DISSERTAÇÃO DE MESTRADO PROFISSIONAL

Proposta de um Framework de Compliance à Lei Geral de Proteção a Dados Pessoais (LGPD): Um estudo de caso para prevenção a fraude no contexto de Big Data

Artur Potiguara Carvalho

Brasília, 08 de setembro de 2021

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

PROPOSAL FOR A COMPLIANCE FRAMEWORK FOR THE GENERAL LAW FOR THE PROTECTION OF PERSONAL DATA (LGPD): A CASE STUDY FOR FRAUD PREVENTION IN THE CONTEXT OF BIG DATA

PROPOSTA DE UM FRAMEWORK DE COMPLIANCE À LEI GERAL DE PROTEÇÃO A DADOS PESSOAIS (LGPD): UM ESTUDO DE CASO PARA PREVENÇÃO DE FRAUDE NO CONTEXTO DE BIG DATA

ARTUR POTIGUARA CARVALHO

ORIENTADORA: EDNA DIAS CANEDO

DISSERTAÇÃO DE MESTRADO PROFISSIONAL EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPEE.MP.012
BRASÍLIA/DF, SETEMBRO - 2021

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO PROFISSIONAL

Proposta de um Framework de Compliance à Lei Geral de Proteção a Dados Pessoais (LGPD): Um estudo de caso para prevenção a fraude no contexto de Big Data

Artur Potiguara Carvalho

Dissertação de Mestrado Profissional submetida ao Departamento de Engenharia Elétrica como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica

Banca Examinadora

Profa. Edna Dias Canedo, Ph.D, CIC/FT/UnB _____
Orientadora

Prof. Fábio Lúcio Lopes de Mendonça, Ph.D, FT/UnB _____
Examinador Interno

Prof. João Souza Neto, Ph.D, Universidade Católica de Brasília (UCB) _____
Examinador Externo

FICHA CATALOGRÁFICA

CARVALHO, ARTUR POTIGUARA
Proposta de um Framework de Compliance à Lei Geral de Proteção a Dados Pessoais (LGPD): Um estudo de caso para prevenção a fraude no contexto de Big Data [Distrito Federal] 2021.
xvi, 200 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2021).
Dissertação de Mestrado Profissional - Universidade de Brasília, Faculdade de Tecnologia.
Departamento de Engenharia Elétrica

1. Governança de Dados
2. Big Data
3. Framework
4. LGPD
5. Privacidade

I. ENE/FT/UnB
II. Proposta de um Framework de Compliance à Lei Geral de Proteção a Dados Pessoais (LGPD): Um estudo de caso para prevenção a fraude no contexto de Big Data

REFERÊNCIA BIBLIOGRÁFICA

CARVALHO, A. P. (2021). *Proposta de um Framework de Compliance à Lei Geral de Proteção a Dados Pessoais (LGPD): Um estudo de caso para prevenção a fraude no contexto de Big Data*. Dissertação de Mestrado Profissional, Publicação: PPEE.MP.012, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 200 p.

CESSÃO DE DIREITOS

AUTOR: Artur Potiguara Carvalho
TÍTULO: Proposta de um Framework de Compliance à Lei Geral de Proteção a Dados Pessoais (LGPD): Um estudo de caso para prevenção a fraude no contexto de Big Data.
GRAU: Mestre em Engenharia Elétrica ANO: 2021

Artur Potiguara Carvalho
Depto. de Engenharia Elétrica (ENE) - FT
Universidade de Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

DEDICATÓRIA

Dedico este trabalho primeiramente a Deus, que plantou em mim a semente da curiosidade. Dedico também este trabalho aos meus familiares, em especial a minha irmã Fernanda, sem a qual talvez eu não tivesse despertado interesse no assunto desta obra. Dedico também aos amigos, colegas de trabalho e de academia que contribuíram de forma direta ou indireta com este trabalho.

AGRADECIMENTOS

Agradeço primeiramente a Deus, do qual entendo que emana toda genuína busca do conhecimento da verdade. Agradeço de forma sincera aos meus pais e irmãos que não só foram privados momentaneamente do meu convívio como auxiliaram de forma direta com o trabalho aqui apresentado. Agradeço também aos amigos que me confortaram com palavras de ânimo nos momentos difíceis. Tenho muito a agradecer à minha professora orientadora Edna Dias Canedo pela paciência, dedicação ímpar, convicção do potencial, e contribuição tão vasta e significativa neste trabalho. Agradeço também às instituições relacionadas a este trabalho, principalmente à Universidade de Brasília e à instituição onde trabalho, pela oportunidade de desenvolver uma pesquisa tão atual e necessária. A todos estes, meu muito obrigado!

"Um pouco de ciência nos afasta de Deus. Muito, nos aproxima."

Louis Pasteur

RESUMO

No cenário de produção de dados em massa ainda em expansão a níveis outrora inimagináveis a proteção à privacidade de dados é um problema que vem ganhando mais destaque a cada dia. Este destaque se reflete também na legislação que tem expandido os direitos dos titulares de dados sobre as suas informações inclusive sobre bases das quais anteriormente eles não possuíam qualquer gerência. Este cenário contribui para o aumento de ações de fraudes, dado que os atacantes dispõem de cada vez mais informações pessoais acessíveis em bases de dados públicas ou de fácil consulta. Faz-se, portanto, necessária a consolidação de boas práticas que, aliadas à experiência prática, auxiliem os projetos de dados em bases massivas a tratar a proteção à privacidade e prevenção a fraude. Esta pesquisa foi conduzida seguindo o método *Design Science Research*, utilizando as técnicas de pesquisa empírica e estudo de caso, a fim de que o *framework* proposto absorvesse o conhecimento acadêmico e a experiência dos profissionais que lidam com os dados, validado através de um caso prático de ingestão de dados. Este trabalho apresenta uma proposta de um *framework* construído a partir das melhores práticas relacionadas à Governança de Dados, Privacidade de Dados e Segurança da Informação. Os resultados encontrados demonstraram a eficácia do *framework* em relação aos aspectos de proteção à privacidade e prevenção a fraude. Além disso, os resultados indicam que as boas práticas já existentes nas disciplinas Governança de Dados, Privacidade de Dados e Segurança da Informação podem auxiliar na proteção à privacidade de dados e por consequência, contribuir para a conformidade com a LGPD.

ABSTRACT

In the scenario of mass data production, still expanding to previously unthinkable levels, data privacy protection is an issue that is gaining more prominence every day. This highlight is also reflected in the regulation that has expanded the rights of data subjects over their information, including bases on which they previously could not manage. This scenario also contributes to the increase in fraudulent actions, as attackers have more and more personal information available on a public or easily accessible data base. It is, therefore, necessary to consolidate good practices that, together with practical experience, help massive data base projects to address privacy protection and fraud prevention. This research was conducted following the Design Science Research method, using empirical research and case study techniques, so that the proposed framework would absorb the academic knowledge and experience of professionals who deal with the data, validated through a practical case of data ingestion. This work presents a proposal for a framework built from the best practices related to Data Governance, Data Privacy and Information Security. The results found demonstrated the effectiveness of the framework in relation to the aspects of privacy protection and fraud prevention. Furthermore, the results indicate that the good practices that already exist in the Data Governance, Data Privacy and Information Security disciplines can help to protect data privacy and, consequently, contribute to compliance with the brazilian data privacy protection law.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

1 INTRODUÇÃO

Vivemos na sociedade do conhecimento (*data-driven*) vislumbrada por Peter Drucker e Peter Senge [54, 162]. Neste cenário, os dados são considerados ativos importantíssimos, como estratégia de negócio, pela grande versatilidade de seu uso [14]. Entretanto, justamente devido às diversas possibilidades de uso dos dados e aos riscos sociais causados pelo uso indevido desses ativos, ampliam-se também os controles e processos formais para o seu tratamento [63]. Isso se dá através do fortalecimento regulatório, em processos de governança corporativa, de Tecnologia da Informação (TI) e de dados, como sentinelas da qualidade e segurança [15, 63, 70, 150]. Aliada à governança de TI, principalmente no que tange à proteção e privacidade de dados pessoais, está a legislação pertinente, que visa a resguardar este importante ativo das organizações. Nesse sentido, em âmbito nacional, temos as leis nº 12.527/2011 (Lei de Acesso à Informação) [40], nº 12.965/2014 (Marco Civil da Internet) [41] e a recente lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais (LGPD) [42].

É comum se pensar na regulação, em especial na governança, como um dispositivo "enrijecedor" que dificulta a correta exploração dos dados, quando na verdade, um dos objetivos de tais ferramentas é permitir a melhora da tomada de decisão e a redução dos atritos operacionais [63].

Iniciativas para o reconhecimento da importância da correta governança de dados (principalmente como facilitadora de processos de análise de dados em *Big Data*) têm se tornado mais frequentes [15, 51, 137]. Seguindo essa tendência, a LGPD [42], sancionada em agosto de 2018, prescreve a governança como uma ferramenta necessária ao tratamento de dados. Além disso, as crescentes preocupações com a proteção à privacidade e segurança das informações, fazem com que a governança seja uma ferramenta cada vez mais necessária e requisitada, não apenas para compliance legal, mas também para atendimento das demandas dos usuários [15, 42, 63, 150]. A regulação, então, surge tanto para o atendimento das demandas dos órgãos de controle e regulamentadores (*compliance* legal) quanto para o atendimento de demandas por privacidade e segurança pelos cidadãos e pelas organizações.

Nesse sentido, o papel dos profissionais que lidam com dados e os processos referentes a estes ativos deve ter como prioridade o controle de riscos operacionais de uma organização [63], levando em consideração aspectos de governança, privacidade e segurança. Isso reduz, a curto prazo, a capacidade e a produtividade dos processos que agregam valor ao produto ou serviço-alvo da organização, uma vez que parte dos esforços são dedicados à implantação de controles e prevenção de riscos [63]. Configura-se assim um aparente conflito entre dois interesses concorrentes: por um lado, um modelo de regulação, por meio da governança de dados e do *compliance* à LGPD, que institui controles e "travas" ao livre acesso a dados; por outro lado, o *advanced analytics*, principalmente no *Big Data*, onde deseja-se a maior quantidade e variedade de informação, extraindo valor exatamente do livre acesso e combinação de dados, por vezes ignorando a fonte ou

natureza desta mesma informação. Pretende-se neste trabalho demonstrar que não se trata, como pode parecer à primeira vista, de modelos conflitantes, mas sim de modelos interdependentes e de cooperação mútua.

Essa interdependência se torna particularmente clara quando levamos em consideração o objeto de estudo deste trabalho, qual seja, a análise de dados pessoais para fins de prevenção a fraudes. Neste contexto, o interesse governamental e institucional precisa se alinhar à proteção efetiva de dados pessoais, para evitar a violação de direitos dos usuários. De fato, se por um lado a prevenção à fraude é uma preocupação constante das organizações, interferindo nos riscos do negócio e por vezes até em sua viabilidade, por outro lado, esse tipo de prevenção não pode se dar de forma a violar os direitos dos consumidores que se utilizam de boa-fé dos serviços.

Ressalta-se que o tratamento de dados pessoais para fins de garantia da prevenção à fraude dispensa o consentimento dos titulares. Dessa forma, o titular pode ter seus dados analisados com a finalidade de prevenção à fraude mesmo que não autorize previamente. Além disso, esse é um dos pouquíssimos casos de tratamentos de dados autorizados por lei [42] que se utilizam de dados pessoais sensíveis. Esses dados sensíveis são especialmente protegidos pela legislação devido seu maior potencial de risco aos titulares, pois são os dados referentes, por exemplo, à origem racial ou étnica, convicção religiosa, opinião política, à saúde, vida sexual, etc.

Essas permissões legais evidenciam a preocupação e necessidade de que as instituições tenham todos os meios disponíveis para fiscalizar, prevenindo ou impedindo fraudes em seus sistemas, e ainda, fornecendo informações para se apurar e identificar os responsáveis por fraudes cometidas. Mas se, por um lado, a LGPD permite o uso de dados sensíveis para o tratamento de dados com o objetivo de prevenção a fraudes, por outro, a permissão legal não isenta o responsável pelo tratamento de agir de forma a preservar a segurança, a privacidade desses dados e o *compliance* à legislação.

Podemos ressaltar, inclusive, que algumas das alternativas legais para flexibilização da regulação precisam ser utilizadas de forma parcimoniosa nesse tipo de tratamento. É o caso, por exemplo, da anonimização. Ocorre que, há um falso entendimento por parte dos controladores de dados de que alguns mecanismos são condição única e suficiente para a proteção da privacidade de dados, principalmente dos dados pessoais [24, 28, 50, 122, 143]. Isso é particularmente evidente em contexto de *Big Data*, onde o grande volume de dados faz com que os controladores procurem alternativas à regulação.

Segundo Carvalho et al. [28], o problema do *compliance* à LGPD e da governança de dados no contexto do *Big Data* tem sido tratado de forma equivocada, principalmente pela má interpretação da lei que pousa sobre ferramentas, como por exemplo a anonimização, induzindo a uma expectativa irreal de proteção do dado pessoal sensível. Especialmente no tratamento de dados que busca a garantia da prevenção à fraude, muitas vezes, a saída interpretativa de flexibilização da regulação pela anonimização deve ser analisada de forma a não prejudicar a finalidade do tratamento. Desta forma, até mesmo para averiguação de que tipos de ferramentas de controle são viáveis de se utilizar no tratamento de prevenção a fraudes, é necessária uma análise mais acurada

da governança, segurança e privacidade que serão envolvidas nesse processo.

Neste trabalho é proposto um *framework* de governança de dados em *compliance* com a LGPD, com foco em *Big Data*, que objetiva um norte sobre o tratamento de prevenção a fraudes seguindo os mais recentes estudos acadêmicos sobre o assunto e a experiência profissional na área. Através das boas práticas combinadas da privacidade de dados, governança de dados e segurança da informação, os requisitos da LGPD, principalmente no que tange à privacidade de dados, serão supridos de forma consistente e sustentável, adaptando-se às necessidades específicas, em cada contexto, da governança em privacidade de dados ponta-a-ponta [137].

1.1 PROBLEMA DE PESQUISA

Este trabalho se debruça nas questões relacionadas ao tratamento de dados pessoais para fins de garantia da prevenção à fraude, analisando as nuances relacionadas ao compliance legal, à governança, à segurança e à privacidade envolvendo esse tipo de tratamento. O tratamento de dados para prevenção à fraude tem previsão no artigo 11, inciso II alínea g da LGPD [42], e é um dos permissivos legais ao tratamento de dados sensíveis, sem o consentimento do titular, conforme teor:

> Art. 11. O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses:
>
> I - quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas;
>
> II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para:
>
> [...]
>
> g) garantia da prevenção à fraude e à segurança do titular, nos processos de identificação e autenticação de cadastro em sistemas eletrônicos, resguardados os direitos mencionados no art. 9º desta Lei e exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais.

Essa previsão legal apresenta várias particularidades que serão destacadas no decorrer do trabalho. Entendendo a importância que o tema ganha na sociedade digital e a urgente necessidade de adequar essas práticas às delimitações legais trazidas pela LGPD, faz-se necessária uma compilação das melhores práticas das disciplinas associadas, a fim de gerar um guia consolidado para projetos de Big Data no contexto brasileiro. Entende-se que o problema apresentado é multidisciplinar pois além de complexo, envolve várias áreas de conhecimento. Podemos citar entre elas a governança de dados e a segurança da informação, expressas diretamente na lei LGPD [42] em seu capítulo XII, seções I e II. Tais disciplinas também são relacionadas ao problema da proteção à privacidade pela literatura da área [15, 23, 50, 63, 70, 84, 85, 86, 128, 137]. Também se inclui a disciplina de privacidade, tema central da lei, uma vez que concentra grande parte das práticas envolvidas no assunto.

A partir do levantamento dessas práticas, é possível se elencar as mais recomendadas pelos profissionais que lidam diretamente com o tratamento de dados, estruturando-as em uma ferramenta guia (*framework*) de orientação desse tipo de tratamento. Dessa forma, foi estabelecido como problema de pesquisa deste trabalho a seguinte questão: Quais as melhores práticas relacionadas à Governança de Dados, Privacidade de Dados e Segurança da Informação, recomendadas pelos profissionais de tecnologias da informação (que lidam com os dados pessoais) e pela literatura pertinente, podem ser elencadas e consolidadas para o tratamento de dados pessoais destinado à otimização da garantia da prevenção a fraude, no contexto de *Big Data*?

Para responder a esta questão de pesquisa, precisamos desmembrá-la em outras subquestões que serão tratados em duas etapas. A primeira etapa consiste no levantamento das melhoras práticas para o tratamento de dados, de forma ampla, de acordo com a literatura de cada disciplina elencada no problema central da pesquisa, repartindo-se em três subquestões conforme a seguir:

RQ.1 Quais são as boas práticas relacionadas à Governança de Dados no contexto de *Big Data* para identificação, classificação e tratamento de dados enquadrados no item g do inciso II do 11º artigo da LGPD?

RQ.2 Quais são as boas práticas relacionadas à Privacidade de Dados no contexto de *Big Data* para identificação, classificação e tratamento de dados enquadrados no item g do inciso II do 11º artigo da LGPD?

RQ.3 Quais são as boas práticas relacionadas à Segurança de Informação no contexto de *Big Data* para identificação, classificação e tratamento de dados enquadrados no item g do inciso II do 11º artigo da LGPD?

Em uma segunda etapa, buscamos selecionar as boas práticas mais recomendadas, conhecidas e aplicadas pelos profissionais que lidam diretamente com o tratamento de dados em *Big Data*. Propomos então a seguinte subquestão:

RQ.4 Quais são as boas práticas relacionadas à Governança de Dados, Privacidade de Dados e Segurança de Informação mais recomendadas pelos profissionais que lidam com tratamento de dados pessoais?

A partir dos resultados obtidos com os subproblemas de pesquisa, estabelecemos a base para a resposta do problema central. Como resultado, propomos um guia na forma de um *framework* voltado ao tratamento de dados pessoais para garantia de prevenção a fraude, a fim de nortear os profissionais de tecnologias da informação acerca das especificidades desse tipo de tratamento. Buscamos ainda realizar a validação do *framework* proposto, analisando se, de fato, essas boas práticas acarretam a melhoria nos processos de garantia de prevenção a fraude. Portanto, buscamos responder ainda a uma última subquestão, qual seja:

RQ.5 O *framework* consolidado a partir dessas boas práticas recomendadas pode otimizar os aspectos de proteção à privacidade e prevenção a fraude de processos de ingestão de dados em *Big Data*?

Com a resposta desta última questão de pesquisa buscamos averiguar se o *framework* construído realiza a otimização da garantia da prevenção a fraude, o fixando como consolidação do resultado obtido com o problema central de pesquisa.

1.2 JUSTIFICATIVA

Uma pesquisa divulgada neste ano (2021) pela Serasa Experian, denominada "Pesquisa Global de Identidade e Fraude 2021" [62], aponta que, desde o início da pandemia decorrente da COVID-19, houve um aumento de 26% no número de empresas que aplicam alguma estratégia de prevenção à fraude, totalizando, atualmente, 8 em cada 10 empresas. Especialmente no Brasil houve um aumento de 11% no investimento em prevenção à fraude entre junho de 2020 e janeiro de 2021.

Os números [62] refletem um aumento das preocupações acerca da garantia de confiabilidade dos negócios em um mundo onde o comércio se torna cada vez mais digital e onde operações fraudulentas podem ser realizadas com cada vez menos rastros. Nesse contexto, a garantia da confiabilidade de tais transações eletrônicas torna-se um dos pilares dos negócios digitais, que, se comprometidos, afetam a própria viabilidade da manutenção desse comércio. E não apenas na esfera econômica, mas também o âmbito penal é atingido por esse tipo de operação, que, em uma análise mais abrangente, compromete a segurança social de forma mais ampla e contundente.

Entendendo os efeitos nefastos que a fraude ocasiona para a segurança pública e para a segurança jurídica desses negócios, a LGPD [42] permite que a garantia de prevenção à fraude disponha de todos os meios possíveis, inclusive do acesso aos dados pessoais sensíveis, a fim de se perseguir a segurança efetiva dessas transações. O desafio, no entanto, estende-se quando levamos em consideração um volume cada vez maior de operações, acarretando em maior complexidade para análise de dados de modo a se prevenir fraudes ou identificá-las a *posteriori*.

De acordo com Brasher, Domingo-Ferrer, Mehmood et. al. e Piras [24, 50, 122, 137], o problema do tratamento de dados no contexto de bases massivas é latente, crescente e complexo, e deve receber atenção das instituições que lidam com esses grandes volumes de dados, principalmente quanto aos riscos à privacidade dos usuários. Desta forma, se por um lado há uma demanda crescente por maior segurança nas operações, apontando para um uso mais flexível dos dados para se apurar fraudes, outros aspectos como a privacidade não podem ser ignorados no tratamento de dados, principalmente em bases de dados massivas.

Assim, neste trabalho será realizada uma discussão da legislação relacionada à proteção da privacidade de dados, com o intuito de enriquecer o arcabouço das instituições com ferramentas de segurança, gerenciamento de riscos e governança de dados. Entende-se também que, em sua grande maioria, as organizações não possuem suas bases devidamente organizadas a fim de se protegerem contra fraudes, levando em conta o respeito à privacidade de dados, principalmente no Brasil, onde a legislação que pousa olhar sobre o tema ainda é muito recente. E mesmo quando a legislação é mais antiga, como no caso da GDPR, a lei é vista como incipiente, segundo a opinião das próprias empresas. É o que revelou a Conferência Mundial de Dados de 2016, em Delray

Beach, na Flórida, onde a recentemente editada GDPR ganhou foco de discussão [15], ou ainda na conferência do ano seguinte (novembro de 2017), em San Diego. Ambas as conferências tinham por objetivo reunir as empresas referências em análise de dados para discussão de problemas e soluções da disciplina.

Outro aspecto que se deve avaliar é que a aplicação de uma abordagem de governança de dados e de gestão de riscos deve impactar positivamente não só a privacidade de dados, mas também os aspectos relacionados à governança e à gestão de riscos em si. Espera-se, portanto, que esta abordagem mais abrangente minimize a ocorrência e os problemas decorrentes de "Pântanos de dados", que são grandes massas de dados heterogêneas, indisponíveis na velocidade requerida, complexa de ser explorada, com má gestão de metadados e sem um modelo de dados adequado [22].

Ainda sobre a exploração do problema da *compliance* à legislação, mais do que a implantação de controles, tratamento de bases e de dados, neste período de adequação, as empresas possuem uma janela de oportunidade para revisar e aprimorar a segurança e a qualidade de seus dados, sejam eles pessoais ou não. Através da revisão dos processos de governança de dados, implantação de iniciativas de gerenciamento de dados mestres e de referência (*Master Data Management* (MDM) [23]), ou da revisão das políticas de acesso e autorização de tratamento de dados, as organizações têm em frente uma brecha para assegurar uma melhor utilização dos ativos de dados pessoais.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é identificar as melhores práticas de Governança de Dados, de Privacidade e de Segurança da Informação no contexto de *Big Data* para a promoção do *compliance* à LGPD, dispostas pela academia em artigos mais recentes, e eleger entre elas, as práticas que, segundo os profissionais que lidam cotidianamente com o tratamento de dados, melhor atendam às necessidades do tratamento de dados para garantia de prevenção a fraudes dispostas na alínea g do inciso II do artigo 11º da mencionada lei. Além disso, iremos propor um *framework* que contemple as disciplinas de Governança de Dados, Privacidade de Dados e Segurança da Informação para nortear os profissionais de tecnologias da informação e demais que lidam com o desenvolvimento do tratamento de prevenção a fraude.

1.3.2 Objetivo Específico

Para atingir o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

1. Realizar uma revisão de literatura em relação aos conceitos de Governança de Dados, LGPD, Privacidade de Dados e Segurança da Informação.
2. Identificar e definir os elementos que irão compor o *framework*;

3. Identificar e/ou definir um modelo adequado de avaliação de Qualidade e Privacidade de dados no contexto proposto;
4. Aplicar o *framework* e o modelo de avaliação em um projeto (estudo de caso prático em uma organização financeira);
5. Realizar os ajustes no *framework*, caso necessário.

1.4 RESULTADOS ESPERADOS

O *framework* objeto deste trabalho se propõe a ser um acelerador de projetos de *Big Data* para tratamento de dados voltados à prevenção a fraude em *compliance* à LGPD, pois compila as melhores práticas das disciplinas de Governança de Dados, LGPD, Privacidade de Dados e Segurança da Informação. Portanto, de forma mais específica, a pesquisa busca se aprofundar nas nuances do tratamento de dados para garantia de prevenção à fraude, descrita na LGPD, propondo um guia com as melhores práticas de governança, segurança e privacidade em *compliance* com a legislação brasileira de proteção de dados.

Além disso, utilizando o *framework* como base, espera-se obter um facilitador para a revisão dos processos de governança de dados, segurança da informação e privacidade de dados de forma mais ampla, que é apresentado como melhor alternativa para o *compliance* à LGPD. Desta forma, o *framework* apresentado ainda visa ser um repositório para consulta dos principais modelos das disciplinas mencionadas, a fim de que equipes menos experientes e projetos diversos do contexto de *Big Data* tenham um suporte didático de melhores práticas.

Esta pesquisa pretende, então, tornar-se um repositório para indexar os trabalhos mais atuais de boas práticas das disciplinas abordadas, produzindo um local de pesquisa para profissionais da área. Ademais, almeja-se contribuir com a discussão sobre as limitações da ferramenta de mascaramento e anonimização de dados a fim de que o uso dessas técnicas seja corretamente apoiado pelas boas práticas das disciplinas abordadas. Espera-se também contribuir para a minimização das lacunas de definições (até o momento existentes) de aspectos do tratamento de dados necessários para o correto cumprimento da lei LGPD [42].

1.5 METODOLOGIA DE PESQUISA

Esta pesquisa seguirá a metodologia *Design Science Research* [181], que é o método comumente utilizado para documentar o acúmulo de conhecimento gerador de artefatos inovadores para as áreas do conhecimento humano. Segundo Vaishnavi et al. [181], em áreas como os Sistemas de Informação, onde a ótica de pesquisa se demonstra multiparadigmática, o conhecimento flui de forma cíclica, interferindo nos processos de produção e sendo consolidados a partir da observação destes referidos processos.

Esta pesquisa também fará uso da técnica de estudo de caso, uma vez que segundo Yin [191]

é a técnica ideal para compreender um fenômeno complexo em seu contexto real de aplicação. Assim, conforme a definição de Dresch [53], o estudo de caso será usado para coletar dados a fim de estabelecer uma base conceitual lógica para a proposição do framework. De acordo com Vaishnavi et al. [181], uma possível metodologia formal para aplicação do *Design Science Research* compreende cinco fases principais: 1. Conscientização do Problema; 2. Sugestão; 3. Desenvolvimento; 4. Avaliação; 5. Conclusão (Figura 1.1). Ao aplicar o *DSR* na presente pesquisa, citaremos as ações de cada fase.

Primeiramente, para a correta identificação do problema, levantar-se-ão por meio de uma revisão da literatura correlata, os principais modelos que compõem o arcabouço de boas práticas nas disciplinas de Governança de Dados, Segurança da Informação e Privacidade, enfatizando modelos com aplicação documentada em projetos de *Big Data*. O objetivo desta etapa é avaliar as ferramentas disponíveis aos profissionais de TI para o alcance do objetivo proposto pela LGPD [42] em seu artigo 11, inciso II, alínea g, a saber, identificação de dados relacionados à prevenção de fraude e à segurança do titular, nos processos de identificação e autenticação de cadastro de sistemas eletrônicos. Ainda como parte da Conscientização do Problema, será aplicado um questionário entre os profissionais de TI e demais, na etapa de desenvolvimento de software, a fim de levantar os principais problemas relacionados ao supracitado artigo da LGPD [42], angariando a experiência destes profissionais, principais lacunas metodológicas e boas práticas experimentadas e aplicadas.

A partir deste conjunto de dados, combinando boas práticas sugeridas e aplicadas em projetos de *Big Data*, será proposta uma versão inicial do *framework*, compilando as práticas priorizadas na fase anterior. O desenho deste artefato também será acompanhado da definição das métricas que serão usadas (em um projeto real) para validar o atingimento do objetivo. Serão usadas métricas de Qualidade de Dados e Privacidade. Como forma de desenvolvimento do *framework* proposto, será eleito um projeto piloto em uma instituição financeira, a fim de aplicar no processo de ingestão e catalogação de dados, o artefato resultado da fase anterior e, de forma iterativa incremental, evoluí-lo. Caso sejam necessários ajustes nas definições das métricas usadas nesta validação, ou até proposições de novas métricas, tal desenvolvimento será realizado também nesta fase.

A fim de validar a aplicabilidade do *framework*, uma ingestão de dados fora do projeto estudo de caso (semelhante em critérios de massa de dados) será submetida às mesmas métricas citadas anteriormente, possibilitando uma análise comparativa. Por fim, uma análise das métricas obtidas confrontadas com os dados reunidos na fase de Conscientização do Problema será consolidada, explicitando os benefícios obtidos com o uso do *framework* no projeto eleito. Considerações acerca da especificidade do projeto em questão serão abordadas nesta fase, de forma a avaliar a possível generalização dos resultados obtidos. Um diagrama explicativo das fases de condução da pesquisa é apresentado na Figura 1.1.

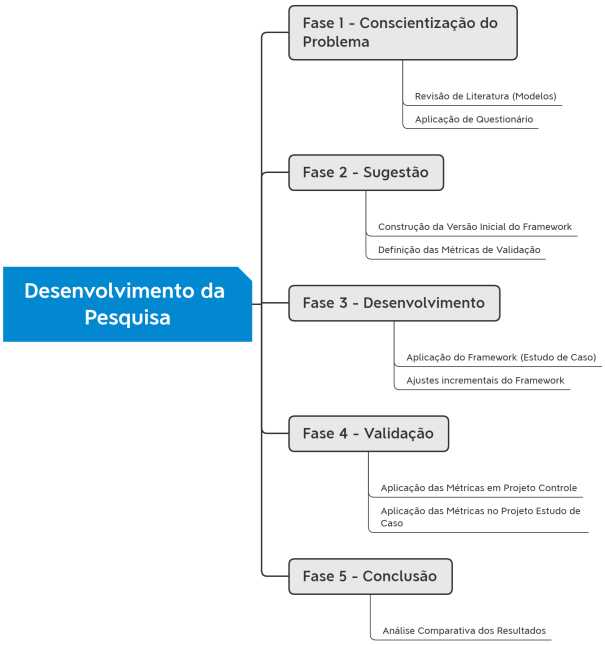

Figura 1.1: Metodologia utilizada nesta pesquisa. (Fonte: Autor)

1.6 PUBLICAÇÕES RESULTANTES DESTA PESQUISA

Este trabalho resultou em diversas publicações, a saber:

1. Potiguara Carvalho, A., Potiguara Carvalho, F., Dias Canedo, E., Potiguara Carvalho, P. H. (2020, June). Big Data, Anonymisation and Governance to Personal Data Protection. In The 21st Annual International Conference on Digital Government Research (pp. 185-195).
2. CARVALHO, Artur Potiguara et al. Anonymisation and Compliance to Protection Data: Impacts and Challenges into Big Data. In: 22th International Conference on Enterprise Information Systems. 2020. p. 31-41.
3. Sâmmara Éllen Renner Ferrão, Artur Potiguara Carvalho, Edna Dias Canedo, Alana Paula Barbosa Mota, Pedro Henrique Teixeira Costa, Anderson Jefferson Cerqueira: Diagnostic of Data Processing by Brazilian Organizations - A Low Compliance Issue. Inf. 12(4): 168 (2021)
4. CARVALHO, Artur Potiguara et al. Anonimisation, Impacts and Challenges into Big Data: A Case Studies. In Enterprise Information Systems: 22nd International Conference, ICEIS 2020, Virtual Event, May 5–7, 2020, Revised Selected Papers (Vol. 1, p. 3). Springer Nature.

1.7 ESTRUTURA DA DISSERTAÇÃO

Este trabalho está dividido em 4 Capítulos, além deste capítulo introdutório.

1. Capítulo 2 - Referencial Teórico: Capítulo que será destinado a apresentar os principais trabalhos acadêmicos da área correlata, dividindo o tema nas três disciplinas principais:
    - Privacidade de Dados: será apresentada uma análise crítica da norma ISO 27701 (sobre controles de segurança relacionados à privacidade de dados pessoais) e sua relação com a LGPD;
    - Segurança da Informação / Anonimização: apresenta uma discussão sobre segurança da informação no contexto de *Big Data*, anonimização como uma ferramenta da segurança de dados e privacidade nesse contexto e algumas reflexões e implicações da dissociação dos dados pessoais do titular em bases massivas.
    - Governança de Dados: dedica-se ao alinhamento dos termos utilizados durante esta pesquisa referentes à disciplina de governança de dados. Serão apresentados os principais modelos consolidados pela literatura e em aplicação nas empresas de TI.
2. Capítulo 3 - Metodologia: Capítulo onde será apresentada a metodologia utilizada para conduzir esta pesquisa em cada uma de suas etapas. Serão apresentados os modelos adotados para coleta de dados, levantamento de indicadores, levantamento e validação dos componentes do *framework* e métodos de análise dos dados obtidos;
3. Capítulo 4 - Análise dos Dados e Resultados: Capítulo referente à apresentação e detalhamento da versão final do *framework*, com especificação dos seus componentes, papéis e responsabilidades, além da aplicação do *framework* proposto em um projeto estudo de caso e da discussão da validação do modelo ante à aplicação das métricas definidas;
4. Capítulo 5 - Conclusão: Capítulo final onde apresentamos as considerações finais sobre o *framework*, a visão de *compliance* à LGPD e Governança de Dados, o projeto estudo de caso e trabalhos futuros;

2 REFERENCIAL TEÓRICO

Essa pesquisa está estruturada no estudo sobre três importantes disciplinas que são levadas em consideração no *compliance* legal do tratamento de dados pessoais (DP), quais sejam: a Governança, a Privacidade e a Segurança da Informação. Norteamos o trabalho buscando a resolução da questão central e das subquestões de pesquisa, as quais se relacionam diretamente às três mencionadas disciplinas.

Este capítulo apresenta uma breve introdução acerca da lei de proteção de dados brasileira (LGPD), apresentando os principais conceitos e os dispositivos relacionados ao tratamento de dados na prevenção a fraude, além dos marcos teóricos relacionados à matéria. O capítulo busca ainda introduzir os conceitos e os principais marcos teóricos relacionados às três disciplinas que são os pilares desta pesquisa, apresentando as principais discussões que envolvem a Governança de Dados (GD), a Privacidade e a Segurança da Informação. Para melhor organizar essas discussões, trataremos as disciplinas em seções específicas. Ressaltamos, no entanto, que, a partir da vigência da LGPD, todas essas disciplinas foram, em alguma medida, norteadas pela lei, e precisam se adequar às suas previsões, além de aprimorar sua aplicação para atingir os objetivos e princípios legalmente estipulados.

A Lei Geral de Proteção de Dados - LGPD, Lei nº 13.709/18 [42], seguindo as demandas nacional e internacional, consolida-se como uma legislação que direciona o tratamento de dados no Brasil, reforçando e concretizando diretrizes já introduzidas pelas leis nº 12.527/2011 (Lei de Acesso à Informação) [40] e nº 12.965/2014 (Marco Civil da Internet) [41]. A legislação brasileira é ainda fortemente influenciada pelo Regulamento Europeu de 2016 (GDPR) [151], que se traduz como um marco internacional na proteção à privacidade dos titulares de DP.

Neste trabalho, seguindo a conceituação trazida pela LGPD (Art. 5, inciso I [42]), tratamos DP como o conjunto de informações relacionadas a um indivíduo, pessoa natural identificada ou identificável. Por sua vez, a definição de DP sensível, é dada como [42]:

> "DP sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural."

A lei LGPD [42] também define os papéis envolvidos na proteção ao DP, sendo eles:

- **Titular**: pessoa natural a quem se referem os DP que são objeto de tratamento;
- **Controlador**: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de DP;

- **Operador**: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de DP em nome do controlador;
- **Encarregado**: pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD);
- **Agente de tratamento**: o controlador e o operador;

Segundo Piras et. al. [137], devido ao extremo avanço das tecnologias tais como *Big Data*, Computação em Nuvem e Internet das Coisas (*IoT*), e o aumento da chance de mal uso dos dados de cidadãos, a privacidade de dados deve ser garantida, tanto pela legislação tal como nos exemplos da LGPD [42] e GDPR [151], como pelas organizações, aplicando métodos e processo preconizados pela literatura para a proteção da privacidade de dados. Como mencionamos, o enfoque deste trabalho é o tratamento de dados que tem por objetivo a garantia de prevenção à fraude, que trata o art. 11, II, "g" da LGPD.

Becker et. al. [16], definiram fraude como o ato de enganar outros para benefício próprio. A fraude, tão antiga quanto o costume de viver em sociedade, encontrada em diferentes naturezas, possui sempre em comum o emprego da desonestidade com o intuito de convencer uma parte inocente da legitimidade de determinada transação [16]. Encontrada principalmente no meio financeiro, a fraude ainda ocorre por questões políticas, prestígio pessoal, autopreservação ou disputa de poder [16]. Entendemos, portanto, que a fraude no contexto tecnológico, possui um aspecto de forjar ou evitar uma identificação, tendo o efeito de "falsidade ideológica" na aplicação de sistemas de informação. Quando os atacantes se passam por um usuário legítimo ou não se identificam a fim de usufruir de privilégios e recursos anteriormente lhes negados, então temos um episódio onde ocorre uma fraude.

Bolton e Hand [21] diferenciam a prevenção a fraude da detecção de fraude. Para os autores [21], a prevenção a fraude descreve medidas para evitar a ocorrência de uma fraude enquanto a detecção a fraude descreve medidas para cessar uma fraude em curso. Prevenção a fraude pode incluir iniciativas como projetos, fibras fluorescentes, desenhos multitons, marcas d'água, tiras de metal luminoso, hologramas em cédulas bancárias, identificações positivas bancárias, sistemas de segurança na internet para transações de cartão de crédito, chips SIM (*Subscriber Identity Module*) de telefonia móvel, senhas computacionais e contas de telebanco [21].

