Requirements Analysts. They cite the following procedures for reducing ambiguity in legal requirements:

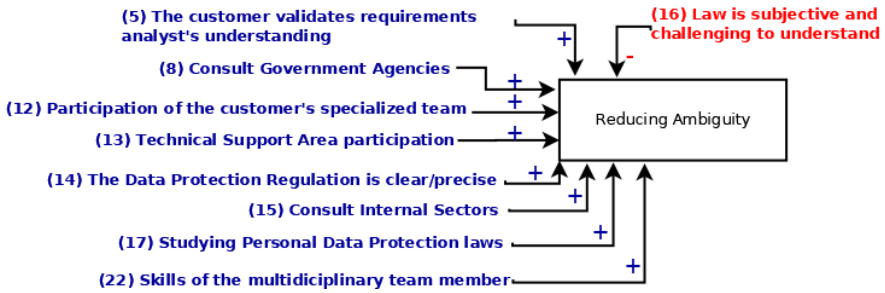
> *Usually, I try to know the real impact that ambiguity that there can have for the Client. I try much more to understand the context of ambiguity to know the possible interpretations. Moreover, together with the Client, measure the level of risk he is exposed to* [Interviewee 14].

> *Do a weekly meeting where they discussed the jobs are coming in and everyone, in this meeting, if the team raised a doubt about the guidance coming from compliance before they start development* [Interviewee 19].

One interviewee mentions no problems with interpretation errors in the Company, as they use the Test-driven Development (Test-driven Developmen (TDD)) methodology. *There is no misinterpretation because we work with TDD before implementation there is acceptance* [Interviewee 18].

Figure 13 shows the factors that influence the Reduced Ambiguity category.

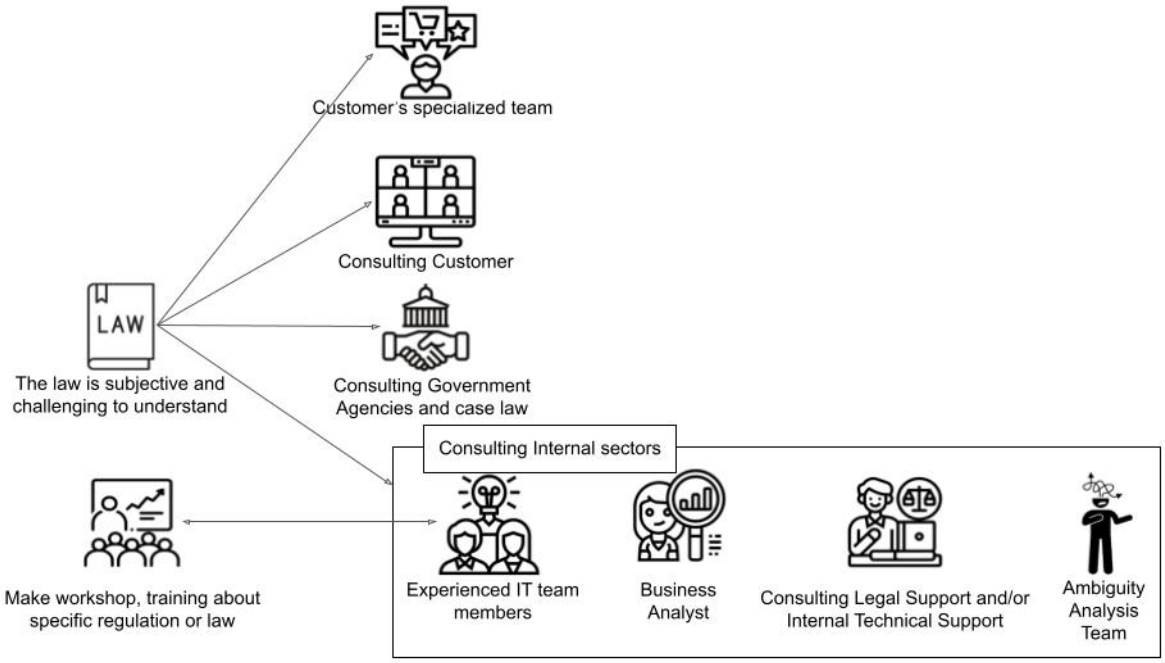
Figure 13 – Factors influencing the Reducing Ambiguity category



Source: The author (2021).

We can summarize the process for Reducing Ambiguity in the following steps, presented in Figure 14. When the law is subjective or difficult to understand, the Requirements Engineer or Business Analyst will consult with those involved in the project to help with understanding. Can consult the Customer's Specialized Team (either a Focal Point or the legal sector, among other stakeholders) or the Customer.Can also consult Government Agencies requesting clarification on the legislation and analyze case law to verify companies that have been sued for any violation of data protection laws and avoid them. Can still consult the Internal sectors and, if necessary, do a workshop, training about specific regulations or law. The Internal Sectors can be the Ambiguity Analysis Team, the Business Analyst, the Legal sector or Internal technical Support, or Experienced IT team members with knowledge of the client's business domain (these cases are more frequent in public companies that operate in projects with related domains).

Figure 14 – How do Companies deal with inherent Legal Requirements ambiguity?



Source: The author (2021).

The following subsection presents the synthesis of the respondent's responses to the third research questions, shown in Section 4.2), in the context of public and private software development companies.

4.6.4 How do companies perform legal requirements specification?

In Brazil, public companies must comply with the Open Data Policy (BRASIL, 2018a), which aims to improve the culture of public transparency. Interviewee 22 states that *"Our main challenge here at the Company is to specify the legal requirements for the data to be public without violating the fundamental right of staff because as a public institution, we have to comply with open data laws*". Therefore, in addition to complying with the Personal Data Protection Law (LGPD), it is necessary to consider other laws that regulate access to information.

Techniques for Requirements Specification in Public Companies include describing use case, user stories, requirements documents, personas, IEEE requirements specification template, the requirement recorded in their tools, **[23] always in natural language**.

> *When it comes to us, the rules, the articles there, everything well pre-established, we follow what is in the user stories* [Interviewee 18].

> *Artifacts written in user stories, business rules, layout specifications, message artifacts* [Interviewee 09].

The techniques used for the Requirements Specification in Private Companies are pretty varied. Private companies mostly follow agile methods and use legal requirements specification techniques: user stories, use cases, own templates, and Lean Inception.

A user story, when improperly defined, can trigger several challenges in agile software development due to incomplete or incorrect documentation (INAYAT et al., 2015). The main difficulties in using user stories to specify requirements are related to sparse detailing of requirements information, difficulty in identifying non-functional requirements, communication and collaboration with users, lack of information for validation of requirements (SOARES et al., 2015). Interviewee 11 states *"We use user stories with acceptance criteria to tell if this functionality is ready. And then, with that, when it comes to development, he takes this user story with the acceptance criteria and is already able to automate his test within the application*". Adaptation of user stories also is used. A private Company specifies user stories with the agile Behavior Driven Development (Behavior Driven Development (BDD)) practices (SOLIS; WANG, 2011). As Interviewee 15 quotes: "*BDD helped a lot in terms of understanding and what we had to interpret*."

Only two private companies use the representation of the expanded use cases, proposed by Phalp et al. (PHALP; VINCENT; COX, 2007) (with main, alternative, and exceptions flows), in natural language to specify legal requirements. However, some common problems, such as ambiguities, incompleteness, and inconsistencies, can arise when describing the requirements through use cases. These problems can cause difficulties in understanding the requirements and, consequently, defects in the software system under development (REGGIO et al., 2018).

A company specifies requirements in its template and registers them in the tool for monitoring the requirements. The Customer approves to the requirements specification before starting the development.

> *We registered the specification in the tool to control Sprints, and that requirement becomes a ticket. Besides, every requirement has an acceptance criterion in real scenarios given a particular input. The system is expected to arrive with that final result. Further-*

more, there is a test scenario for each alternative flow covering that case [Interviewee 20].

People do the grooming [4]*, make the meeting to create the stories, estimate all the stories, plan the entire sprint, and work divided into a sprint. Define the weight of each task and place the weights of each story, and each developer will take the tasks and develop them* [Interviewee 17].

One private company respondents mention the use of the agile technique Lean Inception [5] (CAROLI, 2018).

the Client already has a sense of how to solve the problem, already knows what he wants. Then we spend a week with this client translating all his needs into requirements, in this case, user stories, MVP (Minimum Viable Product), which ends up being the project releases, and a job of prioritizing the requirements to enter the MVPs is done [Interviewee 15].

The process of extracting the legal requirements from the law is manual. **When [14] The data protection is clear/precise**, *" when the legislation is well written, it has entirely drawn the flow data, enters such data, and information processed in such a way, it says the processing rule and will come out in the end. So this will become a likely software requirement*" [Interviewee 05]. Another factor contributing to a precise specification and ambiguity reduction is when we have the **[11] Broader understanding of that customer's business** i.e., context is not entirely new.

The Legal snippet in the requirements specification document is used to perform traceability among the Legal snippet against the requirements. *"For each specified legal requirement, there will be a section in the document that says "to meet the requirement of Law X," and it wrote the full reference of the law*" [Interviewee 04].

If there is a disagreement between what is stated in the law with the Stakeholder requirements, communication and trust are essential pillars. If something has come into conflict or a question has arisen, the Customer has to decide. The law needs to be followed as it is written.

[4] This process breaks out customer requirements, acquired from the stakeholders, into specific work for the team to perform in one work cycle. They met between the PO and the Scrum team (Dev. team and Scrum Master) to discuss the Product Backlog (PB). This time-box meeting is an opportunity for the PO to share user stories and new features with the team. It aims to contribute and discuss future work in order to manage, organize and keep the PB updated (RIBEIRO et al., 2018)

[5] Lean Inception methodology is a sequence of collaborative and dynamic activities that, at the end of the process, quickly obtains the Minimum Viable Product (Minimum Viable Product (MVP)) (CAROLI, 2018).

The Requirements Analyst or Development Team has no autonomy to customize the law; only the Customer can do that. *We ask the customer to decide how the Customer wants the system to be implemented: this or that? The Customer always does it. If he wants to distort it, we specify that it was the Client who is validating* [Interviewee 04]. Then, the Requirements Analyst records what the Customer has requested and is aware. If the legal requirement conflicts with other requirements, *" the whole scenario must be analyzed, the representative stakeholders must be identified to analyze the law, identify the particularities and verify the feasibility of the specification of the legal requirement* "[Interviewee 08].

[5] The Customer validates requirement analyst understanding or **[12] Participation of the Customer's Specialized Team** continuously validates the interpretation of Requirements Analyst or others involved in the project for requirements specification. **[7] Using agile techniques** favors verifying and validating the requirements specification during the product development until the final approval.

> *"The Client receives this set of artifacts and has to validate the requirements specification. To seek the commitment there that what we understand is really what the Customer needs. At least from the perspective of that person who is accompanying us*" [Interviewee 01].

> *"the Client's specialized team always verifies the team's interpretation. So, this is the first phase of verification at the Customer level. When we specify this, we show it to the Customer; the Customer says ok. We say: "this is the direction*" [Interviewee 04].

> *"sprint validations are essential to escape surprises and much rework*" [Interviewee 18].

In such cases, the requirements specification activities were not performed to assess whether the software meets its requirements but to evaluate the system's capacity. Some other times, the verification activities were performed based on subjective or imprecise requirements indicating a gap between industry and academic literature because they are not in use despite techniques.

> *We are reading law and interpreting. It would be costly to create something to try to understand the law automatically* [Interviewee 12].

> *Legislator could use a more formal language to describe these laws and some laws' ambiguities. We know a series of tools that solve a series of problems, and we see people*

making the same mistakes repeatedly because they insist on writing in Natural Language [Interviewee 02].

Some projects are based on the Waterfall method. In these cases, the requirements are specified and documented in a spreadsheet or the Jira software [6]. Finally, the specifications' doubts are resolved with the Customer ([4] Customer involvement in the project), and the Customer validates the requirements. Then, the requirements are divided into sub-requirements and assigned to the development team members. Other techniques used in Waterfall projects are use cases and a few high and low-fidelity prototypes.

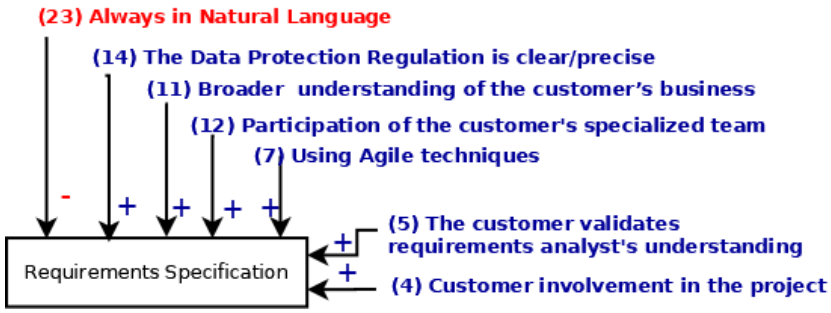
Interviewees from two private companies (C02 e C07) cited that they do not specify requirements:

We receive the requirements already specified. So we take a spreadsheet editor and list the requirements. We try to prioritize it without getting too agile because we do not follow agile. They will always need extra information, as specified. We put all possible observations there on our spreadsheet as well [Interviewee 08].

We do not specify requirements in any document. We organize into epics and register them in a Kanban system, where we can put story cards and have their progress and, on top of what we define tracking. We store everything in Mingle tool to be aware of what a card means [Interviewee 03].

Figure 15 shows the factors that influence the Requirements Specification category.

Figure 15 – Factors influencing the Requirements Specification Categories



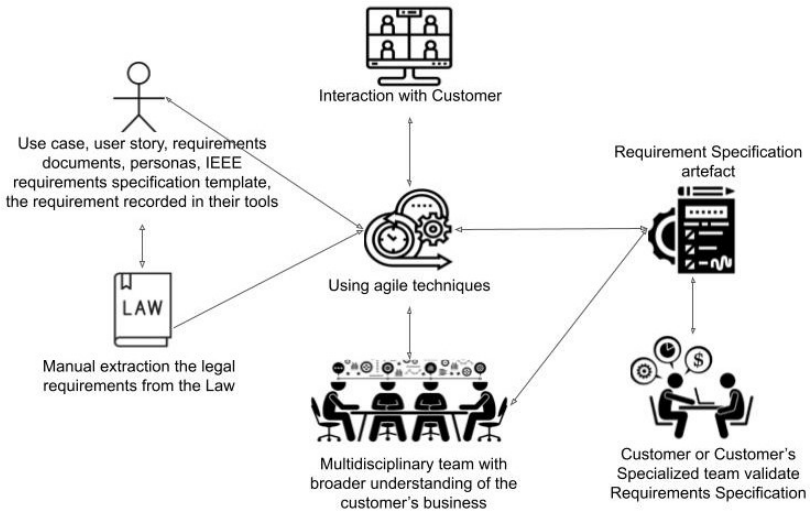
Source: The author (2021).

We can summarize the process for Requirements Specification in the following steps, presented in Figure 16. Regarding requirements specification techniques, the most cited by respondents are use cases, user stories, and templates. Always specified in natural language, and

[6] https://www.atlassian.com/software/jira

extracting these requirements from the legal text is done manually. Therefore, the use of agile techniques is essential to interact with the stakeholders and the client regarding the verification and validation of the specification of legal requirements, supported by a multidisciplinary team. The customer or the customer's specialized team will validate the requirements specification artifact.

Figure 16 – How do companies perform legal requirements specification?



Source: The author (2021).

The following subsection presents the synthesis of the respondent's responses to the fourth research questions, shown in Section 4.2), in the context of public and private software development companies.

4.6.5 How do Companies verify Legal Compliance?

System designers usually are not security or privacy experts or legal experts, so that they may have difficulties in deploying systems that comply with security and privacy requirements as defined in the current legislation (COMPAGNA et al., 2009).

Respondents presented two approaches when they answered the questions. The first deals with the software's legal compliance with data protection laws, such as Lei Geral de Proteção de Dados (LGPD) and GDPR; the second presents the steps for the entire Company to comply with data protection laws. The **[24] Privacy Culture in the company** emerges as a category that encompasses both types of respondents' responses.

"The culture of privacy is essential. It is fundamental. Without the privacy culture, the

> *plan to comply with a data protection law will most likely fail. If it is considered that only the security sector will take care of, it is the responsibility only of the legal sector; it will not work. It must be a natural mechanic—and organizational vision, it is a question of Privacy by Design (PBD) (Privacy by Design)"* [Interviewee 22].

There is still a separation between areas, such as a specific area dealing with privacy and security issues. The law impacts the Company as a whole. Therefore, all sectors need to be aware of privacy and data protection laws, and the legal requirements need to be in line with legislation, as respondents claim **[18] Compliance improves Company image**. Legal compliance must be seen in all organization sectors, and account should be taken of existing processes as implementing a **[24] Privacy Culture in the company** is very important to produce legally compliant software.

> *The compliance process is systemic. It is necessary to analyze all sectors that the data transits within the Institution* [Interviewee 17].

Even with the discussions and GDPR penalties [7] imposed on larger companies by breaching personal data protection laws, some respondents mentioned that **[10] We are not aware of Data Protection Regulations (e.g., GDPR and/or LGPD)**. The fact that some **[19] Company is not concerned about complying legally**, especially with the LGPD, is a factor that negatively influences legal compliance concerning to that **Achieving Legal Compliance** and **Working with Data Protection Regulations**. Thus, there is no broad knowledge of software development individuals on the GDPR and its impact on handling personal and sensitive data. Therefore, employees need to know what personal data is, the processing, the legal bases, and other information related to legal requirements. One way to start the discussion on privacy, protect personal data, and the laws that regulate the Company and the client's domain is through training.

Therefore, all the public companies and three private companies to be LGPD compliant are bringing together security and privacy experts, the Legal sector, and holding lectures, forums, awareness campaigns, good practice guides for implementing the LGPD, and **[3] Training on data privacy and personal data protection regulations (e.g., GDPR and/or LGPD)** for an institutionalized privacy understanding within the Company, making all the employees aware as personal data protection. At the time of the interviews, no one interviewee from public

[7] https://www.enforcementtracker.com/

companies reported that the Company is concerned about GDPR and realized data protection training about GDPR. Additionally, team members need to **[17] Studying personal data protection laws**.

> *"As we studying GDPR, we took courses on GDPR, the whole team. People have experience in handling sensitive data. In a situation where we see a data vulnerability that will be stored and possibly exposed, data is not being justified. We report to the teams that data must either not be stored or has to take a strategy to adhere to the legislation in question*" [Interviewee 03].

Other companies are adapting and will hire an audit firm, as an example, one of the Big Four (Ernst & Young (EY), Deloitte, PricewaterhouseCoopers (PwC), KPMG) to verify that the practices implemented are in compliance with the LGPD and/or GDPR and thus inform their Customers that they have received an "LGPD or GDPR Certification" from an audit firm. Such certification does not exist, but it is as reported by interviewee 12 *"The LGPD has no seal. It has no LGPD certification... What we want at LGPD, initially, is one for an endorsement by a renowned company, a Big Four"*, because **[18] Compliance improves Company image**.

Legal compliance analysis should be done at the early stages of the development process, right after the requirements elicitation phases, because when looking close to delivery, it is challenging to maintain the initial scope.

In the companies participating in the study, legal compliance analysis occurs three times, before the start of development, in the requirements' acceptance phase, with the customer's validation. Some test cases are created in software development companies that use TDD (Test Driven Development). The second level of verification occurs when the functionality has been implemented and performed retesting, the requirement implementation is verified. Moreover, the third level occurs at delivery to the Customer when the legal compliance with the legal framework is verified.

The Customer performs legal compliance verification through the requirements' acceptance and the product itself since it is understood that he is responsible for the legal apparatus's demand and knowledge.

In other cases, Customer demand reaches **[6] There is Legal Support in the Company** that performs the extraction process, sets the rules, and sends to the development team only the necessary adjustments to the system. In this document, there is a mapping of which

requirement meets which part of the legislation. To verify that all legal requirements have been implemented following the law, companies carry out the following actions:

> *We use the requirements document and analyze the requirements specification, whether what was specified there has been implemented.* [Interviewee 11].

> *In other words, based on legislation, or based on criteria that the customer has defined, usability, flow to the system, process, how that legislation will be built within the system* [Interviewee 09].

> *That client is responsible for meeting the demand for her adherence to the Client's normative apparatus because the Company specifically does not make a legal compliance assessment on top of the demand. We trust the Customer's information and accept us* [Interviewee 02].

It is widespread for a law to change and require an adjustment to the system that is in production. The technical factor **[2] Constant changes in the law make legal compliance difficult** is cited by respondents as a challenge when developing systems that need to be in legal compliance. The current laws are very dynamic, complex, and undergo constant variations, so it is essential to follow them to prevent violations at the Federal, State, or Municipal levels. Therefore, this factor negatively influences **Achieving legal compliance**.

> *When there is an external action that will impact the project (an ordinance, a new law, a provisional measure, some reform of some law, something that changed a rule). It happens in a meeting, and then there is a process similar to the initial process. The survey, the decomposition of all the stories' stages, and planning are done again in a more macro way* [Interviewee 08].

However, **[2] Constant changes in the law make legal compliance difficult** becomes a usual practice for those involved in the project because *"This is a common characteristic of what we deal with here since we are dealing with public Clients"* [Interviewee 02].

Broadly, the steps for verifying Legal Compliance in IT Companies consist of analyzing whether the Company meets the legislation of data protection and legal requirements, recording the result of this analysis, and developing an action plan to comply with legislation. One IT Lawyer respondent cited that their customers consult them only when the product is about to enter into a commercial transaction, the company will be sold, or the company will receive some investment. Never when they are in the early stages of the development process.

> *"within what he explained to me in operation, be careful, and what we can do to mitigate this risk... is infrequent, not to mention almost non-existent, that when the Company is going to develop the system, come and talk to you to assess the suitability of the software for legislation. It is never "I want to comply with everything." We mitigate risks. Alternatively, what you are doing in compliance with a legal obligation"* [Interviewee 14].

This process is very similar to an audit and lasts three months to one year. Respondent 19 detail the legal compliance verification processes they participated in:

> *It has a gigantic list of various documents and information that we need: tax, intellectual property, data protection, and the entrepreneur will give all this information. It is not a simple thing. It is infrequent - looking back, if it is all right. If the software meets the tax requirements, intellectual property, data protection, it is not normal* [Interviewee 19].

The steps taken to mitigate risks in Legal Requirements Compliance from Personal Data Protection Law are mapping the flows and processes of receiving, sending, extracting, and storing data.

> *"In my first conversation, I try to understand its entire operation to identify if the treatment is given to that data, from the collection, storage, processing, and enrichment if it is done according to LGPD. So I will raise the aspects that he should, within what he explained to me in operation, be careful, and what we can do to mitigate his risk"* [Interviewee 14].

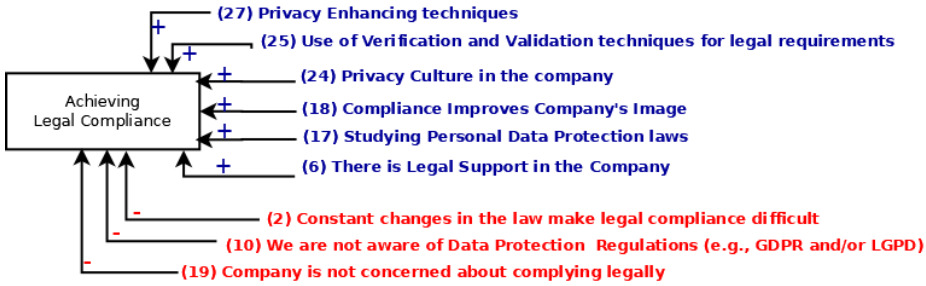
> *"We analyzed the entire cycle, from the moment of how the company trains developers for safe development until post-production after it went into production, the security incidents, as they are being reported. So this was previously defined by the compliance team, and he passed on this information, and, from there, they carried out the work"* [Interviewee 19].

Only **25] Use of Verification and Validation techniques for legal requirements** do not guarantee software privacy requirements specification and legal compliance. It is necessary **[11] Broader understanding of the Customer's business**, and to use **[27] Privacy-enhancing techniques** and guidelines; developers should receive formal education on privacy practices. Esayas and Mahler (ESAYAS; MAHLER, 2015) claims that few specific methods and

techniques for identifying and modeling compliance risks have been developed. Thus, they propose a systematic approach for graphical modeling compliance risks and their documentation facilitating communication among experts from different backgrounds.

Figure 17 shows the factors that influence the Achieving Legal Compliance category (shows in Figure 6).

Figure 17 – Factors influencing the Achieving Legal Compliance category



Source: The author (2021).

4.6.5.1 ***Working with Data Protection Regulation category***

Working with regulated environments is considered by some respondents to be challenging and frustrating, as verifying the understanding and compliance with the law is tiring, as it requires several rounds of verification due to the lack of a legal framework and the entry into force time is concise. Some interviewees consider it is a difficult environment to work in. Thus the Company must provide its employees specific **[3] training on data privacy and personal data protection regulation (e.g., GDPR and/or LGPD))**, and awareness can make those involved in the project perform tasks more correctly and become less frustrating. The achievement of **[3] Training on data privacy and personal data protection regulation (e.g., GDPR and/or LGPD)** with support from **[13] Technical support area participation** positively influences employees **Working with Data Protection Regulations** category.

LGPD Section 6 (VIII) (BRASIL, 2018b) deals with the adoption of measures to prevent the occurrence of damages due to the processing of personal data. Item X of the same article deals with accountability, which consists of demonstrating, by the agent, the adoption of effective measures capable of proving the observance and compliance with the rules on protecting personal data and even the effectiveness of these measures.

Canedo et al. (CANEDO et al., 2020) claim that some organizations use the practice of conducting their practitioners' continuous training to encourage their practitioners' long-term involvement with the organizational data privacy policy. Sirur, Nurse e Webb (SIRUR; NURSE; WEBB, 2018) cites mechanisms and techniques employed during GDPR implementation. One ubiquitous non-technical response was regarding the increase in training and education. Good training trumped rigid and verbose engineering practices as well-trained engineers could make more flexible, informed, and innovative decisions than a static software engineering model.

> *"In terms of privacy, we are doing much training, and we are talking a lot about privacy, internally, in the Company, doing privacy events too*" [Interviewee 04].

> *"Training is a requirement of the law that we have to do this. We do the training and take a minute of attendance to evangelize the whole company*" [Interviewee 13].

> *"The first line is to work in the operational area, providing training and going over what this regulation determines for it to take effect*" [Interviewee 16].

> *"We started creating internal articles for the team and recording recurring situations. We started to find patterns that happened and how we would take these measures. Nevertheless, these measures were wholly based on what the GDPR Law brings us*" [Interviewee 20].

The software must be in line with personal data protection laws, so have an **[14] Data Protection Regulation is clear/precise** positively influences. The law must always prevail above the understanding of the stakeholders. If the need for stakeholders conflicts with the law, "*Law will always prevail because we cannot develop something that does not comply with the legislation in general*" [Interviewee 10].

> "*We have a prior understanding of the Law's interpretation, we talk to this end-user, and we go for the right interpretation of the Law*" [Interviewee 15].

> "*We are making interpretations of what we are doing and whether they comply with the law. Document everything in order to prove that we complied*" [Interviewee 13].

The factor **[1] the time for the entry into force of the law is short and makes the implementation of some features unfeasible** influences the **Working with Data**

Protection Regulation category negatively. A specific case in the health sector is presented by Interviewee 16:

> *"Brazilian National Health Agency (NHA) releases it with a period of three to six months for the systems to adapt. So, we were aware of that, and when she launched this NR (Normative Resolutions), we have to detail the NR precisely and identified the need for changes in the management system"* [Interviewee 16].

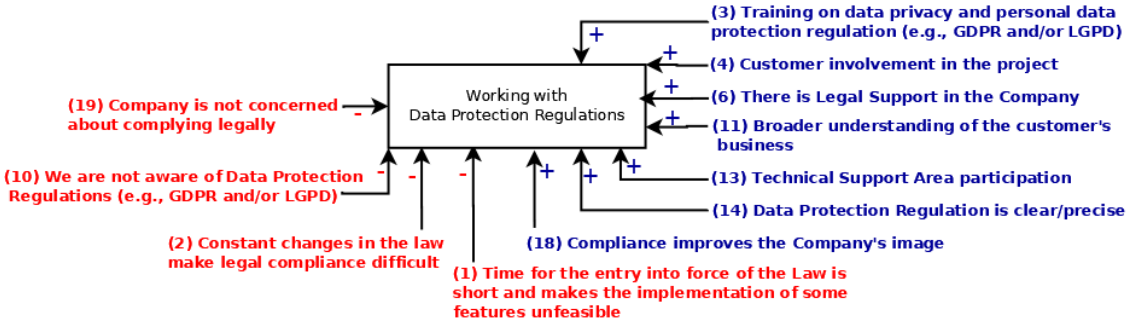
In addition to these factors, private company respondents cited that their projects are custom-made, and for each business, there is a law in force. Furthermore, laws in different spheres (municipal, state, federal, and international) become costly for the team to develop software compliance with specific legislation for each region. Because it is necessary to know the laws individually, it is not feasible for small private companies that do not have a legal department.

> "*Each City and State can create its legislation, making development a little tricky because we will develop on-demand under a single law. It is challenging for you to study and know what point you will customize for a specific region, a specific city*" [Interviewee 20].

Another factor that negatively influences whom **Working with Data Protection Regulation** and makes legal compliance difficult is the **[2] Constant changes in the law make legal compliance difficult**. According to Interviewee 06 *"Sometimes we already had the project ready, started the approval, and then the legislation is changing, and we had to go back and adapt*". Consequently, to avoid interpretation problems and noncompliance, they try to be as conservative as possible in interpretation, not to make the development of the product unfeasible. Figure 18 shows the factors that influence the Working the Law category.

We can summarize the process for legal compliance in the following steps, presented in Figure 19. Regarding legal compliance, the source of requirements will always be legislation, regulations, reading manuals, interpretation of specific legislation for each domain, as the system needs to comply with the legislation. For this understanding to be disseminated throughout the company, it is necessary to have a Culture of Privacy to guarantee data subjects' rights. One way to strengthen this Culture is to carry out training, lectures, courses. Verification, by an auditing company, if the practices implemented in the company comply with the GDPR

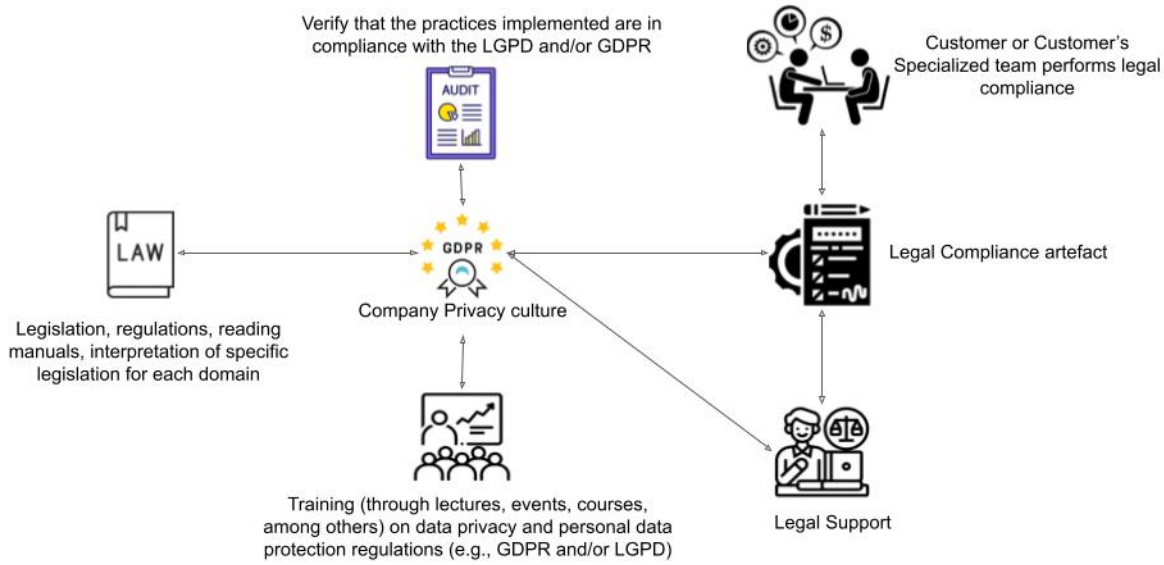
Figure 18 – Factors influencing the Working with Data Protection Regulations category



Source: The author (2021).

and/or LGPD. The legal support area and the client's specialized team verify and validate the legal compliance artifact to its adherence to legislation.

Figure 19 – How do Companies verify legal compliance?



Source: The author (2021).

The following section will briefly discuss the strategies derived from the results of qualitative research based on semi-structured interviews.

4.7 DISCUSSION

Organizations today need to have both lawyers and engineers involved in privacy compliance efforts. According to Swire and Antón (SWIRE; ANTON, 2014), the best results come from collaboration because of the value, knowledge, and expertise that both stakeholder groups (lawyers and engineers) bring.

Companies must provide legal support to ensure that the specification of legal requirements meets the legislation. This process must rely on the advice of the legal area to verify the adequacy of the definitions of data processing to the LGPD and the legal basis to legitimize the processing. It is also necessary to evaluate other laws and regulations applicable to the organization's sector.

Privacy engineers require multidisciplinary knowledge and skills. To be effective, they need to understand both technical and non-technical considerations. Privacy engineers are tasked with managing risks. Privacy engineers must then apply systematic risk analysis, using privacy impact assessments to measure and quantify identified risks. Finally, privacy engineers must design controls to mitigate those risks, including privacy-respecting architectures, effective privacy policies, and a range of data management methods, including minimization, anonymity, aggregation, and the use of privacy-enhancing technologies.

For all demands, **[12] Participation of the Customer's Specialized Team** to provide the necessary inputs for development; usually, that group will say that it was delivered as specified. The Development Team, Business Analyst, and the **[21] The Customer presents the legal framework** will perform a coverage analysis of occurrence against the specification to identify if implemented as written. If not, it was a misunderstanding of the development team, creating a correction demand. If it conforms to the requirements specification but does not meet Customer requirements, it is a maintenance demand prioritized in the next iteration. The development team has access to laws, standards, a procedural document that has to do with that demand, but the **[5] The Customer validates requirements analyst's understanding**. In addition to this specialized Customer team, there is **[13] Technical Support Area participation**, a team that, over the years, as it meets the demands, acquires a broader knowledge of that Client's business. So, it is still another level of information, understanding, and validation of the work developed.

In the following subsections, we describe eight fundamental propositions (**I. P1** - **I. P8**), which can be seen as recommendations that must be considered during the specification

of privacy requirements with reduced ambiguity and verify the legal compliance of software systems.

I. P1 - Specialized Support Areas (Ambiguity Analysis, Anonymization, Legal, among others) are critical for reducing ambiguity in legal requirements specification and compliance with the law

Looking at public and private interviewees, the **[12] Participation of the Customer's Specialized Team** is critical to reducing ambiguity and meeting legal compliance, as they provide advice on Legal and Technical aspects related to the Company domain.

Since software development projects deal with different domains, especially in private companies, there needs to be a **[13] Technical Support Area participation** with experienced professionals (software architects, senior programmers, project managers, and others) interacting with the development team. In Public Companies, as they interact with the same types from the Client's projects, they become knowledgeable about the domain and the Law governing it. However, it is still necessary **[13] Technical Support Area participation**. For this, **[6] There is Legal Support in the company** that can support project members during the development process to specify requirements with less ambiguity and, consequently, have no noncompliance problems with the legislation.

In the study by Sirur, Nurse e Webb (SIRUR; NURSE; WEBB, 2018) no participant successfully cooperated with any legal professionals as part of their compliance process. Thus, without a legal professional of some kind, the average engineer would struggle to utilize the regulations directly.

Similar to the study of Sirur, Nurse e Webb (SIRUR; NURSE; WEBB, 2018), for the most part, **[15] There is no support from the Company's Specialized Legal Area** inside the development teams, and the Legal sector is not just dedicated to software-related privacy, security, or software systems compliance issues. This sector usually handles all the company's legal activities and meets a steady demand if requested by the development team. The composition of multidisciplinary teams allows each member to overcome this absence of the company's legal professional or the development team.

The study revealed that only one Company (C01) has a specialized department for Ambiguity Analysis formed by analysts who do not participate effectively in the project team members. From then on, the documentation starts to come out with better quality, reduce

the rework, reduce the risk of not adequately meeting what the Customer is requesting and, thus, contribute to strengthening the Company's image.

In summary, **[12] Participation of the Customer's Specialized Team** presents the legal framework related to the company's domain. Then, Requirements Analysts, Legal Experts, Development Team Members, and other staff members must come together to carry out the data mapping or data flow to identify technical and legal vulnerabilities. The data mapping steps must be carried out both in the case of legacy software and in the process of adapting existing software to a data protection law or in the development of new software.

The data mapping or data flow identifies the source of the data, the company's sectors that the data transits, analyzing all the data life cycle (how it is collected, processed, analyzed, stored, shared with partners, reused, and discarded). Under art. 5 of the LGPD (BRASIL, 2018b), identify the legal bases that support the treatment of data, and verify that the privacy policy is up to date by presenting the described flow. In this way, the company will be complying with art. 37 of the LGPD (BRASIL, 2018b) "The controller and the operator must keep a record of the person processing personal data that they carry out, especially when based on legitimate interest." Moreover, Art. 30 of the GDPR "Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility. That record shall contain all of the following information [...]".

I. P2 - Promoting the Company's privacy culture improves employees' awareness of privacy

This proposition is about how the organizational perception of the importance of Personal Data Protection, we named **[24] Privacy Culture in the Company**.

Respondents 14, 19, and 22 stated that in the companies they consulted with, the Development Team members, when asked about implementing privacy practices, stated that they contacted the security sector if they had any doubts about privacy aspects. There is still a stigma in thinking that privacy is only a security issue and not the whole company.

In some development teams, the privacy professional joined the development squad. Nevertheless, the culture was still that privacy is a person's responsibility. To change this situation, companies invest in training for developers focused on privacy and security so that people start to create a culture that the security person is there to support, answer questions, and bring new information. However, they are not responsible for safe development. This lack of knowledge about the Personal Data Regulations may happen because the discussion remains

restricted to the area of Information Security or Law Department. As Interviewee 01 said, "I do not know how to speak to you. Maybe (*team member name anonymity preserved*), from the area of Information Security, can help you more".

It is worth mentioning in the interviewees' statements that there is a misunderstanding regarding the definition of privacy and security, as was also presented in the work of Peixoto et al. (PEIXOTO et al., 2020).

Participants reported that organizations generally do not give proper attention to ambiguity in software systems' privacy requirements specification and legal compliance analysis. The company culture is much more than a set of guidelines. Everyone who joins the company, i.e., onboarding, must have privacy and security.

Therefore, **Promoting the Company's privacy culture** is one of the essential factors for specifying unambiguous and precise legal privacy requirements and compliance with the legislation.

> *So, we are trying to create a privacy culture here so that everyone thinks about privacy. We are trying to work for the whole company with this, and there is no sector division* [Interviewee 12].

> *Another point of failure was culture. People have a vision of thinking that security is a security concern, and one of the jobs that we have to do is to show that security is everyone's responsibility and, therefore, the developer has to know about security and develop in a way safety.* [Interviewee 19].

Some authors (CANEDO et al., 2020), (HADAR et al., 2018), (BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019) cites the importance of developing an organizational climate related to privacy in companies to influence the behavior of developers in implementing privacy in the development of software products. Canedo et al. (CANEDO et al., 2020) demonstrate the need for software development organizations to inform their members about LGPD, as organizational privacy characteristics and procedures should be known to all, including software development teams. Omoronyia (OMORONYIA et al., 2010) claims that privacy awareness is a core determinant of the success or failure of privacy infrastructures. If systems and users are not aware of potential privacy concerns, they cannot effectively discover, use or judge the effectiveness of privacy management capabilities.

As Abdullah et al. (ABDULLAH; SADIQ; INDULSKA, 2010) state, the culture of compliance is ingrained in each company's employees' daily rituals, including senior management. Morton

(MORTON, 2005) states that the following five elements are necessary for a firm to have a strong and effective control environment and culture of compliance: strategic vision (also known as compliance goals or objectives); identification of risks; establishment of controls; documentation; and accountability.

In summary, promoting the organizational awareness of ambiguity in privacy requirements specification and legal compliance importance keeps the development team (and any stakeholder) well-informed about privacy, legal compliance, and Personal Data Protection Regulation (as GDPR or LGPD). They promoted team members' training, informing the Customer about the actual state of software privacy and legal compliance. Simulation of privacy disclosure and failures, and show business impact. Regular meetings to discuss privacy practices. External audit to mitigate human problems and verify legal compliance.

> *I say to several Clients, "if you are going to make a privacy policy and leave it there in the drawer". Because it has to be a part of the company's culture for employees to follow, then its implementation will depend a lot on the company's communication team so that it has an endomarketing action, internal communication and that there is in the company's culture because it does not do anything* [Interviewee 14].

I. P3 - Reducing ambiguity in the privacy requirements specification and achieving legal compliance requires cross-functional teams

When constructing the main story (see Figure 6), a specialized support area appears to support all other activities, becoming extremely important in a context with legal requirements, which may have ambiguity and difficulties understanding by stakeholders in the project.

Therefore, **forming multidisciplinary teams with the support of legal experts** are essential factors for specifying precise legal privacy requirements and compliance with the legislation.

In the context of projects that deal with legal requirements, due to the subjectivity of the Law (ambiguity, inconsistencies, cross-references, and others challenges), there may be challenging to understand the legal terms or lack of knowledge of the Client's project domain, or misinterpretation by those involved in the project. These facts can happen during the requirements elicitation or specification sessions, especially when drafting the requirements specification document, **[23] Always in Natural Language**.

Li et al. (LI et al., 2020) identified three main challenges at the Company that hinder GDPR compliance: reliance on manual GDPR tests, limited awareness and knowledge of privacy requirements, and balancing GDPR compliance in a competitive data business. Besides, identified contextual factors contribute to one or more specific GDPR challenges (ambiguity of GDPR, lack of legal training, lack of privacy experience, and lack of shared understanding). Our study also identifies similar factors that minimize the organizational awareness of ambiguity in privacy requirements specification and legal compliance importance.

The lack of **shared understanding among project team members** jeopardizes the requirements specification and, consequently, the legal compliance verification and validation for software systems. During the development process, if any doubts arise because the information is ambiguous, the first thing that team members project' do is talk to someone on the team they imagine has better knowledge or someone who wrote that. Then, an inter-team understanding that is working with the same need. If doubts have not been clarified, then the Project Manager (who usually knows a lot of the business being implemented) or Service Manager is consulted to contact the Customer and clarify it as necessary.

The GDPR consists of ninety-nine articles (REGULATION, 2016), but the entire GDPR is written in legal speak. It is not easy for employees who are not well-versed in legal language or have specific privacy training. Besides, GDPR is often ambiguous (COOL, 2019). Interview 15 and 04 cited this difficulty.

> *"The Legal text is difficult to understand. It has a legal language that is not so trivial to be understood. The Requirements Analyst talks to the end Customer to understand what the Law says, clear up doubts, analyze its impact on the system, and validate with the Customer. This whole process is manual"* [Interviewee 15].

> *"Often, the legal language of some lawyers who write the laws is ambiguous. Furthermore, the systems could be implemented in either way or even in both ways. We ask the Client to decide how the Client wants to implement the system"* [Interviewee 04].

Therefore, to avoid the problems presented above, it is necessary to know the Law; discuss it with the team to have the same vision. In case of doubts about the proper understanding for implementing that Law in the system, consult other sectors Company itself or Regulatory Agencies to clarify ambiguity. Another strategy companies use when faced with ambiguity is to discuss with everyone involved in the project. *The part I would say is the most important, is the direct contact with the Customer* [Interviewee 09].

