Instituto Federal de Educação, Ciência e Tecnologia da Paraíba

Campus João Pessoa

Programa de Pós-Graduação em Tecnologia da Informação

ANDERSON FERNANDO VIEIRA DA BOA MORTE

PRIVACYCHAIN: UM FRAMEWORK PARA COMPATIBILIZAÇÃO DO TRATAMENTO DE DADOS PESSOAIS EM APLICAÇÕES BASEADAS EM DLT COM OS DIREITOS AO ESQUECIMENTO E À RETIFICAÇÃO DA LGPD

DISSERTAÇÃO DE MESTRADO

JOÃO PESSOA – PB

2022

Anderson Fernando Vieira da Boa Morte

PrivacyChain: Um framework para compatibilização do tratamento de dados pessoais em aplicações baseadas em DLT com os direitos ao esquecimento e à retificação da LGPD

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Tecnologia da Informação pelo Programa de Pós-Graduação em Tecnologia da Informação do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB.

Orientador: Prof. Dr. Dênio Mariz Timóteo de Sousa

Coorientador: Prof. Dr. Rostand Edson Oliveira Costa

João Pessoa – PB

2022

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Nilo Peçanha do IFPB, *campus* João Pessoa

B662p Boa Morte, Anderson Fernando Vieira da.
Privacychain : um *framework* para compatibilização do tratamento de dados pessoais em aplicações baseadas em DTL com os direitos ao esquecimento e à retificação da LGPD / Anderson Fernando Vieira da Boa Morte. – 2022.
79 f. : il.
Dissertação (Mestrado - Tecnologia da Informação) - Instituto Federal de Educação da Paraíba / Programa de Pós-Graduação em Tecnologia da Informação (PPGTI), 2022.
Orientação : Profº D.r Dênio Mariz Timóteo de Sousa.
Coorientação : Profº D.r Rostand Edson O. Costa.
1. Programação - *framework*. 2. Proteção de dados - lei. 3. LGPD. 4. DTL. 5. Privacidade. I. Título.
CDU 004.4:340(043)

Lucrecia Camilo de Lima
Bibliotecária – CRB 15/132

Anderson Fernando Vieira da Boa Morte

PrivacyChain: Um framework para compatibilização do tratamento de dados pessoais em aplicações baseadas em DLT com os direitos ao esquecimento e à retificação da LGPD

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Tecnologia da Informação pelo Programa de Pós-Graduação em Tecnologia da Informação do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB.

Aprovado em 28 de março de 2022.

BANCA EXAMINADORA:

Prof. Dr. Paulo Ditarso Maciel Júnior – IFPB

ICP Brasil
Documento assinado digitalmente
GUIDO LEMOS DE SOUZA FILHO
Data: 03/05/2022 10:05:21-0300
Verifique em https://verificador.iti.br

Prof. Dr. Guido Lemos de Souza Filho – UFPB

Avaliador Externo

Prof. Dr. Dênio Mariz Timóteo de Sousa

Orientador

Prof. Dr. Rostand Edson Oliveira Costa

Coorientador

Visto e permitida a impressão

João Pessoa

Prof. Dr. Francisco Petrônio Alencar de Medeiros

Coordenador PPPGTI

Este trabalho é dedicado aos meus pais Antônio Fernando e Mara, aos meus irmãos Alisson e Adriana, à minha esposa Daniela e às minhas filhas Maria Fernanda e Mariana.

AGRADECIMENTOS

Agradeço a Deus pela oportunidade concedida, por colocar todas as coisas em seus devidos lugares, de modo que eu reunisse as melhores condições para o desenvolvimento desse trabalho que tanto me orgulha.

Agradeço à minha esposa, Daniela, pela compreensão e pelo suporte que permitiu que eu pudesse me dedicar aos estudos. Um cheiro carinhoso por cuidar das meninas e por fazê-las — nem sempre com sucesso! — entender que o papai não poderia brincar naquele momento, nem dar a atenção que vocês tanto merecem e precisam.

Agradeço ao IFPB, instituição pública e de qualidade, cuja lembrança me remete ao meu saudoso CEFET-BA, e que a partir de então também guardarei boas lembranças.

Agradeço à Dataprev, pela oportunidade de transferência do Rio para João Pessoa — cidade que me proporcionou um ecossistema fantástico, possibilitando o desenvolvimento deste trabalho.

Agradeço ao Professor Guido Lemos, por me conceder a oportunidade de ser aluno especial na UFPB, ainda em 2018 — onde tive o meu primeiro contato acadêmico com o assunto *blockchain*.

Agradeço ao Professor Francisco Petrônio, pela maneira primorosa em conduzir a PPGTI.

Agradeço ao Professor Paulo Ditarso e aos demais professores da PPGTI pelos valiosos ensinamentos ao longo dessa trajetória.

Agradeço aos meus colegas pela camaradagem, pelos ensinamentos compartilhados e pelos momentos de descontração ao longo dessa caminhada.

Agradeço à minha colega, Anália Meira pelos valiosos incentivos e conselhos durante a execução deste trabalho.

Em especial, agradeço aos meus orientadores Prof. Dr. Dênio Mariz e Prof. Dr. Rostand Costa pela constante dedicação e esmero em me orientar. Foi um processo de busca constante pela excelência, processo que me serve de grande exemplo ao meu aprimoramento pessoal e profissional. Muito obrigado!

RESUMO

Aplicações que fazem uso de DLT (*Distributed Ledger Technology*) em geral e *blockchains* em particular se beneficiam das características de imutabilidade e transparência para registro de dados, mas ao mesmo tempo enfrentam um obstáculo que é prover aos titulares desses dados o exercício dos direitos ao esquecimento (RTBF - *Right to be Forgotten*) e à retificação previstos na LGPD - Lei Geral de Proteção de Dados Pessoais. Ou seja, uma vez que esses dados pessoais são registrados em *blockchains* eles não podem ser removidos ou alterados, o que requer uma abordagem especial para lidar com essas exigências legais de privacidade. Diante deste problema de pesquisa, conduziu-se uma investigação de um conjunto de boas práticas e de técnicas de armazenamento *off-chain* e *commitment criptográfico*, que levou à proposta do *framework* denominado de PrivacyChain, uma ferramenta de software para prover serviços integrados e transparentes para o desenvolvimento de novas aplicações ou ajustes de aplicações legadas através de uma API de serviços. Portanto, dado o crescente uso de *blockchain*, um dos desafios é sua utilização sem prescindir da privacidade e dos direitos ao esquecimento e à retificação previstos na LGPD. A validação e demonstração de viabilidade do PrivacyChain é feita com uma implementação de referência dos serviços em uma API REST, acompanhada de um modelo de uso para guiar e demonstrar a sua utilidade no desenvolvimento de aplicações que usam DLT de forma que mantenham a conformidade com a LGPD no que se refere ao provimento do exercício dos direitos ao esquecimento e à retificação ao titular dos dados pessoais gerenciados.

Palavras-chaves: *blockchain*; DLT; LGPD; RTBF; direito ao esquecimento, privacidade.

ABSTRACT

Applications that make use of DLT (Distributed Ledger Technology) in general and blockchains in particular, benefit from the characteristics of immutability and transparency for data recording, but at the same time they face the obstacle of providing data owners with the exercise of the "right to be forgotten" (RTBF) and "data rectification" provided by the LGPD - General Law for the Protection of Personal Data, in Brazil. That is, once this personal data is recorded on blockchains it cannot be removed or changed, which requires a special approach to dealing with these legal privacy requirements. Faced with this research problem, an investigation of a set of best practices and techniques for off-chain storage and cryptographic commitment was carried out, which led to the proposal of the framework called PrivacyChain, a software tool to provide integrated and transparent services for developing new applications or tweaking legacy applications, through a services API. Therefore, given the growing use of blockchain, one of the challenges is its use without giving up privacy and the rights to forgetfulness and rectification provided for in the LGPD. The validation and feasibility demonstration of PrivacyChain is done with a reference implementation of the services in a REST API, accompanied by a usage model to guide and demonstrate its usefulness in the development of applications that use DLT in a way that maintains compliance with the LGPD with regard to the provision of the exercise of right to be forgotten and data rectification to the holder of the personal data managed.

Keywords: *blockchain*; DLT; LGPD; GDPR; RTBF; *privacy*, PBD.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AEPD	*Agencia Española Protección Datos*
API	*Application Programming Interface*
CNIL	*Commission Nationale de l'Informatique et des Libertés*
CRUD	*Create, Read, Update, Delete*
DLT	*Distributed Ledger Technology*
DSR	*Design Science Research*
GDPR	*General Data Protection Regulation*
GPS	*Global Positioning System*
IPFS	*InterPlanetary File System*
JSON	*JavaScript Object Notation*
LGPD	Lei Geral de Proteção de Dados Pessoais
NPII	*Non-personally Identifiable Information*
PBD	*Privacy by Design*
PII	*Personally Identifiable Information*
PPII	*Potential Personally Identifiable Information*
RTBF	*Right to be Forgotten*
SGBD	Sistema de Gerenciamento de Banco de Dados

SUMÁRIO

INTRODUÇÃO

1.1. Motivação e Definição do Problema de Pesquisa

A Lei 13.709/2018, conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD), surgiu como mecanismo para regular o tratamento de dados pessoais em território nacional, permitindo a garantia dos direitos fundamentais de liberdade e privacidade, não só dos cidadãos brasileiros, mas de qualquer indivíduo cujos dados pessoais sejam aqui tratados. Essa iniciativa adequa a legislação brasileira aos regulamentos já aplicados em outras partes do mundo, como é o caso do GDPR (*General Data Protection Regulation*) no contexto da União Europeia e no qual a LGPD foi baseada (BRASIL, 2018).

A LGPD define "tratamento de dados"[1] como qualquer operação realizada com dados pessoais, tais como: utilização, processamento, armazenamento, distribuição, eliminação e modificação (BRASIL, 2018). O GDPR possui definição semelhante (PARLIAMENT, 2016), mas o regulamento europeu é considerado mais específico e detalhado que a lei brasileira, de forma que se pode afirmar que um tratamento em conformidade com o GDPR também estará em conformidade com a LGPD, embora o contrário não seja necessariamente verdadeiro.

Ambos os regramentos LGPD e GPDR definem perfis. O **Titular** do dado pessoal (*Data Subject* no GDPR) é o detentor do direito fundamental de privacidade e transparência no tratamento de dados. O **Controlador** (*Controller* no GDPR) é o responsável pela tomada de decisões sobre o tratamento. O **Operador** (*Processor* no GDPR) é aquele que realiza a operação de tratamento em nome do controlador. Há também a figura da *Autoridade de Proteção de Dados* (DPA - *Data Protection Authority* no GDPR) que é um órgão governamental que, dentre outras atribuições, possui o dever de fiscalizar e orientar, inclusive tecnicamente, sobre o tratamento de dados pessoais (ITSRIO, 2019).

No âmbito da orientação técnica, instituições como o Parlamento Europeu e a Autoridade Francesa de Proteção de Dados (CNIL - *Commission Nationale de l'Informatique et des Libertés*), diante do crescente número de aplicações baseadas em tecnologia de registro distribuído (DLT – *Distributed Ledger Technology*) e consequentemente do grande volume de dados pessoais tratados por essas, têm publicado documentações no sentido de orientar como realizar esse tratamento de forma a estar em conformidade com a legislação (FINCK, 2019) (CNIL, 2018). Outros estudos no âmbito acadêmico e científico também tratam igualmente dessa adequação (FABER et al., 2019), (ONIK 2019), (POLITOU 2019).

[1] Art. 5º inciso X da Lei 13.709/2018.

A análise dessas recomendações e estudos mostra que o tratamento de dados pessoais realizado por essas aplicações baseadas em DLT - ou *blockchain*, um tipo específico de DLT - deve ser cercado de vários cuidados porque alguns dos direitos dos titulares dos dados elencados na LGPD, tais como o **direito ao esquecimento** (apagamento dos seus dados) e o **direito à retificação** (edição dos seus dados), esbarram em uma característica basilar de uma *blockchain*: a imutabilidade.

Este portanto é o problema a ser analisado por este trabalho: *diante de um cenário em que há cada vez mais aplicações baseadas em blockchain envolvendo o tratamento de dados pessoais, como esse tratamento pode ser realizado em conformidade com a legislação (LGPD), notadamente quanto ao exercício dos direitos ao esquecimento e à retificação, frente à característica de imutabilidade intrínseca da DLT?*

Uma técnica utilizada para viabilizar a conformidade com a legislação do tratamento de dados pessoais realizado por aplicações baseadas em DLT é o armazenamento de dados pessoais *off-chain*. Alguns estudos como (FABER et al., 2019; ONIK et al., 2019; POLITOU et al., 2019) ressaltam a importância desta técnica, considerando-a fundamental para viabilizar essa conformidade. Em linhas gerais, o armazenamento *off-chain* consiste em salvar os dados pessoais em bases de dados locais fora da rede *blockchain*, enquanto um *hash* desses dados é salvo na cadeia de blocos, o que é chamado de armazenamento *on-chain*.

De modo geral, a análise destes estudos a respeito do tratamento de dados pessoais por aplicações baseadas em DLT em aderência aos direitos ao esquecimento e à retificação permite afirmar a adequabilidade de duas técnicas combinadas:

1. Armazenamento *off-chain* de dados pessoais (FABER et al., 2019; ONIK et al., 2019; POLITOU et al., 2019);
2. Armazenamento dos dados pessoais na cadeia através de um compromisso criptográfico (CNIL, 2018).

Neste contexto, o objetivo deste trabalho é propor um *framework* que ofereça serviços que possibilitem que as aplicações baseadas em DLT realizem o tratamento de dados pessoais em conformidade com a LGPD, respeitando-se os direitos ao esquecimento e à retificação sem ter que realizar operações diretamente com a *blockchain*. Este *framework* baseia-se fundamentalmente na utilização das duas técnicas citadas, de maneira combinada.

A metodologia utilizada para desenvolver e propor essas estratégias se baseia na DSR (*Design Science Research*) e engloba a análise de documentos regulatórios relacionados com o GDPR, além da análise do próprio texto da LDPD e de outras fontes na literatura científica, para compilar definições, destacar problemas e desafios inerentes ao processo de tratamento de dados pessoais realizado por aplicações baseadas em DLT, visando garantir os direitos ao esquecimento e à retificação conforme as normas.

Aplicando-se a metodologia proposta, serão prospectadas e consolidadas boas práticas para o tratamento de dados pessoais em aplicações baseadas em DLT, sobretudo para respeito aos direitos ao esquecimento e à retificação da LGPD. Adicionalmente, de forma a materializar essas boas práticas, será proposto um *framework* objetivando facilitar a construção de aplicações compatíveis com tais direitos. A validação da proposta de solução será feita através de uma implementação de referência capaz de demonstrar sua viabilidade de implementação e a sua aplicabilidade ante ao problema de pesquisa.

1.2. Objetivos

1.2.1. Objetivo geral

O objetivo desta pesquisa é propor a especificação e implementação de um *framework* – a ser utilizado por aplicações baseadas em DLT - que forneça serviços que possibilitem o tratamento de dados pessoais para aderência aos direitos ao esquecimento e à retificação.

1.2.2. Objetivos específicos

- Elicitar as exigências legais e regulatórias impostas pela LGPD no que se refere ao exercício dos direitos ao esquecimento e à retificação pelos titulares dos dados pessoais;
- Identificar os problemas comuns enfrentados por aplicações baseadas em DLT ao lidar com dados pessoais considerando os direitos ao esquecimento e à retificação;
- Localizar na literatura científica mecanismos que permitam mitigar, contornar ou resolver os problemas identificados;
- Compilar um conjunto de boas práticas propostas para o desenvolvimento de aplicações baseadas em DLT, compatíveis com os direitos ao esquecimento e à retificação da LGPD;
- Especificar um *framework* que mapeie as boas práticas compiladas com o objetivo de facilitar a construção de aplicações compatíveis com esses direitos;
- Desenvolver uma implementação de referência do *framework* proposto na forma de uma API e apresentar o código implementado como forma a validar a viabilidade da solução proposta;
- Apresentar o modelo de uso do *framework* de modo a atestar a aplicabilidade da solução ante ao problema de pesquisa.

1.3. Estrutura do Documento

Os demais capítulos estão organizados da seguinte maneira.

O Capítulo 2 é dedicado à Fundamentação Teórica e apresenta conceitos fundamentais e tecnologias para melhor compreensão da solução proposta.

O Capítulo 3 dedica-se à apresentação dos Trabalhos Relacionados e discute trabalhos que fundamentam ou se assemelham a esta pesquisa. Apresenta os diferenciais desta pesquisa em relação a esses Trabalhos.

No Capítulo 4 está descrita a Metodologia utilizada como fio condutor da pesquisa. Esta baseia-se na DSR na medida em que, diante de um problema real e relevante, visa a construção de artefatos que podem ser validados, aprimorados e apresentados como alternativa de solução ao problema em questão.

No Capítulo 5 é apresentado o artefato que se propõem a servir como solução ao problema de pesquisa.

O Capítulo 6 apresenta a proposta de validação do artefato na forma de uma Implementação de Referência do *framework* proposto, discute detalhes técnicos acerca da implementação e descreve um modelo de uso do artefato. Ou seja, como sua API pode ser usada para integrar os serviços propostos no desenvolvimento de aplicações baseadas em *blockchain*, visando prover ao titular dos dados pessoais o direito ao esquecimento e o direito à retificação.

Por fim, no Capítulo 7 estão as considerações finais incluindo, uma análise e discussão em torno da pesquisa realizada, destaque das contribuições deste trabalho, trabalhos futuros e a conclusão.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta e discute os elementos conceituais importantes para entendimento da abordagem do problema de pesquisa e da proposta de solução. A seleção desses elementos conceituais baseia-se na avaliação de técnicas aplicáveis à solução do problema, que formam o referencial teórico para esta pesquisa. Para melhor entendimento do restante deste trabalho, esses elementos conceituais são discutidos antecipadamente de forma isolada. Uma discussão sobre a conexão desses conceitos com a solução proposta é apresentada no Capítulo 5.

2.1. LGPD - Lei Geral de Proteção de Dados Pessoais

A LGPD surgiu com o objetivo de regular o tratamento de dados pessoais em território nacional, possibilitando a garantia dos direitos fundamentais de liberdade e privacidade de qualquer indivíduo cujos dados pessoais sejam aqui tratados. Iniciativa que adequa a legislação brasileira a regulamentos como o GDPR no contexto da União Europeia e no qual a LGPD foi baseada (BRASIL, 2018).

O artigo 25 do GDPR apresenta um conceito que vem sendo referenciado na literatura como "privacidade desde a concepção" (do inglês *Privacy by Design* - PBD). Em linhas gerais, a ideia desse artigo, que de forma indireta também está presente na LGPD, é que devem ser priorizadas ações no intuito de se preservar a privacidade do titular dos dados pessoais desde a concepção dos projetos, que incluem o desenvolvimento de produtos de *software*. Na prática, isso significa garantir que a privacidade seja incorporada ao nascimento do sistema e seja tratada durante todo o seu ciclo de vida.

Os artigos 1 e 2 da LGPD realçam o propósito da lei que é a proteção da privacidade do indivíduo. Assim, independente da tecnologia utilizada para o tratamento desses dados, a lei garante ao titular que este possa retificar ou excluir os seus dados pessoais.

A arquitetura das aplicações tem que ser concebida de modo que qualquer tratamento realizado por essas tenha o consentimento do titular e a qualquer tempo esses dados pessoais possam ser retificados ou excluídos a pedido desse.

“A lei é ainda mais rígida quando o tratamento envolve dados sensíveis (dados biométricos, étnicos, religiosos etc.), condicionando o seu tratamento a um consentimento específico e destacado.” (BOA MORTE et al., 2020, p. 6)

2.1.1. Tratamento de Dados Pessoais

O artigo 5º da LGPD define tratamento de dados pessoais como qualquer operação realizada com esses dados. Incluem-se nesse rol, as ações de compartilhamento e armazenamento de dados.

Ao logo deste trabalho serão discutidos como deve ser realizado o compartilhamento de dados pessoais, e como esses dados devem ser armazenados de forma a respeitar-se os direitos do titular, com maior ênfase nos direitos ao esquecimento e à retificação.