Na literatura, o conceito de fraude muitas vezes está ligado ao conceito de autenticidade e autenticação. Segundo Goodrich e Tamassia, autenticação é um conceito ligado à confidencialidade sob o qual há uma determinação de identidade ou papel de alguém [70]. Geralmente é baseada numa combinação de algo que o indivíduo autenticado tem (cartão inteligente, dispositivo de chaves secretas), ou algo que o indivíduo sabe (senha) ou ainda algo que ele é (dados biométricos).

O conceito de autenticidade também está relacionado a uma propriedade desejável dos sistemas: a de não repúdio. Por sua vez, Goodrich e Tamassia [70] conceituam não repúdio como a propriedade que afirmações autênticas emitidas por alguma pessoa ou sistema não podem ser negadas aos

seus consumidores legítimos. O conceito de não repúdio, envolve, entre outros fatores, a propriedade de disponibilidade, sob a qual a informação deve ser acessível e modificável no momento oportuno por aqueles que estejam autorizados a fazer isso [70]. Os autores afirmaram que uma informação extremamente protegida e praticamente inacessível (como uma informação encarcerada em um cofre de ferro fundido em uma montanha tibetana e protegida em tempo integral por um esquadrão de ninjas devotados) pode ser segura, mas tem grande potencial de ferir a disponibilidade [70]. Os autores também afirmaram que a qualidade de uma informação está diretamente associada à sua disponibilidade [70]. Como podemos relacionar disponibilidade ao não repúdio e este à autenticidade, podemos concluir que uma boa forma de mensurar a capacidade de um conjunto de informações de prover insumos às análises de fraude reside em métricas de qualidade de dados.

Um segundo conceito ligado à fraude é o de autorização que, segundo Goodrich e Tamassia [70], define se certo indivíduo tem permissão de acessar recursos com base em uma política de controle de acessos. Os autores ainda fazem uma ressalva que põe em foco o resultado de uma fraude, no trecho em que "Tal autorização deve evitar que um atacante engane o sistema [fraude] para que este permita o seu acesso a recursos protegidos" [70].

É interessante mencionar que Goodrich e Tamassia [70], ao discorrer sobre as características de sistemas de coleta de biometria, relatam que mesmo nestas tecnologias é possível a presença de fraude pois a característica física considerada identificável biometricamente pode ser forjada ou evitada. Isso revela como nos mais diversos sistemas de identificação, mesmo nos considerados mais sofisticados, há vulnerabilidade quanto à fraude. As principais estratégias estabelecidas para o combate e prevenção a fraude estão intimamente ligadas à classificação dos tipos de fraudes empregadas pelos fraudadores. Sobre este assunto, Becker et. al. [16], classificam os tipos de fraude como:

1. **Fraude de Subscrição**: ocorre quando o fraudador assina um serviço como o usuário legítimo, sem a intenção de pagar pelo uso;
2. **Fraude de Intrusão**: ocorre quando o fraudador consegue acesso a uma conta outrora legítima e assina serviços através dela. Difere-se da anterior porque o usuário legítimo pode estar consumindo os serviços legítimos de forma simultânea à intrusão;
3. **Fraudes baseadas em lacunas tecnológicas**: geralmente envolve a exploração de uma má configuração de um serviço tecnológico como uma senha inadequada ou sistemas de defesa permissivos;
4. **Engenharia Social**: ocorre em contraponto à exploração de lacunas tecnológicas, sondando a interação humana, convencendo as pessoas que operam os sistemas a flexibilizar controles e regras de segurança;
5. **Fraudes baseadas em novas tecnologias**: ocorre quando os fraudadores exploram o desconhecimento de suas vítimas de tecnologias recentes, obtendo vantagem da ignorância de suas vítimas;

6. **Fraudes baseadas em novas regulações**: por vezes, regras e normas que buscam a igualdade e justiça acabam sendo exploradas por fraudadores para obter vantagens (em sua maioria, econômicas);

7. **Fraude de identidade de usuário**: representada como o roubo de perfis em redes sociais, cópias de números de cartão de crédito, clonagem de chip da telefonia móvel, hackeamento de sistemas de identificação biométricas ou qualquer artifício que possibilite o fraudador simular a identidade de um usuário legítimo;

Dessa forma, a identificação dos tipos de fraudes auxilia na concepção de estratégias preventivas ou na melhor identificação de fraudes e reparação de lacunas para se evitar novos eventos. Nesse aspecto de soluções de combate e prevenção a fraude, a literatura é extensa ([16, 21, 70, 125, 155, 180] e pode envolver inclusive normas e políticas alheias ao universo da tecnologia, tais como a Lei 7492/86 [38] que trata de crimes contra o sistema financeiro nacional, ou a lei de prevenção à lavagem de dinheiro (Lei 9613/98) [39].

Entendendo os riscos que envolvem a fraude, a Lei de Proteção de Dados Brasileira - LGPD [42], a qual dedicamos maior atenção neste trabalho, permite uma ampla gama de ferramentas de proteção ao viabilizar a utilização de dados sensíveis para esse tipo de tratamento, ainda que não haja o consentimento do titular, conforme teor:

> Art. 11. O tratamento de DP sensíveis somente poderá ocorrer nas seguintes hipóteses:
>
> I - quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas;
>
> II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para:
>
> [...]
>
> g) garantia da prevenção à fraude e à segurança do titular, nos processos de identificação e autenticação de cadastro em sistemas eletrônicos, resguardados os direitos mencionados no art. 9º desta Lei e exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos DP.

É importante ressaltar que a lei expressamente prevê que os direitos do art. 9º devem ser preservados no tratamento de prevenção à fraude. O artigo 9º detalha a aplicação do princípio do livre acesso, que é um dos princípios que regem o tratamento de DP. Na verdade, todos os princípios elencados no art. 6º devem ser observados pelo controlador no momento de tratamento de DP, entretanto, o legislador resolveu ressaltar que, para o tratamento de prevenção a fraudes, deve ser particularmente observado o princípio do livre acesso.

Segundo o art. 6º, IV, o princípio do livre acesso se traduz como a "garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus DP". É necessário pontuar, no entanto, que o princípio do livre acesso, apesar de

ser de essencial aplicação, deve ser respeitado de forma a não exaurir o objetivo do tratamento, que é justamente a prevenção ou identificação de atos fraudulentos. Nesse sentido, o direito de acesso do titular a informações sobre o tratamento de seus dados não pode ser utilizado para criar obstáculos ao tratamento. Até porque, como enfatizamos, o tratamento de dados realizado para prevenção à fraude independe do consentimento do titular. Os demais princípios aplicáveis a esse tratamento de dados estão dispostos no art. 6º e são especificados a seguir:

- **Princípio da finalidade**: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;
- **Princípio da adequação**: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;
- **Princípio da necessidade**: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;
- **Princípio da qualidade dos dados**: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;
- **Princípio da transparência**: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;
- **Princípio da segurança**: utilização de medidas técnicas e administrativas aptas a proteger os DP de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;
- **Princípio da prevenção**: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de DP;
- **Princípio da não discriminação**: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;
- **Princípio da responsabilização e prestação de contas**: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de DP e, inclusive, da eficácia dessas medidas.

Destacamos, no contexto brasileiro de *compliance* à LGPD e da extração de valor de projetos de *Big Data*, o princípio da necessidade, correspondente ao princípio da *Data Minimization* [135] no contexto europeu, o qual exige a minimização da coleta e tratamento de DP ao estritamente necessário. Por seguinte, o tempo de armazenamento do DP coletado também deve ser minimizado. Dessa forma, para *compliance* do tratamento de dados de prevenção à fraude, apesar de dispensado o consentimento do titular, é necessário o cumprimento dos princípios expostos.

Podemos ressaltar que uma das grandes preocupações da LGPD [42] se dá justamente na garantia da privacidade dos cidadãos. Isso se torna explícito logo em seu primeiro artigo, que prevê que o objetivo da lei é "proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural". Assim sendo, abordaremos as nuances legais e teóricas acerca da privacidade, que se consubstancia em um dos pilares de investigação da presente pesquisa.

2.1 PRIVACIDADE

De acordo com Goodrich e Tamassia [70], com o uso cada vez mais popular das redes sociais, as pessoas têm tornado seus dados (principalmente os pessoais) visíveis, pelo menos para uma parcela da internet. Segundo os autores, de forma subversiva, alguns elementos na rede, considerando os fragmentos dispersos, têm a capacidade de recriar de forma assustadoramente completa os perfis pessoais. Por esse motivo, crescem as demandas por maior controle do titular sobre os seus próprios dados, de forma a proteger sua privacidade. Este controle pode se estender sobre questões como quais são as informações disponíveis, ou quem está autorizado a acessá-las, propriedades que por padrão sejam restritivas (a fim de que um usuário desatento ou desinformado não autorize sem ciência a divulgação de seus dados - conhecido como o conceito de *"Privacy by Default"*) ou ainda construção de políticas claras de compartilhamento de informação, para que todo usuário tenha total entendimento da finalidade para qual seus dados serão usados por terceiros (propaganda digital, análise de sentimento, entre outros) [70].

A LGPD, seguindo a tradição internacional, trata desse controle pessoal dos dados como autodeterminação informativa, ou seja, a capacidade do titular de ter um grau de autodomínio sobre seus dados, sendo este um dos fundamentos da proteção de DP (art. 2º, II [42]).

Goodrich e Tamassia [70] também fazem menção ao fato de que nem sempre a privacidade tem sido ameaçada apenas por campos considerados sensíveis e que, através da lista de amigos de certo perfil em uma rede social nos quais haja consentimento da utilização de DP, é possível predizer, com certo grau de precisão, informações sobre esta pessoa tais quais religião, raça, gênero, faixa etária e orientação sexual. E mais, os perfis em diversas redes distintas podem ser mapeados apenas ao se comparar as listas de amigos. De certa forma então, a concessão de uso de informações pessoais pode acarretar na disponibilidade de informações pessoais de toda a lista de amigos e conexões relacionadas.

Atualmente, no escopo de normas internacionais, a NBR/ISO 27701 [86] tem o objetivo de estabelecer, instituir, manter e melhorar continuamente o Sistema de Gestão da Privacidade da Informação (SGPI). Este último trata das questões referentes a riscos e controles de DP. Além disso, a norma define requisitos e guias para auxiliar o tratamento de DP pelos controladores responsáveis e transparência no processamento de tais dados. O SGPI é integrado ao Sistema de Gestão da Segurança da Informação (SGSI), definido na NBR/ISO 27001 [85] e, portanto, as normas 27701 e 27001 são correlacionadas. A norma ainda se direciona a controladores (inclusive de DP

compartilhados) e operadores de DP. Em seus anexos, a norma NBR/ISO/IEC 27701 [86] traz um comparativo dos artigos da lei LGPD e suas diretivas. Nos parágrafos a seguir apresentamos um estudo sobre esta correlação.

2.1.1 NBR/ISO/IEC 27701 e LGPD

Não por coincidência, o primeiro aspecto tratado na norma NBR/ISO 27701 [86] com referência à LGPD é o artigo 50, sobre a governança e boas práticas. Percebe-se então como ambos os documentos dão ênfase à organização processual dos tratamentos de dados com impacto direto nas questões de proteção à privacidade de dados. Segundo a norma 27701 [86], o requisito presente na norma irmã 27001[85], seção 4.1 (a saber, "Entendendo a organização e seu contexto") deve ser expandido, a fim de também autodeterminar o papel da organização como controladora/operadora de DP (DP), incluindo tais definições na concepção do SGPI. Esse requisito representa a atividade de priorização dos aspectos internos e externos próprios da organização que devem ser considerados na concepção do SGSI.

A seção 4.2 ("Entendendo a necessidade e as expectativas das partes interessadas") da norma 27001 [86], onde se definem os envolvidos relevantes e como eles influenciam a implementação do SGSI, deve ser expandida para abarcar os titulares de DP, além das partes que tenham interesse/responsabilidades em tais dados. Tal requisito, ainda que não exatamente delimitado para este fim, pode significar uma sugestão da norma ao uso de métodos de *Privacy by Design*, onde a construção dos sistemas já envolve considerações acerca da privacidade [15].

Ainda envolvendo aspectos introdutórios à governança e boas práticas, o requisito presente na norma 27001 [85], em sua seção 4.3, tal que define e prioriza um escopo (limite de atuação) para o SGSI, inclusive mantendo a comunicação deste escopo definido, deve se expandir para incluir (necessariamente) o tratamento de DP. Esta definição de escopo é crucial, dado que os recursos e insumos disponíveis para implementação de ações de segurança e privacidade são finitos e, dependendo do contexto da organização, escassos.

Um segundo aspecto presente na LGPD e tratado pelo conjunto de normas 27001/27701 é a gestão de riscos. Alguns artigos da lei [42] chegam a definir artefatos a serem produzidos que comprovam a adoção de práticas de gestão de riscos. Um desses artefatos é o Relatório de Impacto a Proteção de DP (RIPD), que segundo a própria LGPD:

> XVII - relatório de impacto à proteção de DP: documentação do controlador que contém a descrição dos processos de tratamento de DP que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco.

No prisma da norma 27701 [86] e da 27001 [85], a avaliação de riscos deve ser um processo que define critérios de aceitação e de desempenho da avaliação de riscos da segurança da informação de forma contínua, produzindo resultados comparáveis, válidos e consistentes. Além disso, os

riscos devem ser identificados, assim como seus responsáveis, e analisados, quanto às suas consequências potenciais, probabilidades realísticas e seus níveis de riscos. Por fim, os riscos devem ser tratados, de acordo com a priorização obtida da análise anteriormente citada, baseados em um processo organizado que selecione de forma apropriada a opção de tratamento mais adequada do risco e defina e implemente todos os controles necessários para tratar o risco. Para o tratamento do risco também é necessário instituir e aprovar: um plano de tratamento do risco (a ser apreciado pelos responsáveis pelo risco); tratamento dos riscos; e aceitação de riscos residuais. Em caráter exemplificativo, ambas as normas [85, 86] relacionam os riscos às dimensões da qualidade de dados, tais como a perda de confidencialidade, integridade e disponibilidade da informação. Podemos então fazer um comparativo entre os riscos e a qualidade de dados. Neste contexto de projetos de dados, um risco seria um evento incerto que impacta direta ou indiretamente na qualidade dos dados, e sua identificação pode estar associada aos controles e medições da qualidade de dados.

Seguindo o mapeamento realizado pela norma 27701 [86] em seu anexo informativo, o RIPD pode ser influenciado por atividades de orientação da direção para segurança da informação através da definição/manutenção da política para segurança da informação. A norma 27002 [84], principalmente na seção 5, trata dos principais aspectos e controles para definição de políticas de segurança da informação. Segundo a norma [84], uma instância deliberativa da organização que possua autoridade decisória deve estabelecer a abordagem corporativa para gerenciar os objetivos da segurança da informação na organização.

Segundo esta mesma definição, uma política de segurança da informação deve definir a segurança da informação no contexto da organização, determinando objetivos e princípios que orientem as atividades de segurança da informação, definindo papéis e responsabilidades, além de determinar também o processo de tratamento de desvios e exceções. Para exemplificar, a norma [84] traz como alguns temas recorrentes dessa política: 1. controle de acesso; 2. classificação e tratamento da informação; 3. segurança física e do ambiente; 4. uso aceitável dos ativos; 5. "mesa limpa e tela limpa"; 6. transferência de informações; 7. dispositivos móveis e trabalho remoto; 8. restrições sobre o uso e instalação de *software*; 9. backup; 10. proteção contra *malwares*; 11. gerenciamento de vulnerabilidades técnicas; 12. Controles criptográficos; 13. segurança nas comunicações; 14. proteção e privacidade da informação de identificação pessoal; 15. relacionamento na cadeia de suprimentos.

Decorre da norma 27002 [84], em sua seção 6.1.1, que as responsabilidades e papéis envolvidos com a segurança da informação (e estendidos à proteção da privacidade de dados pela norma 27701 [86]) devem ser definidos e atribuídos. Para tanto, a norma [84] sugere que os ativos e processos de segurança da informação sejam identificados e definidos; que as entidades responsáveis por cada ativo sejam identificadas e documentadas; que os níveis de autorização sejam claramente definidos e documentados; que as pessoas responsáveis sejam competentes e capazes de cumprir com as responsabilidades, inclusive mantendo-se atualizadas com os desenvolvimentos; e que a coordenação e a visão global dos aspectos de segurança da informação na cadeia de suprimento sejam identificadas e documentadas.

A norma 27701 [86] adiciona a instituição do equivalente na LGPD [42] ao encarregado (em seu artigo 41), sendo ele um ponto de contato dos titulares de DP com relação aos tratamentos de tais dados. No rol de sugestões para escolha do encarregado presente na seção 6.3.1.1 da norma 27701 [86] tem-se: 1. seja independente e assegure uma gestão de riscos de privacidade livre de vieses de outras áreas; 2. seja envolvido em todas as questões relacionadas ao tratamento de DP; 3. possua conhecimento aprofundado em legislação e regulamentação pertinente, além de experiência na prática de proteção de dados; 4. atue como ponto de contato junto às áreas de auditoria; 5. desenvolva ações de conscientização da Alta Direção e dos empregados quanto às obrigações e riscos do tratamento de DP; 6. desenvolva ações com caráter informativo/educativo quanto a avaliações de impacto de privacidade.

As normas (27701 e 27002 [84, 86]) dedicam um tópico para discorrer sobre vulnerabilidades que devem ser consideradas quando os dados são acessados através de dispositivos móveis ou via estações de trabalho remoto. Isto ocorre porque, em ambos os casos, envolve-se de forma mais visceral o tráfego dos dados pela rede mundial de computadores. Portanto, neste cenário, é recomendável que sejam observados cuidados especiais, que as normas definem como "ambientes desprotegidos". Este cuidado especial reforça os requerimentos presentes na LGPD [42] em seu capítulo VII. Outro aspecto importante a se considerar é a conscientização (já mencionada no rol de atribuições do encarregado) de todos os colaboradores da organização quanto ao tratamento de DP. De fato, principalmente se levado em conta o pouco tempo em que essas questões têm ganhado foco nos debates sobre segurança, a aculturação dos princípios de proteção e privacidade tem que ser alavancada por ações de conscientização, educação e treinamento em segurança da informação e privacidade.

Segundo a norma 27002 [84], estes treinamentos devem contemplar aspectos gerais como: comprometimento da organização com segurança da informação e privacidade; *compliance* com políticas, normas, leis, regulamentações, contratos e acordos pertinentes; responsabilização pessoal e comprometimento com a manutenção da proteção à informação; procedimentos e controles básicos de segurança da informação e privacidade; e pontos de contato e recursos para informações adicionais.

Um dos aspectos mais relevantes presentes nas normas e sombreado pela LGPD é a classificação da informação (base para qualquer tratamento de dados). De fato, para implementação de programas de qualidade efetivos, para aprimoramento da governança de dados e tantos outros projetos em dados na organização, faz-se praticamente obrigatório um bom processo de classificação da informação. Para a LGPD [42], a classificação da informação é um tratamento como qualquer outra manipulação da informação (constante do artigo 5º inciso X). Cabe ressaltar que as normas dedicam tópico especial para este tratamento, da mesma forma que existe um tópico especial para o acesso a dados em estações de trabalho remotas).

Tem-se do objetivo da classificação presente na norma 27002 [84], que através dela, é possível priorizar a instituição de controles específicos para os dados de maior importância na organização. Alguns aspectos a se considerar na classificação da informação são: valor; requisitos legais;

sensibilidade; e criticidade. A norma alerta que esta atividade pode impactar a classificação de outros ativos (não de informação) que possuam alguma relação com ativos de informação (a exemplo, um disco rígido usado pelo presidente da organização pode ser mais crítico do que o disco rígido usado por um estagiário devido aos ativos de informação presentes ou ausentes neles). Aliada à classificação, a criação de rótulos pode facilitar a gerência dos ativos informacionais. Entretanto, a LGPD não faz menção direta a nenhuma estratégia de rotulação de dados, mas a viabilidade de seu uso fica subentendida nos artigos sobre princípios de segurança e boas práticas, além dos artigos sobre tratamentos de dados (artigos 5, 6, 46, 47 e 49) [42, 84, 86].

Esta recomendação de classificação de dados também é presente em diversos modelos de governança de dados (conforme abordaremos a seguir na seção específica de modelos de governança de dados 2.3.5 - Modelos/*Frameworks* de Governança de Dados), o que se configura um ponto de junção entre as disciplinas de segurança, privacidade e governança de dados. Este aspecto também demonstra que importância de uma boa classificação de dados tem, impactando diversas ações relativas à dados.

Na seção 6.5.3.1 da norma 27701 (extensão da norma 27002, seção 8.3.1) [84, 86], é possível encontrar algumas sugestões quanto ao uso de mídias removíveis tais como o uso de métodos criptográficos, destruição dos dados quando não mais necessários, ciência da degradação natural da capacidade de armazenamento ao longo do tempo que este meio oferece, controle e registro do uso de mídias removíveis e opção por armazenamento em meios menos vulneráveis. Na seção 6.5.3.2 da norma 27701 e 8.3.2 da norma 27002 [84, 86] encontram-se diretrizes para o descarte das mídias removíveis supracitadas. Segundo as normas [84, 86], deve-se observar o tratamento especial para mídias que contiverem informações confidenciais, dispondo de meios seguros e protegidos para seu descarte, ou para reaproveitamento da mídia (mediante remoção dos dados outrora utilizados). Sugere-se também a identificação de tais itens que requerem descarte seguro. No caso de consumo de serviços de coleta e descarte, deve-se selecionar um fornecedor com experiência e controle implementados adequados. Por fim, deve-se manter uma trilha de auditoria capaz de identificar itens sensíveis descartados.

Além do descarte de mídias físicas, a transferência também é objeto de foco nas normas 27701 e 27002 [84, 86], que sugerem confiabilidade do meio de transporte ou serviço de mensageiros, controle de portadores autorizados a realizar o transporte, procedimentos de verificação da identificação de transportadores (a fim de identificar a autenticidade e autorização destes transportadores), proteção da mídia em transporte a fim de evitar danos físicos, fatores ambientais entre outros e armazenamento de registros descrevendo o conteúdo das mídias, proteção aplicada, tempo gasto no trânsito entre custodiante e destino final.

Quanto ao controle de usuários, as normas (27002 e 27701 [84, 86]) definem que um usuário deve ter uma identificação única (chamada pelas normas de *"ID"*), a fim de identificar, sem margem de confusão, todos os usuários associados à organização. O compartilhamento de tais identificadores só se dará se for necessário por razões operacionais ou de negócios, exigindo a aprovação e documentação de tais casos. A norma também indica a imediata remoção/desabilitação de usuá-

rios quando cabível (deixar a organização, ou mudar de setor, por exemplo), além da constante verificação de redundância de usuários (tanto de forma reativa, isto é, nas bases de usuários, após o cadastro errôneo, como de forma proativa, impedindo tal cadastro equivocado).

Ainda sobre usuários, as normas [84, 86] apresentam diretrizes para o provisionamento de acesso de usuários, sendo desejável a previsão de processos para a obtenção de autorização para uso de serviço ou sistema de informação; segregação por níveis de acesso, através de uma política de acesso; verificação da real concessão de acesso apenas após a aprovação da autorização, manutenção de registro central de direitos de acesso para cada identificador de usuário; e por fim, constante, contínua e periódica validação da permanência da autorização de acesso de acordo com a situação geradora da concessão (por exemplo, a permanência na área que demanda aquele acesso, o desempenho de atividade que justifica o acesso, a própria continuidade da atividade, entre outros).

Segundo as normas em discussão [84, 86], convém que os acessos aos sistemas de informação sejam identificados e protegidos por procedimentos seguros de entrada (*logon*), nos quais não haja informações identificadoras do sistema ou da aplicação, até o sucesso na autenticação. Além disso, deve-se informar a necessidade de acesso por usuários autorizados; evitar fornecer informações que auxiliem um usuário não autorizado; validar informações somente após todos os dados de entrada estarem completos, não informando (no caso de erro) qual o dado falhou na autenticação; resistir a tentativas forçadas de entrada (força-bruta); registrar as tentativas de acesso ao sistema; comunicar de forma tempestiva um evento de segurança; esconder a senha informada durante a digitação e tráfego na rede; possuir mecanismos de expiração de sessões; e restringir o tempo de conexão. A Tabela 2.1 apresenta um resumo comparativo entre as normas e a LGPD.

Tabela 2.1: Tabela comparativa (Normas NBR/ISO/IEC 27701, 27001, 27002 e LGPD. (Fonte: [86, 85, 84, 42])

Subseções 27701	**Subseções 27001**	**Subseções 27002**	**Artigos LGPD**	**Conceito**
5.2.1	4.1	————	Artigo 50º, §1º, Artigo 5 VI, VII e IX	Contexto da Organização
5.2.2	4.2	————	Artigo 50º	Necessidades e expectativas de partes interessadas
5.2.3	4.3	————	Artigo 50º, §1º	Escopo SGSI
5.2.4	4.4	————	Artigo 50º, §2º I	SGSI
5.4.1.2	6.1.2	————	Artigo 38º, Artigo 50º §1º	Risco de Segurança da Informação
5.4.1.3	6.1.3	————	Artigo 38º, Artigo 50º §1º	Risco de Segurança da Informação
6.2.1.1	————	5.1.1	Artigo 38º	Políticas para Segurança da Informação
6.3.1.1	————	6.1.1	Artigo 41º	Responsabilidades e papéis da Segurança da Informação
6.3.2.1	————	6.2.1	Art. 6º, VII, Art. 46, Art. 47, Art. 49	Política dispositivos móveis
6.4.2.2	————	7.2.2	Artigo 50º Caput	Conscientização, educação e treinamento em Segurança da Informação
6.5.2.1	————	8.2.1	Art. 5 X, Art. 6º VII, Art. 46, Art. 47, Art. 49	Classificação da Informação
6.5.2.2	————	8.2.2	Art. 5 X, Art. 6º VII, Art. 46, Art. 47, Art. 49	Rótulos e tratamento da informação
6.5.3.1	————	8.3.1	Art. 5 I, Art. 6º VII, Art. 46, Art. 47, Art. 49	Gerenciamento de mídias removíveis
6.5.3.2	————	8.3.2	Art. 5 X, Art. 6º VII, Art. 46, Art. 47, Art. 49	Descarte de mídias
6.5.3.3	————	8.3.3	Art. 5 X, Art. 6º VII, Art. 46, Art. 47, Art. 49	Transferência física de mídias
6.6.2.1	————	9.2.1	Artigo 46	Registro e cancelamento de usuário
6.6.2.2	————	9.2.2	Artigo 46 e 49	Provisionamento para acesso de usuário
6.6.4.2	————	9.4.2	Artigo 46 e 49	*Log-on*

Uma boa ferramenta para proteção de dados (pessoais ou não) consiste nas técnicas de criptografia de dados. Quanto a controles criptográficos, a norma 27701 [86], estendendo a norma 27002 [84], define que os controles criptográficos devem compor a política da organização, fazendo

com que a Alta Direção esteja envolvida nos aspectos decisórios quanto ao cenário de uso, nível requerido de proteção, gerenciamento de chaves criptográficas, definição e manutenção de papéis e responsabilidades, padrões e boas práticas. Além disso, as normas [84, 86] citam alguns dos objetivos que o uso de criptografia almeja, sendo eles:

1. **Confidencialidade**: uso da criptografia como método de proteção de informações sensíveis ou críticas, armazenadas ou transmitidas.
2. **Integridade/Autenticidade**: uso da criptografia (como em assinaturas digitais ou autenticação de mensagens) a fim de assegurar o emissor da informação e seu conteúdo.
3. **Não repúdio**: uso da criptografia como ferramenta de detecção de um evento ou ação obtendo uma evidência dessa ocorrência.
4. **Autenticação**: uso da criptografia no processo de certificação que um usuário está apto a acessar outras camadas sistêmicas, entidades ou recursos.

Quanto à reutilização ou descarte seguro de equipamentos, a norma 27701 [86] pousa especial atenção a fim de assegurar que controles sejam implementados para que não seja possível acessar de forma indevida DP a partir de equipamentos descartados ou reutilizados. Faz-se necessário, portanto, identificar e inviabilizar os métodos de acesso a este DP previamente armazenado. Além disso, deve-se tratar o risco deste armazenamento considerando que o equipamento efetivamente ainda contém um DP, mesmo que deslocado para outro fim. Os métodos criptográficos podem ser usados para este fim.

Esta preocupação do descarte de equipamentos tem profunda conexão com o chamado "ciclo de vida do dado" definido a seguir em seção específica (2.3.3 - Ciclo de Vida do Dado). Esta relação se dá porque o dado deve ser controlado do momento em que surge (cadastro, carga, registro) até o momento do seu descarte. Alguns atacantes, sabendo da recorrente desatenção com o descarte, exploram esta vulnerabilidade conseguindo importantes informações sobre os dados. Alguns autores chamam essa ação de "revirar o lixo" [34].

As normas [84, 86] também fazem menção a uma política que denominam "Mesa limpa, tela limpa", onde cada informação deve possuir um local adequado de armazenamento, de forma segura, a fim de evitar acesso indevido, principalmente quando estas informações não estão em uso ou quando o ambiente físico que as comporta esteja desocupado. Esta política faz referência às mesas de trabalho de um escritório, que geralmente possuem informações valiosas para um atacante que consiga romper as barreiras físicas e tenha acesso a elas. Uma prática insegura mas comum é anotar senhas em pedaços de papel e deixá-las sobre a mesa, próximas da estação de trabalho. Tal cenário é propício para facilitar a ação de atacantes. Além da sugestão da limpeza das mesas, as estações de trabalho devem ser mantidas bloqueadas ou desligadas, quando fora de uso. Impressões ou cópias tiradas por fotocopiadoras (principalmente as que possam conter informações sensíveis ou pessoais) devem ser recolhidas imediatamente, a fim de reduzir a janela de extravio dessas informações.

Outro ponto comum em políticas de segurança é a necessidade de cópias de segurança (*back-ups*) que possibilitem a restauração da informação em caso de desastres ou perda acidental. As normas [84, 86] citam que o assunto deve ser alvo de uma política governada, de forma que se estabeleça uma estrutura definida de papéis, responsabilidades, escopo e atividades. Deve-se considerar a criticidade, sensibilidade, classificação e requisitos das informações candidatas a compor a cópia de segurança. Deve-se definir também os aspectos temporais dessas cópias como intervalos de tomadas de cópias, período de retenção, níveis de segurança exigidos, eventos que gerem restauração das cópias, entre outros. As normas [84, 86] também citam um plano de *backup* que deve levar em consideração: 1. completude e exatidão das cópias; 2. documentação efetiva sobre o procedimento de restauração; 3. abrangência (cópias completas ou diferenciais) das cópias; 4. aspectos temporais; 5. requisitos de negócio; 6. requisitos de segurança da informação; 7. requisitos de privacidade de DP; 8. criticidade da informação para o plano de continuidade da operação; 9. local de armazenamento da cópia (local remoto); 10. segurança física, ambiental e lógica do local de armazenamento das cópias; 11. validação e verificação periódica das cópias, para ação tempestiva na correção de erros; 12. verificação da necessidade de confidencialidade das informações (para implementação de métodos de encriptação).

Uma estratégia que pode facilitar as rotinas de cópias de segurança e que, na verdade, contribui amplamente para a segurança de sistemas, é a tomada de registros de segurança de eventos para auditoria (*logs*). Tais registros podem ser tomados a partir de atividades de usuários, exceções, falhas e eventos de segurança da informação sendo produzidos, mantidos e analisados criticamente, em intervalos regulares. Alguns exemplos de atividades passíveis de registro são: identificações de usuários; atividades de sistema; eventos de tempo (data, horário, início, fim, período, latência, entre outros); identidade de dispositivos; localização; identificador do sistema; tentativa de acesso a sistemas e recursos, bem sucedidas ou não; alterações de parâmetros e configurações de sistema; uso de privilégios; acesso a aplicações e utilitários de sistema; acesso a arquivos (com o tipo de acesso); endereços e protocolos de rede; alarmes de sistema e controles de acesso; eventos de sistemas de proteção; transações executadas por usuários nas aplicações; acessos/manipulações a DP; alterações de permissões para acessos a DP [84, 86]. Além disso, as normas [84, 86] citam que é necessário que estes registros (*logs*) estejam protegidos de alteração e acesso indevido. Estes registros podem estar relacionados a eventos envolvendo DP tais como acessos, permissões, exclusão, cópia, entre outros. Portanto é necessária a implementação de controles especiais focados na mitigação deste risco.

Um aspecto citado nas normas [84, 86] e que tem vínculo direto com a LGPD [42] é o planejamento da transferência de informações, principalmente no caso de transferência de DP. Isto porque a lei LGPD designa papéis e responsabilidades para quem coleta, armazena, manipula e trata os DP, inclusive podendo envolver diferentes entes desempenhando cada um destes papéis (operador e controlador). Para as normas [84, 86], deve-se estabelecer procedimentos para proteger a informação em trânsito (sensível, crítica ou pessoal) contra interceptação, cópia, modificação, desvio ou destruição. Além disso, deve-se estabelecer procedimentos de detecção e proteção contra códigos maliciosos, protegendo informações sensíveis transferidas em forma de

anexo. Deve-se definir de maneira clara e objetiva o uso aceitável dos recursos eletrônicos de comunicação, além das responsabilidades envolvidas nos processos de transferência de informação. Faz-se necessário definir uma política de retenção e descarte das correspondências de negócio (como mensagens, *e-mails*, reuniões, entre outros), definindo restrições para o uso desses canais de comunicação. É importante também lembrar do tratamento especial que informações sensíveis, críticas ou pessoais devem sofrer a fim de assegurar a proteção à confidencialidade, integridade, privacidade e autenticidade das informações.

Uma das estratégias que também compõem políticas de transferência de informações são os acordos de confidencialidade e não divulgação. Segundo as normas [84, 86], é necessário que este instrumento legal abarque uma definição formal da informação a ser protegida, descrevendo a classificação dessa informação, a validade estimada deste acordo, incluindo os aspectos que perduram mesmo após a expiração deste prazo e incluindo também os procedimentos a serem tomados após este período expirado, assim como as responsabilidades dos signatários, identificação do domínio (proprietário) da informação, segredos comerciais e propriedade intelectual, bem como a relação destes com a proteção da informação confidencial e pessoal, as cláusulas de concessão do uso da informação confidencial e pessoal, direito de auditar e monitorar o tratamento das informações (pessoais e confidenciais), a definição das atividades de notificação e relato de vazamento de informações.

Quanto ao desenvolvimento de aplicações seguras, principalmente no aspecto de compartilhamento de recursos em redes públicas (meios inseguros), as normas [84, 86] definem alguns pontos de consideração:

1. nível de confiança que cada módulo requer, dada certa identidade (autenticação);
2. autorização para, por exemplo, aprovar conteúdos, publicar ou assinar documentos-chave transacionais;
3. comunicação limpa com informação total de autorizações dos dois lados da comunicação a fim do fornecimento/uso do serviço;
4. requisitos de confidencialidade, integridade, evidências de emissão e recebimento de documentos-chave, não-repúdio de fornecimento/uso de serviços, privacidade (informações pessoais) e proteção (informações confidenciais);
5. nível de confiança requerido na integridade de documentos-chave;
6. confidencialidade e integridade de transações com informações críticas para o negócio;
7. grau de investigação apropriado para a verificação de informações fornecidas por um cliente;
8. requisitos apropriados para proteção contra fraudes;
9. prevenção contra perda ou duplicação de informações;
10. responsabilidades associadas a quaisquer transações fraudulentas;

11. avaliar o uso de criptografia para trânsito de informações pessoais por meios inseguros;

Este desenvolvimento seguro de aplicações em redes públicas (e outros aspectos guias dessas ferramentas) devem fazer parte da política de desenvolvimento seguro, onde, de acordo com o contexto da organização, devem-se definir aspectos que disponibilizem um ambiente de desenvolvimento confiável, definir a disciplina de segurança dentro do ciclo de desenvolvimento de *softwares* (tanto na metodologia aplicada de desenvolvimento como na linguagem de programação usada), envolvendo segurança no projeto da solução, pontos de verificação ao longo da construção da solução, repositórios seguros, controle de versões seguro, conhecimento e treinamento em segurança de aplicações e finalmente o fomento para ampliar a capacidade de desenvolvedores de evitar, encontrar e corrigir vulnerabilidades.

Especificamente, a norma 27701 cita nesta subseção (6.11.2.1 [86]) os princípios de *Privacy by Design* e *Privacy by Default*. Segundo Barbieri [15], a alcunha do *Privacy by Design* presente na lei GDPR [151] se refere ao tratamento da privacidade já no momento do projeto, em sua concepção, enquanto a *Privacy by Default* é a diretiva que mantém a privacidade, ou seja, o direito de não ser identificado, como escolha padrão dos usuários, a fim de que o desconhecimento do titular não acarrete na exposição de seus dados.

A norma [86] discorre que, para auxiliar os princípios de *Privacy by Design* e *Privacy by Default*, a política de desenvolvimento seguro deve conter diretrizes sobre proteção de DP e implementação de privacidade no ciclo de vida do desenvolvimento de *software*, além de apresentar requisitos de proteção e privacidade de DP, já na etapa de concepção do *software* baseados numa avaliação de riscos de privacidade ou na avaliação de impacto da privacidade. Tal política deve conter ainda pontos de controle para a proteção de DP como estrutura formal do projeto, conhecimento requerido para privacidade e proteção de DP e, por via de regra, minimização do tratamento de DP. Juntamente com as políticas, as normas [84, 86] definem que é necessário para as organizações estabelecer princípios para o projeto de sistemas seguros, que levem em consideração a engenharia de segurança em todas as camadas da arquitetura (negócio, dados, aplicações e tecnologias) de forma eficaz, para melhorar as normas de segurança no processo de engenharia.

Quanto à informação utilizada para testes de programas, as normas ISO 27701 [86] e ISO 27002[84] sugerem a utilização de dados fictícios, a fim de preservar a segurança da informação produtiva. Entretanto, nem sempre é possível evitar o uso do dado produtivo. Nestes casos, deve-se aplicar controles de acessos efetivos, controles de autorização do uso da cópia dos dados produtivos, deleção imediata após a realização do teste e que seja registrado o uso a fim de formar uma trilha de auditoria. A norma 27701 [86] também alerta sobre a necessidade de manter de forma clara e documentada as responsabilidades compartilhadas e individuais em contratos com fornecedores sobre o tratamento de DP. Para tanto, é preciso definir medidas mínimas técnicas e organizacionais que o fornecedor precisa atender, fornecendo um mecanismo para assegurar que a organização apoie e gerencie a conformidade com todas as legislações e/ou regulamentações aplicáveis.