I. P4 - Training to instruct employees regarding ambiguity analysis in privacy requirements and legal compliance for software systems improves Legal requirements specification

The laws, which we covered in this study, LGPD and GDPR, require companies to be accountable; in other words, they have ways to demonstrate that they are legally compliant. Because it is not enough to adapt, Companies have to show an Adequacy Plan; if there is any failure, this documentation shows that they make the most significant efforts to solve it. Furthermore, it becomes a problem if they do not have that part. It is not clear that the Company was trying to solve it, planning and executing it. Some legal compliant steps are not related to software; for example, the companies have to train their employees.

> *We have a Privacy Committee here, and we keep the Minutes meetings. If the Government arrives here inspecting, we will show "it is here! We are doing committees, training; here is the document you need. Here is the Adequacy Plan"* [Interviewee 12].

> *So it is doing the documentation of all data entry points, in the case of LGPD, of all tables, making a complete mapping, going through this mapping by all teams. You say that this data exists, it is like this, it is used in this way, and then you reference which products are used, why you use them, etc. It is a very documentation process* [Interviewee 13].

Canedo et al. (CANEDO et al., 2020) state that organizations must provide their professionals with specific training related to Data Privacy Laws. Information and Communication Technology (ICT) practitioners have stated that they lack the knowledge necessary to implement privacy principles and LGPD guidelines.

Companies employees' should be trained that address **[27] Privacy-enhancing techniques** (encryption, anonymity, tools to remove, change, hide, or blur data, and tools to inform users after collection (MEAD; MIYAZAKI; ZHAN, 2011)). Other related topics are necessary so that the Development Team has the knowledge and can use these privacy techniques in software development. Privacy by design (organization must incorporate privacy into the system throughout the Software Development Life Cycle (SDLC)). Privacy by default (the most secure privacy settings should be applied by default, with no manual input from the end-user). Fair Information Practices (standards about the collection and use of personal data and privacy issues), Privacy Impact Assessment (used to identify and mitigate privacy risk),

and Data Minimisation (data collected and processed should not be held or further used unless that for reasons that were clearly stated in data privacy agreement).

> *"It is important that these concepts (Personal data (Art. 7 of the LGPD), Sensitive personal data (Art. 11 of the LGPD), anonymization, treatment, consent) are internalized in the company"*[Interviewee 17].

> *"Legal requirements specification without ambiguity is a multi-disciplinary activity, it is necessary that there are professionals from various areas: business, legal and information technology professionals."* [Interviewee 14].

I. P5 - Personal experience compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification

In the companies participating in this study, sometimes, there were no written privacy requirements used as guides. However, one efficient "technique" is tacit knowledge (GERVASI; SAWYER; NUSEIBEH, 2011) (personal wisdom, experience, insight, and intuition) because, through project members' experience, they can identify what may be a risk, know about the Customer domain, and the legislation governing that domain. So these people discuss and come to a consensus on what the Law dictates.

> *The team is always thinking about the data stored. We question what we read, and we take courses, workshops to comply with the GDPR, find out or try to identify whether that data is sensitive or not. This goes a lot with our feeling* [Interviewee 03].

The Privacy by Design (PbD) is a concept developed by Ann Cavoukian (CAVOUKIAN et al., 2009). The idea is to incorporate privacy and personal data safeguards in all projects developed. The seven principles of PnD are 1. Proactive not Reactive; Preventative not Remedial; 2. privacy as the default setting; 3. privacy embedded into the design; 4. full functionality; 5. end-to-end security; 6. visibility and transparency; 7. respect for user privacy. Cavoukian (CAVOUKIAN, 2020) claims that several organizations have transformed these seven principles into specifications, recommendations, and practical suggestions of applications to use in different projects. Perera et al. (PERERA et al., 2019) proposed, using the PbD framework, a guide with a set of best practices to help software engineers ensure users' data privacy during the development of Internet of Things () applications.

> *"We are trying to ensure that all new products that appear are already complying with the LGPD. Like Privacy by Design. We debated the idea and pulled each concept from that idea to the Law's concepts. What is the legal basis? How can we process this? However, this is in the construction process. We started using it in the two new products*" [Interviewee 13].

> *"It does not happen today, but our goal is to follow Privacy by Design and perform the following steps: mapping compliance in LGPD; identifying which data travels between systems; how to integrate systems and restrict unauthorized access; making privacy a standard*" [Interviewee 22].

Thus, the lack of training in **[27] Privacy-enhancing techniques**, legislation related to the domain, or identification of ambiguity (as mentioned in Proposition P4) make the participants use their opinions and knowledge acquired from previous cases, which may result in software applications without privacy implemented.

I. P6 - A systemic view of the Company concerning privacy and protection of personal data positively influences compliance with the specification of the software requirements to the law

The Interviewee 21 (Information technology (IT) Manager) and 22 (IT Lawyer), despite being from private and public companies, respectively, summarize the process of handling Legal Requirements as a partnership between three sectors: Legal, IT, and Governance/Processes (which presents internal procedures, business rules). Interviewee 22 states that "*The Legal and Governance sectors need to reach an understanding and thus present to the IT sector to "translate" into computational language.*" Interviewee 21 states that "*The role is quite a partner, between three individuals who were responsible for this group: me, for the Technology and Information Security part, the Legal Person, who was responsible for the legal items of adequacy, and the process person, who was responsible for mapping and surveying processes and data sources. A well-defined line of how far each went*". Thus, it is clear that "*The compliance process of Personal Data Protection laws is a systemic one. It is necessary to analyze all sectors that the data are scattered within the Institution*" [Interviewee 22].

> *" You can start from training. But sometimes, just training is not enough. Ethical conduct within the company, the organizational vision is more efficient*" [Interviewee 22].

The company needs to demonstrate the technical measures adopted to mitigate the risks and prove its effectiveness to inform that it complies with the LGPD (art. 6, item X). For this, one of the strategies is the Data Protection Impact Assessment Report - Data Protection Impact Assessment Report (DPIA) (in Portuguese, Relatório de Impacto à Proteção de Dados Pessoais), provided for in art. 5 items XVII, of the LGPD: *personal data protection impact report: controller documentation containing the description of the processes for processing personal data that may generate risks to civil liberties and fundamental rights, as well as measures, safeguards, and risk mitigation mechanisms*. The art. 38 of LGPD presents the sections of the DPIA. The report must contain, at least, a description of the types of data collected, the methodology used for the collection, and to guarantee the security of the information and the analysis of the controller regarding measures, safeguards, and risk mitigation mechanisms adopted. There is also a need to develop relevant IT tools to improve employee training and a software management platform compliant for all company applications.

> *"If necessary, to prove in court that you have made adjustments to the LGPD. It is necessary to analyze aspects to knowledge and technology, assess whether the solutions are accessible with the public budget, evaluates if there was training, be careful to inform the employees"* [Interviewee 22].

> *"We can automatically detect which data sources we have, who pulls from each source, to have a map of the data automatically, then with that, know where we can act"* [Interviewee 12].

> *"The company deals with anonymous, pseudo-anonymous data, and we can demonstrate this. Just document all the data tables that exist. Moreover, say here we forbid, and no personal ID, no phone, no this. This source that came from an external base provided it is appropriately transformed, "hashed" and taken out of precision*" [Interviewee 13].

> *"My current team delivers infrastructure to the company's internal teams, and we do it in an automated way, with the cloud. Furthermore, one thing that we care about, at first, is something that the GDPR calls the PII (Personal Identification Information)*" [Interviewee 03].

Privacy policies and terms of use are essential objects for knowing privacy, and companies demonstrate their commitment to transparency in the treatment of personal data. Nevertheless, they are poorly designed, and uninteresting (YAMAUCHI; SOUZA; JUNIOR, 2016). One way

of establishing privacy is through privacy policies, which can vary in complexity, legal sophistication, and service coverage, known for their excessive size and complexity (WILSON et al., 2016). Schaub et al. (SCHAUB et al., 2015), service providers and data consumers should inform users about the information collected, how long it is used, and retained.

For elaborating the clear privacy policy, we need to understand the context of the life cycle of personal data (collection, treatment, storage, and disposal): What are the legal bases, i.e., what are the requirements for data processing (for example, consent of the holder (art. 7, I, LGPD), treatment by the public administration (art. 7, III, LGPD), among others); How the principles of personal data protection laws, and how the holder's rights are met?

> *The company legal sector that regulates these types of policies defines privacy policies* [Interviewee 04].
>
> *Usually, compliance guidance comes because it has all the information about which standards and laws the company is subject to according to the sector in which it operates* [Interviewee 19].

I. P7 - Verification & Validation activities only do not guarantee software privacy requirements specification and legal compliance

Many countries have data protection laws in place (GDPR (REGULATION, 2016)); Argentina, which has the Personal Data Protection Law 25.326 (ARGENTINA, 2000) (in Spanish, Ley de Protección de Los Datos Personales or PDPA), in force since 1994. In Uruguay, the right to data protection is provided by Law 18.331, edited in 2008 (URUGUAY, REPúBLICA ORIENTAL DEL, 2008) (in Spanish, Protección de Datos Personales y Acción de "Habeas Data"). So it is possible to analyze case law to not make the same mistakes as companies punished for breaking a particular law.

Weekly meetings with project members and the client also resolve the ambiguity cause faster. The customer has the most extensive knowledge of the legislation that the system must comply with, so when there is any difficulty understanding, the customer should address the doubt. When the company works with agile methods, this procedure occurs frequently, and the possible problems of noncompliance with the legislation do not reach the system approval phase.

The ISO/IEC 29100 Information technology — Security techniques — Privacy framework (ISO, 2011), and ISO/IEC 27701 Security techniques for privacy information management

— Requirements and guidelines (ISO, 2019) specifies requirements. Guides are establishing, implementing, maintaining, and continually improving a Privacy Information Management System (PIMS) (ISO, 2019).

> *Last year, we did a test to find out what level of effort would be necessary to adapt to the LGPD. So, we determined last year that we would use the ISO/IEC 27701 controls to survey the distance desired to generate the action plan to get closer to compliance* [Interviewee 21].

The LGPD came into force in August 2020. All companies that process data, whether physical or digital, must be in LGPD compliance. When asked how companies are preparing to comply with data protection laws, respondents 14 and 19 responded:

> *Technology companies have been well ahead, although data protection embraces any company. Usually, it is the technology companies that are ahead. They are those whose business model depends on them being complied with because they know that tomorrow all their business can stop. It is different from a school, where the staff says, "you may be breaking the data protection law," but they will not stop school because it is not the main activity of the school; it can be fined* [Interviewee 14].

> *If you take well-organized organizations or anyone invested using investment funds, which has shareholders, they are all moving towards compliance, seeking to understand the Law better and make the necessary internal adjustments* [Interviewee 19].

One Company (C12) is hiring a third company, a consultancy, or a Law firm that will help them analyze if their actions are legally compliant. *So, now we are hiring this third party, this consultancy, it can be one of the Big Four (Ernst & Young - EY, Deloitte, Pricewaterhouse-Coopers - PwC, KPMG), it depends, we are in the hiring phase. The idea is that they help us with this. If a Big Four says, for example, that we comply with the Law, our customers will see this with look kindly* [Interviewee 12].

As Abdullah et al. (ABDULLAH; SADIQ; INDULSKA, 2010) state, many consulting firms, among them Big Four, offer services such as developing a compliance knowledge base, linking regulations to business processes, and relying on comprehensive guidelines and frameworks.

In addition to organizational aspects, other aspects must be taken into account, such as the issues developers face when they attempt to embed privacy into software systems,

presented in Senarath and Arachchilage (SENARATH; ARACHCHILAGE, 2018). The practices to guide software developers to embed privacy into the software systems they design are:

> *Privacy guidelines should be simple, straightforward, and explicit as developers have trouble executing soft decisions.*
>
> *Developers should be given formal education on privacy practices as developers' lack of knowledge affects their personal opinion, which interferes with the way they embed privacy into designs.*
>
> *Privacy guidelines should have steps for evaluation.*
>
> *Privacy requirements should be specified with engineering techniques such as anonymity, as developers find it difficult to relate privacy requirements to engineering techniques.*

The following proposition does not directly concern reducing ambiguity in the specification of legal requirements by the legislation. However, it makes us reflect on the lack of integration between industry and academy, at least from these participants' perspectives.

> **I. P8 - Academic resources (methodologies or tools) for Privacy Requirements Elicitation and Specification are not used by industry**

There are several recognized academic papers on Methodologies for Privacy Requirements Elicitation or Specification evaluated using empirical methods such as, for example, Privacy Safeguard method (PriS) (KAVAKLI et al., 2006), SQUARE for Privacy (BIJWE; MEAD, 2021), Goal-Based Requirements Analysis Method (GBRAM) (ANTON, 1996), among others that interviewees do not cite. There is a methodology called LINDDUN (DENG et al., 2011) aims to support the elicitation privacy requirements of software-intensive systems and design of a Data Flow Diagram (DFD) created based on the high-level system description and followed by mapping privacy threats to the DFD elements to determine the corresponding threats. Privacy Criteria Method (PCM) (PEIXOTO et al., 2019) an approach designed to guide the specification of privacy requirements in agile software development. RSL-IL4Privacy (CARAMUJO et al., 2019) a domain-specific language for specifying privacy policies that can support distinct levels of formality, namely support for multiple modes of presenting privacy requirements to increase integration with a broader variety of authoring and analyzing practices.

However, none of the interviewees mentioned using any similar methodology or tool. As well as in the work of Sirur, Nurse, and Webb (SIRUR; NURSE; WEBB, 2018), in this paper,

none of the companies used academic research in their compliance process, with most finding academic research not pragmatic or transparent enough for immediate use.

4.8 CHAPTER SUMMARY

This Chapter presented a qualitative study on how Brazilian IT companies (public and private) deal with the inherent ambiguity in legal requirements specification and how they perform legal compliance verification.

We present 27 factors classified in personal, organizational, project-related, or technical that affect positively or negatively the categories (Requirements Elicitation, Requirements Specification, Working with Data Protection Regulations, Communication between development team members, Reducing ambiguity, Specialized support area, Achieving legal compliance). Moreover, eight propositions, which can be recommendations considered during the privacy requirement specification with reduced ambiguity, verify the legal compliance of software systems.

5 SURVEY-BASED STUDY

Given the results of literature research, and the twenty-two interviews, presented in Chapter 4, it was possible to identify a set of eight propositions, twenty-seven secondary factors that influence positively or negatively the categories.

However, we still needed a proper understanding of how ambiguity resolution in requirements specification is achieved in software development companies and how to achieve legal software compliance from the early stages of software development (i.e., requirements engineering). These are questions that the current literature has not been able to answer completely. Therefore, we adopted the survey method for this study to gather information regarding software practitioners' experience and expertise. Surveys collect qualitative and quantitative information to provide a snapshot of the current status related to a phenomenon (WOHLIN et al., 2012).

5.1 RESEARCH GOAL

Our primary goal is to conduct descriptive survey questionnaires (more specifically, a self-administered online questionnaire) to collect the software practitioners' perceptions regarding the factors and actions identified from a set of interviews. Moreover, the survey aims to corroborate the practices presented in Chapter 2 and an interview-based study conducted in the industry presented in Chapter 4 to tackle ambiguity and legal compliance in legal requirements specification in the early stages of the software development process aiming to improve legal requirements specification. These factors and actions were identified from a set of interviews (NETTO; SILVA; ARAÚJO, 2021).

Descriptive surveys can be conducted to enable assertions about some populations, like the distribution of specific attributes. The concern is not why the observed distribution exists, but instead what that distribution is (WAGNER et al., 2020).

We summarize the goal of this research as follows:

> ***investigate*** *practices to reduce ambiguity in legal requirements specification compliant with data protection laws,* ***for*** *generating evidence on the benefits and drawbacks of using them for requirements specification, from the* ***viewpoint*** *of IT professionals with industrial experience in RE or system analysis, in the* ***context*** *of software companies.*

Some works related to privacy requirements engineering practices in the industry reinforcing the importance of this research topic. However, we believe that much more should be done to get the actual perceptions of the practitioners about the approaches they use to elicit, specify privacy requirements and verify and validate legal compliance. The contribution of this study is in this direction. Thus, the research questions that guide this survey-based study are:

RQ1. What are the current practices for specifying the legal requirements that the companies use in their daily work?

> *This question investigates the current situation of early requirements engineering practices in organizations (i.e., the techniques, methods, and tools used to specifying legal requirements), report identifiable problems, and how this process takes place*.

RQ2. What are the current practices of IT professionals towards specifying software legal requirements with reduced ambiguity in their daily work?

> *Based on IT professionals' knowledge and experience, this question investigates how frequently and difficult it is to identify and interpret ambiguity in legal requirements. How to specify legal requirements; who is responsible for interpreting or resolving ambiguities in the legal text, and how this process takes place*.

RQ3. What are the current practices towards achieving and verifying legal compliance of software requirements with data protection laws in their daily work?

> *This question investigates the awareness by IT professionals and their company regarding Data Protection Law (i.e., LGPD, GDPR). The techniques, methods, and tools used to validate and verify legal compliance, when and how this process takes place*.

RQ4. What perceptions do IT professionals with industrial experience in RE have about ambiguity resolution in legal requirements specification and the compliance of such requirements with data protection law?

> *This question aims to discover IT professionals' perceptions in some dimensions that make up the privacy culture and validate the eight propositions that emerged from interviews' data*.

Answers to these research questions provide the empirical foundation for the specification of legal requirements with reduced ambiguity and compliance with legislation, based on academic literature and industrial practice perspectives.

5.2 STUDY DESIGN AND PLANNING

We designed the survey protocol following the survey assessment checklist (available in Appendix I), defined by Molléri et al. (MOLLÉRI; PETERSEN; MENDES, 2020), which deals with dimensions: research objectives, study plan, identify the population, instrument design, instrument evaluation, participant recruitment, response management, data analysis and reporting.

The questions cover the four RQs (see the relationship between survey sections and RQs in Appendix I). The survey has 38 questions; thirty-three include closed questions: single-choice (SC), multiple-choice (MC), and five-points Likert Scale (LS). We try to cover all possible answers in multiple-choice questions and not influence the results. We always allow the respondent to state choices not explicitly offered through free text (FT) response fields of the "other" type. In addition, the questionnaire has five open questions. To seek explanatory information, participants can elaborate on their views as the process happens in the company where they are employees. Nevertheless, these questions were not mandatory.

In order to cover the most frequent possible answers, we have collected them from the systematic literature review and snowballing (presented in Chapter 2) and of twenty-two interviews (presented in Chapter 4, in (NETTO; SILVA; ARAÚJO, 2019), and (NETTO; SILVA; ARAÚJO, 2021)).

The target audience of this study is IT professionals with industrial experience in RE or system analysis. Our population is IT software company. Participants were sampled by convenience, using contacts from researchers involved in the study, social media, and the authors of the articles identified in SLM (NETTO; PEIXOTO; SILVA, 2019) how the unit of analysis has IT practitioners that work in software projects.

Participants were invited to the survey through posts in social networks, invited connections to follow the LinkedIn page, academic email lists, and groups. LinkedIn is a convenient forum to find a good representative sample of the population that we wanted to reach (MELLO; SILVA; TRAVASSOS, 2015).

We sent 450 invitations formally to professionals with the most diverse backgrounds through a Cover letter (available in Appendix I). In addition to these invitations, we make

a post on social media and use the following keywords in English and Portuguese: “Requirements Engineering", “Software developer," “Software development," “Privacy Requirements Engineering," “Software engineer," “Software privacy," “EU GDPR," “LGPD," “Data Privacy," “Data Protection Laws." We also use the advanced search on LinkedIn[1] (through the filters: people, locations, and current company) to find professionals with experience in the areas related to this study. We use similar strategies in ResearchGate[2], and Twitter[3].

5.3 DATA COLLECTION

We did a pilot study with a Ph.D. candidate, a Ph.D. researcher with experience in RE, and a Ph.D. professor. The pilot served to verify the understanding of the questions, validate the questions, and verify the participants' time to answer the survey. The responses for the pilot session were not included in the data analysis because they were used for validation. After making the changes suggested by the participants of the survey pilot study, we made the survey available online for eight weeks (April 2021 to May 2021), with versions in Portuguese and English to facilitate the participation of people whose native language to Portuguese.

We use LimeSurvey[4] to host the survey. On the Welcome page, explain the purpose of the survey and how they will be communicated, and present the Survey Principles and Participation, including Informed consent (IC) (available in Appendix I). Access to the rest of the questionnaire was blocked until consent was given. IC informs the overall objective and importance of the research, guaranteed data confidentiality, participation anonymity and voluntary, and the right to withdraw from the research at any moment.

5.4 QUANTITATIVE DATA ANALYSIS

We exported the complete responses to the CSV file and used the R programming language and RStudio for our statistical analysis. To ensure the quality of the data obtained from the questionnaire, we applied sanity checks to find obvious errors in the data. For example, we define variables as factors, transforming Likert scale entries to integers for facilitating ordinal comparisons. To analyze the survey data, we used descriptive statistics (KITCHENHAM;

[1] https://www.linkedin.com/
[2] https://www.researchgate.net/
[3] https://www.twitter.com
[4] https://www.limesurvey.org/

PFLEEGER, 2008) for categorical variables utilized frequencies, percentages, mean, median, and mode. And inferential statistics, testing hypotheses for the distributions of the dependent variable as a function of the independent variables.

We used the Chi-square test (SIEGEL; CASTELLAN, 1975) (indicated for ordinal variables, as in the case of the Likert scale analysis) to check if there is a difference in the distribution of the dependent variable. Between the independent variable levels and if this observed distribution is significantly different from the expected one. However, due to the low number of survey responses, Fisher's Exact Test (SIEGEL; CASTELLAN, 1975), as indicated when there are fewer than 20 observations in the sample or fewer than five counts in a cell, which happens with some categories of data collected in the survey.

From the chi-square test statistics, identifying the association between the variables (p-value > 0.05), three measures of association intensity (SIEGEL; CASTELLAN, 1975) can be calculated: Contingency coefficient, ϕ (phi) and Crámer V^2.

The Kruskal-Wallis (K-W) test is a nonparametric statistical method that compares independent groups of sample data (DANIEL et al., 2000). It was used for studying differences among stratified demographic groups and variations in the perceptions across practitioners. The K-W test requires the sample size of each group to be at least five (DANIEL et al., 2000), the sub-groups with too few observations (less than five) were omitted from this test. We applied the K-W test with a confidence level $\alpha = 0.05$.

When the p-value is less than 0.05, the hypothesis that the samples originate from the same distribution is rejected. This happens when the answers to a survey question differ significantly. We performed the test for each survey question and each corresponding answer option. For subgroups with two independent samples, we used the nonparametric Mann-Whitney test, also called the Wilcoxon rank-sum test (SIEGEL; CASTELLAN, 1975).

In cases where the null hypothesis was rejected, and a significant difference was noticed. Dunn's test (SIEGEL; CASTELLAN, 1975) of multiple comparisons was performed based on rank sums with Bonferroni correction to examine which of the sub-groups significantly differed from the rest (significance level was set to 0.05).

In this section, we report the results related to the four research questions detailed in Section 5.1. Table 7 shows the mapping between the research questions and the questions in the questionnaire as outlined in (NETTO; SILVA, 2021).

RQ1.What are the current legal requirements specification practices companies use in

their daily work?

RQ2. What are the current practices from IT professionals to specify legal software requirements with reduced ambiguity in their daily work?

RQ3. What are the current practices towards achieving and verifying legal compliance of software requirements with data protection laws in their daily work?

RQ4. What perceptions do IT professionals with industrial experience in RE have about ambiguity resolution in legal requirements specification and the compliance of such requirements with data protection law?

Table 7 shows the mapping between the research questions and the questions in the questionnaire as outlined in the Appendix I.

Table 7 – Mapping research questions to questionnaire questions

	Questions in questionnaire
RQ1	Q10 (MC), Q11 (SC), Q12 (MC), Q13 (SC), Q14 (MC), Q15 (FT)
RQ2	Q16 (MC), Q17 (LS), Q18 (MC), Q19 (LS), Q20 (MC), Q21 (MC), Q22 (FT)
RQ3	Q23 (SC), Q24 (MC), Q25 (SC), Q26 (SC), Q27 (SC), Q28 (MC), Q29 (MC), Q30 (MC), Q31 (DT), Q32 (FT), Q33 (MC), Q34 (MC), Q35 (FT), Q36 (FT)
RQ4	Q09 (LS)

MC: Multiple-choice question; SC: Single-choice question; FT: Free-text question (optional); DT: Dichotomous question

Source: The author (2021).

We will present the data for each question using the following format: (*amounting respondents*; *percentage to the total number of respondents*). To ensure the quality of the data obtained from the questionnaire, we applied sanity checks to find obvious errors in the data and used descriptive statistics to analyze the data (KITCHENHAM; PFLEEGER, 2003).

The survey analysis was divided into two parts[5], the first, available in Section 5.5, aims to identify which Methodologies, Techniques, and Tools for legal requirements specification, reduced ambiguity in legal requirements, and legal compliance software is used in the software industry. The second part of the survey evaluated the participants' perceptions about ambiguity in privacy requirements specification and legal compliance. This survey, available in Section 5.6,

[5] https://dorgivalnetto.github.io/survey2021/

aims to validate in practice the propositions that emerged from a set of interviews presented in Chapter 4.

The following section presents the demographic questions for the first part of the survey, followed by analysis obtained regarding each research question.

5.5 SURVEY PART 1 - DEMOGRAPHICS DATA ANALYSIS

In this section, we describe the data analysis from thirty-nine complete responses in this part of the survey in a total of 117 responses (39;33.3%). We formally send 450 invitations; in addition to these invitations, we also post on social networks. The 39 completed answers matched a response rate of 8.7%. Among the 117 responses, in 28 cases, only demographic questions were answered, and in 50 cases, the questionnaire was partly completed. Therefore, we decided to exclude all incomplete responses from our analysis.

Demographic questions were used to filter and form subgroups used to investigate variations in practitioners' answers. Since the survey was online, practitioners from all over the world were able to participate. Some demographic questions were built into the survey design to confirm whether respondents belonged to the targeted population.

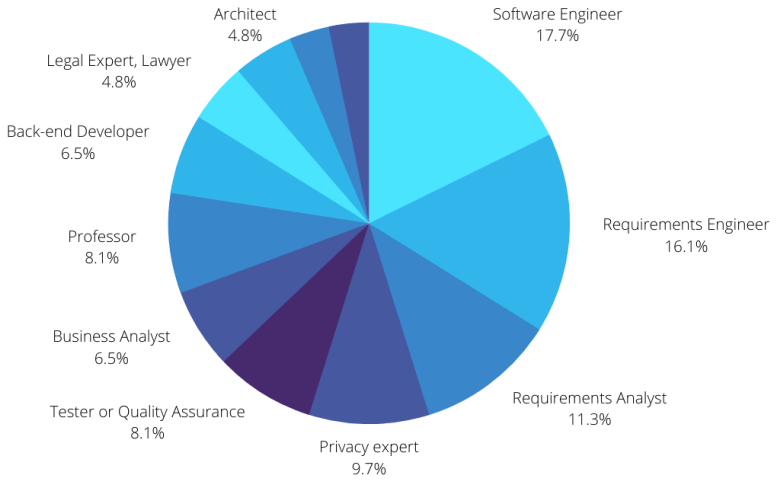
The most significant portion of respondents works in the Americas (30; 76.8%), Europe (7; 18.1%), Asia (2; 5.1%). Brazil is the dominant country with (26; 66.7%). In relation to maximum degree, the respondents are High school (1; 2.56%), Degree (4; 10.25%), Post-degree (12; 30.7%), Master (9; 23%), PhD (11; 28.2%), Post-doctoral (2; 5.1%). Which can be considered a good indicator of the quality of the information gathered in the questionnaire.

Respondents can select all roles that act or acted in the company (multiple-choice question). Figure 20 shows the roles. The most popular job is the Software Engineer (11; 17.7%), following for Requirements Engineer (10; 16.1%), Requirements Analyst (7;11.3%), and Privacy Expert (6; 9.7%).

We analyze the experience in the software industry. Looking at the distribution of participants' professionals years of experience, as shown in Figure 21, we noticed that our sample consisted of a mix of professionals with different levels of experience, assuring a good coverage of experienced participants. Practitioners had, on average, 10.56 years of experience in the software industry (standard deviation 6.97).

Figure 21 show that the majority of participants (18, 46.1%) have more than 11 years in the software industry, nine participants have between six and ten years (23%). Thus (41;

Figure 20 – Demographics: participants roles



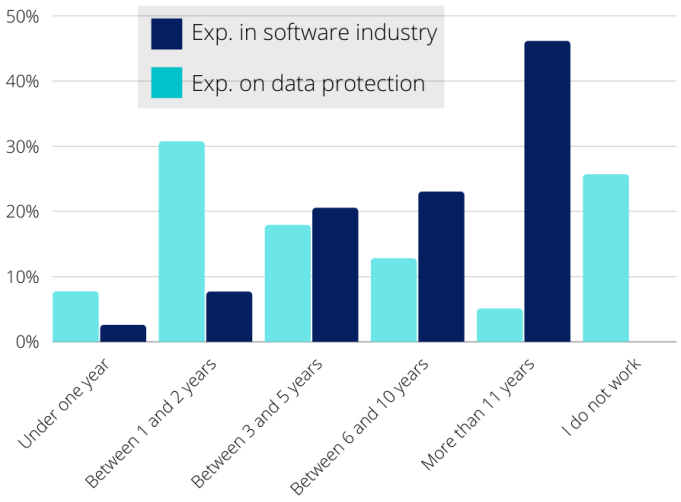
Source: The author (2021).

69.2%) have more than six years of overall industry experience. It was an experienced sample consisting of an analyst to senior-level professionals, which can be considered a good indicator of the information gathered in the questionnaire. Figure 21 also shows the professionals' years of experience in data protection (data privacy) in software projects. Practitioners' experience in data protection projects had, on average, 3.02 years of practical experience (standard deviation 3.26).

The participants have been working on data protection for between one and two years (12; 30.7%), only two participants (5.1% have been working for more than 11 years). The participants who are not working with data protection are (10;25.6%) and those who work under one year (3; 7.7%). We believe that the high rate of respondents reporting that they do not work with data protection is due to the recent entry into force of the LGPD in Brazil in August 2020. Of the ten respondents who do not work with data protection, eight are Brazilian. The others are Pakistani and Canadian and act as professors.

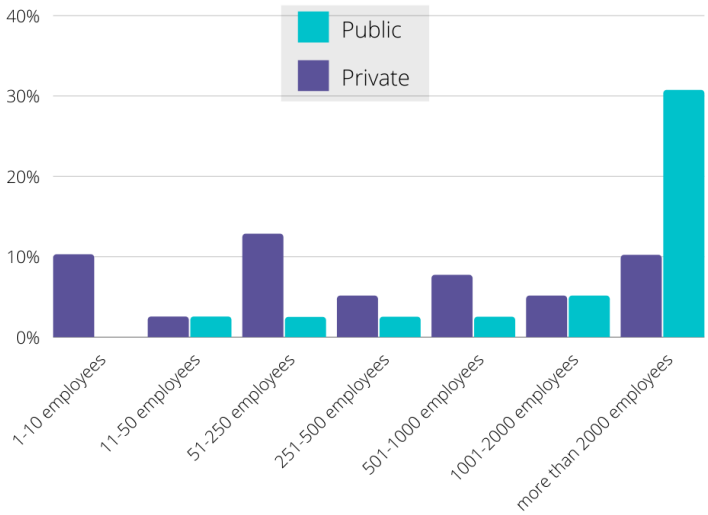
Figure 22 contains the distribution of the respondents according to the size of the organizations working, considering the number of employees and type of organization (public or private). We consider small companies up to 50 employees (5; 12.8% private, 1; 2.5% public); medium, up to 500 employees (7;17.9% private, 2; 5.1% public); large, up to 2000 employees (5; 12.8% private, 3;7.7% public); very large, more than 2000 employees (4;10.2% private, 12;30.7% public). Assuring a good coverage of all the possible organization sizes.

Figure 21 – Demographics: Experience in the software industry, and data protection projects



Source: The author (2021).

Figure 22 – Demographics: public or private, and size of company



Source: The author (2021).

About practice area from the companies, the frequency informed was Software, IT (14;35.9%), Education (10; 25.6%), Legal (4; 10.28%) and Other (E-commerce, Government, Military, Telecommunications, Games, Banking/Finance, Industry (application domain), Healthcare/Medical) (11; 28.2%).

From the demographic questions, we characterized the participants in subgroups considering the *maximum degree*, *Experience in the software industry*, *experience in data protection*

projects, *size of organization*, *type organization* (public or private), and *practice area*. We consider the answer options for each survey question as variables. Then, we performed cross-tabulation between variables and subgroups. Furthermore, analyzed the relationship between them using Fisher's exact test, Mann-Whitney or Kruskal-Wallis test, and post-hoc Dunn's test of multiple comparisons based on rank sums with Bonferroni correction. To facilitate the reading and location of the survey questions in the text, we boldly highlight the statement of the questions.

The following sections present the analysis of the questionnaire to answer the survey research questions. The hypotheses for the survey's questions presented below are:

H_0 - There is no association between the subgroups and the variables.

H_1 - There is an association between subgroups and the variables.

5.5.1 RQ1. What are the current practices for specifying legal requirements that the companies use in their daily work?

We want to understand what techniques are used in practice to specify legal requirements. For this, we define propositions that we want to verify. Each proposition was represented in the questionnaire as an answer option for the participants. Due to space constraints, we will not present all propositions (which can be found in full in the supplementary material (NETTO; SILVA, 2021)). Table 8 presents the propositions for the survey question 10 (represented by S (referring that the proposition is from the survey study) followed by the proposition number).

Table 8 – Propositions about question 10

nº	Propositions
S. P1	Legal requirements are specified via structured list of requirements
S. P2	Legal requirements are specified via informal text or plain text
S. P3	Legal requirements are specified via use case
S. P4	Legal requirements are specified via user stories
S. P5	Legal requirements are specified via prototypes

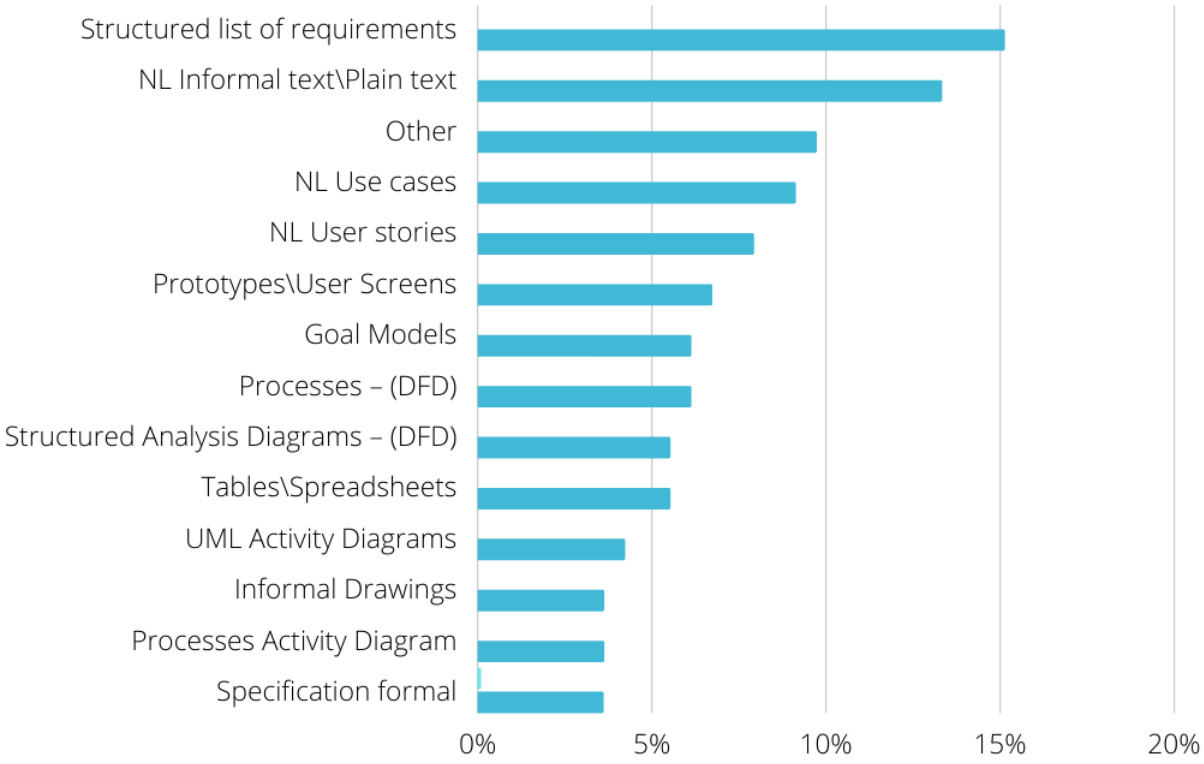
Source: The author (2021).

We asked participants **how they specify legal requirements**, in this question participants can select more than one alternative (multiple-choice question). According to the results, the Figure 23 shows (25; 15%) used *Structured list of requirements*, followed by *informal text or*

plain text (22; 13%), *use cases specification* (15; 9%), *user stories* (13; 7.9%), and *prototypes or user screen* (11; 6.7%).

Whereas *informal text or plain text*, *use case specification* and *user stories* are expressed in Natural language, the largest share of responses (50; 29.9%) write requirements in Natural Language. We categorize the alternatives as *Other* (16;9.7%) selected less than five times (*BPMN, graphical notations, use case diagrams, and entity relationship diagram (ERD)*).

Figure 23 – Approaches to specify legal requirements



Source: The author (2021).

Analyzing Figure 23 and the propositions presented in Table 8, it is possible to verify that they correspond to the five main approaches used in practice to specify legal requirements.

The hypotheses for the tenth question are:

H_0 - There is no association between the subgroups and the approaches to specifying requirements.

H_1 - There is an association between subgroups and approaches to specifying requirements.

We cross-tabulated between the subgroups (*Maximum degree*, *Experience in the Software industry*, *Experience in Data protection*, *Size organization*, *Type organization*, and *Practice area*) and each approach selected in the tenth question.

We used Fisher's exact test, indicated for small samples when fewer than 20 observations were in the model or fewer than five counts in a cell. Pearson's Chi-squared test was also used for variables with a frequency greater than five or more than 20 observations. The significance level used in all analyzes was 5%.

We analyzed the *p*-value for each approach from specify legal requirements and the subgroups *Maximum degree* and *Practice area*. As the *p*-value is larger than .05 for all combinations of approaches to requirements specification and subgroups, we can reject the alternative hypothesis and accept the null hypothesis. Thus, it is possible to state that there is no relationship between the degree (*Maximum degree*) and the legal requirements specification approach selected by the participant. There is also no relationship between the company's practice area (*Practice area*) and the choice of techniques used to specify requirements.

For the other subgroups (*Experience in the software industry*,*experience in data protection projects*, *size of organization*,*type of organization (public or private)*), Fisher's exact tests led to *p*-values that were smaller than 0.05, which means that there is a statistically significant relationship between the two variables. Fisher's exact test only provides the answer to whether or not the variables are correlated. The intensity of this relationship measure, which varies from 0 (absence of association) to 1 (robust association) to Contingency coefficient, phi (-1 (negative association) to 1 (1 positive association)) and Crámer V^2.

There is a statistically significant association between the *Type organization* (public or private) and the *Structured list of requirements* approach whose result of Fisher's exact test (*p*-value = .02367) present the *p*-value < 0.05 (see Table 9). The intensity of this relationship was verified using Phi-Coefficient, Cramer's V and Contingency coefficient. Figure 24 shows the frequency, the chi-square value, the intensity test result, in addition to the result of the Fisher's exact test. We obtained it for *Structured list of requirements* (Phi-Coefficient = 0.379; Contingency Coef. = 0.355; Cramer's V = 0.379 (see Figure 24). As the value of Cramer's V = .379 indicates no strong association between the variables.

Then, the M-W test was performed for subgroups with only two independent samples. In all cases the *p*-value is less than .05 (*Structured list of requirements* (*p*-value = .02024). So we reject the null hypothesis and accept the alternative hypothesis, that is, there is an association between the *Type organization* (Public or Private) and approaches to requirements

Figure 24 – Frequency and inference test results to structured list of requirements

Type_organization	*Structured_list_of_requirements* 0	1	***Total***
Private Sector	4 19 %	17 81 %	21 100 %
Public Sector	10 55.6 %	8 44.4 %	18 100 %
Total	14 35.9 %	25 64.1 %	39 100 %

χ^2=4.139 · df=1 · Cramer's V=0.379 · Fisher's p=0.024

Source: The author (2021).

specification *Structured list of requirements*.

We conducted the analysis for *Natural Language - User stories*, from Fisher's exact test (p-value = .008192) is less than .05 (see Table 9). The intensity of this relationship indicates that there is no strong association between the variables.

Figure 25 – Frequency and inference test results to Natural Language - User stories

Type_organization	*NL_User.stories* 0	1	***Total***
Private Sector	10 47.6 %	11 52.4 %	21 100 %
Public Sector	16 88.9 %	2 11.1 %	18 100 %
Total	26 66.7 %	13 33.3 %	39 100 %

χ^2=5.688 · df=1 · Cramer's V=0.436 · Fisher's p=0.008

Source: The author (2021).

Then, the M-W test was performed to *Type organization* and *Natural Language - User stories* whose p-value = .007515 (see Table 9). So we reject the null hypothesis and accept the alternative hypothesis, that is, there is an association between the *Type organization* (Public or Private) and approach to requirements specification *Natural Language - User stories*.

There is also a statistically significant association between company type (public or private) and the *Prototypes/User Screens*, because the Fisher's exact test (p-value = .03749)

it's smaller than .05 (see Table 9). The strength of the relationship indicates that there is an association, but it is not very strong. Mann-Whitney (M-W) test was performed to *Type organization* and *Prototypes/User Screens* (p-value = .03156)), so we reject the null hypothesis and accept the alternative hypothesis, that is, there is an association between the *Type organization* (Public or Private) and approaches to requirements specification *Prototypes/User Screens*.

Table 9 – Results of M-W test to Type organization (Public or Private) had a significant effect on questionnaire responses

Subgroups	**Variable**	**Fisher's (p-value)**	**Phi**	**Conting.**	**Crammer**	**M-W test (p-value)**
Type Co.	Structured list of requirements	.02367	0.379	0.355	0.379	.02024
Type Co.	NL User Stories	.008192	0.436	0.4	0.436	.007515
Type Co.	Prototype	.03749	0.352	0.332	0.32	.03156

Source: The author (2021).

For subgroups with three or more categories, we perform the Kruskal-Wallis (Kruskal-Wallis (K-W)) test to check if there is a difference between the groups. However, the K-W test only indicates a relationship between the subgroup and independent variable, but it does not examine which subgroups are statistically significant. Therefore, the post-hoc Dunn's test was carried out with Bonferroni's adjustment for the multiple comparisons method.

When we ran Fisher's test, we identified that there is a statistically significant correlation between the (*Experience in Data protection*) and the use of the *Natural Language - Informal text/Plain text approach* (see Table 10). Then, we verified the strength of the association (Contingency Coeff.= .51 and Cramer's V = .593) and performed the K-W test (p-value = .02003), which indicated significance. So, the K-W post-hoc Dunn's test performed with Bonferroni correction showed that this association is not statistically significant. Thus, it is not possible to reject the null hypothesis. Therefore, there is no relationship between the variables *Experience in Data protection* and *Natural Language - Informal text/Plain text*.