2.1.2. Direito ao esquecimento

O Direito ao Esquecimento é um elemento-chave da LGPD e abrange o direito do titular (p.ex. consumidor) solicitar que todos os dados pessoais mantidos pelo controlador (p.ex. empresa) sejam removidos.

No artigo 18 da LGPD, inciso VI, é garantida a "eliminação dos dados pessoais tratados com o consentimento do Titular". Há, entretanto, algumas exceções elencadas no artigo 16, dentre as quais: o cumprimento de obrigação legal ou regulatória pelo controlador, ou a realização de estudo por órgão de pesquisa.

No GDPR esse direito é apresentado no artigo 17. Diferente da LGPD, o GPDR enfatiza a questão do prazo de atendimento a essa solicitação. É ressaltado que o controlador deve "sem demora injustificada" realizar o apagamento dos dados. Estão ressaltados também os motivos pelos quais os dados podem ser apagados, dentre eles:

- os dados não são mais necessários para a finalidade que motivou sua coleta ou processamento;
- o titular retira o consentimento para uso dos seus dados;
- o titular opõe-se ao tratamento dos seus dados.

Na literatura associada a esse direito, observa-se o uso da sigla RTBF (*Right to be Forgotten*) em referência ao direito ao esquecimento (POLITOU 2019) (CASINO et al., 2020) (SHAHAAB, 2020).

As implicações práticas do RTBF em sistemas de informação existentes são enormes. Um dos problemas, por exemplo, relaciona-se com a integridade referencial de dados estruturados em SGBD. Em geral, os SGBD formam uma camada autocontida e autocontrolada abaixo da lógica de negócio do sistema que impõe regras de dependência para inserção, atualização e remoção de dados (p.ex. integridade referencial com uso de chaves estrangeiras). Em muitos casos, a camada lógica superior, implementada no código de programação, é forçada a decidir sobre a remoção de dados com base em um "tudo ou nada" que envolve remover em cascata ou manter o dado. Ao decidir pela remoção, deve-se garantir que outros componentes do sistema não sejam afetados, como por exemplo, a perda de dados históricos. Um dos impactos do RTBF é, portanto, a necessidade de revisão de sistemas existentes para desacoplar logicamente dados pessoais de outros dados que quebrariam o encadeamento lógico dos elementos funcionais do sistema, de forma que seja possível

remover apenas os dados pessoais sem impacto na estrutura dos demais dados que podem ser mantidos.

Sistemas que armazenam dados pessoais em DLT enfrentarão o problema de lidar com a imutabilidade da *blockchain*. Essa estratégia deve ser revista e contornada para adotar mecanismos de armazenamento *off-chain*, discutidos adiante na Seção *2.3 - Armazenamento de dados off-chain*.

2.1.3. Direito à retificação

O artigo 18 da LGPD garante ao titular do dado pessoal, em seu inciso III, o direito à correção de dados incompletos, inexatos ou desatualizados. De acordo com a lei, o Titular pode obter do Controlador a correção dos seus dados pessoais a qualquer tempo e mediante requisição.

Se um controlador (p.ex. empresa) receber do titular (p.ex. um cliente) um pedido de retificação, deve tomar as medidas cabíveis para se certificar de que os dados são precisos e retificá-los, se necessário. As medidas cabíveis dependerão, em particular, da natureza dos dados pessoais e para que serão usados. Quanto mais importante for a exatidão dos dados pessoais, maior deverá ser o esforço da empresa para verificar a veracidade do dado e proceder à sua retificação.

No GDPR é o artigo 16 que confere ao titular o direito à retificação de seus dados pessoais. Nesse contexto, e evidenciando a importância desse direito, (SHAHAAB, 2020) apresenta um trabalho onde a DLT mostra-se apropriada para garantir a pessoas transgêneros o direito à privacidade e confiabilidade quando da solicitação de mudança de gênero.

2.2. DLT - Distributed Ledger Technology

A Tecnologia de Registro Distribuído (DLT - *Distributed Ledger Technology*) constitui-se em uma categoria tecnológica que engloba, além de outras tecnologias, a *blockchain*. Tem como uma de suas principais características a replicação do *ledger* (livro-razão) entre os nós de uma rede *peer to peer* (P2P). Outras características incluem a imutabilidade do *ledger* e a ausência de um nó central controlador. Na *blockchain*, a estrutura de dados distribuída é uma sequência de blocos encadeados por *strings hash*. Outras tecnologias como a Tangle e a *Hashgraph* utilizam-se de outros tipos de estruturas distribuídas (SOARES; COSTA, 2019).

2.2.1. Contextualização

Bitcoin e Ethereum são marcos no universo das criptomoedas. A Bitcoin por ser a pioneira e assim apresentar à comunidade científica a tecnologia que a viabilizou: a *blockchain*; e a Ethereum por introduzir nesse universo o conceito de *smart contracts* (contratos inteligentes) e com isso acrescer automatização ao processo de transferência de ativos digitais. Um terceiro marco pode ser evidenciado pela utilização da DLT em um contexto para além do financeiro das criptomoedas, aplicando-se a áreas como IoT, cidades inteligentes, identidades digitais etc. (GREVE et al., 2018).

Fazem parte deste terceiro marco as implementações da DLT como o Hyperledger Fabric. Este grupo caracteriza-se pela flexibilização de algumas características das criptomoedas, como o controle da entrada de membros à rede. Essas decisões foram tomadas de modo a atender a algumas necessidades da indústria, como prover garantias de: escalabilidade, privacidade, identificação e controle da entrada de membros etc. (GREVE et al., 2018).

2.2.2. Propriedades

Com base em (GREVE et al., 2018) são apresentadas a seguir algumas propriedades da DLT que se julga mais importantes de acordo com o contexto do trabalho em questão.

1) Descentralização - A tecnologia permite com que nós distintos confiem mutuamente entre si, sem que haja a necessidade de uma entidade central garantidora das transações. Todas as transações estão presentes de forma consistente em todos os nós da rede.
2) Transparência - Todas as transações registradas no *ledger* são do conhecimento de todos os nós da rede e podem ser auditadas.
3) Imutabilidade - As transações registradas no livro de registro (*ledger*) são imutáveis e não podem ser refutadas.

2.2.3. DLT permissionadas e não permissionadas

Tanto o Bitcoin como o Ethereum são *blockchains* públicas - também chamadas de não permissionadas, projetadas para atuar em escala global, pela Internet, sem o controle da entrada de membros. Deste modo, o conjunto de nós da rede P2P é desconhecido. Sua composição é dinâmica pois é permitida a entrada e saída de membros aleatoriamente. Como esses nós não precisam de identificação, costumam ser anônimos. Não há confiança entre os nós (GREVE et al., 2018).

No cenário de utilização da *blockchain* de forma a atender às necessidades da indústria, surgiram as *blockchains* privadas ou federadas, também chamadas de permissionadas, com o controle da entrada e saída de membros. Nesta categoria estão, por exemplo, o Hyperledger Fabric e o Multichain – utilizado na prova de conceito do modelo de (ONIK 2019). Neste caso, a composição de membros da rede é conhecida, os nós são identificados, autenticados e autorizados (GREVE et al., 2018).

2.3. Armazenamento de dados *off-chain*

O modelo apresentado em Onik et al. (ONIK 2019) tem como um de seus pilares o armazenamento dos dados pessoais fora do *ledger*, também chamado de fora da cadeia, ou simplesmente de armazenamento *off-chain*. O objetivo desse tipo de armazenamento é contornar a imutabilidade da cadeia.

Em linhas gerais, no contexto do tratamento de dados pessoais, o mecanismo de armazenamento *off-chain* funciona da seguinte forma: Há a separação dos dados pessoais dos dados não pessoais, sendo os pessoais armazenados fora da cadeia, enquanto os não pessoais - e um *hash* dos dados pessoais - são armazenados na cadeia.

2.4. Compromisso criptográfico

Um compromisso criptográfico (também chamado de envelope criptográfico) é um mecanismo que permite "congelar" dados de maneira que seja possível - com informações adicionais - provar o que foi congelado. Além disso, é impossível encontrar ou reconhecer esses dados usando esse único *commitment* (CNIL, 2018).

Um compromisso criptográfico pode ser entendido como o estabelecimento de um contrato entre partes, onde a mensagem original (texto plano) seria colocada em um envelope e lacrada. A ideia é o "comprometimento digital" com a mensagem original, que será ocultada até que seja necessário revelá-la. Uma vez firmado o compromisso (ou seja, o contrato), ele não pode ser alterado. Um *nonce* é utilizado para cada novo contrato, lembrando que *nonce* é um número aleatório que pode ser utilizado somente uma vez (n-*number*, *once*-uma vez). Posteriormente, o conteúdo do envelope pode ser revelado e verificado (GREVE et al., 2018, p.5-7).

Um bom exemplo da aplicação de compromissos criptográficos pode ser encontrado em (REYNERI; KARNIN, 1984). O artigo cujo título é "*Coin Flipping by Telephone*" descreve uma hipotética situação em que é necessário prover confiança em um jogo de cara ou coroa que tem que ser jogado por telefone**.** Em linhas gerais, o exemplo descreve que duas pessoas, Alice e Bob, querem resolver uma disputa com um "cara ou coroa". Presencialmente, a disputa seria facilmente resolvida, no entanto, remotamente, a situação requer que haja algum mecanismo que acrescente confiança ao processo, eliminando a necessidade de se basear na confiança e boa-fé das pessoas. Nesse mecanismo confiável:

1. Alice escolhe um dos lados da moeda, escreve em um pedaço de papel, coloca-o no envelope, assina, lacra e envia para Bob, que deve mantê-lo lacrado. Estabelece-se assim um compromisso com a jogada corrente e assim pode-se lançar a moeda;
2. Alice dá o comando para Bob lançar a moeda;
3. Bob joga a moeda e relata o resultado;
4. Bob abre o envelope onde está anotada a escolha de Alice;
5. Se a escolha de Alice corresponder ao resultado que Bob relatou, Alice ganha. Do contrário Bob é o vencedor.

Merece nota que Bob lança a moeda sem saber a escolha de Alice, ou seja, sem abrir o envelope. Depois de lançada a moeda, com o resultado já sabido, revela-se a escolha de Alice. Bob

não sabe qual a resposta e por isso não poderá trapacear, nem Alice poderá alterar a sua escolha. Com a retirada do lacre do envelope, encerra-se a jogada atual. A retirada do lacre equivalente à utilização do *nonce* citado anteriormente.

Fazendo-se uma analogia da situação narrada acima com o domínio desta pesquisa, o ato de cifrar os dados pessoais pode ser equiparado a uma jogada efetuada entre Alice e Bob. A cada jogada, assim como a cada nova encriptação, estabelece-se um compromisso entre as partes. A função de encriptação corresponde ao ato de lacrar o envelope. A função de verificação (vide Seção 5.4.2.1) equivale à abertura do envelope para verificar o seu conteúdo. A aleatoriedade ao processo é inserida com a adição de um *salt* à mensagem a ser cifrada (vide Seção 5.4.2.1), assim como no jogo entre Alice e Bob a retirada do lacre equivale à criação do *nonce*.

3. TRABALHOS RELACIONADOS

O problema de pesquisa em estudo neste trabalho é como compatibilizar o tratamento de dados pessoais realizado por aplicações baseadas em DLT com os direitos ao esquecimento e à retificação da LGPD. O problema envolve, portanto, como lidar com a característica intrínseca de imutabilidade das DTL frente ao direito ao esquecimento e à retificação do titular dos dados pessoais.

Durante o processo de investigação, um conjunto de trabalhos científicos foi localizado e selecionado usando um protocolo de revisão da literatura, o qual está detalhado no Capítulo 4.

Dois dos trabalhos identificados são explícitos na propositura do que fazer para realizar essa compatibilização: (1) o trabalho de Onik et al. (ONIK 2019) e (2) o relatório CNIL (CNIL, 2018) e serão detalhados a seguir.

O trabalho de Politou et al. (POLITOU 2019) apresenta o estado da arte quanto à essa compatibilização. A Figura 1 sumariza quais técnicas podem ser utilizadas para tal. Dentre as categorias técnicas apresentadas, a saber: (a) uso de *blockchains* editáveis (*Redactable blockchains*), (b) poda (*Pruning*) da rede *blockchain*, (c) encriptação (*Encrypted in blockchain*), e (d) armazenamento *off-chain* (*Off-chain storage*), é possível situar o trabalho aqui descrito no conjunto daqueles que se utilizam da técnica de armazenamento *off-chain* para viabilizar a compatibilização entre imutabilidade e direito ao esquecimento e à retificação.

Figura 1 - Visão geral das soluções para balanceamento entre imutabilidade e RTBF

Fonte: Adaptado de (POLITOU et al., 2019, p.7)

Onik et al. (ONIK 2019) oferece um modelo teórico ante ao problema de pesquisa no contexto do GDPR. O relatório CNIL (CNIL, 2018) é complementar ao trabalho de Onik (ONIK

2019), na medida em que, embora não ofereça uma solução direta ao problema, insere pontos de atenção que devem ser seguidos de modo a compatibilizar o tratamento de dados pessoais em aplicações baseadas em DLT com os requisitos do GDPR. Ademais, o relatório CNIL (CNIL, 2018) destaca que o uso de um compromisso criptográfico dos dados pessoais é preferível ao uso de *hash*. Onik et al. (ONIK 2019) não trata de compromisso criptográfico, já que o modelo apresentado faz uso de *hash*.

Algumas características do modelo proposto em Onik et al. (ONIK 2019):

- é baseado na separação dos dados compartilhados pelo Titular em dados pessoais e dados não pessoais (vide definição adiante). O armazenamento dos dados pessoais é realizado fora da cadeia (*off-chain*), por outro lado o *hash* desses dados pessoais e os dados não pessoais são armazenados na cadeia.
- descreve em pormenores o bloco criado a cada compartilhamento de dados pessoais bem-sucedido entre o Titular e o Controlador, e entre o Controlador e o Operador. Este bloco guarda importantes informações sobre a operação de compartilhamento, se constituindo em um repositório seguro e imutável do registro de informações (*log*) dessas operações. Algumas informações armazenadas no bloco são:
    a) Identificação dos membros da rede que participaram da operação de compartilhamento de dados pessoais: Titular, Controlador e Operador;
    b) *Timestamp* marcando quando ocorreu a operação;
    c) Contador de transações (número de vezes em que ocorreu o compartilhamento de dados entre os membros da rede: Titular, Controlador e Operador);
    d) Termos do compartilhamento (implementados no *smart contract* e que os membros da rede devem aceitar para que o compartilhamento ocorra): engloba termos de privacidade, regulamentação, usabilidade, regras de distribuição, processo de notificação de violação e consenso;

O trabalho de Onik et al. (ONIK 2019) introduz os termos PII, PPII e NPII, que representam conceitos importantes para tratamento de dados pessoais:

a) PII (*Personally Identifiable Information*): subconjunto de atributos do indivíduo que permite a sua identificação inequívoca perante a coletividade, tais como: nome completo, e-mail, número de identidade, dados biométricos etc.
b) PPII (*Potential Personally Identifiable Information*): também conhecidos como quase-identificadores, ou identificadores indiretos. São dados que combinados podem permitir a identificação inequívoca do indivíduo, tais como: parte do nome, parte do endereço, raça,

religião etc.

c) NPII (*Non-personally Identifiable Information*) que engloba as informações não pessoais, portanto, aquelas que não permitem a identificação de um indivíduo.

Onik et al. (ONIK 2019) reforça a importância de não armazenar os PPIIs (quase-identificadores) na cadeia, portanto de forma imutável, pois estes dados combinados podem levar à identificação do indivíduo e a impossibilidade da exclusão desses dados não permitiria a garantia dos direitos ao esquecimento e à retificação. Além disso, sugere o uso de *smart contract* para possibilitar que o Titular tenha conhecimento de todo o tratamento realizado com os seus dados pessoais. O Titular deve inclusive consentir que este tratamento seja realizado.

O relatório CNIL (CNIL, 2018) é uma orientação técnica emitida pela Autoridade Francesa de Proteção de Dados (CNIL) frente ao crescente número de aplicações baseadas em *blockchain* e, consequentemente, do aumento do tratamento de dados pessoais realizado por essas aplicações.

Esse documento de orientação técnica aponta alguns cuidados que se deve tomar quando do tratamento de dados pessoais realizado por aplicações baseadas em *blockchain* e ressalta a dificuldade de responsabilização em caso de incidentes de quebra de privacidade dos dados do Titular quando da utilização de *blockchains* públicas.

Em suma, o trabalho de Onik et al. (ONIK 2019) traz importante contribuição para o referencial teórico desta pesquisa, e que é corroborada pelo CNIL (CNIL, 2018): o uso do armazenamento *off-chain*. O relatório CNIL ainda acrescenta que o uso do compromisso criptográfico dos dados pessoais é preferível ao uso de *hash*. Tal orientação tem como objetivo minimizar os riscos de reversão e vinculação do dado pessoal submetido à função de *hash.*

3.1. Diferenciais desta pesquisa

Um dos diferenciais desta pesquisa está em trazer a discussão especificamente para o contexto da LGPD, diferente das análises realizadas por Onik et al. (ONIK 2019) e (CNIL, 2018) que estão sob o escopo do GDPR, além de avançar na análise trazendo para o arcabouço teórico contribuições como os trabalhos de (FINCK, 2019), (GREVE et al., 2018) entre outros. Além disso, este trabalho avança ao especificar formalmente um conjunto de funções, operações e transações (vide Seção 5.4), especificar uma API REST para oferta de serviços e uma implementação de referência para uso prático no desenvolvimento de aplicações.

A análise desses e de outros materiais citados ao longo da pesquisa resultou na propositura de um conjunto de boas práticas, indicando a adequabilidade da utilização combinada de armazenamento *off-chain* e compromisso criptográfico para enfrentamento ao problema de pesquisa.

Por fim, outra contribuição é a análise de como se daria o tratamento de dados considerando uma aplicação conceitual que se utiliza do *framework*. A utilidade está não apenas em validar a aplicabilidade do *framework* como ferramenta de enfrentamento ao problema de pesquisa, mas também em documentar como o *framework* pode ser usado, deixando uma referência de uso para eventuais aplicações baseadas em DLT que visem estar aderentes à LGPD no que se refere ao exercício do direito ao esquecimento e à retificação.

4. METODOLOGIA

Nesta seção será apresentada a metodologia utilizada para investigar o problema de pesquisa. Essa metodologia aplica técnicas baseadas no conceito de *Design Science Research* (DSR).

A DSR tem como um de seus principais pilares a propositura de artefatos, gerados como resultado-alvo a partir do processo de investigação científica frente a um problema de pesquisa real e relevante, como este que aqui se investiga. Estes artefatos devem ser submetidos à avaliação, validados e aprimorados, servindo como materialização do conhecimento científico, de forma que possam ser utilizados como ferramenta de enfrentamento à classe de problemas de pesquisa a qual se está investigando (HORITA; NETO; SANTOS, 2015).

A decisão pela escolha dessa metodologia foi construída a partir da análise de (ELSHEKEIL; LAOYOOKHONG, 2017). Este documento tem como problema de pesquisa adequar um sistema de informação aos requisitos do GDPR. Utiliza-se da DSR como metodologia e propõe como artefato um *framework* que deve ser seguido de modo a garantir-se a conformidade com o GDPR.

Tal problema de pesquisa assemelha-se ao aqui estudado, na medida em que, trata de uma categoria específica de sistemas de informação (aplicações baseadas em DLT) e trata da conformidade com um regulamento específico de proteção de dados (a LGPD no caso aqui em estudo).

O **Referencial teórico** surge no processo de pesquisa dos meios metodológicos empregados para atingir os objetivos específicos de um problema de pesquisa. É a estrutura que apoia um estudo científico, consiste de conceitos, suas definições e teorias, que são usados para definir, estudar e resolver o problema em questão (UNIVERSITY, 2020).

A Figura 2 mostra a estrutura da metodologia da pesquisa aqui empregada. Ilustra como a partir do problema de pesquisa, utilizando-se de ferramentas e processos, chega-se ao **Referencial teórico** que serve de base ao *framework* proposto como solução.

Por fim, como forma de explicitar o fio condutor da pesquisa, apresenta-se abaixo a relação entre os objetivos específicos descritos na Seção 1.2.2 e as ferramentas e processos da metodologia. Não estão apresentados os objetivos relacionados à implementação e validação da solução, que serão tratados no Capítulos 6.