De fato, as responsabilizações devem constar do processo de gestão de incidentes, para identifi-

cação, registro e notificação de violações à privacidade de DP, além de cabível a comunicação às autoridades, levando em conta a regulamentação/legislação aplicável [86]. Quanto a resposta aos incidentes de violação de DP, a norma 27701 [86] sugere que as notificações de tais incidentes possuam informações suficientes para os decisores capazes de possibilitar (quando necessário) uma análise crítica do evento. Tal notificação deve ser estendida às autoridades de supervisão e ao próprio titular do dado. Aspectos como descrição do incidente, dados afetados, tempo, consequências, passos para resolução do incidente e efeitos do incidente devem ser comunicados.

As normas 27701 [86] e 27702 [84] citam que os requisitos definidos nelas relacionados a responsabilidades de proteção, segurança e privacidade de DP podem guiar cláusulas contratuais com seus clientes a fim de formar uma base para sansões contratuais, no caso de uma violação daquelas responsabilidades. Segundo a norma 27701 [86], convém que a organização mantenha cópias de procedimento e políticas de privacidade, por período apropriado de acordo com o contexto da organização, a fim de embasar análises críticas de políticas e procedimentos (tais como no caso de litígio legal contra cliente e em uma investigação por motivo de auditoria).

As normas [84, 86] sugerem que as organizações, no papel de operadores de DP, disponibilizem aos clientes, antes de celebrar e durante a celebração de um contrato, evidências de que a segurança da informação e da privacidade está totalmente implementada e é consoante aos procedimentos e políticas da organização, com o objetivo de aumentar a transparência no tratamento de DP. Com relação à análise crítica técnica para conformidade, a organização deve incluir métodos e componentes relacionados ao tratamento de DP, tais como: monitoramento contínuo, para verificar que somente o tratamento permitido está sendo executado, ou testes específicos de vulnerabilidade ou invasão [86]. A Tabela 2.2 apresenta os controles tratados neste trabalho.

Tabela 2.2: Tabela comparativa (Normas NBR/ISO/IEC 27701, 27001, 27002 e LGPD (Continuação). (Fonte: [42, 84, 85, 86])

Subseções 27701	**Subseções 27002**	**Artigos da LGPD**	**Conceito**
6.7.1.1	10.1.1	Artigo 46	Política para uso de controles criptográficos
6.8.2.7	11.2.7	Artigo 46	Reutilização ou descarte seguro de equipamentos
6.8.2.9	11.2.9	Artigo 46	"Mesa limpa, tela limpa"
6.9.3.1	12.3.1	Artigo 46	*Back-up*
6.9.4.1	12.4.1	Artigo 46	*Logs*
6.9.4.2	12.4.2	Artigo 46	Proteção de *logs*
6.10.2.1	13.2.1	Artigo 46	Políticas e procedimento para transferência de informações
6.10.2.4	13.2.4	Artigo 46 e 47	Acordos de confidencialidade e não divulgação
6.11.1.2	14.1.2	Artigo 46	Serviços de aplicações seguras em redes públicas
6.11.2.1	14.2.1	Artigo 49	Política de Desenvolvimento Seguro
6.11.2.5	14.2.5	Artigo 49	Princípios para Projetar Sistemas Seguros
6.11.3.1	14.3.1	Artigo 46	Proteção dos Dados Para Teste
6.12.1.2	15.1.2	Artigo 46	Segurança da Informação nos Acordos com Fornecedores
6.13.1.1	16.1.1	Artigo 46	Responsabilidades e Procedimentos
6.13.1.5	16.1.5	Artigo 48 e 50 (g)	Resposta aos Incidentes de Segurança da Informação
6.15.1.1	18.1.1	Artigo 12 §3º, 32, 46 §1º, 49, 50 e 51	Identificação da Legislação Aplicável e de Requisitos Contratuais
6.15.1.3	18.1.3	Artigo 6 §1º	Proteção de Registros
6.15.2.1	18.2.1	Artigo 50	Análise Crítica Independente da Segurança da Informação
6.15.2.3	18.2.3	Artigo 50	Análise Crítica Técnica da Conformidade

A norma 27701 [86] também possui uma série de controles específicos para os controladores de DP presentes em sua seção 7. Estes controles serão discutidos a seguir, sob a luz da lei LGPD [42]. A norma 27701 [86] relembra que é adequado às organizações assegurar que os titulares de DP entendam efetivamente os propósitos para o tratamento de seus dados, através de uma comunicação objetiva e uma documentação clara e detalhada. Tal ação impacta diretamente na capacidade do titular de expressar de forma consistente seu consentimento. Outro ponto de interesse é a demonstração de que a legalidade do tratamento foi devidamente estabelecida antes do tratamento. Algumas origens dessa legalidade são apresentadas pela norma [86]: consentimento

dos titulares de DP; cumprimento de contrato; conformidade legal; proteção dos interesses vitais dos titulares de DP; desempenho de uma tarefa realizada de interesse público; e interesses legítimos do controlador de DP.

Um dos instrumentos legais para justificar o tratamento de DP é a coleta de consentimento (prevista pela LGPD [42]). Sobre o assunto, a norma [86] estabelece que é necessário documentar inequivocamente a requisição de consentimento e os requisitos para obtê-lo. Sugere-se também a correlação entre os propósitos para tratamento e as informações sobre "se" e "como" o consentimento é obtido. Deve-se considerar ainda o tipo de titular de DP e o propósito da coleta de dados que podem incorrer em requisitos adicionais para a coleta do consentimento. Uma vez coletado, esse consentimento deve ser registrado a fim de fornecer, sob solicitação, detalhes como data de concessão, identificação do titular e a declaração de consentimento. São consentimentos válidos aqueles que obedecem a livre vontade do titular, a especificidade do propósito de tratamento e a não ambiguidade e explicitude das informações [86].

Segundo a norma 27701 [86], o tratamento de DP gera ao titular da informação riscos que devem ser considerados através de uma avaliação de impacto à privacidade. Os critérios dessa avaliação são diversos, como tomada de decisão automatizada, tratamento em larga escala de categorias especiais de DP ou monitoramento sistemático de uma área publicamente acessível em larga escala. Cada organização deve, a partir de suas características de negócio, determinar quais elementos são necessários a esta avaliação. A norma [86] define também que um operador de DP que aja em nome da organização deve possuir um contrato que requeira a implementação de controles apropriados, conforme apresentado nas Tabelas 2.3,2.4, 2.5, 2.6, levando em conta o processo de avaliação de riscos de segurança da informação e o escopo de tratamento de DP realizado pelo operador.

Tabela 2.3: Controles e objetivos de Controles (Normas NBR/ISO/IEC 27701). (Fonte: [86])

B.8.2 Condições para coleta e tratamento Objetivo: Documentar e determinar que o tratamento é lícito, com base legal, conforme as jurisdições aplicáveis e com propósitos legítimos e claramente definidos.		
B.8.2.1	Acordos com o cliente	*Controle* A organização deve assegurar, onde pertinente, que o contrato para tratar DP considera os papéis da organização em fornecer assistência com as obrigações do cliente (considerando a natureza do tratamento e a informação disponível para a organização).
B.8.2.2	Propósitos da organização	*Controle* A organização deve assegurar que os DP tratados em nome do cliente sejam apenas tratados para o propósito expresso nas instruções documentadas do cliente.
B.8.2.3	Uso de *marketing* e propaganda	*Controle* A organização não pode utilizar os DP tratados sob um contrato para o propósito de *marketing* e propaganda, sem o estabelecimento de que um consentimento antecipado foi obtido do titular de DP apropriado. A organização não pode fornecer este consentimento como uma condição para o recebimento do serviço.
B.8.2.4	Violando instruções	*Controle* A organização deve informar ao cliente se, na sua opinião, uma instrução de tratamento viola uma regulamentação e/ou legislação aplicável.
B.8.2.5	Obrigação do cliente	*Controle* A organização deve fornecer ao cliente informações apropriadas de tal modo que o cliente possa demonstrar *compliance* com suas obrigações.
B.8.2.6	Registros relativos ao tratamento de DP	*Controle* A organização deve determinar e manter os registros necessários para apoiar a demonstração do *compliance* com suas obrigações (como especificado no contrato aplicável) para tratamento de DP realizado em nome do cliente.

Tabela 2.4: Controles e objetivos de Controles (Normas NBR/ISO/IEC 27701) - Continuação. (Fonte: [86])

B.8.3 Obrigações para os titulares de DP Objetivo: Assegurar que os titulares de DP sejam providos com informações apropriadas sobre o tratamento de seus DP, e que estejam de acordo com quaisquer outras obrigações aplicáveis para os titulares de DP relativas ao tratamento de seus DP.		
B.8.3.1	Obrigações para os titulares de DP	*Controle* A organização deve fornecer ao cliente meios para estar em *compliance* com suas obrigações relativas aos titulares de DP.

Tabela 2.5: Controles e objetivos de Controles (Normas NBR/ISO/IEC 27701) - Continuação. (Fonte: [86])

B.8.4 *Privacy by Design* e *Privacy by Default* Objetivo: Assegurar que os processos e sistemas sejam projetados de forma que a coleta e o tratamento de DP (incluindo o uso, divulgação, retenção, transmissão e descarte) estejam limitados ao que é necessário para o propósito identificado.		
B.8.4.1	Arquivos temporários	*Controle* A organização deve assegurar que os arquivos temporários criados como um resultado do tratamento de DP sejam descartados (por exemplo, apagados ou destruídos), seguindo os procedimentos documentados, dentro de um período especificado e documentado.
B.8.4.2	Retorno, transferência ou descarte de DP	*Controle* A organização deve fornecer a capacidade de retornar, transferir e/ou descartar DP de uma maneira segura. Deve também tornar sua política disponível para o cliente.
B.8.4.3	Controle de transmissão de DP	*Controle* A organização deve sujeitar DP transmitidos sobre uma rede de transmissão de dados a controles apropriados projetados, para assegurar que os dados alcancem seus destinos pretendidos.

Tabela 2.6: Controles e objetivos de Controles (Normas NBR/ISO/IEC 27701) - Continuação. (Fonte: [86])

B.8.5 Compartilhamento, transferência e descarte de DP Objetivo: Determinar se e documentar quando os DP são compartilhados, transferidos para outras jurisdições ou terceiros e/ou divulgados, de acordo com as obrigações aplicáveis.		
B.8.5.1	Bases para transferência de DP entre jurisdições	*Controle* A organização deve informar ao cliente, em um tempo hábil, sobre as bases para a transferência de DP entre jurisdições e de qualquer mudança pretendida nesta questão, de modo que o cliente tenha a capacidade de contestar estas mudanças ou rescindir o contrato.
B.8.5.2	Países e organizações internacionais para os quais os DP podem ser transferidos	*Controle* A organização deve especificar e documentar os países e as organizações internacionais para os quais DP possam, possivelmente, ser transferidos.
B.8.5.3	Registros de DP divulgados para terceiros	*Controle* A organização deve registrar a divulgação de DP para terceiros, incluindo quais DP foram divulgados, para quem e quando.
B.8.5.4	Notificação de solicitações de divulgação de DP	*Controle* A organização deve notificar ao cliente sobre quaisquer solicitações legalmente obrigatórias para divulgação de DP.
B.8.5.5	Divulgações legalmente obrigatórias de DP	*Controle* A organização deve rejeitar quaisquer solicitações para divulgação de DP que não sejam legalmente obrigatórias, consultar o cliente em questão antes de realizar quaisquer divulgações dos DP e aceitar quaisquer solicitações contratualmente acordadas para a divulgação de DP, que sejam autorizadas pelo respectivo cliente.
B.8.5.6	Divulgação de subcontratados usados para tratar DP	*Controle* A organização deve divulgar para o cliente qualquer uso de subcontratados para tratar DP, antes do uso.
B.8.5.7	Contratação de um subcontratado para tratar DP	*Controle* A organização deve somente contratar um subcontratado para tratar DP com base no contrato do cliente.
B.8.5.8	Mudança de subcontratado para tratar DP	*Controle* A organização deve, no caso de ter uma autorização geral por escrito, informar o cliente acerca de quaisquer alterações pretendidas relativas à adição ou substituição de subcontratados no tratamento de DP, dando assim ao cliente a oportunidade de se opor a essas alterações.

Um assunto também abordado pela norma 27701 [86] é a possibilidade de controladoria conjunta de DP. Neste cenário, as responsabilidades e papéis precisam estar claros e de forma transparente (de preferência, no instrumento legal que vincula os dois entes - tais como contratos, cessões, acordos, entre outros - ou outro documento de concorrência similar), contendo os termos e condições para o tratamento combinado. Um exemplo de documento deste seria o "acordo de compartilhamento de dados". Podemos concluir deste cenário que é de suma importância que se registre todas as ações tomadas no tratamento de DP. Segundo a norma [86], a forma de um inventário ou lista de atividades para este fim deve conter: 1) tipo do tratamento; 2) propósito do tratamento; 3) descrição da categoria de DP e dos titulares dos dados; 4) categoria de destinatários para quem o DP será divulgado, incluindo categorização de ente internacional; 5) descrição geral das medidas de segurança técnica e organizacional; e 6) relatório de Avaliação de Impacto de Privacidade.

Mais um aspecto observável pelos controladores de DP é a delimitação das obrigações dos titulares de DP. Sobre o assunto, a norma 27701 [86] sugere que, devido à grande variação de obrigações e meios para apoiar o cumprimento dessas obrigações (dependentes do contexto de cada organização), uma documentação clara e abrangente seja disponibilizada ao titular do DP, a fim de informar de forma acessível e em tempo hábil a abrangência de suas obrigações e de que forma ele pode acompanhar suas solicitações. Frequentemente, para manter o titular atualizado com suas obrigações, será necessário fornecer-lhe informações dos DP. Para tanto, a norma 27701 [86] cita que é necessário identificar os requisitos legais, regulamentares, contratuais e/ou de negócio para cada tipo de informação fornecida.

Segundo a norma 27701 [86], este fornecimento deve ser em tempo hábil, de forma concisa, completa, transparente, inteligível e facilmente acessível, usando uma linguagem curta e clara, apropriada ao público-alvo, além de ser preferencialmente acessível de forma permanente. De posse de suas informações, o titular dos DP pode requerer a alteração ou remoção do consentimento. É responsabilidade da organização controladora fornecer mecanismos efetivos e tempestivos para essas alterações, inclusive informando o titular sobre seus direitos relativos ao consentimento. Informações sobre impossibilidade de mudança imediata do consentimento (por força de lei, ou especificidade do negócio) devem ser amplamente difundidas (de preferência de maneira prévia) aos titulares dos dados, mantendo sempre meios de acompanhamento de sua solicitação [86]. Um dos direitos do titular dos DP é o direito de negar o consentimento ao tratamento de seus dados. As organizações devem assegurar a implementação de medidas apropriadas para permitir que os titulares exercitem esse direito, através da documentação dos requisitos legais e regulamentares relativas às objeções do consentimento. Além disso, é papel da organização controladora informar o titular acerca da capacidade de negar o consentimento [86].

Vale ressaltar que a LGPD garante a prerrogativa a instituição de manter os dados pessoais mesmo sem o consentimento do titular, desde que indispensáveis para a garantia da prevenção a fraude [42], de acordo com o exposto no Capítulo introdutório (Capítulo 1 - Introdução).

A norma 27701 [86] traz diretrizes para que os titulares possam corrigir e excluir seus DP. Para tal, é necessário que a organização implemente políticas, procedimentos e/ou mecanismos com

este fim, nos quais estejam definidos tempo de resposta para tratamento desta solicitação. Há também a necessidade de tratar a possibilidade de disputa sobre a precisão ou correção do DP pelo titular, informando a ele quais as mudanças feitas, quais as razões das mudanças que não foram realizadas, assim como possíveis restrições aplicáveis. As modificações em consentimento ou de dados precisam ser comunicadas aos terceiros colaboradores. Para tanto, a norma 27701 [86] discorre que a organização precisa adotar passos apropriados, mediante a tecnologia disponível, às vezes até por força de lei. Deve-se então manter um canal de comunicação ativo com terceiros, onde as responsabilidades relacionadas sejam passíveis de atribuição de acordo com operações de tratamento de DP e suas manutenções. Também é factível que haja monitoramento e conhecimento do recebimento das informações obtidas por este canal.

Existem diferentes estratégias para se informar ao titular do dado sobre o seu DP tratado. Uma dessas estratégias é o fornecimento de uma cópia deste dado. Convém que isso se dê em um formato estruturado e usado normalmente. Vale destacar que é responsabilidade da organização assegurar que esta cópia faça menção especificamente ao titular relacionado. Caso adotado o mecanismo de anonimização, não se faz necessária a quebra do processo para reidentificação do titular a fim de informá-lo quanto aos seus dados (justificada apenas na implementação deste controle). Outra informação importante é que, caso solicitado pelo titular, uma cópia destes dados pode ser transferida para outra organização de forma direta [86]. As organizações devem atender às solicitações de seus titulares de DP. Estas solicitações legítimas podem tratar de cópia de dados ou apresentação de uma queixa. É necessário que as solicitações sejam acolhidas dentro de prazos apropriados e previamente definidos. Atrasos devem ser comunicados e justificados aos titulares [86]. Principalmente no cenário de *Big Data*, a tomada de decisão automatizada envolvendo DP configuram obrigações específicas e podem afetar especificamente os titulares. Neste contexto, as organizações devem notificar os titulares da existência da tomada de decisão automatizada, permitindo que eles expressem desaprovação da decisão tomada e/ou solicitem intervenção humana. A Tabela 2.7 apresenta os controles resumidamente.

Tabela 2.7: Tabela comparativa (Normas NBR/ISO/IEC 27701, 27001, 27002 e LGPD (Continuação). (Fonte: [42, 84, 85, 86])

Subseções 27701	**Artigos LGPD**	**Conceito**
7.2.1	Art. 9º - I, Art.14º §6º	Identificação e documentação do propósito
7.2.2	Art. 7º - II, Art. 8º §4º, Art. 11º - IIa, Art. 23º, Art. 26 - IV, Art. 34º - I	Identificação de Bases Legais
7.2.3	Art. 5º XII, Art. 7º, Art. 8º, Art. 11º, Art. 14.	Consentimento
7.2.4	Art. 5º XII, Art. 7º, Art. 8º, Art. 11º, Art. 14.	Consentimento
7.2.5	Art. 4º - §3º, Art. 5º XVII, Art. 10º III, Art.32º, Art. 38º	Avaliação do Impacto da Privacidade
7.2.6	Art. 7º, Art. 39º	Contratos com Operadores de DP
7.2.7	Art. 7º - §5º	Controlador Conjunto de DP
7.2.8	Art. 37º	Registros Relativos ao Tratamento de DP
7.3.1	Art. 9º	Determinando e Cumprindo as Obrigações para os Titulares de DP
7.3.2	Art. 9º	Determinando as Informações para os Titulares de DP
7.3.3	Art. 9º	Fornecendo Informações aos Titulares de DP
7.3.4	Art. 8º §5º, Art. 9º §2º	Modificar ou Cancelar o Consentimento
7.3.5	Art. 8º §5º, Art. 9º §2º	Negar Consentimento
7.3.6	Art. 9º	Acesso, Correção e/ou Exclusão
7.3.7	Art. 18º §6º	Obrigações dos Controladores de Informações a Terceiros
7.3.8	Art. 18º II	Fornecendo Cópia do DP Tratado
7.3.9	Art. 18º	Tratamento de Solicitações
7.3.10	Art. 18º	Tomada de Decisão Automatizada

Ainda citando aspectos com impacto especial em ambiente de *Big Data*, a norma 27701 [86] pontua que sejam observados os limites de coleta de DP. Este princípio (chamado de "Minimização de Dados") será abordado posteriormente na seção sobre anonimização de dados (Seção 2.2). Entretanto, para a norma 27701 [86], a coleta deve se restringir ao mínimo adequado, relevante e necessário em relação aos propósitos identificados. Este princípio pode decorrer especificamente do princípio de *Privacy by Default*, onde cada opção de coleta e tratamento de DP seja desabilitada por padrão e somente seja habilitada por uma escolha explícita e consciente do titular de DP.

O tratamento de DP deve ser restrito pelas políticas de privacidade de segurança da informação

e por procedimentos documentados para as suas adoções e conformidade. Este tratamento inclui divulgação, armazenagem e permissões de acesso aos DP. Outro limite para o tratamento é, por padrão, o mínimo necessário para os propósitos identificados [86]. Outro aspecto que deve ser buscado pelas organizações é a precisão dos dados (principalmente os que representam uma pessoa natural que mantém algum tipo de relacionamento com a empresa). Para tanto, é necessária a implementação de políticas, procedimentos e/ou mecanismos que minimizem a imprecisão de DP. Convém que essas políticas, procedimentos e mecanismos também se destinem a responder às questões de imprecisão de DP e que sejam incluídos na informação documentada a fim de que sejam aplicados ao longo de todo o ciclo de vida do DP [86].

A precisão dos dados muitas vezes conduz à análise crítica de minimização de dados. Portanto, é desejável que a organização identifique como os DP específicos e a quantidade de dados coletados e tratados estão alinhados aos propósitos identificados. Em caso de não correspondência, técnicas de minimização de dados como a anonimização podem ser implementadas como alternativa à exclusão de dados. Note que mesmo anonimizado, a norma 27701 [86] trata do grau de correlação do dado anonimizado com o dado original, fazendo menção de que as técnicas de minimização não são necessariamente binárias (mapeável *versus* não-mapeável) mas pode existir uma escala de correlação (ou difusão) entre o dado original e o dado minimizado.

Segundo a LGPD em seu artigo 16º [42], o DP no fim do tratamento deve ser eliminado, ressalvados os casos descritos em seus incisos. Para tanto, a norma 27701 [86] define que a organização deve definir mecanismos para identificar o fim do tratamento de um DP para sua exclusão quando nenhum tratamento adicional for antecipado. Entretanto, alternativamente, algumas técnicas de anonimização podem ser usadas desde que tais técnicas não permitam, de forma razoável, a reidentificação dos titulares.

Segundo a norma 27701 [86], a organização deve levar em consideração a verificação periódica de arquivos temporários que possam conter DP a fim de excluí-los dentro de um período de tempo identificado. Quando a armazenagem não representar informação transitória e volátil (como arquivos temporários), convém que a organização planeje e mantenha atualizadas as políticas de retenção para informações, considerando o requisito para retenção do DP em período não maior do que o necessário. Tais políticas devem considerar requisitos legais, regulamentares e de negócio [86].

A organização deve dar atenção à escolha do método de descarte do DP, uma vez que essa escolha depende de um número de fatores, havendo variações nas suas propriedades e resultados. Alguns pontos de consideração são: a natureza e abrangência do DP descartado, a existência de metadados associados e características físicas da mídia de armazenamento [86]. Como já tratado em outro controle mencionado acima, a transmissão de DP precisa ser controlada em aspectos como autorização do destinatário da informação e adequação dos processos seguidos a fim de assegurar o não comprometimento do dado em trânsito [86]. A transferência de DP deve estar em conformidade (comprovada por meio documental) às legislações e regulamentações pertinentes (inclusive as internacionais) a fim de demonstrar transparência. Pode, portanto, ser necessária a

assinatura de acordos de transferências de informações, e, além disso, análises críticas de acordos já assinados devem ser implementadas por uma autoridade de supervisão designada [86].

A norma [86] também sugere que as identidades de países e organizações internacionais com os quais as quais a organização compartilhe DP, estejam disponíveis para consulta dos clientes, assim como a de entidades que surjam como fruto de subcontratações, a menos nos casos em que, por força de lei, tal especificação não seja possível ou que seja proibida. Esta disponibilização pode ser fruto do registro (desejável) de transferência de DP a terceiros que tenham sido modificados como resultado das suas obrigações no gerenciamento dos controles ou em outras transferências. Além disso, é necessário desenvolver uma política de retenção definindo o período de validade destes registros [86]. Ainda é necessário ao controlador de DP manter registros da divulgação de tais dados, tanto os de operações normais de tratamento de DP como os resultantes de investigações legais ou de auditorias externas [86]. Os controles supracitados para controladores de DP encontram-se resumidos na Tabela 2.8.

Tabela 2.8: Tabela comparativa (Normas NBR/ISO/IEC 27701, 27001, 27002 e LGPD (Continuação). (Fonte: [42, 84, 85, 86])

Subseções 27701	**Artigos da LGPD**	**Conceito**
7.4.1	Art. 6º - III	Limite de Coleta
7.4.2	Art. 16	Limite de Tratamento
7.4.3	Art. 6º - V	Precisão e Qualidade
7.4.4	Art. 6º - III	Objetivos de Minimização de DP
7.4.5	Art. 16º	Anonimização e Exclusão de DP ao Fim do Tratamento
7.4.6	CAPÍTULO VII	Arquivos Temporários
7.4.7	Art. 16º	Retenção
7.4.8	CAPÍTULO VII	Descarte
7.4.9	CAPÍTULO VII	Controle de Transmissão de DP
7.5.1	Art. 7º	Identificando as Bases para a Transferência de DP
7.5.2	CAPÍTULO V	Países e Organizações Internacionais para os quais os DP Podem ser Transferidos
7.5.3	CAPÍTULO V	Registro da Transferência de DP
7.5.4	Art. 37º	Registro de Divulgação de DP para Terceiros

De semelhante modo como houve com controladores de DP, a seção 8 da norma ISO/IEC 27701 [86] trata de controles específicos para operadores de DP. A seguir tecemos alguns comentários sobre estes controles.

Quanto aos acordos com clientes, a norma 27701 [86] expõe uma lista de itens de atenção, entre eles: *Privacy by Design* e *Privacy by Default*; obtenção da segurança do tratamento; notificação de violações envolvendo DP para uma autoridade de supervisão; notificação de violações envolvendo DP para os clientes e titulares de DP; realização de avaliações de impacto da privaci-

dade (AIP);e garantia de assistência pelo operador de DP, no caso em que consultas prévias com autoridades de proteção de DP relevantes sejam necessárias.

A norma 27701 [86] sugere que o contrato entre o cliente e a organização seja guiado pelo objetivo e o tempo de duração do serviço. Desta forma, pode ser necessário determinar o método de tratamento do DP, por razões técnicas. Tais informações devem estar acessíveis ao cliente, dada a conformidade com o propósito dos princípios da limitação e especificação, restringindo o uso de DP ao propósito expresso nas instruções documentadas pelo cliente. Sobre o uso de *marketing* e propaganda, a norma 27701 [86] cita que é desejável que a conformidade dos operadores de DP com os requisitos contratuais dos clientes seja documentada e que a inclusão do dado para este fim só se efetue no caso do consentimento expresso e inequívoco do titular.

Também é papel da organização informar ao cliente se uma instrução de tratamento emitida por ele viola alguma regulamentação e/ou legislação aplicável, dependendo do contexto tecnológico, da instrução em si e do contrato entre a organização e o cliente [86]. Ainda no campo das responsabilidades da organização (no papel de operadora de DP), faz-se necessário munir o cliente de informações suficientes que habilitem a demonstração de conformidade com suas obrigações, tais como contribuição para realização de auditorias [86]. É importante que a organização registre informações referentes ao tratamento dos DP. Algumas informações que podem fazer parte deste registro (segundo a norma [86]) são: 1. Categorias de tratamento realizadas em nome de cada cliente; 2. Transferências para outros países ou organizações internacionais; e 3. Descrição geral das medidas de segurança técnicas e organizacionais.

Segundo a norma 27701 [86], as obrigações do controlador de DP podem ser definidas por legislação, regulamentação e/ou contratos. Tais obrigações podem envolver serviços da organização (no papel de operadora de DP), sendo necessária a documentação da necessidade de consumo destes serviços via cláusulas contratuais. De semelhante modo ao que ocorre com controladores de DP, os operadores devem conduzir verificações periódicas para identificação e remoção de arquivos temporários não usados e que contenham DP, dentro de um período de tempo especificado [86].

No curso normal de uso de DP, o operador se depara com uma necessidade (geralmente final) de descarte da informação. Para tal, o dado pode ser retornado para um cliente, transferido para outra organização ou controlador, excluído (ou outra forma de destruição), desanonimizado ou arquivado. Tal tratamento deve ser gerenciado de maneira segura, assegurando ao cliente que os dados tratados sob um contrato serão apagados, tão logo quanto eles não sejam mais necessários (de acordo com seu propósito de uso). Tal tratamento deve ser parte de uma política de descarte de DP, disponível para os clientes e que cubra o período de retenção do dado antes do seu descarte a fim de proteger um cliente de perda de informações por questões de lapso temporal acidental de um contrato [86]. A norma 27701 [86] cita que a transmissão de DP precisa ser controlada e autorizada, utilizando processos apropriados e assegurando ao dado em trânsito o não comprometimento. Tais requisitos podem constar do contrato entre o operador e o cliente. A transferência de DP pode estar sujeita a requisitos legais, regulamentares, negociais e contratuais. Por isso é necessário que a organização documente a conformidade a esses requisitos como a base para a

transferência [86].

Da mesma forma que ocorre com o papel de controlador do DP, quanto a países e organizações internacionais com as quais a organização compartilhe DP, a norma [86] sugere que as identidades de tais países e organizações estejam disponíveis para consulta dos clientes, assim como entidades que surjam como fruto de subcontratações, a menos nos casos que, por força de lei, tal especificação não seja possível ou que seja proibida. Novamente, um controle compartilhado entre operadores e controladores é a manutenção de registros da divulgação de DP, tanto as de operações normais de tratamento de DP como as resultantes de investigações legais ou de auditorias externas [86]. No cenário onde a organização receba uma obrigatoriedade legal de divulgação de DP (por exemplo, por interesse público), convém que o cliente seja notificado, dentro de um prazo e procedimento acordado, a não ser que a solicitação inclua uma proibição de divulgação (tais como ocorrem em segredo de justiça) [86].

Para a norma 27701 [86], a divulgação do DP sob posse do operador deve ser rejeitada de ofício, a menos que seja vinculada legalmente, mediante consulta ao cliente, ou que seja pré-acordada por vias contratuais. Os detalhes relevantes para implementação do procedimento de divulgação devem constar do contrato. Também deve constar do contrato com o cliente o fornecimento do DP para subcontratado, incluindo os nomes dos subcontratados, países e organizações internacionais envolvidos. Ainda é facultado, por motivo do aumento dos riscos de segurança, que a divulgação seja feita sob um acordo de não divulgação e/ou solicitação do cliente [86].

A delegação do tratamento de DP para um subcontratado se dá mediante autorização escrita do cliente, que pode estar em cláusula contratual no contrato do cliente ou em acordo assinado para tal fim. De qualquer forma, há a necessidade de conhecimento e anuência do cliente, a fim de assegurar os mesmos controles firmados com a organização operadora. Da mesma forma ocorre quando a organização almeja mudança de subcontratado [86]. A Tabela 2.9 apresenta um resumo dos controles destinados aos operadores de DP.

Tabela 2.9: Tabela comparativa (Normas NBR/ISO/IEC 27701, 27001, 27002 e LGPD (Continuação). (Fonte: [42, 84, 85, 86])

Subseções 27701	**Artigos LGPD**	**Conceito**
8.2.1	Artigo 10º, I, II, Artigo 18º	Acordos com o cliente
8.2.2	Artigo 9º, I, II, III, IV, V, VI, VII, Artigo 23º	Propósitos da Organização
8.2.3	Artigo 6º, Artigo 9º, I, II, III, IV, V, VI, VII, Artigo 10º, I	Uso de *Marketing* e Propaganda
8.2.4	Artigo 44º, Artigo 45º	Violando Instruções
8.2.5	Artigo 44º	Obrigações do Cliente
8.2.6	Artigo 37º	Registros Relativos ao Tratamento de DP
8.3.1	Artigo 6º, I, II, III, IV, V, VI, VII, VIII, IX, X, Artigo 7º, I, II, III, IV, V, VI, VII, VIII, IX, X, Artigo 42º	Obrigações para os Titulares de DP
8.4.1	Artigo 46º, Artigo 49º	Arquivos Temporários
8.4.2	Artigo 15º, I, II, III, IV, Artigo 16º, I, II, III, IV, Artigo 46º	Retorno, Transferência ou Descarte de DP
8.4.3	Artigo 6º, VII, VIII, Artigo 37º, Artigo 46º	Controle de Transmissão de DP
8.5.1	Artigo 33º, I II, III, IV, V, VI, VII, VIII, IX, Artigo 34º, I II, III, IV, V, VI	Bases para Transferência de DP
8.5.2	Artigo 33º, I II, III, IV, V, VI, VII, VIII, IX, Artigo 34º, I II, III, IV, V, VI	Países e Organizações Internacionais para os quais os DP Podem ser Transferidos
8.5.3	Artigo 16º, Artigo 37º	Registros de DP divulgados a Terceiros
8.5.4	Artigo 6º, I, VI, Artigo 41º	Notificação de Solicitações de Divulgação de DP
8.5.5	Artigo 4º, III, IV, Artigo 41º	Divulgações Legalmente Obrigatórias de DP
8.5.6	Artigo 41º	Divulgação de Subcontratados Usados para Tratar DP
8.5.7	Artigo 39º, Artigo 41º	Contratação de um Subcontratado para Tratar DP
8.5.8	Artigo 39º, Artigo 41º	Mudança de Subcontratado para Tratar DP

2.2 SEGURANÇA DA INFORMAÇÃO

Segundo Goodrich e Tamassia [70] o conceito de segurança está intimamente ligado à identificação de vulnerabilidades em sistemas de computadores. Os usuários mal intencionados costumam explorar essas vulnerabilidades permitindo desenvolver diversas formas de ataque, com diferentes potenciais de danos avaliados pelos riscos dos ataques. A segurança então se posiciona como a guardiã que, avaliando as vulnerabilidades, as possibilidades de ataques, seus riscos, alcances, potenciais de dano e contramedidas, propõe modelos sólidos e define propriedades de segurança desejáveis, implementando e testando defesas efetivas e monitorando sistemas a fim de validar continuamente a estabilidade dos ambientes e eventuais correções, caso necessárias.

Ainda no capítulo introdutório de seu livro, Goodrich e Tamassia [70] apresentaram o anonimato como um conceito da segurança. Para os autores, a fim de garantir a autenticidade de determinadas transações (podemos acrescentar aqui o contexto de prevenção à fraude, onde essa característica é bastante explorada), os sistemas geralmente vinculam uma propriedade ou característica do indivíduo no mundo real (tal como a biometria), garantindo que, por exemplo, o usuário logado de fato corresponde a determinado indivíduo. Entretanto, essa solução tem um efeito colateral, pois gera nos sistemas uma relação envolvendo as identidades dos indivíduos no mundo real. E estas relações intercambiáveis entre os inúmeros sistemas que utilizam nossos DP para tal acabam nos inter-relacionando entre as plataformas, permitindo a criação de um perfil invasivo obtido pela coleta de dados desenfreada [70]. Como resposta a este efeito colateral desastroso, a segurança propõe a ferramenta da anonimização, utilizada para combater a disseminação desenfreada de identidades reais entre os sistemas. Abordaremos este assunto a seguir, na subseção 2.2.1.

2.2.1 Anonimização de Dados

A anonimização é uma importante ferramenta para implementação da segurança da informação, e está intimamente relacionada à capacidade de identificação de determinado indivíduo. A depender da vinculação entre um determinado dado e um indivíduo, os dados podem ser classificados como *Personally Identifiable Information* (Informação Pessoal Identificável, ou *"PII"*) ou *Auxiliary Data* (Dados Auxiliares, ou *"AD"*). Brasher [24] definiu essas duas categorias, afirmando que a Informação Pessoal Identificável, ou *"PII"*, é a informação estrita capaz de indexar uma pessoa no conjunto de dados, diferenciando-a das demais; enquanto os Dados Auxiliares, ou *"AD"*, que compõe o conjunto de informações que por si só não restringem à unicidade tal indivíduo, mas aliados a demais informações podem contribuir para sua identificação. A título de exemplo, a matrícula de um funcionário é considerada um *PII* enquanto o endereço do mesmo é um *AD*, sendo que o endereço pode identificar certo indivíduo que more sozinho.

A lei prevê duas ferramentas que podem facilitar o fomento à privacidade através do mascaramento de *PII*, e/ou dos *AD*. São elas: a pseudonimização e a anonimização.

Codificado em nossas legislações, há o entendimento de que a anonimização de dados é uma ferramenta única e suficientemente capaz de garantir a proteção de DP. Tal crença é fortemente

discutida pela literatura [24, 50, 63, 122, 156], a qual indica como possível solução a revisão dos processos de governança e gerenciamento de riscos como forma efetiva de *compliance*, reduzindo o ônus do processo produtivo e extração de valor resultante da abdicação do uso dessas tecnologias.

Muitas organizações, principalmente com o início da vigência da LGPD em 2020, depositam uma expectativa irreal na ferramenta de anonimização como uma solução milagrosa para resolver todos os variados e complexos problemas de proteção e privacidade de dados em projetos de Big Data [123].

Isso porque a anonimização se apresenta como uma alternativa à exigência de aplicação dos princípios de proteção de dados a qualquer informação identificada ou identificável de uma pessoa natural (artigo 6º [42]; capítulo II [151]).

É o que ocorre, por exemplo, com o princípio da "Minimização de Dados", ou, no contexto europeu, do "Princípio da Necessidade" (artigo 5º(1)(c)[151]; artigo 6º [42], respectivamente). Sob este princípio, o DP precisa ser adequado, relevante e limitado aos propósitos para os quais foi adquirido, isto é, o tratamento de dados deve se limitar ao mínimo necessário para alcançar seus propósitos.