Analyzing *Prototypes* technique and *experience in the software industry* (Fisher's Exact Test p-value = .001048; Pearson's Chi-squared test p-value = .003081). When we verified the strength of the association (Phi-Coefficient = NA, Contingency Coeff. = .539 and Cramer's V = .64) and performed the K-W test (p-value = .003694) with Bonferroni correction, we

identified that the test is statistically significant, as the value of p for multiple comparisons is still less than .05 (p-value = .0033717) for the groups *Between 3 and 5 years* and *More than 11 years* (see Table 10). Therefore, we reject the null hypothesis and accept the alternative hypothesis. There is an association between *experience in the software industry* (*Between 3 and 5 years* and *More than 11 years*) and Prototype techniques to specify requirements.

Table 10 – Results of K-W test to size company had a significant effect on responses

Subgroups	**Variable**	**Fisher's (p-value)**	**K-W test (p-value)**	**Groups sign. differ**	**Dunn's (p-value)**
Exp. Software industry	Prototypes	.001048	.003694	(Between 3-5 years) and (More than 11 years)	.0033717

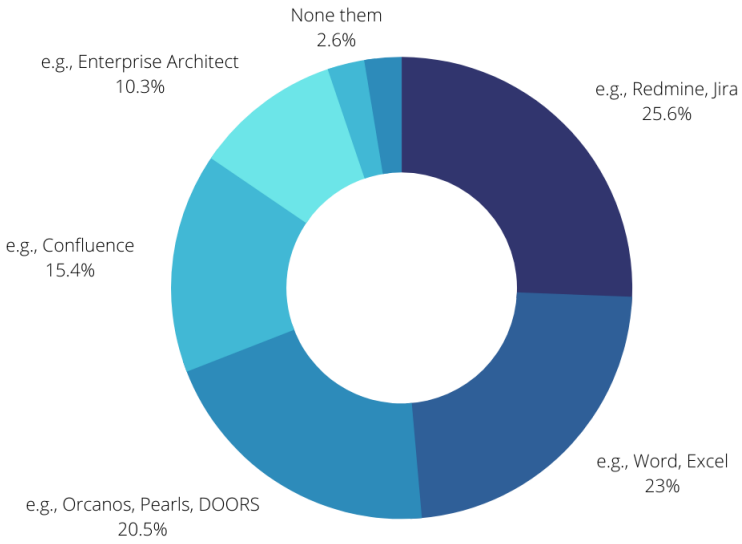
Source: The author (2021).

Through a single-choice question, participants answered **which tools are used to specify legal requirements**.

Figure 26 shows that most participants (10;25.6%) use an *Application Lifecycle Management or Issue Tracking Systems (e.g., Redmine, Jira)*, followed by *Spreadsheets or documents (e.g., Excel, Word)* (9;23.0%), and *A purpose-specific tool for requirements management (e.g., Orcanos, Pearls, DOORS)* (8; 20.5%). Only one participant used the "Other" option (2.6%) and nominated the *P-store methodology* (ANSARI et al., 2021). In this question, the median and mode are 3, equivalent to the alternative *Application Lifecycle Management or Issue Tracking Systems (e.g., Redmine, Jira)*. We performed the cross-tabulation for each subgroup and the variable referring to the tool used by the participants to specify legal requirements. Fisher's exact test had a value greater than .05 in only one case (experience in privacy or data protection, p-value = .7483). For the others, the value was always lower. Then we performed the K-W test for each subgroup, and in all cases, the p-value was equal to or greater than .05. Even so, we performed Dunn's post-hoc test with Bonferroni correction and, for all subgroups, there is no statistically significant relationship between the variables and subgroups.

We used a multiple-choice question to determine the **problems affecting legal requirements engineering projects** (Table 14), considering the personal experiences of respondents. This question was based on the factors that emerged from the interviews (NETTO; SILVA; ARAÚJO, 2021) and in NaPiRE initiative (FERNÁNDEZ et al., 2017). The problems they face are interrelated and deal with ambiguity in legal requirements (AMB), legal compliance

Figure 26 – Tools used to specify legal requirements



Source: The author (2021).

(COMP), and legal requirements specification (SPEC). Analyzing the answers show in Table 14, ordered from top to bottom according to the frequency by the responses, we no noticed a bigger difference between the frequency of the problems. Table 11 presents some propositions for the survey question 12.

Table 11 – Propositions about question 12

nº	Propositions
S. P6	Lack of training on data protection regulation affects projects with legal requirements
S. P7	Constants changes in the law make legal compliance difficult
S. P8	Lack of collaboration between software engineers and lawyers affects projects with data protection requirements
S. P9	Communication flaws between developers and the customers affects projects with data protection requirements
S. P10	There is no one from the Legal Support inside the development team affects projects with data protection requirements
S. P11	Law's entry into force is concise (time-boxing) and makes it infeasible to implement some features that affect projects with data protection requirements

Source: The author (2021).

Analyzing Table 14 and the propositions presented in Table 11 the five main problems, according to the participants, are *Lack of training on data protection regulation* (20; 6.3 %),

Lack of collaboration between software engineers and lawyers (19; 6.01%), *Communication flaws between developers and the customer* (18; 5.69%), *Changing goals, business process, and/or requirements* (17; 6.37%), *Incomplete and/or hidden requirements* (17; 6.37%), and *Inconsistent requirements* (17; 6.37%). Thus, propositions **S. P7**, **S. P10** and **S. P11** are not among the main problems of Legal Requirements Engineering mentioned by the participants.

The hypotheses for Question 12 are:

H_0 - There is no association between the subgroups and the problems in legal requirements engineering projects.

H_1 - There is an association between subgroups and the problems in legal requirements engineering projects.

We cross-tabulated between the subgroups (*Maximum degree*, *Experience in the software industry*, *Experience in Data protection*, *Size organization*, *Type organization*, and *Practice area*) and each problem selected in the 12th question. As for the previous question, we used Fisher's exact test, Pearson's Chi-squared test. The significance level used in all analyzes was 5%. For subgroups with three or more categories, we perform the Kruskal Wallis (K-W) test. The post-hoc Dunn's test was carried out with Bonferroni's adjustment for the multiple comparisons method.

It analyzed the *p*-value for each problem in legal requirements engineering projects and the subgroups. As the *p*-value is more extensive than .05 for all combinations of approaches to requirements specification and subgroups, we can reject the alternative hypothesis and accept the null hypothesis. Thus, it is possible to state that there is no relationship between the *Maximum degree*, *Experience in the software industry*, *Experience in Data protection*, *Size of organization*, and *Practice area* and the problem in legal requirements engineering projects selected by the participants.

For the subgroup *Type organization*, and the variable *Lack of collaboration between software engineers and lawyers* the Fisher's exact test (p-value $= .05616$) present the p-value $< .05$ (see Table 12). There is a statistically significant association Then, the test was performed to verify the intensity of the relationship.

Figure 24 shows the frequency, the chi-square value, the intensity test result, in addition to the result of the Fisher's exact test., which means that there is a statistically significant relationship between the two variables. Fisher's exact test only provides the answer to whether or not the variables are correlated. The intensity of this relationship measure, which varies from 0 (absence of association) to 1 (robust association) to Contingency coefficient, phi and

Crámer V^2.

The intensity of this relationship was verified using Phi-Coefficient, Cramer's V, and Contingency coefficient (see Table 12). The value of Cramer's V = .332 indicates no strong association between the variables.

Figure 27 – Frequency and inference test results to lack of collaboration between software engineers and lawyers

Type_organization	*SE_lawyers* 0	1	***Total***
Private Sector	14 66.7 %	7 33.3 %	21 100 %
Public Sector	6 33.3 %	12 66.7 %	18 100 %
Total	20 51.3 %	19 48.7 %	39 100 %

χ^2=3.080 · df=1 · Cramer's V=0.332 · Fisher's p=0.056

Source: The author (2021).

Then, the Mann-Whitney (M-W) test with continuity correction was performed, the *p*-value is less than .05 (*p*-value = .04204), see Table 12. So we reject the null hypothesis and accept the alternative hypothesis, that is, there is an association between the *Type organization* (Public or Private) and the problem *Lack of collaboration between software engineers and lawyers*.

There is also a statistically significant association between *Type of organization* and *Communication flaws between developers and the customer*, *Incomplete and/or hidden requirements* and *Inconsistent requirements*.

After performing the hypothesis tests for all the variables presented in Table 12, we reject the null hypothesis and accept the alternative hypothesis, that is, there is an association between the *Type organization* (Public or Private) and the variables.

Traceability between software requirements and snippets of legislation is essential for projects in the field of legal requirements. Thus, participants answered **considering the management of traceability among requirements, legal regulations, and document specification, which of the following scenarios best describe the practice in your projects**. Table 13 presents some propositions for the survey question 13. *Trace references are stated inside the requirements, specification and legal regulation artifacts* (11; 28.2%), followed by

Table 12 – Results of M-W test to type of company had a significant effect on responses

Subgroups	Variable	Fisher's (p-value)	Phi	Conting.	Crammer	M-W test (p-value)
Type Co.	Lack of collab. between SE and lawyers	.05616	0.332	0.315	0.332	.04204
Type Co.	Communication flaws between "devs" and the customer	.003869	0.484	0.436	0.484	.002997
Type Co.	Incomplete and/or hidden requirements	.	0.	0.	0.	.04548
Type Co.	Inconsistent requirements		0.	0.	0.	.008301

Source: The author (2021).

A tool is used to record artifacts and trace links (10;25.6%). What caught our attention is that 20.5% of respondents do not know how the traceability between the legal text and the company's requirements happens.

Table 13 – Propositions about question 13

nº	Propositions
S. P12	Trace references are stated inside the requirements, specification and legal regulation artifacts
S. P13	A tool is used to record artifacts and trace links

Source: The author (2021).

Figure 28 shows the percentages of the alternatives very similar to each other:

We performed the cross-tabulation for each subgroup and the variable referring to traceability management among requirements, legal regulations, and document specification. All Fisher's exact test had a *p*-value greater than .05. Nevertheless, we performed the K-W test for each subgroup, and in all cases, the *p*-value was equal to or greater than .05. Even so, we performed Dunn's post-hoc test with Bonferroni correction and, for all subgroups, there is no statistically significant relationship between the variables.

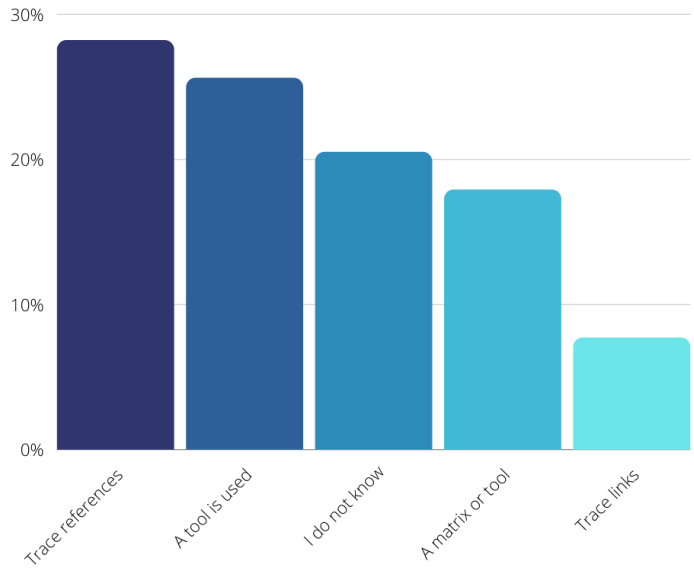
We asked the participants, through a multiple-choice question, **who is responsible for validating the specification of legal requirements**. Figure 29 shows the 96 responses with the following frequency: Customer, Development team members, Data Protection Officer

Table 14 – Problems in legal requirements engineering projects

Problems	**Frequency**	**Pct. of Resp.**
Lack of training on data protection regulations	20	6.3%
Lack of collaboration between software engineers and lawyers	19	6.01%
Communication flaws between developers and the customer	18	5.69%
Changing goals, business processes, and/or requirements	17	5.37%
Incomplete and/or hidden requirements	17	5.37%
Inconsistent requirements	17	5.37%
Standardizing terminology between law, engineering, and business	16	5.06%
Communication flaws within the project development team	15	4.74%
Compliance requirements are purposefully expressed in general terms, omitting implementation-specific details.	13	4.11%
Difficulty understanding domain-specific terms	13	4.11%
Lack of traceability between requirements and legal text	12	3.79%
There is no one from the Legal Support inside the Development Team	12	3.79%
A weak relationship between Customer and project lead.	11	3.48%
The developer may make their wrong interpretation	11	3.48%
Identify the regulations relevant to its specific system	11	3.48%
Interpreting the regulation and translating it into implementable requirements	11	3.48%
Unclear or unmeasurable non-functional requirement	11	3.48%
Constant changes in the law make legal compliance difficult	10	3.16%
Weak knowledge about the customer's domain	10	3.16%
Cross-reference among legal/regulatory documents	9	2.84%
Insufficient support by customer	9	2.84%
Underspecified requirements that are too abstract and allow for various interpretations	9	2.84%
Weak access to customer needs and/or (internal) business information	7	2.21%
Law's entry into force is concise (time-boxing) and makes it infeasible to implement some features.	6	1.89%
Insufficient support by the project lead.	5	1.58%

Source: The author (2021).

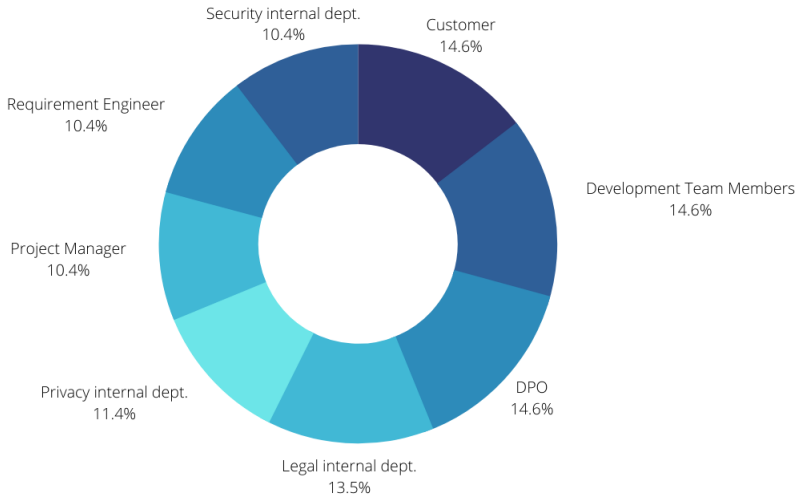
Figure 28 – Frequency of traceability management strategies



Source: The author (2021).

(Data Protection Officer (DPO)) (14; 14.6% each). Followed by Legal internal department (13; 13.5%) and Privacy internal department (11; 11.4%). We did not find any statistically significant relationship between variables and subgroups.

Figure 29 – Responsible for validating the legal requirements specification



Source: The author (2021).

We asked the participants to detail, in a non-mandatory open question, **how is the process, in the company, of validating the specification of legal requirements**. We

had a low amount of responses (14;34.9%). Nevertheless, analyzing respondents' experience, (7;53.8%) have more than 11 years of experience in the software industry, (3;23%) has between 6-10 years, and the others three have 1-5 years experience. The practice areas are *Education* (5;38.4%), *Software, IT* (4;30.7%), the others are *Legal* (2), Government/Military and Telecommunication. We identified that they are experienced professionals and from diversified areas of expertise.

Therefore, the procedures for validating the legal requirement specification presented by the participants occur with a lawyer or legal, Product Owner (PO), DPO, development team, project manager, customer, and quality analyst. One participant indicated that follow the P-STORE Methodology (ANSARI et al., 2021) for efficient privacy requirements elicitation and validation. Another participant cites that [validation is]"*decentralized at first (clients feedback). Recently, there is the creation of a committee for dealing with the LGPD*".

Validation of the specification of legal requirements takes place in three different ways, according to the participants:

> *internal procedures depend on the data handled. Any new business proposal or unforeseen cases arise. A meeting between the stakeholders and a lawyer (who acts as a DPO) can validate and describe the impact document. It is a Legal, DPO, and development team co-work.*
>
> *After analysis and advice from the Legal department, the project manager with the development team executes the validation process. Today the process does not occur systematically, and it is a vague and diversified process across many departments. It follows the traditional systems engineering validation process just observing business rules.*
>
> *The requirement is created by the client, PO, Project Manager, or legal department. After developing the requirement and validating the development team and quality sector. The requirement is validated by the Quality Assurance (QA), PO, the client, and, when necessary, the legal sector validates it.*

5.5.2 RQ2. What are the current practices from IT professionals towards specifying software legal requirements with reduced ambiguity in their daily work?

We used a multiple-choice question to determine the **sources or resources of information/knowledge used to solve or reduce ambiguity in legal requirements** considering respondents' personal experiences. If the alternative was not found in the list, the participants could add any missing value with an open field value option *"Other"*. Table 15 presents some propositions for the survey question 16.

Table 15 – Propositions about question 16

nº	Propositions
S. P14	Discussion between team members helps to solve and/or reduce ambiguity in legal requirements
S. P15	Consulting/reading Laws regulatory sources helps to solve and/or reduce ambiguity in legal requirements
S. P16	Consulting internal Legal sector helps to solve and/or reduce ambiguity in legal requirements
S. P17	Knowledge of the customer's domain helps to solve and/or reduce ambiguity in legal requirements
S. P18	Direct contact with the customer involved with the project helps to solve and/or reduce ambiguity in legal requirements
S. P19	Information from experienced team members helps to solve and/or reduce ambiguity in legal requirements

Source: The author (2021).

In Table 16, the items that were nominated less than five times by participants were categorized into *Other* (i.e., personal experience, weekly meeting, and friends outside the company). Analyzing the answers shown in Table 15 and Table 16, ordered from top to bottom according to the frequency, the **S. P19** proposition is not among the most cited by the participants,therefore, proposition **S. P19** will be replaced. Then, two new propositions appear in the table 30, *Analysis of case law* (new **S. P19**) and *Ask for clarification from the Government Authority* (**S. P20**).

The hypotheses for Question 16 are:

H_0 - There is no association between the subgroups and the sources or resources used to solve or reduce ambiguity in legal requirements.

H_1 - There is an association between subgroups and the sources or resources used to solve or reduce ambiguity in legal requirements.

Table 16 – Sources of information/knowledge used to solve and/or reduce ambiguity

Sources/Resources	Frequency	Pct. of Resp.
Discussion between team members	25	10.72%
Laws regulatory sources	24	10.30%
Consulting internal legal sector	23	9.87%
Analysis of case law	21	9.01%
Ask for clarification from the Government Authority	13	5.57%
Customer domain knowledge	12	5.15%
Direct contact with the customer involved with the project	12	5.15%
Standards	12	5.15%
Ask for clarification to another company sector	11	4.72%
Information from stakeholders	11	4.72%
Organizational procedures	11	4.72%
Communities online	10	4.29%
Information from experienced team members	10	4.29%
The developer may make their wrong interpretation	11	3.48%
Scientific paper	10	4.29%
Books, blogs or white papers online	10	4.29%
Other	7	3.86%
It was reported by a security or privacy audit	6	2.57%
Information from managers	6	2.57%

Source: The author (2021).

Performing the hypothesis tests for each alternative selected by the participants through a cross-tabulation with each subgroup, we only identified a statistically significant relationship in three cases.

There is a statistically significant relationship between *experience in software projects* and the *Consulting internal legal sector* variable. The Table 17 shows the values for Fisher's Exact Test (p-value = .005435) and Pearson's Chi-squared test are less than the significance value. Therefore, we performed the K-W test and Dunn's post-hoc with Bonferroni correction (see Table 17). After adjustment, the p-value remained below .05 and indicated the significance level between the groups with experience from *3-5 years* and *6-10 years*. Analyzing the profile of the 23 respondents who selected this question, (11;47.8%) have more than 11 years of experience in software projects, and (8;34.8%) have between 6 and 10 years. The area of activity of companies is related to Software, IT (8;34.8%), Education (7;30.4%), and legal (3;13%).

Therefore, there is a relationship between the variables. Participants with more experience in software projects consult the company's legal department to assist them in interpret the ambiguity. There is no correlation between the area of expertise or the size of the company.

Table 17 – Results of K-W test to size company had a significant effect on responses

Subgroups	**Variable**	**Fisher's (p-value)**	**K-W test (p-value)**	**Groups sign. differ**	**Dunn's (p-value)**
Exp. in software proj.	Consulting internal legal sector	.005435	.01871	(Between 3-5 years) (Between 6-10 years)	.01317398
Exp. in data protec.	Laws regulatory sources	.02293	.0276	-	-
Exp. in software proj.	Analysis of case law legal sector	.02894	.07199	-	-

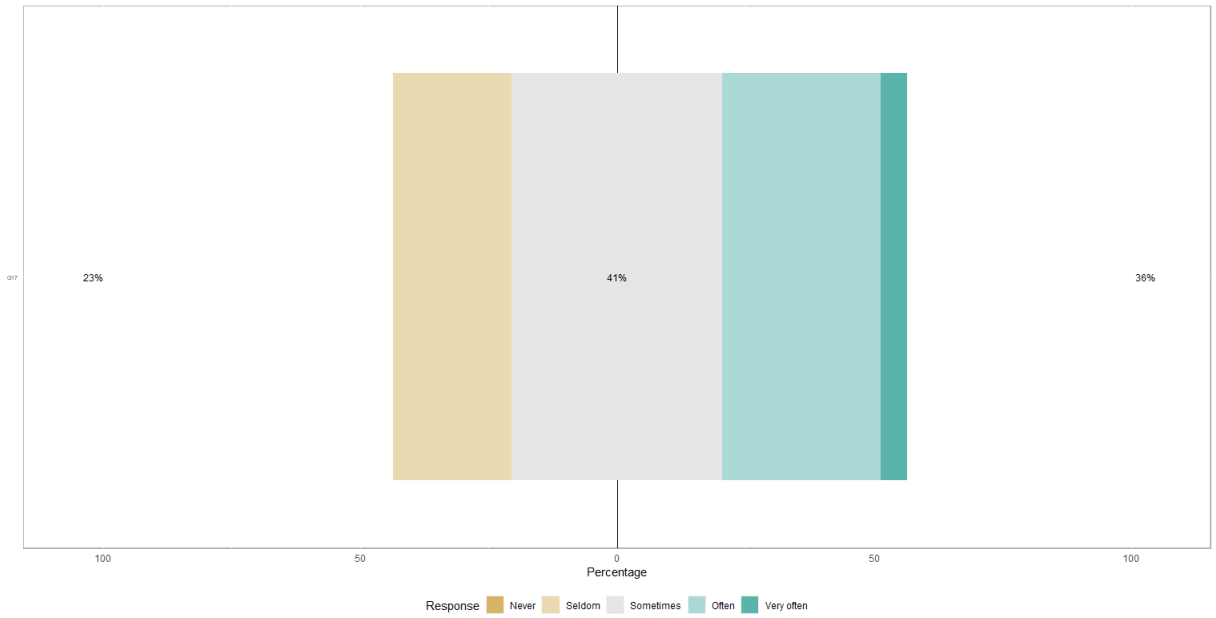
Source: The author (2021).

In the other two cases, there was an indication of a statistically significant relationship in the subgroup *experience in privacy or data protection* with the dependent variables *Laws regulatory sources* and *Analysis of case law*. Table 17 shows that both were analyzed by Fisher's Exact Test (getting a p-value $< .05$), as well as for Pearson's Chi-squared test (p-value $< .05$). Then we performed the K-W test and obtained a p-value less than the significance value, indicating that there is some relationship between the variables. Nevertheless, by performing Dunn's post-hoc with Bonferroni correction, the p-value was adjusted and was not considered to be significant in both cases. Thus, we failed to reject the null hypothesis; there is no relationship between the variables.

Using a five-point Likert scale (1-never, 2-seldom, 3-sometimes, 4- often, 5-very often), we asked the participants **How frequently do you find ambiguity in the legal requirements in your company's projects?**. The Figure 30 shows that the vast majority responded that they face ambiguous legal requirements *sometimes* (16;41%), followed by often (12;37.7%). Only (9;23.7%) said they deal with ambiguity in legal requirements seldom. The median and mean for this question is 3.

About the areas of activity of the companies, *Education* and *Software, IT* are the ones that stand out. Education (Sometimes(4;10.2%) and Often (4;10.2%)), Software, IT (Seldon (5;12.8%) and Often (5;12.8%)).

Figure 30 – How frequently do you find ambiguity in the legal requirements in your company's projects?



Source: The author (2021).

Analyzing the profile of nine respondents who selected *seldom*, we identified that (4; 44.4%) work with privacy or data protection for less than two years, (3;33.3%) have 3-5 years of experience in data protection, (2; 22.2%) do not work with this type of project. The companies are primarily private (6; 66.7%) and operate in the *Software (5; 55.5%), IT, Games, and Banking/Finance sectors*. Despite the respondents' little experience with privacy and data protection projects, there is no significant association between the professionals' experience and the identification of ambiguities.

We analyzed a statistically significant relationship between the dependent variable (answers to question 17) and all the independent variables (the subgroups obtained from the demographic questions). In only one case the p-value was less than .05, *Type organization*.

Table 18 shows the p-value in Fisher's exact test (p-value = .1033) and Pearson's Chi-squared (p-value = .09409 were marginal with respect to significance. So we decided to investigate, performing the Mann-Whitney test, because the *Type organization* variable has only two independent samples (public or private). The Mann-Whitney test obtained p-value = .03399. Thus, there is a relationship between the type's company (private or public) and the participants' frequency of encountering ambiguities in legal requirements.

We ask the participants to answer **how difficult it is to interpret ambiguity in legal requirements**. For this five-point Likert scale question (1-very difficult, 2-challenging, 3-

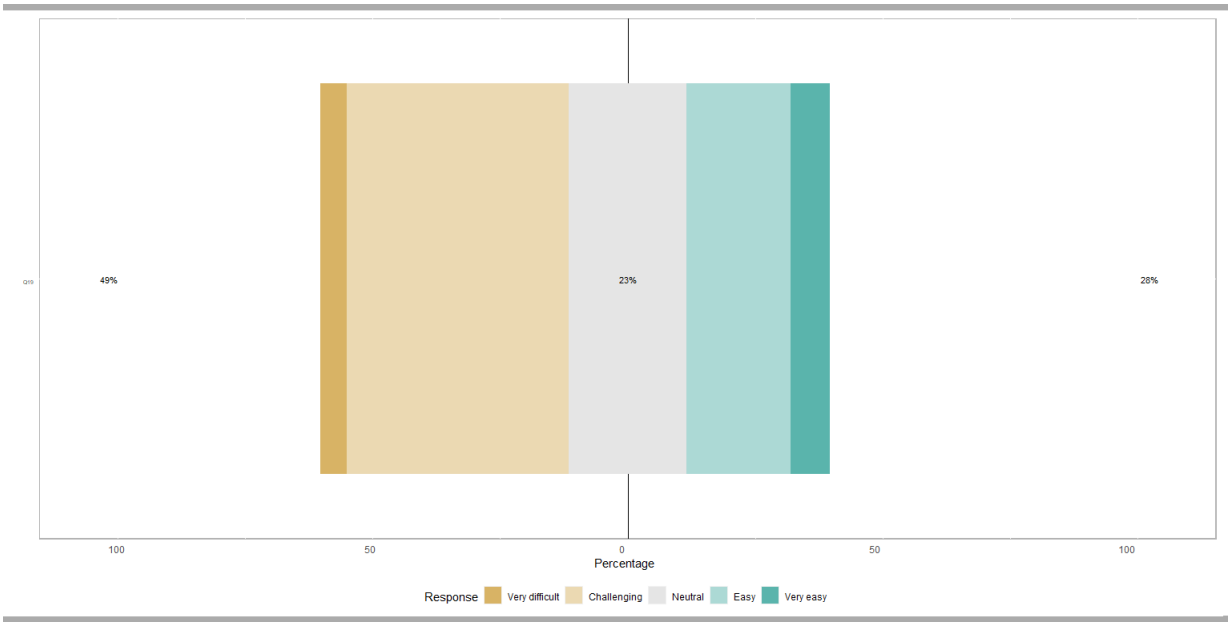
Table 18 – Results of K-W test to size company had a significant effect on responses

Subgroups	Variable	Fisher's (p-value)	Pearson's (p-value)	M-W test (p-value)
Type company	Participants encounter ambiguities in legal req.	.1033	.09409	.03399

Source: The author (2021).

neutral, 4-easy, 5-very easy), and the mode is 2. Figure 31 shows that 48.7% of respondents find it very difficult to interpret ambiguity in legal requirements. There is no statistically significant relationship between the subgroups and how participants perceive ambiguity in legal requirements.

Figure 31 – How difficult is it to interpret ambiguity in legal requirements?



Source: The author (2021).

We asked in a multiple-choice question **what techniques are used to reduce the ambiguity in legal requirements specification**, considering personal experiences from participants. Table 15 presents some propositions for the survey question 18. We got 164 answers to this question. Table 20 presents the techniques most selected by the participants. Those that were selected less than five times, we categorized as "Other" (31; 18.9 %), i.e., *Integral Lawyer inside of the development team*, *Object-Oriented & UML Based Solutions* (5; % each); *Natural Language Process (NLP) Based Solutions*, *Ontology-Based Solutions* (4; 2.4% each).

Boilerplates Based Solutions (sections of code repeated in multiple places with little to no variation) and *Controlled Languages Based Solutions* (3; 1.8% each).

Table 19 – Propositions about question 18

nº	Propositions
S. P21	Basic knowledge about law for software engineers helps to reduce ambiguity in legal requirements specification
S. P22	Training in ambiguity identification techniques helps to reduce ambiguity in legal requirements specification
S. P23	Data dictionary for all domain-specific definitions and acronyms/Glossary helps to reduce ambiguity in legal requirements specification
S. P24	Delegation of a person for tracing laws and legal regulations helps to reduce ambiguity in legal requirements specification
S. P25	Identification of relevant laws and legal regulations and their analysis performed by lawyers helps to reduce ambiguity in legal requirements specification

Source: The author (2021).

Analyzing the answers show in Table 20 and Table 19, ordered from top to bottom according to the frequency, propositions **S. P21** to **S. P25** are the most mentioned by the participants. We analyze, and there is no statistically significant relationship between subgroups and the techniques used to reduce ambiguity.

In order to verify if **practitioners use solutions developed in the academy to deal with ambiguity in legal requirements in the organization**, we created a multiple-choice question. The tools that were selected by the participants are Requirements Engineering Specification Improver (RESI) (3; 7.31%), Context Knowledge & Concepts Ontology (CKCO) (2; 4.88%), Object-Oriented Visualization (2; 4.88%) , Quality analyzer for requirement specification (QuaARS) (1; 2.43%), Word Sense Disambiguation (WSD) (1; 2.44%). The only participant who selected the *I do not use any of them* option and indicated the tool they use for handling ambiguity in legal requirements was Defend Secure Tropos (SecTro) from the Defend European Union (EU) Project (1; 2.43%). Thirty-one participants (75%) do not use any of them.

This finding corroborates Proposition **I. P8** (Academic resources (methodologies or tools) for Privacy Requirements Elicitation and Specification are not used by industry) presented in (NETTO; SILVA; ARAÚJO, 2021) that none of the interviewees mentioned the use of any similar methodology or tool. As well as in the work of Sirur, Nurse, and Webb (SIRUR; NURSE; WEBB, 2018) none of the companies used academic research in their compliance process, with most

Table 20 – Techniques used for reduce ambiguity

Techniques	Frequency	Pct. of Resp.
Basic knowledge about law for software engineers	14	8.5%
Data dictionary for all domain-specific definitions and acronyms\Glossary	12	7.3%
Delegation of a person for tracing laws and legal regulations	12	7.3%
Identification of relevant laws and legal regulations and their analysis performed by lawyers	12	7.3%
Agile requirements specification techniques (i.e., user stories)	10	6.1%
Training in ambiguity identification techniques	9	5.5%
Transformation of legal regulations to legal requirements, iteratively, in cooperation between lawyers and software engineers	9	5.5%
Training in the regulatory domain for the development team	8	4.9%
Basic knowledge about requirements engineering for lawyers (understand the SDLC briefly)	8	4.9%
Critical requirements through vulnerabilities in legal violations of existing software systems discussed in administrative and case law	8	4.9%
Inspections Based Solutions	7	4.3%
Software Requirements Specification (SRS) should contain a section of legal requirements and complete specifications of system requirements related to law	7	4.3%
Reusable catalog of legal requirements that were derived from specific legal texts regarding security and personal data protection	7	4.3%
Integral Lawyer inside of the development team	5	3.0%
Object-Oriented & UML Based Solutions	5	3.0%
Others	31	18.9%

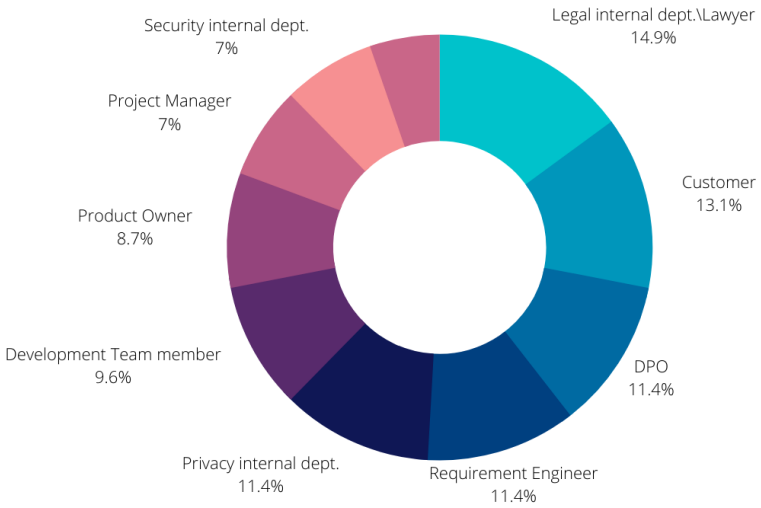
Source: The author (2021).

finding academic research not pragmatic or transparent enough for immediate use.

We asked the participants, through a multiple-choice question, **who is responsible for interpreting or resolving ambiguities in the legal excerpt**. Figure 32 shows the 114 responses with the following frequency: *Legal internal department or Lawyer* (17; 14.9%), *Customer* (15; 13.1%), *Data Protection Officer (DPO)*, *Requirements Engineer* and *Privacy internal department* (13; 11.4% each). Analyzing the relationship between the subgroups and

the variables, we did not find any statistically significant relationship. Thus, we cannot reject the null hypothesis.

Figure 32 – Responsible for validating the legal requirements specification



Source: The author (2021).

Participants responded in one open and non-mandatory question on how to solve an ambiguity in a legal excerpt to be represented as a software requirement. Of the ten answers, five respondents (50%) have more than 11 years of experience, (3;30%) have between 6 to 10 years of experience and work in companies of software, IT (5;50%) and education (3;30%). The ten answers are similar in that they all cite trying to achieve a unified understanding among the development team members, with discussion, documentation, and even searching for similar situations to support the understanding of the ambiguity. If this is not possible, they request the internal legal team, DPO, or security sector to interpret the legal text and solve the ambiguity. The internal Legal team replies with their opinion.

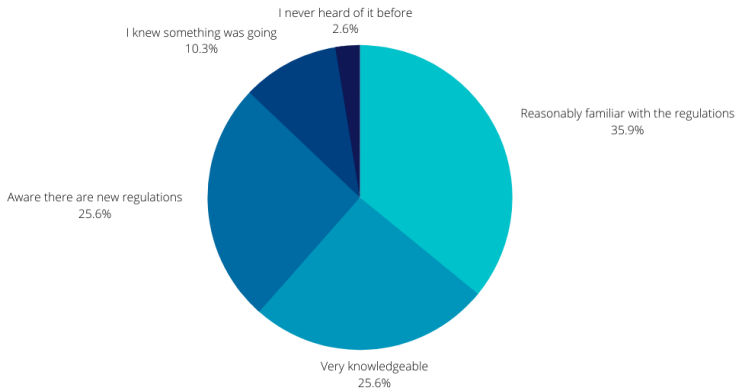
As a Software Engineering with more than ten years of experience in IT projects says: "*When the Project Manager does not have enough information, it looks for the client and the legal sector to be able to solve the doubts*". A detailed discussion and rigorous communication between the development team, the DPO, the legal department, and the client can help solve the ambiguity.

5.5.3 RQ3. What are the current practices towards achieving and verifying legal compliance of software requirements with data protection laws in their daily work?

Respondents answered how they characterize their knowledge about Personal Data Protection Laws in force in the country informed in question 1. For example, Brazilian General Law of Personal Data Protection, in Portuguese: Lei Geral de Proteção de Dados Pessoais (LGPD) or General Data Protection Regulation (EU GDPR). Figure 33 shows fourteen participants (35.9%) consider itself *reasonably familiar with the regulations but have a lot more to learn*. Ten participants responded *I am very knowledgeable about the Data Protection Law* and *I was aware there are new regulations and know some details* (25.6% each). Only one participant *I never heard of it before*, and four *I knew something was going on but do not know any details* (10.3%).

Analyzing the relationship between the subgroups and awareness regarding the Personal Data Protection Laws, we did not find any statistically significant relationship. Thus, we cannot reject the null hypothesis.

Figure 33 – Knowledge about the Personal Data Protection Laws

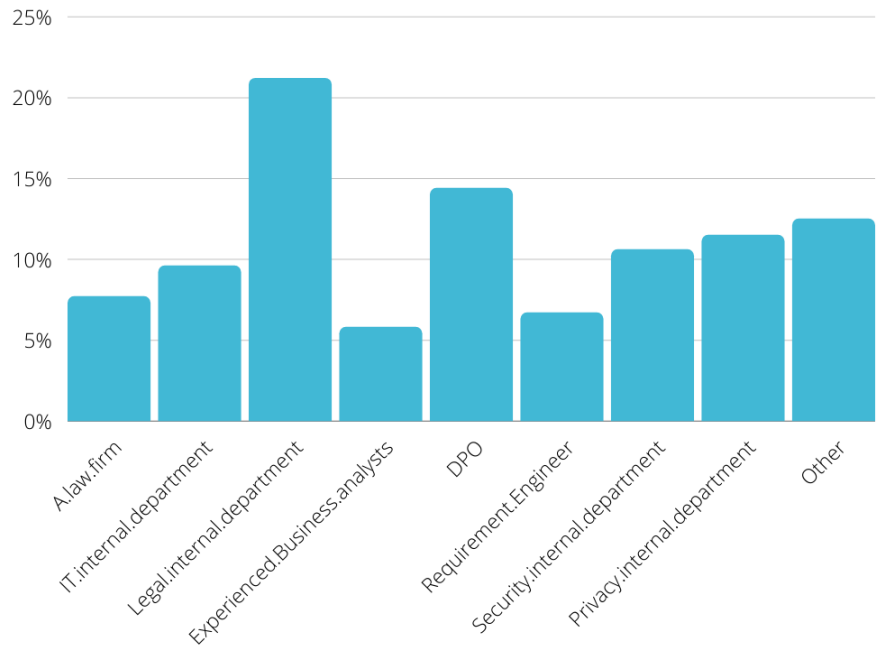


Source: The author (2021).

Participants responded to whom **provides support for your organization with Personal Data Protection Law (i.e., LGPD, GDPR, HIPAA, among others)**. Legal internal department is responsible for providing support in 22 companies (21.2%), followed by DPO

(15; 14.2%) and Privacy internal department (12; 11.5%) (see Figure 34). We categorized those selected less than five times as “Other" (13; 12.5 %).

Figure 34 – Support with Personal Data Protection Law



Source: The author (2021).

Analyzing the relationship between the subgroups and support for your organization with Personal Data Protection Law, we did not find any statistically significant relationship. Thus, we cannot reject the null hypothesis.

Participants answered, **given the company's current data privacy practices, what rate to comply with the Personal Data Protection Law**. Most of the participants affirm that the companies that work comply with the Personal Data Protection Law. *Our existing practices will satisfy some of the LGPD/GDPR, but we will need to make a few changes* (14; 35.9%), followed by *We are compliant in a few areas but need to make significant changes to be compliant* (9; 23.1%), and *We are figuring out who needs to be involved in putting a plan together* (4; 10.25%). Only 17.9% of companies *are already compliant and do not need to change*. And, 12.8% of companies are in non-compliance with the Data Protection Laws: *We have not started on our planning* (4; 10.25%), *We are not compliant at all* (1; 2.6%).

Analyzing the relationship between the subgroups and support for your organization with Personal Data Protection Law, we did not find any statistically significant relationship. Thus, there is a very low probability between subgroups and the rate of compliance with companies'

Personal Data Protection Law.

We asked participants how they handle changing legal requirements after the initial release. Because the legislation often undergoes updates and modifications, either through an amendment or repeal. Table 22 presents some propositions for the survey question 26.

Table 21 – Propositions about question 26

nº	Propositions
S. P26	Team-based discussion helps to deal with changes in legal requirements
S. P27	Regularly change the legal requirements specification helps to deal with changes in legal requirements
S. P28	Discussion with customers helps to deal with changes in legal requirements and to decide the best approach

Source: The author (2021).

Most participants *update our product backlog* (11; 28.2%). Followed by *We regularly change the legal requirements specification* and *We discuss with customers and decide the best approach* with (8; 20.5% each). For 17.9% (7) of the participants *Team-based discussion before the change*. And, only (5; 12.8%) *We only work with change requests*.

Analyzing the answers presented in the previous paragraph, proposition **S. P26** are not the most mentioned by the participants. Therefore, we will change proposition **S. P26** to *updating the product backlog helps to deal with changes in legal requirements*. Analyzing the relationship between the subgroups and how they deal with changing legal requirements after the initial release, we did not find any statistically significant relationship. Thus, there is a very low probability between subgroups and how to deal with legal requirements changes after the initial release.

IT companies must comply with relevant laws and regulations to avoid the risk of costly penalties, lost reputation, and brand damage resulting from non-compliance. Then, we asked participants **how the organization is managing compliance so far**. Most participants answered *Our organization has elected a DPO or similar Officer* (14; 35.9%). Followed by *Our Lawyer(s) say(s) that we are compliant* (9; 23% each). For 17.9% (7) of the participants *Several business processes were re-engineered due to Privacy requirements*. Moreover, only (3; 7.7%) *Management or Auditors have mandated improvements*. We categorized those that were selected less than five times (*I do not know*, *There is not and self-managed*) as “Other" (6; 15.3 %).

Analyzing the relationship between the subgroups and how the organization manages compliance, we did not find any statistically significant relationship. Thus, there is a very low probability between subgroups and how the organization manages compliance so far.

We asked participants **what training and practice should be provided regarding legal requirements and personal data protection laws**. Table 22 presents some propositions for the survey question 28.

Table 22 – Propositions about question 28

nº	Propositions
S. P29	Training and practice about Privacy Principles must be provided to the stakeholders in legal requirements and personal data protection laws
S. P30	Training and practice about Privacy by Design and/or Privacy by Default must be provided to the stakeholders in legal requirements and personal data protection laws
S. P31	Training and practice about legal requirements documentation must be provided to the stakeholders
S. P32	Training and practice about the organization's own internal privacy protocols must be provided to the stakeholders

Source: The author (2021).

Table 23 presents training and practices, frequency, and percentage. In this multiple-choice question, we received 247 responses. Analyzing Table 23 and the propositions presented in Table 22 the four main training and practice that must be offered, according to the participants, are *Privacy Principles (presented in GDPR Article5,and LGPD Article 6)* (18; 7.3%), *Privacy by design, and privacy by default* (18; 7.3%), *Laws and regulations related to the software's subject area to be developed (e.g.,patient record law, Privacy and Electronic Communications Regulation (e-Privacy), Information and Communication Technology (ICT) regulation)* (17; 6.9%), *Concerning data subjects' rights (presented in GDPR Articles 12-23, and LGPD Articles 17-22)* (16; 6.5%). Thus, propositions **P31** and **P32** are not among the most cited by participants.

