Um método para transformação de requisitos legais em padrões de requisitos de software: Um estudo com a LGPD

Cinara Gomes de Melo Carneiro[1], **Taciana N. Kudo**[1], **Renato F. Bulcão Neto**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

cinara_melo@hotmail.com, {taciana,rbulcao}@ufg.br

Abstract. *Recent studies show that requirements analysts struggle with interpreting the LGPD to obtain privacy requirements. In this paper, we propose a method to transform LGPD's legal requirements into a software requirement patterns catalogue. Reusing requirements as patterns can be a viable alternative to assist professionals since the LGPD's legal determinations are recurrent in different software projects. A preliminary evaluation of our patterns catalogue suggests the method's effectiveness.*

Resumo. *Estudos recentes mostram que analistas de requisitos têm dificuldade em interpretar a LGPD para obter requisitos de software de privacidade. Neste artigo, nós propomos um método para auxiliar na transformação de requisitos legais da LGPD em um catálogo de padrões de requisitos de software. Reusar requisitos na forma de padrões pode ser uma alternativa viável para auxiliar profissionais, uma vez que as determinações legais da LGPD são recorrentes em diferentes projetos de software. Uma avaliação preliminar do nosso catálogo de padrões aponta indícios da eficácia do método.*

1. Introdução

A Lei Geral de Proteção de Dados Pessoais (LGPD) estabelece regras para coleta, armazenamento, uso e compartilhamento de dados pessoais, e protege o direito dos titulares desses dados de acessá-los, corrigi-los e excluí-los [BRASIL 2018]. Uma das exigências dessa legislação é considerar a privacidade desde o início do desenvolvimento do produto, seguindo os princípios de *Privacy by Design* [Gürses et al. 2011]. Ou seja, é fundamental que a privacidade dos dados pessoais seja considerada desde a Engenharia de Requisitos.

Pesquisas como as de [Alves and Neves 2021] e [Canedo et al. 2021] mostram que a maioria dos profissionais da área de desenvolvimento de software não possui conhecimento sobre requisitos de privacidade, nem sobre a LGPD. Dessa forma, identifica-se a necessidade de apoiar esses profissionais na identificação dos requisitos de privacidade em conformidade com a LGPD e na adequação dos sistemas de informação.

Dada essa necessidade e a ubiquidade de privacidade em sistemas de software, um Padrão de Requisito de Software (PRS) torna-se uma opção viável. Um PRS é uma abstração que serve como modelo de reúso de requisitos para a elicitação, especificação e validação de requisitos em projetos futuros de aplicações com comportamentos semelhantes [Withall 2007]. O uso de PRS ajuda a mitigar problemas, como requisitos incorretos, omitidos, mal interpretados ou conflitantes, gerando economia de tempo e esforço e consistência ao longo do processo.

Apresentamos neste artigo um método de geração de um catálogo de padrões de requisitos de privacidade (CPRP). Na primeira etapa do método, fez-se uma análise sintática do conteúdo dos 65 artigos da LGPD, resultando em 90 requisitos funcionais e 10 não funcionais de privacidade. Em seguida, realizou-se uma análise semântica de cada RF e RNF [Kudo et al. 2023], conforme a gramática de um metamodelo chamado SoPaMM (*Software Pattern MetaModel*) [Kudo et al. 2019], transformando-os em padrões agrupados conforme os capítulos da LGPD.

Uma avaliação preliminar com um profissional com conhecimento de LGPD e Engenharia de Software considerou a completude e corretude do CPRP em relação aos requisitos legais da LGPD. Embora ainda não conclusivos, os resultados obtidos sugerem a eficácia do método. A partir da produção de um CPRP com o método proposto, espera-se auxiliar analistas na elicitação, especificação e validação de requisitos de privacidade por meio do reúso e instanciação dos padrões em projetos de software com características similares no tratamento de dados pessoais.