Um segundo princípio, por exemplo, que rege o processamento de DP é o princípio do legítimo interesse. De acordo com este, o dado deve ser usado, levando em consideração as expectativas racionais dos sujeitos aos quais os dados se referem, baseado na relação destes com o controlador dos dados (artigo 7º, IX, e 10º [42]; artigo 6º (1)(f) [151]; Preâmbulo da GDPR, ponto 47 [151]). Então, o legítimo interesse une o processamento do DP ao propósito ao qual o dado foi coletado. Infere-se desses princípios, dentre outros, que a regulamentação estabelece guias para o uso e impõe limites para o processamento de dados de pessoas naturais. Dados estes limites, é praticamente impossível pensar em projetos de *Big Data* envolvendo DP sem envolver algum tipo de estratégia de anonimização de dados.

De fato, esses princípios não são aplicados aos dados após a anonimização, uma vez que os sujeitos referentes aos dados após este processamento não podem ser mais identificados (artigo 12º [42]; preâmbulo GDPR, ponto 26 [151]).

Uma vez anonimizados, os dados podem ser coletados, tratados e explorados sem submissão a uma política de minimização de dados, e podem ser usados até mesmo para propósitos diversos do originalmente acordado. Ressalte-se, ainda, que as regulações fragilizam o conceito de dado anônimo enquanto dado que não pode mais ser relacionado a uma pessoa natural identificável. Isso porque a impossibilidade de vinculação entre o dado e o indivíduo não precisa ser absoluta, mas deve ser obstaculizada pelo tempo, custo ou outros fatores que dificultam a reidentificação. A relativização que faz a lei, prescindindo de um caráter absoluto para a anonimização, leva a crer que o dado anonimizado não pode mais gerar um vazamento da privacidade, o que nem sempre é verdade.

Portanto, apesar da inegável utilidade da ferramenta, a anonimização apresenta riscos e não pode

ser entendida como a solução final para os problemas de segurança da informação. É o que demonstra os autores Brasher, Domingo-Ferrer, Mehmood et. al. e Ryan e Brinkley [24, 50, 122, 156].

A crença na infalibilidade da anonimização pode ser resultado de uma falta comum de uma cultura de governança de dados institucionalizada [24, 43, 156, 29, 142], o que também acaba trazendo à tona outros problemas, tais como a desinformação sobre os riscos que persistem após a aplicação da anonimização.

Um relatório da empresa MicroStrategy em 2019, coletando dados de países como Brasil, Alemanha, Japão, Inglaterra e Estados Unidos, informou que apenas 38% das companhias consideraram reter mais do que a metade dos seus dados controlados. Apenas 16% dos respondentes disseram que "a implantação de *analytics* na organização está no nível de maturidade para incluir uma arquitetura sofisticada para análise de autoatendimento com governança, estruturas de segurança, acesso a *Big Data* e tecnologias móveis e preditivas apoiadas por um centro de excelência para treinamento e suporte" [124]. Corroborando com esse relatório, um estudo da *Harvard Business Review* publicou em 2017 que apenas 3% dos dados das companhias pesquisadas atendiam a padrões básicos de qualidade [173].

Assim sendo, podemos destacar quatro características relacionadas à anonimização que impõem desafios ao uso desse mecanismo. A regulação assimila os três primeiros, porém o quarto desafio pode ser objeto de análise.

A primeira característica se refere ao conceito de dado anônimo. Para ser considerado anônimo, um dado não deve ser passível de vinculação a um indivíduo. Essa impossibilidade de identificação deve ser ampla, considerando não apenas o nome, mas qualquer informação capaz de identificar ou tornar possível a identificação de um indivíduo (artigo 5º, I, [42]; artigo 4º, I, [151]).

A segunda característica é a dificuldade em se determinar o que é considerado identificado ou identificável para definir o anonimato de um conjunto de dados. Ambas as regulações dão destaque ao fato de que o caráter "razoável" deve ser levado em conta (artigo 5º, III [42]; ponto 26 [151]):

> "A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios." (artigo 12º, § 1º, [42]).

Desta forma, até mesmo as estruturas legais e normativas preconizam a existência de uma zona dúbia entre os dados anônimos, onde talvez com grandes quantidades de dados seja possível reidentificar indivíduos a partir de dados anônimos.

O terceiro aspecto é a dificuldade de determinar o anonimato de determinado dado ao longo do tempo, afinal, a identificação depende de critérios que são voláteis no tempo, mudando de acordo com os avanços tecnológicos, ou até mesmo com as condições particulares de análise. Como mencionado, a regulação considera os avanços tecnológicos para determinar o limite do razoável

ao anonimato do dado. Entretanto, este aspecto produz uma nuvem de incerteza quanto aos critérios práticos do grau de definição do anonimato.

Por fim, o quarto aspecto é que os dados anônimos envolvendo projetos de *Big Data* têm uma maior possibilidade de reidentificação. Isto acontece porque, no contexto de *Big Data* (e de dados massivos), a disponibilidade de grandes quantidades de dados torna a possibilidade de interconexão de tais dados mais provável, mesmo quando lidamos com metadados ou fragmentos de dados. A tecnologia disponível para reidentificar os indivíduos anônimos na base se configura como uma ameaça à privacidade dos dados. Portanto, algumas técnicas conhecidas de anonimização, tais como mascaramento, mesmo que efetivas se aplicadas a um conjunto menor de dados ou a bases de dados isoladas, dificilmente são únicas e suficientes para o contexto do *Big Data* [141].

Este quarto fator, especialmente o último ponto, contribui para o fato de que não é possível sustentar a crença de que anonimização é um método eficaz e suficiente para garantir a privacidade em contextos de *Big Data*. Ainda que seja inferido, a partir das leis supracitadas, que os dados anônimos não podem ser mais identificados (relacionados a um indivíduo identificado/identificável), tornando o tratamento de tais dados mais maleável, deve-se levar em consideração a gestão consciente desses dados. Isto porque o conhecimento sobre riscos envolvendo ferramentas de anonimização induzem ao questionamento da suficiência de tais técnicas à proteção do DP, principalmente quando o tratamento de tais dados é afastado das boas práticas de gerenciamento.

Em 2016, Farvera e da Silva [43] discutiam as ameaças veladas à privacidade de dados na era do *Big Data*. No mesmo ano, Mehmood et al. [122] conceituava os métodos e técnicas para proteção e encriptação de dados dentro do *Big Data*, assim como classificavam algumas formas de anonimização. Portanto, é possível observar que desde este período, já haviam estudos discutindo as aplicações e ameaças da anonimização no ambiente de *Big Data*. De acordo com Mehmood et al. [122], é possível destacar dois tipos de dados: a Informação Pessoal Identificável (*Personally Identifiable Information* - "PII") e o Dado Auxiliar (*Auxiliary Data* - "AD"). Os "PII" identificam de forma unívoca os indivíduos e os registros em uma base de dados. Estes dados podem incluir o que os autores chamam de "quase-identificadores" (*quasi-identifiers*), que são atributos que não identificam univocamente um registro por si mesmos, mas, se associados a outros dados externos, têm o potencial de reidentificar os registros, portanto constituem ameaças à segurança de DP. Os "AD" também podem revelar as pessoas referentes aos dados. Estes dois tipos de dados devem ser tratados de forma separada pela anonimização, de acordo com os riscos inerentes a cada um. Para exemplificar essa descrição, Mehmood et al. [122] apresentaram as informações associativas ("quase-identificadores") de registros de uma aplicação médica e de uma aplicação de avaliação de filmes, conforme apresentado na Figura 2.1.

Ainda em 2016, Lin et al. [113] apresentou um modelo considerando a privacidade diferencial (outra forma de proteger a privacidade de dados). Citando a fragilidade dos métodos de anonimização, Lin et al. [113] aplicou este modelo de privacidade diferencial para uma rede de sensores corporais usando *Big Data* sensitivo. No referido trabalho, Lin et al. [113] combinaram estraté-

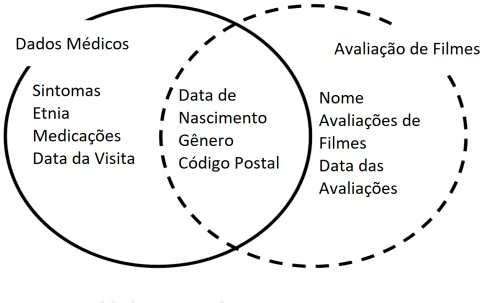

Figura 2.1: Quase-identificadores e registros de ligação (Fonte: [122])

gias de anonimização, agregação e adição de ruídos - números 3 e 4 (Figura 2.2) - para fortalecer a privacidade de determinado conjunto de dados. O esquema adotado pelos autores considerou apenas a informação disponível internamente na base de dados, ignorando possíveis ataques envolvendo AD disponíveis na *internet*, por exemplo. Os autores também discutiram o risco da perda de dados através do processo de anonimização.

Para tentar conciliar a extração de valor de plataformas *Big Data* e a conformidade com a lei GDPR [151], Hintze and Eman [79] defenderam a anonimização como uma possível solução. Os autores defenderam a adoção generalizada da anonimização após o tratamento do DP, para a permanência do dado com o gestor. Alertando sobre o uso de anonimização em *Big Data*, Brasher [24] apresentou algumas fragilidades do "atual" processo dessa ferramenta em bases de dados massivas. A pesquisa de Brasher [24] apresentou as cinco formas mais comuns das técnicas de anonimização: (1) **Supressão** (*Suppression*), (2) **Generalização** (*Generalization*), (3) **Agregação** (*Aggregation*), (4) **Adição de Ruído** (*Noise Addition*) e (5) **Substituição** (*Substitution*), como exposto na Figura 2.2.

1) **Supressão** é o processo que exclui qualquer PII da base [24].

2) **Generalização** mistura os identificadores PII, sem haver exclusão de qualquer informação, reduzindo sua capacidade de associação [24].

3) Na **Agregação**, ambos os tipos de dados (PII e AD) passam por algum tratamento de redução que mantém alguma propriedade do dado (média, distribuição estatística, ou outra) e também reduz sua capacidade de associação [24].

4) **Adição de Ruído** adiciona algum dado não-produtivo para confundir a associação entre os PII/AD e seus sujeitos [24].

5) **Substituição**, que é similar à Generalização, mas difere-se por: ao invés de misturar os identificadores, esta estratégia mescla os valores do dado em si, substituindo o conjunto de dados original por outros parâmetros. Essa estratégia pode ser aplicada para ambos os tipos de dados, Informação Pessoal Identificável e Dados Auxiliares [24].

Finalmente, em 2019, a avaliação de Brasher foi resumida por Domingo-Ferrer [50], que apresentou os problemas da anonimização e suas peculiaridades nas plataformas de *Big Data*. Domingo-

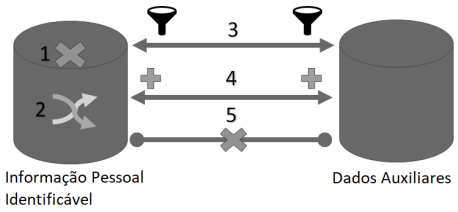

Figura 2.2: Técnicas de Anonimização. (Fonte: adaptado de [24, 122])

Ferrer critica a Supressão (estratégia 1 da Figura 2.2). De acordo com o autor, anonimizar dados no *Big Data* não é suficiente, pois religar os identificadores deletados é uma tarefa trivial no contexto de dados massivos, especialmente quando há a inclusão de dados externos na análise. As preocupações sobre o impacto social dessa proteção insuficiente são tão grandes quanto aparecem na grande mídia [50]. O autor explicou que, para uma proteção da privacidade eficiente, deve ser considerado o balanço entre dois aspectos: a perda de utilidade e o ganho da privacidade nos dados baseados em PII. Um suposto ganho de privacidade ocorre ao custo da perda de utilidade. Quando uma porção de dados suprimidos é descartada, uma quantidade menor de informação válida está disponível para exploração [50]. Então, a anonimização em *Big Data* permanece limitada [50]. Domingo-Ferrer apresentou três principais limitações para os atuais processos de anonimização em *Big Data*:

1) A confiança no controlador do dado, garantida e codificada nas Regulações, que é debilitada devido à lacuna de critérios de uma gestão litigável para o tratamento da confidencialidade.

2) O custo de utilidade da anonimização de dados, que pode incorrer na dificuldade de agregar e explorar o dado anonimizado.

3) A fragilidade dos métodos de anonimização, que satisfazem um conjunto insuficiente de Controles de Distribuição Estatística (SDC).

É importante ressaltar que Mehmood et al. [122] e Domingo-Ferrer [50] concordaram sobre a relação custo-benefício entre privacidade pela anonimização e utilidade, e sua relação negativa no contexto específico de *Big Data*.

2.3 GOVERNANÇA DE DADOS

Alguns trabalhos têm direcionado a solução da *compliance versus Big Data* à governança de dados, entendendo que o correto desenvolvimento dos controles preconizados pelas boas práticas, assim como o correto conhecimento e classificação dos dados contidos nas bases de dados, guia o impacto positivo na proteção à privacidade de dados [14, 15, 50, 63, 64, 137, 156, 182].

De acordo com Fernandes e Abreu [63], dentre os aspectos essenciais da governança corporativa nas organizações modernas, destacam-se o papel da segurança e qualidade de dados. Para os autores, as empresas modernas podem ser valoradas pela sua capacidade de proteger seus dados, revesti-los de qualidade, e produzir informações confiáveis, precisas, acessíveis e disponíveis

tempestivamente.

Sobre a evolução das técnicas e uso de dados, Rego [150] mencionou a crescente disparidade entre a evolução com que os profissionais de dados propõem soluções de tratamento de dados e a dinamicidade dos requisitos dos ambientes de negócios nas empresas. Segundo Rego [150], principalmente a partir década de 2010, o papel da governança de dados tem sido replicar as práticas bem sucedidas da década passada, assim como reconhecer os erros passados, esboçando através do conhecimento de tais o caminho futuro para a aderência às necessidades e evoluções das áreas de negócios.

Em 2014, Fernandes e Abreu [63] alegavam que a governança de dados se tornaria um requisito regulatório, e que as práticas de governança de dados deveriam ser comprovadas a órgãos fiscais e de auditorias regulares [63]. De fato, o advento da LGPD [42] pode ser visto como o início de um movimento para a implantação de requisitos legais para governança de dados, privacidade, segurança e qualidade de dados e informações nas organizações [182]. Ainda sobre previsões, Fernandes e Abreu [63] também viam um papel centralizador (como o encarregado citado pela LGPD no artigo 23 [42]), responsável por relatar (inclusive publicamente) riscos relativos à qualidade de dados e gerenciar o uso da informação.

Segundo Fernandes e Abreu [63], há uma diferença entre a Governança da Tecnologia da Informação e Comunicação (TIC) e a Governança de dados: enquanto a primeira destina-se à manutenção do portfólio de serviços, projetos e infraestrutura de TIC, a segunda é composta por um grupo multifuncional especialista nos assuntos de negócios e técnicos, capaz de se responsabilizar por decisões inerentes a dados. Para os autores [63], a Governança de Dados demanda conhecimentos específicos (distintos da Governança de TIC) e um conhecimento especial da estruturação dos dados na organização a fim de que seja possível compreender os dados e as técnicas empregadas para planejar, modelar, criar, manter, integrar e distribuir dados.

Segundo Thomas [178], a Governança de Dados é um sistema de tomada de decisão e responsabilidades para os processos relacionados aos dados, executado de acordo com políticas, normas e restrições (também mantidas por este sistema). Rego [150] apresentou a Gestão de Dados (principalmente após o período de declínio da Administração de Dados, tido pelo autor sob a alcunha de "Terceiro estágio") como o movimento com o propósito de delimitar uma função mais abrangente que a Administração de Dados (segundo o próprio autor, baseada na tríade Modelagem de Dados, Administração de Bancos de Dados e Gestão de Modelos de Dados), não limitada à administração/gestão dos metadados no ciclo de desenvolvimento de sistemas, mas sim com foco no ciclo de vida do dado, gerenciando melhor dados e metadados. Para o autor [150], essa inovação permite o compartilhamento de responsabilidades entre as áreas de TI (desenvolvimento) e as demais áreas de negócio das empresas.

A *Data Management International* (DAMA) define Governança de Dados como uma disciplina que trata do planejamento, supervisão e controle sobre o gerenciamento de dados e de uso de dados. A Governança de Dados é o exercício da autoridade, controle e tomada de decisão compartilhada sobre a gestão de ativos de dados, através do *Planejamento do Gerenciamento de Dados*

e *Supervisão do Gerenciamento de Dados* [23].

O foco da Governança de Dados, principalmente nos tempos atuais com o movimento para preservação da privacidade e segurança de dados e autonomia do titular de dados, pode ser (sem se restringir a) a privacidade, a conformidade e a segurança da informação. Entretanto, é comum que assuntos como arquitetura, integração de dados, qualidade de dados, inteligência artificial e de negócio, DevOps entre outros, sejam envolvidos neste processo [63].

Segundo o *Data Governance Institute* (DGI) [178], a implantação de um programa de Governança de Dados deve ser guiada por um levantamento das necessidades de Gestão de Dados dentro da organização, tal qual permita uma delimitação de escopo para a atuação da Governança de Dados. Desta forma é possível perceber na literatura [23, 63, 150, 178] que a governança (gestão, como sinônimo) de dados envolve, de maneira abrangente, aspectos como conformidade legal, cadeia de responsabilização, definições estratégicas internas e gestão de relacionamento externo às organizações no que diz respeito ao dado, metadado, privacidade, segurança e qualidade.

Podemos, portanto, diferenciar a Governança de Dados da Gestão de Dados pelo seguinte aspecto: enquanto a Governança tem por objetivo a autoridade e controle sobre os ativos de dados, a Gestão implementa as ações de controle definindo o planejamento, a execução e a fiscalização das políticas, práticas e projetos que envolvem tratamento de dados e informações. Apesar dessa definição, é comum o uso de ambos os termos como sinônimos, ora representando ações de controle, ora exercendo as ações de tratamento sobre o dado.

2.3.1 Objetivos da Governança de Dados

Apesar da variação das visões da governança e dos modelos, são objetivos comuns de projetos em Governança de Dados [178]:

- Permitir um aproveitamento de uso de dados maior para tomada de decisão.
- Reduzir o atrito operacional causado por falta de consumo de dados.
- Auxiliar a proteção das necessidades das partes interessadas.
- Empregar uma visão holística no tratamento de problemas de dados.
- Definir e manter padrões, processos, metodologias de forma corporativa.
- Reduzir custos e aumentar a eficácia de ações institucionais.
- Assegurar a transparência dos processos.

A DAMA define algumas tarefas para as duas áreas que compõem a Governança de Dados, sendo elas [23]:

- Para o Planejamento do Gerenciamento de Dados:
  - Planejamento e Gestão de Dados.
  - Identificação de necessidades estratégicas de dados.

  - Desenvolvimento e manutenção da estratégia de gerenciamento de dados.
  - Estruturação e implementação da função de gerenciamento de dados, bem como a orientação dos profissionais responsáveis por esta função.
  - Identificação e nomeação dos profissionais que exercerão os papéis inseridos no contexto de dados, tais como administradores de dados, administradores de metadados, administradores de bancos de dados, dentre outros.
  - Estabelecimento da função de Gestor da Informação e orientação do trabalho relativo a esta função.
  - Desenvolvimento, revisão e aprovação das políticas de dados, normas e procedimentos, revisão e aprovação da arquitetura de dados.
  - Estimativa do valor dos ativos de dados e os custos associados ao seu gerenciamento.
- Para a Supervisão do Gerenciamento de Dados:
  - Supervisão das áreas e dos profissionais relacionados ao Gerenciamento de Dados.
  - Coordenação das atividades de Governança.
  - Gerenciamento e resolução dos problemas de dados.
  - Monitoramento e garantia da conformidade regulatória, contemplando políticas, normas e arquitetura de dados.
  - Supervisão e gestão de projetos e serviços de dados.
  - Promoção do valor dos ativos de dados.

2.3.2 Dados

Em relação a definição formal do elemento primordial da Governança/Gestão de Dados, segundo Rego [150], quando nos referimos a "dados" nas organizações, geralmente temos em mente a base da matéria-prima necessária para utilizar informações como insumos para decisões tempestivas e acuradas. E, frequentemente, isso só é possível através de um ciclo de evolução capaz de fazer o dado amadurecer para seu uso [150]. O autor [150] define então quatro estágios graduais e progressivos de amadurecimento: **Dado**, **Informação**, **Conhecimento** e **Sabedoria**.

De acordo com Rego [150], o Dado é a representação de um fato fornecido através de caracteres primitivos e isolados (textos, números, imagens, sons, vídeos), essencialmente desprovidos de contexto. O autor [150] ainda faz menção aos Metadados, sendo estes, representações do significado do dado, tanto de forma técnica, tais como a estrutura, formato, tamanho e restrições "físicas" inerentes ao dado, como de forma subjetiva, tais como informações sobre definições, conceitos, relevância e regras de negócio dos dados envolvidos (metadados de negócio). Barbieri [15] complementa a definição de dado como a representação na sua forma mais orgânica e elementar (como um *byte* 00100110 representando o número 38), informando bem pouco sobre o

fato ocorrido. Ainda sobre Metadados, Rego [150] defende que uma boa manutenção de metadados contribui diretamente para a melhoria da qualidade das informações e para o amadurecimento da cadeia de evolução dos dados.

Dando seguimento as quatro etapas de evolução do dado, a Informação se traduz na introdução de um dado contexto ao dado, processados com algum significado e reduzindo a incerteza sobre alguma coisa, estado ou evento [150]. Rego [150] defende que o uso de metadados para leitura e interpretação dos dados (dado + metadado) amadurece a análise evoluindo para o nível de informação. Barbieri [15] concordou com essa definição, acrescentando que a informação provê mais detalhes sobre o fato do que o dado "frio".

O nível de amadurecimento do dado de Conhecimento é o insumo para soluções de problemas e tomadas de decisão [150]. Portanto, a informação acrescentada de significado, premissas, padrões de comportamento, tendências e valores agregados, obtidos por regras de manipulação e características intrínsecas, aumenta o amadurecimento do dado, elevando o nível de conhecimento [150]. Já para Barbieri [15], o conhecimento é o ato de entender a coisa, por meio da razão, do experimento ou da experiência. Envolve sinapses cerebrais, agregando elementos à interpretação.

Por fim, segundo Rego [150], a sabedoria é o uso prático (e contínuo) do conhecimento com eficácia e eficiência. Possui como premissas a avaliação constante do conhecimento adquirido em cada processo, bem como da qualidade e confiabilidade das ações tomadas baseadas em dados. Segundo o autor [150], poucas empresas atingem esse nível, apesar da inteligência analítica (nível de amadurecimento anterior) já ser uma realidade dentro do mercado há um bom tempo. O autor [150] também cita a imperícia dos profissionais de dados na extração e uso de dados como uma possível causa do fracasso no atingimento deste nível de maturidade.

Ainda sobre a definição de sabedoria, Barbieri acrescenta o elemento tempo, a experiência acumulada, a vivência e o empirismo, as percepções e os pontos de vistas populares [15]. Para o autor [15], a sabedoria só é obtida de forma empírica através da aplicação prática do conhecimento no tempo, definindo o ponto ótimo de eficiência na extração de valor deste conhecimento. Barbieri [15] também faz uma breve discussão sobre a confusão na literatura dos termos dado e informação que, apesar de possuírem definições distintas, frequentemente são utilizados como sinônimos principalmente no contexto de governança de dados/informações e na visão de controle destes recursos. Apesar disso, o autor cita que termos em ascensão como *Big Data*, *Chief Data Officer* (CDO), ou Encarregado, como as leis brasileiras o refere, entre outros, têm fortalecido o uso do termo dado para endereçar os ativos de conhecimento.

Rego [150] mencionou que 85% das empresas ainda estão no estágio de tratamento dos ativos de dados como informação ou conhecimento, ainda não sendo capazes de atingir o nível de sabedoria (competência + experiência) com ações práticas fomentadas por dados. O autor [150] também defende que o principal, mas não único, papel responsável por fomentar este ciclo de maturidade é o "cientista de dados", analisando grandes volumes de dados e descobrindo novas tendências e conjuntos de informações e combinações que agreguem valor às empresas. Barbieri [15] apresentou uma classificação do dado em quatro categorias a partir da sua forma e gerência,

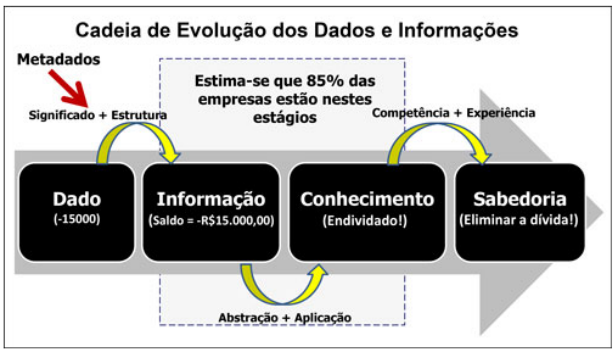

Figura 2.3: Cadeia de Evolução dos Dados e Informações. (Fonte: [150])

sendo elas:

1. **Dados Mestres**: Dados bases/pilares das empresas. Costumam ser mais estáveis, podem ser chamados de dados de fundação e dão origem aos dados transacionais [15].
2. **Dados Referenciais**: Dados "meio-primos" dos Dados Mestres. São mais voláteis que os dados mestres e, portanto, requerem uma gerência especial. Geralmente têm o objetivo de padronização, obtidos de fontes externas definidas por entidades oficiais (podendo também ser gerados internamente) e têm forte associação com dados mestres [15].
3. **Dados Transacionais**: Dados dinâmicos, resultados das transações de negócio das empresas. Geralmente definem relações entre os dados mestres e têm valores e cálculos como atributos [15].
4. **Dados Históricos**: Originalmente todos os demais tipos de dados, quando guardados em uma linha do tempo, representam métricas ou medidas capturadas (a partir de dados transacionais) e compõem *Data Warehouses* e *Data Marts* permitindo tomadas de decisão. A principal dimensão desses dados é o tempo que permite a remontagem da variação temporal do dado [15].

Além das quatro categorias de gerência de dados, os dados podem ser divididos por seu formato armazenado, sendo eles: **Estruturado**, que possui um formato rígido e definido; **Semi-estruturado**, possuindo alguma liberdade de forma de armazenamento; e **Não estruturado**, que possui formatos diversos de armazenamento. Outras possíveis divisões seriam quanto à origem dos dados (**Internos** ou **Externos**), ou ainda quanto à gênese (**Primários** ou **Derivados**) sendo Primários os dados básicos e Derivados os dados produzidos a partir dos básicos [15].

2.3.3 Ciclo de Vida do Dado

Segundo Brackett e Earley [23], o dado segue determinadas etapas (sequenciais ou não) de uso as quais denominamos ciclo de vida, que podem compreender: extração, exportação, importação, migração, validação, edição, atualização, limpeza, transformação, conversão, integração, segregação, agregação, referenciação, revisão, relatação, análise, garimpo, armazenamento, recuperação, arquivação, restauração e eliminação.

Brackett e Earley [23] e Rego [150] afirmaram que o ciclo de vida do dado possui forte e direta relação com o ciclo de desenvolvimento das aplicações, uma vez que geralmente o tratamento de dados ocorre por meio do uso (e desenvolvimento/teste) das aplicações. Apesar disso, o ciclo de vida do dado é (a partir de certo ponto) independente do ciclo da aplicação e pode, inclusive, perdurar, mesmo que a aplicação já tenha encerrado seu ciclo [150]. Rego ainda defende que apenas após a etapa de entrega dos dados (correspondente às etapas finais do ciclo de desenvolvimento do software) é que de fato os dados começam a agregar valor ao negócio [150]. Quanto à correspondência dos ciclos de vida do dado e das aplicações, Rego apresenta uma comparação visual de tais ciclos, conforme apresentado na Figura 2.4.

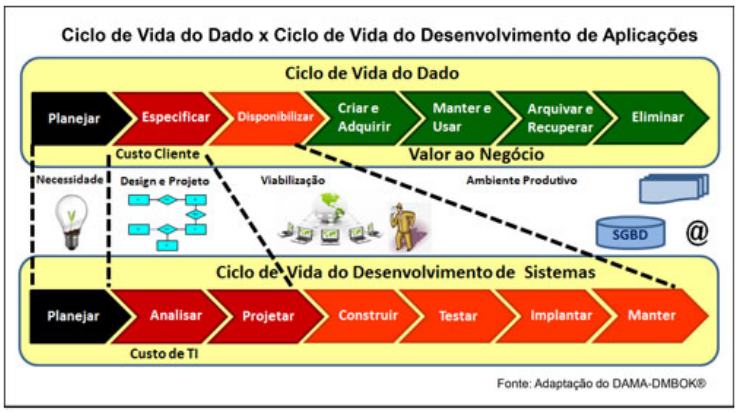

Figura 2.4: Ciclo de Vida do Dado versus Ciclo de Vida do Desenvolvimento de Aplicações. (Fonte: [150] adaptado de [23])

2.3.4 Qualidade de Dados

Rego [150] mencionou que apenas com dados de qualidade os investimentos realizados em gestão e governança de dados podem resultar em retorno financeiro e/ou social para as organizações. No mesmo sentido, Barbieri [15] mencionou que a propriedade única dos dados (como ativo das empresas) de poderem ser replicados, copiados e manipulados entre camadas transacionais e informacionais (níveis operacionais e gerenciais), confere um aspecto de "profilaxia" aos dados: garantir a qualidade aos dados, evitando a proliferação de elementos incompletos, campos sem significado, dados com erro ou com precisões duvidosas contribui sensivelmente para tomadas de decisões corretas, fundamentais no mundo informacional.

Apesar da divergência da literatura sobre as dimensões que compõem a qualidade de dados, o autor [150] mencionou oito dimensões de qualidade de dados e metadados baseadas nas onze dimensões presentes no guia DAMABOK [23]. São elas:

1. **Aderência ao negócio**: grau de conformidade dos dados e metadados com requisitos de informação e regras de negócio da empresa.
2. **Unicidade**: inverso ao grau de redundância da informação. Indica que não há repetição de conteúdo ou conceitos, exclusividade de dados e metadados.
3. **Integridade**: grau de conformidade com todas as restrições de integridade definidas no

metadado, necessárias para que o dado possa ser considerado confiável. Permite a representação das regras de negócio, sem as quais há prejuízo à confiabilidade.

4. **Confiabilidade**: grau de atualidade, corretude e utilidade livre risco de efeitos negativos. Indica que não há erros de transformação, disponibilização e até digitação.
5. **Manutenibilidade**: grau de facilidade com o qual os dados e metadados podem ser mudados quando necessário. A ferramenta de Modelos de Dados quando utilizadas com modelos flexíveis pode contribuir para baixo custo de manutenção.
6. **Performance**: grau de conformidade do tempo de resposta e acesso aos dados com o requisito de uso.
7. **Legibilidade**: grau de facilidade de entendimento, compreensão e utilização de dados e metadados. Novamente, a ferramenta de Modelos de Dados, se usada definindo nomenclaturas adequadas às estruturas, acarreta em metadados legíveis.
8. **Disponibilidade**: grau de conhecimento e acessibilidade de dados e metadados, no momento necessário, para quem tem o devido acesso. Envolve conceitos de governança e segurança da informação.

Já Barbieri [15] define as seguintes oito dimensões da qualidade de dados:

1. **Completude**: grau de completeza dos atributos de dados.
2. **Unicidade**: grau de não redundância.
3. **Razoabilidade**: grau de conformidade com aspectos lógicos da informação. Dessa forma, existiria uma avaliação se a informação é possível, ou se é verossímil.
4. **Integridade**: grau de fidelidade da informação (duas fontes distintas divergentes).
5. **Temporalidade**: grau de disponibilidade e conformidade formal na disposição do dado.
6. **Validade**: grau de coerência de valores (estrutural).
7. **Cobertura**: grau de atingimento dos objetivos dos dados.
8. **Precisão**: grau de acurácia da informação (dados seus objetivos).

2.3.5 Modelos/*Frameworks* de Governança de Dados

Para lidar com o precioso ativo de dados e a complexa diversidade de ações necessárias para sua proteção/tratamento, a comunidade científica e o mercado têm empreendido esforços para determinar modelos e *frameworks* para gestão e governança de dados. Segundo Fernandes e Abreu [63], podem ser apresentadas pelo menos três referências de soluções para utilização neste contexto: *CobiT* [83], *DAMA DMBOK*[23] e *DGI Framework*[178].

O modelo *Control Objectives for Information and related Technology* (CobiT), apesar de ser um *framework* de governança e gestão de TI, é uma referência de padrões e boas práticas relacio-

nadas a tecnologias da informação para alinhamento da TI às estratégias de negócio através do gerenciamento e governança corporativa [83]. Para Fernandes e Abreu [63], duas práticas deste modelo (versão 5) estão diretamente relacionadas com questões de gestão de informações, sendo elas:

- **APO01.06 - Definir a propriedade de informações (dados) e sistemas (do processo APO01 - Gerenciar o *framework* de gestão de TI)**. Segundo essa prática, é necessário definir e garantir a manutenção de responsabilidades pela propriedade de informações e sistemas de informações. Além disso, cada responsável deve tomar decisões sobre a classificação das informações e sistemas em suas alçadas para otimizar a proteção, integridade e consistência dos dados.
- **APO03.02 - Definir a arquitetura de referência (do processo APO03 - Gerenciar a arquitetura corporativa)**. Com relação a esta prática, é necessário o estabelecimento e manutenção de um modelo de informação corporativo alinhado à estratégia corporativa com o qual seja possível a disponibilização de informações úteis e de valor para a tomada de decisão. Este modelo é uma ferramenta de comunicação e compartilhamento de informações na empresa. Além disso, faz-se necessária a compilação de um dicionário corporativo, a fim de prover a sincronia de regras de sintaxe e semântica de dados. Esta ferramenta também serve de guia para a classificação dos dados, definição de níveis de segurança apropriados e requisitos de retenção e destruição de dados (conforme o ciclo de vida dos dados apresentado - Figura 2.4).

O *CobiT* também possui uma publicação de um guia sobre estruturação do pensamento acerca da informação e questões relacionadas à governança e gerenciamento da informação [83]. Segundo o *CobiT 5: Enabling Information* [83], a informação é um ativo corporativo e um importante habilitador da governança e do gerenciamento. São ainda objetivos da chamada governança da informação: alinhar as metas corporativas para alcançar as necessidades, condições e opções das partes interessadas por meio da aquisição e do gerenciamento de recursos de informação; priorizar o gerenciamento da informação com foco à tomada de decisão; e monitorar e controlar o desempenho e conformidade dos recursos de informação em relação às políticas, padrões, arquiteturas e procedimentos [83]. Ainda segundo o guia, há uma sugestão de métricas para avaliar a qualidade de itens de informação, dentre as quais podem ser citadas [63]:

1. Tempo decorrido desde a última atualização do Plano Estratégico de TI;
2. Quantidade de incidentes significativos relacionados à TI não identificados na avaliação de riscos;
3. Percentual de políticas suportadas por padrão e práticas efetivas de trabalho;
4. Quantidade de horas de treinamento/aprendizado por pessoa;
5. Grau de satisfação dos usuários de negócio com o treinamento e o manual do usuário;

As versões mais atuais do *framework CobiT*, tais como a versão 2019, publicada em dezembro

de 2018, possuem estruturas específicas para a gestão e governança de dados, em detrimento da versão aqui utilizada, que encapsula tais resultados na gestão e governança de TI.

O *DAMABOK* [23], outro modelo relacionado por Fernandes e Abreu [63], define algumas metas, funções, atividades, entregas primárias, papéis, princípios, tecnologias e questões culturais envolvidas na Gestão de Dados. Segundo a *DAMA*, o *DAMABOK* [23] é um corpo de conhecimento (semelhante ao *PMBOK* para gestão de projetos ou o *SWEBOK* para engenharia de software) que tem por principal objetivo apoiar a profissão do gestor de dados. O guia, portanto, é inteiramente dedicado a descrever como as empresas devem aplicar boas práticas para o tratamento dos dados, dentre os quais, os DP, relacionando as responsabilidades compartilhadas entre TI e negócio, ou descrevendo domínios de conhecimento que devem ser desenvolvidos na organização para alcançar maturidade no tratamento de dados [23]. O escopo de atuação do guia pode ser resumido de acordo com a Figura 2.5.

Figura 2.5: Funções da Gestão de Dados - Escopo sumarizado. (Fonte: [23])

Rego [150] esclareceu que a nova versão (v2) do *DAMABOK* (ainda não traduzida pelo capítulo Brasil da *DAMA*) traz uma nova função a ser acrescentada - Integração de Dados - pela qual se define estratégias, ferramentas, políticas, papéis, métodos e controles para a multi-operação de dados nos mais heterogêneos formatos, recursos de armazenamento, sistemas e afins. Além disso, para a *DAMA* [23], a Gestão de Dados pode ser resumida de acordo com as Tabelas 2.10 e 2.11.