Next, we ask **who should receive training** on the issues identified above. Figure 35 shows the percentage of all alternatives. Most subjects indicated that *All employees should have a basic understanding of privacy and information security* (33; 24.1%). Followed by Developers (27; 19.7%), Architects (19; 13.8%), and testers (15;10.9%) should be competent in the topics of data protection, secure coding, privacy, and security by design and by default. The options presented by those directly involved with the implementation and quality of the code

Table 23 – Training and Practices

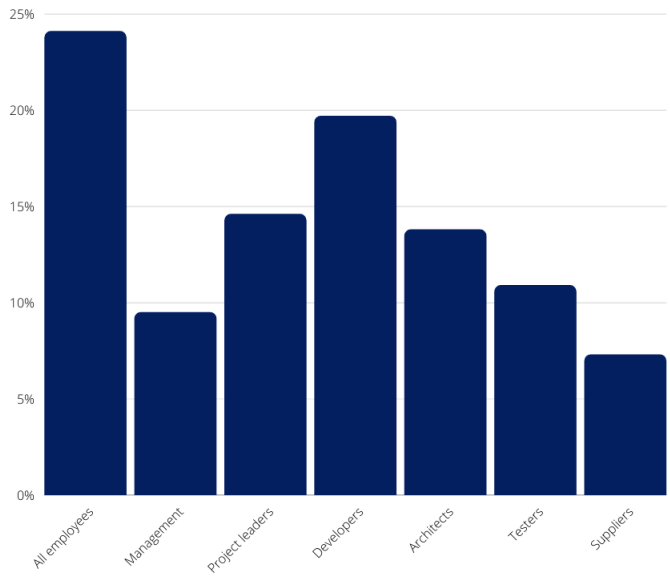
Training and Practices	**Freq**	**Pct.of.Resp**
Privacy Principles (presented in GDPR Article 5,and LGPD Article 6)	18	7.3%
Privacy by design, and privacy by default	18	7.3%
Laws and regulations related to the software's subject area to be developed (e.g.,patient record law, Privacy and Electronic Communications Regulation (e-Privacy), ICT regulation)	17	6.9%
Concerning data subjects' rights (presented in GDPR Articles 12-23, and LGPD Articles17-22)	16	6.5%
Conditions for consent (presented in GDPR Articles 7 and 8, and LGPD Article 8)	14	5.7%
Data protection impact assessment (DPIA) and prior consultation	14	5.7%
Documentating requirements	14	5.7%
Information Privacy Framework (e.g., ISO/IEC 29132:2017, ISO/IEC 27701:2019,National Institute of Standards and Technology (NIST) Privacy Framework)	14	5.7%
Mandatory business/sector/industry requirements and code of conduct	14	5.7%
Data Protection Officer (DPO), appointment, job descriptions, an overview of tasks codes of conduct, and certification	13	5.2%
Notification of personal data and information security breaches to the supervisory authority and data breach notification to the data subject	13	5.2%
The organization's own internal privacy protocols	13	5.2%
The lawfulness of processing (presented in GDPR Article 6, and LGPD Article 7)	12	4.8%
Processing special categories of personal data (presented in GDPR articles 9 and 10,and LGPD Article 11 and 14)	11	4.4%
The organization's information privacy requirements and guidelines	11	4.4%
On the duties of data controllers and data handlers (presented in GDPR Articles 24-43,and LGPD 37-40)	10	4.0%
Penalties and sanctions of Data Protection Laws (presented in GDPR Article 84, and LGPD 52-54)	8	3.2%
Records of the data processing activity	8	3.2%
Roles and organization relating to privacy	8	3.2%
I do not know	1	0.4%

Source: The author (2021).

had a higher percentage than the management roles. The *Project leaders should be competent in data protection topics by design and by default* (20; 14.6%), and *Management should be competent in assessing the impact and consequences of privacy, risk assessment, management responsibilities, and handling of risks to privacy* (13; 9.5%).

Few participants find it essential that suppliers participate in training related to privacy and data protection. *Suppliers should be competent in data protection by design and, by default, data processing agreements, incident response handling, and emergency response. The suppliers should have readable, standardized, and updated privacy documentation to comply with LGPD/GDPR* (10; 7.3%).

Figure 35 – Who should receive training?



Source: The author (2021).

Analyzing the relationship between the subgroups and support for the organization with Personal Data Protection Law, Fisher's exact test led to p-value = .01379 (see Table 24). This means that there is a statistically significant relationship between *Size company* and the variable *Project leaders should be competent in data protection topics by design and by default*. Therefore, there is a high probability that large (1000-2000) and medium (51-250) companies want *Project leaders* to receive training on privacy and data protection laws and should be competent in topics by design and by default.

So we decided to analyze this relationship more specifically and identified that ten subjects fit into these company size groups ((1000-2000) and (51-250)). They are Brazilian, 70% have between 1 to 2 years of experience in privacy or data protection. Six (60%) are of *Software IT*

Table 24 – Results of K-W test to size company had a significant effect on responses

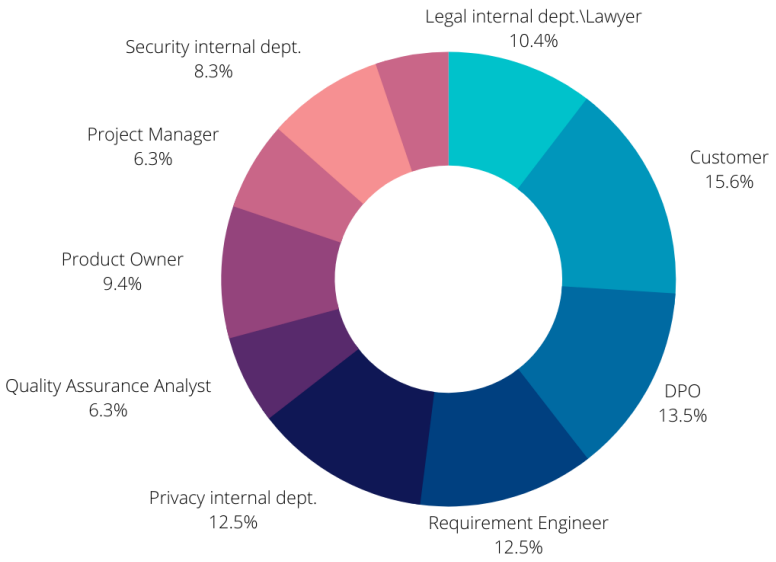
Subgroups	Variable	Fisher's (p-value)	K-W test (p-value)	Groups sign. differ	Dunn's (p-value)
Size company	Project leaders	.01379	.04343	(1000-2000) and (51-250)	.04657215

Source: The author (2021).

companies being five privates and one public. Of these, 90% also selected that *All employees should have a basic understanding of privacy and information security*.

We asked the participants, **in the Company, who is responsible for deciding whether the system complies with the legislation**. Figure 36 shows that the *customer* (15; 15.6%) is the most cited. Next are sectors or employees of the company itself as *DPO* (13; 13.5%), *Requirement Engineer* and *Privacy internal department* (12; 12.5 each), and *Legal internal department/Lawyer* (10; 10.4%). We did not find any statistically significant relationship between the subgroups and the variables.

Figure 36 – Responsible for deciding whether the system complies with the legislation



Source: The author (2021).

From the interviews, we identified that having the support of someone who knows the legal area is essential for a development team that deals with legal requirements. Thus, we ask **if an individual with legal knowledge participates inside the development team** (someone who has taken courses, training related to legislation, or a Bachelor of Law Degree). The answers show that (26; 66.7%) respondents did not have support a legal expert on the

development team. Similarly, (13; 33.3%) has a member of the development team with legal expertise. Participants who responded that they do not have the support of an individual with legal knowledge on the development team (26; 66.7%) are primarily from the *Software IT* industry (12; 46.2%).

Performing the analysis of respondents who mentioned having someone with legal knowledge as part of the development team, with sub-groups identified from demographic questions, we found that the vast majority (9;69.2%) work in large companies (up to 2000 The participants) or very large (more than 2000 employees). Employees from small (up to 50 employees) and medium companies (up to 500 employees) represent only (4;30.8%). Regarding the practice area of the participating companies, five companies in the Education area (38.5%), Legal (3;23.1%), E-commerce and Software, IT (2;15.4% each)There were no considerable differences between the Type organization for this question: Public companies (6; 46.2%) or Private (7; 53.8%).

We asked the participant to **explain the organization's procedure when software needs to comply with legislation**. Eight participants answered this open-ended question. The vast majority of respondents have more than 11 years of experience (5; 62.5%), two with 6 to 10 years of experience (25%) and (1;12.5%) with 3-5 years of experience in projects of software. 62.5% are from public companies and 37.5% private companies. Due to the low adherence to this question, we cannot affirm that these are the procedures presented in all public or private companies. However, some of the steps corroborate the findings of the interviews, presented in (NETTO; SILVA; ARAÚJO, 2021).

When software needs to comply with legislation, the organization makes a specification of compliance requirements. Then a search for information on regulation. Translate into requirements with support from the legal sector or legal experts. The data protection officer can sometimes review it. Then implement the requirements and validate them with the responsible sectors (or customers).

Considering the stages of the software development life cycle (SDLC), participants responded 32 times (33%) that verification of requirements in relation to legal compliance occurs in the *Requirement collection and analysis* stage. Followed by *Design* (18;18.5%), *Testing* (15; 15.5%), *Feasibility study* (10;10.3%) In this question, participants were able to select more than one alternative.

We asked the participants to detail, in a non-mandatory open question, **how to verify in the company that all legal requirements are compliant with the legislation**. We

had a low amount of responses, just six. Nevertheless, analyzing respondents' experience, (4; 83.4%) have more than 11 years of experience in the software industry, one has between 6-10 years, and one has 3-5 years experience. The roles are diverse: Professor (2), Security Engineer, Privacy expert, Software Engineer, and DPO. The practice areas are *Education*, *Software, IT*, *Banking/Finance*, and *Legal*.

In summary, the verification takes place through a "*Consultation with legal sector*", *the requirements engineering team ensures this*, *Reviews by the data protection officer*, and *Privacy, security, data governance tool*. The Security Engineer mentions: "*A set of best practices are followed. If questions arise, the Legal team is queried*", but does not mention what the *" are best practices"*. Moreover, the Software Engineer cites: "*The requirements are tested and validated by the quality sector. After that, the requirements are validated by the Legal department, customer, or other sector involved with the legal requirement.*"

We also asked the participants **when is it validated that all legal requirements are following the legislation** (multiple-choice question), **considering the stages of the software development life cycle (SDLC)**. Participants responded 30 times (30.1%) that verification of requirements in relation to legal compliance occurs in the *Requirement collection and analysis* stage. Followed by *Testing* (20; 20.6%), *Design* (14;14.4%), and *Coding* (9;9.3%).

Three participants who answered the previous open-ended question also answered how it is validated if all legal requirements are compliant with the law. In summary, the validation takes place through a "*Consultation with legal sector*" and *the customer and the expert ensure this*. The Security Engineer mentions: "*If questions arise, the Legal team is queried. For Terms and Conditions changes, the Legal team needs to review the changes*". Nevertheless, analyzing respondents' experience, all have more than 11 years of experience in the software industry.

5.6 SURVEY 2 - DEMOGRAPHICS DATA ANALYSIS

In this section, we describe the data analysis from (50;42.7%) complete responses in a total of 117 responses. Among the 117 responses, in 28 cases, only demographic questions were answered, and in 39 cases, the questionnaire was partly completed. Therefore, we decided to exclude all incomplete responses from our analysis.

We formally send 450 invitations; in addition to these invitations, we also post on social networks. The method steps were presented in Section 5.2. The 50 completed answers matched a response rate of 11.1%.

Demographic questions were used to filter and form sub-groups used to investigate variations in practitioners' perceptions. Since the survey was conducted online, practitioners from all over the world were able to participate. Some demographic questions were built into the survey design to confirm whether respondents belonged to the targeted population.

The participants were asked to inform their country, they indicated in Brazil (66%), followed by Canada and United Kingdom (6% each), Spain (4%), Bolivia, Chile, Germany, Iceland, India, Italy, Pakistan, Spain, Portugal, United States (2% each).

Concerning educational background. More than 50% have MSc (10; 20%) or PhD (15; 30%). Moreover, have Post-degree (14; 28%) and Degree (8; 16%). Which can be considered a good indicator of the quality of the information gathered in the questionnaire.

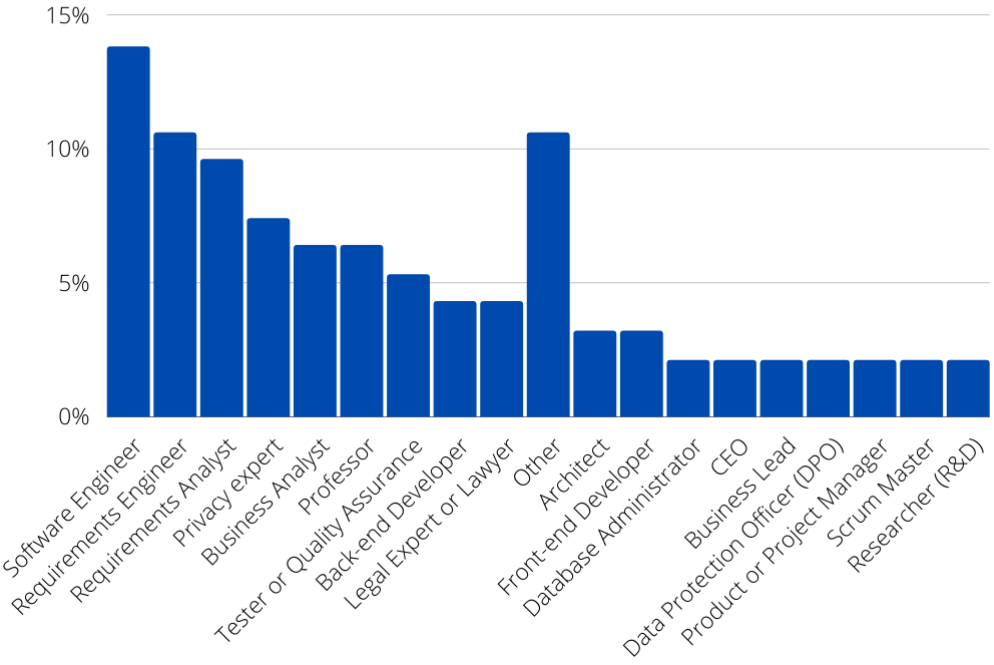
When asked about their roles, the participants can choose all that apply (i.e., multiple answer questions), a total of 94 answers (see Figure 37). The most cited roles were Software Engineer (13; 13.8%), Requirements Engineer (10; 10.6%), Requirements Analyst (10; 9.6%), and Privacy expert (7; 7.4%). Besides these, single responses were recorded for the "Others" (10; 10.6%) roles of Chief Data Officer (CDO), Designer (Interaction Designer, UX designer) or Specialist in Human-Computer Interaction, IT Consultant, Security Analyst, Security Engineer, Team Lead, Network Admin, Datacenter analyst, Cyber Security Manager, IT support analyst.

Among the 50 respondents, we identify 26 associated with multiple roles, as in Franch et al. (FRANCH et al., 2021) We observed 14 subjects associated with more than two roles; nine subjects were associated with three roles; one participant with four roles; one with five roles. And one, with eight roles (Database Administrator, Business Analyst, Architect, Privacy expert, Product or Project Manager, Requirements Analyst, Requirements Engineer, and Software Engineer). This participant operates in the Banking/Finance sector. They were assuring a good coverage of all the possible roles.

Looking at the distribution of participants' professionals years of experience, as shown in Figure 38, we noticed that our sample consisted of a mix of professionals with different levels of experience, assuring a good coverage of experienced participants. Practitioners had, on average, 9.36 years of experience in the software industry (standard deviation 6.30). Figure 38 shows that the majority of participants (22; 44%) have more than 11 years, 12 participants have between six and ten years (24%). Thus (34; 66%) have more than six years of overall industry experience. It was an experienced sample consisting of an analyst to senior-level professionals, which can be considered a good indicator of the information gathered in the questionnaire.

Figure 39 shows the participants' professionals years of experience in data protection (data

Figure 37 – Demographics: participants roles



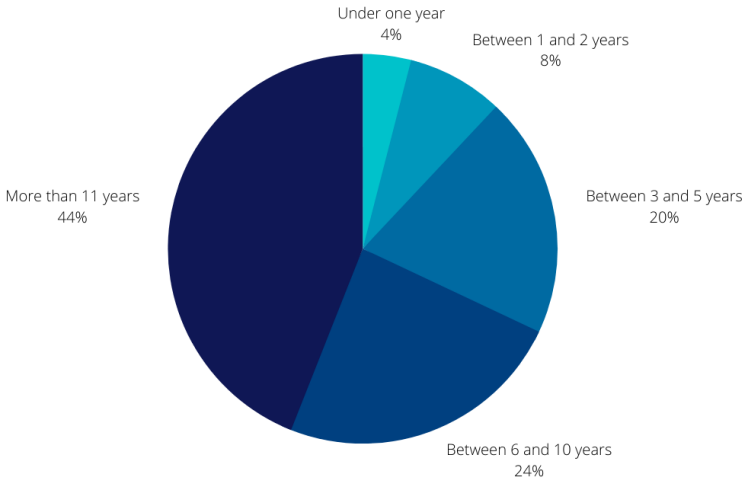
Source: The author (2021).

privacy) in software projects. Practitioners' experience in data protection projects had, on average, 2.82 years of practical experience (standard deviation 3.67).

The participants have been working on data protection for between one and two years (14; 28%), only three participants (6% have been working for more than 11 years). The participants that are not working with data protection are (15;30%). We believe that the high rate of respondents reporting that they do not work with data protection is due to the recent entry into the LGPD in Brazil in August 2020. Of the 15 respondents who do not work with data protection, 12 are Brazilian. The other three are from Spain, Pakistan, and Canada and reported that they act as teachers.

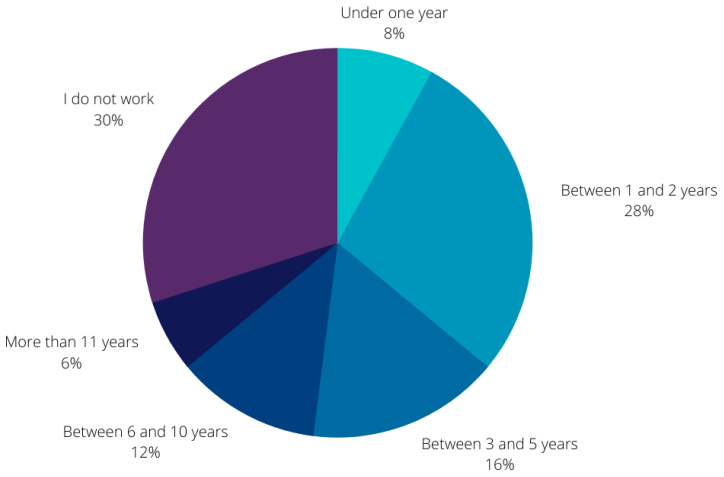
Figure 41 contains the distribution of the respondents according to the size of the organizations currently working, considering the number of employees and type of organization. In relation to type for the companies, Public (22; 44%), Private (28; 56%). We consider small companies up to 50 employees; medium up to 500 employees; large up to 2000 The participants; very large more than 2000 employees. Respondents that work in very large companies are 21 (42%). The percentages in all the others are few similar, assuring a good coverage of all the possible organization sizes.

Figure 38 – Demographics: experience in software projects



Source: The author (2021).

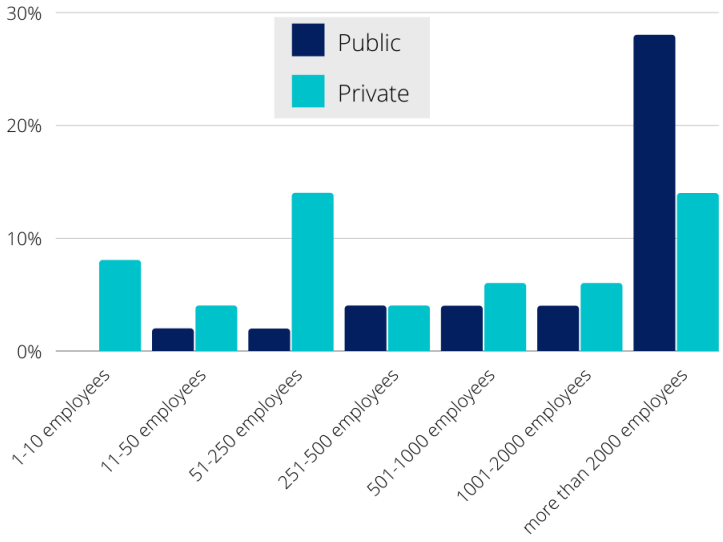
Figure 39 – Demographics: experience in privacy or data protection projects



Source: The author (2021).

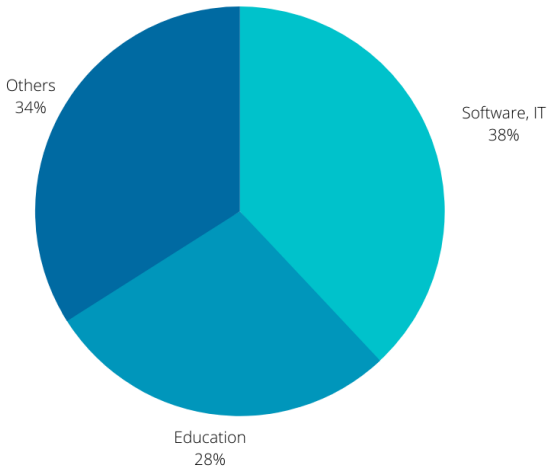
About the practice area, Software, IT (19; 38%), and Education (14; 28%) are the most cited. We include the other areas of activity (Automotive, Transport, Telecommunications, E-commerce, Legal, Games, Banking/Finance, Government, Military, Industry (application domain), Healthcare, Medical) in a category called Other (17; 34%).

Figure 40 – Demographics: distribution of responses for organization size *vs* type



Source: The author (2021).

Figure 41 – Demographics: practice area



Source: The author (2021).

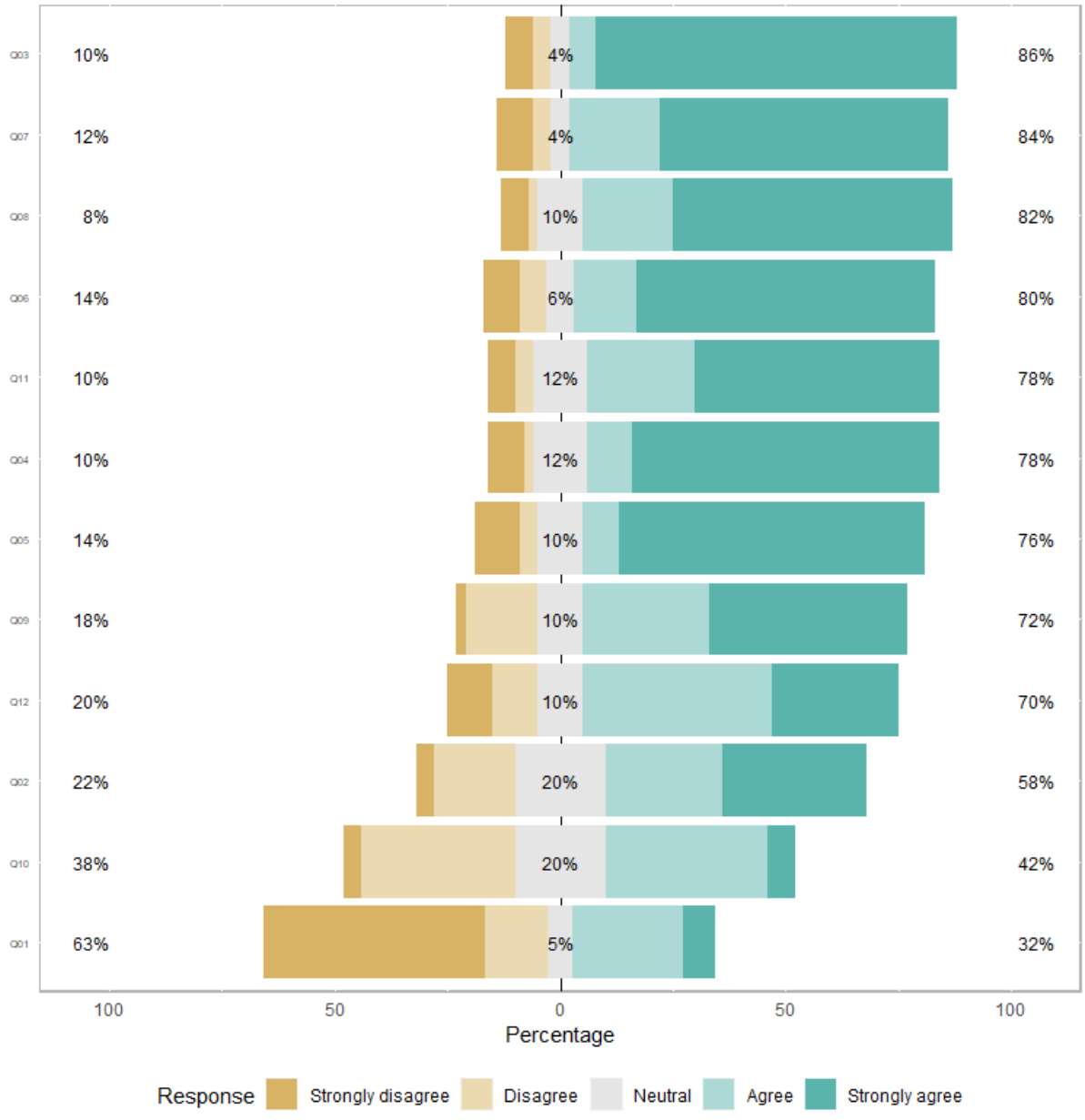
5.6.1 RQ4. What perceptions do IT professionals with industrial experience in RE have about ambiguity resolution in legal requirements specification and the compliance of such requirements with data protection law?

This section presents the results of our statistical analysis about personal perceptions about ambiguity in privacy requirements specification and legal compliance. The questions are divided

into a five-point Likert scale (1 - Strongly disagree (STD); 2 - Somewhat (SWD); 3 - Neither agree nor disagree (NAD); 4 - Somewhat agree (SWA); 5 - Strongly agree (STA)), and a choice based on agreement and participant's believes.

One of the objectives of the Likert scale questions is to validate the propositions that emerged from the interviews presented in (NETTO; SILVA; ARAÚJO, 2021) and summarized in Table 27. Figure 42 shows the percentage of participants' responses to each question about the point on the Likert scale.

Figure 42 – Professionals' perceptions regarding privacy culture for ambiguity resolution and legal compliance



Source: The author (2021).

Table 26 presents the frequency, percentage of responses for each option of the Likert scale, central tendency of responses through the median and mode.

Table 25 – IT professional's perceptions regarding legal requirements

	STD	SWD	NAD	SWA	STA	Mode
I believe that privacy and data protection is the responsibility only of the Information Technology department.	28 (56%)	5 (10%)	3 (6%)	10 (20%)	4 (8%)	1
I believe that the content of my company's privacy policy represents the practices we used.	2 (4%)	9 (18%)	10 (0,2%)	13 (26%)	16 (32%)	5
I believe that awareness-raising initiatives (for example, training, campaigns, training courses) for privacy are essential.	3 (6%)	2 (4%)	2 (4%)	3 (6%)	40 (80%)	5
I believe that privacy should be addressed from conception (Privacy by Design) when planning its activities	4 (8%)	1 (2%)	6 (12%)	5 (10%)	34 (68%)	5
I believe that Specialized Support Areas (Ambiguity Analysis, Anonymization, Legal, among others) are critical for reducing ambiguity in legal requirements specification and compliance with the law.	5 (10%)	2 (4%)	5 (10%)	4 (8%)	34 (68%)	5
I believe that promoting the company's privacy culture improves employees' awareness of privacy.	4 (8%)	3 (6%)	3 (6%)	7 (14%)	33 (66%)	5
I believe that reducing ambiguity in the privacy requirements specification requires cross-functional teams	4 (8%)	2 (4%)	2 (4%)	10 (20%)	32 (64%)	5
I believe that achieving legal compliance requires cross-functional teams.	3 (6%)	1 (2%)	5 (10%)	10 (20%)	31 (62%)	5
I believe that training that instructs all employees in ambiguity analysis in privacy requirements and legal compliance for software systems improves Legal requirements specification.	1 (2%)	8 (16%)	5 (10%)	14 (28%)	22 (44%)	5

Source: The author (2021).

We consider the answer to be predominantly "agree" if SWA plus STA is greater than the sum of the percentage of SWD and STD. Similarly, "disagree" is considered predominant if

Table 26 – IT professional's perceptions regarding legal requirements (continued)

	STD	SWD	NAD	SWA	STA	Mode
I believe that the tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification.	2 (4%)	17 (34%)	10 (20%)	18 (36%)	3 (6%)	4
I believe that a systemic view of the company concerning privacy and protection of personal data positively influences compliance with the specification of the software requirements to the law.	3 (6%)	2 (4%)	6 (12%)	12 (24%)	27 (54%)	5
I believe that Verification & Validation activities only do not guarantee software privacy requirements specification and legal compliance.	5 (10%)	5 (10%)	5 (10%)	21 (42%)	14 (28%)	4

Source: The author (2021).

the sum of the percentage of SWD and STD is greater than the sum of SWA and STA.

The top three measures which the respondents indicated agreement: **P3** - *I believe that awareness-raising initiatives (for example, training,campaigns, training courses) for privacy are essential* (43; 86%), **P7** - *I believe that reducing ambiguity in the privacy requirements specification requires cross-functional teams* (42; 84%), and **P8** - *I believe that achieving legal compliance requires cross-functional teams* (41; 82%).

The three measures that most respondents perceived as disagree are **P1** - (*I believe that privacy and data protection is the responsibility only of the IT (Information Technology) department*) is the only one that the disagree responses (STD and SWD) are higher (33;63%) than the other answers. The other alternatives that had relevant disagreement were: **P10** - *I believe that the tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification* (19; 38%), *I believe that the content of my company's privacy policy represents the practices we used* (11, 22%).

We cross-tabulated between the subgroups (*Maximum degree*, *Experience in the software industry*, *Experience in Data protection*, *Size organization*, *Type organization*, and *Practice area*) and each personal perception of ambiguity in the specification of the privacy requirements and legal compliance. Affirmed by the participants in the ninth question.

Question 10 (I believe that the tacit knowledge compensates for the lack of guidelines for the reduced ambiguity in the privacy requirements specification) has a minor difference

of answer between Somewhat disagree (SWD) (17;34%), neither agree nor disagree (NAD) (10;20%), and somewhat agree (SWA) (18;36%). We can consider tacit knowledge (personal wisdom, experience, insight, and intuition) as a factor that helps to reduce ambiguity in the specification of legal requirements, as they can identify risk, know the customer domain, and the governing legislation that domain.

To check if there is a difference in the distribution of the dependent variable between the levels of the independent variable and if this observed distribution is significantly different from the expected one, we performed the Fisher's Exact Test. Fisher's Exact Test is indicated when fewer than 20 observations are in the sample or fewer than five counts in a cell. Most of Fisher's tests performed led to *p*-values that were larger than 0.05, which means that there is not a statistically significant relationship between these two variables.

The only case that the *p*-value is less than 0.05 occurs between *I believe that the tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification* (Fisher's Exact Test: *p*-value = .03818; Pearson's Chi-squared: *p*-value = .001252) and practice area from the company, indicating that there is a relationship statistically significant.

Then, the intensity of this relationship was verified using Cramer's V and Contingency coefficient. The results range from 0 (no association) to 1 (very strong association). We obtained Cramer's V = 0.602 and Contingency coefficient = 0.769, which indicates an association, but it is not very strong.

Then, we performed the Kruskal-Wallis test (*p*-value = .2209) and the Dunn post-hoc test with Bonferroni adjust, which did not indicate significance between the relationship. Therefore, there is no relationship between the company's area of operation and the response to the *I believe that the tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification*.

Thus, analyzing the Figure 42 and Table 26 we identified that the participants agree agree with ten of the twelve propositions presented in Table 27. The propositions **I. P9** to **I. P12** emerged from the data when we were performing the constant comparison method to prepare the survey, so they did not appear in the study of the interviews.

Proposition **I. P6** - *Tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification* had a low agreement rate, lower than 50%. We believe that this happened, as the participants may not have understood the term tacit knowledge. The proposal aimed to verify whether the professional's experience (personal

Figure 43 – Frequency and inference test results to question 10

Practice_area	*Q10* 1	2	3	4	5	***Total***
Automotive, Transport	0 0 %	0 0 %	0 0 %	0 0 %	1 100 %	1 100 %
Banking/Finance	0 0 %	0 0 %	0 0 %	2 100 %	0 0 %	2 100 %
E-commerce	1 33.3 %	0 0 %	1 33.3 %	1 33.3 %	0 0 %	3 100 %
Education	0 0 %	6 42.9 %	3 21.4 %	5 35.7 %	0 0 %	14 100 %
Games	0 0 %	0 0 %	0 0 %	1 100 %	0 0 %	1 100 %
Government, Military	1 50 %	0 0 %	0 0 %	1 50 %	0 0 %	2 100 %
Healthcare, Medical	1 100 %	0 0 %	0 0 %	0 0 %	0 0 %	1 100 %
Industry (application domain)	0 0 %	1 100 %	0 0 %	0 0 %	0 0 %	1 100 %
Legal	0 0 %	2 50 %	0 0 %	2 50 %	0 0 %	4 100 %
Software, IT	0 0 %	9 47.4 %	6 31.6 %	3 15.8 %	1 5.3 %	19 100 %
Telecommunications	0 0 %	0 0 %	0 0 %	2 100 %	0 0 %	2 100 %
Total	3 6 %	18 36 %	10 20 %	17 34 %	2 4 %	50 100 %

χ^2=72.510 · df=40 · Cramer's V=0.602 · Fisher's p=0.039

Source: The author (2021).

wisdom, experience, insight, and intuition) can assist in identifying risk, know about the customer domain and the legislation governing that domain. Similar to the work by Peixoto et al. (PEIXOTO et al., 2020), which identified that the *empirical knowledge about informational privacy* is a positive personal factor indicated which respondents had practical knowledge about personal data.

The proposition **I. P9** - *I believe that privacy and data protection is the sole responsibility of the Information Technology department* has a high disagreement rate (66%). Nevertheless, such a response was expected. Nevertheless, selecting disagree demonstrates that participants perceive the need and importance of privacy and data protection must be addressed across all company sectors and not just the responsibility of the Legal and IT area. It is being considered a positive factor.

Table 27 – Propositions and agreement percentage

Proposition	Agreement (%)
I. P1 - Specialized Support Areas (Ambiguity Analysis, Anonymization, Legal, among others) are critical for reducing ambiguity in legal requirements specification and compliance with the law	76
I. P2 - Promoting the company's privacy culture improves employees' awareness of privacy	80
I. P3 - Reducing ambiguity in the legal requirements specification and achieving legal compliance requires cross-functional teams	84
I. P4 - Achieving legal compliance requires cross-functional teams	82
I. P5 - Training to instruct employees regarding ambiguity analysis in privacy requirements and legal compliance for software systems improves Legal requirements specification	72
I. P6 - Tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification	42
I. P7 - A systemic view of the company concerning privacy and protection of personal data positively influences compliance with the specification of the software requirements to the law	78
I. P8 - Verification & Validation activities only do not guarantee software privacy requirements specification and legal compliance	70
I. P9 - Privacy and data protection are the responsibility only of the IT (Information Technology) department.	32
I. P10 - The content of my company's privacy policy represents the practices we used.	58
I. P11 - Awareness-raising initiatives (for example, training,campaigns, training courses) for privacy are essential.	86
I. P12 - Privacy should be addressed from conception (Privacy by Design) when planning its activities.	78

Source: The author (2021).

5.7 THREATS TO VALIDITY

In this section, we discusses the main types of validity namely: face validity, content validity, construct validity, criterion validity, as defined by Kitchenham and Pfleeger (KITCHENHAM; PFLEEGER, 2008).

Face Validity: evaluates the appearance of the questionnaire in terms of feasibility, readability, consistency of style and formatting, and the clarity of the language used (TAHERDOOST, 2016).

Content Validity: involves evaluation of a new survey instrument in order to ensure that it includes all the items that are essential and eliminates undesirable items to a particular construct domain (BOUDREAU; GEFEN; STRAUB, 2001).

We mitigate the face and content validity asking two researchers with more than 15 years of experience in the RE field, analyzed the questionnaire. Moreover, we used an empirically evaluated checklist (MOLLÉRI; PETERSEN; MENDES, 2020) to guide our survey design. The details of our checklist evaluation can be found in (NETTO; SILVA, 2021). To avoid instrumentation threats, we also conducted pilots of the questionnaire with one Ph.D. candidate, one Ph.D. researcher with experience in RE, and one Ph.D. professor to ensure its correct understanding and find possible defects.

Construct Validity: concerns how well an instrument measures the construct it is designed to measure (KITCHENHAM; PFLEEGER, 2008). In order to ensure that the instrument properly measures what it is supposed to measure, we based the constructs, propositions, and questions of the survey were elaborated from analyzing a set of 22 interviews ((NETTO; SILVA; ARAÚJO, 2019) (NETTO; SILVA; ARAÚJO, 2021)), and in the Systematic Literature Mapping (SLM) (NETTO; PEIXOTO; SILVA, 2019).

Criterion validity: measure how well one instrument compares with another instrument or predictor (KITCHENHAM; PFLEEGER, 2008). The questionnaire was based on the factors that emerged from the interviews (NETTO; SILVA; ARAÚJO, 2021). Nevertheless, there was no other instrument to make comparisons that deal with ambiguity in legal requirements specification and compliant with data protection laws.

5.8 LIMITATIONS

The survey homepage as well as cover letter explicitly mentioned that the target audience of this study is IT professionals with industrial experience in RE or system analysis. Participants were also informed of the overall objective and importance of the research, guaranteed data confidentiality, participation anonymity, and the right to withdraw from the research at any time, through the Informed Consent (IC) (available in (NETTO; SILVA, 2021)).

The constructs, propositions, and questions of the survey were elaborated from analyzing a set of 22 interviews. Therefore, we cannot guarantee that those study participants from the interviews did not respond to the questionnaire because the survey was carried out entirely anonymously by the participants. However, as we had 39 responses in the first part of the

survey and 50 responses in the second part, if any respondent has answered the survey, it would not influence the final result.

Looking at the participants' work experience distribution, we noticed that our sample consisted of professionals with different experience levels, assuring a good coverage of experienced participants. According to their experience in the software industry, 44% have more than 11 years, and 24% have between six and ten years. Thus (34; 66%) have more than six years of overall industry experience, it was an experienced sample consisting of the analyst to senior-level professionals.

Due to the fact of having only 50 responses at Survey 5.6 and 39 responses at Survey 5.5, does not allow for a generalization of the results, only observations about the current state of the practice. However, we believe that the results already provide significant insights into the current state of practice.

We took several measures to avoid fake answers; we activated detection of responses coming from the same IP address, making it impossible for the same person to submit the survey twice without changing the IP of their device. Second, we organized the online questionnaire into sections presented on different web pages, impossible to skip sections. The work involved in answering the survey would discourage people who were not interested in the survey subject.

To prevent participants from considering the invitation email unnecessary, we sent the message through the institutional University email. We included a Cover Letter explaining the academic purpose of the survey and provided a link to the DARE research group page[6], link to survey page, link to supplementary material (NETTO; SILVA, 2021), and signed the emails with the authors' names.

As Dillman et al. (DILLMAN; SMYTH; CHRISTIAN, 2014) stated, one of the main problems of online surveys is having meager participation rates. In our case, we sent out 450 formal invitations. In the case of individuals identified over LinkedIn, the email contents were sent as a direct message and made posts on LinkedIn, Twitter, and ResearchGate, and received 39 responses at Survey 5.5, response rate 8.7%. On Twitter, we were able to see that there were 691 impressions (times people saw this tweet on Twitter), but we cannot say that these users viewed the survey. Nevertheless, in other social networks (LinkedIn and ResearchGate, we could not accurately verify the number of interactions with the publication. Finding a suitable sampling frame (i.e., the current population) is challenging in surveys for which no exhaustive register of the target population exists (DILLMAN; SMYTH; CHRISTIAN, 2014). This problem is

[6] http://www.cin.ufpe.br/dare/

also reported by (PALOMARES; QUER; FRANCH, 2017). Therefore, it is impossible to know the percentage of responses to which the survey announcement arrived, but just the percentage of people who opened the survey.

At Survey 5.5, from the 95 respondents that started to answer, only 39 completed it, which represents a low completion rate (42.7%). We realized a high percentage of the non-completed attempts in a specific section of the survey that dealt with legal compliance, privacy, and data protection law (GDPR and LGPD). We believe that the participants were unaware of the topic and, as they were unable to advance the questionnaire without answering the questions, they ended up dropping out. In addition, the length and complexity of the survey might have influenced the completion rate.

5.9 CHAPTER SUMMARY

We conduct a cross-sectional survey through a self-administered questionnaire online since the participants belong to different countries and the Brazilian States. The survey aims to collect the software practitioners' perceptions regarding the factors and actions to achieve ambiguity resolution and legal compliance in a software requirements specification. Additionally, it intends to identify new factors and actions that can promote or mitigate the factors.

This Chapter, presented thirty-two propositions that can be recommendations and define an initial theory explaining how the privacy requirement specification with reduced ambiguity and verifying software systems' legal compliance has been addressed in practice. As benefits for the academy, we can mention a detailed research method that can guide other researchers, and this study can be replicated. Benefits this study to practitioners, presents an overview of how privacy is taken into account in companies, encourages positive factors, and mitigates the negative ones.

6 DEFINITION OF THEORY

A theory provides explanations and understanding in terms of basic concepts and underlying mechanisms, which constitute an essential counterpart to the knowledge of passing trends and their manifestation (HANNAY; SJOBERG; DYBA, 2007), (WOHLIN et al., 2012). From the practical perspective, theories should be helpful and explain or predict phenomena that occur in software engineering and gives us input to decision-making regarding choice of technology and resource management (SJØBERG et al., 2008). From a scientific perspective, theories should guide and support further research in software engineering to facilitate the communication of ideas and knowledge. Helps develop and consolidate common research agendas (SJØBERG et al., 2008).

Based on our results, we developed an theory of how IT Companies address ambiguity resolution and compliance with Data Protection Laws in the requirements specification. Consulting the literature for theories that deal with ambiguity in the specification of legal requirements or compliance, we have not found any theory.

Sjøberg (SJØBERG et al., 2008) mention that since Software Engineering (SE) is an applied discipline, SE theories should, at least ultimately, be helpful to the software industry. Our research establishes a theory that can be used as starting point for further studies and more detailed investigations. Practitioners can use the results as a guide for selecting methods and techniques suitable to specify legal requirements with reduced ambiguity compliant with data protection laws.

We used the structure outlined in the theory of Sjøberg (SJØBERG et al., 2008): (1) defining the constructs, what are the essential elements; (2) defining the propositions, how do the constructs interact; (3) providing explanations to justify the theory, why are the propositions as specified; (4) determining the scope of the theory, what is the universe of discourse in which the theory is applicable; as well as (5) testing the theory through empirical research.

The *constructs* are the basic elements of the theory. Concepts from grounded theory analysis are candidate constructs for a theory, and the central categories are candidates for constructs are identified (SJØBERG et al., 2008). In this study, the constructs emerged from the interview data in (NETTO; SILVA; ARAÚJO, 2019), and (NETTO; SILVA; ARAÚJO, 2021), and were validated in this survey. The *propositions* are the interactions among the constructs. Relationships that had clear support from the data were candidates for being included in the theory's propositions.