- Para **elicitar as exigências legais e regulatórias** impostas pela LGPD no que se refere aos direitos ao esquecimento e à retificação para titulares de dados pessoais foram utilizadas as ferramentas [F03] - Texto da LGPD com o processo “analisar legislação” e a ferramenta [F04] - Junção de [F01], [F02] e [F03] com o processo “análise conjunta”. Os resultados estão presentes ao longo do Capítulo 5 (5.1 - onde estão presentes definições acerca dos dados tratados).

Em outras palavras, esse objetivo foi alcançado a partir da análise do texto da LGPD e dos documentos de orientação técnica. Os resultados estão presentes ao longo do Capítulo 5, onde estão esmiuçados conceitos como a classificação dos dados pessoais em PI, PPII e NPII (5.1).

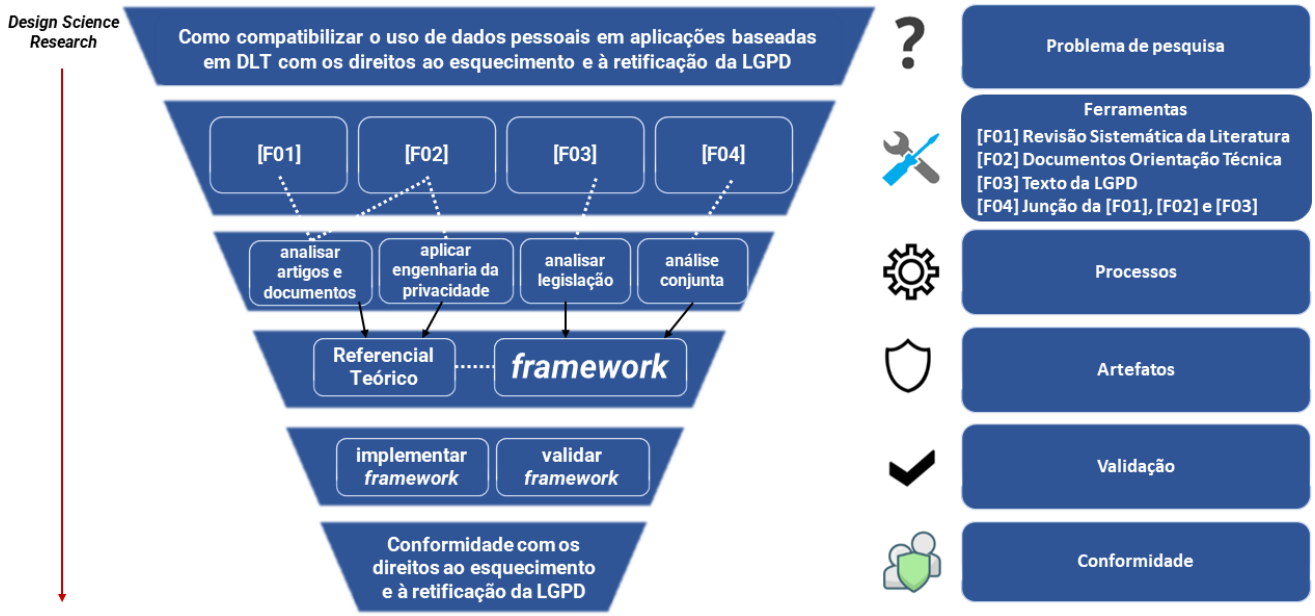
- Para **identificar os problemas comuns enfrentados por aplicações baseadas em DLT ao tratar dados pessoais** considerando os direitos ao esquecimento e à retificação foram utilizadas as ferramentas [F01] - Revisão Sistemática da Literatura e [F02] - Documentos de Orientação Técnica com o processo "analisar artigos e documentos". Os resultados estão presentes no Capítulo 5.

Em outros termos, o principal problema comum a aplicações DLT quanto ao tratamento de dados pessoais é fazê-lo em conformidade com a LGPD (p.ex. direito ao esquecimento e à retificação), considerando a imutabilidade da cadeia. Esse conhecimento foi obtido a partir da análise dos artigos na Revisão Sistemática e dos documentos de orientação técnica. Os resultados permeiam todo o Capítulo 5, pois norteiam a arquitetura, os serviços e recursos da solução proposta.

- Para **localizar na literatura científica mecanismos que permitam mitigar, contornar ou resolver os problemas identificados** foi utilizada a ferramenta [F01] - Revisão Sistemática da Literatura com o processo "analisar artigos e documentos". Os resultados são apresentados no Capítulo 5, pois esta análise resulta na propositura do Referencial Teórico.
- Para compilar um **conjunto de boas práticas propostas para desenvolvimento de aplicações compatíveis com LGPD** foram utilizadas as ferramentas [F01] - Revisão Sistemática da Literatura e [F02] - Documentos de Orientação Técnica com o processo "analisar artigos e documentos". Os resultados são apresentados no Capítulo 5, pois as principais boas práticas estão no Referencial Teórico.
- Para **especificar um *framework*** que **mapeie as boas práticas compiladas com o objetivo de facilitar a construção de aplicações compatíveis com esses direitos** foram utilizadas as ferramentas [F01], [F02] e [F03] com o processo "análise conjunta dos artigos, documentos e legislação". Os resultados estão presentes ao longo do Capítulo 5.

No restante desta seção, será delimitado o escopo da pesquisa e serão apresentadas as ferramentas utilizadas para enfrentamento ao problema. Serão também apresentados os processos executados (derivados dessas ferramentas). Estes processos servirão de base para construção do artefato proposto como solução ao problema de pesquisa e que será validado em sequência (vide capítulos 5 e 6). A conexão entre as ferramentas, processos, artefatos e mecanismo de validação está ilustrada na Figura 2 abaixo.

Figura 2 - Estrutura da Metodologia da Pesquisa



Fonte: O Autor

4.1. Delimitação do Escopo da Proposta

O problema alvo de estudo desta pesquisa é a compatibilização dos direitos ao esquecimento e à retificação da LGPD no tratamento de dados pessoais por aplicações baseadas em DLT, considerando-se a imutabilidade intrínseca desta tecnologia.

Um ponto que deve ser destacado como não escopo é que o objetivo dessa pesquisa não é a utilização da DLT como barramento (*backbone*) para o compartilhamento de dados pessoais em conformidade com as normas. Embora algumas características da tecnologia DLT, como transparência e segurança, contribuam para a conformidade de um sistema de *software* com a LGPD, pois estes itens são princípios onde a lei está fundamentada (vide Tabela 2).

Delimitado o escopo, segue-se com a apresentação das ferramentas, processos e artefato produzido baseados na metodologia DSR, ante ao problema de pesquisa. Neste ponto, é feita uma análise mais detalhada das ferramentas e processos. Nas seções seguintes, o foco estará sobre a proposta de um artefato como solução para o problema e a sua validação.

4.2. Revisão Sistemática da Literatura

A primeira das ferramentas utilizadas na investigação do problema de pesquisa é e Revisão Sistemática da Literatura – RSL.

A RSL foi realizada tendo como base o guia disponibilizado em (OKOLI; DUARTE; MATTAR, 2019). Este guia apresenta oito passos para realização da RSL que foram seguidos adequando-os ao escopo desta pesquisa como se segue:

1 – Identifique o objetivo

Objetivo: Buscar artigos científicos que proponham - de forma direta e não apenas abordando a questão - alternativas de solução ao problema de pesquisa, ou seja, como garantir os direitos ao esquecimento e à retificação previstos na LGPD em aplicações baseadas em DLT que realizem tratamento de dados pessoais.

Identificado o objetivo, deriva-se deste a questão de pesquisa. Chamada de Q1.

1.1 – Elaborar a questão de pesquisa

Q1: Como garantir os direitos ao esquecimento e à retificação em aplicações baseadas em DLT que realizam tratamento de dados pessoais?

Observação: Data de realização da busca: 01 de outubro de 2020.

2 – Estabelecer o protocolo de busca

A seguinte *string* de busca da Q1 foi utilizada:

((DLT OR *blockchain*) AND (GDPR OR LGPD) AND (RTBF OR *immutability* OR "*right to be forgotten*"))

3 - Definir os critérios de Inclusão

Os seguintes critérios de **inclusão (I)** de artigos foram utilizados:

- **(I)** O ano de publicação dos artigos devem estar no intervalo entre 2019 e 2021;
    - Identificou-se que alguns motores de busca já listam artigos com o ano de publicação 2021. Por essa razão, este ano foi incluído nos critérios de inclusão de artigos;
    - O GDPR entrou em vigor em 25 de maio de 2018. Desta forma, incluiu-se o ano seguinte (2019) à entrada em vigor do regulamento nos critérios de inclusão de artigos;
    - A LGPD entrou em vigor em 18 de setembro de 2020. O ano de 2020 está no intervalo dos critérios de inclusão de artigos;

- **(I)** Os artigos devem estar escritos em português ou inglês;
- **(I)** Os artigos selecionados devem atender à *string* de busca seja no Título, Palavras-Chave (*keywords*), Resumo (*abstract*) ou Conteúdo.

4 – Realizar a Busca na Literatura

Seguem-se os mecanismos de busca utilizados e a quantidade de publicações que atenderam aos critérios de pesquisa em cada um deles, como mostrado na Tabela 1.

Tabela 1 - Quantidade de artigos obtidos por veículo de publicação

Veículo de Publicação ou Portal	Quantidade de Artigos
ACM Digital Library[2]	20
Science Direct[3] Adicionalmente foram selecionados os critérios "*Review articles*" e "*Research articles*", critérios específicos a este mecanismo de busca.	49
IEEE Xplore[4]	10
Google Acadêmico[5] Desse total, 20 artigos foram pré-selecionados baseados na análise do Título da publicação, de forma a verificar se este atendia ao objetivo da pesquisa.	1280
Periódicos Capes[6]	39
Inseridos manualmente	2

Ao final desta fase foram pré-selecionados 140 artigos. A Figura 3 sumariza como estes artigos estão distribuídos considerando-se cada uma das fontes de pesquisa utilizadas.

2 dl.acm.org

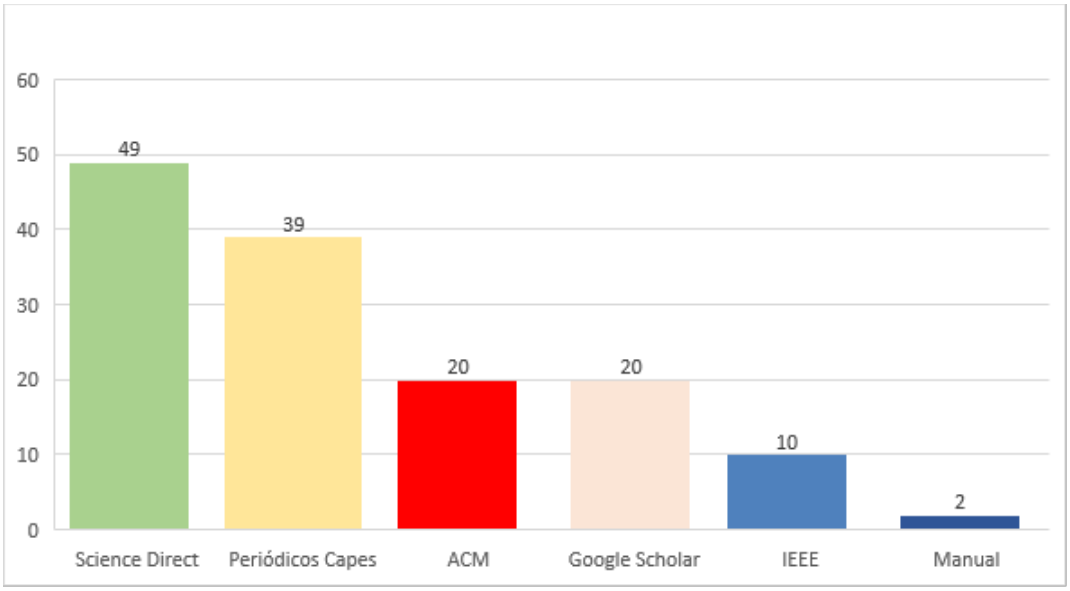
3 sciencedirect.com

4 ieeexplore.ieee.org

5 scholar.google.com

6 periodicos.capes.gov.br

Figura 3 - Distribuição de artigos selecionados por Fonte de Pesquisa



Fonte: O Autor

5 – Extração dos dados

Os artigos selecionados foram exportados em formato BibTex para a ferramenta StArt (*State of the Art through Systematic Review*), desenvolvida pelo Laboratório de Pesquisa em Engenharia de *Software* (LaPES), do Departamento de Ciência da Computação da Universidade Federal de São Carlos (UFSCar) (LAPES, 2016)**.**

6 - Avaliar a qualidade metodológica

O seguinte critério de **inclusão (I)** de artigos foi considerado:

- **(I)** O artigo deve propor uma alternativa direta de solução ao problema
    - Isto é, artigos que apenas abordam a questão sem oferecer uma alternativa direta de solução ao problema de pesquisa não foram selecionados;

Os seguintes critérios de **exclusão (E)** foram considerados – baseados na análise de (POLITOU 2019):

- **(E)** O artigo propõe *blockchains* editáveis (*Redactable blockchains*) como alternativa de solução ao problema de pesquisa;
- **(E)** O artigo propõe a técnica de poda (*pruning*) como alternativa de solução ao problema de pesquisa;
- **(E)** O artigo propõe a técnica de encriptação de dados *on-chain* (*encripted in blockchain*) como alternativa de solução ao problema de pesquisa.

Submetendo-se os 140 artigos pré-selecionados na fase **4 - Busca da bibliografia** aos critérios de inclusão e exclusão acima mencionados, restaram 13 artigos.

7 – Sintetização dos resultados

Essa fase consiste em analisar novamente os artigos segundo critérios qualitativos. Analisou-se o resumo dos artigos de forma mais acurada tendo em vista os objetivos de pesquisa e chegou-se à lista de 02 artigos abaixo elencados:

- *Privacy-aware blockchain for personal data sharing and tracking*; (ONIK et al., 2019)
- BPDIMS: *A Blockchain-based Personal Data and Identity Management System*; (FABER et al., 2019)

8 – Documentar a RSL

Realizou-se a revisão sistemática da literatura acadêmica de acordo com o escopo da pesquisa, chegando-se a dois documentos que oferecem uma alternativa direta de solução ao problema: Onik et al. (ONIK 2019) e FABER (FABER et al., 2019).

Onik et al. (ONIK 2019) é um dos trabalhos que serviram de base à pesquisa. Suas principais contribuições, além dos diferenciais quanto a este trabalho são apresentados de forma mais detalhada no Capítulo 3 - Trabalhos Relacionados.

FABER (FABER et al., 2019) não está no rol de trabalhos que servem de base à pesquisa pois:

> O escopo difere do problema de pesquisa aqui investigado, visto que oferece um modelo genérico – baseado em *blockchain* – para o gerenciamento de dados e identidades. O problema de pesquisa aqui tratado refere-se a aplicações baseadas em *blockchain* em conformidade com os direitos ao esquecimento e à retificação. Diferente de utilizar *blockchain* como um barramento para o tráfego de dados pessoais em conformidade com o GDPR, como a pesquisa de (FABER et al., 2019) propõe;

A análise de Onik et al. (ONIK 2019) contribuiu para a propositura do Referencial Teórico desta pesquisa.

4.3. Documentos de Orientação Técnica

Outra ferramenta utilizada na investigação do problema de pesquisa é a Análise dos Documentos de Orientação Técnica emitidos por algumas Autoridades Nacionais de Proteção de Dados Pessoais, com destaque para o Relatório CNIL e o Guia emitido pela Agência Espanhola de Proteção de Dados (AEPD - *Agencia Española Protección Datos*).

Na Seção de Trabalhos Relacionados, o Relatório CNIL (CNIL, 2018) foi apresentado como um dos trabalhos em que esta pesquisa se baseia, justamente por conter orientações técnicas

específicas acerca de cuidados que devem ser tomados quando do tratamento de dados pessoais realizado por aplicações baseadas em *blockchain*. A análise desse relatório contribuiu para o Referencial Teórico, principalmente quanto ao uso de compromissos criptográficos.

Ampliando um pouco mais o campo de análise, não restringindo-o a aplicações baseadas em DLT como no caso do Relatório CNIL (CNIL, 2018), mas a sistemas de informação de um modo geral, chega-se ao guia emitido pela Agência Espanhola de Proteção de Dados (AEPD - *Agencia Española Protección Datos*). Este guia contribui para a garantia da privacidade de forma geral, englobando inclusive os direitos ao esquecimento e à retificação. Esse documento baseia-se na Engenharia da Privacidade — processo sistemático orientado a riscos que objetiva traduzir em termos práticos e operacionais os princípios de *Privacy By Design* (PBD) (SPANISH DATA PROTECTION AGENCY, 2019).

No GDPR, os princípios de PBD encontram-se explicitamente presentes no artigo 25. Estão também relacionados aos itens do artigo 5° (vide Tabela 2). Na LGPD é possível evidenciá-los, embora de maneira não explícita.

Tabela 2 - Princípios de *Privacy by Design* (PBD) versus Artigo 5º do GDPR

#	Privacy by Design	Artigo 5º GDPR
1	Proativo, não reativo. Prevenir, não remediar	Tratamento lícito, leal e transparente
2	Privacidade como configuração padrão	Propósito específico
3	Privacidade embutida no *design*	Minimização dos dados
4	Funcionalidade total - soma positiva, não soma zero	Limitação do armazenamento (não mais do que o necessário)
5	Segurança de ponta a ponta - Total Proteção do ciclo de vida	Integridade e confidencialidade (perda não autorizada, ilegal, acidental, destruição, dano, medidas técnicas ou organizacionais)
6	Visibilidade e transparência - mantenha aberto	Acurácia (exato, atualizado, apagado ou retificado)
7	Respeito pela privacidade do usuário - mantenha centrado no usuário	Prestação de contas (responsabilização + facilidade de localização da informação (*findability*) + demonstrar conformidade)

No Guia espanhol (SPANISH DATA PROTECTION AGENCY, 2019), a partir desses 07 princípios do PBD derivam-se metas, estratégias e, no nível mais baixo, técnicas (tecnologias) – também chamadas PETS (*Privacy Enhancing Technologies*) - que são usadas para implementar padrões de *design* de privacidade concretamente por meio de uma dessas tecnologias, de forma isolada ou em conjunto.

Tabela 3 - Metas *versus* Estratégias de Privacidade

Metas de Privacidade	Estratégias	
	Orientadas a dados	**Orientadas a processos**
Desvinculação	MINIMIZE, ABSTRAIA, SEGREGUE, OCULTE	
Controle		CONTROLE, IMPONHA, DEMONSTRE
Transparência		INFORME

Fonte: (SPANISH DATA PROTECTION AGENCY, 2019, p.17)

Essas oito estratégias estão divididas em dois grupos: Estratégias orientadas a dados e Estratégias orientadas a processos. Um sumário dessas estratégias é descrito na Tabela 4:

Tabela 4 - Estratégias de privacidade

#	Orientação a	Estratégias	Descrição
1	DADOS	MINIMIZE	Limite ao máximo possível os dados necessários ao tratamento
2		OCULTE	Proteja os dados pessoais, ou torne-os desvinculados ou indistinguíveis. Garanta que não sejam disseminados publicamente
3		SEGREGUE	Segregue o máximo possível os dados pessoais necessários ao tratamento
4		ABSTRAIA	Limite ao máximo possível o detalhamento de quais dados pessoais estão sendo tratados
5	PROCESSOS	INFORME	Informe ao Titular dos dados pessoais sobre o tratamento dos seus dados de forma contínua e apropriada
6		CONTROLE	Forneça ao Titular o controle adequado sobre o tratamento de seus dados pessoais
7		IMPONHA	Comprometa-se a tratar os dados pessoais de forma a facilitar a privacidade (*privacy-friendly*) e imponha de forma adequada à organização
8		DEMONSTRE	Demonstre que o tratamento dos dados pessoais é realizado de forma a facilitar a privacidade (*privacy-friendly*)

Por sua vez, essas estratégias são associadas a técnicas cuja aplicabilidade deve ser verificada diante da situação real de privacidade a que se esteja sendo submetido. A Tabela 5 apresenta as técnicas associadas a cada estratégia.