O artigo está assim organizado: a Seção 2 descreve fundamentação teórica; a Seção 3 analisa trabalhos relacionados; a Seção 4 apresenta o método proposto e a sua aplicação para geração do CPRP; e a Seção 5 discute considerações finais.

2. Fundamentação Teórica

Esta seção descreve o metamodelo SoPaMM como importante fundamento para compreensão desta pesquisa.

O metamodelo SoPaMM descreve como padrões de software devem ser criados, relacionados e classificados [Kudo et al. 2019]. O SoPaMM serve de referência para a construção de modelos terminais (ou padrões de software), e estes podem ser usados para gerar modelos de aplicação (ou especificações do mundo real). Os principais elementos que compõem a gramática do metamodelo SoPaMM são apresentados a seguir. Detalhes podem ser encontrados em [Kudo et al. 2023].

***Catalogue*:** permite reunir padrões de software de um domínio específico;

***SoftwarePatternBag* (SPB):** é utilizado para agrupar padrões de software por categoria;

***SoftwarePattern* (SP):** metaclasse para gerar classes de padrões de software, p.ex., *Functional Requirement Pattern* (FRP), *Non-Functional Requirement Pattern* (NFRP) e *Acceptance Test Pattern* (ATP). Todo padrão de software possui os seguintes atributos: autor, contexto, forças, problema, solução, fonte e versão.

***Functional Requirement Pattern* (FRP):** é uma composição de elementos *Feature*, descritos como histórias de usuário. Cada *Feature* é composta por elementos **Scenario**, descritos como cenários de teste de aceitação que, por sua vez, contêm elementos *Example*, com dados de exemplo para testes.

***Non-Functional Requirement Pattern* (NFRP):** é uma composição de propriedades do sistema de software (restrições comportamentais ou atributos de qualidade) descritas por atributos textuais (p.ex., nome e descrição).

***Acceptance Test Pattern* (ATP):** reúne soluções de teste para comportamentos recorrentes de diferentes cenários. No metamodelo SoPaMM, cada ATP deve ser associado a um FRP e contém um ou mais casos de teste, os quais usam dados de entrada e saída da classe *Example*, vinculada a um *Scenario* de um FRP em particular.

***CatalogueRelationship*, *SPBRelationship* e *SPRelationship*:** permitem, respectivamente, relacionar catálogos de padrões, agrupamentos e padrões de software.

3. Trabalhos Relacionados

Enquanto há pesquisas que dão suporte à elicitação, modelagem e especificação de requisitos de privacidade [Ferrão and Canedo 2022, Saraiva and Soares 2023, Mai et al. 2018, Neves Camêlo and Alves 2023, Peixoto et al. 2023], este trabalho propõe uma abordagem de reúso de requisitos de privacidade baseada nos preceitos legais da LGPD. Essa abordagem, materializada como um catálogo de padrões de requisitos, apoia não apenas a atividade de especificação, mas também de elicitação e validação de requisitos de privacidade.

A elicitação é beneficiada porque os analistas podem reusar as definições dos padrões de requisitos de privacidade do catálogo para seus projetos de software. A especificação também é apoiada porque o catálogo permite a instanciação dos padrões de requisitos sob uma notação de histórias de usuário. Por fim, a validação de requisitos de privacidade também é considerada nesta pesquisa, uma vez que cada padrão de requisito de privacidade constante no catálogo possui comportamento associado na forma de padrões de teste de aceitação, podendo assim beneficiar, não apenas analistas de requisitos, mas também analistas de testes.

Um destaque ao inventário de dados pessoais (IDP) do trabalho de Saraiva e Soares[Saraiva and Soares 2023], que foi elaborado conforme previsto no Art. 37 da LGPD, embora não seja explicitamente declarado nesse artigo. O Art. 37 foi também incluído no catálogo de padrões proposto como um padrão, chamado FRP_MANTER_REGISTRO_OPERAÇÕES. Portanto, a abordagem de Saraiva e Soares [Saraiva and Soares 2023] serve de complemento para esse padrão.