Furthermore, the relationships were validated using questionnaires and compared with the literature. Finally, the relationships that were supported by all the data and that included the candidate constructs were aggregated into propositions (SJØBERG et al., 2008). The *explanations* describe why the propositions are as specified. An explanation is a relationship among constructs, and other categories, which are not central enough to become constructs (SJØBERG et al., 2008). The scope of the theory describes the universe in which the theory is applicable (SJØBERG et al., 2008). The scope of our theory is IT companies with legal requirements projects.

Sjøberg et al. (SJØBERG et al., 2008) also propose a graphical representation inspired by UML class diagrams. Our theory is presented in Figure 44 which shows the constructs of theory for reducing ambiguity in legal requirements specification and achieving specifications compliant with data protection laws. A construct is represented as a class or an attribute of a class. The relationships are specified further into propositions of the theory. As suggested by Sjøberg et al. (SJØBERG et al., 2008), we structure the constructs into technology, activity, and actors. The main actors we describe in our theory, presented in Table 28 are Requirements Engineers (**C8**), Specialized Support (**C9**) (lawyers, DPO, or another specialist in software domain), and the Customer (**C10**).

Based on our analysis of twenty-two interviews (NETTO; SILVA; ARAÚJO, 2021), we identified codes, concepts, and categories. The six categories (*Requirements specification*, *Reducing ambiguity*, *Communication between development team members*, *Achieving legal compliance*, *Working with data protection regulation*, *Specialized support area*) we arrived at became constructs in our proposed initial theory. The category Requirements Elicitation is not part of the theory, because the survey did not analyze issues related to the requirements elicitation activity, as the focus of the work is the specification of legal requirements.

The category *Working with data protection regulation* is represented in the construct *Legal Compliance Techniques* (**C3**), and *Specialized Support Area* is represented through the actor (**C9**). All main constructs of our theory are summarised in Table 28 which also makes the scope of our theory explicit.

For each category, we includes propositions that established essential interactions between the constructs. In the questionnaire, the answer possibilities corresponded directly to the propositions and resulted in the closed multiple-choice survey question. For example, " **Question 12** *Considering your personal experiences, which of the following problems in legal requirements engineering affects your projects?*"" with the choices "*A weak relationship between Customer*

and project lead," "*Changing goals, business processes, and/or requirements*," "*Lack of collaboration between system engineers and lawyers*," and "Other." The participants evaluated these propositions and, after analyzing the data, we were able to validate the propositions with industry professionals. In this way, we carried out the theory design, presented in Figure 44.

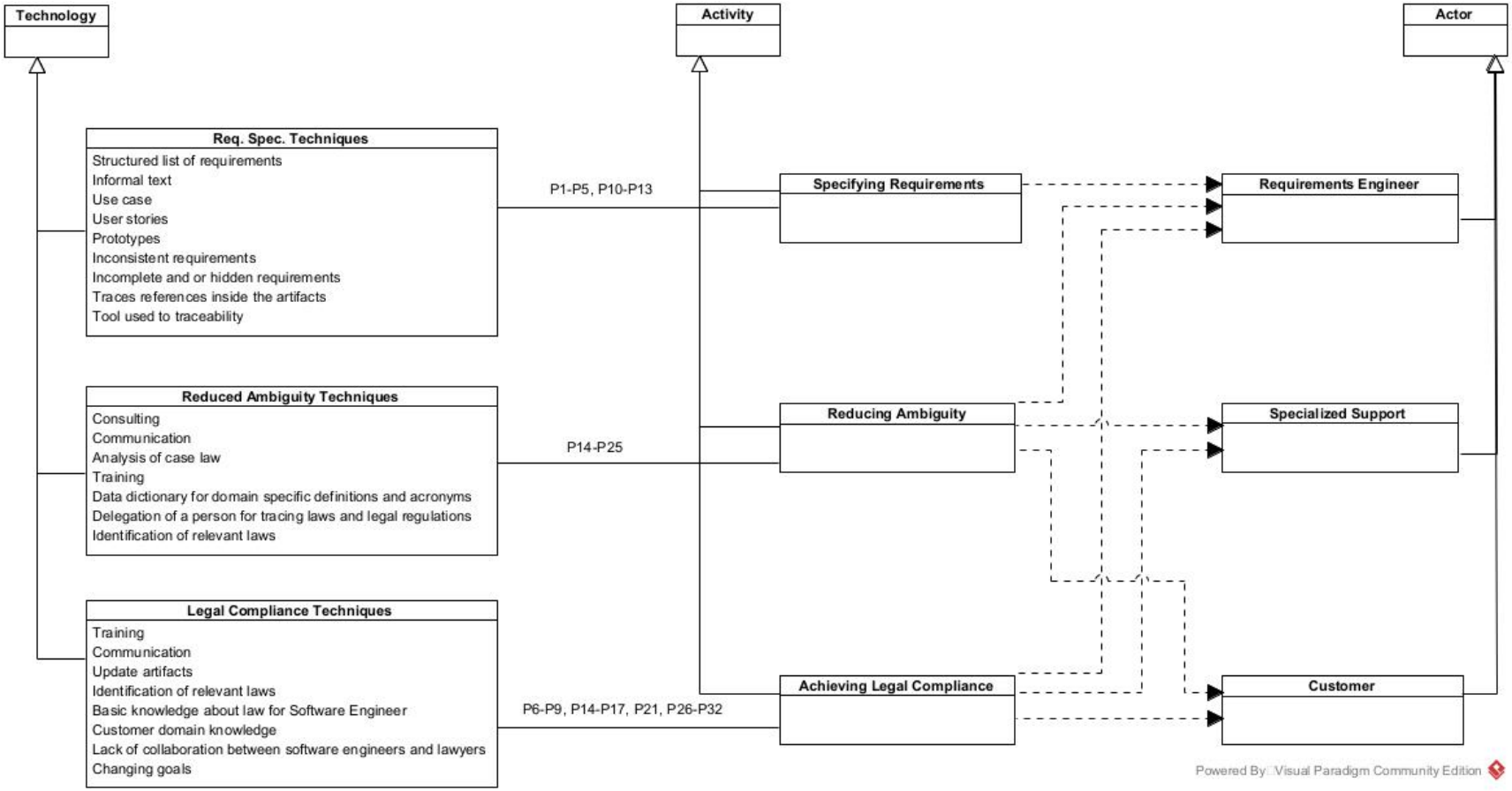
Table 28 – Constructs and scope of the theory

nº	**Constructs**	**Type**
C1	Requirements specifications techniques	Technology
C2	Reduced ambiguity techniques	Technology
C3	Legal compliance techniques (achieving legal compliance, and working with Data Protection Regulation)	Technology
C4	Communication	Technology
C5	Specifying Legal Requirements	Activity
C6	Reducing Ambiguity	Activity
C7	Achieving Legal Compliance	Activity
C8	Requirements Engineer	Actor
C9	Specialized support	Actor
C10	Customer	Actor

Scope
The scope of our theory is IT companies with legal requirements projects.

Source: The author (2021).

Figure 44 – Theory representation of how IT Companies address ambiguity resolution and compliance with Data Protection Laws in requirements specification



Source: The author (2021).

When describing the theory, we introduce the constructs and propositions, identifying them in the text with **C** and **P**, respectively. This section describes the constructs and the explanations for the propositions (presented in Table 30) that relate to that construct.

Table 29 – Propositions about legal requirements specification with reduced ambiguity and compliant with data protection laws

nº	Propositions
S. P1	Legal requirement are specified via structured list of requirements
S. P2	Legal requirement are specified via informal text or plain text
S. P3	Legal requirement are specified via use case
S. P4	Legal requirement are specified via user stories
S. P5	Legal requirement are specified via prototypes
S. P6	Lack of training on data protection regulation affects projects with data protection requirements
S. P7	Changing goals, business process, and/or requirements
S. P8	Lack of collaboration between software engineers and lawyers affects projects with data protection requirements
S. P9	Communication flaws between developers and the customers affects projects with data protection requirements
S. P10	Incomplete and/or hidden requirements
S. P11	Inconsistent requirements
S. P12	Trace references are stated inside the requirements specification and legal regulations artifacts
S. P13	A tool is used to record artifacts and trace links
S. P14	Discussion between team members helps to solve and/or reduce ambiguity in legal requirements
S. P15	Consulting/reading Laws regulatory sources helps to solve and/or reduce ambiguity in legal requirements
S. P16	Consulting internal Legal sector helps to solve and/or reduce ambiguity in legal requirements
S. P17	knowledge of the customer's domain helps to solve and/or reduce ambiguity in legal requirements
S. P18	Direct contact with the customer involved with the project helps to solve and/or reduce ambiguity in legal requirements
S. P19	Analysis of case law helps to solve and/or reduce ambiguity in legal requirements
S. P20	Ask for clarification form Government Authority helps to solve and/or reduce ambiguity in legal requirements

Source: The author (2021).

Table 30 – Propositions about legal requirements specification with reduced ambiguity and compliant with data protection laws (continued)

nº	Propositions
S. P21	Basic knowledge about law for software engineers helps to reduce ambiguity in legal requirements specification
S. P22	Training in ambiguity identification techniques helps to reduce ambiguity in legal requirements specification
S. P23	Data dictionary for all domain specific definitions and acronyms/Glossary helps to reduce ambiguity in legal requirements specification
S. P24	Delegation of a person for tracing laws and legal regulations helps to reduce ambiguity in legal requirements specification
S. P25	Identification of relevant laws and legal regulations and their analysis performed by lawyers helps to reduce ambiguity in legal requirements specification
S. P26	Updating the product backlog helps to deal with changes in legal requirements
S. P27	Regularly change the legal requirements specification helps to deal with changes in legal requirements
S. P28	Discussion with customers helps to deal with changes in legal requirements and to decide the best approach
S. P29	Training and practice about Privacy Principles must be provided to the stakeholders in legal requirements and personal data protection laws
S. P30	Training and practice about Privacy by Design and/or Privacy by Default must be provided to the stakeholders in legal requirements and personal data protection laws
S. P31	Training and practice about Laws and regulations related to the software's subject area to be developed (e.g.,patient record law, Privacy and Electronic Communications Regulation (e-Privacy), ICT regulation) must be provided to the stakeholders
S. P32	Training and practice about Concerning data subjects' rights (presented in GDPR (Articles 12-23), and LGPD (Articles 17-22) must be provided to the stakeholders

Source: The author (2021).

6.1 EXPLANATIONS ABOUT CONSTRUCT "REQUIREMENTS SPECIFICATION TECHNIQUES"

We analyzed the practitioners' practices and **Requirements Specification Techniques** (**C1**) that the companies use in their daily work. The **Specifying Legal Requirements** (**C5**) activities are performed by the **Requirements Engineer** (**C8**) actor.

Legal requirements are specified textually by most participants. Either through a structured list of requirements (**S. P1**), informal or plain text (**S. P2**), use cases (**P3**) or user story (**S. P4**). Thus, the concern with ambiguity in the specification of legal requirements becomes even

more present.

According to Oran et al. (ORAN et al., 2021), three of the most used techniques in the industry to specify requirements are use cases, user stories (due to the growth of agile development), and prototypes, as presented in the result of our survey. In the work by Wagner et al. (WAGNER et al., 2019), structured requirements list (**S. P1**) and use case (**S. P3**) are the main techniques used to document functional requirements. According to Wagner et al. (WAGNER et al., 2019), the documentation of non-functional requirements (as in the case of legal and privacy requirements) happens in a textual form, as was corroborated in the data in this survey.

The representation of the expanded use cases (**S. P3**) proposed by Phalp et al. (PHALP; VINCENT; COX, 2007), (with main, alternative, and exceptions flow) uses natural language to specify legal requirements. However, some common problems, such as ambiguities, incompleteness, and inconsistencies, can arise when describing the requirements through use cases. These problems can cause difficulties in understanding the requirements and, consequently, defects in the software system under development (REGGIO et al., 2018).

User story (**P4**), when improperly defined, can trigger several challenges in agile software development due to incomplete or incorrect documentation (INAYAT et al., 2015). The main difficulties in using user stories to specify requirements are related to sparse detailing of requirements information, difficulty in identifying non-functional requirements, communication and collaboration with users, lack of information for validation of requirements (SOARES et al., 2015).

The use of prototypes (SNYDER, 2003) (**S. P5**) is quite widespread to represent the requirements in the requirements elicitation sessions, through a visual representation, it facilitates the understanding of the system's functionalities for both the requirements engineer and for the customer or other stakeholders. In the requirements specification, prototypes are used as specification techniques for activities related to User-Centered Design (REGGIO et al., 2018). De Lucia et al. (LUCIA; QUSEF, 2010) recommend the use of prototypes to document the requirements for communication and knowledge sharing between stakeholders and agile teams.

Thus, all propositions **S. P1** to **S. P5** use any natural language technique to specify legal requirements. It is also possible to relate the propositions to factor F23 (Always in Natural Language) from the study of interviews, which negatively influences the specification of legal requirements.

Two of the main issues affecting the specification of legal requirements are incomplete and/or hidden requirements (**S. P10**) and Inconsistent requirements (**S. P11**). The data support these propositions. *Inconsistent requirements* (17; 5.37%), *Incomplete and/or hidden requirements* (17; 5.37%) are the fifth and sixth most-cited problems, respectively. Similar to the responses of our survey, the work by Kalinowski et al. (KALINOWSKI et al., 2016) identified that the most critical reported RE problems are related to incomplete/hidden requirements, underspecified requirements, communications flaws between the project team and the customer, and communication moving targets and time boxing problems. In addition to these, they cited insufficient support by the customer and inconsistent requirements.

The management of traceability between legal requirements, the legal text, and the requirements document is crucial. *Lack of traceability between requirements and legal text* is a problem cited by 12 (3.79%) participants, and happens through propositions (**S. P12**) *trace references are stated inside the requirements specification and legal regulations artifacts*, and *a tool is used to record artifacts and trace links* (**S. P13**).

Explanations for propositions **S. P1** - **S. P5**:

E1 - *The legal requirements are specified using techniques* ((**S. P1**) to (**S. P4**)) *that allow all involved stakeholders to establish a common understanding of the system's functionalities. Moreover, they present different informational needs for each team role to perform their specific tasks properly.*

E2 - *Prototypes* (**S. P5**) *allow that all involved stakeholders have a common understanding of the system's functionalities, meet the customer's expectations, and comply with data protection regulations.*

Explanations to propositions **S. P10** - **S. P13**:

E3 - *Incomplete and/or hidden requirements* (**S. P10**) *and Inconsistent requirements* (**S. P11**) *are critical problems that affect the specification of legal requirements and legal compliance, as they lead to ambiguous requirements specifications.*

E4 - *Maintaining the traceability between the legal text and the requirements specification document* (**S. P12** - **S. P13**) *facilitates verification if the law undergoes any changes and assists the Requirements Engineer in verifying and legal compliance.*

6.2 EXPLANATIONS ABOUT CONSTRUCT “REDUCED AMBIGUITY TECHNIQUES"

Specifying legal requirements (**C5**) with **Reduced Ambiguity Techniques** (**C2**) is a challenge, and **Communication** (**C4**) between the project's stakeholders and **Specialized Support** (**C9**) is essential.

Techniques that can be used to reduce ambiguity are *consulting internal legal sector* (**S. P16**), and *ask for clarification from the Government Authority* (**S. P20**) about the interpretation of the legal text that will be operationalized in a software requirement.

The most cited sources of knowledge used to help resolve or reduce ambiguity in the specification of legal requirements are the *discussion among team members* (**S. P14**) and the *direct contact with the customer involved with the project* (**S. P18**). Both propositions highlight the importance of **communication** to reduce ambiguity in the specification of legal requirements. We can relate the proposition **S. P14** with the factors **F09** (*Team members have the same understanding*) and **F22** (*Skill of each multidisciplinary team member*), from the study of the interviews. In the same way, we can report the proposition **S. P18** to the factor **F04** (*Customer involvement in the project*).

According to Fricker et al. (FRICKER et al., 2016), requirements communication is the process of transmitting a customer's needs to a development team to implement a solution. Requirements communication problems can cause productivity losses or even design failures. Méndez et al. (FERNÁNDEZ et al., 2016) highlight that the critical engineering requirements (RE) problems are related to communication problems and incomplete/hidden or unspecified requirements.

Communication flaws between developers and the customer (**S. P9**) is the third most cited problem (18; 5.69%) that affects projects with legal requirements. Participants also selected intra-team communication flaws (*Communication flaws within the project development team* (15; 4.74%)) or have *Difficulty understanding domain-specific terms* (13; 4.11%), that could lead to *The developer may make their wrong interpretation* (11; 3.48 %), *knowledge of the customer's domain* (**P17**), and *direct contact with the customer involved with the project* (**P18**) resolves these issues quickly. Communication is essential to solving the problems of interpreting legal requirements and therefore specifying the correct and unambiguous specification and the compliance of legal requirements with legislation.

Another contribution from the legal sector supporting the requirements engineer is in the *analysis of case law* (**S. P19**), as there are older data protection laws in other countries whose

processes were performed in companies or software that failed to comply with the law. Thus, the analysis of jurisprudence can facilitate legal compliance and the understanding of legislation, as it is possible to verify the court's understanding of those legal requirements. Kosenkov et al. (KOSENKOV et al., 2021) claim that the interpretation of regulatory requirements typically needs the legal knowledge of legal experts. The interdisciplinary nature of regulatory RE makes it essential to establish a dialog between legal experts and requirements engineers for an effective regulatory RE process (BOELLA et al., 2014).

Training in ambiguity identification techniques (**S. P22**) for the Requirements Engineer and Basic knowledge about law for the software engineer (**S. P21**) can help you identify regulatory law sources (**S. P15**) and specification of legal requirements with reduced ambiguity.

Other techniques that can be used to reduce ambiguity are (**S. P24**) *delegation of a person for tracing laws and legal regulations* supporting the requirements specification process. *Identification of relevant laws and legal regulations and their analysis performed by lawyers* (**P25**).

Furthermore, creating a (**P23**) Data Dictionary for all domain-specific definitions and acronyms/glossaries. As stated by Otto and Antòn (OTTO; ANTóN, 2007), it is needed to support requirements engineers, policymakers, and auditors in establishing a unified glossary for the system specification, design documents, and compliance audit artifacts.

Explanation to proposition **S. P9**:

> **E5** - *Communication flaws between developers and the customer* (**S. P9**) *is one of the main problems affecting projects with legal requirements. Therefore, the knowledge of the customer's domain can help solve and/or reduce ambiguity in legal requirements*.

Explanations to propositions **S. P16**, **S. P19**, and **S. P20**:

> **E6** - *The company's legal department* (**S. P16**) *can analyze case laws* (**S. P19**) *or consult external Government Authorities* (**S. P20**) *and propose a solution to the ambiguity that was presented*.

Explanations to propositions **S. P14**, **S. P17** and **S. P18**:

> **E7** - *The knowledge of the customer's domain* (**S. P17**) *helps in the discussion between team members* (**S. P14**) *in search of ambiguity resolution. However, should the ambiguity remain, the direct contact with the customer involved with the project* (**S. P18**) *resolves these issues quickly.*

Explanations to propositions **S. P15**, and **S. P21** - **S. P25**:

E8 - Training in ambiguity identification techniques (**P22**), basic knowledge of legislation (**S. P21**) and how to consult/reading Laws regulatory sources (**S. P15**) should be provided for the Requirements Engineer in order to help to solve and/or reduce ambiguity in legal requirements.

E9 - The specialized support area or the legal sector may (**S. P24**) delegate a person for identification and analysis of relevant laws and legal regulations (**S. P25**) for supporting the requirements specification process.

E10 - The specialized support area or the legal sector can create a (**P23**) data Dictionary for all domain-specific definitions and acronyms/glossary to help to reduce ambiguity in legal requirements specifications.

6.3 LEGAL COMPLIANCE TECHNIQUES

Achieving Legal Compliance (**C7**) is an activity that must be performed with the support of the **Specialized Support** (**C9**) (company's legal sector, which may involve lawyers, DPO, among other professionals specialized in the domain of the requirements that will be implemented).

In order to achieve legal compliance with the legal requirements specification, those involved must have *basic knowledge about law for Software Engineers* (**P21**), relating to Privacy and Personal Data Protection. Therefore, the *lack of training on data protection regulations* (P6) is the primary concern of twenty practitioners in legal requirements engineering projects. The legal requirements have peculiarities that need to be explored in training to become aware of the practices. Most of the laws are constantly updated (*Constant changes in the law make legal compliance difficult* (10;3.16%)). There will likely be a change in the processes requiring training with employees and customers. The lack of training on data protection regulations can lead to difficulties interpreting legislation and, therefore, an ambiguous specification of requirements and, consequently, non-compliance with data protection laws.

In this way, **training** or campaigns can be carried out by the company, or the professional can carry them out so that they have a (**P21**) *basic knowledge about law for Software Engineer* and improve their skills. The training and practices most mentioned by the participants are (**P29**) Training in privacy principles, (**P30**) Training in Privacy by Design (PbD), (**P31**) Laws

and regulations related to software subject area to be developed, (**P32**) Concerning data subject's right.

Consulting internal legal sector (**P16**) in order to clarify doubts and assist in the elaboration of solution proposals for specifying legal requirements and achieving legal compliance is trivial. Therefore, when there is (**P8**) *lack of collaboration between software engineers and lawyers*, it can hinder the legal compliance activity of the requirements specification. Therefore, *lack of collaboration between software engineers and lawyers* (**P8**) (19; 6.01%) is the second problem most identified by the participant. The way the legal text is written and structured (*Cross-reference among legal/regulatory documents* (9;2.84%) makes it difficult for users who do not have the legal knowledge to understand. support from the company's lawyers or legal department, even to *Identify the regulations relevant to its specific system* (**P25**) (11; 3.48%) and laws regulatory sources (**P15**) (24 ; 10.3%) in order to prevent *The developer may make their wrong interpretation* (11; 3.48%).

Thus, as in the activity related to the specification of requirements with reduced ambiguity, the (**P17**) *knowledge of the customer's domain* communication is essential for Legal Compliance. Everyone involved must have the same understanding of the legal requirements that will be implemented. Therefore, the (**P14**) *Discussion between team members* reduces non-compliance issues. *Communication flaws between developers and the customer* (**P9**) can lead to non-compliance, as the participation of the customer or the customer's specialized support team helps in the verification and validation of the requirements specification and, consequently, in the achievement of legal compliance.

As laws undergo frequent changes, the adequacy of artifacts related to requirements specification (**P26** *update from product backlog* and **P27** *regularly change the legal requirements specification*) in order to meet the new functionalities is decided with the participation of the customer (**P28**) *we discuss with customer and decide the best approach* . *Changing goals, business processes, and/or requirements* (**P7**) (17; 5.37%) is the fourth problem most selected by the participants. Any change in the project influences the verification of legal compliance, as it is necessary to validate and verify that the legal requirements were correctly specified according to the customer's needs and the law to avoid non-compliance.

The *Compliance requirements are purposefully expressed in general terms, omitting implementation-specific details* (13; 4.11%), therefore, it is important to have the participation and collaboration of legal experts, discussion among team members and the client for *Interpreting the regulation and translating it into implementable requirements* (11; 3.48%) and also avoid

Inconsistent requirements (17; 5.37%), *Incomplete and/or hidden requirements* (17 ; 5.37%).

Explanations to propositions **S. P8**, **S. P15** - **S. P16**, and **S. P25**:

E11 - A lack of collaboration between software engineers and lawyers affects projects with legal requirements (**S. P8**). Therefore, similar to what is done for ambiguity reduction, the company's legal department (**S. P16**) should be consulted to identify and analyze relevant laws and legal regulations (**S. P25**) (**S. P15**) and should be provided for the Requirements Engineer to help achieve legal compliance.

Explanations to propositions **S. P7**, **S. P26** - **S. P28**:

E12 - The Requirements Engineer needs to be aware of the updates that the laws undergo (**S. P7**) and regularly update the requirements specification (**S. P27**), the product backlog (**S. P26**), and discuss with the customer the impact of changes to the software (**S. P28**) to ensure legal compliance.

Explanations to propositions **S. P6**, **S. P21**, and **S. P29** - **S. P32**:

E13 - Lack of training on data protection regulation (**S. P6**) may lead to a specification of non-compliant legal requirements.

E14 - The Requirements Engineer and all stakeholders involved in the project must have basic knowledge about law for help to legal compliance (**S. P21**).

E15 - Initial training and practices should involve topics related to Privacy Principles (**S. P29**), Privacy by Design (**S. P30**), legislation specific to the domain of the software being specified (**S. P31**), and user rights from Privacy and Data Protection Regulations.

6.4 EVALUATION OF THEORY

Sjøberg et al. (SJØBERG et al., 2008) define concepts to evaluating of theories:

Testability - The degree to which a theory is constructed such that empirical refutation is possible;

Empirical Support - The degree to which a theory is supported by empirical studies that confirm its validity;

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 assumption, and relates to that which is already well understood;

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

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

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

Evaluating the theory defined in this Chapter, according to the criteria defined by Sjøberg et al. (SJØBERG et al., 2008), analyzing the **testability** we believe that the constructs and propositions of the theory are clear and precise, that is, they are understandable, internally consistent and free from ambiguities at least from the point of view of authors these study. The theory's scope conditions are explicitly specified. We intend to test the theory through interviews and test whether the theory's propositions are supported in other projects.

The theory has not been tested against **empirical support** by practitioners or researchers. Nevertheless, its constructs and propositions emerged from data from empirical research (interviews and surveys). Thus, we believe that the theory is partially supported empirically. It is possible to carry out several studies to confirm or complement our theory.

We consider the **explanatory power** of the theory low because many factors influence the specification of legal requirements with reduced ambiguity and in compliance with data protection laws.

A theory was derived from 22 interviews and 39 responses from an online survey. We used grounded theory to identify the factors, and then we developed propositions that were tested in a survey and gave rise to the theory. So, despite having a large body of knowledge, we have attempted to use a minimum of constructs and propositions in this theory. We consider the **parsimony** of the theory as moderate.

The scope of this theory is restricted. Thus, we consider the **generality** of the theory to be low.

This theory emerged from the opinions of professionals, either through interviews or surveys. Thus, we believe that the theory can be used in decision-making in projects that deal with data

protection laws' specifications and legal requirements. We consider the **utility** of the theory as high.

6.5 CHAPTER SUMMARY

This Chapter, presented an initial theory of how IT companies address ambiguity resolution and compliance with data protection laws in the requirements specification. Relationships that were supported by all data and that included the candidate constructs were aggregated into propositions. We elaborate explanation through a relationship among constructs, and other categories, which are not central enough to become a constructs. Then, we performed an evaluation, according to the criteria of Sjøberg et al. (SJØBERG et al., 2008).

7 MITIGATION ACTIONS AND GUIDELINES

Wiegers and Beatty (WIEGERS; BEATTY, 2013) describes that every IT professional needs to acquire a set of techniques that can be used to face possible challenges in a project, and the lack of this guide makes the professional need to discover an approach based on what seems reasonable at the moment. This forces engineers to choose among diverse, sometimes contradicting, approaches and their best to integrate them (GÜRSES, 2014).

Engelmann (ENGELMANN et al., 2020) defined a set of best practices to support development teams in the elicitation of user requirements. Notary et al. (NOTARIO et al., 2015) analyzed existing best practices in the analysis and design stages of the system development life-cycle, introduced a systematic methodology for privacy engineering, named PRIPARE, that merges and integrates them, and described its alignment with current standardization efforts.

Thus, from the studies presented above in literature (Systematic Literature Mapping and Snowballing) and practice (Interviews and Survey), we identified problems and challenges to specify legal requirements with reduced ambiguity and compliant with data protection regulations. Thus, based on these studies, we identify problems, define mitigation actions and describe guidelines for implementing them.

Table 31 to 39 shows the problems identified in the literature and practice (interviews and survey), frequency and percentage of responses, mitigation actions and guidelines, based on the constructs presented in Table 28.

These mitigation actions and guidelines can assist researchers in investigating best practices to prevent the problems mentioned. To practitioners, this study provides further insights into problems and how to prevent them. It is essential to highlight that mitigation actions and guidelines should not be considered a standard script that everyone must follow. However, as a guide, as each project has its particularities, not all of them can be applicable.

Table 31 presents the problems and mitigation actions related to category **Communication between stakeholders** and lack of collaboration between software engineers and lawyers, identified as one of the main problems affecting legal requirements engineering projects (shown in Table 14). The Business Analyst mediates the relationship between customer and user and the development team. Therefore, an analyst must be explicitly allocated to the project to maintain continuous monitoring during all phases until the delivery of the product (ENGELMANN et al., 2020).

Table 31 – Problems and mitigation actions related to Communication between stakeholders

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Communication flaws within the project development team	15;4,74%	Keeping developers well-informed about privacy. Standardization of terminology between law, engineering, and business Use agile communication techniques (Sprint Planning Meeting, Daily Meeting, Sprint Review Meeting, Sprint Retrospective) Assign a requirements engineer and a project manager to the project (with a high degree of experience and expertise). Involve the development team, validation, and verification team in the requirements phase Conduct training on basic legal knowledge of law for software engineers and the development team.
Communication flaws between team/developers and the customer	18;5.69%	Use agile communication techniques (Sprint Planning Meeting, Daily Meeting, Sprint Review Meeting, Sprint Retrospective) with the participation of the Customer's specialized team. Raise the level of abstraction with the customer using Prototyping Techniques. Introduce an early feedback loop with the customer. Conduct regular meetings with the customer with the participation of the Customer's specialized team. Planning and executing training to improve stakeholders skills on the business domain. Educate customers about the importance of privacy and data protection.
Lack of collaboration between software engineers and lawyers	19;6.01%	Provide legal support to software development in the company and delegate a legal person for tracing laws and legal regulations. Assign a requirements engineer to the project (with a high degree of experience and expertise). Use agile communication techniques (Sprint Planning Meeting, Daily Meeting, Sprint Review Meeting, Sprint Retrospective) with the participation of lawyers or representatives of the legal sector. Provide basic requirements engineering knowledge for lawyers (understand the SDLC briefly). Conduct training on basic legal knowledge of law for software engineers and the development team.

Source: The author (2021).

Table 32 present the problems and mitigation actions related to **Achieving legal compliance**.

Table 32 – Problems and mitigation actions related to Achieving legal compliance

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Lack of training on data protection regulations	20 ; 6.3%	The Company must define an institutional privacy training schedule on personal data protection regulations Conduct training on Data Protection Impact Assessment (DPIA) and prior consultation. Conduct training on Laws and regulations related to the software's subject area (e.g., patient records law, Privacy and Electronic Communications Regulation (e-Privacy)). Conduct training on the rights of data subjects (GDPR Articles 12-23, and LGPD Articles 17-22). Conduct training on Conditions for consent (GDPR Articles 7-8, and Article 8 of LGPD).
Compliance requirements are purposefully expressed in general terms, omitting implementation-specific details.	13;4.11%	Involve the Requirements Engineer, development team, and validation and verification team in the requirements phase. Customer, DPO, Development team member, internal Legal Department or internal Privacy Department validates understanding of the legal requirements. Conduct training on Information Privacy Framework Training (e.g., ISO/IEC 29132:2017, NIST Privacy Framework). Conduct regular meetings with the customer with the participation of the Customer's specialized team.
Lack of a Culture of Compliance	20 ; 6.3%	Conduct training on Penalties and sanctions of Data Protection Laws (GDPR Article 84 and LGPD Articles 52-54). Plan and carry out training and/or awareness actions on respecting users' privacy and that they are aware of the adequacy of these actions to GDPR/LGPD. Use Privacy-enhancing techniques.
Difficulty understanding domain-specific terms	13;4.11%	Internal Legal department, DPO or Internal Privacy department support data protection laws. Conduct regular meetings with the customer with the participation of the Customer's specialized team. Conduct training on basic legal knowledge of law for Software Engineers and the Development Team. Form multidisciplinary teams. Define a data dictionary for all domain-specific definitions and acronyms (Glossary).

Source: The author (2021).

The organization's information privacy requirements and guidelines outline the importance of privacy to users and the organization. Showing that complying with Data Protection laws, in addition to being mandatory, can be considered a differential against competitors.

Compliance culture plays a vital role in inculcating compliance (ABDULLAH; SADIQ; INDULSKA, 2010). A good culture, though difficult to achieve, can promote a positive attitude towards legal compliance activity at all levels within an organization (MORTON, 2005). To foster a culture of privacy, an organization must clearly articulate privacy as an organizational priority; communicate key privacy and security messages; educate across the organization; raise awareness of the importance of registering privacy incidents and breaches, and make privacy information and guidance readily accessible (POWER, 2007).

Table 33 and 34 presents the problems and mitigation actions related to **Working with Data Protection Regulation** identified in the interviews. Constant changes in the law (changing goals, business processes, and/or requirements), laws subjective and difficult to understand make legal compliance difficult.

Table 33 – Problems and mitigation actions related to Working with Data Protection Regulation

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Constant changes in the law (Changing goals, business processes, and/or requirements) make legal compliance difficult	10;3.16%	Regularly update the product backlog and legal requirements specification. Explain impact of changes to Customers and decide the best approach to updating the product backlog Assign a requirements engineer and a project manager to the project (with a high degree of experience and expertise). Conduct regular meetings with the customer with the participation of the Customer's specialized team. Internal Legal Department, DPO, or internal Privacy Department provides support with data protection laws. Privacy by design and default training.

Source: The author (2021).

Lack of Perception of Compliance as a Value-add embedded in their business processes, those organizations, having documented their business activities, argue that they see no returns for the time-consuming and expensive documentation. It is required that regulations and legislations are interpreted to, and mapped to, business processes by experts who deeply understand the organization's legal and operational aspects (ABDULLAH; SADIQ; INDULSKA, 2010).

Table 34 – Problems and mitigation actions related to Working with Data Protection Regulation (continued)

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Lack of Perception of Compliance as a Value-add	- ; -	Introduce Data Protection Laws-related processes and rules for new employees (onboarding). Conduct training on the lawfulness of processing (presented in GDPR Article 6, and LGPD Article 7). Employees (Developers, Project leaders, Architects) should understand privacy and information security. Plan and carry out training and/or awareness actions on respecting for the privacy of users, and that they are aware of the adequacy of these actions to GDPR/LGPD. Conduct training in Processing special categories of personal data (GDPR Articles 9 and 10, and LGPD Articles 11 and 14). Training on data controllers and handlers (GDPR Articles 24-43, and LGPD 37-40).
We are not aware of Data Protection Regulations (F10) / Company is not concerned about complying legally (F19)	- ; -	Perform actions to promote organizational awareness of the importance of privacy. Elect a DPO or lawyer to help comply with Data Protection Laws. Conduct training on Penalties and sanctions of Data Protection Laws (GDPR Article 84 and LGPD 52-54). Create policies for your organization and keep them updated. Conduct Training on the organization's own internal privacy protocols Conduct Training on Mandatory business/sector/industry requirements and code of conduct. Plan and carry out training and/or awareness actions on respecting for the privacy of users, and that they are aware of the adequacy of these actions to GDPR/LGPD.

Source: The author (2021).

Table 35 presents the problems when there is no one from the Legal Support within the Development Team, and when the legal sector is not just dedicated to software-related privacy, security, or software systems compliance issues (C16). Besides, mitigation actions related to **Specialized Support Area** identified in the interviews. Therefore, it is important that stakeholders have knowledge, even if basic, about how data protection laws work. Nevertheless, iteratively, transform legal regulations into legal requirements in cooperation between lawyers and Software Engineers.

Table 35 – Problems and mitigation actions related to Specialized Support Area

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
There is no one from the Legal Support within the Development Team	10;3.16%	Identification of relevant laws and legal regulations and their analysis carried out by lawyers Conduct training on basic legal knowledge of law for software engineers and the development team. Critical requirements discussed in administrative and case law through vulnerabilities in existing legal breaches of software systems.
The legal sector is not just dedicated to software-related privacy, security, or software systems compliance issues (C16).	10;3.16%	Iteratively, transform legal regulations into legal requirements in cooperation between lawyers and Software Engineers. Participation of the customer's specialized team Participation of Technical support area Privacy by design and privacy by default training. Identification of relevant laws and legal regulations and their analysis carried out by lawyers. Multidisciplinary team according to the needs of the project. Conduct training on basic legal knowledge of law for software engineers and the development team.

Source: The author (2021).

Table 36 and 37 presents the Problems and mitigation actions related to **Reducing Ambiguity** identified in the interviews. When non-functional requirements are unclear or unmeasurable, must have on the participation of experienced professionals such as Software Architects and Business Analysts. Moreover, provide training on Legal Requirements Documentation(Structured list of requirements, use case, user stories, prototypes, among others). To deal with underspecified requirements that are too abstract and allow for various interpretations, the Customer, DPO, Development Team member, internal Legal Department, or internal Privacy Department validates understanding of the legal requirements.

In order to prevent the developer may making misinterpretations, conduct peer reviews with appropriate inspection methods (e.g., checklists), ideally involving different stakeholders (e.g., users, designers, and testers) in the verification and validation. To reduce the difficulty of understanding domain-specific terms, define a data dictionary for all domain-specific definitions and acronyms (Glossary) and conduct regular meetings should be conducted with the Customer's specialized team to participate.

Table 36 – Problems and mitigation actions related to Reducing ambiguity

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Non-functional requirements are unclear or unmeasurable	11;3.48%	Participation of the Technical support area. Provide training on Legal Requirements Documentation (Structured list of requirements, use case, user stories, prototypes, among others). Critical requirements discussed in administrative and case law through vulnerabilities in existing legal breaches of software systems.
Underspecified requirements that are too abstract and allow for various interpretations	10;3.16%	Provide training on Legal Requirements Documentation(Structured list of requirements, use case, user stories, prototypes, among others). Conduct training in ambiguity identification techniques. Iteratively, transform legal regulations into legal requirements in cooperation between lawyers and Software Engineers. Customer, DPO, Development Team member, internal Legal Department or internal Privacy Department validates understanding of the legal requirements.
The developer may make misinterpretation	10;3.16%	Conduct peer reviews with appropriate inspection methods (e.g., checklists), ideally involving different stakeholders (e.g., users, designers, and testers) in the verification and validation. Conduct training on basic legal knowledge of law for software engineers and the development team. Internal Legal department, DPO, or Internal Privacy department, Customer and Requirements Engineer interprets or resolves ambiguities. Conduct training in ambiguity identification techniques. Reuse knowledge acquired from other similar systems as a basis for defining the requirements. Define a data dictionary for all domain-specific definitions and acronyms (Glossary)

Source: The author (2021).

Furthermore, when the law is subjective and difficult to understand, ask for clarification from another company sector and consult Government Agencies and other internal sectors. Support from departments specialized in Ambiguity Analysis formed by analysts who do not effectively participate in the team and encourage integration between teams working on similar topics mitigate actions to reduce ambiguity.

Table 37 – Problems and mitigation actions related to Reducing ambiguity (continued)

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Difficulty understanding domain-specific terms	13;4.11%	Conduct training on Laws and regulations related to the software's subject area (e.g., patient records law, Privacy and Electronic Communications Regulation (e-Privacy)). Internal Legal department, DPO or Internal Privacy department support data protection laws. Internal Legal department, DPO, or Internal Privacy department, Customer and Requirements Engineer interprets or resolves ambiguities. Conduct regular meetings with the customer with the participation of the Customer's specialized team. Delegation of a person for tracing laws and legal regulations. Conduct training on basic legal knowledge of law for Software Engineers and the Development Team. Form multidisciplinary teams. Define a data dictionary for all domain-specific definitions and acronyms (Glossary).
The law is subjective and difficult to understand (F16)	- ; -	Consult Government Agencies and other internal sectors / Government Authority. Direct contact with the customer involved with the project and with the participation of the Customer's specialized team. Ask clarification another sector of the company. Assign a requirements engineer and a project manager to the project (with a high degree of experience and expertise). Reusable catalog of legal requirements that were derived from specific legal texts regarding security and personal data protection. Create a Software Requirements Specification (SRS) template containing a legal requirements section and complete specifications of law-related system requirements. Department specialized in Ambiguity Analysis formed by analysts who do not effectively participate in the team. Encouraging integration between teams working on similar topics. Customer, DPO, Development Team member, internal Legal Department or internal Privacy Department validates understanding of the legal requirements.

Source: The author (2021).

Table 38 and 39 presents the problems and mitigation actions related to **Requirements Specification** identified in the interviews.

Lack of traceability between requirements and legal text is one of the main problems identified. Thus, to mitigate it, one must create a Software Requirements Specification (SRS) template containing a legal requirements section and complete specifications of law-related system requirements. Conduct peer reviews with appropriate inspection methods (e.g., checklists), ideally involving different stakeholders (e.g., users, designers, and testers) in the verification and validation steps.

To deal with incomplete and/or hidden requirements (**P10**), should conduct peer reviews with appropriate inspection methods (e.g., checklists), ideally involving different stakeholders (e.g., users, designers, and testers) in the verification and validation. Customer, DPO, Development Team member, internal Le-gal Department or internal Privacy Department validates the legal requirements specification.

To avoid inconsistent requirements (**P11**), the verification team and Product Owner should discuss the specification to identify and adjust any deviations before the specification goes into development. Conduct regular meetings with the Customer with the participation of the Customer's specialized team. Furthermore, invest more time in the requirements specification, using scenarios and prototypes to gather requirements more thoroughly.

Table 38 – Problems and mitigation actions related to Requirements Specification

Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Lack of traceability between requirements and legal text	12;4.11%	Create a Software Requirements Specification (SRS) template containing a legal requirements section and complete specifications of law-related system requirements. Provide training for the verification team to operate the CASE tools adopted to record artifacts and trace links. Conduct peer reviews with appropriate inspection methods (e.g., checklists), ideally involving different stakeholders (e.g., users, designers, and testers) in the verification and validation. Provide training on Legal Requirements Documentation (Structured list of requirements, use case, user stories, prototypes, among others).

Source: The author (2021).

Table 39 – Problems and mitigation actions related to Requirements Specification (continued)

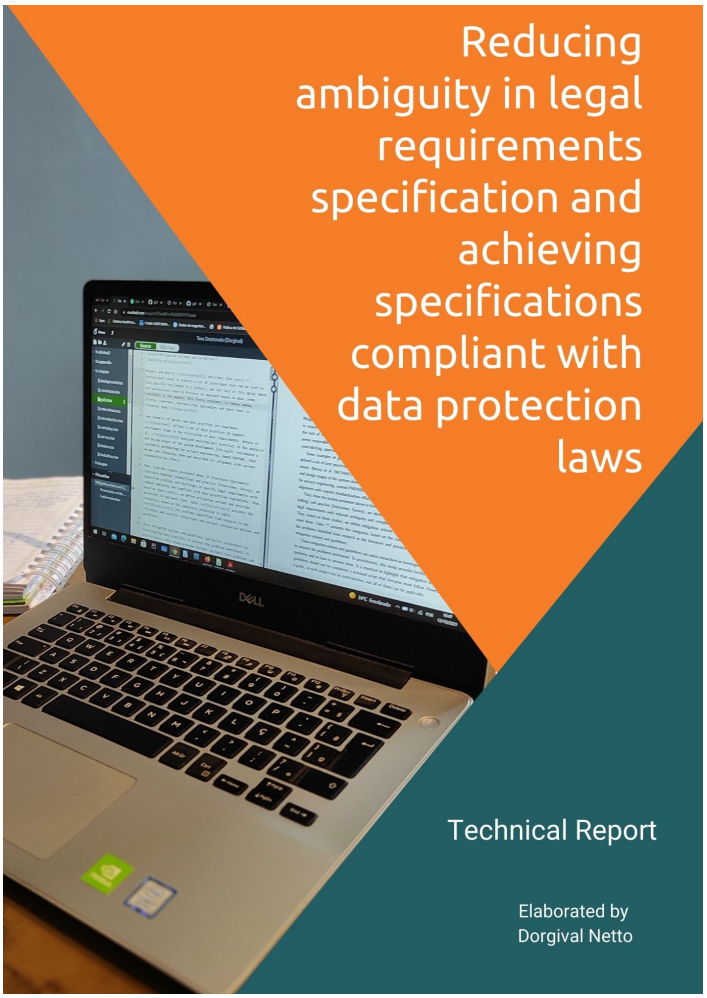
Problems	**(Freq. ; Perc.)**	**Mitigation actions**
Incomplete and/or hidden requirements (P10)	17;5.37%	Conduct peer reviews with appropriate inspection methods (e.g., checklists), ideally involving different stakeholders (e.g., users, designers, and testers) in the verification and validation. Customer, DPO, Development Team member, internal Legal Department or internal Privacy Department validates the legal requirements specification. Conduct training on basic legal knowledge of law for software engineers and the development team. Conduct regular meetings with the Customer with the participation of the Customer's specialized team. The verification team and Product Owner should discuss the specification to identify and adjust any deviations before the specification goes into development. Assign a req. eng. and a proj. manager to the project (with a high degree of experience and expertise). Provide training on Legal Requirements Documentation (Structured list of requirements, use case, user stories, prototypes, among others). Invest more time in the requirements specification, using scenarios and prototypes to gather requirements more thoroughly.
Inconsistent requirements (P11)	17;5.37%	Create a Software Requirements Specification (SRS) template containing a legal requirements section and complete specifications of law-related system requirements. Create a knowledge base of recurring defects. Assign a requirements engineer and a project manager to the project (with a high degree of experience and expertise). Customer, DPO, Development Team member, internal Legal Department or internal Privacy Department validates the legal requirements specification. The verification team and Product Owner should discuss the specification to identify and adjust any deviations before the specification goes into development. Conduct regular meetings with the customer with the participation of the Customer's specialized team. Conduct training on basic legal knowledge of law for software engineers and the development team. Invest more time in the requirements specification, using scenarios and prototypes to gather requirements more thoroughly.