Tabela 5 - Estratégias de privacidade *versus* PETS

#	Orientação a	Estratégias	Técnicas ou Padrões (PETS)
1	DADOS	MINIMIZE	Supressão de Dados Anonimização e **Pseudoanonimização**
2		OCULTE	Supressão de Dados Anonimização e **Pseudoanonimização** Criptografia/**Encriptação** Autenticação
3		SEGREGUE	Segregação Física e Lógica dos Dados
4		ABSTRAIA	Agregação de Dados Generalização de Dados Perturbação de Dados *K-anonimity*
5	PROCESSOS	INFORME	Política de Privacidade Termos de Uso e Cookies Dashboard de Privacidade
6		CONTROLE	Dashboard de Privacidade
7		IMPONHA	Política de Privacidade Termos de Uso e Cookies Controle de Acesso
8		DEMONSTRE	Logs de Auditoria

Fonte: (SPANISH DATA PROTECTION AGENCY, 2019, p.24)

A análise dos conceitos de Engenharia da Privacidade e PBD introduzidos pelo guia emitido pela AEPD permite afirmar que:

- o Relatório CNIL (CNIL, 2018), um documento que trata especificamente da conformidade ao GDPR de aplicações baseadas em *blockchain*, utiliza-se dos conceitos de PBD e consequentemente da aplicação dos conceitos de Engenharia da Privacidade para tecer suas recomendações;
- a aplicação dos conceitos de Engenharia da Privacidade em aplicações baseadas em *blockchain* que realizem tratamento de dados pessoais não prescinde a utilização dos itens que compõem o Referencial Teórico, a saber: armazenamento *off-chain,* armazenamento *on-chain* preferencialmente sob a forma de compromisso criptográfico;
- As estratégias e PETS decorrentes da utilização da sistemática da Engenharia da Privacidade estão contemplados no Referencial Teórico – vide Tabela 6 a seguir.

Tabela 6 - PETS *versus* Referencial Teórico

Estratégias	Técnicas ou Padrões (PETS)	Aplicabilidade à solução
MINIMIZE	Supressão de Dados Anonimização e Pseudoanonimização	Anonimização de dados através da técnica de compromisso criptográfico
OCULTE	Supressão de Dados Anonimização e Pseudoanonimização Criptografia/Encriptação Autenticação	Anonimização de dados através da técnica de compromisso criptográfico.
SEGREGUE	Segregação Física e Lógica dos Dados	Separação dos dados em PII, PPII e NPII.
ABSTRAIA	Agregação de Dados Generalização de Dados Perturbação de Dados K-Anonimity	Obs. A depender do domínio da aplicação alguma dessas técnicas de abstração de dados podem se mostrar adequadas.
INFORME	Política de Privacidade Termos de Uso e Cookies Dashboard de Privacidade	A indexação dos dados pessoais no banco de dados do *framework* – vide seções 5.4.2.3 e 5.4.2.4 - permite que seja implementado o informe ao Titular de quaisquer operações efetuadas com os seus dados pessoais.
CONTROLE	Dashboard de Privacidade	A indexação dos dados pessoais no banco de dados do *framework* – vide seções 5.4.2.3 e 5.4.2.4 - permite que seja implementado o informe ao Titular de quaisquer operações efetuadas com os seus dados pessoais.
IMPONHA	Política de Privacidade Termos de Uso e Cookies Controle de Acesso	A indexação dos dados pessoais no banco de dados do *framework* – vide seções 5.4.2.3 e 5.4.2.4 - permite que seja implementado o informe ao Titular de quaisquer operações efetuadas com os seus dados pessoais.
DEMONSTRE	Logs de Auditoria	A indexação dos dados pessoais no banco de dados do *framework* – vide seções 5.4.2.3 e 5.4.2.4 - permite que seja implementado o informe ao Titular de quaisquer operações efetuadas com os seus dados pessoais.

Vale nota – conforme relatado também no Capítulo 3 - Trabalhos Relacionados – que todas as estratégias orientadas a dados (MINIMIZE, OCULTE, SEGREGUE e ABSTRAIA) estão contempladas no Referencial Teórico na medida em que:

- a decisão de não salvar os PPIIs na cadeia tem como objetivo minimizar os riscos de vinculação e consequente identificação do Titular por meio da combinação dos seus PPIIs;
- um dos objetivos do uso de compromissos criptográficos também visa a redução dos riscos de vinculação e consequente identificação do Titular dos dados pessoais

A descrição e a obtenção de maiores referências acerca dos PETS apresentados na Tabela 6 podem ser encontradas em (SPANISH DATA PROTECTION AGENCY, 2019).

4.4. Processos

Finalizando a seção de metodologia da pesquisa, segue-se a descrição dos processos utilizados para contextualizar o processo de consolidação do conhecimento que serviu de base à construção da Proposta de Solução a ser detalhada ao longo do Capítulo 5.

4.4.1. Analisar artigos e documentos

A análise dos artigos levantados na Revisão Sistemática (vide Seção 4.2) e dos documentos emitidos pelas Autoridades de Proteção de Dados, notadamente o CNIL, permitiu a proposição do Referencial Teórico desta pesquisa, composto pelo armazenamento de dados pessoais *off-chain* e o uso de compromisso criptográfico. Esse Referencial Teórico serve de base à construção do *framework.*

4.4.2. Aplicar engenharia da privacidade

A aplicação da engenharia da privacidade permitiu consolidar o conhecimento obtido por meio da análise dos artigos e documentos – vide Seção 4.1 - reafirmando a necessidade de aplicação do Referencial Teórico. Tal conhecimento está materializado na Tabela 6 - PETS *versus* Referencial Teórico.

4.4.3. Analisar legislação

A análise do texto da LGPD permeia toda a pesquisa e materializa-se no Capítulo 5 - Proposta de Solução.

4.4.4. Análise conjunta

Compreende uma análise crítica que tem como ferramentas toda a documentação utilizada durante a pesquisa: artigos, documentos e legislação. Também se materializa ao longo do Capítulo 5 - Proposta de Solução.

5. PROPOSTA DE SOLUÇÃO

Este capítulo apresenta o PrivacyChain, um *framework* para persistência de dados pessoais em aplicações baseadas em DLT que ofereçam ao titular os direitos ao esquecimento e à retificação da LGPD. Este *framework* é, portanto, o artefato proposto como solução ao problema de pesquisa abordado neste trabalho.

Ademais, este capítulo **discute e detalha o processo de concepção deste *framework***, desde a visão conceitual/arquitetural macro até a descrição dos recursos, isto é, dos serviços providos pelo *framework*.

Este capítulo está organizado da forma a seguir. A Seção 5.1 discorre sobre as motivações em propor este tipo de artefato para enfrentar o problema de pesquisa. A Seção 5.2 apresenta a arquitetura e visão conceitual do PrivacyChain, sua divisão em camadas e os serviços providos por cada camada, além de discutir como se dá a interação entre o *framework* e uma aplicação que dele se utilize. A Seção 5.3 discorre sobre os domínios do *framework*. A Seção 5.4 apresenta em detalhes os recursos disponíveis no *framework.*

5.1. O *Framework* PrivacyChain

Um *framework* é definido como um conjunto integrado de artefatos de *software* (como classes, objetos e componentes) que colaboram para fornecer uma arquitetura reusável para uma família específica de aplicações relacionadas SOMMERVILLE (2011). Nesta pesquisa, a principal motivação em propor um *framework* como instrumento de enfrentamento ao problema em questão é a possibilidade de **reuso** da solução. Outras motivações são: a incorporação ao *framework* de **boas práticas de programação** e de recomendações para facilitar a **conformidade com as regras impostas pela LGPD**.

Segue-se a terminologia dos dados tratados por este *framework*, importante para o entendimento da visão arquitetural do *framework* apresentada adiante.

- PII (*Personally Identifiable Information*) é a categoria que reúne o subconjunto de atributos do indivíduo que permite a sua identificação inequívoca perante a coletividade (p.ex. nome completo, e-mail, número de identidade, CPF, dados biométricos etc.);
- PPII (*Potential Personally Identifiable Information),* também conhecidos como quase identificadores, ou identificadores indiretos. São dados que combinados podem permitir a identificação inequívoca do indivíduo, tais como: parte do nome, parte do endereço, raça, religião etc.
- NPII (*Non-personally Identifiable Information)* engloba as informações não pessoais, portanto, aquelas que não permitem a identificação de um indivíduo.

Considerando que as informações com potencial de identificação pessoal (PPII) apresentam risco de identificação do indivíduo, elas serão tratadas da mesma forma que as informações de identificação pessoal (PII). Assim, para fins de simplificação será usado neste trabalho o termo PI (*Personal Information*) para referir-se a uma informação que pode ser PPII ou PII.

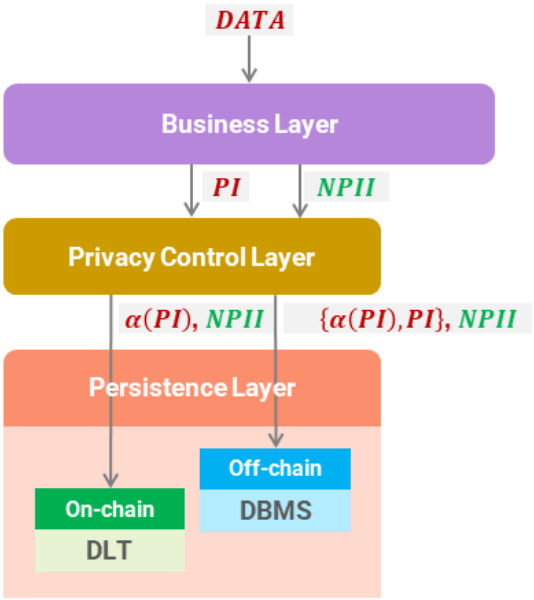
5.2. Visão Conceitual do PrivacyChain

Esta seção explica como o *framework* PrivacyChain foi idealizado em termos conceituais. Para resolver o problema de pesquisa foi idealizado que o *framework* deveria ser composto por camadas que deveriam executar determinadas funções. Esta visão conceitual balizou a criação da arquitetura da solução e está representada na Figura 4.

Conceitualmente, o *framework* PrivacyChain é composto pelas seguintes camadas: *Business Layer*, *Privacy Control Layer* e *Persistence Layer*.

A decisão em dividir o *framework em* camadas parte do princípio da separação de responsabilidades, de forma que cada uma das camadas tenha a responsabilidade de prover algum serviço ou recurso com o intuito final de possibilitar a persistência de dados pessoais por aplicações DLT respeitando-se os direitos ao esquecimento e à retificação.

Figura 4 – Visão Conceitual do *framework* PrivacyChain



Fonte: O Autor

5.2.1. Business Layer

A *Camada de Negócio* (*Business Layer - BL*) não pertence ao *framework* PrivacyChain. O serviço de **classificação dos dados** nas categorias PI e NPII provido por esta camada deve ser realizado pela aplicação que irá se utilizar do *framework* e deve partir da análise dos dados pessoais a serem tratados por essa aplicação. Os dados classificados como NPII e PI são entregues à *Camada de Controle de Privacidade* de dados pessoais para processamento.

Ressalta-se que não há impedimento para que a camada de negócio entregue dados NPII para persistência *on-chain*. Na verdade, sempre será possível acionar funcionalidades da camada de persistência para armazenar qualquer informação *on-chain*. Esses casos não trariam ameaças aos direito ao esquecimento e à retificação da LGPD, uma vez que não envolvem dados pessoais e, apesar de não poderem ser removidos (devido à imutabilidade da DLT), sempre poderão ser deixados para trás com o novo registro de eventuais alterações desses dados em outras transações na DLT, mantendo o identificador de registro da nova transação de forma *off-chain* e descartando o identificador do registro anterior.

Neste ponto é importante salientar que quanto à classificação de dados pessoais, a LGPD especifica que este é um papel dos chamados agentes de tratamento (controlador e operador) e do Encarregado de Proteção de Dados (DPO - *Data Protection Officer*). Na prática, entretanto, o que se observa é que esta tarefa é delegada à àrea de TI e à Área de Negócio que elaboram o Inventário de Dados Pessoais – IDP. Ao final do processo, o DPO revisa o IDP, assumindo um papel de consultoria quanto à classificação dos dados. Mais detalhes acerca das diretrizes apontadas pela LGPD quanto à tarefa de classificação de dados pessoais podem ser encontrados em (ME/SGD, 2021).

5.2.2. Privacy Control Layer

A *Camada de Controle de Privacidade* de dados (*Privacy Control Layer - PCL*) recebe como **entrada** os dados classificados como PI ou NPII e tem como função o provimento da lógica de persistência, de forma a não permitir que os dados pessoais sejam salvos na cadeia. Para tal, oferece os seguintes serviços:

a. Anonimização dos dados pessoais: Gera o apontador $\mathbf{PI'} = \boldsymbol{\alpha}(\boldsymbol{PI})$, que é um compromisso criptográfico dos dados pessoais (PI);

b. Controle de persistência, de forma que:

i. o apontador $\mathbf{PI'}$ seja salvo *on-chain* e *off-chain*, sendo esta a chave de ligação entre esses repositórios. Numa analogia com banco de dados relacionais, o PI' seria a chave primária de uma tabela e a chave estrangeira da tabela relacionada, constituindo-se, portanto, na ligação entre os repositórios;

ii. os dados não pessoais (NPII) sejam salvos *on-chain* e *off-chain*[7]*;*

c. Os dados pessoais (PI) sejam salvos *off-chain*.

A PCL foi pensada de forma a ser uma camada *database agnostic e DLT agnostic*, sendo, portanto, independente da base de dados e da DLT escolhida quando da fase de implementação.

5.2.3. Persistence Layer

A Camada de Persistência (*Persistence Layer – PL*) tem como entrada o **PI'**, os dados pessoais (PI) e os dados não pessoais (NPII). O serviço provido por esta camada é a persistência dos dados *on-chain* com base nas decisões efetuadas pela PCL e o controle desses registros em uma base de dados local para recuperação futura dos registros para fins de rastreamento, operações de retificação e esquecimento.

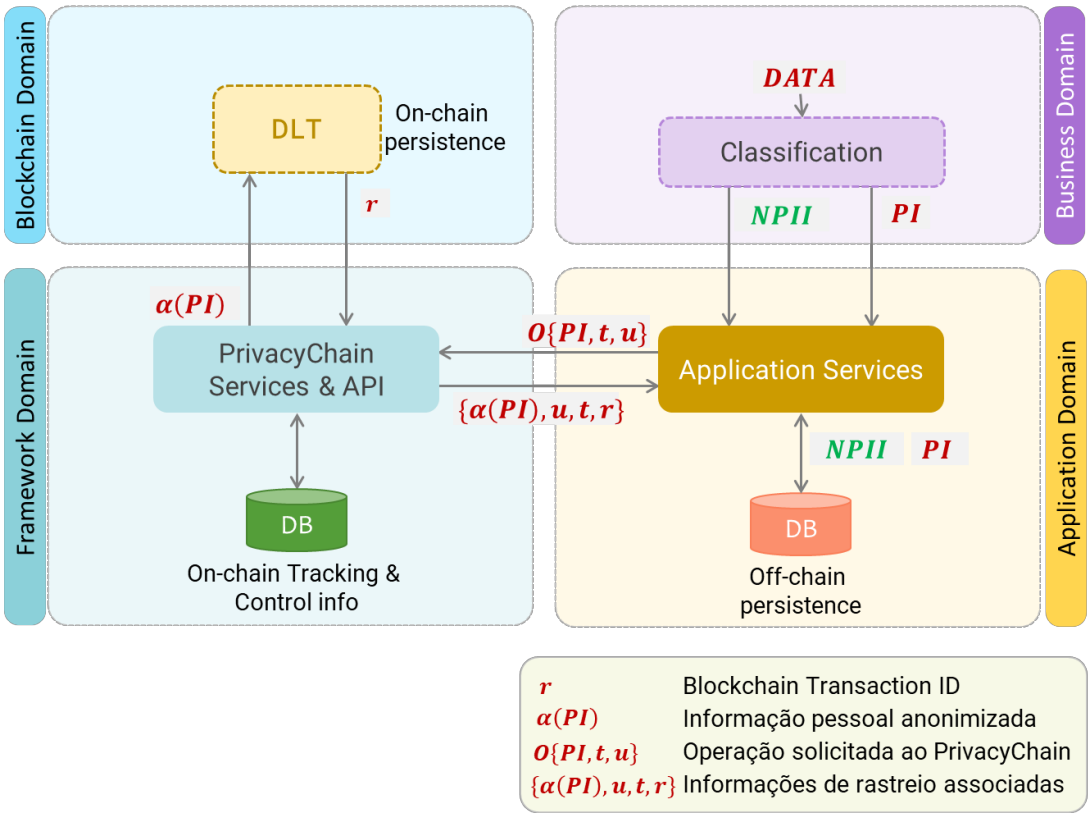
[7] Por serem dados não pessoais não há maiores implicações se estes forem persistidos dentro ou fora da cadeia.

5.3. Arquitetura do PrivacyChain

A Figura 5 ilustra a arquitetura do *framework* PrivacyChain. Nesta arquitetura estão demonstrados os domínios envolvidos em toda a cadeia de uso do *framework*. O termo "domínio" aqui empregado tem o sentido de "área de competência" ou "blocos de atribuição de responsabilidade".

A arquitetura conceitual do PrivacyChain envolve diferentes domínios: Negócio, Aplicação, *Framework* e *Blockchain*. Esses domínios são ilustrados na Figura 5 abaixo e discutidos em detalhes a seguir.

Figura 5 – Arquitetura do *framework* PrivacyChain



Fonte: O Autor

5.3.1. Domínio do Negócio

O Domínio de Negócio (*Business Domain*) relatado a seguir e ilustrado na Figura 5 tem relação direta com a *Businness Layer* presente na Figura 4.

É possível observar que as operações de classificação de dados (NPII e PI) são realizadas no contexto do negócio (*Business Domain*), o que não envolve ainda nenhuma implementação, codificação ou chamada de API.

É ainda no contexto do negócio que deve ser feita a classificação de dados que será implementada diretamente pela aplicação. Essa classificação pode ser dividida em classificação expandida ou resumida. Segue-se a descrição desses recursos.

- A ***classificação expandida*** de uma entidade E em uma tupla que contém dados $PII, PPII, NPII$ é denotada pela operação $E \rightarrow [\![PII, PPII, NPII]\!]$. O resultado desta operação é a classificação dos atributos da entidade E, de tal maneira que seja possível identificar quais deles são $PII, PPII$ ou $NPII$.
- A ***classificação resumida*** de uma entidade E é operação $D \Rightarrow [\![PI, NPII]\!]$, resultando em uma tupla que contém dados $PI, NPII$ da entidade E. O resultado é semelhante à classificação expandida (que usa na sua notação uma seta "simples" ao invés de "dupla"), mas a classificação dos atributos da entidade D fundirá atributos do tipo $PII, PPII$ em PI apenas, de tal maneira que não será possível identificar, a partir de PI, quais atributos são PII ou $PPII$.

Assume-se que estas operações são feitas manualmente uma única vez para cada entidade e que a aplicação modela seus dados adequadamente de acordo com essa classificação, de maneira a conhecer de cada entidade quais dados são PI ou NPI.

Nesta pesquisa, uma "entidade" tem significado semelhante ao utilizado na área de banco de dados. Trata-se de uma coisa, pessoa, lugar, unidade, objeto ou qualquer item sobre o qual os dados são organizados na forma dos seus atributos. Assim, um "cliente" é uma entidade que tem atributos como "nome", "endereço", "CPF" etc. Embora este trabalho foque no direito ao esquecimento e no direito à retificação, não será necessário restringir que entidades devam ser apenas pessoas. Ou seja, mantém-se a generalização da entidade para qualquer tipo de dado (pessoas, coisas etc.) sem prejuízo nas operações sobre esses dados. Ainda, o termo genérico "dados" será usado para referenciar qualquer unidade de dados quando não houver relevância sobre o fato de representar uma entidade, um atributo ou uma transformação causada por uma operação sobre eles.

5.3.2. Domínio da Aplicação

No domínio da aplicação (*Application Domain*), o que está em destaque é a persistência dos dados fora da cadeia (*off-chain*), ou seja, no banco de dados da própria aplicação e o acionamento dos serviços[8] do *framework*.