Alguns aspectos decorrem da definição de Gestão de Dados presente no *DAMABOK* [23]:

1. Um dos objetivos da Gestão de Dados é otimizar o valor do dado na organização, bem como o enriquecimento do valor e a obtenção efetiva de vantagem através da exploração

Gestão de Dados

Tabela 2.10: Diagrama de contexto de Gestão de Dados - Definições. (Fonte: [23])

Definição	O planejamento, execução e fiscalização das políticas, práticas e projetos para adquirir, controlar, proteger, entregar e enriquecer o valor dos ativos de dados e informações
Missão	Garantir disponibilidade, qualidade e segurança necessárias para todos os *stakeholders*
Objetivos	1. Entender quais são as informações necessárias para a organização e seus *stakeholders* 2. Capturar, armazenar, proteger, garantir a integridade dos ativos de dados 3. Incrementar continuamente a qualidade dos dados e informações 4. Assegurar a privacidade e confidencialidade, visando evitar o acesso não autorizado ou inapropriado 5. Maximizar a efetividade do uso dos ativos de dados e informações

Tabela 2.11: Diagrama de contexto de Gestão de Dados - Fluxo. (Fonte: [23])

• **Entradas:** – Estratégias do negócio – Atividades do negócio – Atividades de TI – Questões de dados • **Fornecedores:** – Executivos – Criadores de dados – Fontes externas – Entidades Regulatórias • **Participantes:** – Criadores de dados – Consumidores de informações – Gestores de dados – Profissionais de dados – Executivos	• **Funções:** – Governança de dados – Gestão de Arquitetura de dados – Desenvolvimento de dados – Gestão de operações de dados – Gestão da segurança de dados – Gestão de Dados mestres e de referência – Gestão de DW e BI – Gestão da documentação e conteúdo – Gestão de metadados – Gestão da qualidade de dados • **Ferramentas:** – Ferramentas de modelagem de dados – Sistemas de gerenciamento de banco de dados – Ferramentas de qualidade e integração de dados – Ferramentas de BI – Ferramentas de gerenciamento de documentação – Ferramentas de repositório de metadados	• **Entregas primárias:** – Estratégias de dados – Arquitetura de dados – Serviços de dados – Banco de dados – Dados, informação – Conhecimento e competência • **Consumidores:** – Trabalhadores de escritório – Trabalhadores do conhecimento – Gerentes – Executivos – Clientes • **Métricas:** – Métricas de valor de dados – Métricas de qualidade de dados – Métricas de Gestão de Dados

desse valor;

2. A Gestão de Dados é responsável por definir a atuação (papéis, políticas, ferramentas, técnicas) da qualidade de dados;
3. A privacidade de dados também é responsabilidade da Gestão de Dados;
4. A Gestão de Dados engloba boa parte da empresa (de consumidores de informação aos executivos) em diferentes níveis (tático, operacional e estratégico);

Ainda sobre o modelo *DAMA* [23], há uma definição resumida de governança de dados, de acordo com as Tabelas 2.12 e 2.13:

Governança de Dados

Tabela 2.12: Diagrama de contexto de governança de dados - Definições. (Fonte: [23])

Definição	O exercício de autoridade e controle (planejamento, monitoramento e engajamento) sobre o gerenciamento de ativos de dados
Objetivos	1. Definir, aprovar e comunicar estratégias de dados, políticas, padrões, arquitetura, procedimentos e métricas 2. Acompanhar e forçar o cumprimento de regulatórios e conformidades com políticas de dados, padrões, arquiteturas e procedimentos 3. Patrocinar, acompanhar e supervisionar as entregas de projetos e serviços de Gestão de Dados 4. Gerenciar e resolver questões relacionadas a dados 5. Entender e promover o valor dos ativos de dados

Fernandes e Abreu [63] mencionaram que os modelos *CobiT* e *DAMABOK* são complementares e podem ser usados em conjunto para estabelecimento da governança e gerenciamento da informação (usando o alto grau de detalhamento prático do *DAMABOK* aliado ao conceito de habilitador encontrado no modelo *CobiT*).

Por fim, Fernandes e Abreu [63] fazem menção ao modelo *DGI framework* [82] consolidado pelo *Data Governance Institute* (DGI) que, segundo os autores, auxilia as empresas a empreenderem iniciativas de Governança de Dados e compreenderem questões vitais como: missão da GD, escopo de trabalho, papéis funcionais envolvidos e suas interações e processos a serem executados. O *DGI framework* [82] possui dez componentes, conforme apresentado na Figura 2.6.

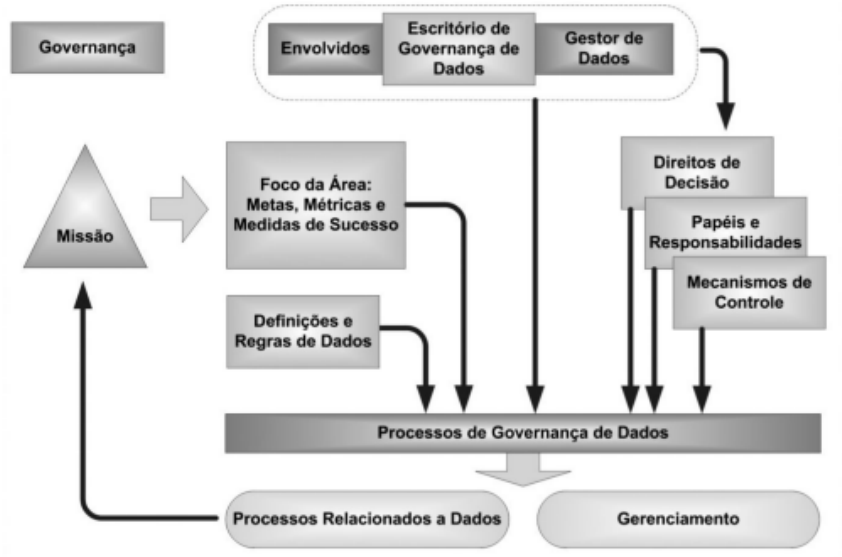

Figura 2.6: Componentes do *framework* do DGI. (Fonte: [63])

1. **Missão:** A missão da GD geralmente se divide em três partes, sendo elas, definir e alinhar

Tabela 2.13: Diagrama de contexto de governança de dados - Fluxo. (Fonte: [23])

• **Entradas:** – Metas do negócio – Estratégias do negócio – Objetivos de TI – Estratégias de TI – Necessidades de dados – Questões de dados – Requisitos regulatórios • **Fornecedores:** – Executivos de negócios – Executivos de TI – Gestores de dados – Entidades Regulatórias • **Participantes:** – Gestores executivos de dados – Coordenadores de Gestão de Dados – Profissionais de dados – Executivos de Gestão de Dados – CIO	• **Atividades:** 1. Planejamento da Gestão de Dados (P) (a) Entender as necessidades estratégicas de dados da organização (b) Desenvolver e manter a estratégia de dados (c) Estabelecer papéis dos profissionais e das organizações de dados (d) Identificar e apontar gestores de dados (e) Estabelecer organizações de gestão e governança de dados (f) Desenvolver e aprovar políticas, padrões e procedimentos de dados (g) Rever e aprovar arquitetura de dados (h) Planejar e patrocinar projetos e serviços de Gestão de Dados (i) Estimar o valor dos ativos de dados e custos associados 2. Controle da Gestão de Dados (C) (a) Supervisionar equipe e organizações profissionais de dados (b) Coordenar atividades de governança de dados (c) Gerenciar e resolver questões relacionadas a dados (d) Monitorar e forçar o cumprimento dos regulatórios (e) Monitorar e forçar a conformidade com as políticas, padrões e arquiteturas de dados (f) Supervisionar projetos e serviços de Gestão de Dados (g) Comunicar e promover o valor dos ativos de dados • **Ferramentas:** – Intranet – E-mail – Ferramentas de metadados – Repositório de metadados – Ferramenta de gestão de questões – *Dashboards* KPI (*Key Performance Indicator*) em Governança de dados Atividades: (P) - Planejamento, (C) - Controle	• **Entregas primárias:** – Políticas de dados – Padrões de dados – Questões solucionadas – Projetos e serviços de Gestão de Dados – Dados e informações qualificadas – Valor do dado reconhecido • **Consumidores:** – Produtores de dados – Trabalhadores do conhecimento – Gestores e executivos – Profissionais de dados – Clientes • **Métricas:** – Valor do dado – Custo de Gestão de Dados – Objetivos atingidos – Quantidade de decisões tomadas – Cobertura/Representação de Gestão de Dados – Quantidade de profissionais de dados – Maturidade em processos de Gestão de Dados

regras proativamente, fornecer serviços de dados a todas as partes interessadas e reagir e resolver problemas de inconformidades às regras;

2. **Foco:** objetivos, métricas e medidas de sucesso, estratégica de financiamento e patrocínio;
3. **Definição de regras de dados:** políticas de dados, normas, conformidade e regras de negócio. O programa de GD é capaz de criar, reunir, identificar lacunas e sobreposições e alinhar e priorizar regras;
4. **Decisões:** estabelecimento do processo de tomada de decisão, além de temporalidade e atores deste processo;
5. **Papéis e Responsabilidades:** quem deve fazer o quê e quando;
6. **Controles:** controle de riscos do gerenciamento de dados;

7. **Partes Interessadas:**

   (a) **Envolvidos:** todos os profissionais incluídos na GD;

   (b) **Escritório de Governança de Dados:** criado para facilitar e apoiar atividades da GD. Coleta as métricas e medidas de sucesso, elabora relatórios e os distribui para as partes interessadas;

   (c) **Administradores e gestores de dados:** conjunto de intervenientes de dados que decidem as questões de dados. Podem definir políticas, normas e recomendações aproveitadas pelos níveis superiores;

8. **Processos:** Alinhamento de políticas, requisitos e controles, processos de decisão, papéis e responsabilidades, governança, gerenciamento de mudanças, definição de dados, resolução de problemas, requisitos de qualidade de dados, integração de GD e TI, gerenciamento de envolvidos, comunicações e medição e reporte do valor.

2.4 TRABALHOS CORRELATOS

Nesta seção serão apresentados os principais trabalhos relacionados com o tema, assim como a relação destes com o propósito desta pesquisa.

Em 2015, Priebe e Markus [144] apresentaram um trabalho sobre uma proposta de metodologia aplicável a projetos de dados, ciência de dados e governança em *Big Data*. Ao propor a metodologia de *Business Information Modeling* (BIM), os autores classificam 10 maneiras possíveis de "explorar" modelos de negócios, sendo elas: Análises de Impacto e Requisitos, Harmonização de Dados, Mapeamento de Dados, Ferramentas ETL (*Extract, Transform, Load*), Gerenciamento de Dados Mestres e de Referência, Suporte a Testes, Gerenciamento de Qualidade de Dados, Catálogo e Linhagem de Dados, Curadoria de Dados e Suporte à Consistência de Relatórios. Os autores documentam que seu modelo foi utilizado e evoluído durante a aplicação no Banco da Irlanda, onde os autores consideram ter validado uma versão da ferramenta funcional. Até a data da publicação, não havia validações quanto aos contextos de ciência de dados e de *Big Data*.

ElShekeil e Laoyookhong [60] apresentaram em 2017 um *framework* que tinha o objetivo de encadear as atividades de privacidade por projeto requerida pela GDPR [151], apresentando as principais boas práticas levantadas por uma revisão da literatura. A versão aprimorada de seu trabalho apresentava as seguintes fases: Preparação (Sistema, Contexto de Negócio, Ambiente circunvizinho e Elaboração do Relatório de Impacto a Privacidade de DP), Avaliação (*Accountability, Purpose Limitation, Storage Limitation, Integrity and Confidentiality, Data Minimization, Accuracy, Lawfulness, Fairness and Transparent*, além do levantamento dos requisitos de privacidade) e Implementação (dos requisitos levantados acima). Os autores submeteram seu *framework* a um estudo de caso aplicado ao desenvolvimento de um *"chatbot"* e concluíram que, apesar de necessidades de aprimoramentos, de maneira geral o seu *framework* contribui para a conformidade com o artigo 25 da GDPR [151].

Já em 2018, Ayala-Rivera e Pasquale [12] propuseram um guia para elicitação de requisitos de privacidade de dados para conformidade legal com os princípios de tratamento de DP impostos pela GDPR [151]. A abordagem proposta pelas autoras é composta de seis fases: Auditoria de Dados, Análise de *"Gap"*, Planejamento e Preparação, Revisão do Plano, Execução e Revisão Pós-Implementação. Também segundo as autoras [12], sua abordagem foi avaliada como uma contribuição para o objetivo de conformidade com a GDPR e também auxilia a comunicação dos requisitos apropriados para as soluções.

Em 2019, Piras et al. [137], como parte de um *framework* completo e um conjunto de ferramentas para compliance da GDPR [151], publicaram sobre o módulo de governança de dados chamado de *Data Management Scope* (DSM). Tal proposição tinha por objetivo preencher algumas lacunas práticas da legislação relacionadas à privacidade de dados. Em especial, o módulo DSM implementa ações que contribuem com os princípios de Privacidade por Padrão e Privacidade por Projeto (resumido no artigo como PbD). Portanto, o trabalho buscava estruturar as atividades de implementação e análise necessárias para alcance da PbD e, se possível, automatizá-las. Vale ressaltar que os autores não restringem a aplicabilidade de sua ferramenta a ambientes de *"small data*", o que nos possibilita a comparação com o tema da pesquisa corrente.

Piras et al. [137] apresentaram uma interessante relação dos conceitos ligados à privacidade de dados através do fluxo de informações implementado na ferramenta (tais como a relação de precedência entre a avaliação de impacto à privacidade e análise de "minimização" de dados, ou a precedência da organização e catalogação de dados a qualquer atividade de proteção de dados). Entretanto, os detalhes técnicos e procedimentais dessa implementação ficaram restritos à ferramenta proposta.

Por fim, Piras et al. [137] concluíram que nem todos os passos relativos à implementação da PbD podem ser automatizados (uma vez que o *framework* proposto delega aos papéis do "analista de negócio" e do "analista de privacidade/segurança" ações relativas à confecção do relatório de análise de impacto, requerido pela GDPR [151]), mas que o uso da ferramenta proposta auxilia (opinião coletada por três *workshops* de avaliação qualitativa de usuários) na implementação de tais práticas. A ferramenta também carece de averiguação quanto à aplicabilidade no contexto da LGPD [42], uma vez que a lei brasileira possui alguns pontos ligeiramente divergentes da sua correspondente europeia [151].

No início deste mesmo ano, Stoll publicava sua proposta de *framework* para governança da privacidade de dados [168], também baseada nos requisitos impostos pela lei Europeia [151]. O "modelo de governança holística de privacidade de dados" proposto por Stoll [168] atende, segundo a autora, principalmente os requisitos da GDPR [151] de: registro do tratamento de dados (dado pelo acompanhamento do modelo); controle do propósito, escopo e contexto do tratamento de dados visível no relatório de impacto e risco da proteção de dados (atendidos na integração de tais requisitos às políticas de privacidade de dados, objetivos e políticas corporativas); os próprios relatórios de impactos e riscos à proteção de dados (explícitos no modelo); direitos dos titulares de dados (integrados às políticas de privacidade de dados); instrução de colaboradores e operado-

res (também integradas às políticas); além do monitoramento constante e efetivo das medidas de privacidade de dados (explícitas no modelo).

Jain et. al. [88] propuseram um modelo de segurança à privacidade de dados em ambientes de *Big Data* ainda na camada de ingestão de Dados (*MapReduce*). Para os autores [88], manter na fase de ingestão de Dados conceitos relacionados à segurança de dados e privacidade de dados (através de encriptação e decriptação de dados) contribui para melhoria geral da segurança e privacidade dos sistemas. No trabalho, os autores [88] concluíram que incluir este processamento adicional pode incrementar o tempo de resposta dos servidores, havendo, entretanto, melhorias em conceitos como uso de processamento, perda de informação e uso de memória, se comparados a outros métodos de segurança.

É importante citar que Priebe e Markus [144], Piras et al. [137] e Stoll [168] concordam em destacar o papel de uma boa governança corporativa de dados para o sucesso de iniciativas de proteção à privacidade de dados, nos mais variados contextos de aplicação.

Outro trabalho relevante (dado o recorte desta pesquisa) produzido em 2019 foi o dos autores Ruan et. al. [155], que propuseram um modelo de detecção de fraude cooperativo (principalmente de aplicativos móveis) para tratar de fraudadores que tomam vantagem da transmissão de ligações telefônicas móveis para esconder seus ataques. O trabalho dos autores [155] também visa prevenir vazamento de dados (ataque à privacidade) de usuários de telefonia móvel. Os autores [155] concluem o trabalho avaliando que seu modelo é bem sucedido num cenário de uso real (com dados reais) e que sua proposta é aplicável aos conjuntos de dados reais. Além disso, eles concluem que seu modelo coopera de forma positiva para manutenção da privacidade de usuários da telefonia móvel, otimizando a detecção de fraudes com alto grau de precisão em situações do mundo real, mesmo protegendo a privacidade dos dados [155].

De acordo com estes trabalhos levantados, há ainda na literatura relativa uma lacuna teórica e prática sobre guias que auxiliem as organizações a implementar projetos de dados com o objetivo de prevenção a fraude, em plataformas analíticas massivas de dados (*Big Data*), promovendo a conformidade legal com os requisitos de proteção à privacidade de dados (principalmente nos termos das leis LGPD [42] e GDPR[151]), fazendo uso das boas práticas de governança de dados, tendo todos estes aspectos unificados e considerados de maneira integrada em um *framework*.

Na Tabela 2.14 apresentamos um quadro-resumo dos assuntos abordados pelos trabalhos correlatos e sua relação com esta pesquisa:

Tabela 2.14: Tabela comparativa trabalhos correlatos. (Fonte: Autor)

	Aplicação em *Big Data*	***Framework* de Privacidade**	**Modelo de Governança**	**Prevenção à Fraude**
Priebe e Markus [144]	X		X	
ElShekeil e Laoyookhong [60]		X		
Ayala-Rivera e Pasquale [12]		X		
Piras et al. [137]		X	X	
Stoll [168]		X	X	
Jain et. al. [88]	X	X		
Ruan et. al. [155]		X		X
Nossa proposta	**X**	**X**	**X**	**X**

3 METODOLOGIA

Neste capítulo é apresentada a metodologia utilizada para conduzir essa pesquisa, bem como os mecanismos desenvolvidos e o processo de desenvolvimento de tais mecanismos para a validação dos resultados obtidos.

3.1 REVISÃO BIBLIOGRÁFICA

Inicialmente, foi conduzida uma revisão da literatura para identificar os principais modelos relacionados à Segurança, Privacidade e Governança de Dados. A pesquisa incluiu as bases digitais *Computer Science Bibliography* (DBLP), *Association for Computing Machinery* (ACM), *Springer* e *Institute of Electrical and Electronics Engineers* (IEEE) Xplore, que é uma biblioteca digital associada à organização de mesmo nome. O campo de pesquisa foi delimitado elegendo-se os artigos mais recentes, assim considerados os artigos publicados entre 2015 e 2020, em conferências internacionais, jornais científicos ou livros científicos, indexados pelas bases digitais citadas. A Tabela 3.1 mostra a quantidade de artigos relacionados de acordo com cada base digital.

Tabela 3.1: Levantamento de modelos da revisão de literatura. (Fonte: Autor)

Base Digital	Qtd. Estudos relacionados à Governança	Qtd. Estudos relacionados à Privacidade	Qtd. Estudos relacionados à Segurança
DBLP	16	138	105
ACM	9	1	6
Springer	20	13	15
IEEE Xplorer	82	86	233

Dentre os artigos selecionados, como forma de priorizar os estudos relevantes a essa pesquisa, foram analisados os artigos com aplicabilidade ao contexto de *Big Data* e aplicabilidade ao contexto de prevenção à fraude no sistema financeiro (mesmo cenário de aplicação do estudo de caso). Desta forma, estabelecemos o recorte da nossa pesquisa entre os trabalhos identificados nas mencionadas bases.

3.2 CLASSIFICAÇÃO DE ATIVIDADES

Após uma leitura inicial dos estudos mencionados, foram identificadas e classificadas cada atividade referente às disciplinas de Governança (coluna “Gov”), Proteção à Privacidade (coluna “Priv”) e Segurança/Anonimização da Informação (coluna “Ano/Seg”), conforme apresentado na Tabela 3.2. As atividades também foram classificadas quanto à natureza de atuação dentro

de um ciclo PDCA (*"Plan", "Do", "Check", "Act"*, ou "Planejamento", "Desenvolvimento", "Controle" e "Ação"), conforme apresentado na Tabela 3.2.

Tabela 3.2: Atividades identificadas nos estudos. (Fonte: Autor)

	Atividade	**Disciplina**			**PDCA**				**Referências**
		Gov	**Priv**	**Ano/Seg**	**Planejamento**	**Desenvolvimento**	**Controle**	**Ação**	
A1	Arquitetura de governança (des)centralizada do controlador	X			X				[105]
A2	Definição do Acesso e Propriedade do Dado	X	X	X	X				[31, 71, 86, 89, 93, 105, 119, 128, 157, 179]
A3	Análise do Ambiente Regulatório	X			X				[23, 105, 153, 166]
A4	Análise de Utilidade do Dado	X			X	X			[19, 23, 45, 49, 51, 56, 57, 65, 67, 68, 97, 98, 105, 119, 131, 154, 160, 166]
A5	Análise do Custo-benefício (ROI) do Dado	X	X		X				[19, 23, 45, 49, 51, 56, 57, 67, 68, 97, 98, 105, 129, 130, 131, 165]
A6	Aplicação de Casos de Uso de Dados	X			X				[96, 105]
A7	Análise de Conformidade Legal	X					X		[23, 97, 99, 102, 105, 106, 163, 165, 166, 170]
A8	Monitoração de Processos	X	X					X	[100, 105, 153, 186]
A9	Desenho da Linhagem/-Fluxo/Mapeamento de Dados	X			X	X			[9, 72, 102, 105, 110, 115, 118, 139, 144, 157, 164, 163]
A10	Processo de Qualidade de Dados	X					X		[3, 4, 13, 52, 58, 78, 89, 90, 105, 119, 129, 133, 144, 145, 153, 160, 165, 179]
A11	Política de Dados e Privacidade	X	X	X	X				[1, 23, 33, 66, 68, 86, 93, 97, 100, 108, 128, 153, 159, 189]
A12	Desenho da Estratégia de Dados	X			X				[23, 153, 160]
A13	Planejamento do Ciclo de Vida do Dado	X			X				[8, 23, 95, 108, 110, 140, 148, 153]
A14	Gestão dos Metadados	X			X				[17, 23, 47, 73, 114, 139, 153, 169, 171]
A15	Controle da Infraestrutura	X	X	X		X			[23, 25, 75, 153]
A16	Arquitetura de Segurança da Informação			X	X				[1, 10, 20, 55, 59, 61, 75, 89, 92, 108, 111, 126, 139, 145, 153]
A17	Princípios de Dados e Valor de Dados	X		X	X		X		[9, 17, 68, 95, 130, 146, 147, 153, 154, 160, 165, 187]
A18	Compreender Necessidades Corporativas Estratégicas de Dados	X			X				[23, 160, 165]
A19	Definição de Papéis e Responsabilidades	X	X		X				[23, 86]

Continua na próxima página

Tabela 3.2 – continuando da última página

	Atividade	Disciplina			PDCA				Referências
		Gov	Priv	Ano/Seg	Planejamento	Desenvolvimento	Controle	Ação	
A20	Organização da Gestão e Governança de Dados	X			X				[9, 23]
A21	Gestão de Dados	X			X				[23, 106, 112, 128, 137, 160]
A22	Arquitetura de Dados	X			X				[23, 102, 111, 148]
A23	Projetos e Serviços de Gestão de Dados	X			X		X		[23, 139]
A24	Capacitação de Recursos Humanos	X	X	X	X		X		[23, 86]
A25	Coordenar as Atividades de Governança de Dados	X					X		[23]
A26	Resolução de conflitos	X		X			X		[23, 36, 146, 147]
A27	Modelagem de Dados	X		X		X			[20, 23, 58, 59, 68, 95, 115, 138, 139]
A28	Classificação de Dados	X	X	X		X	X		[17, 58, 65, 86, 159, 160, 163, 175, 187]
A29	Análise de Requisitos e Impacto	X	X				X		[44, 72, 144, 160, 190]
A30	Harmonização de Dados	X					X		[144]
A31	Integração de Dados	X				X	X		[1, 2, 52, 97, 102, 106, 138, 144, 167, 169, 183, 186]
A32	Gestão de Dados Mestres	X				X			[23, 144]
A33	Curadoria de Dados	X			X	X			[23, 144]
A34	Avaliação de Riscos de Segurança da Informação		X	X			X	X	[2, 52, 61, 86, 107, 164, 166, 170, 189]
A35	Planejamento e Uso de Métodos Criptográficos		X	X	X	X			[3, 10, 13, 32, 33, 61, 75, 78, 86, 88, 89, 90, 91, 101, 106, 114, 145, 157, 158, 160, 172, 177, 183]
A36	Definição de Estratégia de Cópias de Segurança (*Backup*)			X	X	X			[86]
A37	Auditoria		X	X			X		[66, 86, 146, 166, 175]
A38	Controle da Transferência de Dados		X	X		X			[32, 86, 89, 179]
A39	Política de Desenvolvimento e Teste Seguro	X	X	X	X				[86, 96, 132, 147]
A40	Definição do Contexto da Organização	X	X	X	X				[23, 86, 137]
A41	Coleta do Consentimento do Uso dos Dados		X	X	X	X			[31, 68, 86, 105, 166]
A42	Definição do Propósito do Tratamento		X		X	X			[6, 25, 86, 111, 128, 160, 166]
A43	Mapeamento de Fontes Externas de Dados		X		X	X			[25, 44]
A44	Definição dos Objetivos da Segurança da Informação			X	X				[25]
A45	Definição dos Serviços de Autenticação			X	X	X			[3, 11, 13, 25, 65, 75, 78, 81, 86, 89, 114, 116, 145, 149, 155, 157, 169, 175, 184]

Continua na próxima página

Tabela 3.2 – continuando da última página

	Atividade	Disciplina			PDCA				Referências
		Gov	Priv	Ano/Seg	Planejamento	Desenvolvimento	Controle	Ação	
A46	Deidentificação (*Unlinkability*)			X		X			[25, 33, 98, 147]
A47	Desenvolvimento do Serviço de Proteção a Dados			X		X			[25, 77, 78, 92, 106, 175]
A48	Uso de Ferramentas de Privacidade Diferencial e K-Anonimização			X		X			[7, 18, 19, 26, 30, 35, 37, 44, 45, 49, 51, 56, 57, 67, 69, 94, 97, 100, 109, 117, 119, 120, 136, 152, 155, 166, 174, 185, 188]
A49	Uso de Ferramentas de Inteligência Artificial Aplicadas e Aprendizado de Máquina à Segurança (tais como Modelo Oculto de Markov, Computação Bayesiana Aproximada e Redes Bayesianas, Aprendizado Federado, Redes Neurais, Blockchain, Mineração de Dados, entre outras)			X		X			[19, 55, 69, 76, 80, 90, 106, 107, 121, 131, 134, 167, 176, 188, 192]
A50	Definição de Métricas de Riscos à Privacidade		X				X		[8, 26, 31, 44, 72, 87, 97, 103, 119, 145, 160, 161, 163, 190]
A51	Uso de Ferramentas de Mascaramento de Dados e Embaralhamento		X	X		X			[69, 88, 89, 104]
A52	Privacidade por Projeto (*Privacy by Design*) e Privacidade por padrão (*Privacy by Default*)		X		X				[9, 48, 86, 137, 160, 166]
A53	Gestão do Controle de Acesso		X	X	X				[10, 20, 33, 71, 74, 81, 86, 92, 100, 111, 127, 157, 169, 175, 179, 184, 189]
A54	Publicação de Dados (*Statistical Disclosure Control Method*)		X		X				[18, 45, 97, 102, 103, 104, 109, 111, 117, 119, 131, 136, 152, 160]
A55	Uso de Ferramentas de Criptografia Homomórfica			X		X			[61, 114, 145, 172]
A56	Geração de Dados Sintéticos		X	X		X			[27, 44, 131]
A57	Definição da Política de Segurança da Informação			X	X				[36, 47, 61, 77, 86, 99]
A58	Definição dos Requisitos de Segurança			X	X				[133, 147]

Através da leitura inicial dos estudos descritos na revisão da literatura, foram identificadas 58 atividades, distribuídas entre as disciplinas de Governança de Dados (36 atividades), Proteção à

Privacidade de Dados (26 atividades) e Segurança da Informação/Anonimização de Dados (30 atividades). Observe que algumas das atividades pertencem a mais de uma disciplina (portanto a soma entre as disciplinas é superior ao total de atividades classificadas).

3.3 CONSTRUÇÃO DO *FRAMEWORK*

Com base neste levantamento inicial, foram desenvolvidos três questionários, um para cada disciplina, sendo elas: Governança de Dados, Proteção à Privacidade de Dados e Segurança da Informação/Anonimização de Dados. Esses questionários tiveram como objetivo identificar, no escopo da empresa estudo de caso e na opinião dos profissionais de Tecnologias da Informação que lidam com dados, quais das atividades indicadas na revisão de literatura têm eficácia prática no contexto dessa pesquisa: proteção à privacidade em ambientes de dados massivos aplicados à prevenção a fraude. Os questionários estiveram disponíveis entre 14 de junho de 2021 e 30 de junho de 2021, contando com a participação de todas as Superintendências de Tecnologia da Informação da empresa objeto do estudo de caso. Os questionários estão disponibilizados em anexo nos apêndices I.1, I.2 e I.3.

Vinte e oito (28) participantes responderam aos questionários, o que é equivalente a 10% (28/283) do efetivo interno de empregados da área de Tecnologia da Informação da empresa, e abrangeu representantes de todas as subáreas da Diretoria de Tecnologia (Governança, Desenvolvimento e Produção). A Análise do resultado desta aplicação será apresentada na seção 4.2. As questões dos questionários são de múltipla escolha em escala *likert* [5], e na analise os resultados, a fim de construir o *framework*, consideramos as atividades que obtiveram, em média, respostas favoráveis à sua aplicabilidade, seja no prisma de proteção à privacidade, como no prisma de prevenção a fraude. O resultado da aplicação pode ser verificado na Figura 3.1.

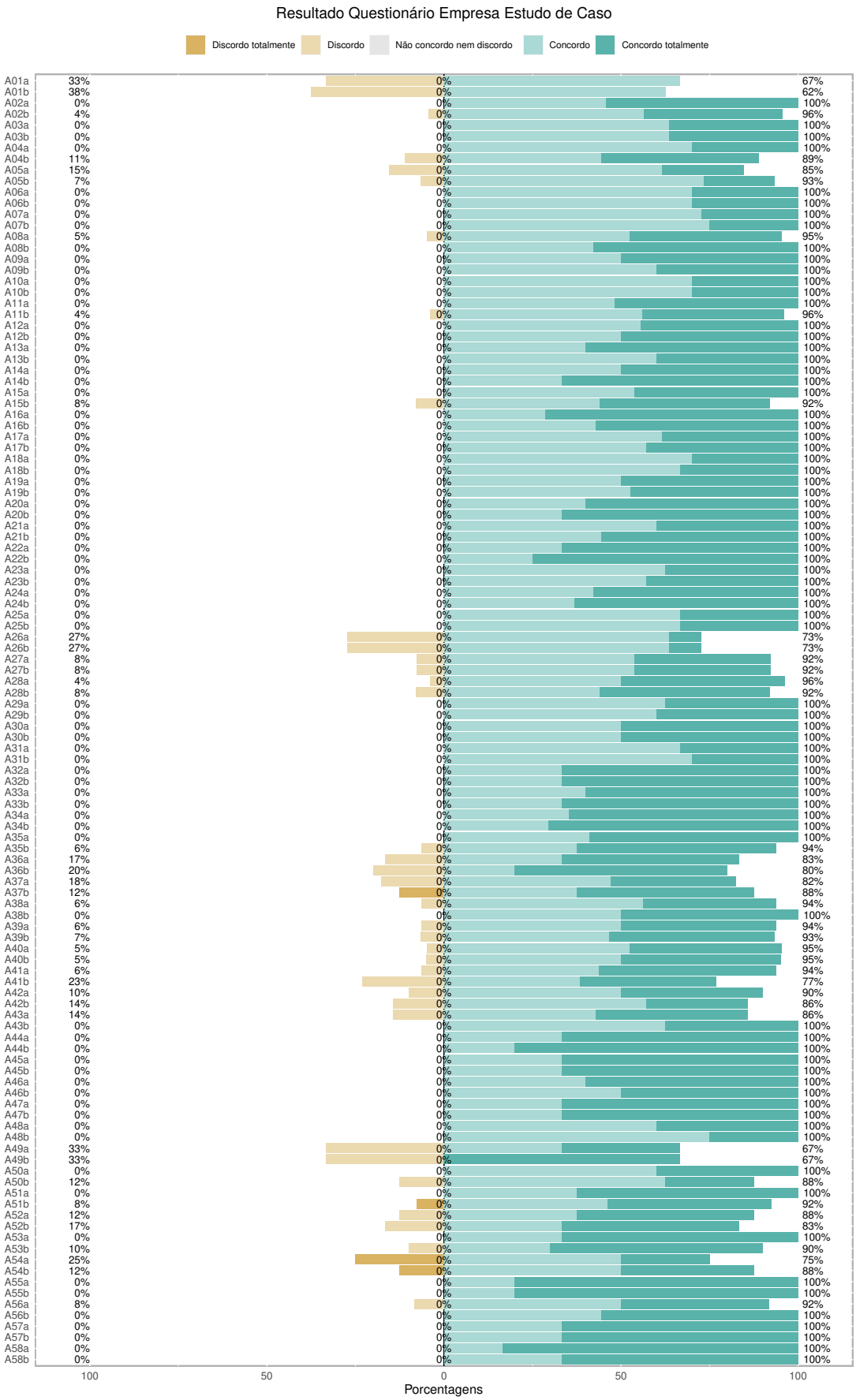

Figura 3.1: Respostas do questionário aplicado na empresa estudo de caso. (Fonte: Autor)

Assim, foi atribuído um valor numérico a cada resposta, de acordo com o apresentado na Tabela 3.3.

Tabela 3.3: Valor numérico e semântico das opções da escala *Likert*. (Fonte: Autor)

Opção *Likert*	Valor Numérico	Valor Semântico	Força Valor Semântico
Concordo Totalmente	5	Favorável	Forte
Concordo	4	Favorável	Fraco
Não concordo nem Discordo	3	Neutro	Neutro
Discordo	2	Desfavorável	Fraco
Discordo Totalmente	1	Desfavorável	Forte

Após essa atribuição numérica, foram consideradas cada atividade descrita na Tabela 3.2 e as médias dos valores numéricos das respostas obtidas na execução do questionário com os profissionais de TI. A maior média entre proteção à privacidade e proteção à fraude foi acatada. Por fim, as atividades que obtiveram médias superiores a 4, ou seja, atividades que foram consideradas de maneira geral aplicáveis à proteção à privacidade ou à prevenção a fraude, compuseram a primeira versão do *framework* proposto, conforme as Figuras 4.9, 4.11, 4.12, 4.13 e 4.14, apresentadas na Seção 4.3 (Framework) do Capítulo 4 (Análise dos Dados e Resultados) específica sobre o *framework*.

A Tabela 3.4 apresenta as atividades, com suas respectivas médias calculadas e a indicação (colorida) das atividades descartadas (as atividades em tons escuros foram descartadas, e as atividades em tons médios foram tratadas com atenção por possuírem médias inferiores à mediana da distribuição).

Tabela 3.4: Médias numéricas obtidas das atividades. (Fonte: Autor)

Atividade	Média Privacidade	Média Fraude	Atividade	Média Privacidade	Média Fraude
A1	**3,272727273**	3,181818182	**A30**	**3,900000000**	**3,900000000**
A2	**4,321428571**	4,071428571	**A31**	4,200000000	**4,300000000**
A3	**4,363636364**	**4,363636364**	**A32**	**4,500000000**	**4,500000000**
A4	**4,181818182**	4,000000000	**A33**	4,600000000	**4,666666667**
A5	3,571428571	**3,761904762**	**A34**	4,647058824	**4,705882353**
A6	**4,181818182**	**4,181818182**	**A35**	4,588235294	4,352941176
A7	**4,272727273**	3,909090909	**A36**	**4,000000000**	3,857142857
A8	4,333333333	**4,428571429**	**A37**	4,000000000	**4,058823529**
A9	**4,363636364**	4,272727273	**A38**	**4,250000000**	4,235294118
A10	**4,181818182**	**4,181818182**	**A39**	**4,235294118**	4,176470588
A11	**4,518518519**	4,222222222	**A40**	**4,037037037**	4,000000000
A12	**4,300000000**	4,200000000	**A41**	**4,294117647**	3,705882353
A13	**4,600000000**	4,400000000	**A42**	**4,200000000**	3,700000000
A14	**4,500000000**	**4,500000000**	**A43**	3,80000000	**4,100000000**
A15	**4,407407407**	4,222222222	**A44**	**4,666666667**	4,500000000
A16	**4,714285714**	4,571428571	**A45**	**4,666666667**	**4,666666667**
A17	4,125000000	**4,176470588**	**A46**	**4,333333333**	4,000000000
A18	**4,300000000**	4,200000000	**A47**	**4,666666667**	**4,666666667**
A19	4,350000000	**4,400000000**	**A48**	**4,166666667**	3,833333333
A20	**4,600000000**	4,500000000	**A49**	3,333333333	**3,500000000**
A21	**4,400000000**	**4,400000000**	**A50**	**4,400000000**	3,800000000
A22	**4,500000000**	4,400000000	**A51**	**4,625000000**	4,000000000
A23	**4,100000000**	4,000000000	**A52**	**4,000000000**	3,700000000
A24	4,500000000	**4,550000000**	**A53**	**4,500000000**	4,400000000
A25	**4,200000000**	**4,200000000**	**A54**	3,400000000	**3,800000000**
A26	**3,352941176**	**3,352941176**	**A55**	**4,500000000**	**4,500000000**
A27	**3,941176471**	**3,941176471**	**A56**	**3,937500000**	3,875000000
A28	**4,333333333**	4,222222222	**A57**	**4,666666667**	**4,666666667**
A29	**4,100000000**	4,050000000	**A58**	**4,833333333**	4,666666667

3.4 VALIDAÇÃO DO *FRAMEWORK*

Para realizar a validação deste *framework*, foi desenvolvido um estudo de caso com o objetivo de mensurar quais benefícios práticos podem ser obtidos ao aplicar as atividades propostas no *framework*. O método usado para mensurar estes benefícios hipotéticos foi: aplicar medidas de qualidade de dados em dois projetos de ingestão de dados, a fim de mensurar a evolução no aspecto de prevenção a fraude, e medidas de privacidade de dados, a fim de validar a proteção à privacidade de dados e aspectos de *compliance* à LGPD.

Utilizando dados produtivos relacionados com os processos de prevenção a fraude operantes na empresa do estudo de caso, foi realizada a ingestão de dados (através de arquivos delimitados gerados a partir da extração dos dados de sistemas transacionais) em duas etapas: uma sem considerar as atividades sugeridas pelo *framework*, e outra etapa que considerou as práticas propostas no *framework*. Nos dois ambientes resultantes dessa carga, foram tomadas medidas de qualidade de dados, que, segundo o exposto no referencial teórico deste trabalho, pode ser considerado um método de quantificar a utilidade do dado para processos de prevenção a fraude. Além disso, foram tomadas as medidas propostas por Domingo-Ferrer et. al. [51], com adaptações (expostas a seguir).

A qualidade de dados interfere diretamente na disponibilidade da informação, contribuindo também de forma direta na propriedade dos sistemas de não repúdio, e esta, por sua vez, influencia (mais uma vez, diretamente) a autenticidade da informação [70]. Esta cadeia de relações diretas configura a qualidade de dados como um possível mensurador da utilidade destes dados para a proteção a fraudes em sistemas.