Source: The author (2021).

7.1 GUIDE FOR PROFESSIONALS

We developed a version of the guide (see Figure 45) for professionals in order to facilitate the use and dissemination of the results of this thesis. The guide is available through the link[1] and in Supplementary Material (NETTO; SILVA, 2021) and will be updated as new validation steps of this work are carried out and, consequently, new knowledge is generated.

Figure 45 – Guide for Professionals



Source: The author (2021).

7.2 GUIDE EVALUATION

Practitioners have not evaluated the mitigation actions and the guide presented in the previous sections through empirical studies concerning their adequacy to address specific problems

[1] https://www.flipsnack.com/ambguide/mitigation-actions.html

in specific contexts. Therefore, we designed an assessment to be carried out by industry professionals about information quality (IQ), presented in Appendix H.

Gharib et al. (GHARIB; GIORGINI; MYLOPOULOS, 2018) present that IQ is a key success factor for most business processes since low-quality information may result in undesirable outcomes, or it might even prevent the business process from achieving its goals.

For the elaboration and evaluation of the guide with professionals about the quality of information, we will use the attributes of information quality: clarity, readability, accuracy, completeness, reliability, consistency, relevancy, usefulness, understandability, interpretability, informativeness (FERREIRA, 2011). Below we present the definitions for each attribute of information quality to be used in the evaluation.

Clarity: refers to the ability to present facts, things, data in a clear, distinct, and intelligible way. Information is qualified as clear or obscure (DELONE; MCLEAN, 1992).

Readability: refers to the sharpness of the calligraphic or typographic representation of the information record to allow it to be easily read. The information qualifies as readable, or ineligible (DELONE; MCLEAN, 1992).

Accuracy: refers to information free from error or misunderstanding. Information can be qualified as accurate or imprecise (DELONE; MCLEAN, 1992).

Completeness: characterizes how there is no lack of parts or elements of those that constitute it or those that it must have. Information may be qualified as complete or incomplete (DELONE; MCLEAN, 1992).

Reliability: comprises the ability to deliver as promised, safely and accurately. The information can be qualified as trustworthy or untrustworthy (DELONE; MCLEAN, 1992).

Consistency: indicates the existence of logical consistency and conformity to facts. Information is qualified as coherent or inconsistent (JARKE; VASSILIOU, 1997).

Relevancy: refers to the applicability of the information about what is being considered or discussed, indicates that the information has a significant and demonstrable influence on the subject in question (DELONE; MCLEAN, 1992).

Usefulness: characterizes information that has some use. The information can be qualified as useful or useless (DELONE; MCLEAN, 1992).

Understandability: is the capacity of information to be understood, learned, understood. Information qualifies as understandable or incomprehensible (DELONE; MCLEAN, 1992).

Interpretability: refers to the degree of difficulty the user may have to understand, correctly use and analyze the information provided (JARKE; VASSILIOU, 1997).

Informativeness: is the ability to provide meaningful data and information for the intended purpose (DELONE; MCLEAN, 1992).

The questionnaire, based on Lee et al. (LEE et al., 2002), can be used as an evaluation instrument, uses a five-point Likert scale, where 1 - strongly disagree and 5 - strongly agree, and a choice based on agreement and participant's beliefs. The instrument outline is presented in Appendix H.

7.3 CHAPTER SUMMARY

This Chapter presented actions to mitigate the problems identified in the literature (Systematic Literature Mapping, Snowballing) and practice (Interviews, Survey to specify legal requirements with reduced ambiguity and compliant with data protection regulations. Next, we developed a guide for professionals that summarizes the findings of this thesis and aims to disseminate the results to be used in practice.

8 CONCLUSION AND FUTURE WORKS

This chapter aims to present the final considerations on the main topics covered in this thesis, including the contributions reached and indications for future work.

Ambiguity in legal requirements and software requirements specification is a well-known problem in both academic and industry communities. Privacy is a matter that deserves attention from everyone within the company because it is a point of vulnerability in the actions that it performs in the company. Below we present the answers to each research question.

RQ1. *What are the existing approaches to deal with the ambiguity in the specification of legal requirements?*

We respond to RQ1 by performing a Systematic Mapping of the Literature on Security and Privacy in Requirements Engineering. Then, we carried out a snowballing in works that deal with techniques for treating and reducing ambiguity in the specification of legal requirements. This second study helped us to define the research problem and identify related work.

RQ2. *How do organizations deal with ambiguity in legal requirements specification and achieve law compliance?*

To answer RQ2, we investigate how public and private companies treat legal compliance and address ambiguity in specifying legal requirements through a study with 22 professionals from public and private companies through semi-structured interviews.

From the interviews, we identified 27 factors that influence (positively or negatively) the categories that emerged from the data. The relationships between factors and categories generated eight propositions, which can be seen as recommendations during the privacy requirement specification with reduced ambiguity and compliance with data protection laws.

The Requirements elicitation session occurs through interviews. Therefore, closer contact with the client is necessary because the customer presents the legal framework (laws, norms, standards, among others) related to its activity area. Training on data privacy and personal data protection regulation helps the requirements analyst to identify software requirements. The requirements analyst needs to count com multidisciplinary team according to the needs of the project and members with a broader understanding of the customer's business, the legal support in the company, participation of the customer's specialized team, and participation from technical support area.

The procedures for reducing ambiguity in legal requirements vary among companies, consult

the customer or Government Agencies and other Internal Sectors, analyze case law, request the participation of the Customer's Specialized Team. Team member studying personal data protection laws.

Techniques for Requirements Specification in Public Companies from describing use case, user stories, requirements document, IEEE requirements specification template, and the requirement recorded in their tools are always natural language. In Private Companies, specify requirements with user stories with acceptance criteria. The Customer or Customer's Specialized Team continuously validates the requirements analyst's interpretation or others involved in the project for requirements specification. We emphasize that user requirements specification techniques can improve and reduce ambiguity.

Most companies verify legal compliance after the requirements specification, mostly with well-defined acceptance criteria, when the functionality is written in user stories. Lawyers should validate the system from a legal perspective, and it should confirm that the entire system complies with the law. Validation with examining the impact of law-related requirements on the system by developers and customers should check whether such scope of the system is satisfactory for customers and feasible for the development team in given project constraints.

So, we affirm that it is not enough to have only guidelines to specify requirements with reduced ambiguity and compliance with the legislation. It is necessary to have a culture of privacy in the company. As laws are changed frequently, new regulations emerge, and employees in all company sectors need to be aware that it is necessary to keep up-to-date. This cultural change begins with understanding the principles and rules of the Personal Data Protection Laws, such as the Lei Geral de Proteção de Dados (LGPD). Training and campaigns are essential for improving awareness concerning privacy compliance techniques employed, such as accountability (records and audit trails) and data mapping.

RQ3. *What practices are defined in academia and industry to address ambiguity in legal requirements and specify legal compliance systems?*

We conduct a cross-sectional survey through a self-administered questionnaire online since the participants belong to different countries and the Brazilian States. The survey aims to collect the software practitioners' perceptions regarding the factors and actions to achieve ambiguity resolution and legal compliance in a software requirements specification.

We surveyed 39 professionals to corroborate the practices to tackle ambiguity and legal compliance represented in the systematic literature review, snowballing, and the twenty-two interviews. Moreover, identify the techniques, methods, and tools used to specifying legal

requirements with reduced ambiguity, and validate and verify legal compliance. Then, we surveyed 50 professionals to verify the software practitioners' perceptions regarding the factors and actions identified from a set of interviews.

From the survey result, we elaborated explanations for the constructs and propositions. We developed an initial theory of how IT Companies address ambiguity resolution and compliance with Data Protection Laws in the requirements specification. Our study Establishes a theory that can be used as starting point for further studies for more detailed investigations. Practitioners can use the results as a guide on selecting suitable methods and techniques to specify legal requirements with reduced ambiguity compliant with data protection laws.

RQ4. *How to produce a requirements specification with reduced ambiguity and legal compliance?*

Analyzing the results of previous studies, we identified problems that affect ambiguity in the specification of legal requirements and compliance with legislation. Thus, we define mitigation actions and describe guidelines to implement them.

8.1 PRACTICAL IMPLICATIONS

As benefits for the academy, we can mention the detailed research method can guide other researchers in carrying out their studies. The research method includes the steps for conducting and analyzing the set of interviews (Chapter 4), the survey (Chapter 5), and the definition of the theory (Chapter 6).

Systematic literature mapping and snowballing can provide evidence for further research. They can be updated in the future. Furthermore, the interviews-based study and the survey can be replicated to validate the results in more companies and/or other countries.

Benefits of this study to practitioners include presenting how ambiguity is addressed by companies and how they achieve legal compliance of software requirements. Based on the factors discovered in the interviews-based study, companies can encourage positive factors, and mitigate the negative ones. Moreover, practitioners can compare these results with their own experiences and practices and improve their own processes and techniques.

This thesis made use of multiple research methods, such as systematic literature mapping, snowballing, interview, and survey. Such research strategies enabled the generation of empirical results based on evidences, which can support the software practitioners to improve their actions to achieve ambiguity resolution and legal compliance in a software requirements

specification. Additionally, software researchers can identify challenges in the area that can be addressed in the future.

8.2 FUTURE WORK

As future works, we also intend to update the Systematic Literature Mapping to cover the period 2016 to 2021, comparing the findings as (AKHIGBE; AMYOT; RICHARDS, 2019) and (NETTO; PEIXOTO; SILVA, 2019). The purpose is to corroborate the practices presented in the literature and industry to tackle ambiguity in legal requirements specification and achieve legal compliance in the early stages of the software development process to provide guidelines and best practices to improve legal requirements specification compliant with the law.

We present thirty-two propositions that can be recommendations and define an initial theory explaining industry practices to reduce ambiguity and achieve legal compliance of requirement specifications with data privacy laws. One future work is to use the theory to analyze and, perhaps, improve approaches aimed at promoting regulatory compliance in RE as, for example, the work of Kosenkov et al., (KOSENKOV et al., 2021).

The mitigation actions and guidelines require additional industry experts to evaluate their adequacy to address specific problems in specific contexts. Also, it is necessary to evaluate their application in controlled experiments and industrial case studies and analyse if they fit the particularities of industrial context and the practices used therein.

The Guide for Professionals proposed need to be evaluated regarding the quality of information, by using the questionnaire presented in Appendix H, for example, and attributes such as clarity, readability, accuracy, completeness, reliability, consistency, relevance, usefulness, understandability, interpretability, and informativeness.

REFERENCES

ABDULLAH, N. S.; SADIQ, S.; INDULSKA, M. Emerging challenges in information systems research for regulatory compliance management. in: Pernici b (hrsg) advanced information systems engineering, s 251-265. In: . [S.l.: s.n.], 2010. ISBN 978-3-319-98176-5.

ABU-NIMEH, S.; MEAD, N. R. Privacy risk assessment in privacy requirements engineering. *2009 Second International Workshop on Requirements Engineering and Law*, 2009.

AKHIGBE, O.; AMYOT, D.; RICHARDS, G. A systematic literature mapping of goal and non-goal modelling methods for legal and regulatory compliance. *Requirements Engineering*, v. 24, 12 2019.

ANSARI, M. T. J.; BAZ, A.; ALHAKAMI, H.; ALHAKAMI, W.; KUMAR, R.; KHAN, R. A. P-store: Extension of store methodology to elicit privacy requirements. *Arabian Journal for Science and Engineering*, Springer, p. 1–24, 2021.

ANTON, A. I. Goal-based requirements analysis. In: IEEE. *Proceedings of the second international conference on requirements engineering*. [S.l.], 1996. p. 136–144.

ARGENTINA. *PERSONAL DATA PROTECTION ACT ACT 25.326*. 2000. Disponível em: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/norma.htm>.

AYALA-RIVERA, V.; PASQUALE, L. The grace period has ended: An approach to operationalize gdpr requirements. In: . [S.l.: s.n.], 2018. p. 136–146.

BANO, M. Addressing the challenges of requirements ambiguity: A review of empirical literature. In: IEEE. *2015 IEEE Fifth International Workshop on Empirical Requirements Engineering (EmpiRE)*. [S.l.], 2015. p. 21–24.

BEDNAR, K.; SPIEKERMANN, S.; LANGHEINRICH, M. Engineering privacy by design: Are engineers ready to live up to the challenge? *The Information Society*, Routledge, v. 35, n. 3, p. 122–142, 2019. Disponível em: <https://doi.org/10.1080/01972243.2019.1583296>.

BEECHAM, S.; HALL, T.; RAINER, A. Software process improvement problems in twelve software companies: An empirical analysis. *Empirical software engineering*, Springer, v. 8, n. 1, p. 7–42, 2003.

BERRY, D. M.; KAMSTIES, E. Ambiguity in requirements specification. In: _____. *Perspectives on Software Requirements*. Boston, MA: Springer US, 2004. p. 7–44. ISBN 978-1-4615-0465-8. Disponível em: <https://doi.org/10.1007/978-1-4615-0465-8_2>.

BERRY, D. M.; KAMSTIES, E.; KRIEGER, M. From contract drafting to software specification: Linguistic sources of ambiguity. In: . [S.l.: s.n.], 2003.

BHATIA, J.; BREAUX, T.; REIDENBERG, J.; NORTON, T. A theory of vagueness and privacy risk perception. In: . [S.l.: s.n.], 2016. p. 26–35.

BIJWE, A.; MEAD, N. Adapting the square process for privacy requirements engineering. 05 2021.

BLIX, F.; ELSHEKEIL, S. A.; LAOYOOKHONG, S. Data protection by design in systems development: From legal requirements to technical solutions. In: IEEE. *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*. [S.l.], 2017. p. 98–103.

BOELLA, G.; HUMPHREYS, L.; MUTHURI, R.; ROSSI, P.; TORRE, L. van der. A critical analysis of legal requirements engineering from the perspective of legal practice. *2014 IEEE 7th International Workshop on Requirements Engineering and Law, RELAW 2014 - Proceedings*, p. 14–21, 08 2014.

BOUDREAU, M.-C.; GEFEN, D.; STRAUB, D. W. Validation in information systems research: A state-of-the-art assessment. *MIS quarterly*, JSTOR, p. 1–16, 2001.

BRASIL. *LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011*. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>.

BRASIL. *LEI Nº 13.709 Lei Geral de Proteção de Dados Pessoais (LGPD)*. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>.

BREAUX, T. D.; ANTÓN, A. I. A systematic method for acquiring regulatory requirements: A frame-based approach. *RHAS-6), Delhi, India*, 2007.

BREAUX, T. D.; ANTÓN, A. I.; SPAFFORD, E. H. A distributed requirements management framework for legal compliance and accountability. *Computers and Security*, v. 28, n. 1–2, p. 8–17, fev.–mar. 2008.

CANEDO, E. D.; CALAZANS, A. T. S.; MASSON, E. T. S.; COSTA, P. H. T.; LIMA, F. Perceptions of ict practitioners regarding software privacy. *Entropy*, v. 22, n. 4, 2020. ISSN 1099-4300. Disponível em: <https://www.mdpi.com/1099-4300/22/4/429>.

CARAMUJO, J.; SILVA, A.; MONFARED, S.; RIBEIRO, A.; CALADO, P.; BREAUX, T. Rsl-il4privacy: a domain-specific language for the rigorous specification of privacy policies. *Requirements Engineering*, v. 24, 03 2019.

CAROLI, P. Lean inception: como alinhar pessoas e construir o produto certo. *São Paulo: Editora Caroli*, 2018.

CASELLAS, N. *Legal ontology engineering: Methodologies, modelling trends, and the ontology of professional judicial knowledge*. [S.l.]: Springer Science & Business Media, 2011. v. 3.

CAVOUKIAN, A. Understanding how to implement privacy by design, one step at a time. *IEEE Consumer Electronics Magazine*, IEEE, v. 9, n. 2, p. 78–82, 2020.

CAVOUKIAN, A. et al. Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, v. 5, p. 12, 2009.

Centers for Medicare & Medicaid Services. *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. 1996. Online at http://www.cms.hhs.gov/hipaa/.

COMPAGNA, L.; KHOURY, P. E.; KRAUSOVá, A.; MASSACCI, F.; ZANNONE, N. How to integrate legal requirements into a requirements engineering methodology for the development of security and privacy patterns. *Artificial intelligence and law*, Springer Netherlands, Dordrecht, v. 17, n. 1, p. 1–30, 2009. ISSN 0924-8463.

COOL, A. Impossible, unknowable, accountable: Dramas and dilemmas of data law. *Social Studies of Science*, v. 49, n. 4, p. 503–530, 2019. PMID: 31057059. Disponível em: <https://doi.org/10.1177/0306312719846557>.

CORBIN, J.; STRAUSS, A. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. [S.l.]: Sage publications, 2014.

CRESWELL, J. W. *Research design: Qualitative, quantitative, and mixed methods approaches*. [S.l.]: Sage, 2013.

DANIEL, W. W. et al. *Applied nonparametric statistics (revised ed.)*. [S.l.]: Duxbury, 2000.

DELONE, W. H.; MCLEAN, E. R. Information systems success: The quest for the dependent variable. *Information systems research*, INFORMS, v. 3, n. 1, p. 60–95, 1992.

DENG, M.; WUYTS, K.; SCANDARIATO, R.; PRENEEL, B.; JOOSEN, W. A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requir. Eng.*, v. 16, p. 3–32, 03 2011.

DEY, I. Grounding grounded theory. *Grounding Grounded Theory*, p. 249–269, 1999.

DEY, I. *Qualitative data analysis: A user friendly guide for social scientists*. [S.l.]: Routledge, 2003.

DILLMAN, D. A.; SMYTH, J. D.; CHRISTIAN, L. M. *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. [S.l.]: John Wiley & Sons, 2014.

EASTERBROOK, S.; SINGER, J.; STOREY, M.-A.; DAMIAN, D. *Selecting Empirical Methods for Software Engineering Research*. [S.l.]: Springer, 2007.

EKMAN, L. W.; BILLGREN, P. Compliance challenges with the general data protection regulation. 2017.

ENGELMANN, L. K. et al. Boas práticas para apoio ao processo de elicitação de requisitos de usuário no contexto da engenharia de software. Pontifícia Universidade Católica do Rio Grande do Sul, 2020.

ESAYAS, S.; MAHLER, T. Modelling compliance risk: a structured approach. *Artificial Intelligence and Law*, v. 23, p. 271–300, 09 2015.

FERNANDES, M.; SILVA, A. R. da; GONÇALVES, A. *Specification of Personal Data Protection Requirements: Analysis of Legal Requirements based on the GDPR Regulation*. 2018.

FERNÁNDEZ, D. M.; WAGNER, S.; KALINOWSKI, M.; FELDERER, M.; MAFRA, P.; VETRO, A.; CONTE, T.; CHRISTIANSSON, M.-T.; GREER, D.; LASSENIUS, C. et al. Naming the pain in requirements engineering: contemporary problems, causes, and effects in practice. 2016.

FERNÁNDEZ, D. M.; WAGNER, S.; KALINOWSKI, M.; FELDERER, M.; MAFRA, P.; VETRÒ, A.; CONTE, T.; CHRISTIANSSON, M.-T.; GREER, D.; LASSENIUS, C. et al. Naming the pain in requirements engineering. *Empirical software engineering*, Springer, v. 22, n. 5, p. 2298–2338, 2017.

FERREIRA, O. C. A. Atributos de qualidade da informação. In: _____. [S.l.: s.n.], 2011. p. 111. ISBN Dissertação (Mestrado)-Universidade de Brasília.

FIRESMITH, D. Common requirements problems, their negative consequences, and the industry best practices to help solve them. *J. Object Technol.*, v. 6, n. 1, p. 17–33, 2007.

FRANCH, X.; GLINZ, M.; MENDEZ, D.; SEYFF, N. A study about the knowledge and use of requirements engineering standards in industry. *IEEE Transactions on Software Engineering*, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2021. ISSN 2326-3881. Disponível em: <http://dx.doi.org/10.1109/TSE.2021.3087792>.

FRICKER, S. A.; SCHNEIDER, K.; FOTROUSI, F.; THUEMMLER, C. Workshop videos for requirements communication. *Requirements Engineering*, Springer, v. 21, n. 4, p. 521–552, 2016.

GERVASI, V.; SAWYER, P.; NUSEIBEH, B. Unknown knowns: Tacit knowledge in requirements engineering. In: IEEE COMPUTER SOCIETY. *2011 IEEE 19th International Requirements Engineering Conference*. [S.l.], 2011. p. 329–329.

GERVASI, V.; ZOWGHI, D. On the role of ambiguity in re. In: SPRINGER. *International Working Conference on Requirements Engineering: Foundation for Software Quality*. [S.l.], 2010. p. 248–254.

GHANAVATI, S.; AMYOT, D.; RIFAUT, A. Legal goal-oriented requirement language (legal grl) for modeling regulations. In: *Proceedings of the 6th international workshop on modeling in software engineering*. [S.l.: s.n.], 2014. p. 1–6.

GHARIB, M.; GIORGINI, P.; MYLOPOULOS, J. Analysis of information quality requirements in business processes, revisited. *Requirements Engineering*, Springer, v. 23, n. 2, p. 227–249, 2018.

GHARIB, M.; MYLOPOULOS, J.; GIORGINI, P. Copri - a core ontology for privacy requirements engineering. In: _____. [S.l.: s.n.], 2020. p. 472–489. ISBN 978-3-030-50315-4.

GJERMUNDRØD, H.; DIONYSIOU, I.; COSTA, K. privacytracker: a privacy-by-design gdpr-compliant framework with verifiable data traceability controls. In: SPRINGER. *International Conference on Web Engineering*. [S.l.], 2016. p. 3–15.

GLASER, B. G. Theoretical sensitivity. mill valley. CA: Sociology Press, 1978.

GLINZ, M. On non-functional requirements. In: . [S.l.: s.n.], 2007. p. 21 – 26. ISBN 978-0-7695-2935-6.

GOVERNATORI, G. Law, logic and business processes. In: IEEE. *2010 Third International Workshop on Requirements Engineering and Law*. [S.l.], 2010. p. 1–10.

GÜRSES, S. Can you engineer privacy? *Communications of the ACM*, ACM New York, NY, USA, v. 57, n. 8, p. 20–23, 2014.

GURSES, S.; ALAMO, J. D. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security Privacy*, v. 14, p. 40–46, 03 2016.

GUTIÉRREZ, J. F.; BONACHE, C. P.; QUER, C. Industrial practices on requirements reuse: An interview-based study. In: SPRINGER. *Requirements Engineering: Foundation for Software Quality, 26th International Working Conference, REFSQ 2020: Pisa, Italy, March 24–27, 2020: proceedings*. [S.l.], 2020. p. 78–94.

HADAR, I.; HASSON, T.; AYALON, O.; TOCH, E.; BIRNHACK, M.; SHERMAN, S.; BALISSA, A. Privacy by designers: Software developers' privacy mindset. In: *Proceedings of the 40th International Conference on Software Engineering*. New York, NY, USA: Association for Computing Machinery, 2018. (ICSE '18), p. 396. ISBN 9781450356381. Disponível em: <https://doi.org/10.1145/3180155.3182531>.

HANNAY, J. E.; SJOBERG, D. I.; DYBA, T. A systematic review of theory use in software engineering experiments. *IEEE transactions on Software Engineering*, IEEE, v. 33, n. 2, p. 87–107, 2007.

HOFFMANN, A.; SCHULZ, T.; HOFFMANN, H.; JANDT, S.; ROSSNAGEL, A.; LEIMEISTER, J. M. Towards the use of software requirement patterns for legal requirements. 2012.

HOSSEINI, M. B.; HEAPS, J.; SLAVIN, R.; NIU, J.; BREAUX, T. Ambiguity and generality in natural language privacy policies. In: *2021 IEEE 29th International Requirements Engineering Conference (RE)*. [S.l.: s.n.], 2021. p. 70–81.

INAYAT, I.; SALIM, S. S.; MARCZAK, S.; DANEVA, M.; SHAMSHIRBAND, S. A systematic literature review on agile requirements engineering practices and challenges. *Computers in human behavior*, Elsevier, v. 51, p. 915–929, 2015.

ISO, B. *IEC 29100, 2011. BS ISO/IEC29100: Information technology—security techniques—privacy framework*. [S.l.], 2011. Disponível em: <https://www.iso.org/standard/45123.html>.

ISO, B. *ISO/IEC 27701:2019 Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines*. 2019. Disponível em: <https://www.iso.org/standard/71670.html>.

JARKE, M.; VASSILIOU, Y. Data warehouse quality: A review of the dwq project. In: *IQ*. [S.l.: s.n.], 1997. p. 299–313.

KALINOWSKI, M.; FELDERER, M.; CONTE, T.; SPÍNOLA, R.; PRIKLADNICKI, R.; WINKLER, D.; FERNÁNDEZ, D. M.; WAGNER, S. Preventing incomplete/hidden requirements: reflections on survey data from austria and brazil. In: SPRINGER. *International Conference on Software Quality*. [S.l.], 2016. p. 63–78.

KAMSTIES, E.; BERRY, D. M.; PAECH, B.; KAMSTIES, E.; BERRY, D.; PAECH, B. Detecting ambiguities in requirements documents using inspections. In: *Proceedings of the first workshop on inspection in software engineering (WISE'01)*. [S.l.: s.n.], 2001. p. 68–80.

KAMSTIES, E.; PEACH, B. Taming ambiguity in natural language requirements. In: *Proceedings of the Thirteenth international conference on Software and Systems Engineering and Applications*. [S.l.: s.n.], 2000.

KASURINEN, J.; TAIPALE, O.; SMOLANDER, K. Test case selection and prioritization: risk-based or design-based? In: *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. [S.l.: s.n.], 2010. p. 1–10.

KAVAKLI, E.; KALLONIATIS, C.; LOUCOPOULOS, P.; GRITZALIS, S. Incorporating privacy requirements into the system design process: the pris conceptual framework. *Internet research*, Emerald Group Publishing Limited, 2006.

KERRIGAN, S.; LAW, K. Logic-based regulation compliance-assistance. In: . [S.l.: s.n.], 2003. p. 126–135.

KHAN, N. F.; IKRAM, N. Security requirements engineering: A systematic mapping (2010-2015). *2016 International Conference on Software Security and Assurance (ICSSA)*, 2016.

KITCHENHAM, B.; PFLEEGER, S.; PICKARD, L.; JONES, P.; HOAGLIN, D.; EMAM, K. E.; ROSENBERG, J. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering*, v. 28, n. 8, p. 721–734, ago. 2002. ISSN 0098-5589.

KITCHENHAM, B.; PFLEEGER, S. L. Principles of survey research part 6: data analysis. *ACM SIGSOFT Software Engineering Notes*, ACM New York, NY, USA, v. 28, n. 2, p. 24–27, 2003.

KITCHENHAM, B. A.; CHARTERS, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. [S.l.], 2007. Disponível em: <https://www.elsevier.com/___data/promis_misc/525444systematicreviewsguide.pdf>.

KITCHENHAM, B. A.; PFLEEGER, S. L. Personal opinion surveys. In: _____. *Guide to Advanced Empirical Software Engineering*. London: Springer London, 2008. p. 63–92. ISBN 978-1-84800-044-5. Disponível em: <https://doi.org/10.1007/978-1-84800-044-5_3>.

KIYAVITSKAYA, N.; KRAUSOVá, A.; ZANNONE, N. Why eliciting and managing legal requirements is hard. In: . [S.l.: s.n.], 2008. p. 26 – 30.

KOOPS, B.-J.; LEENES, R. Privacy regulation cannot be hardcoded. a critical comment on the 'privacy by design'provision in data-protection law. *International Review of Law, Computers & Technology*, Taylor & Francis, v. 28, n. 2, p. 159–171, 2014.

KOSENKOV, O.; UNTERKALMSTEINER, M.; MENDEZ, D.; FUCCI, D. Vision for an artefact-based approach to regulatory requirements engineering. *arXiv preprint arXiv:2108.13059*, 2021.

KOTONYA, G.; SOMMERVILLE, I. *Requirements engineering: processes and techniques*. [S.l.]: John Wiley & Sons, Inc., 1998.

KUTYLOWSKI, M.; LAUKS-DUTKA, A.; YUNG, M. Gdpr challenges for reconciling legal rules with technical reality. In: SPRINGER. *European Symposium on Research in Computer Security*. [S.l.], 2020. p. 736–755.

LAMSWEERDE, A. V. Elaborating security requirements by construction of intentional anti-models. *Proceedings. 26th International Conference on Software Engineering*.

LE, T.; LE, C.; JEONG, H. D.; GILBERT, S. B.; CHUKHAREV-HUDILAINEN, E. Requirement text detection from contract packages to support project definition determination. In: *Advances in informatics and computing in civil and construction engineering*. [S.l.]: Springer, 2019. p. 569–576.

LEE, Y. W.; STRONG, D. M.; KAHN, B. K.; WANG, R. Y. Aimq: a methodology for information quality assessment. *Information Management*, v. 40, n. 2, p. 133–146, 2002. ISSN 0378-7206. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378720602000435>.

LI, Z. S.; WERNER, C.; ERNST, N.; DAMIAN, D. *GDPR Compliance in the Context of Continuous Integration*. 2020.

LUCIA, A. D.; QUSEF, A. Requirements engineering in agile software development. *Journal of emerging technologies in web intelligence*, Citeseer, v. 2, n. 3, p. 212–220, 2010.

MASSEY, A. K.; HOLTGREFE, E.; GHANAVATI, S. Modeling regulatory ambiguities for requirements analysis. In: MAYR, H. C.; GUIZZARDI, G.; MA, H.; PASTOR, O. (Ed.). *ER*. Springer, 2017. (Lecture Notes in Computer Science, v. 10650), p. 231–238. ISBN 978-3-319-69904-2. Disponível em: <http://dblp.uni-trier.de/db/conf/er/er2017.html#MasseyHG17>.

MASSEY, A. K.; RUTLEDGE, R. L.; ANTÓN, A. I.; HEMMINGS, J. D.; SWIRE, P. P. *A strategy for addressing ambiguity in regulatory requirements*. [S.l.], 2015.

MASSEY, A. K.; RUTLEDGE, R. L.; ANTóN, A. I.; SWIRE, P. P. Identifying and classifying ambiguity for regulatory requirements. In: GORSCHEK, T.; LUTZ, R. R. (Ed.). *RE*. IEEE Computer Society, 2014. p. 83–92. ISBN 978-1-4799-3033-3. Disponível em: <http://dblp.uni-trier.de/db/conf/re/re2014.html#MasseyRAS14>.

MASSEY, A. K.; SMITH, B.; OTTO, P. N.; ANTÓN, A. I. Assessing the accuracy of legal implementation readiness decisions. In: IEEE. *2011 IEEE 19th International Requirements Engineering Conference*. [S.l.], 2011. p. 207–216.

MAXWELL, J. C.; ANTóN, A. I.; SWIRE, P. P.; RIAZ, M.; MCCRAW, C. M. A legal cross-references taxonomy for reasoning about compliance requirements. *Requir. Eng.*, v. 17, n. 2, p. 99–115, 2012. Disponível em: <http://dblp.uni-trier.de/db/journals/re/re17.html#MaxwellASRM12>.

MEAD, N.; MIYAZAKI, S.; ZHAN, J. Integrating privacy requirements considerations into a security requirements engineering method and tool. *Int. J. of Information Privacy*, v. 1, p. 106 – 126, 01 2011.

MELLO, R. M. de; SILVA, P. C. da; TRAVASSOS, G. H. Investigating probabilistic sampling approaches for large-scale surveys in software engineering. *J. Softw. Eng. Res. Dev.*, v. 3, p. 8, 2015. Disponível em: <http://dblp.uni-trier.de/db/journals/jserd/jserd3.html#MelloST15>.

MERRIAM, S. B.; TISDELL, E. J. *Qualitative research: A guide to design and implementation*. [S.l.]: John Wiley & Sons, 2015.

MOLLÉRI, J. S.; PETERSEN, K.; MENDES, E. An empirically evaluated checklist for surveys in software engineering. *Information and Software Technology*, Elsevier, v. 119, p. 106240, 2020.

MORTON, J. C. The development of a compliance culture. *Journal of investment compliance*, Emerald Group Publishing Limited, 2005.

NEKVI, M.; MADHAVJI, N. Impediments to regulatory compliance of requirements in contractual systems engineering projects. *ACM Transactions on Management Information Systems*, v. 5, p. 1–35, 12 2014.

NEKVI, M. R. I.; MADHAVJI, N. H.; FERRARI, R.; BERENBACH, B. Impediments to requirements-compliance. In: SPRINGER. *International Working Conference on Requirements Engineering: Foundation for Software Quality*. [S.l.], 2012. p. 30–36.

NETTO, D.; PEIXOTO, M. M.; SILVA, C. Privacy and security in requirements engineering: Results from a systematic literature mapping. In: *Proceedings 22nd Workshop on Requirements Engineering - WER*. [S.l.: s.n.], 2019.

NETTO, D.; SILVA, C. Supplementary material. 2021. Disponível em: <"https://dorgivalnetto.github.io/phd_thesis/">.

NETTO, D.; SILVA, C.; ARAÚJO, J. Identifying how the brazilian software industry specifies legal requirements. In: *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*. [S.l.: s.n.], 2019. p. 181–186.

NETTO, D.; SILVA, C.; ARAÚJO, J. How information technology companies address ambiguity resolution and compliance with data protection laws in requirements specification? *submitted in evaluation*, 2021.

NOTARIO, N.; CRESPO, A.; MARTÍN, Y.-S.; ALAMO, J. M. D.; MÉTAYER, D. L.; ANTIGNAC, T.; KUNG, A.; KROENER, I.; WRIGHT, D. Pripare: integrating privacy best practices into a privacy engineering methodology. In: IEEE. *2015 IEEE Security and Privacy Workshops*. [S.l.], 2015. p. 151–158.

OMORONYIA, I.; FERGUSON, J.; ROPER, M.; WOOD, M. A review of awareness in distributed collaborative software engineering. *Software: Practice and Experience*, Wiley Online Library, v. 40, n. 12, p. 1107–1133, 2010.

ORAN, A. C.; SANTOS, G.; GADELHA, B.; CONTE, T. A framework for evaluating and improving requirements specifications based on the developers and testers perspective. *Requirements Engineering*, Springer, p. 1–28, 2021.

OTTO, P. N. Reasonableness meets requirements: Regulating security and privacy in software. *Duke LJ*, HeinOnline, v. 59, p. 309, 2009.

OTTO, P. N.; ANTóN, A. I. Addressing legal requirements in requirements engineering. In: *RE*. IEEE Computer Society, 2007. p. 5–14. ISBN 0-7695-2935-6. Disponível em: <http://dblp.uni-trier.de/db/conf/re/re2007.html#OttoA07>.

PALOMARES, C.; QUER, C.; FRANCH, X. Requirements reuse and requirement patterns: a state of the practice survey. *Empirical Software Engineering*, Springer, v. 22, n. 6, p. 2719–2762, 2017.

PANDIT, N. R. The creation of theory: A recent application of the grounded theory method. *The qualitative report*, Fort Lauderdale, v. 2, n. 4, p. 1–15, 1996.

PEIXOTO, M.; FERREIRA, D.; CAVALCANTI, M.; SILVA, C.; VILELA, J.; ARAÚJO, J.; GORSCHEK, T. On understanding how developers perceive and interpret privacy requirements research preview. In: MADHAVJI, N.; PASQUALE, L.; FERRARI, A.; GNESI, S. (Ed.). *Requirements Engineering: Foundation for Software Quality*. Cham: Springer International Publishing, 2020. p. 116–123. ISBN 978-3-030-44429-7.

PEIXOTO, M.; SILVA, C.; LIMA, R.; ARAúJO, J.; GORSCHEK, T.; SILVA, J. Pcm tool: Privacy requirements specification in agile software development. In: *Anais Estendidos do X Congresso Brasileiro de Software: Teoria e Prática*. Porto Alegre, RS, Brasil: SBC, 2019. p. 108–113. ISSN 2177-9384. Disponível em: <https://sol.sbc.org.br/index.php/cbsoft_estendido/article/view/7666>.

PERERA, C.; BARHAMGI, M.; BANDARA, A.; AJMAL, M.; PRICE, B.; NUSEIBEH, B. Designing privacy-aware internet of things applications. *Information Sciences*, v. 512, 09 2019.

PHALP, K. T.; VINCENT, J.; COX, K. Assessing the quality of use case descriptions. *Software Quality Journal*, Springer, v. 15, n. 1, p. 69–97, 2007.

POWER, E. M. Developing a culture of privacy: A case study. *IEEE Security & Privacy*, IEEE, v. 5, n. 6, p. 58–60, 2007.

RABINIA, A.; GHANAVATI, S. Fol-based approach for improving legal-grl modeling framework: A case for requirements engineering of legal regulations of social media. In: *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*. [S.l.: s.n.], 2017. p. 213–218.

REGGIO, G.; LEOTTA, M.; RICCA, F.; CLERISSI, D. Dusm: A method for requirements specification and refinement based on disciplined use cases and screen mockups. *Journal of Computer Science and Technology*, Springer, v. 33, n. 5, p. 918–939, 2018.

REGULATION, P. Regulation (eu) 2016/679 of the european parliament and of the council. *Regulation (eu)*, v. 679, p. 2016, 2016.

REIDENBERG, J. R.; BHATIA, J.; BREAUX, T. D.; NORTON, T. B. Ambiguity in privacy policies and the impact of regulation. *The Journal of Legal Studies*, University of Chicago Press Chicago, IL, v. 45, n. S2, p. S163–S190, 2016.

RIBEIRO, F.; FERREIRA, A. L.; TERESO, A.; PERROTTA, D. Development of a grooming process for an agile software team in the automotive domain. In: SPRINGER. *World Conference on Information Systems and Technologies*. [S.l.], 2018. p. 887–896.

ROBSON, C. *Real world research: A resource for social scientists and practitioner-researchers*. [S.l.]: Wiley-Blackwell, 2002.

RUNESON, P.; HÖST, M.; RAINER, A.; REGNELL, B. *Case study research in software engineering-guidelines and examples Wiley*. 2012.

RUNESON, P.; HöST, M. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, Kluwer Academic Publishers, Hingham, MA, USA, v. 14, p. 131–164, April 2009. ISSN 1382-3256. Disponível em: <http://portal.acm.org/citation.cfm?id=1519313.1519324>.

SCHAUB, F.; BALEBAKO, R.; DURITY, A. L.; CRANOR, L. F. A design space for effective privacy notices. In: *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*. USA: USENIX Association, 2015. (SOUPS '15), p. 1–17. ISBN 9781931971249.

SEAMAN, C. B. Qualitative methods. In: *Guide to advanced empirical software engineering*. [S.l.]: Springer, 2008. p. 35–62.

SENARATH, A.; ARACHCHILAGE, N. A. G. *Why developers cannot embed privacy into software systems? An empirical investigation*. 2018.

SHAH, U. S.; JINWALA, D. C. Resolving ambiguities in natural language software requirements: a comprehensive survey. *ACM SIGSOFT Software Engineering Notes*, ACM New York, NY, USA, v. 40, n. 5, p. 1–7, 2015.

SIEGEL, S.; CASTELLAN, N. *Estatística não-Paramétrica Para Ciências do Comportamento*. McGraw-Hill, 1975. ISBN 9788536313580. Disponível em: <https://books.google.com.br/books?id=LKmNnQEACAAJ>.

SIENA, A. *Engineering law-compliant requirements: the nomos framework*. Tese (Doutorado) — University of Trento, 2010.

SIENA, A.; INGOLFO, S.; PERINI, A.; SUSI, A.; MYLOPOULOS, J. Automated reasoning for regulatory compliance. In: SPRINGER. *International Conference on Conceptual Modeling*. [S.l.], 2013. p. 47–60.

SIRUR, S.; NURSE, J. R.; WEBB, H. Are we there yet? understanding the challenges faced in complying with the general data protection regulation (gdpr). In: *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*. [S.l.: s.n.], 2018. p. 88–95.

SJØBERG, D. I. K.; DYBÅ, T.; ANDA, B. C. D.; HANNAY, J. E. Building theories in software engineering. In: _____. *Guide to Advanced Empirical Software Engineering*. London: Springer London, 2008. p. 312–336. ISBN 978-1-84800-044-5. Disponível em: <https://doi.org/10.1007/978-1-84800-044-5_12>.

SNYDER, C. *Paper prototyping: The fast and easy way to design and refine user interfaces*. [S.l.]: Morgan Kaufmann, 2003.

SOARES, H. F.; ALVES, N. S.; MENDES, T. S.; MENDONÇA, M.; SPÍNOLA, R. O. Investigating the link between user stories and documentation debt on software projects. In: IEEE. *2015 12th International Conference on Information Technology-New Generations*. [S.l.], 2015. p. 385–390.

SOLIS, C.; WANG, X. A study of the characteristics of behaviour driven development. In: *2011 37th EUROMICRO Conference on Software Engineering and Advanced Applications*. [S.l.: s.n.], 2011. p. 383–387.

SOMMERVILLE, I.; SAWYER, P. *RE: a good practice guide*. [S.l.]: John Wiley and Sons, 1997.

SOUAG, A.; MAZO, R.; SALINESI, C.; COMYN-WATTIAU, I. Reusable knowledge in security requirements engineering: a systematic mapping study. *Requirements Engineering*, v. 21, n. 2, p. 251–283, 2015.

STRANDBERG, P. E. Ethical interviews in software engineering. In: IEEE. *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. [S.l.], 2019. p. 1–11.

SWIRE, P.; ANTON, A. Engineers and lawyers in privacy protection: Can we all just get along. *IAPP Privacy Perspectives*, 2014.

TAHERDOOST, H. *Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research*. [S.l.], 2016. Disponível em: <https://ideas.repec.org/p/hal/journl/hal-02546799.html>.

TANKARD, C. What the gdpr means for businesses. *Network Security*, Elsevier, v. 2016, n. 6, p. 5–8, 2016.

TIKKINEN-PIRI, C.; ROHUNEN, A.; MARKKULA, J. Eu general data protection regulation: Changes and implications for personal data collecting companies. *Computer Law Security Review*, v. 34, n. 1, p. 134–153, 2018. ISSN 0267-3649. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0267364917301966>.