[8] $O\{PI, t, u\}$, onde t representa o tempo em que a operação ao *framework* é disparada.

5.3.3. Domínio do Framework

No domínio do *framework,* destaca-se o armazenamento de informações de rastreio e controle (*tracking and control*) na base de dados interna do *framework*, que mantém todas as referências de registros na *blockchain* solicitados através da sua API, o que permite a associação entre os dados da aplicação (*off-chain*) e os dados registrados na blockchain (*on-chain*).

5.3.4. Domínio da Blockchain

O domínio da *Blockchain* refere-se à rede *blockchain* em si, usada para a persistência dos registros *on-chain*.

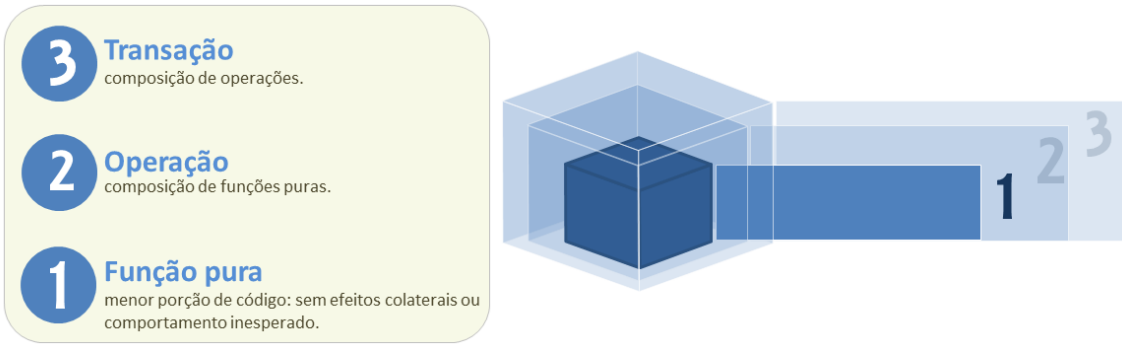
5.4. Os recursos do PrivacyChain

Os recursos oferecidos pelo *framework* PrivacyChain são classificados em: função pura, operação e transação.

O conceito de funções puras vem do paradigma de programação funcional que parte da premissa que os programas devem ser construídos como uma coleção de funções puras. Neste paradigma de programação, as funções são tratadas como o principal bloco de código e assim podem ser passadas como parâmetro para outras funções, servindo como base para que outras funções possam ser construídas como uma composição de funções puras (CUNNINGHAM, 2019; MCBRIDE, 2018).

Valendo-se do paradigma de programação funcional, no PrivacyChain, as operações são uma composição de funções puras. Por sua vez, as transações são formadas por operações, como ilustrado na Figura 6.

Figura 6 – Hierarquia dos recursos do PrivacyChain



Fonte: O Autor

A principal razão para a adoção do paradigma de programação funcional está no fato de que as funções puras são construídas de forma a não gerar efeitos colaterais, ou comportamento

inesperado. Como todo o restante do código é construído a partir de funções puras, o esperado é que o comportamento da aplicação como um todo seja também sem efeitos colaterais e sem comportamentos inesperados. Em suma, espera-se a construção de código com maior qualidade.

A Tabela 7 sumariza a classificação dos recursos do *framework* PrivacyChain em função pura, operação e transação.

Tabela 7 – Classificação dos recursos do PrivacyChain

Recurso	Função Pura	Operação	Transação
Anonimização Simples	*		
Anonimização segura (*commitment*)		*	
Verificação de anonimização segura		*	
Seleção de *Blockchain* Padrão	*		
Registro *on-chain*	*		
Registro *off-chain*	*		
Recuperação *on-chain*	*		
Indexação *on-chain*		*	
Indexação *on-chain* com anonimização		*	
Desindexação *on-chain*		*	
Desafio de registro seguro imutável		*	
Retificação *off-chain* (Direito à retificação)			*
Retificação *on-chain* (Direito à retificação)			*
Remoção *off-chain* (Direito ao Esquecimento)			*
Remoção *on-chain* (Direito ao Esquecimento)			*

5.4.1. Funções puras

No PrivacyChain os recursos classificados como funções puras são: anonimização simples, seleção de *blockchain* padrão, registro *on-chain*, registro *off-chain* e recuperação *on-chain*. Seguem-se suas descrições em detalhe.

5.4.1.1. Anonimização Simples

A ***anonimização simples*** dos dados D é a função pura $A = \alpha(D, h)$ que anonimiza D gerando um identificador não inversível A que representa D de forma anônima, usando a função de hash h. Ou seja, de posse de A não é possível na prática[9] deduzir D. A função de hash h é um parâmetro opcional e, quando omitido, assume-se a SHA-256.

[9] O termo "não é possível na prática (executar uma tarefa)" é usado para denotar que não há viabilidade computacional na execução da tarefa. Ou seja, não é possível executá-la em tempo razoável.

A anonimização simples não assume nenhuma semântica sobre a entrada D, o que significa que D pode ser uma entidade inteira (uma tupla com todos os seus atributos) ou um único atributo da entidade.

5.4.1.2. Seleção de Blockchain Padrão

A função pura $\Delta(\beta)$ faz com que o *framework* adote a *blockchain* β como ***blockchain padrão*** para operações. Esta *blockchain* será usada sempre que o parâmetro β for omitido na entrada em quaisquer operações que envolvem uma *blockchain*, como será visto adiante.

5.4.1.3. Registro on-chain

O ***registro on-chain***, denotado por $T_\beta = W(d, \beta)$, representa a persistência do *array* de bytes d em uma transação registrada na *blockchain* β. O resultado T_β é o *transaction ID* gerado pelo registro da transação, que é uma string única de bits atribuída a cada transação registrada na *blockchain* β. A omissão do parâmetro de entrada β causará a adoção da *blockchain default*, previamente selecionada pela operação de *Seleção de Blockchain Padrão*.

Esta função pura não faz nenhum juízo sobre a natureza das informações contidas em D. Na prática, será atribuição do chamador fazer o devido tratamento da informação, anonimizando-a antes, quando necessário.

5.4.1.4. Registro off-chain

O registro *off-chain* é a funão pura de persistência *off-chain* de uma entidade de forma parcial ou total. Na prática é uma operação de INSERT no SGBD da aplicação. Sua implementação é trivial e ficará sob responsabilidade da aplicação (fora do *framework*). Entretanto, é importante ressaltar que, se o registro *off-chain* se der sobre um atributo do tipo PI, então um registro *on-chain* deve ser disparado via *framework* para o mesmo atributo, visando a persistência *on-chain*.

5.4.1.5. Recuperação on-chain

A ***recuperação on-chain***, denotada por $d = R(T_\beta, \beta)$ é a função pura que recupera na *blockchain* β o array de bytes d, registrado sob o identificador de transação T_β.

5.4.2. Operações

No PrivacyChain os recursos classificados como Operações são: anonimização segura, verificação de anonimização segura, indexação *on-chain*, indexação *on-chain* com anonimização, desindexação *on-chain* e desafio de registro seguro imutável. Seguem-se suas descrições.

5.4.2.1. Anonimização segura (commitment)

A ***anonimização segura*** dos dados D é a operação $A = \gamma(D, s, h)$ que anonimiza D com um "salt" s. Na prática, o resultado será uma anonimização simples sobre a concatenação de D com o salt s, ou seja, $\alpha(D \parallel s, h)$.

Esta operação representa o mecanismo de compromisso criptográfico, que serve para evitar que o processo de anonimização seja revertido por dedução, principalmente quando o domínio de entrada é reduzido.

Por exemplo, suponha-se que o resultado de um exame médico de um paciente (uma informação pessoal e confidencial) tenha apenas dois valores possíveis: "POS" (positivo) e "NEG" (negativo) e que se deseja registrar o resultado do exame na *blockchain*. Caso fosse feita uma anonimização simples sobre o resultado do exame teríamos algo como:

- $\boldsymbol{\alpha}(\text{"}\boldsymbol{NEG}\text{"}, \boldsymbol{SHA256}) =$ *cc525c516e6cc498cb173275fab4a776cd4f9ca989eae60ebdb58e1da566359e*
- $\boldsymbol{\alpha}(\text{"}\boldsymbol{POS}\text{"}, \boldsymbol{SHA256}) =$ *cf3a0304f88cb8cd18dc0b7c117543d2052e5e90223201fae582c5f6ae31c5db*

O resultado (*hash*) seria então armazenado na *blockchain*. Caso não se conhecesse os possíveis valores do exame, seria inviável derivar a informação original a partir do *hashs* apresentados acima e a anonimização simples seria segura. Entretanto, sabendo-se que a entrada se restringe a apenas um conjunto reduzido valores ("POS", "NEG"), é trivial deduzir a partir do *hash* se o valor de entrada foi "POS" ou "NEG".

Para resolver este problema, a operação de anonimização segura visa aumentar a entropia da entrada para aplicar uma camada de confidencialidade. Ou seja, a informação de entrada é adicionada a um salt s, antes da anonimização simples.

Internamente, a operação $A = \gamma(D, s, h)$ é realizada da seguinte forma:

a) concatena-se $\boldsymbol{D}$ com o salt $\boldsymbol{s}$, representado como $\boldsymbol{Y = D \parallel s}$;

b) aciona-se a operação de anonimização simples do resultado $\boldsymbol{Y}$, ou seja $\boldsymbol{A = \alpha(Y)}$.

O salt s e a função de *hash* h devem ser associados a D e mantidos na forma da tupla $\langle D|s|h\rangle$ para permitir uma eventual verificação da anonimização segura no futuro.

5.4.2.2. Verificação de anonimização segura

A ***verificação de anonimização segura*** $\Gamma(A, D, s, h)$ é o processo de verificar se $A = \gamma(D, s, h)$, ou seja, se A é o resultado da anonimização segura de D com o salt s e hash h. Esta operação retorna TRUE se a verificação for bem-sucedida ou FALSE caso contrário.

Internamente, a operação $\Gamma(A, D, s, h)$ é realizada da seguinte forma:

a) calcula-se operação $\boldsymbol{A' = \gamma(D, s, h)}$;

b) se $\boldsymbol{A' = A}$ retorna-se TRUE, ou retorna-se FALSE caso contrário.

5.4.2.3. Indexação on-chain

A ***indexação on-chain***, denotada por $I(L_E, t, d, \beta)$ é a operação que registra na *blockchain* β os dados d de uma entidade E identificada pelo localizador L_E, associando-se ainda o carimbo de tempo t. A operação não faz juízo sobre d, que pode ser uma informação anonimizada ou não.

Na prática, efetua-se o registro *on-chain* $T_\beta = W(d, \beta)$ e em seguida persiste-se localmente (em um serviço de armazenamento interno ao *framework*) a tupla $\langle L_E | t | \beta | T_\beta \rangle$, de maneira a permitir recuperação posterior.

Observe-se que d não é persistido localmente neste serviço de armazenamento interno ao *framework*, mas sim na *blockchain*, podendo ser recuperado com a operação de recuperação *on-chain* $d = R(T_\beta, \beta)$, onde β e T_β são obtidos a partir de L_E e t na tupla persistida $\langle L_E | t | \beta | T_\beta \rangle$.

Na persistência local da tupla $\langle L_E | t | \beta | T_\beta \rangle$, considera-se que o par $\{L_E, t\}$ deve ser único, de maneira que a execução de uma segunda operação de indexação usando o mesmo par $\{L_E, t\}$ irá sobrepor os valores de β e T_β na tupla $\langle L_E | t | \beta | T_\beta \rangle$, perdendo valores anteriormente registrados.

5.4.2.4. Indexação on-chain com anonimização

A ***indexação on-chain com anonimização simples***, denotada por $x = I_R(L_E, t, d, \beta)$, é a operação de indexação *on-chain* que executa uma anonimização simples dos dados d antes de proceder com a indexação. Ou seja, usa $d' = \alpha(d, h)$ no lugar de d. Neste caso, a indexação on-chain persiste o atributo adicional de controle h e a tupla de controle passa a ser $\langle L_E | t | \beta | T_\beta | h \rangle$.

A indexação on-chain com anonimização segura, denotada por $I_S(L_E, t, d, \beta)$, é a operação de indexação on-chain que executa uma anonimização segura dos dados d antes de proceder com a indexação. Ou seja, usa $d' = \gamma(d, s, h)$ no lugar de d. Neste caso, a indexação on-chain persistirá também os atributos adicionais de controle s, h e a tupla de controle passa a ser $\langle L_E | t | \beta | T_\beta | s | h \rangle$.

Por exemplo, suponha-se que uma aplicação tenha os seguintes dados pessoais:

```
{
   "cpf":"72815157071",
   "datetime":1620766937,
   "exam":"HIV",
   "result":"POS"
}
```

Para indexar esses dados on-chain de forma anonimizada e segura com a ajuda do framework PrivacyChain, poderia acionar a indexação da seguinte forma:

1) obter uma forma canônica para os dados[10] D={"cpf":"72815157071","exam":"HIV","datetime":1620766937,"result":"POS"}

2) definir um *salt* aleatório s. Por exemplo, pode-se usar uma função que gera um UUIDv4[11] `s=UUIDv4()="d35c2c63-2d15-43b1-9024-b347cbe473c8"`

3) fazer uma *anonimização segura* de D usando o *salt* s e o hash $h = \text{SHA-256}$ $A = \gamma(D, s, h)$

4) indexar o dado anonimizado A na blockchain ETHEREUM. Isso fará implicitamente o registro on-chain $T_\beta = W(A, ETHEREUM)$ e registrará a tupla de controle $\langle L_E | t | \beta | T_\beta \rangle$

   d. $I("72815157071", 1620766937, A, ETHEREUM)$,

   e. persistindo a tupla $\langle "72815157071" | 1620766937 | ETHEREUM | T_\beta \rangle$

5) Adicionar s e h na tupla de controle persistida, atualizando-a para $\langle L_E | t | \beta | T_\beta | s | h \rangle$

   f. $\langle "72815157071" | 1620766937 | ETHEREUM | T_\beta | "d35c2c63 - 2d15 - 43b1 - 9024 - b347cbe473c8" | SHA256 \rangle$

5.4.2.5. Desindexação on-chain

A **desindexação *on-chain*** de uma entidade E identificada pelo localizador L_E, denotada por $\Delta(L_E, t)$, é a operação que consiste em dissociar uma indexação *on-chain* anterior para localizador L_E no tempo t. Quando t for omitido, todos registros de indexação registrados para L_E serão dissociados. Na prática, a dissociação de um índice para L_E consiste na remoção da tupla $\langle L_E | t | \beta | T_\beta | \ldots \rangle$ resultante de uma operação de indexação anterior, que está persistida no serviço de armazenamento interno operado pelo *framework*.

5.4.2.6. Desafio de registro seguro imutável

O ***desafio de registro seguro imutável*** consiste em verificar em um momento t_2 se um valor A que foi registrado em um momento t_1, sendo $t_1 < t_2$, na *blockchain* β sob o identificador de transação T_β é uma anonimização segura dos dados D.

Esta operação pode ser implementada com as seguintes etapas:

---

[10] Uma forma canônica é uma maneira única de representar a mesma informação. Por exemplo, em um JSON, basicamente elimina-se espaços, tabs, *newlines* que não afetam os dados.

[11] Um UUID versão 4 é um identificador universalmente exclusivo gerado por meio de números aleatórios. Os UUID da versão 4 devem ser produzidos usando um gerador de números aleatórios seguro.

a) recupera-se da *blockchain* $\boldsymbol{\beta}$ o dado $\boldsymbol{A'}$ registrado sob o identificador de transação $\boldsymbol{T_\beta}$

- $\boldsymbol{A' = R(T_\beta, \beta)}$

b) calcula-se o resultado de uma anonimização segura sobre $\boldsymbol{D}$ com *salt* $\boldsymbol{s}$ e hash $\boldsymbol{h}$

- $\boldsymbol{A = \gamma(D, s, h)}$

c) retorna TRUE se $\boldsymbol{A = A'}$ ou FALSE caso contrário.

5.4.3. Transações

As transações providas pelo PrivacyChain são o objetivo desta pesquisa, ou seja, o provimento dos direitos ao esquecimento e à retificação. Seguem-se suas descrições.

5.4.3.1. Direito à Retificação

A transação de direito à retificação constitui-se na prática na operação de retificação de dados pessoais na aplicação (retificação *off*-chain) e por conseguinte – como será descrito a seguir – na retificação lógica desses dados pessoais na *blockchain* (retificação *on-chain*).

5.4.3.1.1. Retificação *off-chain*

É a operação de substituir o valor de um ou mais atributos de uma entidade persistida *off-chain*. Na prática é uma operação de UPDATE no SGBD da aplicação. Sua implementação é trivial e ficará sob responsabilidade da aplicação (fora do *framework*). Entretanto, é importante ressaltar que, se a retificação se der sobre um atributo do tipo PI, então uma *retificação on-chain* deve ser disparada via *framework* para o mesmo atributo, visando a atualização *on-chain*.

5.4.3.1.2. Retificação *on-chain*

A retificação de dados da *blockchain* é, na prática, impossível, considerando-se sua característica de imutabilidade. Entretanto, o *framework* fará uma retificação lógica *on-chain* através da desindexação dos atributos anteriores, seguida da indexação dos novos atributos retificados desta entidade.

Essa desindexação $\Delta(L_E, t)$ deve ser realizada passando-se o localizador L_E da entidade em questão e omitindo-se o atributo t (tempo), de forma que todos os registros anteriores relacionados à entidade E sejam dissociados.

Segue-se um exemplo de modo a clarificar esta operação de retificação, com a intenção de trocar a informação "fone" do usuário U do valor X para o valor Y, faz-se:

a) a desindexação de U através da operação $\boldsymbol{\Delta(L_U, t)}$;

b) a anonimização simples ou segura de Y: simples seria $\boldsymbol{Y' = \alpha(Y)}$;

c) registro *on-chain*: $\boldsymbol{T = \delta(Y', \beta)}$;

d) indexa o *transaction* ID gerado com o usuário U: $\boldsymbol{I(L_U, t, "fone\ anonimizado",\ T)}$ (onde $\boldsymbol{L_U}$ é o identificador do usuário – p.ex. CPF).

5.4.3.2. Direito ao Esquecimento

A transação de direito ao esquecimento constitui-se na prática na operação de remoção de dados pessoais na aplicação (remoção *off*-chain) e, por conseguinte, na remoção lógica desses dados pessoais na *blockchain* (remoção *on-chain*).

5.4.3.2.1. Remoção *off-chain*

É a operação de remover uma entidade persistida *off-chain*. Na prática é uma operação de DELETE no SGBD da aplicação. Sua implementação é trivial e ficará sob responsabilidade da aplicação (fora do *framework*). Entretanto, é importante ressaltar que, se a remoção se der sobre uma entidade que possui algum atributo do tipo PI, então uma remoção *on-chain* deve ser disparada para o *framework* para este atributo.

5.4.3.2.2. Remoção *on-chain*

A remoção de dados da *blockchain* é, na prática, impossível, considerando-se sua característica de imutabilidade. Entretanto, o *framework* implementará a remoção *on-chain* através da desindexação dos atributos da entidade removida. O efeito da desindexação é que não haverá mais nenhum elo entre os dados *off-chain* e os dados *on-chain*, o que garante que o *framework* não poderá mais recuperar dados na *blockchain*. Ou seja, na prática, nem os dados serão efetivamente removidos, nem o registro da transação que persistiu os dados será desfeito na *blockchain*. Contudo como a transação do registro terá sido irremediavelmente "deixada para trás", o esquecimento dos dados é garantido.

Para alcançar o efeito do esquecimento de uma entidade E (p.ex.: um cliente) identificada pelo localizador L_E (p.ex.: um CPF), aciona-se a operação de ***desindexação on-chain***, denotada por $\Delta(L_E, t)$. Com isso, quaisquer informações persistidas na *blockchain* não poderão ser associadas ao localizador L_E, uma vez que não existirá conexão entre eles. Isto é, o valor do identificador da transação T_β registrado no repositório do *framework* será perdido.

6. IMPLEMENTAÇÃO DE REFERÊNCIA

Como forma de validar a proposta de solução, os recursos do *framework* PrivacyChain foram materializados na forma de uma implementação de referência (IR), como uma prova de conceito para demonstrar a adequabilidade da solução proposta ante ao problema de pesquisa abordado.