4. Método Proposto

Este trabalho propõe um método para a extração de padrões de requisitos de privacidade por meio da análise sintática e semântica dos artigos da LGPD.

4.1. Sobre a Análise Sintática

Entende-se por análise sintática o estudo da estrutura gramatical das frases e orações, envolvendo a identificação e análise das funções das palavras, a ordem das palavras na frase e as relações sintáticas entre elas. Nesta pesquisa, a análise sintática foi empregada para estruturar os requisitos extraídos da LGPD, de modo a garantir que os requisitos sejam compostos por sujeito, verbo de ação e objeto.

Com o objetivo de identificar direitos e deveres, a fim de extrair os requisitos legais, foram elaborados os seguintes procedimentos:

1. Localizar os atores em cada trecho da LGPD (artigo, artigo e inciso ou artigo e parágrafo). Um ator pode ser o **Titular dos Dados** (pessoa física a quem se referem os dados pessoais); o **Controlador** (pessoa física ou jurídica a quem competem as decisões referentes ao tratamento de dados pessoais); o **Operador** (pessoa física ou jurídica que realiza o tratamento de dados em nome do controlador); e o **Encarregado** (pessoa indicada pelo controlador e operador para intermediar comunicações entre controlador, titulares dos dados e a Autoridade Nacional de Proteção de Dados - ANPD). Esses atores desempenham papéis distintos na proteção de dados pessoais, conforme estabelecido pela LGPD. O ator que representa o sujeito da frase é utilizado inicialmente para construir o requisito.

2. Identificar o verbo de ação, que representa uma ação realizada por um ator.
3. Descobrir qual termo recebe a ação do verbo, ou seja, o objeto.
4. Classificar o requisito como funcional (RF) ou não-funcional (RNF).
5. As posições dos atores são invertidas e os passos anteriores são seguidos.

Segue um exemplo de aplicação da análise sintática do artigo 9º, inciso I da LGPD:

"*Art. 9º - O titular tem direito ao acesso facilitado às informações sobre o tratamento de seus dados, que deverão ser disponibilizadas de forma clara, adequada e ostensiva acerca de, entre outras características previstas em regulamentação para o atendimento do princípio do livre acesso:*

I - finalidade específica do tratamento;"

1. O ator do trecho tratado é o **Titular dos Dados**. Mas, ressalta-se que, implicitamente, o **Controlador** também é ator.
2. No trecho "o titular tem direito de acessar informações", o verbo de ação por parte do titular considerado foi **acessar**.
3. No trecho analisado, pode-se formular a seguinte pergunta: "O titular tem direito de acessar o quê?". Com base na análise do art. 9º inciso I, o objeto vinculado ao verbo de ação em questão é **a finalidade específica do tratamento**.
4. Após seguir os três passos anteriores, obteve-se o requisito de software: "**O titular acessa a finalidade específica do tratamento**". Neste passo, o requisito é classificado como RF.
5. Com a inversão da posição dos atores no RF em questão, o **Controlador** passou a ser o sujeito. Ou seja, além de "O titular acessa a finalidade específica do tratamento", foi também criado o requisito "O controlador disponibiliza a finalidade específica do tratamento", também classificado como RF.
   Além disso, há uma informação adicional no art. 9º, que exige a forma pela qual a finalidade deve ser disponibilizada — "*deverão ser disponibilizadas de forma clara, adequada e ostensiva*"— ou seja, esse texto representa uma locução adverbial. Assim sendo, uma locução adverbial pode sugerir a criação de um RNF, já que indica a maneira pela qual a ação deve ser realizada. Dessa maneira, foi possível identificar mais um requisito de software: "**O controlador deve disponibilizar a finalidade de forma clara, adequada e ostensiva**".
   Ao realizar essa inversão, foram executados os 4 primeiros passos (identificar ator, identificar verbo de ação, identificar objeto e classificar os requisitos).