TJONG, S. F. Avoiding ambiguity in requirements specifications. *no. February*, 2008.

URUGUAY, REPúBLICA ORIENTAL DEL. *Ley Nº 18.331 PROTECCIÓN DE DATOS PERSONALES Y ACCIÓN DE "HABEAS DATA"*. 2008. Disponível em: <https://legislativo.parlamento.gub.uy/temporales/leytemp7711203.htm>.

USMAN, M.; FELDERER, M.; UNTERKALMSTEINER, M.; KLOTINS, E.; MENDEZ, D.; ALEGROTH, E. Compliance requirements in large-scale software development: An industrial case study. In: SPRINGER. *International Conference on Product-Focused Software Process Improvement*. [S.l.], 2020. p. 385–401.

VILELA, J.; CASTRO, J.; MARTINS, L. E. G.; GORSCHEK, T. Integration between requirements engineering and safety analysis: A systematic literature review. *Journal of Systems and Software*, Elsevier, v. 125, p. 68–92, 2017.

VOLLSTEDT, M.; REZAT, S. An introduction to grounded theory with a special focus on axial coding and the coding paradigm. In: _____. [S.l.: s.n.], 2019. p. 81–100. ISBN 978-3-662-59350-9.

WAGNER, S.; FERNÁNDEZ, D. M.; FELDERER, M.; VETRÒ, A.; KALINOWSKI, M.; WIERINGA, R.; PFAHL, D.; CONTE, T.; CHRISTIANSSON, M.-T.; GREER, D. et al. Status quo in requirements engineering: A theory and a global family of surveys. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, ACM New York, NY, USA, v. 28, n. 2, p. 1–48, 2019.

WAGNER, S.; MENDEZ, D.; FELDERER, M.; GRAZIOTIN, D.; KALINOWSKI, M. Challenges in survey research. In: *Contemporary Empirical Methods in Software Engineering*. [S.l.]: Springer, 2020. p. 93–125.

WEBEL, C.; STEGLICH, R. Modeling legal requirements. In: WILEY ONLINE LIBRARY. *INCOSE International Symposium*. [S.l.], 2017. v. 27, n. 1, p. 418–433.

WIEGERS, K.; BEATTY, J. *Software requirements*. [S.l.]: Pearson Education, 2013.

WILSON, S.; SCHAUB, F.; DARA, A. A.; LIU, F.; CHERIVIRALA, S.; LEON, P. G.; ANDERSEN, M. S.; ZIMMECK, S.; SATHYENDRA, K. M.; RUSSELL, N. C.; NORTON, T. B.; HOVY, E.; REIDENBERG, J.; SADEH, N. The creation and analysis of a website privacy policy corpus. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, 2016. p. 1330–1340. Disponível em: <https://www.aclweb.org/anthology/P16-1126>.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. New York, NY, USA: Association for Computing Machinery, 2014. (EASE '14). ISBN 9781450324762. Disponível em: <https://doi.org/10.1145/2601248.2601268>.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLÉN, A. Experimentation in software engineering. 2012. *Google Scholar Google Scholar Digital Library Digital Library*, 2012.

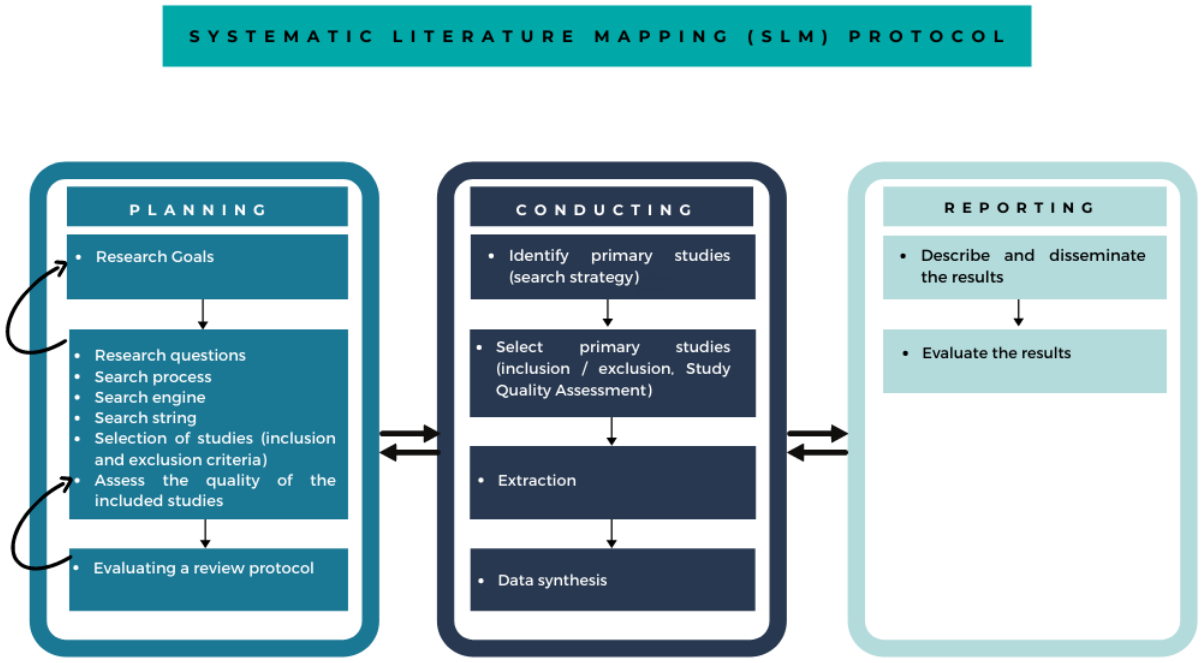
YAMAUCHI, E. A.; SOUZA, P. C. de; JUNIOR, D. P. S. Prominent issues for privacy establishment in privacy policies of mobile apps. In: *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2016. (IHC '16). ISBN 9781450352352. Disponível em: <https://doi.org/10.1145/3033701.3033727>.

ZENI, N.; KIYAVITSKAYA, N.; MICH, L.; CORDY, J. R.; MYLOPOULOS, J. Gaiust: supporting the extraction of rights and obligations for regulatory compliance. *Requirements engineering*, Springer, v. 20, n. 1, p. 1–22, 2015.

APPENDIX A – SYSTEMATIC LITERATURE MAPPING

This section shows an extensive bibliographic study on Privacy and Security in Requirements Engineering. Figure 46 shows the Systematic Literature Mapping (SLM) steps. SLM was performed in the first year of the Ph.D. to delimit the problem. Therefore, we cover research that addresses privacy and security in requirements engineering. Full SLM information has been published in (NETTO; PEIXOTO; SILVA, 2019).

Figure 46 – Systematic Literature Mapping protocol



Source: The author (2021).

Some research has already made efforts to understand privacy and security in RE. For example, Souag et al. (SOUAG et al., 2015) perform a systematic mapping study about Security Requirements Engineering (SRE) covering an interval from 2000 to 2013 identified 30 methods and categorized them in a set of five main types of knowledge forms of representation that were (re) used by SRE approaches: security patterns, taxonomies and ontologies, templates and profiles, catalogs and generic models and mixed. However, this systematic mapping takes into account only aspects of security.

Khan and Ikram (KHAN; IKRAM, 2016) carried out systematic mapping of the literature in the field of SRE from 2010 to 2015. They present 15 problem clusters: domain security (74 papers, 29%), methodologies (17 papers, 7%), integration of security, lack of evaluation,

architecture, documents, legal requirements (e.g., threats, human/environment not considered, automatic support, change, ontologies) and further divided into subcategories that comprise more specific related problems.

Abu-Nimeh and Mead (ABU-NIMEH; MEAD, 2009) affirm that despite the overlap between Privacy Requirements Engineering (PRE) and SRE, each addresses a different set of problems. As a result, security risk assessment techniques used in SRE may be unsuitable for assessing privacy risks. Moreover, it is not yet evident how to achieve this systematically through the various stages of the RE process (LAMSWEERDE,).

Motivated by this scenario, this section intends to present a Systematic Literature Mapping (SLM), which aimed at understanding the state of the research on the privacy and security of Requirements Engineering. The SLM was chosen because it is the most appropriate method to provide a broad overview of a research area (KITCHENHAM; CHARTERS, 2007). The SLM catch papers from the year 2000 to 2016.

The research question that guides this study is **RQ: What is the current state of privacy and security research in Requirements Engineering?**

The following specific research questions (RQ) were used to guide the synthesis of results:

RQ1: What research topics are investigated about privacy and security in requirements engineering?

RQ2: What research methods are used for privacy and security in requirements engineering?

RQ3: What types of study about privacy and security are in requirements engineering?

RQ4: What is the research problem about privacy and security in requirements engineering?

RQ5: What trends or future work are presented by primary studies on privacy and security in requirements engineering?

A.0.1 Research protocol

The SLM followed the procedures indicated by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007). Two Ph.D. students conducted the SLM, and an experienced graduate professor and researcher with expertise in Requirements Engineering validated the procedures.

The articles were individually analyzed by the two doctoral students and later discussed in weekly meetings for acceptance or exclusion. For the answers to the research questions, the researchers elaborated on them.

A.0.2 Search process

The rigor of the search process is a factor that distinguishes systematic literature review or mapping from other types of reviews (KITCHENHAM; CHARTERS, 2007). The goal of an SLM is to find as many primary studies addressing the issue of possible research using an unbiased search strategy. The identification of the related research occurred in five automatic search engines: Ei COMPENDEX [1], IEEExplorer [2], ACM Digital library [3], Scopus [4], Science Direct [5]. We choose these search engines because they are relevant sources for the Software Engineering area.

We developed a search string, with relevant synonyms, for the identification of the related research through automatic search:

> *(privacy OR security) AND (“requirements engineering" OR “requirements approach" OR “requirements methodology" OR “requirements process").*

We have thoroughly tested various terms and synonyms to get the search string used. It is essential to clarify that we apply the string to titles and abstracts for some search engines because when we perform differently, we find many irrelevant works. Therefore, we adapted the search string according to the specific criteria of each search engine, as can be seen below.

> IEEE: *(((privacy) OR (security)) AND ((“requirements engineering") OR (“requirements approach") OR (“requirements methodology") OR (“requirements process"))). Obs: search for metadata.*
>
> ACM: *recordAbstract:(“privacy" OR “security") AND (“requirements engineering" OR “requirements approach" OR “requirements methodology" OR "requirements process").*
>
> SCOPUS: *TITLE-ABS-KEY(privacy) OR TITLE-ABS-KEY(security) AND TITLE-ABS-KEY(“requirements engineering") OR TITLE-ABS-KEY(“requirements approach") OR*

[1] www.engineeringvillage2.org/
[2] ieeexplore.ieee.org/
[3] dl.acm.org
[4] www.scopus.com/
[5] www.sciencedirect.com/

TITLE-ABS-KEY(“requirements methodology") OR TITLE-ABS-KEY(“requirements process").

ScienceDirect: *TITLE-ABSTR-KEY((“privacy" or “security")) and TITLE-ABSTR-KEY(“requiremer engineering" or “requirements approach" or “requirements methodology" or “requirements process").*

Ei COMPENDEX: *(((((privacy)WN KY) OR ((security)WN KY)) AND ((((“requirements engineering")WN KY) OR ((“requirements approach")WN KY) OR ((“requirements methodology")WN KY) OR ((“requirements process")WN KY))))).*

We found papers were from the year 2000 to 2016. It is important to note that this research does not show the full effect of all the papers published in 2016 because the search and selection occurred between July and September of 2016.

A.0.3 Selection of studies

Once we get only potentially relevant studies, they need to be evaluated, for which it is necessary to indicate some inclusion and exclusion criteria. These criteria are intended to identify primary studies that provide direct evidence on the research question (KITCHENHAM; CHARTERS, 2007).

We defined the inclusion and exclusion criteria, based on the RQ, to achieve consistent results:

1. Inclusion Criteria: I1 Peer-reviewed studies; I2 Accessible studies; I3 Original studies in the languages: English, Portuguese and Spanish.

2. Exclusion Criteria: E1 Duplicated studies (only one copy included); E2 Gray literature (Short papers, presentations, reports, dissertations, theses, secondary and tertiary studies); E3 Studies that do not focus on privacy or security; E4 Studies that do not focus on RE; E5 Publications whose text was not available (through search engines or by contacting the authors).

First, the studies have been checked using the exclusion criteria. If a paper could meet any of the exclusion criteria, in turn, if E1 OR E2 OR E3 OR E4 OR E5 is true, then the paper must be removed. Another case for a duplicate E1 is when a journal article follows

a conference paper. In such cases, we select the higher-valued publication, i.e., journal over conference [13]. Subsequently, the inclusion criteria were observed. Thus, it was verified if I1 AND I2 AND I3 could meet. If so, papers must be selected, and if any criteria are not met, the article is excluded.

The selection process occurred in three different steps:

1. Step1: reading titles, keywords, and abstract; considering the inclusion and exclusion criteria.

2. Step 2: reading introduction and conclusion; considering the inclusion and exclusion criteria.

3. Step 3: the studies included are thoroughly read, excluding irrelevant papers for the research questions.

A.0.4 Data extraction

Data extraction should be designed to collect all the information needed to address the mapping issues (KITCHENHAM; CHARTERS, 2007). We performed the data extraction with a spreadsheet. We contained the following fields: Identifier, source, year, affiliations, list of authors, title, keywords, main subject (security or privacy), answers to research questions, subjective extraction.

A.0.5 Threats to Validity

The mapping protocol follows a few steps to ensure that the search is as accurate and objective as possible. However, potential limitations may arise. We used the categorization of threats presented by Wohlin et al. (WOHLIN et al., 2012).

Construct validity is related to the generalization of the result to the concept or theory behind the study execution (WOHLIN et al., 2012). The search string used may not include all existing synonyms for the term “Privacy and Security in Requirements Engineering" and may be insufficient to capture all studies in the area. To minimize threats of this nature, we used synonyms for the key constructs, and we realized a manual search to find papers that were not found in automatic search.

Internal validity is related to a possible wrong conclusion about causal relationships between treatment and outcome (WOHLIN et al., 2012). To mitigate personal bias in the study, two Ph.D. students conducted the SLM, and an experienced graduate professor and researcher with expertise in Requirements Engineering validated the procedures. We also selected only peer-reviewed papers.

External validity is concerned with the degree to which the primary studies are representative for the review topic (WOHLIN et al., 2012). In the case of a literature mapping, if the identified literature is not externally valid, neither is the synthesis of its content (VILELA et al., 2017). We excluded gray literature papers.

Conclusion validity (WOHLIN et al., 2012) the research protocol was carefully designed and validated by the authors to minimize the risk of exclusion of relevant studies. Besides, we used many synonyms for the constructs of this paper to improve the high coverage of possibly relevant studies from automatic search.

A.0.6 Data analysis

We conducted an extensive bibliographic study on Privacy and Security in Requirements Engineering.

Initially, through the automatic search, as shown in Table 1, we found 2658 papers. Excluding duplicate articles (1446), we get 1212 unique papers. Afterward, read the title and the abstract. We excluded 630 studies, based on the exclusion criteria being: Gray literature (134 papers); Does not focus on privacy or security (245 articles); Does not present focus on RE (241 papers); Could not be accessed (08 papers), Non-English, Spanish or Portuguese written papers (02 paper). In step 1, we selected 582 papers to be analyzed in the next step. Of the 582 papers from the previous stage, 284 were excluded, resulting in 298 being selected to participate in Step 3 (see Table 1). Of the excluded studies, it is possible to observe the following data: It was not peer-reviewed (2); Duplicates (19); Gray literature (28); Does not focus on privacy or security (78); Does not present focus on RE (97); Could not be accessed (60).

For the third stage (see Table 40), the studies resulting from the previous step were read, and those that presented answers to some of the research questions were selected. At the end of the process, we choose 267 papers after excluding 31 studies. The complete list of the

selected studies list [6] is available online. Of the excluded studies, we observe the following data: Does not focus on privacy or security (21); Does not the present focus on RE (10).

Table 40 – Paper selection engines research

Search engines	Articles	Step 1	Step 2	Step 3
ACM Digital library	76	36	30	26
Ei COMPENDEX	1002	12	06	04
IEEExplorer	425	183	114	95
Science Direct	33	22	14	14
Scopus	1122	329	134	128
Total	**2658**	**582**	**298**	**267**

Source: The author (2021).

We found papers were from the year 2000 to 2016. The pivotal year of publication was 2014 with a total of 36 papers (13.5%), followed by 2012 with 28 (10.5%), 2008 with 27 (10.1%) (see Figure 47). It is important to note that this research does not show the full effect of all the papers published in 2016 because the search and selection occurred between July and September of 2016.

Figure 47 – Papers published by year

Source: The author (2021).

[6] https://doi.org/10.6084/m9.figshare.7789637

We categorize the papers according to the central theme, namely: security, privacy, or both (see Table 41). Security was the theme that presented the highest number, with 202 (75.7%) of papers.

Table 41 – Central theme

Main theme	Frequency	Percentage
Privacy	44	16.5%
Security	202	75.7%
Privacy & Security	21	7.8%
Total	**267**	**100%**

Source: The author (2021).

A.0.7 First research question

The first question, RQ1, ask what the research topics investigated about privacy and security in RE are. Based on the similarity of the problem and how many studies reported the problem, we grouped the number of studies greater than four as a cluster. The studies which were stand alone in terms of the problem that they were reporting or problem reported by less than four studies were put into the “other" category. Table 1 shows the list of research topics and “other" categories listed individually or together with other topics. Table 1 presents eight research topics identified from the classification that the authors show in their studies. The most prominent research topic was “Requirements Elicitation" in 84 (31.46%) papers. “Requirements Modeling" was the second most frequent topic in 39 (14.61%) of the papers. It was followed by “Requirements Analysis" in 32 (11.99%) papers. We divide the papers into periods of years:

2000 - 2005: Thirty-two papers (privacy - 5, security - 25, both - 2). The main research topics were “Requirements Modeling" (10), “Requirements Elicitation" (5), and “Requirements Engineering Process" (5). In 2000 just one paper about Security Requirements Elicitation. The year with the highest number of publications was 2005, with 16 papers representing 50% of publications. 2006 - 2010: Ninety-nine papers (privacy - 22, security - 77). The main research topics were “Requirements Modeling" (22), “Requirements Elicitation" (15), and “Requirements Analysis" (14). In this period, the year with the highest number of publications was 2008, with 27 papers representing 27% of published papers.

Table 42 – Research Topics

Research topics	**Frequency**	**Percentage**
Requirements elicitation	128	47.94%
Requirements modeling	41	15.35%
Requirements analysis	30	11.24%
Requirements specification	22	8.24%
Requirements engineering process	19	7.12%
Requirements standards	08	3.00%
Requirements design	06	2.24%
Other	13	4.87%
Total	**267**	**100.00%**

Source: The author (2021).

2011 - 2016: 136 papers are in this range (privacy -18, security - 104, and both - 14). The main research topics were “Requirements Elicitation" (61), “Requirements Identification" (18), and “Requirements Analysis" (13). In this period, the year with the highest number of publications was 2014, with 36 papers representing 26.47% of the published papers. In the first six months of 2016, we identified five published papers whose research topics are Requirements Elicitation and Specification.

A.0.8 Second research question

The second question, RQ2, asks what research methods used for privacy and security in RE are. Table 43 shows the list of research methods listed individually or in combination with other methods.

The most prominent research method was “Applying the method to an example or simulation with 115 papers, 17 in the privacy field, 88 in security, and 10 in papers that addressed privacy and security together. “Case Study/Focus Group" also presented good results with a total of 70 papers, 15 in the privacy area, 49 in security, and 6 in papers addressing both areas. Some methods have only one occurrence. They are: “Experts evaluation" (security), “Interview" (security), “Literature study, structured analysis" and, “brainstorming" (security).

Table 43 – Research Methods

Research method	**Privacy**	**Security**	**Privacy & Security**	**Total**
Study/comparative analysis of models or approaches	1	2	0	3
Applying the method to an example or simulation	17	88	10	115
Experts evaluation	0	1	0	1
Interview	0	1	0	1
Case Study / Focus Group	15	49	6	70
Literature study, structured analysis	0	1	0	1
Usability study or user study	0	2	0	2
Observational study	0	0	02	2
Experiment	1	7	0	8
Does not present a formal method or did not make clear the used method	9	47	2	58
Survey	1	4	1	6
Total	**44**	**202**	**21**	**267**

Source: The author (2021).

A.0.9 Third research question

The third question RQ3 asks what types of study about privacy and security in RE are. Table 44 shown the types of studies. The variable type of study was based on Petersen et al. [14]:

> *Evaluation Research: Techniques implemented (applied) in practice, and an evaluation of the method conducted (solution implementation).*
>
> *Opinion Papers: These papers express the opinion of somebody whether a specific technique is right or wrong or how things should have been done.*
>
> *Philosophical Papers: These papers sketch a new way of looking at existing things by structuring the field in the form of taxonomy or conceptual framework.*
>
> *Solution Proposal: A solution to a problem can be either novel or a significant extension of an existing technique. A small example of a good line of argumentation shows the potential benefits and the applicability of the solution (but no empirical data).*

Validation Research: Techniques investigated are novel and have not yet been implemented in practice. Techniques used are, for example, experiments.

Solution Proposal with 204 papers was the type of study that presented the highest number of results, followed by Evaluation Research with 28 articles and Validation Research with 14 papers. The Solution Proposal has been the type of study with the highest number of results, demonstrating a lack of studies that carry out validation with formal methods, such as controlled experiments.

Table 44 – Type of study

Type of study	**Privacy**	**Security**	**Privacy & Security**	**Total**
Evaluation Research	6	20	2	28
Experience Papers	2	7	0	9
Opinion Papers	4	3	0	7
Philosophical Papers	2	3	0	5
Solution Proposal	29	158	17	204
Validation Research	1	11	2	14
Total	**44**	**202**	**21**	**267**

Source: The author (2021).

A.0.10 Forth research question

The fourth question, RQ4, asks the research problem about privacy and security in RE. We grouped the papers according to the research topic (Table 1) covered in RQ1, and we performed a characterization of the research problems of this research topic. Selected Studies List[7] is available online.

Requirements Elicitation is the most cited research topic. These papers aim to derive privacy and security requirements and guidelines for specific contexts, such as mobile technologies, goal-oriented approaches, and legal requirements, contributing to security and privacy users' data protection. In this category, the study proposes a methodology to determine the software requirements by analyzing the natural language of privacy policies [SC0178]1. Define a Goal-Oriented approach to elicitation and formal description of security requirements and

[7] https://doi.org/10.6084/m9.figshare.7789637

incorporate fault tolerance into system requirements models through the partial satisfaction of security objectives [SC0327]. Define a method for eliciting security objectives and then suggest composing these goals into consistent security requirements [SC0205].

Requirements Elicitation and Legal Requirements is a broad field of research. The Secure Tropos framework allows obtaining high-level security requirements and automatically verifying system requirements specified in the formal modeling language [SD021]. A paper presents a framework called "Water Marking Requirements" that business analysts can use to align the requirements of various jurisdictions [IEEE198]. Other papers define a methodology for directly extracting access rights and obligations from regulatory texts [SCOPUS144]. A paper aims to define an approach that identifies software requirements by analyzing privileged documents, appointments, and online rights [IEEE007].

Requirements Modeling is the second most identified research topic in the papers captured in this mapping. In this category, the existing approaches to specifying and enforcing access control policies do not provide methodological support while determining these policies. Therefore, [EI001] defines a modeling language to specify and analyze access control policies about the organization and security and permission requirements of system administrators. Derive semantic goals models extracted from privacy policy documents [IEEE030]1. Presenting a methodology that incorporates basic privacy requirements into the design process also describes a systematic way of analyzing the impact of privacy objectives on the organizational process and the systems that support the process [IEEE067]1. Present an approach that assists navigation, indexing, and modeling security goals formulated in Natural Language (NL) and provides a valuable tool for critically assessing and refining NL text [ACM033].

Requirements analysis is one of the most identified research topics in the papers captured in this mapping. In this category, the research problems addressed are identified the assets, threats, and vulnerabilities of a system, helping developers to analyze and extract the requirements at the early stage of development. Applying RE's principles and best practices over privacy policy analysis to analyze the relationship between the various participants, possible attacks, threats, and vulnerabilities; and use the techniques of misuse cases, tree attack, and risk assessment to obtain the elements [IEEE283]. Use privacy arguments as a means of generally reviewing privacy requirements to allow the system to adapt at runtime to privacy requirements [IEEE194].

Papers in the research topic **Requirement Specification** identify the issues, types, and methods of security requirements, such as [SC0615], who use a framework to derive a set of

requirements specifications. [SCOPUS020] Automatically generates a security policy from a more structured specification of the system objectives.

Requirements Engineering has an activity called **Requirements Management** that seeks to control evolution and changes, as well as enable the tracking of requirements throughout the development process. One of the papers that treats this topic aims to define a metamodel for tracking compliance between different models of User Requirements Notation (URN) models of the HIC (Health Information Custodians) and privacy legislation [SC0712]. Another paper presents a tool (SecMER) that can automatically detect changes in requirements and violations of security properties [SCOPUS085].

At the level of **Requirements Design**, the papers have as research problems to use approaches of the RE to define and evaluate models of access control about the security requirements of the organization and to analyze the impact of the privacy requirements of the organizational objectives [ACM010], [EI001], [IEEE095].

One of the papers whose theme is **Requirements Reuse** aims to develop a repository with all sources of relevant security requirements for the organization to avoid unnecessary efforts to identify, understand and relate security aspects to requirements sources [IEEE134].

Using **Requirements Standards** can significantly reduce the time spent in the requirements elicitation phase. Some papers use requirements standards to support the Security Requirements Specification process [IEEE152]. Legal Requirements also appear related to requirements standards. The paper presents an organizational-level security standard to assist legal and security experts in capturing, modeling, and setting security standards [SC0355].

Papers whose research topic is **RE Process** aims to integrate existing tools and techniques (i* (i-star), NFR framework, misuse case, abuse case) with risk analysis to improve the process of RE for Privacy and Security [SC0715]. Extend RUP as security requirements in elicitation, analysis, and specification activities [ACM047].

Papers dealing with Privacy and Security **Requirements Evolution** aim to investigate the challenges of analyzing the impact of evolutionary changes on system security. The international standard for secure application development, Common Criteria (CC), is regularly cited in the papers either as a certification parameter or as a guide that can be used to verify security requirements. A paper aims to integrate CC into the RE Process for requirements security through the definition of a tool that allows applying the SREPPLine approach systematically and intuitively, as well as compliance with standards (CC, ISO / IEC 270001 and IEEE 830: 1998) without the need to know these standards, reducing the participation of

specialists [SC0164].

The misuse cases technique is used in security to determine the actions that any actor can perform to harm the system. Papers establish a framework (MOSRE) consisting of use cases and misuse cases to identify security requirements [SC0618]. Misuse cases are used to define a framework to detect threats such as risk assessments that arise from misuse of stakeholder permissions over resources [SC0192]. In this paper, the authors propose translating Tropos models into Misuse Cases diagrams to integrate security analysis from the earliest stages into all stages of the process development [ACM057].

Ontologies are used in Software Engineering to represent concepts within a domain and its relationships. Ontologies can be used as a source to specify knowledge of security requirements efficiently [SC0475]. Another paper proposes to include it in the elicitation, analysis, and validation process to engineering security requirements [SC0714]. Some papers relate Common Criteria to ontologies to defining a Goal-Oriented ontology model for CC requirements [SC0271]1. This paper proposes to use Object Constraint Language (OCL) to formalize security requirements in a model-driven approach to critical applications [IEEE241].

A.0.11 Fifth research question

The fifth question, RQ5, asks what trends or future work about Privacy and Security in RE. Selected Studies List[8] is available online. Many papers also present the need for tool development that supports the proposed methodology. When they already have tools, the papers present the need for usability improvements for a better user experience [SCOPUS085] [IEEE213]. Some papers also claim that it is possible to use an approach to other types of requirements that are not about privacy or security [ACM011] or indicate the need to extend the work to support other security standards [ACM012] or to apply the different method types of security [IEEE326]. Other papers will discuss how security research can be extended or adapted to support privacy [IEEE131].

One paper addresses Requirements Reuse, developing a framework that addresses the adequacy of reusable requirements presented in requirements catalogs [IEEE144].

Requirements Modeling addresses extending modeling activities to the design, coding, and testing phase [SC0637]. A paper provides to investigate modeling techniques to conduct a RE process in a planned way in a framework that uses the models created to identify functional

[8] https://doi.org/10.6084/m9.figshare.7789637

requirements [ACM013].

As a future works, Requirements Specification says that it is necessary to evaluate the efficient use of Security Requirements in the final stages of the software development life cycle. Integrate approaches to measures to ensure compliance with security quality policies and the profile of the attacker [IEEE023].

The papers of Requirements Elicitation note that a future direction is to analyze the priority of requirements and build a system of compatible privacy legislation and develop standards of security requirements. Address need to define a complete list of security requirements to address vulnerabilities [IEEE251]. Identify more complex threat patterns that lead to the violation of security properties [SC0192]. Develop a method to identify criteria for assessing functional requirements derived from non-functional legislation [SCOPUS143]. Develop a support tool to extract legal requirements related to privacy and security from the specification expressed in natural language [SC0355].

APPENDIX B – SNOWBALLING

Subsequently, a more specific bibliographic study on ambiguity in legal requirements was carried out using the snowballing technique (Wohlin, 2014). To identify works from an initial set that had in its title or abstract the keywords ambiguity, "legal requirements," "regulatory requirements," "requirements engineering," and identify problems that related to legal requirements, ambiguity, and legal compliance. The initial set is composed by (Reidenberg et al., 2016; Massey et al., 2014; Breaux and Gordon, 2013; Maxwell et al., 2011; Breaux et al., 2008).

This section describes the research questions, procedures, and methods for carrying out this study's snowballing planning stage. It investigates how the ambiguity present in the specification of legal requirements and compliance with the data protection law has been dealt with in the literature. Snowballing aims to assist in the definition of related works and answering the research question "*How is the ambiguity present in the legal text dealt with in the legal requirements specifications of systems compliant with the legislation?*"

B.0.1 Research Goals

The specific research questions that guided this snowballing were:

1. RQ1. What are the strategies used in Requirements Engineering to deal with the ambiguity present in legal requirements in developing systems in legal compliance?

   This question aims to obtain an overview of the techniques used to identify or reduce ambiguity in legal requirements.

2. RQ2. What are the approaches used in specifying legal requirements with reduced ambiguity in developing systems in legal compliance?

   This question aims to obtain an overview of the approaches used to specify legal requirements with reduced ambiguity.

3. RQ3. What are the challenges and limitations related to ambiguity in specifying legal requirements when developing systems in legal compliance?

   This question aims to identify the benefits and limitations reported in the papers about ambiguity in the legal requirements specification.

B.0.2 Study design and planning

We used the guidelines proposed by Wohlin (WOHLIN, 2014) to perform snowballing. Snowballing refers to using an article's reference list or article citations to identify additional articles. The procedures for snowballing are presented below.

The first challenge is to identify an initial set of articles to use in the snowballing procedure. Wholin (WOHLIN, 2014) mentions that identifying several relevant and highly cited articles can be an alternative to define the initial set if many articles are found.

The inclusion and exclusion criteria for the initial set and the articles identified through backward and forward snowballing are:

1. Inclusion criteria are I1 - Primary studies; I2 - Studies that aim to reduce or eliminate ambiguity in Legal Requirements Engineering; I3 - Studies relating ambiguity and specification of legal requirements; I4 - Studies relating ambiguity and elicitation of legal requirements.

2. Exclusion criteria are E1 - Secondary studies; E2 - Short papers (less than three pages); E3 - Duplicate studies (only one copy of each study was included); E4 - Articles not written in English; E5 - Not published in a peer-reviewed event; E6 - Gray literature (technical reports, project reports, others); E7 - Paper not available.

B.0.3 Data collection

To calibrate the search string and avoid research bias in database selection, we used Google Scholar, as suggested by Wholin (WOHLIN, 2014). The search string has the following keywords: *ambiguity*, "*legal requirements*," "*regulatory requirements*," "*requirements engineering*," we configure the display of articles captured in the search as Sort by relevance, and we do not define a period. The search and analysis was carried out in the second half of 2017, between June and October.

We obtained 110 candidate articles for the initial set that were evaluated using the inclusion and exclusion criteria presented in the previous section. We define an identifier for each article in the format C001, indicating that these are candidates for inclusion.

Analyzing the articles and classifying them according to the inclusion and exclusion criteria, we selected five articles that make up the initial snowballing set. All articles have at least one

author in common. Table 45 shows the list of articles that make up the initial set.

Table 45 – Papers of initial set

Code	Title	Authors	Year	References
C001	Legal requirements, compliance and practice: an industry case study in accessibility	Travis D. Breaux, Annie I. Antón, Kent Boucher, Merlin Dorfman	2008	26
C002	A legal cross-references taxonomy for identifying conflicting software requirements	JC Maxwell, AI Antón, P Swire	2011	37
C003	Regulatory requirements traceability and analysis using semi-formal specifications	Travis D. Breaux, David G. Gordo	2013	37
C004	Identifying and classifying ambiguity for regulatory requirements	A K Massey, R L Rutledge, A I Antón, Pater P. Swire	2014	18
C023	Ambiguity in privacy policies and the impact of regulation	Joel R. Reidenberg, Jaspreet Bhatia, Travis D. Breaux, and Thomas B. Norton	2016	22

Source: The author (2021).

The following sections present the three iterations that were performed in this snowballing.

B.0.4 Iteration 1

Backward Snowballing

In the first iteration of backward snowballing, the references of articles included in the initial set are analyzed to identify more articles to include in the study. Each article is evaluated one at a time. To facilitate identifying the articles added in the backward snowballing step, we named each one using [BW - 1 – C01]. The first part represents backward, the second the iteration, and the candidate's last identifier for inclusion. To illustrate the performance of Iteration 1 of Backward Snowballing, we used the article by Breaux et al. (2008).

This article identified as C001 has 26 references. Analyzing them, we identified that 21 should be excluded based on the title, as they do not fit the research question, and five articles

are candidates for inclusion. We evaluate the full text of the five candidate articles to avoid using an article in snowballing that could later be excluded. The inclusion is based on the full article. All five articles were assessed as relevant and included in the study. The five articles added to the list are:

[BW - 1 - C01] T.D. Breaux, M.W. Vail, A.I. Antón. "Towards compliance: extracting rights and obligations to align requirements with regulations," IEEE Int'l Conf. Engr. Req., pp. 49-58, 2006.

[BW - 1 - C02] T.D. Breaux, A.I. Antón. "A systematic method for acquiring regulatory requirements: a frame-based approach," 6th Int'l Workshop on Requirements for High Assurance Systems, Delhi, India, 2007.

[BW - 1 - C03] T.D. Breaux, A.I. Antón. "Analyzing regulatory rules for privacy and security requirements," IEEE Trans. Soft. Engr., Special Issue on Soft. Eng. for Secure Sys., 34(1): 5-20, 2008.

[BW - 1 – C04] S. Ghanavati, D. Amyot, L. Peyton, "Towards a framework for legal compliance tracking in healthcare," 19th Int'l Conf. Adv. Sys. Engr., pp. 218-232, 2007.

[BW - 1 – C05] F. Massacci, M. Prest and N. Zannone. "Using a security requirements engineering methodology in practice: the compliance with the Italian data protection legislation," Computer Standards & Interfaces, 27(5):445 455, 2005.

Forward Snowballing

In forward snowballing, we will evaluate the articles that cite the articles that make up the initial set. Again, we use Google Scholar to identify citations. Forty-three works cite article C001. Performing the analysis of these articles, we identified that ten articles had been included, and 33 articles were excluded through the evaluation of the title as they do not fit the research question of this study. Therefore, we will not include articles in the first iteration of forwarding snowballing for article C001. Five works cite the article identified as C023. Performing the analysis of these articles, we identified that four were excluded by evaluating the title. One article [FW - 1 – C02] was considered relevant and included in the study. [FW - 1 – C02] Bhatia, Jaspreet et al. A theory of vagueness and privacy risk perception. In: Requirements Engineering Conference (RE), 2016 IEEE 24th International. IEEE, 2016. p. 26-35.

Iteration 1 Summary

The Google Scholar search returned 110 results, which were analyzed, and we include five in the study. Of these five articles, we evaluate 156 candidates (26 from C001, 37 from C002, 37 from C003, 34 from C004, 22 from C023) in the backward snowballing step. Six new articles were included [BW – 1 – C01] to [BW – 1 – C06]. In forward snowballing, we analyze 123 candidates, and nine new articles were included [FW – 1- C01] and [FW – 1- C02]. We evaluate a total of 233 articles and included 14 articles in the study in iteration 1.

B.0.5 Iteration 2

The 14 new articles identified in the first iteration will be analyzed, first concerning backward snowballing and then forward snowballing.

Backward Snowballing

We will analyze the references of 14 articles. The article identified by [BW – 1 – C01] includes 28 references; 26 were excluded based on the title. Two references were analyzed in more detail in the text, reading the full article, but it was decided not to include them. [BW – 1 – C02] includes 19 articles. Of these, 12 were excluded through title analysis, two articles belong to the study, and five had been excluded in previous stages. [BW – 1 – C06] includes 34 references; 30 were excluded based on title, three references were previously included, and one reference was included [BW - 2 – C01].

Forward Snowballing

We will analyze the articles that cite these 14 articles. [BW – 1 – C06] was cited because it includes 74 references, 41 were excluded in previous steps, 25 were excluded based on title, seven references had been previously included, and one reference was included [FW - 2 – C02]. The article [FW - 1 - C04] published in October 2017 was cited only once due to snowballing in the second half of 2017. Therefore, no added article to the study as of [FW – 1 – C04].

Iteration 2 Summary

In iteration 2, 485 candidates were evaluated (28 of [BW - 1 - C01], 19 of [BW - 1 - C02], 44 of [BW - 1 - C03], 22 of [BW - 1 - C04], 14 from [BW - 1 - C05], 34 from [BW - 1 - C06]) in backward snowballing. Two new articles were added [BW – 2 – C01] to [BW – 2 – C02]. In forward snowballing, 933 candidates were analyzed (27 of [FW - 1 - C01], 58 of [FW - 1 - C02], 27 of [FW - 1 - C03], 19 of [FW - 1 - C04], 82 of [FW - 1 - C05], 32 of [FW - 1 - C06], 20 of [FW - 1 - C07], 35 of [FW - 1 - C08], 24 of [FW - 1 - C09]), and two new articles were included [FW – 2- C01] and [FW – 2- C02]. A total of 1418 articles were evaluated and four

were included in the study.

B.0.6 Iteration 3

As in iteration 3, only two articles were identified in backward snowballing and two articles in forward snowballing, and the analysis was simpler to perform in the backward stage. Nevertheless, in the forward stage, due to the two articles being a little older (2007 and 2001), there was a more significant amount of papers citing them.

Backward Snowballing

[BW – 2 – C01] includes 47 references; 23 were previously excluded, eight were added to the study previously, three were gray literature, and 13 were excluded based on the title. No new articles were included in the study. [BW – 1 – C02] includes 28 references; 17 were excluded based on title, three references were previously included, one reference was gray literature, and seven articles were previously excluded.

[FW – 2 – C01] includes 49 references; 23 were previously excluded, eight were added to the study previously, three were gray literature, and 13 were excluded based on the title. No new articles were included in the study. [FW – 1 – C02] includes 35 references; 17 were excluded based on title, nine references were previously included, nine articles were previously excluded.

Forward Snowballing

In the forward snowballing stage, most of the articles mentioning these four works had been previously evaluated. Therefore, no articles were included in this step.

B.0.7 Results

The results of this study indicated that there is a more significant number of studies dealing with the detection of ambiguities (Rabinia and Ghanavati, 2017; Massey et al., 2017; Bhatia et al., 2016; Ferrari et al., 2016; Massey et al., 2015; Massey et al., 2014; Kamsties, 2005; Berry and Kamsties, 2004; Berry and Kamsties, 2001) rather than avoiding or reducing them (Umber and Bajwa, 2011; Polpinij, 2009; Popescu et al. ., 2007; Boyd et al., 2005). We have identified some systematic literature reviews dealing with work on requirements engineering for legal compliance (Otto and Antón, 2007; Cliver and Winter, 2009; Shamsaei et al., 2011; Ghanavati et al., 2011; Akhigbe et al., 2018).

A list of the 24 articles included in this study can be found in Appendix A – Snowballing Articles. We present in Table 4 the articles and classify them if they aim to identify, reduce or avoid ambiguity. We capture articles that deal with other aspects of legal requirements during the snowballing process, such as cross-reference, requirements extraction, and legal requirements tracking. Therefore, some articles are unclassified in terms of ambiguity.

Table 1 – Classification of snowballing articles

Code	Title	Classification
C001	Legal requirements, compliance and practice: an industry case study in accessibility	Reduce\Avoid
C002	legal cross-references taxonomy for identifying conflicting software requirements	Identify
C003	Regulatory requirements traceability and analysis using semi-formal specifications	Identify
C004	Identifying and classifying ambiguity for regulatory requirements	Identify
C023	Ambiguity in privacy policies and the impact of regulation	Reduce\Avoid
[BW - 1 C01]	Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations	
[BW - 1 C02]	Systematic Method for Acquiring Regulatory Requirements: A Frame-Based Approach;	
[BW - 1 C03]	Analyzing Regulatory Rules for Privacy and Security Requirements	
[BW - 1 C04]	Towards a framework for tracking legal compliance in healthcare	
[BW - 1 C05]	Using a security requirements engineering methodology in practice: the compliance with the Italian data protection legislation	Reduce\Avoid
[BW - 1 C06]	Evaluating existing security and privacy requirements for legal compliance	Reduce\Avoid
[FW-1-C01]	Toward multilevel textual requirements traceability using model driven engineering and information retrieval	
[FW-1-C02]	A Theory of Vagueness and Privacy Risk Perception	Identify\Reduce
[FW-1-C03]	Managing Ambiguity and Traceability in Regulatory Requirements: A Tool-supported Frame-based-Approach	Reduce
[FW-1-C04]	Modeling Regulatory Ambiguities for Requirements Analysis	Reduce
[FW-1-C05]	Ambiguity and tacit knowledge in requirements elicitation interviews.	Reduce
[FW-1-C06]	A Strategy for Addressing Ambiguity in Regulatory Requirements	Identify\Reduce

Source: The author (2021).

Table 2 – Classification of snowballing articles (continued)

Code	Title	Classification
[FW-1-C07]	Ambiguity in Privacy Policies and the Impact of Regulation	
[FW-1-C08]	Based Approach for Improving Legal-GRL Modeling Framework: A Case for Requirements Engineering of Legal Regulations of Social	
[FW-1-C09]	Analyzing privacy requirements: a case study of healthcare in Saudi Arabia	
[BW-2-C01]	Addressing Legal Requirements in Requirements Engineering	
[BW-2-C02]	Detecting Ambiguities in Requirements Documents Using Inspections	Identify
[FW - 2 -C01]	Managing Legal Texts in Requirements Engineering;	Identify
[FW - 2 -C02]	A critical analysis of legal requirements engineering from the perspective of legal practice;	

Source: The author (2021).