A forma de avaliação aqui utilizada é dita "Descritiva" (LACERDA et al., 2013) e visa a construção de um argumento convincente a respeito da **utilidade** do artefato proposto. (LACERDA et al., 2013) elenca outras possíveis formas de validação de artefatos, tais como: Observacional (Estudo de Caso), Experimental (Experimento Controlado), ou Teste (*Black box/ White Box)*. No entanto, a natureza conceitual (abstrata) da aplicação escolhida para validar o PrivacyChain reforça a adequabilidade da avaliação **Descritiva** escolhida.

Desta forma, a implementação de referência demonstra a **exequibilidade** da proposta de solução para o problema desta pesquisa. Ou seja, mostra que o PrivacyChain é viável na forma um artefato de *software* e que pode ser integrado com aplicações baseadas em DLT que precisem prover o exercício do direito ao esquecimento e à retificação ao titular dos dados pessoais gerenciados. Dito de forma mais abrangente o PrivacyChain constiui-se em um artefato de *software* que resolve a classe de problemas tratada nesta pesquisa.

Este capítulo está organizado da seguinte forma. A Seção 6.1 aponta referências externas onde o código implementado está hospedado e registrado. A Subseção 0 descreve aspectos técnicos e decisões de projeto sobre a implementação de referência. Na Seção 6.3 está a descrição da API. Por fim, na Seção 6.4 está descrito o Modelo de Uso do PrivacyChain.

6.1. Disponibilidade do código fonte e Registro

O código-fonte está disponível publicamente no repositório GitHub através do endereço https://github.com/abmorte/PrivacyChain.

O código-fonte está registrado no INPI[12] através do identificador *hash* 24cdba463467c85f736c71c7b94b344be22ced7450dbe25837a648348d694f48e561ccddca90f50929bb9b8503faa67d16f03e12e17b8a1f75c3901277957a14 (número de registro do INPI ainda não disponível na data do fechamento deste texto).

[12] https://www.gov.br/inpi/pt-br

6.2. Aspectos Técnicos da API PrivacyChain

O PrivacyChain foi implementado na forma de uma **API (*Application Programming Interface*)**. Cada recurso do PrivacyChain está implementado como um *endpoint* nesta API. Fundamentalmente, a implementação do PrivacyChain foi realizada utilizando-se do *framework* Python **FAST API (versão 0.68.0)**[13], da *personal* Ethereum **Ganache (versão 2.5.4)**[14] – também chamada *local test chain* - e do banco de dados **PostgreSQL (versão 13.3)**. Uma biblioteca que merece ser mencionada devido a sua importância para a interação com o Ethereum é a **Web3.py (versão 5.23.0)**[15]. A decisão de se utilizar o FAST API é devida principalmente às suas características quanto à facilidade de codificação, testes e geração de documentação. Quanto ao Ganache a decisão de utilizar-se dessa *personal* Ethereum deve-se principalmente à facilidade em efetuar testes. Outras vantagens dignas de menção são:

1. Disponibiliza 10 carteiras cada uma com uma quantia inicial de 100 Ethers (ETH[16]);
2. Permite o desenvolvimento local simulando uma *blockchain* pública através de uma interface gráfica.

Para que o PrivacyChain utilize a Ethereum pública (*mainnet*) ao invés do Ganache — como atualmente está implementado — é necessário utilizar-se de um provedor de acesso ao nó remoto (*remote node provider*), tal como o Infura[17] (“Web3.py”, [s.d.]).

A Figura 7 tem como objetivo apresentar as diferenças de código entre uma chamada local ao Ganache (*local provider*) e uma chamada a um nó remoto via Infura (*remote provider*).

[13] https://fastapi.tiangolo.com/

[14] https://www.trufflesuite.com/ganache

[15] https://web3py.readthedocs.io/en/stable/

[16] Criptomoeda nativa da plataforma Ethereum.

[17] O Infura (https://infura.io/) é um serviço popular no ecossistema Ethereum, usado para facilitar a conexão com redes Ethereum (*mainnet and testnets*) e também com o IPFS (FERREIRA et al., 2022).

Figura 7 – Chamadas local e remota do método HTTPProvider da biblioteca Web3.py

```
from web3 import Web3

 # Local Provider - código atual do PrivacyChain
 w3 = Web3(Web3.HTTPProvider('http://127.0.0.1:7545'))


 # Remote Provider - chamada à Ethereum pública via Infura
 w3 = Web3(Web3.HTTPProvider('"https://mainnet.infura.io/v3/YOUR_PROJECT_ID))
```

6.3. Descrição da API PrivacyChain

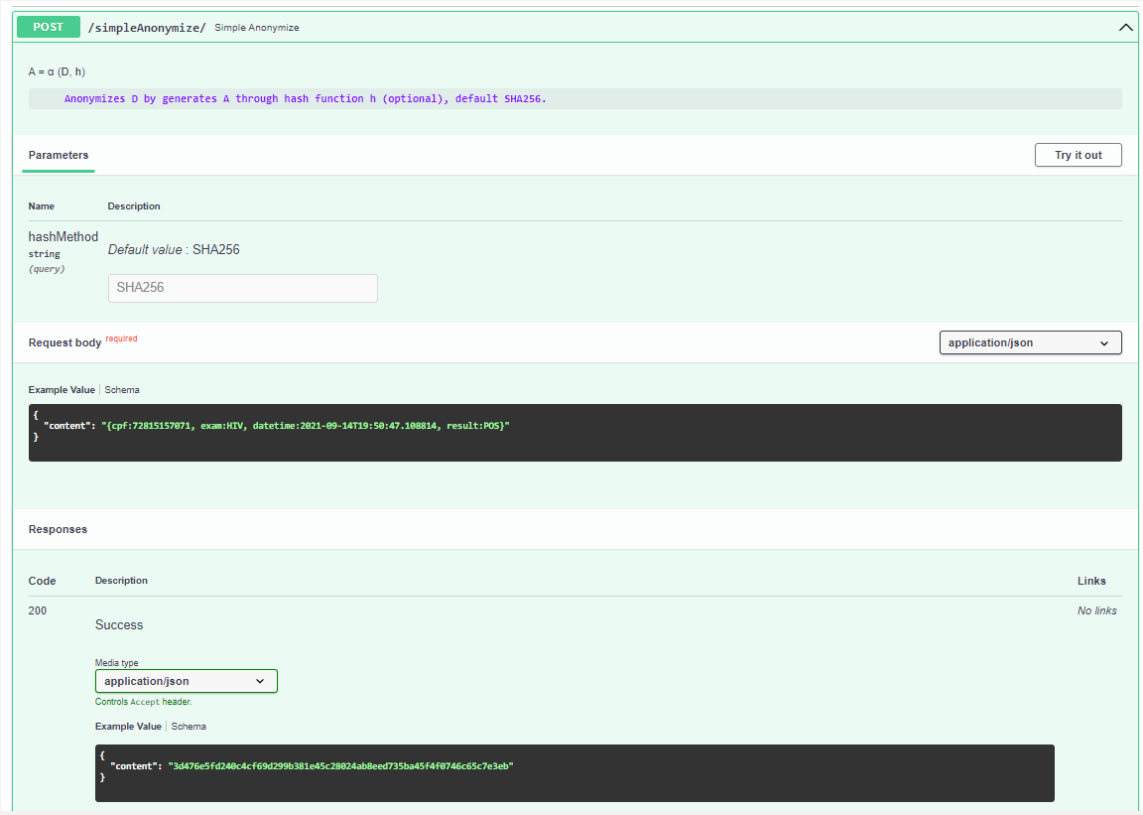
Na Tabela 8 estão relacionados os recursos do PrivacyChain, as referências para suas especificações formais e seus respectivos *endpoints* na API da IR. Desta forma, é possível fazer a associação entre a especificação formal de cada recurso, descrita ao longo do Capítulo 5, e sua implementação sob a forma de um *endpoint* na API.

Tabela 8 – Correspondência entre as funções, operações e transações PrivacyChain e o *endpoint* correspondente na API da IR

Recurso	Tipo	*Endpoint* na API
Anonimização Simples - vide 5.4.1.1	Função Pura	`/simpleAnonymize/`
Anonimização segura (*commitment*) – vide 5.4.2.1	Operação	`/secureAnonymize/`
Verificação de anonimização segura – vide 5.4.2.2	Operação	`/verifySecureAnonymize/`
Seleção de *Blockchain* Padrão - vide 5.4.1.2	Função Pura	`/setDefaultBlockchain/`
Registro *on-chain* – vide 5.4.1.3	Função Pura	`/registerOnChain/`
Registro *off-chain* - vide 5.4.1.4	Função Pura	`/registerOffChain/`
Recuperação *on-chain* - vide 5.4.1.5	Função Pura	`/getOnChain/`
Indexação *on-chain* - vide 5.4.2.3	Operação	`/indexOnChain/`
Indexação *on-chain* com anonimização - vide 5.4.2.4	Operação	`/indexSecureOnChain/`
Desindexação *on-chain* - vide 5.4.2.5	Operação	`/unindexOnChain/`
Desafio de registro seguro imutável - vide 5.4.2.6	Operação	`/verifySecureImmutableRegister/`
Retificação *off-chain* (Direito à Retificação) - vide 5.4.3.1.1	Transação	`/rectifyOffChain/`
Retificação *on-chain* (Direito à Retificação) - vide 5.4.3.1.2	Transação	`/rectifyOnChain/`
Remoção *off-chain* (Direito ao Esquecimento) - vide 5.4.3.2.1	Transação	`/removeOffChain/`
Remoção *on-chain* (Direito ao Esquecimento) - vide 5.4.3.2.2	Transação	`/removeOnChain/`

A associação entre a especificação formal do recurso e sua implementação sob a forma de *endpoint* fica explícita na especificação do *framework* PrivacyChain, disponível em formato OpenAPI através do endereço http://localhost:8000/doc (vide ANEXO A – Instalação do *PrivacyChain*). A Figura 8 abaixo ilustra essa associação para o *endpoint* **`/simpleAnonymize/`**.

Figura 8 – Especificação OpenAPI do *endpoint* simpleAnonymize



6.4. Modelo de Uso do PrivacyChain

Esta seção visa demonstrar a utilidade do artefato PrivacyChain através do manual de utilização (modelo de uso) dessa ferramenta. Aqui descreve-se o modelo (padrão) de uso do *framework* PrivacyChain. O modelo constitui-se em um conjunto de instruções e deve ser utilizado por aplicações que desejem consumir os recursos do PrivacyChain.

Conforme relatado nos capítulos anteriores, em especial no Capítulo 5 - Proposta de Solução, o PrivacyChain destina-se a prover serviços para aplicações baseadas em DLT que realizam tratamento de dados pessoais, com a finalidade de garantir que o titular dos dados possa exercer os direitos ao esquecimento e à retificação presentes na LGPD. Com essa finalidade em foco, dois pontos merecem destaque:

i. O PrivacyChain pode ser utilizado tanto por aplicações legadas quanto por novas aplicações, não havendo distinção quanto à forma de uso do *framework*.

ii. Na implementação de referência do PrivacyChain, os dados anonimizados foram

persistidos em uma *blockchain* de testes Ethereum. No entanto, conceitualmente não há restrições à utilização do PrivacyChain com uma *blockchain* privada ou outra *blockchain* de escolha da aplicação do usuário (vide 2.2.3).

6.5. Integração dos serviços do PrivacyChain na aplicação

Para utilização do PrivacyChain é necessário a execução sequencial das etapas: 1) Preparar (classificar) dados pessoais e; 2) adequar a aplicação para acionamento dos serviços do PrivacyChain através de chamadas à API REST do PrivacyChain.

Uma vez que essas etapas tenham sido executadas, o PrivacyChain estará apto a prover o **direito ao esquecimento** e o **direito à retificação,** alvos da pesquisa aqui realizada.

Além da descrição detalhada dessas etapas, convém uma apresentação e discussão sobre os artefatos necessários ao uso do PrivacyChain. As instruções detalhadas para a instalação da sua implementação de referência estão disponíveis no ANEXO A.

6.5.1. Etapa 1: Preparar dados pessoais

Após a instalação do PrivacyChain, conforme orientações disponíveis no ANEXO A, é necessária uma etapa que se encarregue da **preparação dos dados pessoais** a serem registrados na *blockchain*. Para tal é preciso:

1. Listar - conforme contexto de negócio da aplicação - os dados pessoais que serão armazenados na *blockchain*[18];
2. Selecionar um dado que identifique de forma atômica o titular dos dados pessoais, que será a chave locator a ser utilizada nos *endpoints* de registro na *blockchain*.

6.5.2. Etapa 2: Adequar aplicação para uso dos serviços do PrivacyChain

Essa adequação consiste em construir blocos de código que contenham:

1. Método CRUD utilizado pela aplicação;
2. Chamada ao *endpoint* correspondente no PrivacyChain.

Por exemplo, dada uma aplicação que realize a inserção de registros médicos em uma base de dados local e em uma rede *blockchain* é uma boa prática de programação que essas duas operações ocorram dentro de um bloco *try/catch*, pois estas fazem parte de uma mesma unidade de trabalho, de forma que uma não pode ocorrer dissociada da outra.

---

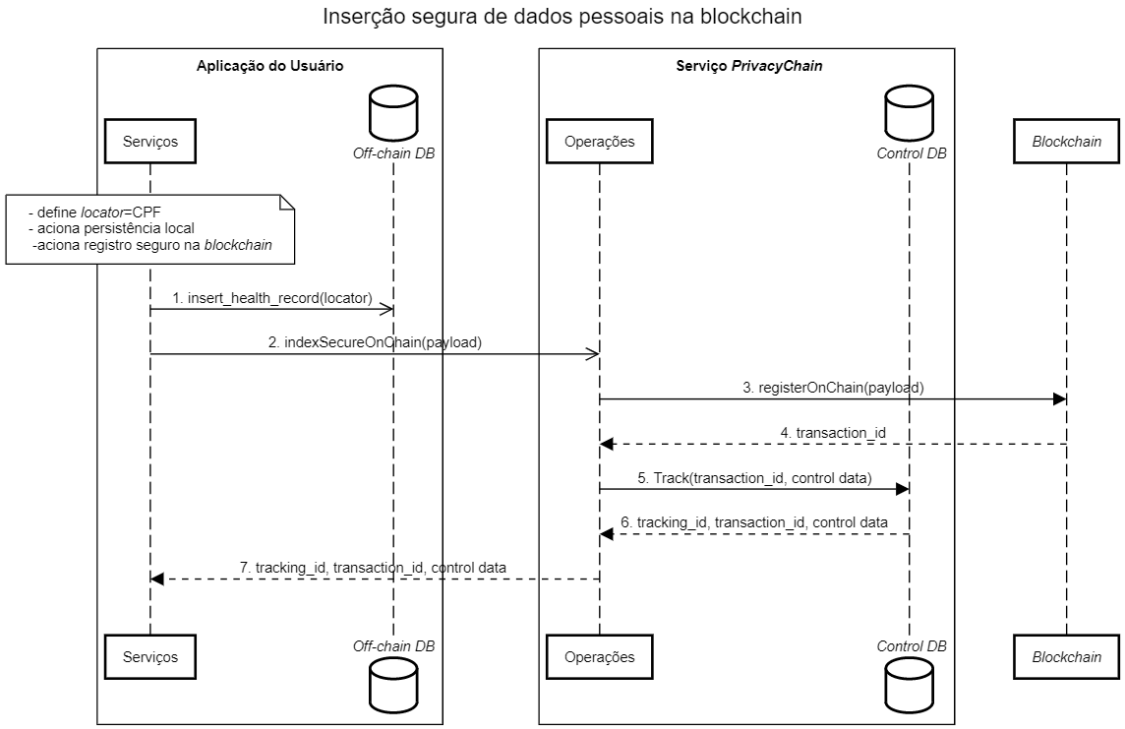
[18] A Seção 5.2.1 - Business Layer traz importantes informações sobre o que recomenda a LGPD quanto à classificação de dados pessoais.

A seguir, nas figuras Figura 9, Figura 11 e Figura 13 encontra-se o pseudocódigo ilustrando como deve se dar essa adequação respectivamente para as ações de inserir, esquecer e retificar dados pessoais em aplicações baseadas em DLT que se utilizem do *framework* PrivacyChain.

6.5.3. Inserção segura de dados pessoais (endpoint indexSecureOnChain)

O diagrama de sequência apresentado na Figura 9 ilustra a inserção segura de dados pessoais na *blockchain*. O processo é composto pelo registro na base local da aplicação *Off-chain DB* via 1. insert_health_record e pelo método 2. indexSecureOnChain que realiza a chamada ao *endpoint* **`/indexSecureOnChain`**. Internamente, este grava na cadeia por meio da chamada à 3. registerOnChain, retornando o identificador da transação 4. transaction_id. Compõe ainda o processo, a persistência das informações de controle em *Control DB* através de 5. Track. Estas informações de controle são retornadas ao *endpoint* por meio de 6. tracking_id, control_data, e desse à aplicação via 7. tracking_id, control_data. Considera-se que estes registros médicos contêm dados pessoais e que a inserção na *blockchain* é realizada com anonimização segura.

Figura 9 – Diagrama de sequência para inserção segura de dados pessoais na *blockchain*



Verifica-se na Figura 9 que a chave locator, que identifica atomicamente o titular dos dados pessoais, é passada como parâmetro ao método de inserção de dados na base da aplicação - insert_health_record(locator). Ela também faz parte do *payload* (objeto JSON) do *endpoint*

/indexSecureOnChain, utilizado para registro com anonimização segura (vide Seção 5.4.2.1). Segue-se descrição das chaves do *payload*:

1. ***to_wallet***: carteira de origem dos dados pessoais;
2. ***from_wallet****:* carteira de destino dos dados pessoais;
3. ***content****:* dados pessoais em formato canônico;
4. ***locator****:* identificador do titular dos dados pessoais;
5. ***datetime****:* carimbo de tempo indicando o momento do registro dos dados pessoais na *blockchain*.
6. ***salt****:* valor adicional utilizado para anonimização segura.

Adicionalmente, segue-se na Figura 10 um trecho de código em Python de forma a demostrar como pode ser implementado um *client* para consumo do *endpoint* ***/indexSecureOnchain***.

Figura 10 – Código de exemplo do *client* para consumo do *endpoint indexSecureOnchain*

```
import requests

url = "http://localhost:8000/indexSecureOnChain/"

payload = {
    "to_wallet": "0x1eca7eD6322B410219Ef953634442AF33aB05BA3",
    "from_wallet": "0x190e97032E45A1c3E1D7E2B1460b62098A5419ab",
    "content": "{cpf:72815157071, exam:HIV, datetime:2021-09-14T19:50:47.108814, result:POS}",
    "locator": "72815157071",
    "datetime": "2021-09-25T10:58:00.000000",
    "salt": "e3719002-8c09-4c8f-8da3-9f5ce34c2d76"
}
headers = {"Content-Type": "application/json"}

response = requests.request("POST", url, json=payload, headers=headers)

print(response.text)
```

6.5.4. Direito ao Esquecimento (endpoint removeOnChain)

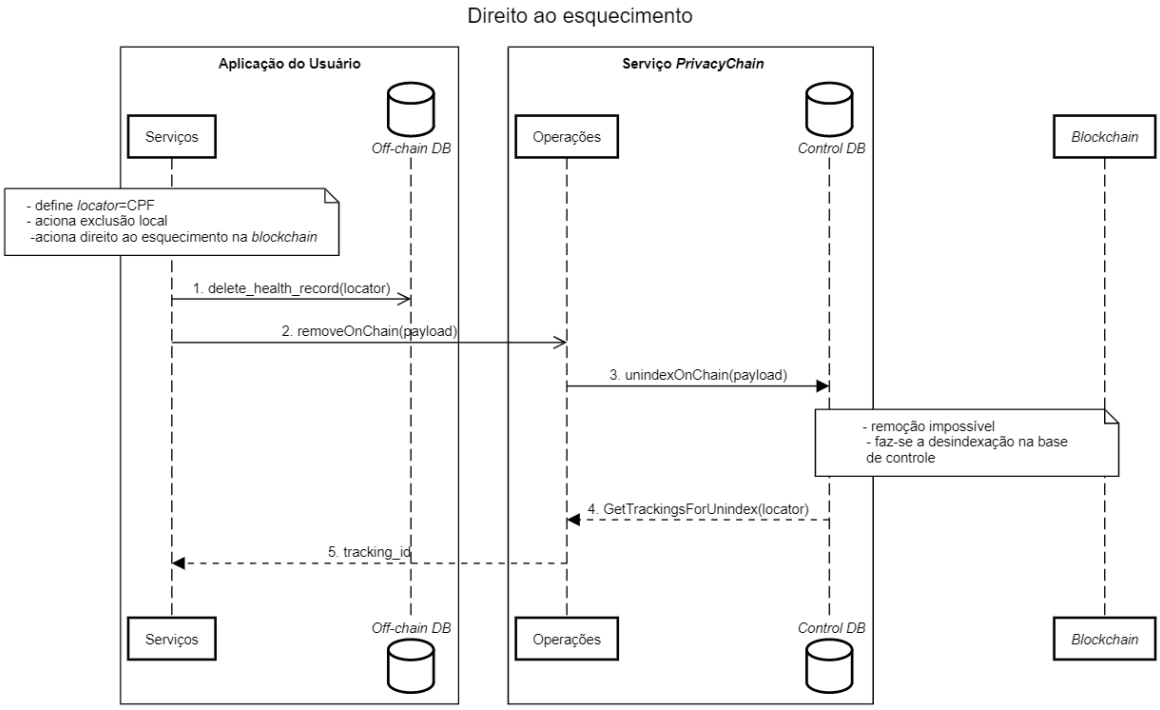
Para que se possa exercer o **direito ao esquecimento** em uma aplicação baseada em DLT faz-se necessário executar os seguintes passos:

1. fazer uso de um método que exclua os dados pessoais de sua base de dados local;
2. realizar chamada ao *endpoint removeOnChain*.