Essas medidas foram aplicadas com o intuito de quantificar a efetividade do modelo em otimizar a proteção à fraude e da privacidade dos projetos de *Big Data*. Conforme exposto na Seção 2.3 (Governança de Dados) em sua Sub-Seção 2.3.4 (Qualidade de Dados) do Capítulo 2 (Referencial Teórico), a literatura diverge em quais as métricas ideais para quantificar a qualidade de dados de determinada base. Porém, os modelos apresentados convergem para uma delimitação de dimensões. Neste sentido, a fim de definir uma possível estimativa de qualidade de dados com o objetivo de validar a aplicação do *framework*, propomos a união das dimensões descritas por Rego [150] e Barbieri [15], de acordo com:

1. **Completude**: Avaliada através da quantidade de campos nulos do atributo.
2. **Unicidade**: Avaliada através da quantidade de repetições de valores do atributo.
3. **Razoabilidade**: Avaliada através da quantidade de valores com erro lógico.
4. **Integridade**: Avaliada através da quantidade de valores que ferem a integridade referencial lógica entre as duas tabelas.
5. **Validade**: Avaliada pela quantidade de valores que não correspondem ao domínio do atributo.
6. **Performance**: avaliada pelos tempos de inserção e consulta ao dado.

As métricas apresentadas definem o modelo inicial de validação composto de métricas de qualidade de dados. Essas métricas são o suporte para validação do modelo no quesito prevenção à fraude, uma vez que dados de qualidade embasam de maneira mais efetiva qualquer análise de fraude. Quanto à validação da aplicação do *framework* no quesito privacidade, adotamos o modelo proposto por Domingo-Ferrer [51], conforme a equação:

$$CM1(X, Y) = 1 - \rho_1^2 \tag{3.1}$$

onde, CM1 é a métrica de confidencialidade, X é o conjunto de atributos identificadores (*PII*) presentes na base, Y é o conjunto de atributos não identificadores e ρ_1 seja a maior correlação canônica entre o conjunto de atributos identificadores X e o conjunto de atributos não identificadores Y. Com relação à proposição de Domingo-Ferrer et. al. [51], adaptamos o cálculo para um coeficiente de contingência de Pearson, uma vez que avaliamos que os dados pessoais representam uma variável qualitativa (descritiva) que deve ser tratada estatisticamente como tal. Portanto, para correlações de variáveis qualitativas teremos a fórmula 3.2 e 3.3 de correlação:

$$C^o = \sqrt{\frac{(k\chi^2)}{[(k-1)(\chi^2+n)]}} \tag{3.2}$$

onde, C^o é o coeficiente de correlação das variáveis, k é a menor quantidade (contagem) entre os elementos de X e Y, n é o tamanho da população e χ^2 é dado pela equação 3.3:

$$\chi^2 = \sum \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right] \tag{3.3}$$

onde, O_{ij} é o valor de frequência obtido de cada combinação de elementos de X e Y e E_{ij} é o valor esperado.

Além deste cálculo, também utilizamos uma pequena lista de checagem, avaliando os quesitos mínimos para tratamento de dados pessoais com o objetivo de conformidade com a LGPD. A Tabela 3.5 apresenta o modelo desta lista de checagem.

Tabela 3.5: Lista de checagem dos critérios para tratamento de dados para prevenção à fraude, em conformidade com a LGPD. (Fonte: Autor)

	Questão	**X**	**V**
CK1	O propósito do tratamento do dado foi concebido?		
CK2	O propósito do tratamento do dado foi documentado?		
CK3	O propósito do tratamento do dado está acessível?		
CK4	A linhagem (origem e destino) do dado é documentada?		
CK5	O dado é classificado (dado pessoal, dado sensível)?		
CK6	A classificação do dado está acessível?		

4 ANÁLISE DOS DADOS E RESULTADOS

Neste capítulo apresentamos o detalhamento do *framework* proposto, a análise dos dados obtidos com as múltiplas aplicações dos questionários disponíveis nos anexos I.1, I.2 e I.3, bem como o resultado da aplicação estudo de caso.

4.1 LEVANTAMENTO DAS ATIVIDADES

Conforme apresentado no Capítulo 3 (Metodologia), as principais atividades das disciplinas de Governança de Dados, Privacidade de Dados e Segurança da Informação/Anonimização de Dados foram identificadas a partir de uma revisão da literatura da área. As atividades identificadas podem ser consultadas na Tabela 3.2.

4.2 QUESTIONÁRIOS

Inicialmente os questionários apresentados nos anexos I.1, I.2 e I.3 foram aplicados na empresa estudo de caso entre o período de 14 a 30 de junho de 2021, sendo respondido por 28 participantes. Os questionários tiveram por objetivo coletar a opinião dos profissionais de Tecnologia da Informação (TI) que lidam com dados na organização acerca da aplicabilidade das atividades levantadas em uma revisão de literatura focada na proteção à privacidade de dados e no contexto de prevenção a fraudes.

Em momento posterior, os mesmos questionários foram aplicados em outras instituições, como forma comparativa das respostas obtidas na empresa alvo do estudo de caso. Esta segunda aplicação dos questionários obteve 17 respostas, sendo disponibilizada entre os dias 25 de julho e 7 de agosto de 2021. A aplicação dos questionários para as demais instituições incluiu empresas de pequeno, médio e grande porte, dos setores privado, público e misto, de vários segmentos (financeiro, educacional, governo, prestação de serviços entre outros). De forma semelhante à primeira aplicação, diferentes profissionais das disciplinas de Governança, Privacidade e Segurança participaram como respondentes.

Os questionários foram construídos em duas partes: um primeiro conjunto de questões referentes à atividade da organização e do respondente e uma segunda parte expondo para apreciação as atividades identificadas na literatura, conforme apresentado no capítulo 3 (Metodologia). O primeiro conjunto de questões, além de delimitar um comparativo demográfico para avaliação deste trabalho, teve por objetivo indicar se os profissionais respondentes da pesquisa se avaliam como aptos a tratar dados tendo em vista a vigência da LGPD [42] e os demais requisitos de privacidade para tratamento de dados. A Figura 4.1 apresenta as respostas às primeiras questões da empresa estudo de caso. As questões de 1 a 3 (Q1 a Q3) não aparecem na avaliação porque dizem respeito

ao respondente do questionário e à instituição representada.

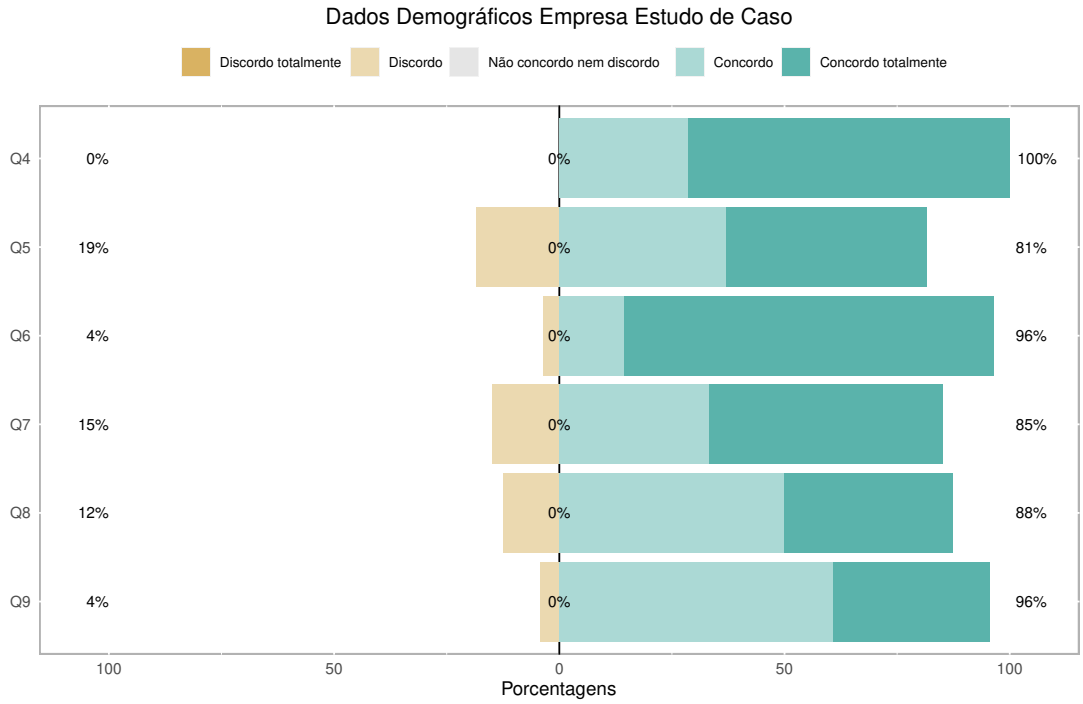

Figura 4.1: Respostas do questionário aplicado na empresa estudo de caso (demográfico). (Fonte: Autor)

De acordo com os resultados obtidos, é possível perceber que, de maneira geral, os profissionais da empresa estudo de caso avaliam que suas atividades envolvem, de maneira direta ou indireta, análise e prevenção a fraudes (Q4 e Q5 da Figura 4.1) e manipulação de dados pessoais (Q6 e Q7 da Figura 4.1), além de considerarem que sua empresa vem adotando medidas para o *compliance* à LGPD (Q8 da Figura 4.1) e alegarem ter algum conhecimento da LGPD (Q9 da Figura 4.1).

Outro aspecto interessante é que, apesar de alguns respondentes alegarem não possuir conhecimento da LGPD (Q9 da Figura 4.1) em sua totalidade, há opinião consensual de que as atividades da empresa envolvem o tratamento direto de dados pessoais (Q6 da Figura 4.1). Este dado revela que ainda é necessário a empresa disseminar o conhecimento a respeito da regulação. A Figura 4.2 apresenta as respostas das primeiras questões aplicadas nas demais instituições.

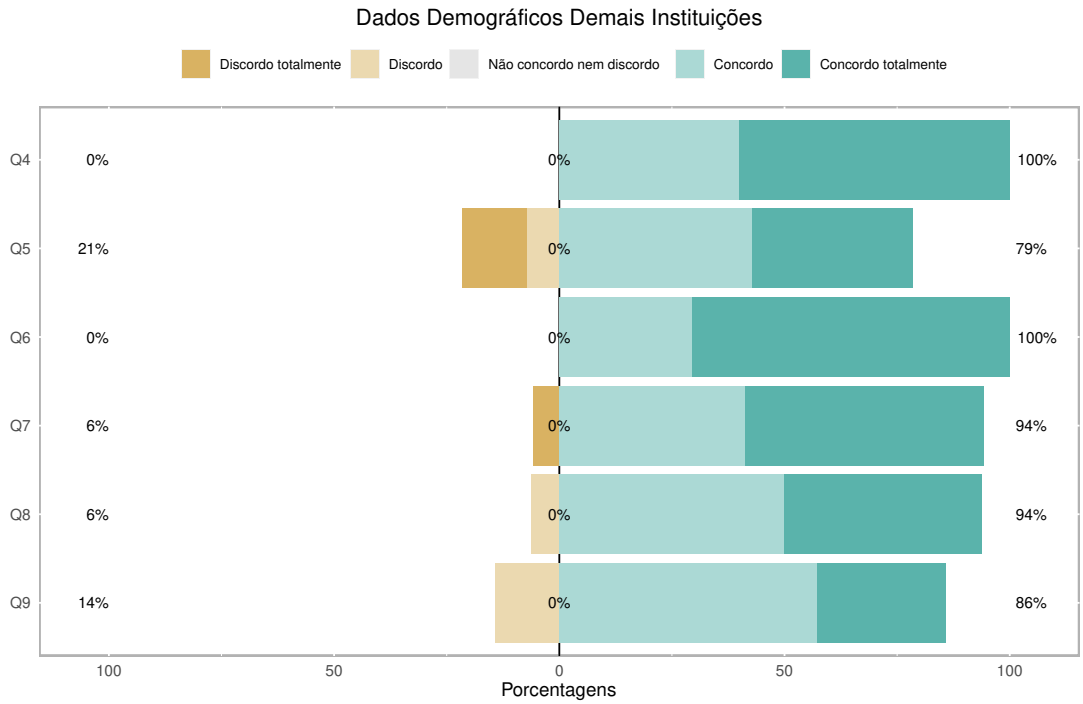

Figura 4.2: Respostas do questionário aplicado nas demais instituições (demográfico). (Fonte: Autor)

Em linhas gerais, os profissionais das demais instituições também entendem que as atividades das suas empresas (Q4 da Figura 4.2) e as suas próprias atividades individuais (Q5 da Figura 4.2) envolvem de alguma forma a prevenção a fraude, demonstrando que o assunto extrapola o setor financeiro. Além disso, foi unânime a opinião de que as diversas empresas tratam, em algum nível, com dados pessoais (Q6 da Figura 4.2), e foi bastante expressiva (94%) a opinião de profissionais que consideravam suas atividades como inseridas no tratamento de dados pessoais (Q7 da Figura 4.2). De maneira geral, as instituições também têm desenvolvido ações para o *compliance* à LGPD (Q8 da Figura 4.2) e ações de conscientização da regulação (Q9 da Figura 4.2).

A crítica feita à empresa alvo do estudo de caso também se aplica às demais instituições, pois houve respondentes que avaliaram não conhecer a legislação (LGPD), mesmo tratando em algum nível de dados pessoais. Pode-se deduzir que esta tem sido uma dificuldade enfrentada pela maioria das organizações. A seguir apresenta-se o segundo conjunto de questões, separadas pelas disciplinas referenciadas.

Nesta segunda etapa, as atividades descritas na Tabela 3.2 foram apreciadas de acordo com sua aplicabilidade ao contexto de proteção da privacidade de dados nos moldes da LGPD (sufixo a) e no contexto de prevenção a fraude (sufixo b). Deixaremos de referenciar as questões diretamente (Q1, Q2, Q3) pois as atividades foram arguidas dos entrevistados em diferentes questões, de acordo com o questionário preenchido. Exemplificando a forma de referência às atividades, a questão marcada como A21a investigou a aplicabilidade da atividade de "Gestão de Dados" para a proteção da privacidade dos dados, enquanto a questão A21b verificava essa mesma atividade

no contexto de prevenção a fraude.

4.2.1 Governança de Dados

As questões de governança de dados foram aplicadas aos profissionais que consideraram que sua participação na empresa envolve atividades como definição de políticas, procedimentos, manuais, normas e guias, além das áreas de garantia da qualidade, gestão de contratos, arquitetura (referencial) de software e análise de dados (*BI*). A Figura 4.3 apresenta a percepção desses profissionais com relação às atividades levantadas no grupo da disciplina governança.

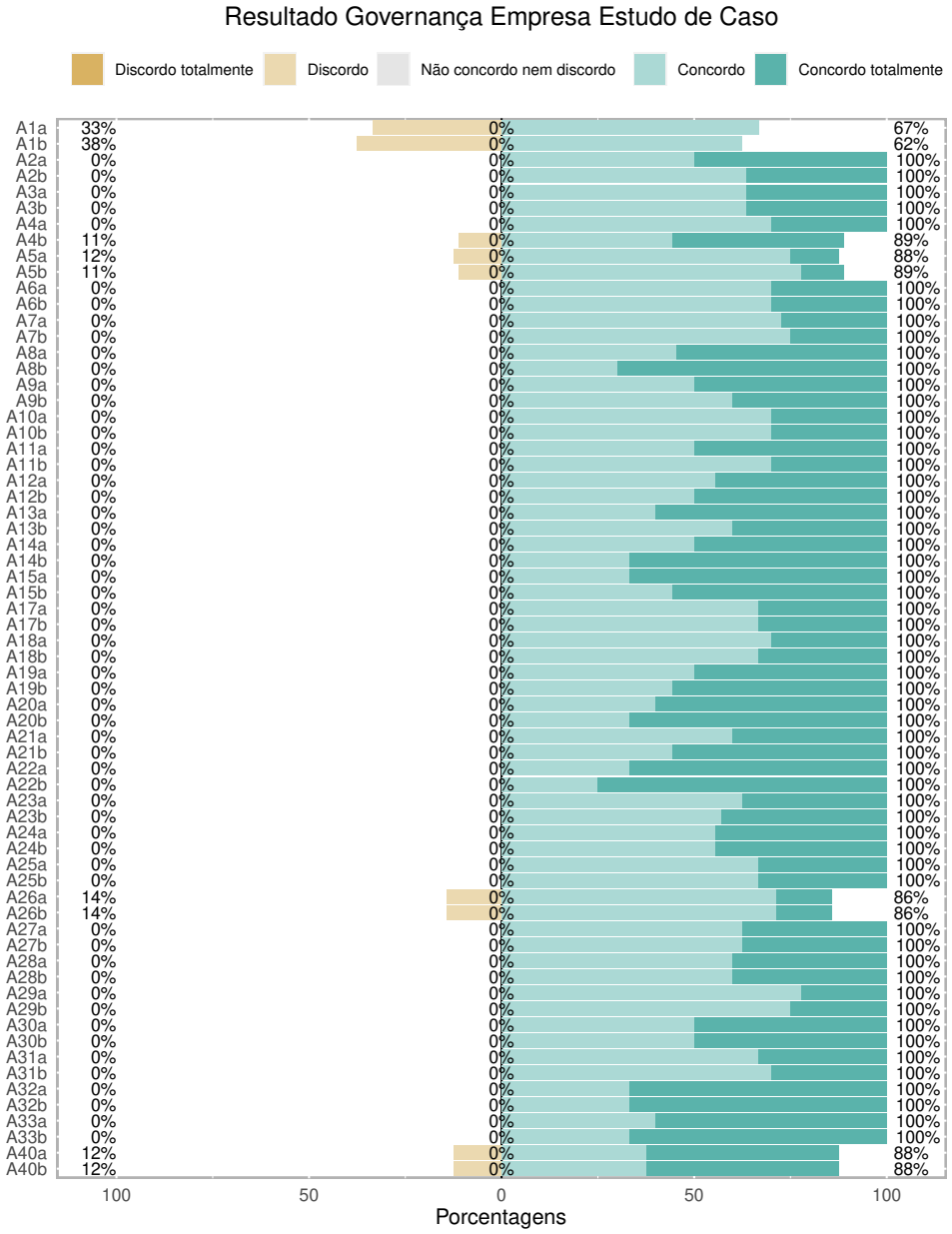

Figura 4.3: Respostas das questões de Governança da empresa estudo de caso. (Fonte: Autor)

Podemos observar que, dadas algumas avaliações que consideraram as atividades não aplicáveis

ao contexto de proteção da privacidade de dados e prevenção a fraude, houve atividades exclusivas dessa disciplina que foram descartadas do *framework*. É o caso da atividade A1 ("Arquitetura de governança (des)centralizada do controlador") (vide Tabela 3.2). Uma possível interpretação dessa avaliação é de que, na opinião destes profissionais, a escolha da arquitetura de governança (centralizada no controlador ou descentralizada e diluída entre as várias instituições colaborando com a operação do dado) poderia estar representada em outras atividades de cunho mais amplo, como a A1 ([Definição da] "Política de Dados e Privacidade").

Outra atividade que também foi descartada do modelo foi a A5, a saber "Análise do Custo-benefício (ROI) do Dado". Podemos entender então que, na opinião dos profissionais entrevistados, as questões de proteção à privacidade e prevenção a fraude podem ganhar prioridade se comparadas a análises de custo-benefício baseadas apenas em retorno financeiro. Uma possível explicação para este resultado reside no fato de que a empresa estudo de caso opera na modalidade de economia mista, e tem em vários de seus aspectos fundadores a questão social imprimida. A própria empresa vem operando serviços que, se analisados no aspecto apenas econômico, representam prejuízo para a instituição, tais como operações financeiras de pagamento de benefícios sociais do governo. O resultado dessa cultura de bem estar social é, portanto, a redução da prioridade de análise de custo-benefício, que talvez não seja observada em instituições inseridas em outros contextos.

A atividade A26 ("Resolução de conflitos") também foi descartada. Apesar de representar uma necessidade de qualquer contexto onde há conflito de interesses (como observado na dicotomia privacidade *versus* personalização), as práticas pertencentes a esta atividade também podem ser distribuídas nas demais.

A atividade A40 ("Definição do Contexto da Organização"), apesar de ter sido considerada por 12% dos entrevistados como não aplicável a nenhum dos contextos, também foi avaliada pelos entrevistados da disciplina Privacidade de Dados 4.2.2 e foi mantida no modelo. Algumas atividades foram fortemente recomendadas pelos profissionais entrevistados, dentre elas: A8 ("Monitoração de Processos"), A14 ("Gestão de Metadados"), A20 ("Organização da Gestão e Governança de Dados"), A22 ("Arquitetura de Dados") e A32 ("Gestão de Dados Mestres"). Essas recomendações podem ser baseadas na tendência apontada na LGPD e discutida ao longo deste documento de que a implantação e a melhoria de processos de Governança de Dados auxiliam na proteção da privacidade. A maioria das atividades destacadas pelos profissionais são atividades puramente da Governança e Gestão de Dados, o que reforça o argumento acima apresentado.

De acordo com a Figura 4.4, podemos comparar o resultado obtido com a empresa estudo de caso e com as demais instituições que responderam ao questionário (na segunda aplicação).

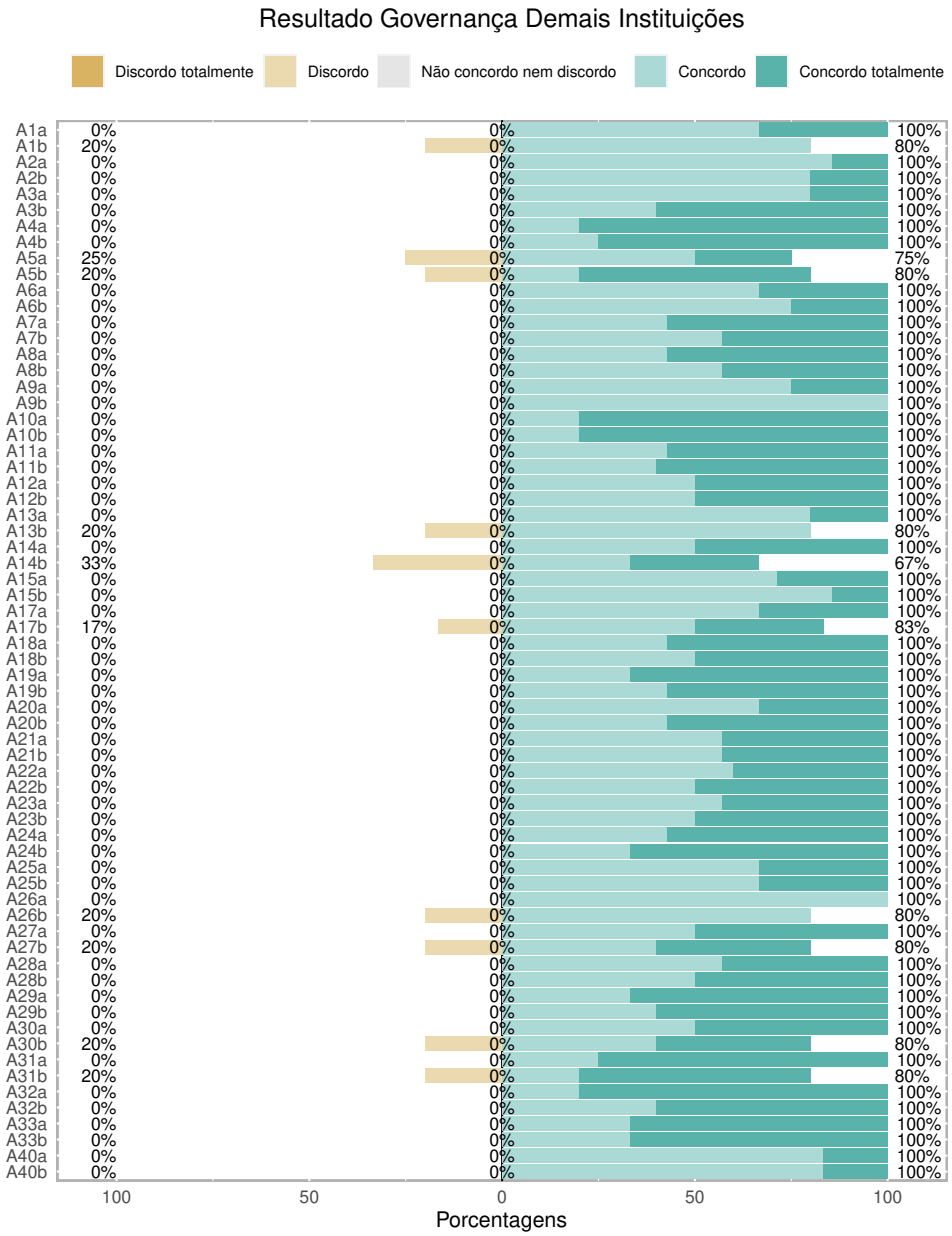

Figura 4.4: Respostas das questões de Governança das demais instituições. (Fonte: Autor)

Pelo comparativo, podemos perceber que as atividades A1 ([Definição da] “Política de Dados e Privacidade”), A5 (“Análise do Custo-benefício (ROI) do Dado”) e A26 (“Resolução de conflitos”) também foram consideradas pelos profissionais fora da empresa estudo de caso como não aplicáveis aos contextos propostos. Entretanto, de forma distinta dos profissionais da empresa estudo de caso, os profissionais das demais instituições elegeram a atividade A40 (“Definição do Contexto da Organização”) como aplicável, o que é consoante às definições presentes tanto na regulação (LGPD, ISOs) como na literatura da área. Além disso, houve divergências de opinião nas atividades A13 (“Planejamento do Ciclo de Vida do Dado”), A14 (“Gestão dos Metadados”), A17 (“Princípios de Dados e Valor de Dados”), A27 (“Modelagem de Dados”), A30 (“Harmonização de Dados”) e A31 (“Integração de Dados”).

Para os profissionais das demais instituições, as atividades com maior impacto nos processos de proteção à privacidade de dados e prevenção a fraudes são: A4 (“Análise de Utilidade do Dado”), A5 (“Análise do Custo-benefício (ROI) do Dado”), A10 (“Processo de Qualidade de Dados”), A31 (“Integração de Dados”) e A32 (“Gestão de Dados Mestres”).

4.2.2 Privacidade de Dados

As questões da disciplina de privacidade de dados foram submetidas no contexto da empresa estudo de caso aos profissionais da tecnologia da informação que lidam diretamente com o desenvolvimento de aplicações e programas, ou seja, àqueles que lidam diretamente com os dados através da construção das soluções de TI. A Figura 4.5 apresenta o resultado da aplicação deste questionário.

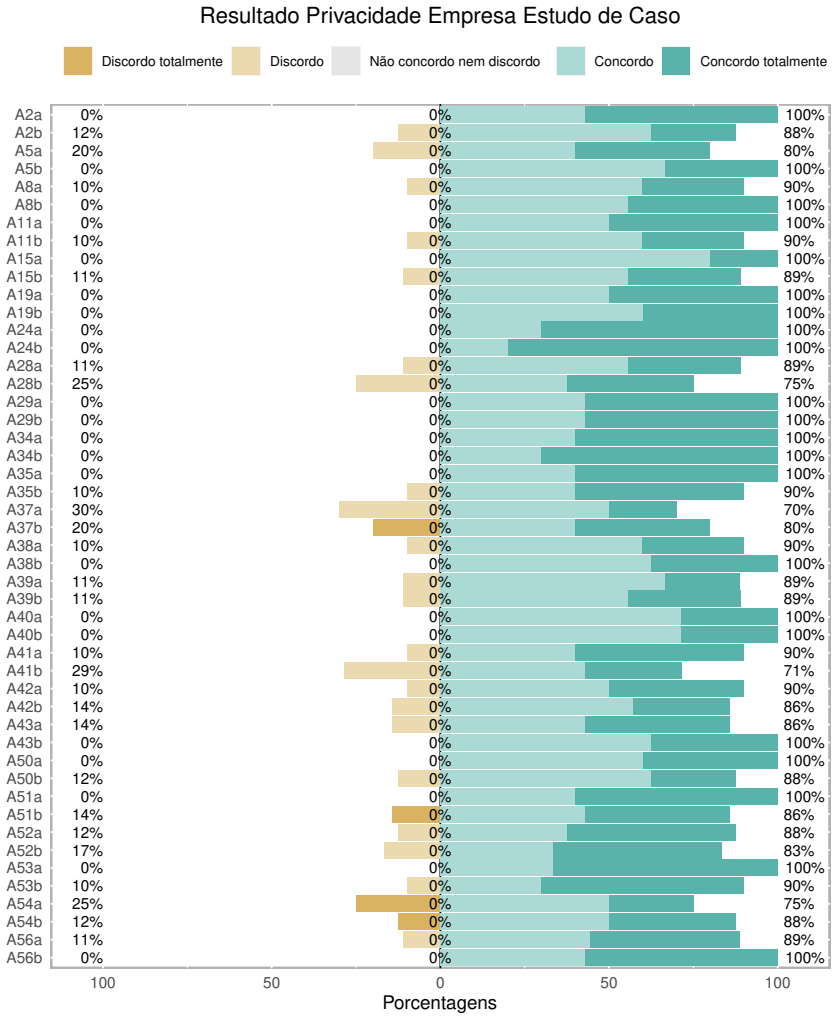

Figura 4.5: Respostas das questões de Privacidade da empresa estudo de caso. (Fonte: Autor)

As atividades que foram descartadas a partir da avaliação deste questionário foram as atividades A54 (“Publicação de Dados (*Statistical Disclosure Control Method*)”) e A56 (“Geração de Dados Sintéticos”). Uma possível interpretação da opinião dos profissionais seria que tais atividades

podem estar compreendidas nas demais, especialmente na atividade A39 (“Política de Desenvolvimento e Teste Seguro”), que poderia trazer diretrizes sobre publicação de dados e uso de dados sintéticos.

Algumas das atividades deste grupo dividiram opiniões, tendo tanto avaliações expressivas de discordância da aplicabilidade como de forte recomendação. Foram elas: A37 (“Auditoria”), A51 (“Uso de Ferramentas de Mascaramento de Dados e Embaralhamento”) e A54 (“Publicação de Dados (*Statistical Disclosure Control Method*)”). Como o critério usado para o descarte das atividades foi o da média aritmética dos valores das respostas (conforme as tabelas 3.3 e 3.4), as atividades citadas foram preservadas no modelo.

Para os profissionais da empresa estudo de caso, é possível destacar as seguintes atividades como as de maior impacto para os processos de proteção à privacidade e prevenção a fraude: A24 (“Capacitação de Recursos Humanos”), A34 (“Avaliação de Riscos de Segurança da Informação”), A51 (“Uso de Ferramentas de Mascaramento de Dados e Embaralhamento”) e A53 (“Gestão do Controle de Acesso”). De acordo com a Figura 4.6, podemos comparar o resultado obtido com a empresa estudo de caso e com as demais instituições que responderam ao questionário (na segunda aplicação).

Quanto à disciplina de privacidade, a opinião dos profissionais da empresa estudo de caso divergiu com maior intensidade da opinião dos profissionais das demais instituições se comparada a disciplina de Governança. As atividades A54 e A56, descartadas do modelo, foram avaliadas pelos profissionais das demais instituições como aplicáveis por unanimidade. Outras divergências foram as atividades A11 (“Política de Dados e Privacidade”), A15 (“Controle da Infraestrutura”), A35 (“Planejamento e Uso de Métodos Criptográficos”), A38 (“Controle da Transferência de Dados”), A39 (“Política de Desenvolvimento e Teste Seguro”), A40 (“Definição do Contexto da Organização”), A41 (“Coleta do Consentimento do Uso dos Dados”) e A43 (“Mapeamento de Fontes Externas de Dados”), todas avaliadas como não aplicáveis.

Houve divergências também nas atividades destacadas, que para os profissionais das demais instituições seriam: A8 (“Monitoração de Processos”), A11 (“Política de Dados e Privacidade”), A35 (“Planejamento e Uso de Métodos Criptográficos”) e A37 (“Auditoria”).

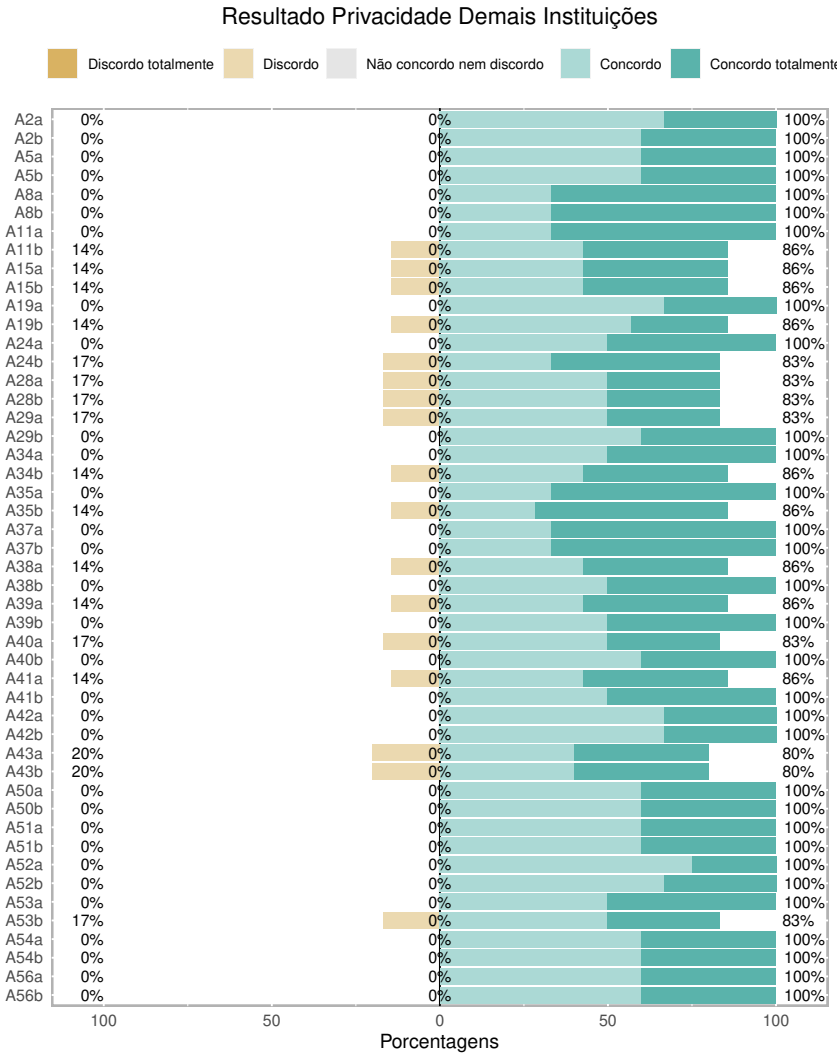

Figura 4.6: Respostas das questões de Privacidade das demais instituições. (Fonte: Autor)

4.2.3 Segurança da Informação / Anonimização

O conjunto de questões das disciplinas de Segurança da Informação e Anonimização de Dados, quando apreciada pelos profissionais da tecnologia da informação da empresa estudo de caso, contou com respondentes ligados ao suporte da operação de sistemas e infraestrutura, ao monitoramento de ambientes, suporte ao *deployment* e produção. Tais profissionais lidam diariamente com atendimento a incidentes que requerem aplicações de boas práticas de segurança e anonimização. Além disso, estes profissionais também podem ter, pela natureza de seu trabalho, contato com os dados pessoais e sensíveis. A Figura 4.7 representa a consolidação das opiniões destes profissionais quanto às atividades sugeridas nos contextos de prevenção a fraudes e proteção da privacidade.

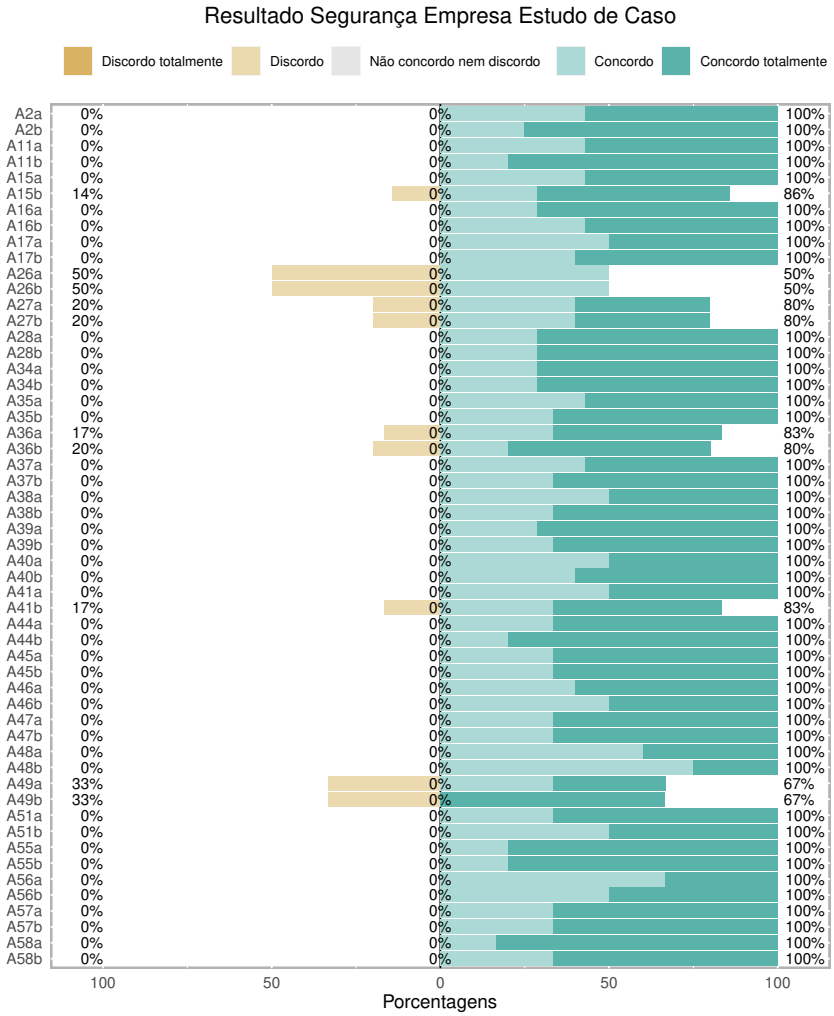

Figura 4.7: Respostas das questões de Segurança/Anonimização da empresa estudo de caso. (Fonte: Autor)

As atividades descartadas do modelo a partir das opiniões impressas na Figura 4.7 foram: A26 (“Resolução de conflitos”), A27 (“Modelagem de Dados”) e A49 (“Uso de Ferramentas de Inteligência Artificial Aplicadas e Aprendizado de Máquina à Segurança (tais como Modelo Oculto de *Markov*, Computação Bayesiana Aproximada e Redes Bayesianas, Aprendizado Federado, Redes Neurais, *Blockchain*, Mineração de Dados, entre outras)”).