Após seguir os cinco passos do método de análise sintática do artigo 9º, inciso I da LGPD, resultaram em três requisitos de software, a saber:

RF 01. O titular acessa a finalidade específica do tratamento;
RF 02. O controlador disponibiliza a finalidade específica do tratamento;
RNF 01. O controlador disponibiliza a finalidade de forma clara, adequada e ostensiva.

Ressalta-se que há requisitos que podem ser identificados como pré-requisitos de outros. Por exemplo, ao extrair o requisito "**O Controlador disponibiliza os direitos do titular**" referente ao art. 18, inciso VII da LGPD, é possível inferir o requisito "**O Controlador gerencia os direitos do titular**". Isto ocorre porque, para disponibilizar esses direitos, o referido ator necessita realizar atividades como cadastro, consulta, alteração e exclusão, a fim de garantir a integridade dos dados pessoais.

A seção do método que realiza a análise sintática dos 65 artigos da LGPD produziu 100 requisitos de software de privacidade, sendo destes 90 RF e 10 RNF. Estes requisitos foram analisados, como se descreve em seguida, para convertê-los em padrões de requisitos de privacidade conforme a gramática do metamodelo SoPaMM.

4.2. Sobre a Análise Semântica

A etapa de análise semântica do método considera o significado e o contexto dos termos dos requisitos de privacidade elicitados de modo a estruturá-los segundo a gramática do metamodelo SoPaMM. Para isso, aplica-se o método descrito por [Kudo et al. 2023] para construção de catálogos de padrões de software baseados no SoPaMM, assim como o CPRP mencionado neste trabalho. O método consiste em sete passos, a saber:

1. Se possível, reunir especificações de requisitos e de casos de teste e catálogos de padrões.
2. Estudar cada requisito e caso de teste e classificá-lo.
3. Se um padrão existente puder ser aplicado ao requisito ou caso de teste, registre esse fato e prossiga.
4. Se um padrão existente não se adequar perfeitamente, analisar se o requisito ou o caso de teste pode produzir um novo padrão novo e mais especializado. Caso afirmativo, classificá-lo como padrão de requisito funcional (FRP) ou não funcional (NFRP) ou padrão de teste de aceitação (ATP). Sugerir um nome a esse padrão e adicione-o à lista de padrões candidatos.
5. Depois de passar por todas as especificações, revise os padrões candidatos em busca de duplicatas ou sobreposições e resolva essas inconsistências.
6. Agrupar os padrões de software resultantes (SP) em *bags* de padrões de software (SPB). Observe que um SPB permite a composição de vários e diferentes tipos de SP, como FRP, NFRP e ATP.
7. Escrever cada padrão de software definido.

Considerando os passos supracitados, descreve-se, de uma forma geral, como cada passo desse método foi realizado. Em relação aos passos 1 a 4 do método de [Kudo et al. 2023], foram reunidos e analisados os 90 RF e 10 RNF resultantes da análise sintática dos trechos da LGPD. Dado que não foram encontrados padrões de requisitos de privacidade que pudessem ser reaproveitados para cada RF e RNF, todos os requisitos de privacidade extraídos da LGPD foram transformados em seus respectivos padrões (FRP e NFRP), uma vez que essa lei abrange um domínio transversal, no qual qualquer domínio de software que precise se adequar à legislação deve adotar esses requisitos. Observou-se também que para cada FRP poderia ser desenvolvido, no mínimo, um ATP, pois toda funcionalidade do software pode ser testada.

Cada padrão candidato foi nomeado e adicionado à lista de padrões candidatos. Após revisá-los, identificou-se não haver sobreposição ou duplicidade entre esses padrões (passo 5), de acordo com o conhecimento da pesquisadora sobre a legislação. Pelo passo 6, os padrões FRP, NFRP e ATP foram agrupados em *bags* que, neste caso, os padrões foram agrupados conforme o capítulo da LGPD a que se referem, p.ex. "**DO TRATAMENTO DE DADOS PESSOAIS**". Por fim, no passo 7, o catálogo contendo seus agrupamentos de padrões e a descrição de cada padrão foram especificados conforme a gramática do metamodelo SoPaMM descrita na Seção 2.