Para ilustrar melhor como utilizar esse *endpoint*, considera-se a aplicação que realiza a manutenção de registros médicos mencionada na seção anterior. Nesse caso, a aplicação deve excluir os registros médicos em sua base de dados local e realizar a chamada ao *endpoint removeOnChain*. Conforme já mencionado, é uma boa prática de programação que essas duas operações ocorram dentro de um bloco *try/catch*.

A Figura 11 demonstra a situação acima mencionada.

Figura 11 – Diagrama de sequência - direito ao esquecimento



O direito ao esquecimento na *blockchain* é realizado através da desindexação dos dados pessoais no ***database framework***, de forma que se perde a referência a esses dados pessoais na rede *blockchain*.

O *payload* (objeto JSON) utilizado como parâmetro de entrada do *endpoint* removeOnChain tem as seguintes chaves:

- ***`Locator:`*** identificador do titular dos dados pessoais;
- ***`datetime:`*** carimbo de tempo indicando o momento da exclusão dos dados pessoais na *blockchain*.

Adicionalmente, segue-se na Figura 12 um trecho de código em Python de forma a demostrar a implementação de um *client* para consumo do *endpoint removeOnchain.*

Figura 12 - Código de exemplo do *client* para consumo do *endpoint removeOnchain*

```
import requests

url = "http://localhost:8000/removeOnChain/"

payload = {
    "locator": "72815157071",
    "datetime": "2021-09-14T19:50:47.108814"
}
headers = {"Content-Type": "application/json"}

response = requests.request("POST", url, json=payload, headers=headers)

print(response.text)
```

6.5.5. Direito à Retificação (endpoint rectifyOnChain)

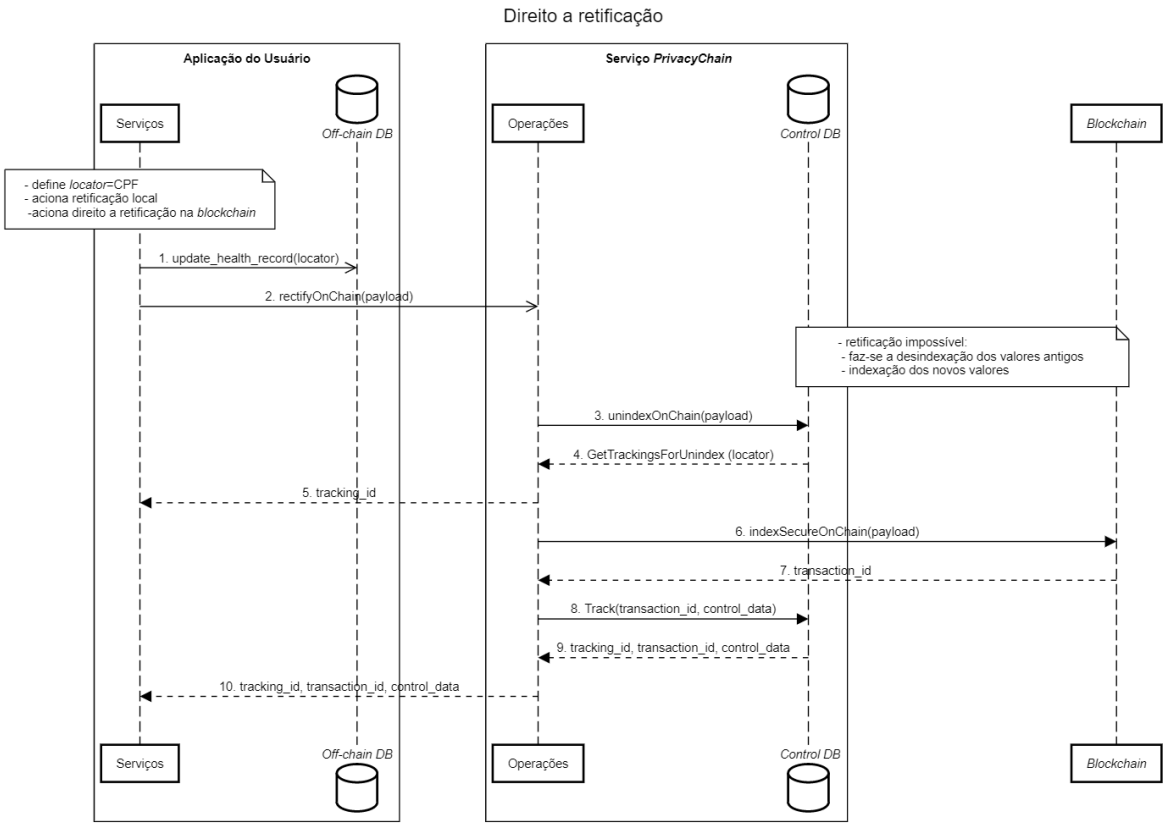
Para que se possa exercer o direito à retificação em uma aplicação baseada em DLT faz-se necessário executar os seguintes passos:

1. fazer uso de um método que retifique os dados pessoais em sua base de dados local;
2. realizar chamada ao *endpoint rectifyOnChain*.

Para ilustrar como utilizar esse *endpoint*, considera-se a aplicação que realiza a manutenção de registros médicos mencionada na seção anterior. Nesse caso, a aplicação deve retificar os registros médicos em sua base de dados local e realizar a chamada ao *endpoint rectifyOnChain*. Conforme já mencionado, é uma boa prática de programação que essas duas operações façam parte de uma mesma unidade de trabalho, ou seja, ocorram dentro de um bloco *try/catch*.

A Figura 13 demonstra a situação acima mencionada.

Figura 13 – Diagrama de sequência - direito à retificação



O direito à retificação na *blockchain* é realizado através da:

1. desindexação dos antigos dados pessoais no *database framework*, de forma que se perde a referência a esses antigos dados na rede *blockchain*;
2. reindexação no *database framework* dos novos dados pessoais inseridos na rede *blockchain*.

O *payload* (objeto JSON) utilizado como parâmetro de entrada do *endpoint rectifyOnChain* tem as seguintes chaves:

- ***`content`***: novos dados pessoais (alvo da retificação) em formato canônico;
- ***`salt`***: valor adicional utilizado para anonimização segura;
- ***`to_wallet`***: carteira de origem dos dados pessoais;
- ***`from_wallet`***: carteira de destino dos dados pessoais;
- ***`locator:`*** identificador do titular dos dados pessoais;

- ***datetime***: carimbo de tempo indicando o momento da retificação dos dados pessoais na *blockchain*.

Adicionalmente, segue-se na Figura 14 um trecho de código em Python de forma a demostrar como pode ser implementado um *client* para consumo do *endpoint rectifyOnchain*.

Figura 14 - Código de exemplo do *client* para consumo do *endpoint rectifyOnchain*

```
import requests

url = "http://localhost:8000/rectifyOnChain/"

payload = {
    "content": "{cpf:72815157071, exam:HIV, datetime:2021-09-14T19:50:47.108814, result:POS}",
    "salt": "e3719002-8c09-4c8f-8da3-9f5ce34c2d76",
    "to_wallet": "0x1eca7eD6322B410219Ef953634442AF33aB05BA3",
    "from_wallet": "0x190e97032E45A1c3E1D7E2B1460b62098A5419ab",
    "locator": "72815157071",
    "datetime": ""
}
headers = {"Content-Type": "application/json"}

response = requests.request("POST", url, json=payload, headers=headers)

print(response.text)
```

7. CONSIDERAÇÕES FINAIS E CONCLUSÕES

Os direitos ao esquecimento e à retificação da LGPD estão inseridos em um contexto mais amplo e de grande importância, cujo objetivo é a garantia dos direitos fundamentais de liberdade e privacidade.

O panorama que convencionou-se chamar de 4.ª Revolução Digital envolve a inédita disponibilidade de uma grande capacidade de **processamento**[19] de uma **quantidade massiva de dados**[20]**,** impulsionada pela possibilidade de **obtenção de conhecimento** através do uso de tecnologias como Inteligência Artificial, *Machine Learning* e *Deep Learning* (POLITOU et al., 2022). Neste contexto, um dos principais desafios é valer-se desse panorama sem, contudo, prescindir do direito à privacidade.

O foco desta pesquisa foi o problema de aplicações que usam DLT para registro de dados pessoais, para as quais a característica intrínseca de imutabilidade da DLT é um obstáculo para a garantia dos direitos ao esquecimento e à retificação pelo titular, exigidos pela LGPD.

A vigência da LGPD trouxe consigo a exigência da adequação do tratamento de dados pessoais realizado pelas aplicações de *software*, de forma que este tratamento se dê respeitando-se os direitos do titular elencados na legislação. Ou seja, os operadores e controladores devem adequar-se a estas novas exigências legais, o que envolve não apenas uma restruturação e reavaliação jurídica nos processos internos da organização, mas também uma sobrecarga técnica na reformulação de sistemas legados e reorientação para novos sistemas. O *framework* PrivacyChain contribui nesta última atividade, ao permitir o tratamento dos dados pelas aplicações que fazem uso de DLTs, através do uso de uma API bem definida.

Como solução para o problema, o *framework* PrivacyChain foi proposto como um conjunto de serviços que podem ser acionados de forma transparente por aplicações para implementar a retificação e esquecimento mesmo usando DLT. Os serviços foram formalmente especificados e materializados em uma implementação de referência na forma de uma API REST como prova de conceito e validação da sua viabilidade e utilidade, usando como base pacotes de software, bibliotecas e SGBD de uso comum e de código aberto. O código fonte da implementação de referência do PrivacyChain está disponível em um repositório público com documentação para sua instalação e uso com licença MIT.

[19] vide o uso de tecnologias como Cloud, 5G, computação móvel ubíqua etc.

[20] vide o uso de tecnologias como BigData, IoT etc.

7.1. Resultados e Contribuições

O artigo (BOA MORTE et al., 2020), publicado no SBRC 2020, teve escopo mais genérico do que esta pesquisa, englobando a compatibilização do uso de dados pessoais em aplicações baseadas em DLT com todos os direitos e princípios elencados na LGPD, não apenas com os direitos ao esquecimento e à retificação. Este artigo reflete a fase de estudos iniciais desta pesquisa, que envolviam a revisão da literatura dentro do escopo definido, do estudo da legislação e seu impacto em sistemas de informação.

O restante das atividades desta pesquisa teve foco mais específico e tratou da compatibilização unicamente com os direitos ao esquecimento e à retificação.

Os seguintes resultados foram alcançados:

1. Propositura do Referencial Teórico (vide Subseção 4.4.1);
2. Revisão Sistemática da Literatura (RSL – vide Subseção 4.2);
3. Especificação do *framework* PrivacyChain e formalização dos seus recursos (vide Capítulo 5);
4. Implementação de referência do *framework* PrivacyChain na forma de uma API REST para integração com aplicações (vide Capítulo 6);
5. Manual de uso do *framework* PrivacyChain, com orientações de como usar seus recursos para incluir em aplicações o suporte aos direitos ao esquecimento e à retificação (vide Capítulo 6);
6. Registro do código-fonte no INPI através do identificador 24cdba463467c85f736c71c7b94b344be22ced7450dbe25837a648348d694f48e561ccddca90f50929bb9b8503faa67d16f03e12e17b8a1f75c3901277957a14;
7. Disponibilização do código-fonte no repositório GitHub https://github.com/abmorte/PrivacyChain.

7.2. Trabalhos Futuros

O *framework* PrivacyChain foi proposto como solução para o problema de compatibilizar aplicações com o direito ao esquecimento e à retificação do titular dos dados tratados. Sua implementação de referência, entretanto, focou em materializar uma prova de conceito suficiente para sua validação como solução, em função da delimitação do escopo e limitações de tempo. Entretanto, o PrivacyChain tem potencial para aplicações profissionais e comerciais, mediante algumas melhorias a serem implementadas como trabalhos futuros.

7.2.1. Compatibilidade transparente com múltiplas DLTs

A implementação de referência do PrivacyChain adotou a *blockchain* Ethereum. Uma evolução deste *framework* pode implementar ou integrar o uso de *brokers de blockchain*, vide (PIRES et al., 2018). Um broker de *blockchain* permite transparência de uso de através da configuração de um "*pool de blockchains*" e adoção de critérios configuráveis como "menor preço por transação" ou "menor tempo de confirmação da transação". O objetivo é permitir o tratamento de dados pessoais em conformidade com o direito ao esquecimento e à retificação não apenas na DLT Ethereum, como fora na implementação aqui realizada, mas em múltiplas DLTs.

7.2.2. Uso do protocolo IPFS como repositório distribuído off-chain

A base relacional onde são armazenadas informações de rastreio e controle[21] dos dados pessoais constitui-se em um ponto único centralizado dentro de um contexto naturalmente distribuído (aplicações DLT), representando um possível ponto de vulnerabilidade, uma vez que acrescenta à solução questões relativas à segurança, disponibilidade, integridade etc.

Os autores em (POLITOU et al., 2022) narram que o IPFS tem sido empregado em projetos *blockchain* para o armazenamento de dados pessoais *off-chain*, de forma a estar em conformidade com o direito ao esquecimento do GDPR e propõem um protocolo que endereça questões relativas ao apagamento de dados sob IPFS[22].

A proposta de trabalho futuro, portanto, é a utilização do protocolo IPFS - em substituição à base de dados relacional empregada nesta pesquisa - para viabilizar o armazenamento distribuído de dados pessoais *off-chain* em aplicações baseadas em DLT, de modo a estar em conformidade com os direitos ao esquecimento e à retificação da LGPD.

7.2.3. Direito ao esquecimento e à retificação em Identidades Auto-Soberanas

Os autores em (POLITOU et al., 2022) citam que, com o advento da COVID-19, alguns governos estão introduzindo o conceito de **passaporte de imunidade**. Esses documentos visam permitir aos indivíduos viajar entre países e participar de grandes eventos, atestando que determinado indivíduo foi totalmente vacinado. Questões relativas à privacidade emergem. Por exemplo, o uso indevido do passaporte, quando não usado para viagens ou participação em eventos, mas para acessar instalações e serviços onde não seria necessário usar uma forma de identidade.

Sugere-se a adequabilidade do uso de uma abordagem descentralizada, baseada em *blockchain*, para o gerenciamento desses passaportes, onde o *ledger* distribuído serviria como uma

[21] vide subseções 5.2.3., 5.4.2.3

[22] https://ipfs.io/

trilha de auditoria da vacinação, e os indivíduos poderiam permitir seletivamente que outros acessem ou não o seu rastreamento. Isto posto, o gerenciamento de identidades pode ser um problema a resolver. Alguns pesquisadores sugerem o uso de identidades auto-soberanas (baseadas em *blockchain*) para permitir aos indivíduos maior controle dos seus dados.

Diante deste cenário, portanto, a proposta de trabalho futuro é o estudo do Exercício do direito ao esquecimento e à retificação em Identidades Auto-Soberanas.

7.2.4. Privacidade em Registros Eletrônicos de Saúde

Os autores em (POLITOU et al., 2022) narram o uso de aplicativos *mobile* para o rastreio das pessoas que tiveram contato com um indivíduo contaminado com a COVID-19. Essencialmente, a solução consiste em mapear e compartilhar a localização do indivíduo via Bluetooth embarcado em seu dispositivo *mobile*.

Obviamente, manter o controle de quais pessoas alguém contatou levanta questões de privacidade, pois pode revelar dados como localização, preferências religiosas, sexuais etc. Pode-se utilizar um modelo descentralizado através do qual o dispositivo *mobile* enviará os *tokens* de seus contatos recentes. O servidor ou a arquitetura distribuída funcionam como um quadro de avisos anônimo. Os usuários teriam a opção de baixar localmente esses *tokens* e determinar se entraram em contato com alguém com diagnóstico de COVID-19.

A descrição desse cenário serve como exemplo de aplicação, mas a proposta de trabalho futuro pode ser mais abrangente e englobar outros registros de saúde. Isto posto, portanto, a proposta de trabalho futuro é a realização de Estudos quanto à privacidade no contexto dos Registros Eletrônicos de Saúde (*Electronic Health Records - EHR*).

7.2.5. Validações de Segurança

Por fim, de forma a transformar a prova de conceito implementada em um produto de *software*, uma importante tarefa a ser efetuada é o acréscimo de validações de segurança ao código, mitigando possíveis ataques ao *framework.*

REFERÊNCIAS BIBLIOGRÁFICAS

AEPD. **A Guide to Privacy by Design**. Disponível em: <https://www.aepd.es/sites/default/files/2019-12/guia-privacidad-desde-diseno_en.pdf>. Acesso em: 20 set. 2020.

BRASIL. **Lei nº 13.709, de 30 de agosto de 2018**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm#art65>. Acesso em: 20 dez. 2019.

BOA MORTE, A. ANÁLIA MEIRA, ROSTAND COSTA, DÊNIO MARIZ. Uma Análise Sobre o Uso de DLTs no Tratamento de Dados Pessoais: Aderência aos Princípios e Direitos elencados na. **W*Blockchain* SBRC - Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, p. 14, 2020.

CASINO, F. et al. Immutability and Decentralized Storage: An Analysis of Emerging Threats. **IEEE Access**, v. 8, p. 4737–4744, 2020.

CNIL. *Blockchain*. Solutions for a responsible use of the *blockchain* in the context of personal data. **CNIL Report**, p. 10, 2018.

CR´EPEAU, C. Commitment. Disponível em: <http://crypto.cs.mcgill.ca/~crepeau/PDF/Commit.pdf>. Acesso em: 19 set. 2020.

CUNNINGHAM, H. C. CSci 555: Functional Programming Functional Programming in Scala Functional Data Structures. p. 38, 2019.

ELSHEKEIL, S. A.; LAOYOOKHONG, S. GDPR Privacy by Design. p. 1–49, 2017.

FABER, B. et al. BPDIMS: A *Blockchain*-based Personal Data and Identity Management System. v. 6, p. 6855–6864, 2019.

FERREIRA, C. M. S. et al. A middleware for systems consumes Ethereum data in soft real-time: a Semantic Web approach. p. 122–127, 2022.

FINCK, M. *Blockchains* and the General Data Protection Regulation. ***Blockchain* Regulation and Governance in Europe**, n. July, p. 88–116, 2019.

GAUR, N. et al. Hands-On *Blockchain* with Hyperledger: Building decentralized applications with Hyperledger Fabric and Composer. [s.l.] Packt Publishing Ltd, 2018.

GREVE, F. et al. *Blockchain* e a Revolução do Consenso sob Demanda. **Minicursos do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, p. 52, 2018.