Uma possível interpretação para a não aplicabilidade da atividade A49 aos contextos definidos pode residir na quantidade de ferramentas nem sempre usuais aos respondentes, o que pode ter gerado um falso sentimento de impropriedade. Além disso, tais ferramentas poderiam ser sugeridas nas demais atividades, tais como A39 (“Política de Desenvolvimento e Teste Seguro”) ou A47 (“Desenvolvimento do Serviço de Proteção a Dados”). Relativo a não aplicabilidade das atividades A26 e A27, um estudo mais direcionado seria necessário para investigar o motivo dessa opinião, pois ambas as atividades têm importância documentada na literatura [20, 23, 36, 58, 59, 68, 95, 115, 138, 139, 146, 147]. No exemplo da atividade A27 de Modelos de dados, várias das demais atividades detêm, no artefato modelos de dados, insumos para seu desenvolvimento.

Um aspecto interessante observado na opinião dos profissionais relacionados a esta disciplina na empresa estudo de caso foi que houve mais recomendações intensas (resposta “Concordo plenamente”, na escala escolhida) proporcionalmente do que as respostas das demais disciplinas. Deste modo, podemos citar como destaques, uma vez que fortemente recomendadas pelos respondentes, as atividades A11 (“Política de Dados e Privacidade”), A44 (“Definição dos Objetivos da Segurança da Informação”), A55 (“Uso de Ferramentas de Criptografia Homomórfica”) e A58 (“Definição dos Requisitos de Segurança”). De acordo com a Figura 4.8, podemos comparar o resultado obtido com a empresa estudo de caso e com as demais instituições que responderam ao questionário (na segunda aplicação).

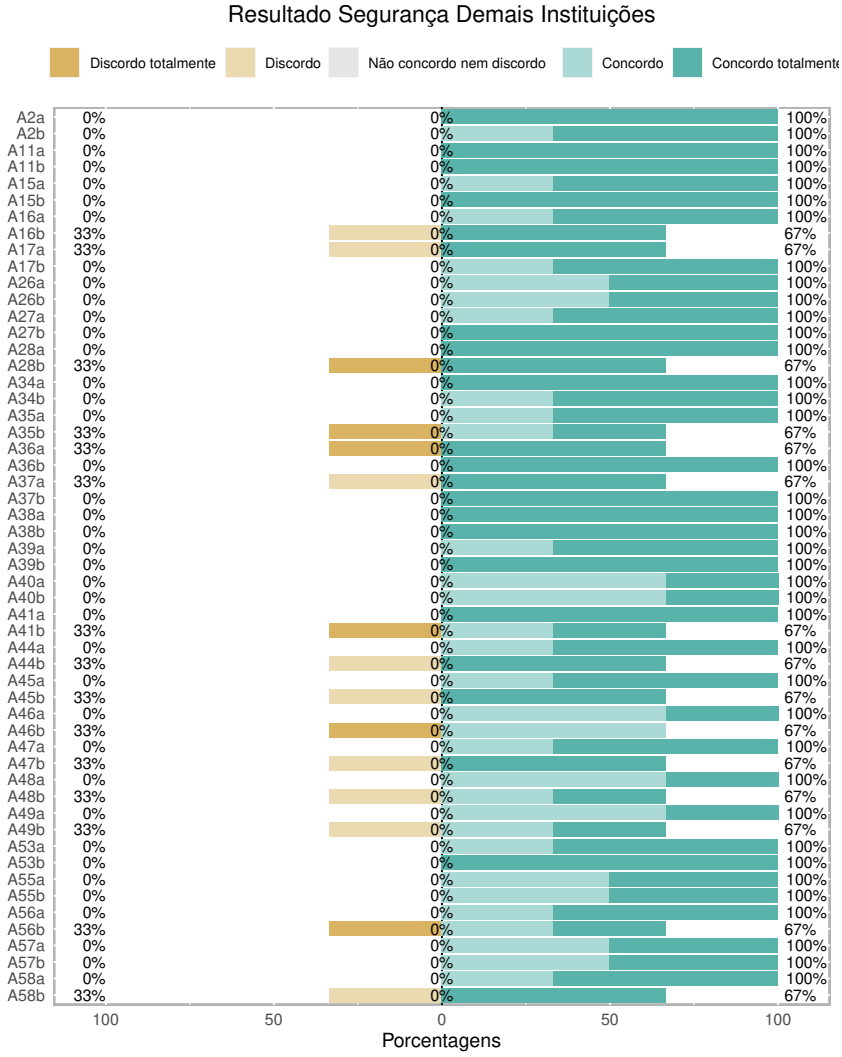

Figura 4.8: Respostas das questões de Segurança/Anonimização das demais instituições. (Fonte: Autor)

Quanto à segurança, houve divergências nas atividades A16 (“Arquitetura de Segurança da Informação”), A28 (“Classificação de Dados”), A35 (“Planejamento e Uso de Métodos Criptográficos”), A36 (“Definição de Estratégia de Cópias de Segurança (*Backup*)”), A41 (“Coleta do Consentimento do Uso dos Dados”), A44 (“Definição dos Objetivos da Segurança da Informação”), A45 (“Definição dos Serviços de Autenticação”), A46 (“Deidentificação (*Unlinkability*)”), A47

("Desenvolvimento do Serviço de Proteção a Dados"), A48 ("Uso de Ferramentas de Privacidade Diferencial e K-Anonimização"), A56 ("Geração de Dados Sintéticos") e A58 ("Definição dos Requisitos de Segurança"). Um estudo detalhado seria necessário para investigar a raiz dessas divergências.

A consolidação da tabela 3.2 representa a resposta à primeira etapa de questões problema desta pesquisa, a saber RQ.1, RQ.2 e RQ.3. Portanto, quanto às boas práticas relacionadas à Governança de Dados no contexto de *Big Data* para identificação, classificação e tratamento de dados enquadrados no item g do inciso II do 11º artigo da LGPD (RQ.1) temos as seguintes atividades (relacionadas pelos seus índices): **A2, A3, A4, A6, A7, A8, A9, A10, A11, A12, A13, A14, A15, A17, A18, A19, A20, A21, A22, A23, A24, A25, A28, A29, A31, A32, A33, A39 e A40**, de acordo com a tabela 4.1. Desta lista, os profissionais recomendaram fortemente as atividades **A8, A14, A20, A22 e A32**.

Tabela 4.1: Atividades do modelo - Governança. (Fonte: Autor)

	Atividade	**PDCA**				**Referências**
		Planejamento	**Desenvolvimento**	**Controle**	**Ação**	
A2	Definição do Acesso e Propriedade do Dado	X				[31, 71, 86, 89, 93, 105, 119, 128, 157, 179]
A3	Análise do Ambiente Regulatório	X				[23, 105, 153, 166]
A4	Análise de Utilidade do Dado	X	X			[19, 23, 45, 49, 51, 56, 57, 65, 67, 68, 97, 98, 105, 119, 131, 154, 160, 166]
A6	Aplicação de Casos de Uso de Dados	X				[96, 105]
A7	Análise de Conformidade Legal			X		[23, 97, 99, 102, 105, 106, 163, 165, 166, 170]
A8	Monitoração de Processos				X	[100, 105, 153, 186]
A9	Desenho da Linhagem/-Fluxo/Mapeamento de Dados	X	X			[9, 72, 102, 105, 110, 115, 118, 139, 144, 157, 164, 163]
A10	Processo de Qualidade de Dados			X		[3, 4, 13, 52, 58, 78, 89, 90, 105, 119, 129, 133, 144, 145, 153, 160, 165, 179]
A11	Política de Dados e Privacidade	X				[1, 23, 33, 66, 68, 86, 93, 97, 100, 108, 128, 153, 159, 189]
A12	Desenho da Estratégia de Dados	X				[23, 153, 160]
A13	Planejamento do Ciclo de Vida do Dado	X				[8, 23, 95, 108, 110, 140, 148, 153]
A14	Gestão dos Metadados	X				[17, 23, 47, 73, 114, 139, 153, 169, 171]
A15	Controle da Infraestrutura		X			[23, 25, 75, 153]
A17	Princípios de Dados e Valor de Dados	X		X		[9, 17, 68, 95, 130, 146, 147, 153, 154, 160, 165, 187]

Continua na próxima página

Tabela 4.1 – continuando da última página

	Atividade	PDCA				Referências
		Planejamento	Desenvolvimento	Controle	Ação	
A18	Compreender Necessidades Corporativas Estratégicas de Dados	X				[23, 160, 165]
A19	Definição de Papéis e Responsabilidades	X				[23, 86]
A20	Organização da Gestão e Governança de Dados	X				[9, 23]
A21	Gestão de Dados	X				[23, 106, 112, 128, 137, 160]
A22	Arquitetura de Dados	X				[23, 102, 111, 148]
A23	Projetos e Serviços de Gestão de Dados	X		X		[23, 139]
A24	Capacitação de Recursos Humanos	X		X		[23, 86]
A25	Coordenar as Atividades de Governança de Dados			X		[23]
A28	Classificação de Dados		X	X		[17, 58, 65, 86, 159, 160, 163, 175, 187]
A29	Análise de Requisitos e Impacto			X		[44, 72, 144, 160, 190]
A31	Integração de Dados		X	X		[1, 2, 52, 97, 102, 106, 138, 144, 167, 169, 183, 186]
A32	Gestão de Dados Mestres		X			[23, 144]
A33	Curadoria de Dados	X	X			[23, 144]
A39	Política de Desenvolvimento e Teste Seguro	X				[86, 96, 132, 147]
A40	Definição do Contexto da Organização	X				[23, 86, 137]

Relacionado às boas práticas relacionadas à Privacidade de Dados no contexto de *Big Data* para identificação, classificação e tratamento de dados enquadrados no item g do inciso II do 11º artigo da LGPD (RQ.2), foram levantadas as atividades (listadas pelos seus índices): **A2, A4, A8, A11, A15, A19, A24, A28, A29, A34, A35, A37, A38, A39, A40, A41, A42, A43, A50, A51, A52 e A53**, de acordo com a tabela 4.2. Dentre elas, as mais recomendadas pelos profissionais foram: **A24, A34, A51 e A53**.

Tabela 4.2: Atividades do modelo - Privacidade. (Fonte: Autor)

	Atividade	PDCA				Referências
		Planejamento	Desenvolvimento	Controle	Ação	
A2	Definição do Acesso e Propriedade do Dado	X				[31, 71, 86, 89, 93, 105, 119, 128, 157, 179]
A4	Análise de Utilidade do Dado	X	X			[19, 23, 45, 49, 51, 56, 57, 65, 67, 68, 97, 98, 105, 119, 131, 154, 160, 166]

Continua na próxima página

Tabela 4.2 – continuando da última página

	Atividade	PDCA				Referências
		Planejamento	Desenvolvimento	Controle	Ação	
A8	Monitoração de Processos				X	[100, 105, 153, 186]
A11	Política de Dados e Privacidade	X				[1, 23, 33, 66, 68, 86, 93, 97, 100, 108, 128, 153, 159, 189]
A15	Controle da Infraestrutura		X			[23, 25, 75, 153]
A19	Definição de Papéis e Responsabilidades	X				[23, 86]
A24	Capacitação de Recursos Humanos	X		X		[23, 86]
A28	Classificação de Dados		X	X		[17, 58, 65, 86, 159, 160, 163, 175, 187]
A29	Análise de Requisitos e Impacto			X		[44, 72, 144, 160, 190]
A34	Avaliação de Riscos de Segurança da Informação			X	X	[2, 52, 61, 86, 107, 164, 166, 170, 189]
A35	Planejamento e Uso de Métodos Criptográficos	X	X			[3, 10, 13, 32, 33, 61, 75, 78, 86, 88, 89, 90, 91, 101, 106, 114, 145, 157, 158, 160, 172, 177, 183]
A37	Auditoria			X		[66, 86, 146, 166, 175]
A38	Controle da Transferência de Dados		X			[32, 86, 89, 179]
A39	Política de Desenvolvimento e Teste Seguro	X				[86, 96, 132, 147]
A40	Definição do Contexto da Organização	X				[23, 86, 137]
A41	Coleta do Consentimento do Uso dos Dados	X	X			[31, 68, 86, 105, 166]
A42	Definição do Propósito do Tratamento	X	X			[6, 25, 86, 111, 128, 160, 166]
A43	Mapeamento de Fontes Externas de Dados	X	X			[25, 44]
A50	Definição de Métricas de Riscos à Privacidade			X		[8, 26, 31, 44, 72, 87, 97, 103, 119, 145, 160, 161, 163, 190]
A51	Uso de Ferramentas de Mascaramento de Dados e Embaralhamento		X			[69, 88, 89, 104]
A52	Privacidade por Projeto (*Privacy by Design*) e Privacidade por padrão (*Privacy by Default*)	X				[9, 48, 86, 137, 160, 166]
A53	Gestão do Controle de Acesso	X				[10, 20, 33, 71, 74, 81, 86, 92, 100, 111, 127, 157, 169, 175, 179, 184, 189]

Por fim, referente às boas práticas relacionadas à Segurança de Informação no contexto de *Big Data* para identificação, classificação e tratamento de dados enquadrados no item g do inciso II

do 11º artigo da LGPD (RQ.3), temos as atividades (listadas pelos seus índices): **A2, A11, A15, A16, A17, A24, A28, A34, A35, A36, A37, A38, A39, A40, A41, A44, A45, A46, A47, A48, A49, A51, A53, A55, A57 e A58**, destacando-se em recomendações as atividades: **A11, A44, A55 e A58**, de acordo com a tabela 4.3.

Tabela 4.3: Atividades do modelo - Segurança/Anonimização. (Fonte: Autor)

	Atividade	**PDCA**				**Referências**
		Planejamento	**Desenvolvimento**	**Controle**	**Ação**	
A2	Definição do Acesso e Propriedade do Dado	X				[31, 71, 86, 89, 93, 105, 119, 128, 157, 179]
A11	Política de Dados e Privacidade	X				[1, 23, 33, 66, 68, 86, 93, 97, 100, 108, 128, 153, 159, 189]
A15	Controle da Infraestrutura		X			[23, 25, 75, 153]
A16	Arquitetura de Segurança da Informação	X				[1, 10, 20, 55, 59, 61, 75, 89, 92, 108, 111, 126, 139, 145, 153]
A17	Princípios de Dados e Valor de Dados	X		X		[9, 17, 68, 95, 130, 146, 147, 153, 154, 160, 165, 187]
A24	Capacitação de Recursos Humanos	X		X		[23, 86]
A28	Classificação de Dados		X	X		[17, 58, 65, 86, 159, 160, 163, 175, 187]
A34	Avaliação de Riscos de Segurança da Informação			X	X	[2, 52, 61, 86, 107, 164, 166, 170, 189]
A35	Planejamento e Uso de Métodos Criptográficos	X	X			[3, 10, 13, 32, 33, 61, 75, 78, 86, 88, 89, 90, 91, 101, 106, 114, 145, 157, 158, 160, 172, 177, 183]
A36	Definição de Estratégia de Cópias de Segurança (*Backup*)	X	X			[86]
A37	Auditoria			X		[66, 86, 146, 166, 175]
A38	Controle da Transferência de Dados		X			[32, 86, 89, 179]
A39	Política de Desenvolvimento e Teste Seguro	X				[86, 96, 132, 147]
A40	Definição do Contexto da Organização	X				[23, 86, 137]
A41	Coleta do Consentimento do Uso dos Dados	X	X			[31, 68, 86, 105, 166]
A44	Definição dos Objetivos da Segurança da Informação	X				[25]
A45	Definição dos Serviços de Autenticação	X	X			[3, 11, 13, 25, 65, 75, 78, 81, 86, 89, 114, 116, 145, 149, 155, 157, 169, 175, 184]
A46	Deidentificação (*Unlinkability*		X			[25, 33, 98, 147]
A47	Desenvolvimento do Serviço de Proteção a Dados		X			[25, 77, 78, 92, 106, 175]

Continua na próxima página

Tabela 4.3 – continuando da última página

	Atividade	PDCA				Referências
		Planejamento	Desenvolvimento	Controle	Ação	
A48	Uso de Ferramentas de Privacidade Diferencial e K-Anonimização		X			[7, 18, 19, 26, 30, 35, 37, 44, 45, 49, 51, 56, 57, 67, 69, 94, 97, 100, 109, 117, 119, 120, 136, 152, 155, 166, 174, 185, 188]
A49	Uso de Ferramentas de Inteligência Artificial Aplicadas e Aprendizado de Máquina à Segurança (tais como Modelo Oculto de *Markov*, Computação Bayesiana Aproximada e Redes Bayesianas, Aprendizado Federado, Redes Neurais, *Blockchain*, Mineração de Dados, entre outras)		X			[19, 55, 69, 76, 80, 90, 106, 107, 121, 131, 134, 167, 176, 188, 192]
A51	Uso de Ferramentas de Mascaramento de Dados e Embaralhamento		X			[69, 88, 89, 104]
A53	Gestão do Controle de Acesso	X				[10, 20, 33, 71, 74, 81, 86, 92, 100, 111, 127, 157, 169, 175, 179, 184, 189]
A55	Uso de Ferramentas de Criptografia Homomórfica		X			[61, 114, 145, 172]
A57	Definição da Política de Segurança da Informação	X				[36, 47, 61, 77, 86, 99]
A58	Definição dos Requisitos de Segurança	X				[133, 147]

A listagem das atividades das três disciplinas supracitadas responde à pergunta de pesquisa RQ.4, pois representam as práticas mais recomendadas pelos profissionais que lidam com os dados. Podemos então perceber que as boas práticas relacionadas à Governança de Dados, Privacidade de Dados e Segurança de Informação mais recomendadas pelos profissionais que lidam com tratamento de dados pessoais (RQ.4) foram (listadas pelos seus índices): **A8, A11, A14, A20, A22, A24, A32, A34, A44, A51, A53, A55, A56 e A58**. A partir do descarte das atividades julgadas não aplicáveis, de acordo com a opinião dos profissionais da empresa estudo de caso, o *framework* foi construído conforme o especificado a seguir.

4.3 FRAMEWORK

Esta seção destina-se a detalhar o *framework* resultado deste trabalho. A seguir apresentaremos as fases previstas no nosso modelo, bem como os papéis envolvidos e os artefatos sugeridos. Como todo modelo, sugerimos que a aplicação em um cenário real seja acompanhada de uma

análise crítica da viabilidade de aplicação de todas as fases/atividades propostas no *framework*. O conjunto de boas práticas propostas tem sua eficácia comprovada se aplicadas dissociadas do modelo, conforme apresentamos no Capítulo 3 (Metodologia), não havendo obrigatoriedade de aplicação integral do modelo.

4.3.1 Estrutura

O modelo proposto se baseia nos pilares de boas práticas de três principais disciplinas: Governança de Dados (GD), Privacidade de Dados e Segurança da Informação. O enfoque nessas três disciplinas resulta da análise da lei LGPD [42] que endereça os requisitos (ora implícitos, ora explícitos) envolvendo essas três disciplinas. Desta forma, a disposição da visão geral do modelo sugerido é um ciclo PDCA, conforme o proposto por Deming [46], que envolve em todas as suas etapas as três disciplinas basilares. Além disso, relacionamos essas disciplinas em momentos distintos dos projetos de dados, de acordo com o teor geral de suas atividades. De forma geral, as atividades da GD ocorrem nos períodos iniciais do projeto, enquanto planejamento e projeto. De forma semelhante, as atividades da Privacidade tendem a se concentrar no período de desenvolvimento e testes. Por fim, as atividades da Segurança da Informação fazem parte da operação/manutenção dos dados.

Apesar dessa organização (que guiou a disposição dos círculos concêntricos das disciplinas), é preciso perceber que há atividades de todas as disciplinas em todos os momentos do projeto, e que não há restrições para tratar questões da disciplina de Segurança no período de planejamento e projeto, como ocorre com a atividade A36 de *Backup*, por exemplo. A relação temporal aqui delineada é apenas deduzida pela concentração de atividades das disciplinas.

Por seguir o ciclo de Deming [46], o modelo proposto é interativo e incremental, o que representa que as atividades de proteção à fraude e de *compliance* à LGPD devem ser constantes e evolutivas. Como toda atividade de Segurança, as atividades propostas neste modelo visam à proteção de dados em um cenário orgânico e em constante alteração. Portanto, as adaptações das soluções propostas devem ser monitoradas e constantemente evoluídas. A Figura 4.9 apresenta a disposição das disciplinas e etapas do *framework* proposto:

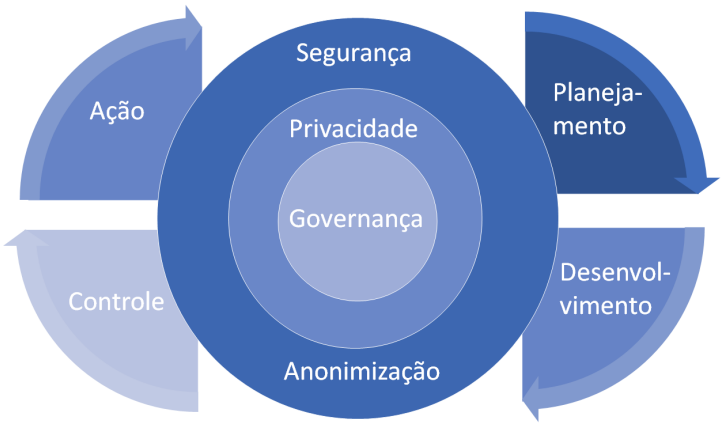

Figura 4.9: Visão Geral do *Framework*. (Fonte: Autor)

Como um ciclo PDCA [46], o modelo indicado possui quatro fases que se encadeiam e interligam, mantendo um ciclo virtuoso das práticas. As fases podem ser descritas como:

1. **Planejamento**: Fase que representa as atividades de projeto, geralmente envolvendo atividades de definições de políticas, processos, responsáveis, requisitos, modelos, recursos, entre outros. Pode receber insumos da execução do ciclo anterior.
2. **Desenvolvimento**: Fase que representa a implementação dos dados, bases e infraestrutura. Envolve atividades de codificação, exploração de dados, construção de *dashboards*, modelos preditivos, classificação de *features*, extração de dados ("extração, transformação e carga"), ciência de dados, entre outras.
3. **Controle**: Fase que representa a monitoração, teste, validação e verificação das soluções propostas e medições em geral. Pode envolver atividades de análise, monitoramento, tratamento de incidentes, manutenção e continuidade de serviços.
4. **Ação**: Fase que representa a análise crítica da solução e proposição de melhorias. Envolve as atividades de melhoria contínua e evolução. Fornece insumos para a fase de Planejamento do próximo ciclo.

A seguir é apresentada uma breve explicação dos itens que se inserem no *framework*, para melhor compreensão de como será apresentado o modelo proposto, conforme apresentado na Figura 4.10.

Figura 4.10: Visão Detalhada do *Framework*: Modelo. (Fonte: Autor)

1. Corresponde à fase do modelo em detalhamento (Planejamento, Desenvolvimento, Controle ou Ação)
2. Corresponde ao sequenciamento lógico das atividades. No exemplo, a atividade AY depende das saídas da atividade AX para ser desenvolvida.
3. Corresponde à disciplina de agrupamento de atividades (Governança, Privacidade ou Segu-

rança/Anonimização)

4. Corresponde à lista de atividades elencadas nas fases e disciplinas correspondentes.

Além da figura explicativa, para cada atividade detalhamos uma breve descrição, seus artefatos de entrada, artefatos de saída e responsáveis sugeridos, conforme o modelo:

Atividade AX - Título da Atividade AX

Descrição:

Descrição da Atividade AX.

Entradas: Artefatos sugeridos como insumos para o desenvolvimento de AX.

Saídas: Artefatos sugeridos como resultados do desenvolvimento de AX.

Responsáveis: Papéis de responsabilidade principal pelo desenvolvimento de AX (conforme papéis definidos na LGPD [42].

Assim sendo, apresentamos os componentes do *framework*, detalhando cada fase, as atividades que compõem as fases, seus artefatos, responsáveis, entradas, saídas e sequenciamento.

4.3.2 Planejamento

A fase de planejamento indica o início de cada novo ciclo do *framework*. Corresponde a uma "preparação" para o desenvolvimento do ciclo que se inicia, atualizando a documentação existente e incluindo novas documentações quando necessárias. Consiste nas atividades de consolidação de políticas (ou revisão delas), projeto de processos, configuração de ferramentas de análise e projeto, definições de escopo, alocação de recursos, entre outros. A Figura 4.11 apresenta o detalhamento das atividades da fase de planejamento.

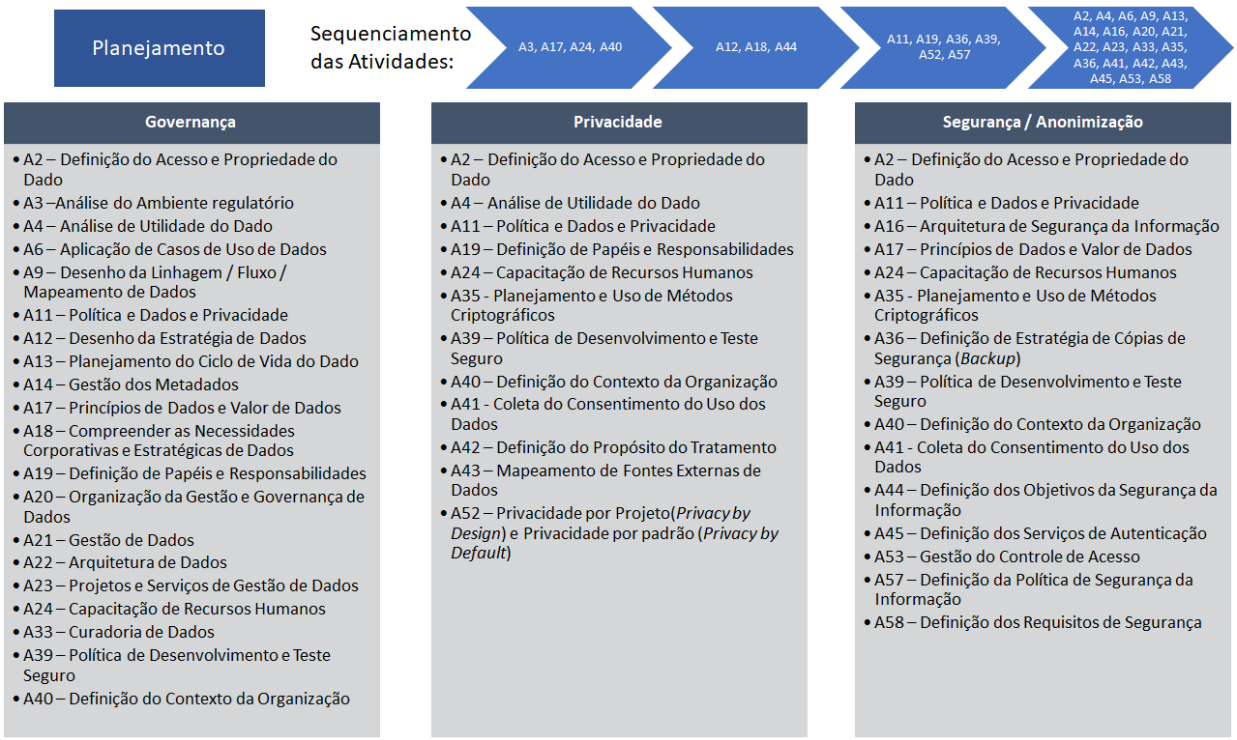

Figura 4.11: Visão Detalhada do *Framework*: Planejamento. (Fonte: Autor)

Atividade A2 - Definição do Acesso e Propriedade ao Dado

Descrição:

Atividade que visa delimitar os agentes autorizados a tratar (nos temos da LGPD [42]) os dados, bem como os papéis que possuem propriedade sobre cada dado tratado pela organização (Figura 4.11).

Entradas: Modelo de dados.

Saídas: Catálogo de dados, Política de Dados e Política de Segurança da Informação.

Responsáveis: Controlador.

Atividade A3 - Análise do Ambiente Regulatório

Descrição:

Atividade que visa levantar os requisitos de conformidade legal envolvidos no negócio e com os dados relacionados (Figura 4.11).

Entradas: Legislação pertinente.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A4 - Análise de Utilidade do Dado

Descrição:

Atividade que visa planejar as informações em seu contexto para definir aspectos que impactam a sua utilidade (dado a finalidade relacionada na coleta das informações) (Figura 4.11).

Entradas: Propósito do Tratamento, Modelo de Dados.

Saídas: Catálogo de dados.

Responsáveis: Controlador.

Atividade A6 - Aplicação de Casos de Uso de Dados

Descrição:

Atividade que visa planejar a utilização do dado, auxiliando no planejamento do propósito de coleta do dado (Figura 4.11).

Entradas: Modelo de Dados.

Saídas: Catálogo de dados.

Responsáveis: Controlador.

Atividade A9 - Desenho da Linhagem/Fluxo/Mapeamento dos Dados

Descrição:

Atividade que visa definir o fluxo de informações inter-sistemas a fim de controlar os variados tratamentos sofridos pelo dado (Figura 4.11).

Entradas: Propósito do Tratamento, Modelo de Dados.

Saídas: Catálogo de dados, Arquitetura de Dados.

Responsáveis: Controlador.

Atividade A11 - Política de Dados e Privacidade

Descrição:

Atividade que visa definir a política de dados e de privacidade da organização. Será feita em forma contributiva com outras tarefas (que também compõem a política de dados da organização) como as atividades A12 (estratégia), A17 (princípios) e A19 (papéis e responsabilidades) (Figura 4.11).

Entradas: Estratégia de Negócio, Política de Dados.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A12 - Desenho da Estratégia de Dados

Descrição:

Atividade que visa alinhar os objetivos estratégicos de dados aos objetivos estratégicos da organização (geralmente decompondo os objetivos de negócio que envolvem tratamento de dados). Compõem a parcela inicial da Política de Dados (Figura 4.11).

Entradas: Estratégia de Negócio, Política de Dados.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A13 - Planejamento do Ciclo de Vida do Dado

Descrição:

Atividade que visa consolidar o projeto do tratamento do dado, considerando todas as fases do dado, desde seu surgimento até seu eventual descarte. Pode compor os documentos de Política de Dados ou de Arquitetura de Dados, dependendo da abrangência da informação (corporativa ou departamental) (Figura 4.11).

Entradas: Política de Dados, Propósito do Tratamento, Modelo de Dados.

Saídas: Política de Dados, Catálogo de Dados.

Responsáveis: Agente de Tratamento.

Atividade A14 - Gestão dos Metadados

Descrição:

Atividade que visa definir o processo de controle dos metadados da organização. Geralmente envolve ferramentas de versionamento de modelos de dados, de catálogo de dados e de arquitetura de dados (Figura 4.11).

Entradas: Política de Dados, Propósito do Tratamento.

Saídas: Modelo de Dados, Catálogo de Dados.

Responsáveis: Agente de Tratamento.

Atividade A16 - Arquitetura da Segurança da Informação

Descrição:

Atividade que visa definir os recursos e infraestrutura necessários para os processos de Segurança da Informação. Podem incluir padrões e boas práticas para o desenvolvimento estrutural de aplicações, entre outros (Figura 4.11).

Entradas: Política de Dados, Política de Segurança da Informação.

Saídas: Política de Segurança da Informação.

Responsáveis: Agente de Tratamento.

Atividade A17 - Princípios de Dados e Valor de Dados

Descrição:

Atividade que visa definir os princípios de dados relevantes para a organização alinhados à estratégia e objetivos do negócio. Dessa análise também resulta o valor dos dados para organização (Figura 4.11).

Entradas: Objetivos de Negócio.

Saídas: Política de Dados, Política de Segurança da Informação.

Responsáveis: Controlador.

Atividade A18 - Compreender as Necessidades Corporativas e Estratégicas de Dados

Descrição:

Atividade que visa definir os requisitos negociais de dados (de forma ampla e corporativa) tais quais como guias de iniciativas de dados da organização e corroborando com a estratégia e objetivos gerais da organização (Figura 4.11).

Entradas: Política de Dados.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A19 - Definição e Papéis e Responsabilidades

Descrição:

Atividade que visa definir quais serão os agentes dos tratamentos de dados em suas variadas formas assim como os responsáveis pela salvaguarda destes ativos. Pode envolver partes interessadas, fornecedores, terceiros contratados como operadores de dados, entre outros (Figura 4.11).

Entradas: Política de Dados.

Saídas: Política de Dados, Política da Segurança da Informação.

Responsáveis: Controlador.

Atividade A20 - Organização da Gestão e Governança de Dados

Descrição:

Atividade que visa definir as instâncias de controle e decisão de assuntos relativos a dados. Pode resultar na criação de comitês de gestores de dados (técnicos e negociais), Conselhos de Governança de Dados, Escritórios de Governança de Dados, Conselho de Encarregados (*DPO - Data Protection Officer*), entre outros (Figura 4.11).

Entradas: Política de Dados.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A21 - Gestão de Dados

Descrição:

Atividade que visa planejar, executar e fiscalizar as políticas, práticas e projetos para adquirir, controlar, proteger, entregar e enriquecer o valor dos ativos de dados e informações [23] (Figura 4.11).

Entradas: Política de Dados.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A22 - Arquitetura de Dados

Descrição:

Atividade que visa planejar as necessidades de informações da organização de forma corporativa, desenvolvendo e mantendo um modelo corporativo de dados, um modelo de arquitetura de tecnologia de dados, um modelo de integração de dados, um modelo de arquitetura de *Data Warehouse* (DW) e *Business Intelligence* (BI), taxonomias e domínios e metadados (Figura 4.11).

Entradas: Política de Dados.

Saídas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados.

Responsáveis: Controlador.

Atividade A23 - Projetos e Serviços de Gestão de Dados

Descrição:

Atividade que visa operacionalizar os projetos e serviços de dados, planejando os recursos, infraestrutura, ferramentas e tecnologias para operação de dados (Figura 4.11).

Entradas: Política de Dados, Arquitetura de Dados, Modelos de Dados, Catálogo de Dados.

Saídas: Serviços de Dados (sem artefatos planejados).

Responsáveis: Controlador.

Atividade A24 - Capacitação de Recursos Humanos

Descrição:

Atividade que visa acompanhar e manter capacitados de forma técnica e gerencial os profissionais que tratam dados, para operar e controlar os dados. Resulta em atividades de *workshops*, treinamentos, capacitação, *benchmarking*, contratação, terceirização, acordos de cooperação, entre outros (Figura 4.11).

Entradas: Política de Dados.

Saídas: Serviços de Dados (sem artefatos planejados).

Responsáveis: Controlador.

Atividade A33 - Curadoria de Dados

Descrição:

Atividade que visa identificar e incentivar os responsáveis pela salvaguarda dos dados bem como os especialistas do tratamento do dado a fim de desenvolver e disseminar o conhecimento dos processos de negócios e fluxos envolvendo os dados corporativos (Figura 4.11).

Entradas: Política de Dados, Catálogo de Dados, Arquitetura de Dados, Modelos de Dados.

Saídas: Política de Dados, Catálogo de Dados, Arquitetura de Dados.

Responsáveis: Controlador.

Atividade A35 - Planejamento do Uso de Métodos Criptográficos

Descrição:

Atividade que visa identificar os contextos corporativos que demandam o uso de criptografia (dados produtivos em ambientes não controlados como desenvolvimento ou teste, ou fornecimento de dados a terceiros). A atividade resulta na adoção de padrões e procedimentos para o uso corporativo (independente de requisito departamental) de criptografia (Figura 4.11).

Entradas: Política de Dados, Catálogo de Dados, Arquitetura de Dados, Modelos de Dados.

Saídas: Política de Dados, Catálogo de Dados, Arquitetura de Dados, Política da Segurança da Informação.

Responsáveis: Controlador.

Atividade A36 - Definição da Estratégia de Cópias de Segurança (*Backup*)

Descrição:

Atividade que visa definir como serão feitas as cópias de segurança, e definir aspectos como local de armazenamento, periodicidade das tomadas das cópias, estilo de tomada de cópias (*full*, incremental, híbrido), entre outros aspectos. (Figura 4.11).

Entradas: Política da Segurança da Informação, Política de Dados.

Saídas: Política da Segurança da Informação.

Responsáveis: Controlador.

Atividade A39 - Política de Desenvolvimento Seguro e Teste Seguro

Descrição:

Atividade que visa identificar as necessidades negociais e corporativas de segurança de desenvolvimento e testes. Aqui definem-se ameaças internas à segurança dos dados, padrões e procedimentos para o desenvolvimento e teste seguros, diretivas de arquitetura, papéis e responsabilidades da segurança da informação, entre outros (Figura 4.11).

Entradas: Política de Dados, Estratégia de Negócio.

Saídas: Política Segurança da Informação.

Responsáveis: Controlador.

Atividade A40 - Definição do Contexto da Organização

Descrição:

Atividade que visa identificar o contexto onde a organização está inserida e o impacto deste contexto nas questões relacionadas a dados. Esta análise pode impactar as questões de contratos com terceiros e fornecedores, políticas de dados e segurança, arquitetura de dados, necessidades negociais e legais, entre outros (Figura 4.11).

Entradas: Política de Dados, Estratégia de Negócio.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A41 - Coleta do Consentimento do Uso dos Dados

Descrição:

Atividade que visa planejar, executar, controlar e manter os procedimentos de coleta de consentimento do Uso de Dados, desenhando os possíveis casos de uso dos dados, os envolvidos, as responsabilidades e questões legais (Figura 4.11).