A partir dos requisitos **RF 01**, **RF 02** e **RNF 01** resultantes na análise do artigo 9º, inciso I, foram elaborados os seguintes padrões:

- **FRP_Acessar_Finalidade** e **ATP_Acessar_Finalidade**, para o **RF 01**, conforme ilustram as Figuras 1 e 2.
- **FRP_Disponibilizar_Finalidade** e **ATP_Disponibilizar_Finalidade**, relativo ao **RF 02**.
- **NFRP_Disponibilizar_Finalidade_De_Forma_Clara_Adequada_Ostensiva**, referente ao **RNF 01**, como mostra a Figura 3.

Dado que esses padrões foram extraídos de requisitos relativos ao art. 9°, que se encontra no capítulo II - DO TRATAMENTO DOS DADOS PESSOAIS, seção I - Dos Requisitos para o Tratamento dos Dados Pessoais, da LGPD, criou-se uma *bag* nomeada **SPB_Tratamento_de_Dados_Pessoais_REQUISITOS**, em que o nome da *bag* é a junção do nome do capítulo (Tratamento de dados Pessoais) com a sessão (Requisitos), em seguida os padrões foram agrupados nessa *bag*.

A Figura 1 mostra a estrutura de um FRP, utilizando como exemplo o padrão **FRP_Acessar Finalidade**. Para todo FRP identificado, são definidos os seus atributos (autor, contexto, forças, nome, problema, solução, fonte e versão – 1A), pelo menos um elemento *Feature* (1B), utilizando a sintaxe da história do usuário (AS, I CAN, SO THAT) e, para cada elemento *Feature*, um ou mais cenários (1C) na linguagem *Gherkin* (GIVEN, WHEN, THEN), com seus respectivos *Examples* (1D).

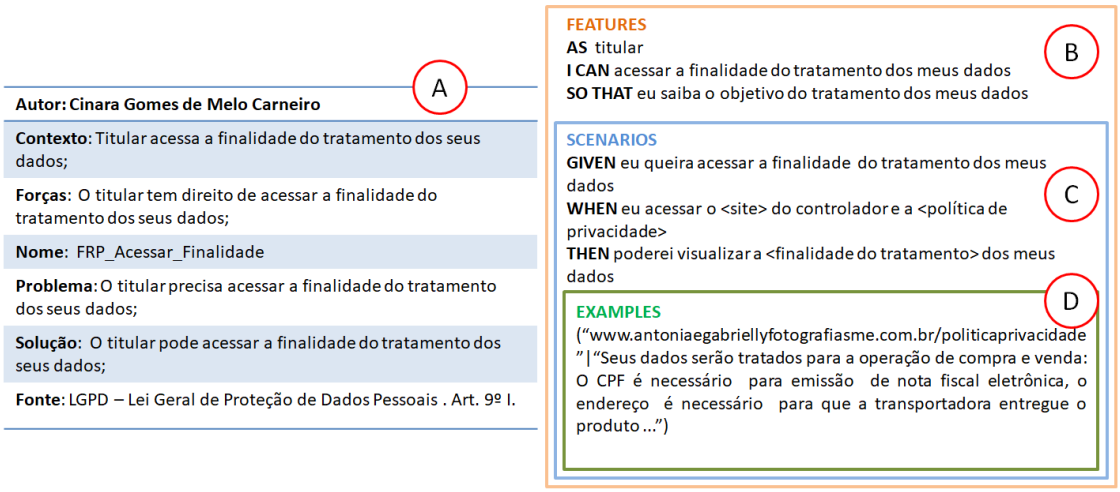

Figura 1. Descrição do Padrão FRP_Acessar Finalidade.