HORITA, F. E. A.; NETO, V. V. G.; SANTOS, R. P. DOS. Design Science Research em Sistemas de Informação e Engenharia de *Software*: Conceitos, Aplicações e Trabalhos Futuros. **Método de pesquisa para avanço da ciência e tecnologia.**, n. April 2019, p. 141–172, 2015.

IBM. **Perishable Goods Network**. Disponível em: <https://github.com/hyperledger-archives/composer-sample-networks/tree/master/packages/perishable-network>. Acesso em: 12 out. 2020.

ITSRIO. Lei Geral De Proteção De Dados Pessoais (Lgpd) E Setor Público. p. 40, 2019.

LACERDA, D. P. et al. Design Science Research: A research method to production engineering. **Gestão & Produção**, v. 20, n. 4, p. 741–761, 2013.

LAPES. **StArt - State of the Art through Systematic Review**. Disponível em: <http://lapes.dc.ufscar.br/tools/start_%0Atool. [*software*]>. Acesso em: 10 out. 2020.

MCBRIDE, W.-M. **Beginning Ethereum Smart Contracts Programming**. Berkeley, CA: Apress, 2019.

MCBRIDE, M. **Functional programming in Python**. 1. ed. [s.l.] Axlesoft Ltd, 2018.

ME/SGD. **Guia de Elaboração de Inventário de Dados Pessoais - LGPD**. Brasília: [s.n.]. Disponível em: <https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/guias/guia_inventario_dados_pessoais.pdf/view >. Acesso em: 25 jul. 2021.

NAKAMOTO, S. **Bitcoin: A Peer-to-Peer Electronic Cash System**. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Acesso em: 1 set. 2020.

OFFICE, I.-I. C. **Right to rectification**. Disponível em: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/#ib2>. Acesso em: 12 out. 2020.

OFFICE, I.-I. C. **Data protection impact assessments**. Disponível em: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/>. Acesso em: 17 out. 2020.

OKOLI, C.; DUARTE, T. POR:DAVID W. A.; MATTAR, R. TÉCNICA E INTRODUÇÃO:JOÃO. Guia Para Realizar uma Revisão Sistemática de Literatura. **EaD em Foco**, v. 9, n. 1, p. 1–40, 2019.

ONIK, M. M. H. et al. Privacy-aware *blockchain* for personal data sharing and tracking. **Open Computer Science**, v. 9, n. 1, p. 80–91, 1 jan. 2019.

PIRES, M. et al. Uma Abordagem Baseada em Brokers para Registro de Transações em Múltiplos Livros Razão Distribuídos. **Workshop em Blockchain: Teoria, Tecnologias e Aplicações (WBlockchain_SBRC)**, v. 1, n. 1/2018, 2018.

POLITOU, E.; CASINO, F.; ALEPIS, E.; PATSAKIS, C. *Blockchain* Mutability: Challenges and Proposed Solutions. **IEEE Transactions on Emerging Topics in Computing**, 16 jul. 2019.

POLITOU, E. et al. **Privacy and Data Protection Challenges in the Distributed Era**. 1. ed. Cham: Springer International Publishing, 2022. v. 26

POLITOU, E. et al. Blockchain Mutability: Challenges and Proposed Solutions. **IEEE Transactions on Emerging Topics in Computing**, 16 jul. 2019.

REYNERI, J. M.; KARNIN, E. D. Coin Flipping by Telephone. **IEEE Transactions on Information Theory**, v. 30, n. 5, p. 775–776, 1984.

SHAHAAB, A. Managing Gender Change Information on Immutable *Blockchain* in Context of GDPR. **The Journal of The British *Blockchain* Association**, v. 3, n. 1, p. 1–8, 2020.

SOARES, M.; COSTA, R. Autoidentificação Voluntária e Verificável de Participantes em Aplicações Baseadas em Livros-Razão Distribuídos. **W*Blockchain* SBRC - Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, p. 14, 2019.

SOMMERVILLE, I. **Engenharia de software**. São Paulo: Pearson Prentice Hall, 2011.

UNIVERSITY, S. H. **Organizing Academic Research Papers: Theoretical Framework**. Disponível em: <https://library.sacredheart.edu/c.php?g=29803&p=185919>. Acesso em: 15 out. 2020.

Web3.py. Disponível em: <https://web3py.readthedocs.io/en/stable/quickstart.html#test-provider>. Acesso em: 18 fev. 2022

WIKI, C. **Commitment scheme**. Disponível em: <https://cryptography.fandom.com/wiki/Commitment_scheme>. Acesso em: 19 set. 2020

ANEXOS

ANEXO A – INSTALAÇÃO DO *PRIVACYCHAIN*

O funcionamento da implementação de referência do PrivacyChain requer a instalação e configuração de vários pacotes de software existentes, de um gerenciador de **banco de dados** e o acesso a uma *blockchain*.

O banco de dados é a base relacional utilizada para a indexação e controle dos dados pessoais anonimizados inseridos na *blockchain*. Na implementação de referência do PrivacyChain, é utilizada o PostgreSQL[23] cujo *script* de criação está disponível adiante.

A *blockchain* é usada para persistir os dados pessoais anonimizados. A implementação de referência adotou uma *blockchain* de teste da rede Ethereum, que é implementada e instanciada localmente através do software *client* Ganache. Após a instalação do Ganache, a interface para acesso à rede Ethereum estará acessível a partir da URL http://localhost:7545.

Segue a transcrição das instruções de instalação do PrivacyChain para o ambiente Linux Ubuntu 20.04:

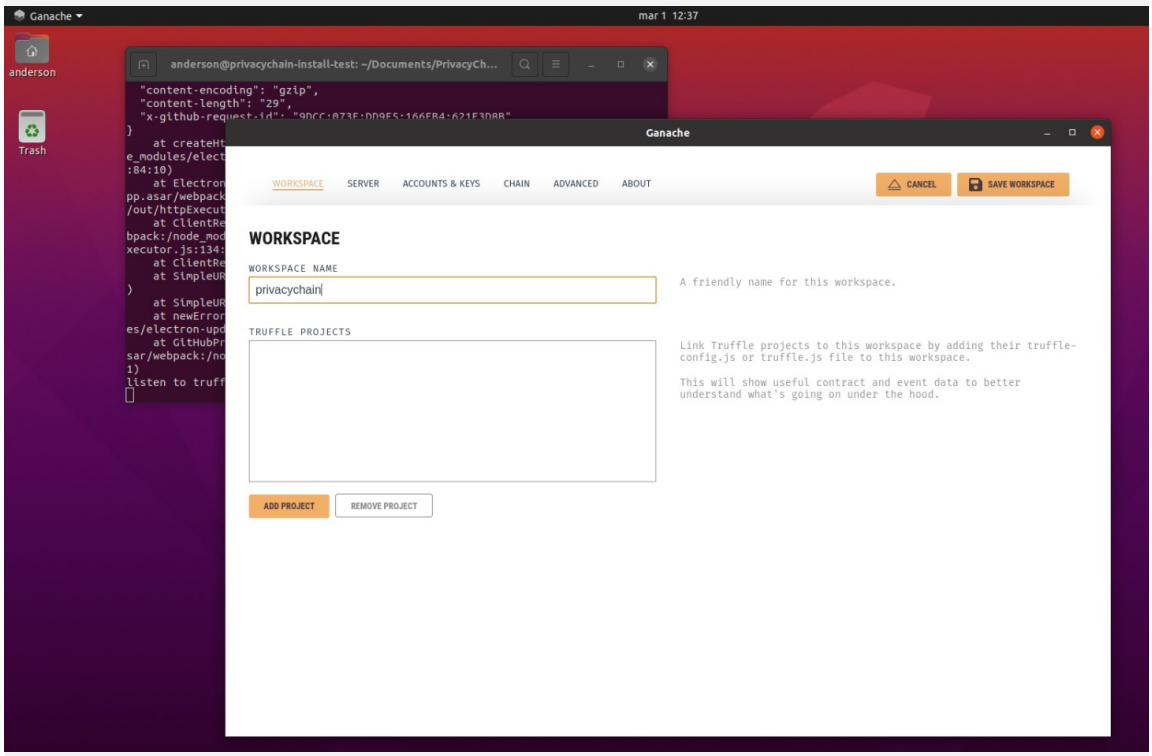
```
#
# instruções para instalação do ambiente de desenvolvimento PrivacyChain
#

# baixando informações sobre os pacotes que serão instalados
sudo apt update && sudo apt -y upgrade

## instalando o Python 3.9
sudo apt install -y python3.9

## instalando o Git
sudo apt install -y git

# instalando bibliotecas de desenvolvimento Python 3.9
sudo apt-get install -y python3.9-venv python3.9-dev

# atualizando a biblioteca python3-pip
sudo apt install -y python3-pip
python3.9 -m pip install --upgrade pip

## obtendo o código-fonte do PrivacyChain
git clone https://github.com/abmorte/PrivacyChain

# Preparando ambiente Python
cd PrivacyChain

# criando ambiente virtual .venv
python3.9 -m venv .venv

# ativando ambiente virtual .venv
source .venv/bin/activate
```

[23] https://www.postgresql.org/download/

```
# instalando requirements para o Ubuntu 20.04.4 LTS
pip install -r requirements/ubuntu20044lts.txt

# Iniciando extensão ASGI (Asynchronous Server Gateway Interface) para Python
# A interface Swagger UI da API estará acessível em http://localhost:8000/docs
uvicorn app.main:app --reload

# Instalando o PostgreSQL
sudo apt install -y postgresql postgresql-contrib

# mudando usuário para a conta postgres
sudo -i -u postgres

# acessando o prompt do Postgres
psql

# criando os objetos de banco de dados
\i /PrivacyChain/script.sql

# saindo o psql para o shell
\q

# obtendo o arquivo de instalação do Ganache
sudo wget https://github.com/trufflesuite/ganache-ui/releases/download/v2.5.4/ganache-2.5.4-linux-x86_64.AppImage

# alterando as permissões do arquivo de instalação do Ganache
chmod 777 ganache-2.5.4-linux-x86_64.AppImage

# executando o instalador do Ganache
./ganache-2.5.4-linux-x86_64.AppImage
```

```
# Observação:
# A interface visual do Ganache é chamada Ganache UI (“is a desktop application
# supporting both Ethereum and Corda technology”). A sua ferramenta de linha de
comando é # chamada ganache-cli, formalmente conhecida como TestRPC - a fast and
customizable blockchain emulator. Para instalação do ganache-cli, consultar
# https://github.com/trufflesuite/ganache-cli-archive/blob/master/README.md

# manualmente, via Ganache UI, deve-se criar um workspace
# com o nome privacychain (em minúsculas), vide imagem abaixo:
```

Figura 15 – Tela de criação do workspace no Ganache



Tendo tudo ocorrido sem erros, consulte a API do PrivacyChain abrindo a seguinte URL em um navegador local:

- Para acessar a interface OpenAPI[24]: http://localhost:8000/docs (Figura 16);
- Para acessar a interface redoc: http://localhost:8000/redoc (Figura 17).

[24] Anteriormente conhecido como especificação Swagger – vide https://swagger.io/

Figura 16 – Interface OpenAPI do PrivacyChain

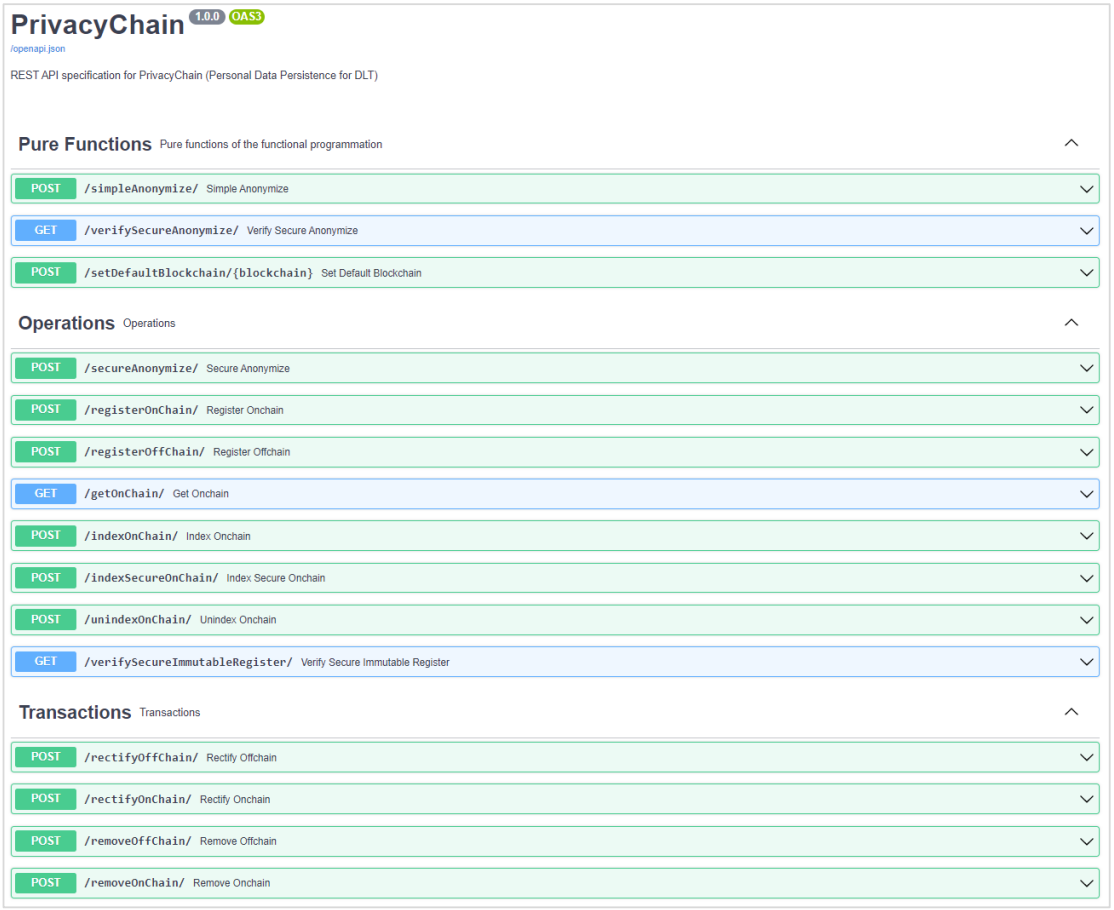
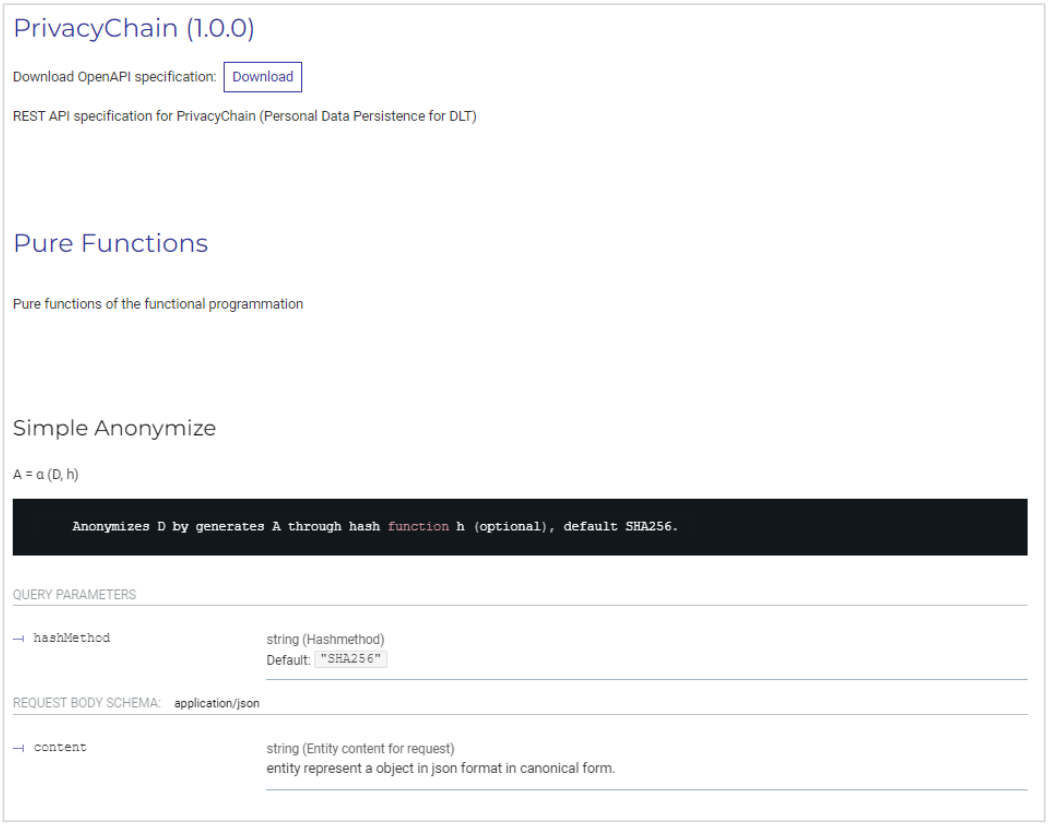


Figura 17 – Interface Redoc do PrivacyChain



ANEXO B – MODELO DE DADOS

O script visualizado na Figura 18 a seguir é uma transcrição do arquivo `script.sql`, localizado no diretório raiz do repositório https://github.com/abmorte/PrivacyChain.

A tabela *tracking* é criada através da execução desse script e armazena informações de rastreio e controle dos registros persitidos na *blockchain* quando da utilização do *endpoints* do *framework* **PrivacyChain**. Isso permite a associação entre os dados da aplicação (*off-chain*) e os dados registrados na *blockchain* (*on-chain*).

Figura 18 – *Script* SQL de criação das tabelas de controle do PrivacyChain

```
CREATE TABLE privacychain.tracking
(
    tracking_id integer NOT NULL GENERATED BY DEFAULT AS IDENTITY ( INCREMENT 1 START 1 MINVALUE 1 MAXVALUE 2147483647 CACHE 1 ),
    canonical_data character varying COLLATE pg_catalog."default",
    anonymized_data character varying COLLATE pg_catalog."default",
    blockchain_id integer,
    transaction_id character varying COLLATE pg_catalog."default",
    salt character varying COLLATE pg_catalog."default",
    hash_method character varying COLLATE pg_catalog."default",
    tracking_dt timestamp without time zone DEFAULT now(),
    locator character varying COLLATE pg_catalog."default",
    CONSTRAINT tracking_pkey PRIMARY KEY (tracking_id)
)
TABLESPACE pg_default;
ALTER TABLE privacychain.tracking
    OWNER to postgres;
COMMENT ON TABLE privacychain.tracking
    IS 'Metadata repository for tracking personal data entered in the blockchain';
COMMENT ON COLUMN privacychain.tracking.tracking_id
    IS 'Sequential code of the personal data tracking table';
COMMENT ON COLUMN privacychain.tracking.canonical_data
    IS 'personal data in canonical format';
COMMENT ON COLUMN privacychain.tracking.anonymized_data
    IS 'anonymized personal data';
COMMENT ON COLUMN privacychain.tracking.blockchain_id
    IS 'blockchain identifier where anonymized personal data is persisted';
COMMENT ON COLUMN privacychain.tracking.transaction_id
    IS 'transaction identifier on the blockchain where anonymized personal data is persisted';
COMMENT ON COLUMN privacychain.tracking.salt
    IS '36-character random string generated using the uuid4 function - see RFC 4122';
COMMENT ON COLUMN privacychain.tracking.hash_method
    IS 'String identifying the method used in the anonymization of personal data. Domain: ''MD5'', ''SHA1'', ''SHA256'', ''SHA512''';
COMMENT ON COLUMN privacychain.tracking.tracking_dt
    IS 'timestamp identifying when the tuple is persisted in the database';
COMMENT ON COLUMN privacychain.tracking.locator
    IS 'identification of entity holding personal data';
```

ENTREGA DA VERSÃO FINAL DE DISSERTAÇÃO

Eu, PROF. DR. Dênio Mariz Timóteo de Sousa, autorizo o aluno(a) Anderson Fernando Vieira da Boa Morte a entregar a versão final da dissertação de mestrado, à secretaria do PPGTI, que foi por mim analisada e está de acordo com os apontamentos feitos pelos membros da banca de apresentação do referido aluno.

Prof. Dr. Dênio Mariz
Orientador

João Pessoa, 02 de março de 2022