APPENDIX C – SNOWBALLING PAPERS

61

Apêndice A – Artigos do Snowballing

Conjunto Inicial

Código	Título	Autores	Ano	Cita
C001	Legal requirements, compliance and practice: an industry case study in accessibility	Travis D. Breaux, Annie I. Antón, Kent Boucher, Merlin Dorfman	2008	26
C002	A legal cross-references taxonomy for identifying conflicting software requirements	JC Maxwell, AI Antón, P Swire	2011	37
C003	Regulatory requirements traceability and analysis using semi-formal specifications	Travis D. Breaux, David G. Gordon	2013	37
C004	Identifying and classifying ambiguity for regulatory requirements	A K Massey, R L Rutledge, A I Antón, Pater P. Swire	2014	18
C023	Ambiguity in privacy policies and the impact of regulation	Joel R. Reidenberg, Jaspreet Bhatia, Travis D. Breaux, and Thomas B. Norton	2016	22

Iteração 2

Backward – Iteração 2

Código	Título	Cita	É citado por
[BW - 1 - C01]	Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations;	28	256
[BW - 1 - C02]	A Systematic Method for Acquiring Regulatory Requirements: A Frame-Based Approach;	19	28
[BW - 1 - C03]	Analyzing Regulatory Rules for Privacy and Security Requirements	44	304
[BW - 1 - C04]	Towards a framework for tracking legal compliance in healthcare	22	125
[BW - 1 - C05]	Using a security requirements engineering methodology in practice: the compliance with the Italian data protection legislation	14	95
[BW - 1 - C06]	Evaluating existing security and privacy requirements for legal compliance	34	73

Forward – Iteração 2

Código	Título	Cita	É citado por
[FW - 1 - C01]	Toward multilevel textual requirements traceability using model-driven engineering and information retrieval	27	11
[FW - 1 - C02]	A Theory of Vagueness and Privacy Risk Perception	58	7
[FW - 1 - C03]	Managing Ambiguity and Traceability in Regulatory Requirements: A Tool-supported Frame-based Approach	27	1
[FW - 1 - C04]	Modeling Regulatory Ambiguities for Requirements Analysis	19	1
[FW - 1 - C05]	Ambiguity and tacit knowledge in requirements elicitation interviews.	82	13
[FW - 1 - C06]	A Strategy for Addressing Ambiguity in Regulatory Requirements	32	1
[FW - 1 - C07]	Ambiguity in Privacy Policies and the Impact of Regulation	20	15

[FW - 1 - C08]	FOL-Based Approach for Improving Legal-GRL Modeling Framework: A Case for Requirements Engineering of Legal Regulations of Social Media	35	0
[FW - 1 - C09]	Analyzing privacy requirements: a case study of healthcare in Saudi Arabia	24	3

Iteração 3
Backward – Iteração 3

Código	Título	Cita	É citado por
[BW - 2 - C01]	Addressing Legal Requirements in Requirements Engineering	47	149
[BW - 2 - C02]	Detecting Ambiguities in Requirements Documents Using Inspections	28	144

Forward – Iteração 3

Código	Título	Cita	É citado por
[FW - 2 - C01]	Managing Legal Texts in Requirements Engineering;	49	5
[FW - 2 - C02]	A critical analysis of legal requirements engineering from the perspective of legal practice;	35	14

APPENDIX D – INTERVIEW - INVITATION LETTER

Invitation letter to participate in the interview doctoral research

Caro, Sr(a). _____

Sou Dorgival Netto, estudante de doutorado no Programa de Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco (UFPE), orientando da professora Drª Carla Silva.

Venho por meio deste e-mail solicitar a sua participação em uma entrevista que faz parte de um estudo exploratório sobre Ambiguidade na Especificação de Requisitos Legais em empresas governamentais e privadas. Nosso objetivo é coletar informações em organizações governamentais e empresas de desenvolvimento de software sobre **(1)** o processo de especificação de requisitos legais no desenvolvimento de software, **(2)** como a ambiguidade presente no texto legal é tratada na especificação dos requisitos**.**

A entrevista terá duração de 30 a 45 minutos e poderá ser realizada por Skype, Google Meet ou Hangouts.

O resultado da pesquisa será utilizado para a minha tese de doutorado, podendo também ser apresentado em encontros ou em revistas científicas.

Contato dos Pesquisadores:

Dorgival Pereira da Silva Netto – dpsn2@cin.ufpe.br
Currículo Lattes **(http://lattes.cnpq.br/6404552479445485)**
Prof.ª Drª. Carla Silva
Currículo Lattes **(http://lattes.cnpq.br/0581226769296441)**

Desde já muito obrigado pela atenção.

[]'s

APPENDIX E – INTERVIEW - CONSENT TERM

Termo de Consentimento Livre e Esclarecido (TCLE)

Você está sendo convidado(a) para participar, como voluntário, de uma entrevista. Após ser esclarecido(a) sobre as informações a seguir, no caso de aceitar fazer parte do estudo, assine ao final deste documento, que está em duas vias. Uma delas é sua e a outra é do pesquisador responsável. Em caso de recusa você não será penalizado(a) de forma alguma. Em caso de dúvida você pode procurar os pesquisadores responsáveis:

Carla Taciana Lima Lourenço Silva Schuenemann, DSc.

Doutora em Ciência da Computação pela Universidade Federal de Pernambuco. Professora do Centro de Informática da UFPE (Universidade Federal de Pernambuco) desde 2011. Orientadora de estudantes de doutorado.

Currículo Lattes: http://lattes.cnpq.br/0581226769296441

Endereço: Universidade Federal de Pernambuco. Av. Jornalista Anibal Fernandes, S/N, Cidade Universitária, CEP: 50740-560 - Recife, PE – Brasil.

E-mail: ctlls@cin.ufpe.br

Dorgival Pereira da Silva Netto

Doutorando em Ciência da Computação pela UFPE (Universidade Federal de Pernambuco). Professor do Instituto Federal de Mato Grosso do Sul (IFMS) desde 2016.

Currículo Lattes: http://lattes.cnpq.br/6404552479445485

Endereço: Universidade Federal de Pernambuco. Av. Jornalista Anibal Fernandes, S/N, Cidade Universitária, CEP: 50740-560 - Recife, PE – Brasil.

E-mail: dpsn2@cin.ufpe.br

Esta pesquisa de natureza acadêmica sob o título ***Addressing Ambiguity in Legal Requirements Engineering of Software Systems*** tem como objetivo investigar como a as organizações governamentais tratam a ambiguidade presente no texto legal, na especificação de requisitos legais visando garantir a conformidade (do inglês, *compliance*) com a legislação.

Entre os benefícios esperados da pesquisa espera-se obter o entendimento do processo de tratamento da ambiguidade na especificação de requisitos legais no desenvolvimento de software na organização pesquisada. A partir desse entendimento, propor diretrizes para que o conhecimento advindo das práticas das organizações possam ser reusadas por outras empresas e que as diretrizes auxiliam o analista de requisitos a identificar a ambiguidade presente nos requisitos e no texto legal e a especificar um sistema em conformidade com a legislação.

Os participantes da pesquisa serão submetidos a uma entrevista sobre o processo de elicitação e especificação de requisitos de software que devem estar em conformidade com uma determinada legislação. O conteúdo das entrevistas não terá nenhuma influência na avaliação do funcionário no desempenho das suas atividades na organização. A entrevista será gravada para posterior documentação. Se o participante sentir-se constrangido durante o andamento da discussão, tem toda a liberdade de sair, sem ser penalizado de nenhuma forma.

É direito dos sujeitos participantes, e dever da equipe de pesquisadores, mantê-los(las) informados(as) sobre o andamento da pesquisa, mesmo que de caráter parcial ou temporário.

Não há despesas pessoais para os sujeitos participantes em nenhuma etapa da pesquisa, como também não há compensações financeiras ou de qualquer outra espécie relacionadas à sua participação.

Esse estudo não envolve nenhum risco físico, mas os assuntos tratados poderão causar-lhe dor emocional, constrangimento ou emoções desagradáveis ao responder a pergunta. Por isso, como mencionado anteriormente, você poderá encerrar a entrevista no momento em que desejar. Essa pesquisa não envolve remuneração para nenhuma das partes, por isso sua participação é voluntária e gratuita. Além disso, a carga horária de trabalho será respeitada e os encontros serão marcados nos momentos em que os sujeitos apresentarem disponibilidade.

Há garantia incondicional quanto a preservação exclusiva da finalidade científica do manuseio dos dados obtidos. Os dados e imagens obtidos durante este estudo, poderão ser utilizados em futuros eventos científicos.

CONSENTIMENTO DA PARTICIPAÇÃO DA PESSOA COMO PARTICIPANTE

Eu, _____,
abaixo assinado, concordo em participar do Estudo de Caso Exploratório sobre Ambiguidade na Especificação de Requisitos Legais, como voluntário.

Fui devidamente informado e esclarecido pelo pesquisador **Dorgival Pereira da Silva Netto** sobre a pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Também foi me garantido que posso recusar a participar da pesquisa, ou retirar meu consentimento a qualquer momento, mesmo após o início dos trabalhos, sem precisar justificar, sem que isto leve a qualquer prejuízo em minha relação com a organização.

Estou ciente e fui esclarecido de que minha privacidade será respeitada, ou seja, qualquer informação ou elemento que possa de qualquer forma me identificar será mantido em sigilo.

Enfim, tendo sido orientado quanto ao teor de todo o conteúdo aqui mencionado e compreendido a natureza e o objetivo do já referido estudo, manifesto meu livre consentimento em participar, estando totalmente ciente de que não há nenhum valor econômico ou material a receber, ou a pagar, por minha participação.

Local e data: _____

_____

Nome e Assinatura do Participante

_____

Dorgival Pereira da Silva Netto

Entrevistador

APPENDIX F – INTERVIEW SCRIPT

Roteiro da Entrevista

Data da entrevista:___ /___ /____

Horário da entrevista: das ..... às .....

Apresentação

Autoapresentação

Sou Dorgival Pereira da Silva Netto, doutorando em Ciência da Computação pela Universidade Federal de Pernambuco. Estou realizando a minha tese que trata da Ambiguidade na Especificação de Requisitos Legais de Software. Nosso objetivo é coletar informações em organizações governamentais e empresas de desenvolvimento de software sobre **(1)** o processo de especificação de requisitos legais no desenvolvimento de software, **(2)** como a ambiguidade presente no texto legal é tratada na especificação dos requisitos. As perguntas que farei nessa entrevista exigirão que você faça reflexões sobre o assunto. Quanto mais detalhada for a sua resposta, mais informações teremos para analisar. Por favor, não tenha pressa ao responder, todos os detalhes nos interessam. Talvez, durante a entrevista, você tenha a impressão de que algumas perguntas se repetem, mas, na verdade, o que pretendemos é considerar e aprofundar todos os ângulos possíveis de cada tema. O objetivo, portanto, é lhe ajudar a lembrar de detalhes relevantes.

Agradecimento ao participante

Pedir permissão para gravação da entrevista

Estimativa de tempo da entrevista (30 a 45 minutos)

Caracterização da Organização

Cidade da Unidade

Ano de fundação

Quantidade de empregados

Tipo (nacional, multinacional, regional)

Tipo de produto (software próprio, software por encomenda, os dois),

Q1. {**Caracterização do Participante**} Fale um pouco de você: sua formação, idade, trajetória profissional.

Probe: Quanto tempo de experiência você tem na área de TI?

Probe: Quanto tempo de experiência você tem na área de compliance?

Q2. Conte-me como ocorre a organização das equipes/times na sua organização.

Probe: Orientação (por equipe de trabalho, por projeto, ambos).

Probe: A equipe é multidisciplinar?

Q3. Conte-me como ocorre a elicitação de requisitos na sua organização.

Probe: Quais as técnicas utilizadas na extração de requisitos?

Probe: Quais são os(as) papéis/funções dos envolvidos da organização que participam da sessão de elicitação?

Q4. Conte-me como foi a sua participação na equipe de um projeto de software que deveria atender/estar em conformidade com um conjunto de normas/legislações?

Probe: Há algum guideline recomendado para uso na empresa?

Probe: Como foi a interação com o cliente?

Q5. Qual o procedimento da organização quando um requisito de software precisa estar em conformidade (do inglês, *compliant*) com uma legislação?

Probe: Há algum indivíduo com conhecimento jurídico que participa da equipe?

Probe: Cite exemplos das legislações que os sistemas desenvolvidos pela sua organização devem atender?

Q6. Como é o processo de extração dos requisitos a partir do texto legal?

Probe: Quem realiza?

Probe: Quais os passos?

Probe: Há alguma automatização? Mineração?

Probe: Como você identifica as diferentes fontes de requisitos (leis, clientes, stakeholders)?

Probe: Como você lida com requisitos conflitantes entre requisitos legais (legislação) e requisitos de software (clientes e stakeholders)?

Q7. *Como esse trecho legal é mapeado em um requisito?*

Probe: como esse trecho legal é operacionalizado em um requisito?

Q8. Quando há uma ambiguidade no trecho legal, como ela é identificada?

Probe: Quais os procedimentos utilizados pelo time do projeto?

Probes: Como é decidida a interpretação deste trecho legal para o sistema?

Probe: Qual o grau de dificuldade associado à interpretação legal?

Q9. Como o cliente reage às questões pendentes ou às ambiguidades, quando confrontado sobre a decisão da ambiguidade?

Probe: Em algum projeto houve erro de interpretação? Com qual frequência ocorre?

Probe: Se o cliente muda de ideia após a decisão ou se equivoca na interpretação, o que acontece?

Q10. Conte-me como ocorre a especificação de requisitos na sua organização.

Probe: Como os requisitos são documentados?

Probe: Quais técnicas de documentação você utiliza?

Q11. Conte-me como ocorre a obtenção da conformidade legal do sistema em relação à legislação que ele deve atender?

Probe: Quando, no processo de desenvolvimento de software, é verificado se todos os requisitos legais estão de acordo com a legislação?

Probe: Como é verificado se todos os requisitos legais estão de acordo com a legislação? Quais os passos?

Probe: Quando, no processo de desenvolvimento de software, é validado se todos os requisitos legais estão de acordo com a legislação?

Probe: Como é validado se todos os requisitos legais estão de acordo com a legislação?

Probe: Quem são os responsáveis por decidir se o sistema atende a legislação?

Q12. Como a empresa ou o setor está se preparando para atender a Lei Geral de Proteção de Dados (Lei 13.709/2018) que foi sancionada em 14 de Agosto de 2018?

Q13. Há preocupações da empresa em relação ao Regulamento Geral de Proteção de Dados (GDPR). Legislação Europeia que entrou em vigor em maio de 2018?

Q14. Como são definidas as Políticas de Privacidade para o sistema que foi desenvolvido?

Q15. Você gostaria de adicionar alguma informação ou observação que não foi perguntada, mas que você considere importante para o tratamento da ambiguidade na especificação de requisitos legais?

Q14. Pode me indicar algum contato para também contribuir com a pesquisa?

Agradecer a participação

APPENDIX G – INTERVIEW SCRIPT - SPECIALIZED SUPPORT

Roteiro da Entrevista

Data da entrevista:___ /___ /____

Horário da entrevista: das ..... às .....

Apresentação

Autoapresentação

Sou Dorgival Pereira da Silva Netto, doutorando em Ciência da Computação pela Universidade Federal de Pernambuco. Estou realizando a minha tese que trata da Ambiguidade na Especificação de Requisitos Legais de Software. Nosso objetivo é coletar informações em organizações governamentais e empresas de desenvolvimento de software sobre **(1)** o processo de especificação de requisitos legais no desenvolvimento de software, **(2)** como a ambiguidade presente no texto legal é tratada na especificação dos requisitos. As perguntas que farei nessa entrevista exigirão que você faça reflexões sobre o assunto. Quanto mais detalhada for a sua resposta, mais informações teremos para analisar. Por favor, não tenha pressa ao responder, todos os detalhes nos interessam. Talvez, durante a entrevista, você tenha a impressão de que algumas perguntas se repetem, mas, na verdade, o que pretendemos é considerar e aprofundar todos os ângulos possíveis de cada tema. O objetivo, portanto, é lhe ajudar a lembrar de detalhes relevantes.

Agradecimento ao participante

Pedir permissão para gravação da entrevista

Estimativa de tempo da entrevista (30 a 45 minutos)

Caracterização da Organização

Cidade da Unidade

Ano de fundação

Quantidade de empregados

Tipo (nacional, multinacional, regional)

Tipo de produto (software próprio, software por encomenda, os dois),

Q1. {**Caracterização do Participante**} Fale um pouco de você: sua formação, idade, trajetória profissional.

Probe: Qual a sua experiência na área de compliance?

Probe: Qual a sua experiência na área de TI? Ou realizando projetos que apoiam a TI?

Q2. Conte-me como é estruturado o setor/equipe de conformidade legal da sua organização.

Probe: Orientação (por equipe de trabalho, por projeto, ambos).

Probe: Quais os papéis dos membros do time?

Q3. Conte-me como foi a sua participação na equipe de um projeto de um software que deveria atender/estar em conformidade com um conjunto de normas/legislações?

Probe: Há algum guideline recomendado para uso na empresa?

Probe: Como foi a interação com a equipe de desenvolvimento?

Q4. Qual o procedimento da organização quando um software precisa estar em conformidade (do inglês, *compliant*) com uma legislação?

Probe: Há algum indivíduo com conhecimento jurídico que participa da equipe?

Probe: Cite exemplos das legislações que os sistemas desenvolvidos pela sua organização devem atender?

Q5. Como é o processo de extração de trechos legais que o software deve atender?

Probe: Quem realiza?

Probe: Quais os passos?

Probe: Há alguma automatização? Mineração?

Probe: Como você identifica as diferentes fontes de requisitos (leis, clientes, stakeholders)?

Probe: Como você lida com requisitos conflitantes entre requisitos legais (legislação) e requisitos de software (clientes e stakeholders)?

Q6. Quando há uma ambiguidade no trecho legal:

Probe: Quais os procedimentos utilizados?

Probes: Como é decidida a interpretação deste trecho legal para o sistema?

Probe: Qual o grau de dificuldade associado à interpretação legal?

Q7. Como o cliente reage às questões pendentes ou às ambiguidades, quando confrontado sobre a decisão da ambiguidade?

Probe: Em algum projeto houve erro de interpretação? Com qual frequência ocorre?

Probe: Se há a necessidade de realizar uma modificação nas funcionalidades do software após a decisão ou há equívoco na interpretação, o que acontece?

Q8. Conte-me como ocorre a obtenção da conformidade legal do sistema em relação à legislação que ele deve atender?

Probe: Quando é verificado se todos os requisitos legais estão de acordo com a legislação?

Probe: Como é verificado se todos os requisitos legais estão de acordo com a legislação? Quais os passos?

Probe: Quando é validado se todos os requisitos legais estão de acordo com a legislação?

Probe: Como é validado se todos os requisitos legais estão de acordo com a legislação?

Probe: Quem são os responsáveis por decidir se o sistema atende a legislação?

Q9. Como a empresa ou o setor está se preparando para atender a Lei Geral de Proteção de Dados (Lei 13.709/2018) que foi sancionada em 14 de Agosto de 2018?

Q10. Há preocupações da empresa em relação ao Regulamento Geral de Proteção de Dados (GDPR). Legislação Europeia que entrou em vigor em maio de 2018?

Q11. Como são definidas as Políticas de Privacidade para o sistema que foi desenvolvido?

Q12. Você gostaria de adicionar alguma informação ou observação que não foi perguntada, mas que você considere importante para o tratamento da ambiguidade na especificação de requisitos legais?

Q13. Pode me indicar algum contato para também contribuir com a pesquisa?

Agradecer a participação

APPENDIX H – QUESTIONNAIRE TO EVALUATE THE GUIDE

Clarity

The information is clear;

Readability

The information is legible;

Accuracy

The information is accurate;

The information is correct;

The information is reliable;

The information is believable;

The information is meaningful;

Completeness

The information is complete;

The information includes all necessary values;

The information covers the needs of our tasks;

The information is sufficiently complete for our needs;

The information has sufficient breadth and depth for our tasks

Reliability

The information is from a trustable source of information;

The information is trustable;

The information is unreliable;

The information comes from a reliable source;

The information is concerned with the degree of accuracy;

Consistency

The information is presented consistently;

Relevancy

The information is useful to our work;

The information is relevant to our work;

The information is helpful for task at hand;

The information is applicable to our work;

The information is appropriate for our work

Understandability

The information is easy to understand;

The information is easy to comprehend (find, receive);

The information is easy to interpret

Usefulness

The information is useful;

Interpretability

It is easy to interpret what this information means;

This information is difficult to interpret;

It is difficult to interpret the coded information;

This information is easily interpretable;

Informativeness

The information is meaningful;

APPENDIX I – SURVEY SCRIPT

Survey about How Information Technology Practitioners Perform Requirements Specification Addressing Ambiguity Resolution and Compliance with Data Protection Laws

Dear survey participant, thank you very much for sparing 15 minutes of your time to answer our questionnaire.

We are researchers investigating ambiguity resolution in legal requirements specification and the compliance of such requirements with data protection laws under the perspective of IT professionals with practical experience in Requirements Engineering, Business Analysis, or System Analysis. This survey results will help us understand how companies (public and private) deal with the inherent ambiguity in the specification of legal requirements and how they achieve legal compliance of such requirements.

For more information about the processing of your personal data, please click here (https://dorgivalnetto.github.io/survey2021/).

For further information / questions, please contact:

MSc. Dorgival Netto (dpsn2@cin.ufpe.br - primary contact - Federal University of Pernambuco)

Prof. Dr. Carla Silva (supporting scientist - Federal University of Pernambuco)

There are 37 questions in this survey.

Demographics

This section aims to identify the professional profile responding and the background in the area of data privacy.

1
What is your country?
*

Please write your answer here:

2
What is your role?
*

❗ Check all that apply
Please choose **all** that apply:

- [ ] Database Administrator
- [ ] Business Analyst
- [ ] Architect
- [ ] CEO, President, or Owner
- [ ] Back-end Developer
- [ ] Chief Data Officer (CDO)
- [ ] Audit Officer
- [ ] Front-end Developer
- [ ] Business Lead
- [ ] Data Protection Officer (DPO)
- [ ] Designer (Interaction Designer, UX designer) or Specialist in Human-Computer Interaction
- [ ] Legal Expert, Lawyer
- [ ] Privacy expert
- [ ] Product Owner
- [ ] Product or Project Manager
- [ ] Requirements Analyst
- [ ] Requirements Engineer
- [ ] Scrum Master
- [ ] Software Engineer
- [ ] Tester or Quality Assuranc
- [ ] Other:

3

What is your maximum degree?

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ High school
◯ Degree
◯ Post-degree
◯ Master
◯ PhD
◯ Post-doctoral

4

How long have you been working in the software industry?

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ Under one year
◯ Between 1 and 2 years
◯ Between 3 and 5 years
◯ Between 6 and 10 years
◯ More than 11 years

5

How long have you been working on data protection (data privacy) in software projects?

*

Choose one of the following answers
Please choose **only one** of the following:

- ◯ Under one year
- ◯ Between 1 and 2 years
- ◯ Between 3 and 5 years
- ◯ Between 6 and 10 years
- ◯ More than 11 years
- ◯ I do not work

6

What is the size of organization your work?

*

Choose one of the following answers
Please choose **only one** of the following:

- ◯ 1-10 employees
- ◯ 11-50 employees
- ◯ 51-250 employees
- ◯ 251-500 employees
- ◯ 501-1000 employees
- ◯ 1001-2000 employees
- ◯ more than 2000 employees

7

Define the type of your organization from the options below:

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ Public Sector

◯ Private Sector

8

What is your company's practice area?

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ Automotive, Transport

◯ Banking/Finance

◯ E-commerce

◯ Education

◯ Games

◯ Government, Military

◯ Healthcare, Medical

◯ Industry (application domain)

◯ Insurance

◯ Legal

◯ Manufacturing, Supply,

◯ Software, IT

◯ Telecommunications

◯ Other

Personal Perceptions about Ambiguity in Privacy Requirements Specification and Legal Compliance

This section aims to discover IT professional's perceptions regarding privacy

9
The following question aims to identify professionals' perceptions regarding privacy culture for ambiguity resolution and legal compliance.

The questions are divided into five different response scales, and a choice must be based on agreement with what the volunteer believes.

1 - Strongly agree; **2** - Somewhat agree; **3** - Neither agree nor disagree; **4** - Somewhat disagree; **5** - Strongly disagree;

*

Please choose the appropriate response for each item:

	1	2	3	4	5
I believe that privacy and data protection are the responsibility only of the IT (Information Technology) department.	◯	◯	◯	◯	◯
I believe that the content of my company's privacy policy represents the practices we used.	◯	◯	◯	◯	◯
I believe that awareness-raising initiatives (for example, training, campaigns, training courses) for privacy are essential.	◯	◯	◯	◯	◯
I believe that privacy should be addressed from conception (Privacy by Design) when planning its activities.	◯	◯	◯	◯	◯

	1	2	3	4	5
I believe that Specialized Support Areas (Ambiguity Analysis, Anonymization, Legal, among others) are critical for reducing ambiguity in legal requirements specification and compliant with the law.	◯	◯	◯	◯	◯
I believe that promoting the company's privacy culture improves employees' awareness of privacy.	◯	◯	◯	◯	◯
I believe that reducing ambiguity in the privacy requirements specification requires cross-functional teams.	◯	◯	◯	◯	◯
I believe that achieving legal compliance requires cross-functional teams.	◯	◯	◯	◯	◯
I believe that training that instructs all employees in ambiguity analysis in privacy requirements and legal compliance for software systems improves Legal requirements specification.	◯	◯	◯	◯	◯
I believe that the tacit knowledge compensates for the lack of guidelines for reduced ambiguity in the privacy requirements specification.	◯	◯	◯	◯	◯

	1	2	3	4	5
I believe that a systemic view of the company concerning privacy and protection of personal data positively influences compliance with the specification of the software requirements to the law.	◯	◯	◯	◯	◯
I believe that Verification & Validation activities only do not guarantee software privacy requirements specification and legal compliance.	◯	◯	◯	◯	◯

1 - Strongly agree; 2 - Somewhat agree; 3 - Neither agree nor disagree; 4 - Somewhat disagree; 5 - Strongly disagree;

Requirements Specification

This section aims to identify which Methodologies, Techniques, and Tools for Privacy Requirements Specification are used in the Software Industry

10
In which approach was expressed the Privacy Requirements Specification?
*

Check all that apply
Please choose **all** that apply:

- [ ] A structured list of requirements
- [ ] Informal Drawings
- [ ] Goal Models
- [ ] Graphical Notations
- [ ] Natural Language - Use cases
- [ ] Natural Language - Informal text/Plain text
- [ ] Natural Language - User stories
- [ ] UML Diagrams - Use Case Diagrams
- [ ] UML Diagrams - Activity Diagrams
- [ ] UML Diagrams - Class Diagrams
- [ ] UML Diagrams - Sequence Diagrams
- [ ] UML Diagrams - State Machines Diagrams
- [ ] Processes - Activity Diagram
- [ ] Processes - Data Flow Diagram (DFD)
- [ ] Processes - BPMN, BPML
- [ ] Prototypes/User Screens
- [ ] Requirements Specification for Developer (RSD) Approach
- [ ] Structured Analysis Diagrams - Data Flow Diagram (DFD)
- [ ] Structured Analysis Diagrams - Entity-Relationship Diagram (ERD)
- [ ] Tables\ Spreadsheets
- [ ] Specification formal
- [ ] Other:

11
Which kind of requirements tools are used to specify the Privacy Requirements?

*

Choose one of the following answers
Please choose **only one** of the following:

◯ A purpose-specific tool for requirements management (e.g., Orcanos, Pearls, DOORS)

◯ A diagramming tool for requirements (e.g., Enterprise Architect)

◯ Spreadsheets or documents (e.g., Word/Excel)

◯ Content Management Systems (e.g., Confluence)

◯ Application Lifecycle Management or Issue Tracking Systems(e.g., Redmine, Jira)

◯ Other

Choose the answer that most closely applies.

12
Considering your personal experiences, which do the following problems in legal requirements engineering affect your projects?
*

❗ Check all that apply
Please choose **all** that apply:

- [ ] A weak relationship between Customer and project lead.
- [ ] Changing goals, business processes, and/or requirements.
- [ ] Lack of collaboration between system engineers and lawyers
- [ ] Communication flaws within the project development team
- [ ] Communication flaws between developers and the customer
- [ ] Compliance requirements are purposefully expressed in general terms, omitting implementation-specific details.
- [ ] Constant changes in the law make legal compliance difficult
- [ ] Cross-reference among legal/regulatory documents
- [ ] The developer may make their wrong interpretation
- [ ] Difficulty understanding domain-specific terms
- [ ] Identify the regulations relevant to its specific system
- [ ] Incomplete and/or hidden requirements
- [ ] Inconsistent requirements
- [ ] Insufficient support by customer
- [ ] Insufficient support by the project lead.
- [ ] Interpreting the regulation and translating it into implementable requirements.
- [ ] Lack of training in data protection regulations
- [ ] Law's entry into force is concise (timeboxing) and makes it infeasible to implement some features.
- [ ] Missing traceability between requirements and legal text.
- [ ] Standardizing terminology between law, engineering, and business.
- [ ] Training about privacy is tedious or repetitive
- [ ] There is no one from the Legal Support inside the Development Team
- [ ] Underspecified requirements that are too abstract and allow for various interpretations
- [ ] Unclear/unmeasurable non-functional requirement.
- [ ] Weak access to customer needs and/or (internal) business information
- [ ] Weak knowledge about the customer's domain.

☐ Other:

13
Considering the management of traceability among requirements, legal regulations, and document specification, which of the following scenarios best describes the practice in your projects?
*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ A matrix or tool is used to record only the trace links.

◯ A tool is used to record artifacts and trace links.

◯ Trace references are stated inside the requirements, specification and legal regulation artifacts.

◯ Trace links are only recovered to produce final documentation.

◯ Don’t know

◯ Other

Choose the answer that most closely applies.

14
Who is responsible for validating the specification of legal requirements?
*

Check all that apply
Please choose **all** that apply:

- [ ] Customer
- [ ] Development Team Members
- [ ] Data Protection Officer (DPO)
- [ ] Experienced Business analysts
- [ ] Legal internal department
- [ ] Product Owner
- [ ] Project Manager
- [ ] Quality Assurance Analyst
- [ ] Requirement Engineer
- [ ] Security internal department
- [ ] Privacy internal department
- [ ] Other:

15
Please elaborate on your view. How is the process of validating the specification of legal requirements in your company?

Please write your answer here:

Ambiguity

This section aims to identify which Methodologies, Techniques, and Tools for reduced ambiguity in Privacy Requirements are used in the Software Industry.

Whereas ambiguity has been defined as a statement with more than one interpretation (Chantree et al., 2006 (http://Whereas ambiguity has been defined as a statement with more than one interpretation (Chantree et al., 2006))). A requirement is ambiguous if it has multiple interpretations despite the reader's knowledge of the context (Kamsties, 2001 (http://publica.fraunhofer.de/dokumente/N-6409.html)).

An example of ambiguity in the HITECH Act (http://federal.elaws.us/cfr/title45.part170.section170.302), 45 CFR Subtitle A, § 170.302

45 CFR Subtitle A, § 170.302(p): Emergency access. Permit authorized users (who are authorized for emergency situations) to access electronic health information during an emergency.

The definition of an “emergency situation” or what it means to be “authorized for emergency situations” is not provided (Massey et al., 2014 (https://www.cc.gatech.edu/~akmassey/documents/papers/akmassey-re2014.pdf)).

16
Considering your personal experiences, what sources/resources of information/knowledge are you using to solve/reduce ambiguity in legal requirements?
*

❗ Check all that apply
Please choose **all** that apply:

- [ ] Analysis of case law.
- [ ] Ask for clarification to another Company Sector.
- [ ] Ask for clarification from the Government Authority.
- [ ] Books / Blogs or white papers online.
- [ ] Communities online.
- [ ] Consulting Internal Legal Sector.
- [ ] Customer domain knowledge.
- [ ] Direct contact with the customer/customer involved with the project.
- [ ] Discussion between team members.
- [ ] Friends outside the organization.
- [ ] He was reported by a security/privacy audit.
- [ ] Information from managers.
- [ ] Information from experienced team members.
- [ ] Information from stakeholders.
- [ ] Laws/regulatory sources.
- [ ] Organizational procedures.
- [ ] Scientific papers.
- [ ] Standards.
- [ ] Tacit Knowledge.
- [ ] Weekly meeting.
- [ ] Other:

17
How frequently do you encounter ambiguity in legal requirements in your project's company?
*

Please choose **only one** of the following:

◯ 1
◯ 2
◯ 3
◯ 4
◯ 5

1 - never, 2 - seldom, 3 - sometimes, 4 - often, 5 - very often

18
Considering your personal experiences, what techniques are used to reduce the ambiguity that the specification of legal privacy requirements may have?
*

❗ Check all that apply
Please choose **all** that apply:

- [ ] ARES - Agile Requirements Specification.
- [ ] Basic knowledge about law for software engineers
- [ ] Basic knowledge about requirements engineering for lawyers (understand the SDLC briefly).
- [ ] Boilerplates Based Solutions (sections of code that are repeated in multiple places with little to no variation)
- [ ] Controlled Languages Based Solutions.
- [ ] Data dictionary for all domain-specific definitions and acronyms \ Glossary.
- [ ] Delegation of a person for tracing laws and legal regulations.
- [ ] Critical requirements through vulnerabilities in existing software systems, legal violations discussed in administrative and case law.
- [ ] Identification of relevant laws and legal regulations and their analysis performed by lawyers.
- [ ] Inspections Based Solutions
- [ ] Integral Lawyer inside of the development team
- [ ] Lawyers can prepare a separate document describing the law and legal regulation analysis in detail.
- [ ] Natural Language Process (NLP) Based Solutions
- [ ] Ontology-Based Solutions
- [ ] Object-Oriented & UML Based Solutions.
- [ ] Software Requirements Specification (SRS) should contain a section of legal requirements and complete specifications of system requirements related to law.
- [ ] Reusable catalog of legal requirements that were derived from specific legal texts regarding security and personal data protection.
- [ ] Team specialized in handling ambiguity
- [ ] To attach auxiliary annotations to ambiguous sections to flag them for further analysis in collaboration with the proper stakeholders.
- [ ] Training in ambiguity identification techniques.
- [ ] Training in the regulatory domain for the development team.

☐ Transformation of legal regulations to legal requirements, iteratively, in cooperation between lawyers and software engineers.

☐ Using the supplemental documents to identify similar and related legal texts will also help address the identification problem.

☐ We are maintaining traceability between the derived requirements and the source in the legal text.

☐ Other: ____

19
In your view, how difficult is it to interpret ambiguity in legal requirements?

*

Please choose **only one** of the following:

◯ 1
◯ 2
◯ 3
◯ 4
◯ 5

1 - very difficult, 2 - challenging, 3 - neutral, 4 - easy, 5 - very easy

20
Do you use any of these tools for handling ambiguity in Privacy Requirements in your organization?
*

Check all that apply
Please choose **all** that apply:

- [ ] CKCO (Context Knowledge & Concepts Ontology)
- [ ] NAI (Nocuous Ambiguity Identification)
- [ ] NL2OCL (Natural Language to Object Constrained Language)
- [ ] Object-Oriented Visualization
- [ ] QuaARS (Quality analyzer for requirement specification)
- [ ] RESI (Requirements Engineering Specification Improver)
- [ ] RSLingo's RSL (Requirements Specification Language).
- [ ] RSLingo4Privacy Studio
- [ ] SBVR Tools (Semantics of Business Vocabulary and Rules)
- [ ] SREE (Systemized Requirements Engineering Environment)
- [ ] WSD (Word Sense Disambiguation)
- [ ] I don't use any of them
- [ ] Other:

21
Who is responsible for interpreting or resolving ambiguities in the legal excerpt?
*

Check all that apply
Please choose **all** that apply:

- [ ] Customer
- [ ] Development Team Members
- [ ] Data Protection Officer (DPO)
- [ ] Experienced Business analysts
- [ ] Legal internal department\Lawyer
- [ ] Product Owner
- [ ] Project Manager
- [ ] Quality Assurance Analyst
- [ ] Requirement Engineer
- [ ] Security internal department
- [ ] Privacy internal department
- [ ] Other:

22
Please elaborate on your view. How do you solve an ambiguity in a legal excerpt that needs to be represented as a software requirement in your company?

Please write your answer here:

Software Legal Compliance

This section aims to identify which Methodologies, Techniques, and Tools for Software Legal Compliance are used in the Software Industry

23
How would you characterize your awareness regarding the Personal Data Protection Laws (i.e., Brazilian General Law of Personal Data Protection, in Portuguese: Lei Geral de Proteção de Dados Pessoais (LGPD) / General Data Protection Regulation (EU GDPR)?

Consider a Data Protection Act in force in the country you answered in question 1.

*

Choose one of the following answers
Please choose **only one** of the following:

◯ I never heard of it before

◯ I knew something was going on but do not know any details

◯ I was aware there are new regulations and know some details

◯ I am reasonably familiar with the regulations but have a lot more to learn

◯ I am very knowledgeable about the Data Protection Law.

Consider a Data Protection Act in force in the country you answered in question 1.

24
Who provides support for your organization with Personal Data Protection Law (i.e., LGPD[1], GDPR[2], HIPAA[3], among others)?

*

❗ Check all that apply
Please choose **all** that apply:

- [ ] A law firm
- [ ] All Development Team Members
- [ ] Business management
- [ ] IT internal department
- [ ] One of the Big Four (Deloitte, PWC, EY, KPMG)
- [ ] Legal internal department
- [ ] Experienced Business analysts
- [ ] Data Protection Officer (DPO)
- [ ] Requirement Engineer
- [ ] Security internal department
- [ ] Privacy internal department
- [ ] Other:

1. General Law of Personal Data Protection (in Portuguese: Lei Geral de Proteção de Dados Pessoais (LGPD)) - Brazil
2. General Data Protection Regulation (GDPR) - European Union
3. Health Insurance Portability and Accountability Act (HIPAA) - United States

25
How do you rate your company to comply with Personal Data Protection Law (i.e., LGPD[1], GDPR[2], HIPAA[3], among others), given your current data privacy practices?

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ We are already compliant and do not need to change.

◯ Our existing practices will satisfy some of LGPD/GDPR, but we will need to make a few changes.

◯ We are compliant in a few areas but need to make significant changes to be compliant.

◯ We are figuring out who needs to be involved in putting a plan together.

◯ We have not started on our planning.

◯ We are not compliant at all.

1. General Law of Personal Data Protection (in Portuguese: Lei Geral de Proteção de Dados Pessoais (LGPD)) - Brazil
2. General Data Protection Regulation (GDPR) - European Union
3. Health Insurance Portability and Accountability Act (HIPAA) - United States

26
How do you deal with changing legal requirements after the initial release?

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ We regularly change the legal requirements specification.
◯ We update our product backlog.
◯ We only work with change requests.
◯ Test-driven analysis for TDD projects.
◯ We discuss with customers and decide the best approach
◯ Team-based discussion before the change.

Legislation often undergoes updates and modifications, either through an amendment or repeal.

27
Which of the following practices describe how your organization is in managing compliance so far?

*

❗ Choose one of the following answers
Please choose **only one** of the following:

◯ Our Lawyer (s) say(s) that we are compliant.
◯ Our organization has elected a Data Protection Officer (DPO) or similar Officer.
◯ Management or Auditors have mandated improvements
◯ Several business processes were re-engineered due to Privacy requirements.
◯ Other

28
Regarding legal privacy requirements and personal data protection laws, what training and practice should be provided?
*

❗ Check all that apply
Please choose **all** that apply:

- [ ] Concerning data subjects' rights (presented in GDPR Articles 12-23, and LGPD Articles 17-22).
- [ ] Conditions for consent (presented in GDPR Articles 7 and 8, and LGPD Article 8).
- [ ] Data protection impact assessment (DPIA) and prior consultation.
- [ ] Data Protection Officer (DPO), appointment, job descriptions, an overview of tasks codes of conduct, and certification.
- [ ] Legal requirements documentation
- [ ] Information Privacy Framework (e.g., ISO/IEC 29132:2017, ISO/IEC 27701:2019, National Institute of Standards and Technology (NIST) Privacy Framework).
- [ ] Laws and regulations related to the software's subject area to be developed (e.g., patient record law, Privacy and Electronic Communications Regulation (ePrivacy), ICT regulation).
- [ ] Mandatory business/sector/industry requirements and code of conduct.
- [ ] Notification of personal data and information security breaches to the supervisory authority and data breach notification to the data subject.
- [ ] On the duties of data controllers and data handlers (presented in GDPR Articles 24-43, and LGPD 37-40).
- [ ] Penalties and sanctions of Data Protection Laws (presented in GDPR Article 84, and LGPD 52-54).
- [ ] Privacy Principles (presented in GDPR Article 5,and LGPD Article 6).
- [ ] Privacy by design, and privacy by default.
- [ ] Processing special categories of personal data (presented in GDPR articles 9 and 10, and LGPD Article 11 and 14).
- [ ] Records of the data processing activity.
- [ ] Roles and organization in the organization relating to privacy.
- [ ] The lawfulness of processing (presented in GDPR Article 6, and LGPD Article 7)
- [ ] The organization's information privacy requirements and guidelines.
- [ ] The organization's own internal privacy protocols.
- [ ] Other:

IT companies must comply with relevant laws and regulations to avoid the risk of costly penalties, lost reputation, and brand damage resulting from non-compliance.

29
Who should receive training?

*

Check all that apply
Please choose **all** that apply:

- [ ] All employees should have a basic understanding of privacy and information security.
- [ ] Management should be competent in assessing the impact and consequence of privacy implications, risk assessment, management responsibilities, and handling of risks relating to privacy.
- [ ] Project leaders should be competent in data protection topics by design and by default.
- [ ] Developers should be competent in the topics of secure coding and privacy and security by design.
- [ ] Architects should be competent in data protection by design and by default.
- [ ] Testers should be competent in data protection by design and by default.
- [ ] Suppliers should be competent in data protection by design and, by default, data processing agreements, incident response handling, and emergency response. The suppliers should have readable, standardized, and updated privacy documentation to comply with LGPD/GDPR.

30
In the Company, who is responsible for deciding whether the requirements specification complies with the legislation?

*

Check all that apply
Please choose **all** that apply:

- [ ] Customer
- [ ] Development Team Members
- [ ] Data Protection Officer (DPO)
- [ ] Experienced Business analysts
- [ ] Legal internal department\Lawyer
- [ ] Product Owner
- [ ] Project Manager
- [ ] Quality Assurance Analyst
- [ ] Requirement Engineer
- [ ] Security internal department
- [ ] Privacy internal department
- [ ] Other:

31
Is there an individual with legal knowledge[1] who participates in the development team?

1. Who has taken courses, training related to legislation, or is Bachelor Law Degree.

*

Choose one of the following answers
Please choose **only one** of the following:

- ◯ Yes
- ◯ No

1. Who has taken courses, training related to legislation, or is Bachelor Law Degree.

32
What is the organization's procedure when software that will be developed needs to comply with legislation?

Please write your answer here:

Verification & Validation

33
Considering the stages of the SDLC (Software Development Life Cycle), when is it verified that all legal requirements are following the legislation?

*

Check all that apply
Please choose **all** that apply:

- [ ] Requirement collection and analysis
- [ ] Feasibility study
- [ ] Design
- [ ] Coding
- [ ] Testing
- [ ] Installation/Deployment
- [ ] Maintenance

34
Considering the stages of the SDLC (Software Development Life Cycle), when is it validated that all legal requirements are compliant with the legislation?
*

❗ Check all that apply
Please choose **all** that apply:

- [ ] Requirement collection and analysis
- [ ] Feasibility study
- [ ] Design
- [ ] Coding
- [ ] Testing
- [ ] Installation/Deployment
- [ ] Maintenance

35
How is it verified that all legal requirements are compliant with the legislation?

Please write your answer here:

36
How is it validated if all legal requirements are compliant with the law?

Please write your answer here:

37
If you would like to be notified about the research progress and get access to the results from this survey, please enter your email address. Note: your answers will not be associated with your email address, AND we will use your email address for the sole purpose of sending you the results.

Please write your answer here:

Thank you so much for taking the time out of your day to complete our survey. I appreciate your participation! Once again, we are extremely grateful for your contributing your valuable time, your honest information.

Please contact Dorgival Netto for further assistance.

Best regards.

10-12-2021 – 02:45