Para a construção de um ATP, deve-se definir os mesmos atributos de um FRP, i.e., autor, contexto, forças, nome, problema, solução, fonte e versão (vide Figura 2A). Depois, elaborar casos de teste que utilizem os cenários específicos de um FRP ao qual o ATP está vinculado. Por exemplo, o cenário **FRP_Acessar_Finalidade** da Figura 1C foi relacionado ao **ATP_Acessar_Finalidade**, conforme ilustra a Figura 2B. Os exemplos ligados a este cenário no FRP, item D da Figura 1, são empregados como entrada e saída para o caso de teste, item C da Figura 2.

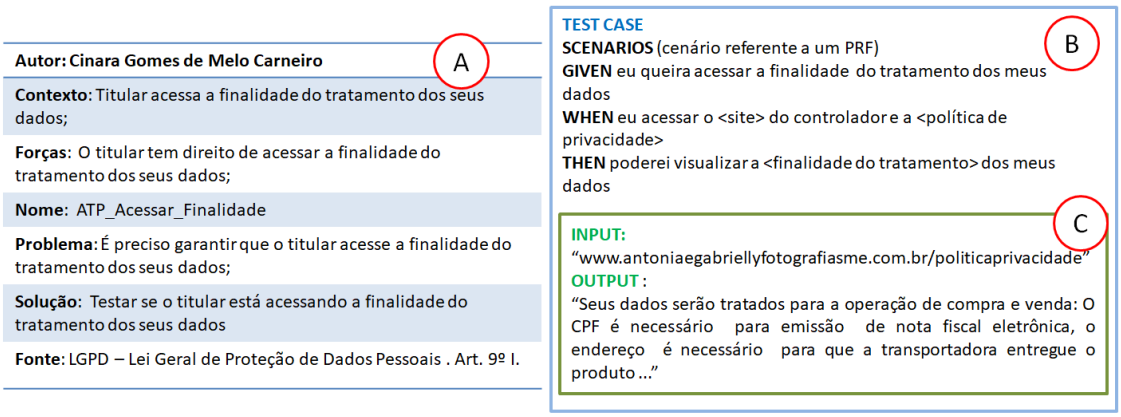

Figura 2. Descrição do Padrão ATP_Acessar Finalidade.

Por fim, para cada NFRP foram criados os seus respectivos atributos, i.e., autor, nome, problema, solução, fonte e versão (Figura 2A) e elementos do tipo *Propriedade do Sistema*, para cada restrição comportamental ou atributo de qualidade (Figura 2B).

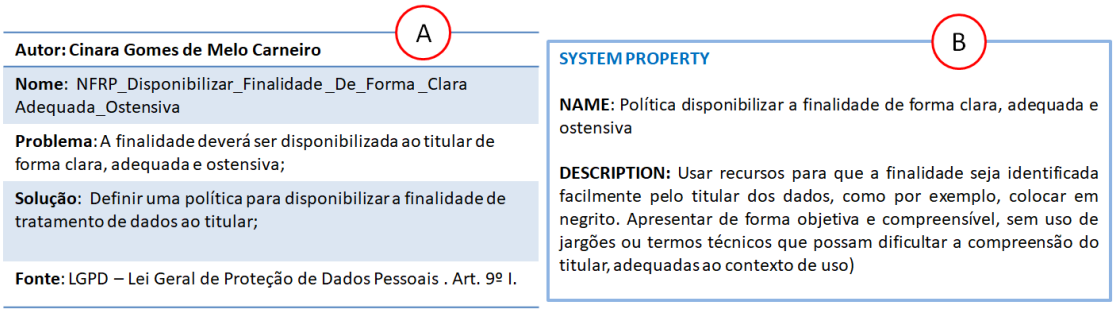

Figura 3. Descrição do Padrão NFRP_Disponibilizar_Finalidade_De_Forma_Clara_Adequada_Ostensiva

O CPRP produzido conta com 90 PRF, 10 PRNF e 91 PTA. Devido a sua extensão, disponibilizamos o catálogo completo no endereço `https://zenodo.org/records/10782808`. A construção do catálogo foi realizada por meio da ferramenta TMEd (*Terminal Model Editor*), desenvolvida pelo grupo de pesquisa [Kudo et al. 2023].

Em seu estágio atual, a qualidade do CPRP foi avaliada por um profissional com conhecimentos de LGPD e Engenharia de Software. A avaliação envolveu dois aspectos do CPRP: o quanto abrange os requisitos legais da LGPD (completude) e o quanto fornece os resultados corretos em relação à LGPD (corretude). Para cada padrão constante no catálogo, foi atribuída uma pontuação de 1 a 5, segundo a escala *Likert*.

A maioria dos padrões avaliados examinados apresentou alta corretude (valor 5), com cerca de 84,5% (160 de 189). Em relação à completude, 80% (24 de 30) dos artigos contemplados no CPRP foram avaliados com a pontuação máxima também. Além disso, a ausência de pontuações inferiores a 4 para corretude e completude indica um alto nível de qualidade nessas dimensões. A descrição completa da avaliação do CPRP desenvolvido será publicada oportunamente.

5. Considerações Finais

Este trabalho apresentou um método que analisa artigos da LGPD e os traduz em um catálogo de padrões de requisitos de privacidade (CPRP). Esses padrões podem ser reusados e instanciados em diferentes projetos onde a proteção de dados pessoais é transversal.

Como ameaças à validade, a análise semântica do método baseia-se no conhecimento da autora sobre a LGPD, apesar de possuir formação complementar compatível. Também não houve um estudo da literatura sobre métodos de análise sintática, o que poderia ter efeito sobre a qualidade do CPRP. Por fim, não foi realizada uma avaliação em uso do CPRP, algo que se planeja realizar como trabalho futuro.

Referências

Alves, C. and Neves, M. (2021). Especificação de requisitos de privacidade em conformidade com a LGPD: resultados de um estudo de caso. In *Anais do WER21 - Workshop em Engenharia de Requisitos, Brasilia, BSB, Brasil, August 23-27, 2021*.

BRASIL (2018). Lei no 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). seção 1, Brasília, ano 155, n. 157, p. 59, 15 ago. 2018.

Canedo, E. D. et al. (2021). Agile teams' perception in privacy requirements elicitation: LGPD's compliance in Brazil. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pages 58–69. IEEE.

Ferrão, S. É. R. and Canedo, E. D. (2022). Uma taxonomia para requisitos de privacidade e sua aplicação no Open Banking Brasil. In *Anais do WER22 - Workshop em Engenharia de Requisitos, Natal - RN, Brazil, August 23-26, 2022*.

Gürses, S., Troncoso, C., and Diaz, C. (2011). Engineering privacy by design. *Computers, Privacy & Data Protection*, 14(3):25.

Kudo, T. N., Bulcão-Neto, R. F., and Vincenzi, A. M. (2019). A conceptual metamodel to bridging requirement patterns to test patterns. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*, pages 155–160.

Kudo, T. N., Bulcão-Neto, R. F., Neto, V. V. G., and Vincenzi, A. M. R. (2023). Aligning requirements and testing through metamodeling and patterns: Design and evaluation. *Requirements Engineering*, 28(1):97–115.

Mai, P. X. et al. (2018). Modeling security and privacy requirements: a use case-driven approach. *Information and Software Technology*, 100:165–182.

Neves Camêlo, M. and Alves, C. (2023). G-Priv: A guide to support LGPD compliant specification of privacy requirements. *iSys - Brazilian Journal of Information Systems*, 16(1):2:1 – 2.

Peixoto, M. et al. (2023). Evaluating a privacy requirements specification method by using a mixed-method approach: results and lessons learned. *Requirements Engineering*, 28(2):229–255.

Saraiva, J. and Soares, S. (2023). Adoption of the LGPD inventory in the user stories and BDD scenarios creation. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering, SBES 2023*, pages 416–421. ACM.

Withall, S. (2007). *Software requirement patterns*. Pearson Education.