Entradas: Política de Dados, Estratégia de Negócio.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A42 - Definição do Propósito do Tratamento

Descrição:

Atividade que visa planejar, executar, controlar e manter os procedimentos de tratamento de dados, confeccionando a documentação e promovendo sua divulgação interna para incentivar o conhecimento dos corretos processos e métodos de tratamentos de dados (Figura 4.11).

Entradas: Política de Dados, Estratégia de Negócio.

Saídas: Política de Dados, Catálogo de Dados, Arquitetura de Dados, Política de Segurança, Modelos de dados.

Responsáveis: Controlador.

Atividade A43 - Mapeamento de Fontes Externas de Dados

Descrição:

Atividade que visa documentar e planejar a aquisição de dados externos, mantendo uma fonte de consulta atualizada e consistente da origem de informações (corporativas) produzidas fora da organização (Figura 4.11).

Entradas: Política de Dados, Arquitetura de Dados.

Saídas: Catálogo de Dados, Arquitetura de Dados.

Responsáveis: Controlador.

Atividade A44 - Definição dos Objetivos da Segurança da Informação

Descrição:

Atividade que visa definir, a partir da Estratégia e dos Objetivos de Negócio, as necessidades corporativas da Segurança da Informação. Tem forte relacionamento com a Estratégia e Princípios de Dados, definidos na Política de Dados (Figura 4.11).

Entradas: Política de Dados, Estratégia de Negócio.

Saídas: Política de Segurança da Informação.

Responsáveis: Controlador.

Atividade A45 - Definição dos Serviços de Autenticação

Descrição:

Atividade que visa definir quais os métodos, processos, ferramentas e técnicas serão utilizados para garantir a legitimidade de transações a partir do critério de identidade. Atividade relevante

para os processos de prevenção à fraude (Figura 4.11).

Entradas: Política de Dados, Estratégia de Negócio.

Saídas: Política de Segurança da Informação.

Responsáveis: Controlador.

Atividade A52 - Privacidade por Projeto (*Privacy by Design*) e Privacidade por padrão (*Privacy by Default*)

Descrição:

Atividade que visa estabelecer, documentar e incentivar a instituição destes princípios no tratamento de dados. Tais princípios são tratados de forma especial (em relação aos demais princípios de dados) pois não só estão expressos literalmente na lei de proteção de dados pessoais europeia [151] e indiretamente na lei brasileira [42], mas também fazem parte de várias recomendações de boas práticas, inclusive com efeitos práticos (Figura 4.11).

Entradas: Estratégia de Negócio.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A53 - Gestão do Controle de Acesso

Descrição:

Atividade que visa planejar e definir as ferramentas, processos, métodos e técnicas serão aplicadas para o controle de acessos aos sistemas e recursos. Pode ser desenvolvida em conjunto com a segurança corporativa para incluir aspectos como segurança física, segurança predial, entre outros. Impacta diretamente as atividades de prevenção à fraude (Figura 4.11).

Entradas: Estratégia de Negócio, Política de Dados.

Saídas: Política de Dados, Política de Segurança da Informação.

Responsáveis: Controlador.

Atividade A57 - Definição da Política de Segurança da Informação

Descrição:

Atividade que visa estruturar, documentar, planejar, divulgar e manter a política da Segurança da Informação. Dentre os vários conteúdos que devem compor esta política (já citados nas atividades anteriores), podemos destacar ainda: os procedimentos, técnicas, métodos e ferramentas adotados para a segurança da informação a nível corporativo, os responsáveis e os papéis desempenhados nos processos de segurança da informação, os processos de controle internos e externos envolvidos, padrões e boas práticas de segurança, entre outros (Figura 4.11).

Entradas: Estratégia de Negócio, Política de Dados.

Saídas: Política de Segurança da Informação.

Responsáveis: Controlador.

Atividade A58 - Definição dos Requisitos de Segurança

Descrição:

Atividade que visa definir através das fontes de requisitos corporativos (negócio, legislação pertinente, contexto da organização, fornecedores e partes interessadas, objetivos de dados e de segurança) os requisitos relevantes para o desenho dos processos de Segurança da Informação (Figura 4.11).

Entradas: Estratégia de Negócio, Política de Dados.

Saídas: Política de Segurança da Informação.

Responsáveis: Controlador.

4.3.3 Desenvolvimento

A fase de desenvolvimento concentra as atividades de operação dos dados. É nessa fase que desenvolvemos os projetos de dados e onde há a extração de seu valor, seja manipulando as bases, publicando os dados em visões e relatórios que são insumos para processos decisórios, seja através da exploração por meio de inteligência artificial. Portanto, essa fase concentra medidas de segurança para um uso racional do dado, como aplicações de métodos criptográficos e de anonimização, classificação e categorização de dados, mascaramento e embaralhamento. A Figura 4.12 apresenta o detalhamento das atividades da fase de desenvolvimento:

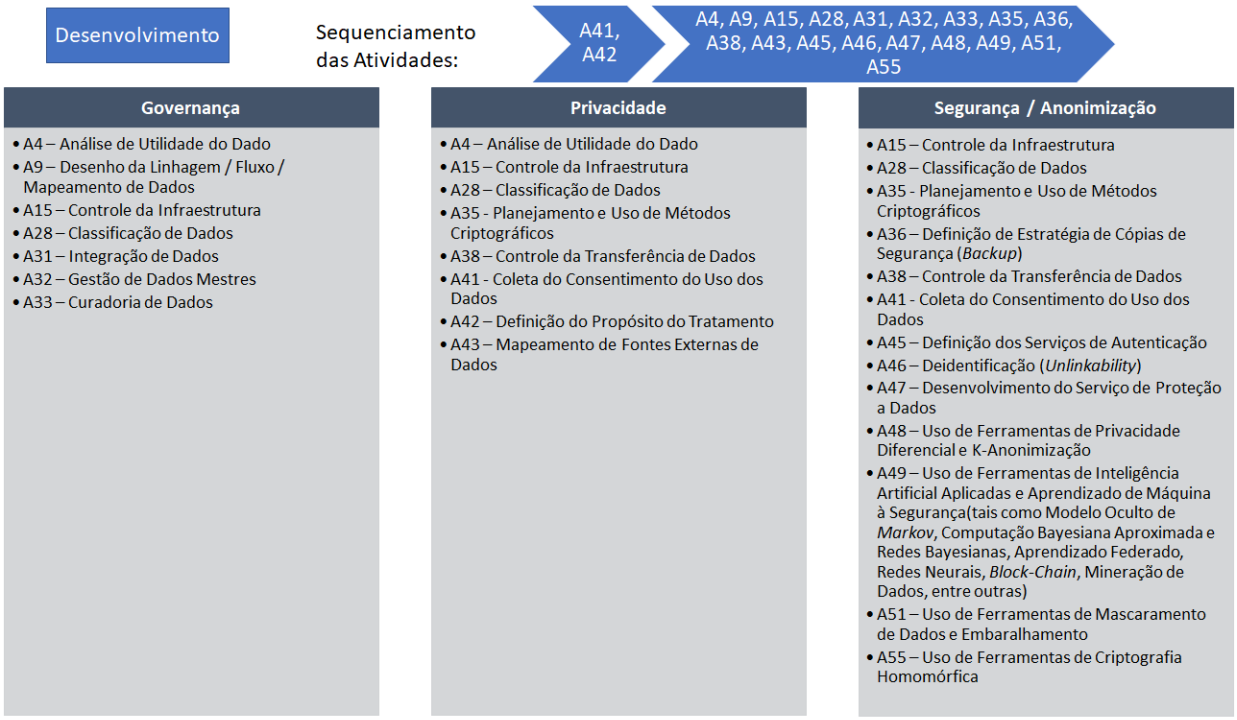

Figura 4.12: Visão Detalhada do *Framework*: Desenvolvimento. (Fonte: Autor)

Como já referenciado no capítulo sobre a metodologia desta pesquisa (Capítulo 3 - Metodologia), algumas atividades podem representar mais de uma disciplina, pelo seu caráter multidisciplinar natural. Portanto, as atividades A4, A9, A33, A35, A36, A41, A42 e A45 que já foram explicadas na fase Planejamento (Subseção 4.3.2), não serão citadas de forma redundante aqui. Na fase de Desenvolvimento, essas atividades representam a implementação das tarefas anteriormente planejadas, resultando nos projetos e tratamentos de dados. Para melhor detalhamento sobre elas, retorne à seção Planejamento (Subseção 4.3.2).

Atividade A15 - Controle da Infraestrutura

Descrição:

Atividade que visa definir os processos de planejamento, acompanhamento, alocação e desalocação de recursos de infraestrutura necessários para a operação de dados e segurança da informação. Pode envolver (sem se limitar) o planejamento de capacidade de armazenagem em disco, servidores, contratos com *datacenters*, disponibilização de nuvens privadas e públicas, ferramentas de software e hardware, disponibilização de *links* de comunicação de rede, entre outros (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Arquitetura de Dados.

Responsáveis: Agente de Tratamento.

Atividade A28 - Classificação de Dados

Descrição:

Atividade que visa planejar, executar, manter e aprimorar a classificação de dados de acordo com as necessidades do negócio e de demandas externas à organização. O controle de dados pessoais sensíveis pode ser uma das demandas que requerem a correta classificação dos dados. Além da classificação, esta atividade é responsável por manter a documentação dessa classificação disponível e atualizada, a fim de que este artefato se torne útil como base para os projetos de dados (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Arquitetura de Dados, Catálogo de Dados, Modelos de Dados.

Responsáveis: Controlador.

Atividade A31 - Integração de Dados

Descrição:

Atividade que visa planejar, executar, fiscalizar e aprimorar os processos de integração de dados entre sistemas, a fim de que haja na organização um modelo padronizado, repetível, seguro e com baixa incidência de erros de integração. A melhor técnica de integração depende do contexto de aplicação e deve ser analisada por uma equipe multidisciplinar que esteja atenta a todas as partes envolvidas. É, portanto, papel das estruturas de Governança e Gestão de Dados definir qual estratégia padronizada será adotada na organização, compondo e recebendo apoio dessa equipe multidisciplinar. Em face da LGPD, aspectos de segurança à privacidade devem ser considerados nos diálogos de integração de dados (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Arquitetura de Dados, Política de Dados.

Responsáveis: Controlador.

Atividade A32 - Gestão de Dados Mestres [e de Referência]

Descrição:

Atividade que visa planejar, implementar e controlar os processos para garantir a consistência dos dados mestres e de referência, provendo as fontes desses dados com confiabilidade, diminuindo custos e riscos por meio de reuso de dados e padrões e suportando *BI* e integrações entre sistemas [23] (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Arquitetura de Dados, Catálogo de Dados, Modelos de Dados.

Responsáveis: Controlador.

Atividade A38 - Controle da Transferência de Dados

Descrição:

Atividade que visa planejar, implementar e controlar os processos de transferência da informação, mantendo a segurança na transferência da informação dentro da organização e com quaisquer entidades externas (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Arquitetura de Dados, Política de Dados.

Responsáveis: Controlador.

Atividade A46 - Deidentificação (*Unlinkability*)

Descrição:

Atividade que visa planejar os cenários de uso do dado que requerem deidentificação de dados dos titulares da informação. Este processo é necessário para conformidade com a LGPD [42], uma vez que se espera que nem todos os dados, pelo menos de sistemas legados, estejam sobre o alcance do consentimento do usuário para seu uso. Desta forma, parte da informação deve ser deidentificada para prosseguimento do tratamento. Essa atividade visa responder às seguintes perguntas: Quando aplicar as regras de deidentificação? Quais regras aplicar? Em que cenários? E como serão aplicadas essas regras? (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Catálogo de Dados, Modelos de Dados, Política de Dados.

Responsáveis: Agente de Tratamento.

Atividade A47 - Desenvolvimento do Serviço de Proteção a Dados

Descrição:

Atividade que visa planejar, executar, fiscalizar e aprimorar os serviços de proteção aos dados, considerando os padrões, boas práticas, métodos, técnicas, ferramentas e processos definidos nas Políticas de Dados e de Segurança da Informação e as implementando nos cenários de projetos de dados (Figura 4.12).

Entradas: Política de Dados e Política de Segurança da Informação.

Saídas: Serviço de Proteção a Dados (sem artefato).

Responsáveis: Controlador.

Atividade A48 - Uso de Ferramentas de Privacidade Diferencial e K-Anonimização

Descrição:

Atividade que visa identificar a viabilidade do uso de ferramentas de privacidade diferencial e k-anonimização nos cenários de projetos de dados. Pode incluir definições de configurações dos ambientes e ferramentas, padrões de infraestrutura e compartilhamento de recursos, entre outros. Esta atividade não indica, contudo, que todas as empresas e projetos deverão fazer uso de tais ferramentas (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Projetos de Dados (sem artefato).

Responsáveis: Agente de Tratamento.

Atividade A49 - Uso de Ferramentas de Inteligência Artificial Aplicadas e Aprendizado de Máquina à Segurança, tais como Modelo Oculto de *Markov*, Computação Bayesiana Aproximada e Redes Bayesianas, Aprendizado Federado, Redes Neurais, *Blockchain*, Mineração de Dados, entre outras

Descrição:

Atividade que visa identificar a viabilidade do uso de ferramentas de inteligência artificial e aprendizado de máquina aplicáveis à segurança da informação nos cenários de projetos de dados. Pode incluir definições de configurações dos ambientes e ferramentas, padrões de infraestrutura e compartilhamento de recursos, entre outros. Esta atividade não indica, contudo, que todas as empresas e projetos deverão fazer uso de tais ferramentas (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Projetos de Dados (sem artefato).

Responsáveis: Agente de Tratamento.

Atividade A51 - Uso de Ferramentas de Mascaramento de Dados e Embaralhamento

Descrição:

Atividade que visa identificar cenários de uso para as ferramentas de mascaramento e embaralhamento de dados. Ao contrário do definido nas últimas atividades, recomendamos o uso das ferramentas propostas nesta atividade para a proteção de dados em ambientes de desenvolvimento/testes, a fim de que, de forma mínima, os dados produtivos caso trafegados em ambientes inseguros não sejam trivialmente identificados (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Política de Dados, Política de Segurança da Informação, Arquitetura de Dados.

Responsáveis: Agente de Tratamento.

Atividade A55 - Uso de Ferramentas de Criptografia Homomórfica

Descrição:

Atividade que visa identificar a viabilidade do uso de ferramentas de criptografia homomórfica nos cenários de projetos de dados. Pode incluir definições de configurações dos ambientes e ferramentas, padrões de infraestrutura e compartilhamento de recursos, entre outros. Esta atividade não indica, contudo, que todas as empresas e projetos deverão fazer uso de tais ferramentas (Figura 4.12).

Entradas: Arquitetura de Dados, Modelos de Dados, Catálogo de Dados, Projetos de Dados, Política de Dados e Política de Segurança da Informação.

Saídas: Projetos de Dados (sem artefato).

Responsáveis: Agente de Tratamento.

4.3.4 Controle

A fase de controle concentra as atividades de monitoramento, medição, avaliação e verificação. As atividades listadas aqui têm o objetivo de mensurar a qualidade, o alcance, a completude e

eficiência das atividades definidas e implementadas anteriormente. Tais medições são insumos para o processo de melhoria contínua e adaptação de novos cenários em cada ciclo de execução do *framework*. A Figura 4.13 e as descrições seguintes apresentam o detalhamento das atividades da fase de controle.

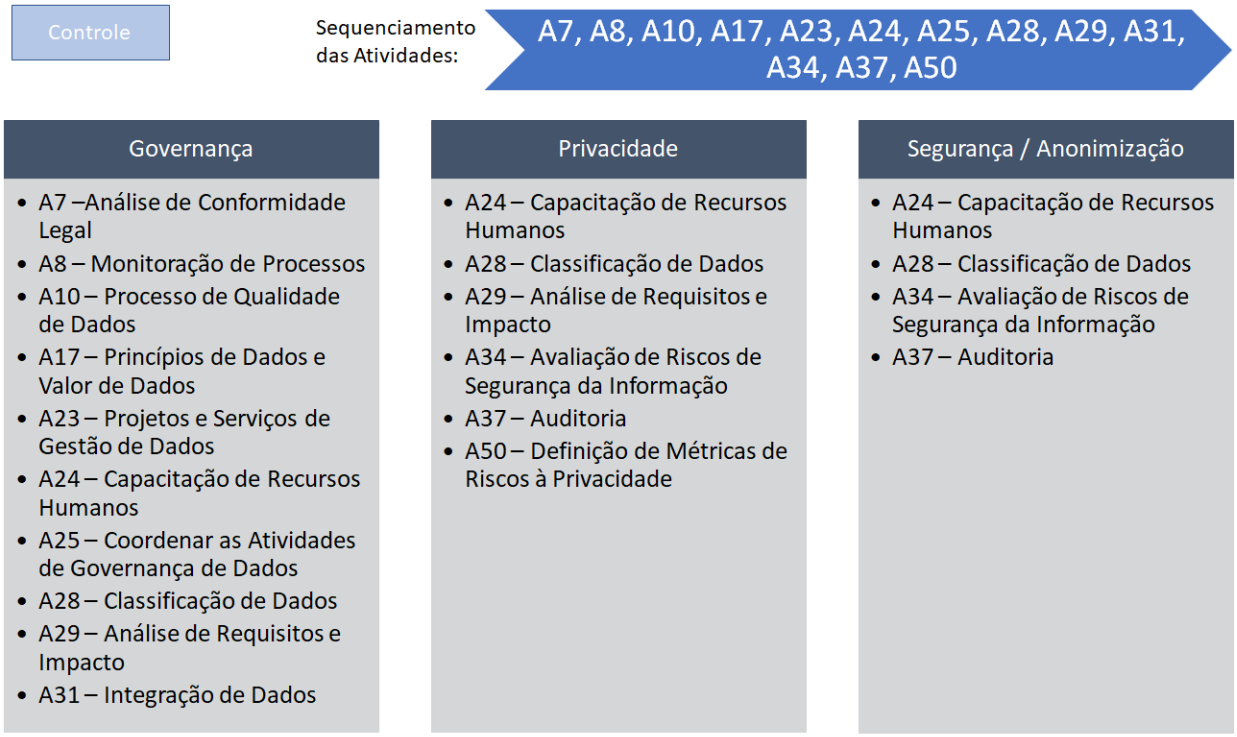

Figura 4.13: Visão Detalhada do *Framework*: Controle. (Fonte: Autor)

As atividades A17, A23, A24 e A28 já foram explicadas nas fases anteriores Planejamento (Subseção 4.3.2) e Desenvolvimento (Subseção 4.3.3). Na fase de Controle, essas atividades representam a implementação das métricas e controles definidos anteriormente. Para melhor detalhamento sobre elas, retorne às seções Planejamento (Subseção 4.3.2) e Desenvolvimento (Subseção 4.3.3).

Atividade A7 - Análise de Conformidade Legal

Descrição:

Atividade que visa descrever a necessidade de constante acompanhamento do ambiente legal envolvendo o tratamento de dados, não apenas em legislações aplicadas a dados, como no regramento específico de cada nicho de informação. As equipes e especialistas que lidam com dados devem atentar sempre para a legislação pertinente e aprimorar com tempestividade os processos inconformes (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Ambiente Legal Específico.

Saídas: Política de Dados, Política de Segurança de Dados, Projetos de Dados.

Responsáveis: Todos.

Atividade A8 - Monitoração de Processos

Descrição:

Atividade que visa planejar e implementar os critérios para avaliação das atividades aplicadas a fim de coletar a efetividade (eficácia e eficiência) dos processos definidos bem como o atingimento dos objetivos propostos. Gera insumos para os ciclos posteriores através da análise dos resultados das medições (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados.

Saídas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados.

Responsáveis: Controlador.

Atividade A10 - Processo de Qualidade de Dados

Descrição:

A Qualidade de Dados é um conjunto de atividades importante com efeitos em diversos domínios e disciplinas, sendo uma das atividades de maior destaque deste modelo (e um dos métodos de validação também). Para tal, esta atividade preconiza o planejamento, implementação, controle, fiscalização e otimização da qualidade de dados, instituindo medidas, processos, ferramentas, envolvidos, critérios e escopo para o constante aprimoramento da qualidade de dados tratados (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados, Catálogo de Dados, Modelo de Dados, Projetos de Dados.

Saídas: Projetos de Dados (sem artefatos).

Responsáveis: Todos.

Atividade A25 - Coordenar as Atividades de Governança de Dados

Descrição:

Atividade que visa acompanhar a implantação e evolução da governança de dados, garantido a adaptação dos processos de governança de dados aos ambientes voláteis e em constante evolução. Envolve a definição de medidas de desempenho de processos, atingimento de objetivos, cumprimento de metas de processos, acompanhamento de projetos, entre outros (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados, Catálogo de Dados, Modelo de Dados, Projetos de Dados.

Saídas: Política de Dados.

Responsáveis: Controlador.

Atividade A29 - Análise de Requisitos e Impacto

Descrição:

Além dos requisitos de dados, negociais, corporativos e legais, com a entrada em vigor da LGPD [42], as organizações podem ser convocadas pela ANPD a prestar o relatório de impacto à proteção de dados pessoais. Para tal, é preciso planejar, controlar, documentar e tornar disponível uma análise sempre atualizada dos riscos à privacidade de dados, bem como os planos de respostas a esses riscos. Avaliações de requisitos de segurança também fazem parte desta e de outras atividades como a atividade A34 (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados, Catálogo de Dados, Modelo de Dados, Projetos de Dados.

Saídas: Política de Dados, Política de Segurança da Informação, Relatório de Impacto à Proteção de Dados Pessoais.

Responsáveis: Controlador.

Atividade A34 - Avaliação de Riscos de Segurança da Informação

Descrição:

Atividade que visa planejar, implementar, controlar e otimizar os critérios de avaliação dos riscos da Segurança da Informação, categorizando tais riscos e desenvolvendo um plano de resposta aos riscos. Definições comuns dessa atividade são: os ativos a serem avaliados; quando os riscos podem ocorrer; qual o impacto, severidade e probabilidade; quais mecanismos de resposta ao risco; quais planos de contingência; e quais serviços e recursos serão afetados (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados, Catálogo de Dados, Modelo de Dados, Projetos de Dados.

Saídas: Política de Segurança da Informação, Política de Dados.

Responsáveis: Controlador.

Atividade A37 - Auditoria

Descrição:

Atividade que visa definir de maneira ampla como será aplicada a auditoria aos processos, definindo responsáveis, fases de execução, planos de melhoria e respostas a inconformidades (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados, Catálogo de Dados, Modelo de Dados, Projetos de Dados.

Saídas: Política de Segurança da Informação, Política de Dados.

Responsáveis: Controlador.

Atividade A50 - Definição de Métricas de Riscos à Privacidade

Descrição:

Em conjunto com as atividades A29 e A34, essa atividade visa explicitar o tratamento e acompanhamento do risco à privacidade da informação; e pode constar do Relatório de Impacto a Proteção de Dados Pessoais (Figura 4.13).

Entradas: Política de Dados, Política de Segurança de Dados, Arquitetura de Dados, Catálogo de Dados, Modelo de Dados, Projetos de Dados.

Saídas: Política de Segurança da Informação, Política de Dados.

Responsáveis: Controlador.

4.3.5 Ação

A fase de ação é responsável pela análise crítica das métricas definidas/coletadas na fase de Controle (Seção 4.3.4) e consolida as definições de melhorias para o próximo ciclo. Os processos são auditados, monitorados, controlados e geram insumos para resolução de pontos de melhoria de performance que são incorporados em evolução de planejamento e documentação. O detalhamento da fase pode ser visualizado na Figura 4.14.

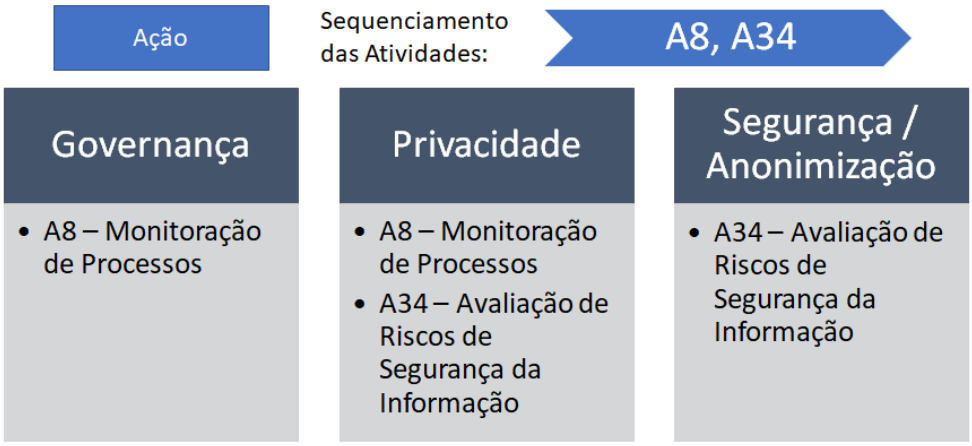

Figura 4.14: Visão Detalhada do *Framework*: Ação. (Fonte: Autor)

As atividades A8 e A34 já foram explicadas na fase Controle (Seção 4.3.4). Na fase de Ação, essas atividades representam a consolidação, interpretação e análise das medições obtidas, planejamento das sugestões de melhoria e preparação para a execução do próximo ciclo. Para melhor detalhamento sobre elas, retorne à seção Controle (Seção 4.3.4).

Exemplos dos artefatos aqui sugeridos foram preenchidos para o estudo de caso e podem ser consultados através da *url* <https://drive.google.com/drive/folders/1Xy3mnXn2DZi3orWMbkyMu-SqNWh7crWw?usp=sharing>.

4.4 ESTUDO DE CASO

Para a validação do *framework* e do conjunto de atividades sugeridas pelos profissionais da área de Tecnologia da Informação, foi realizada uma comparação em dois projetos de ingestão de dados em *Big Data* utilizando a mesma base de dados. O primeiro projeto (Projeto 1) realizou a ingestão dos dados sem observar as práticas sugeridas no *framework* (Seção 4.3 - Framework, deste capítulo), seguindo basicamente apenas os *scripts* de ingestão (extração de dados, criação das estruturas no destino e carga).

Em ambas as bases resultantes (Projeto 1 e 2), foram executadas as métricas de qualidade de dados e privacidade de dados especificadas no Capítulo 3 - Metodologia, como forma de estimar a adequação das bases ao contexto de proteção à privacidade de dados e de prevenção a fraudes. A comparação de tais métricas indica se o uso do *framework* proposto atende ao objetivo de melhorar os aspectos de proteção à privacidade e prevenção a fraude de processos de ingestão de dados em *Big Data*, conforme o apresentado na subquestão de pesquisa RQ.5. A seguir apresentamos um breve resumo do contexto dos projetos e das métricas obtidas, além das comparações

pertinentes. O modelo lógico dos dados envolvidos nos projetos pode ser visualizado na Figura 4.15.

Figura 4.15: Descrição dos dados do projeto estudo de caso. (Fonte: Autor)

Conforme apresentado na Figura 4.15, os dados envolvidos na ingestão fazem referência a informações de clientes e transferências bancárias, passíveis da análise de fraude e lavagem de dinheiro pela empresa. Os dados possuem informações pessoais (*PII*), tais como endereço de rede, número de cadastro de pessoas físicas, número de conta, além de informações gerais que podem levar a identificação de indivíduos (*AD*), tais como gerente da conta, ordem da conta (titular ou controlador), data de transação, entre outros.

Esses dados já compõem a solução de análise e prevenção a fraude da empresa, atualmente em tecnologia de baixa plataforma (servidores de menor capacidade, não sendo em *mainframe*) e no contexto de identificação de fraudes à *posteriori*. A ingestão destes dados em *Big Data* é um passo para implementar uma solução de intervenção em tempo real a transações para detecção de fraudes e bloqueio de transações suspeitas. É, portanto, imprescindível que os dados ingeridos possuam um nível de qualidade aceitável para a análise fim.

O processo de ingestão mencionado é composto das etapas de extração das informações contidas no sistema de prevenção a fraude, resultando em arquivos delimitados (csv), transporte desses arquivos via rede interna para o cluster de *Big Data* (representado neste exemplo pela máquina virtual *"Centos7 CDP-DC Trial VM"* disponível em <https://www.cloudera.com/downloads/hortonworks-sandbox.html>) e, por fim, a carga destes arquivos no sistema HDFS utilizando a tecnologia Cloudera Data Flow (antigo HortonWorks Ambari). O sistema presente no estudo de caso é integrado por bancos de dados em tecnologia Oracle versão 11g de baixa plataforma. A Figura 4.16 apresenta o esquemático da ingestão.

A especificação da máquina que representou no exemplo proposto no ambiente *Big Data* é apresentada na Figura 4.17. Os dados de origem possuem a seguinte volumetria: 60782 registros de

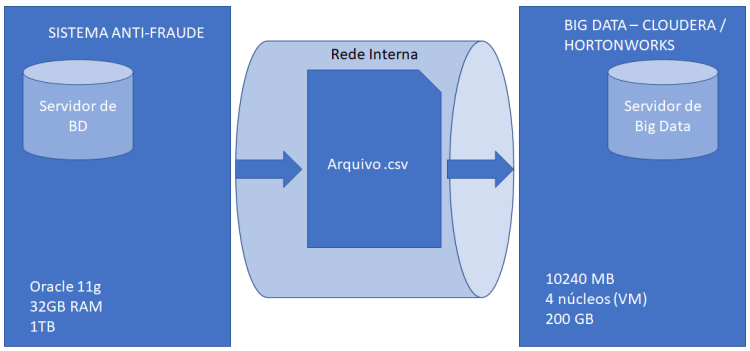

Figura 4.16: Processo de ingestão de dados. (Fonte: Autor)

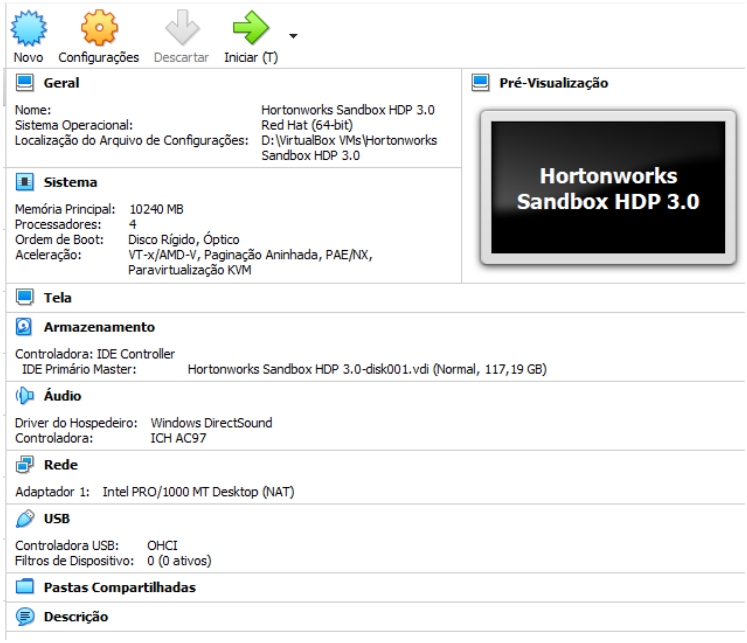

Figura 4.17: Especificações da VM *Big Data*. (Fonte: Autor)

cliente (“AX_CLIENTE”) totalizando 6,05 Megabytes de informação e 3617 registros de transferências suspeitas (“TB_ INDICIOS_TRANSFERENCIA”) totalizando 940 Kilobytes de informação. Após seguir o processo de ingestão (Projeto 1), foram tomadas as medidas de qualidade e privacidade conforme apresentado na Tabela 4.4.

Tabela 4.4: Métricas de Qualidade sobre Tabelas - Projeto 1. (Fonte: Autor)

Tabela	**Métrica de Unicidade**	**Métrica de Integridade**
AX_CLIENTE	77,6085025172%	0,0230331348%
TB_INDICIOS_TRANSFERENCIA	100%	0,8017694222%

A base resultante da ingestão do Projeto 1 apresentou um repositório de clientes com 77% de unicidade, isto é, 23% dos registros de clientes estão duplicados nesta base. Além disso, a base resultante apresenta fortes problemas de integridade, uma vez que os dados de conta no repositório de clientes correspondem a menos de 1% das contas presentes no repositório de transações. Além de métricas sobre as tabelas, foram aferidas métricas sobre as colunas de dados, conforme

apresentado nas Tabelas 4.5 e 4.6.

Tabela 4.5: Métricas de Qualidade sobre Colunas - Projeto 1 - parte 1. (Fonte: Autor)

Tabela/Coluna	**Métrica de Completude**	**Métrica de Razoabilidade**	**Métrica de Validade**
AX_CLIENTE.CD_CLIENTE	98,5587838505%	98,5587838505%	100%
AX_CLIENTE.NR_CONTA	100%	100%	100%
AX_CLIENTE.NR_ORDEM	100%	99,9983547761%	100%
AX_CLIENTE.CD_ORIGEM_CONTA	100%	100%	100%
AX_CLIENTE.DS_ORIGEM_CONTA	100%	100%	100%
AX_CLIENTE.CD_EMPRESA	100%	100%	100%
AX_CLIENTE.CD_DEPENDENCIA	100%	100%	100%
AX_CLIENTE.NM_DEPENDENCIA	100%	100%	100%
AX_CLIENTE.NR_MATRICULA_GERENTE	55,9919054983%	55,9919054983%	100%
AX_CLIENTE.NM_GERENTE	55,9919054983%	55,9919054983%	100%
AX_CLIENTE.ST_CONTRATO_UNICO	99,9983547761%	99,9983547761%	100%
AX_CLIENTE.ST_KIT_SERVICO	99,9983547761%	99,9983547761%	100%

Tabela 4.6: Métricas de Qualidade sobre Colunas - Projeto 1 - parte 2. (Fonte: Autor)

Tabela/Coluna	**Métrica de Completude**	**Métrica de Razoabilidade**	**Métrica de Validade**
TB_INDICIOS_TRANSFERENCIA.DT_OCORRENCIA	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_IP	52,3638374343%	52,1979541056%	100%
TB_INDICIOS_TRANSFERENCIA.NR_CPF	78,4628144871%	78,4628144871%	100%
TB_INDICIOS_TRANSFERENCIA.NR_CONTA_DEBITO	99,9170583356%	99,9170583356%	100%
TB_INDICIOS_TRANSFERENCIA.NO_CLIENTE	99,7235277855%	99,7235277855%	100%
TB_INDICIOS_TRANSFERENCIA.NR_CONTA_CREDITO	99,9170583356%	99,9170583356%	100%
TB_INDICIOS_TRANSFERENCIA.NO_CORRESPONDENTE	99,7235277855%	99,7235277855%	100%
TB_INDICIOS_TRANSFERENCIA.DS_MOTIVO	83,3010782416%	100%	100%
TB_INDICIOS_TRANSFERENCIA.VL_TRANSACAO	99,9170583356%	99,9170583356%	100%
TB_INDICIOS_TRANSFERENCIA.CD_CANAL	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.ST_INDICIO	97,5393972906%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_MATRICULA_ANALISTA	97,5393972906%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DT_ANALISE	97,5393972906%	100%	100%
TB_INDICIOS_TRANSFERENCIA.ST_CORRESPONDENTE	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.SQ_EVTSEQ	99,9170583356%	99,9170583356%	100%
TB_INDICIOS_TRANSFERENCIA.SQ_INDICIO	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_AUTENTICACAO	0%	0%	100%

Das métricas apresentadas (Tabelas 4.5 e 4.6), é possível perceber que a base resultante do projeto 1 possui problemas de completude uma vez que alguns campos alcançam marcas de metade dos

registros preenchidos, e problemas de razoabilidade, isto é, o conteúdo do campo nem sempre é razoável ou condiz com uma informação esperada para o tipo do dado. Entretanto, todas as informações aferidas eram válidas, ou seja, o tipo da informação condizia com o tipo de dado documentado, por exemplo, os campos data estavam em atributos documentados como de datas, códigos numéricos em atributos do tipo *"NUMBER"*, e assim por diante. De forma complementar, foram extraídas métricas de *performance* (tanto de carga, como de consulta) conforme apresentado nas Figuras 4.18, 4.19, 4.20 e 4.21.

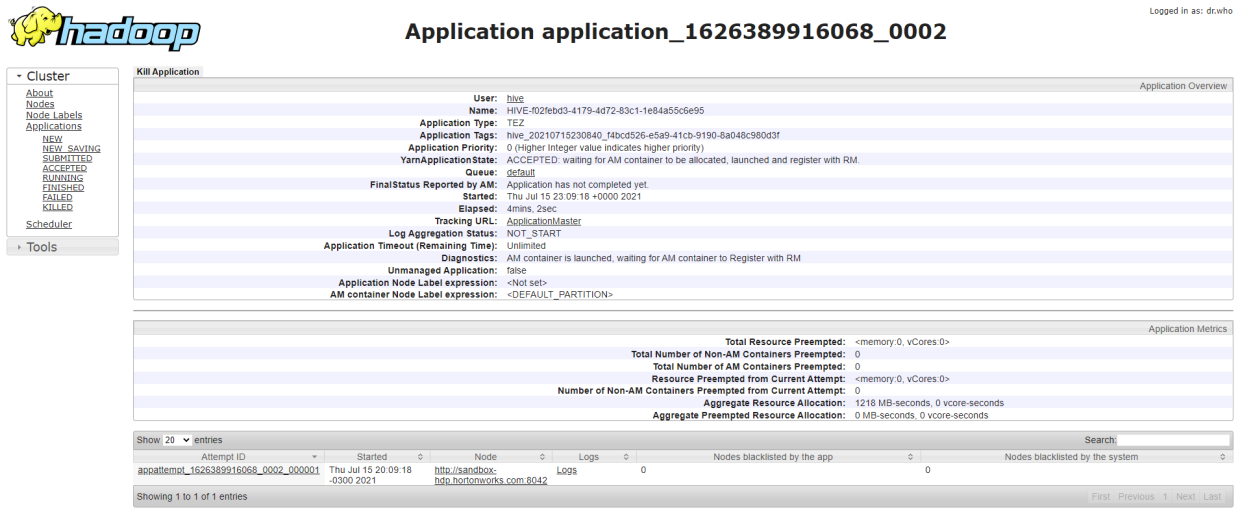

Figura 4.18: Métrica de *Performance* (carga AX_CLIENTE) do Projeto 1. (Fonte: Autor)

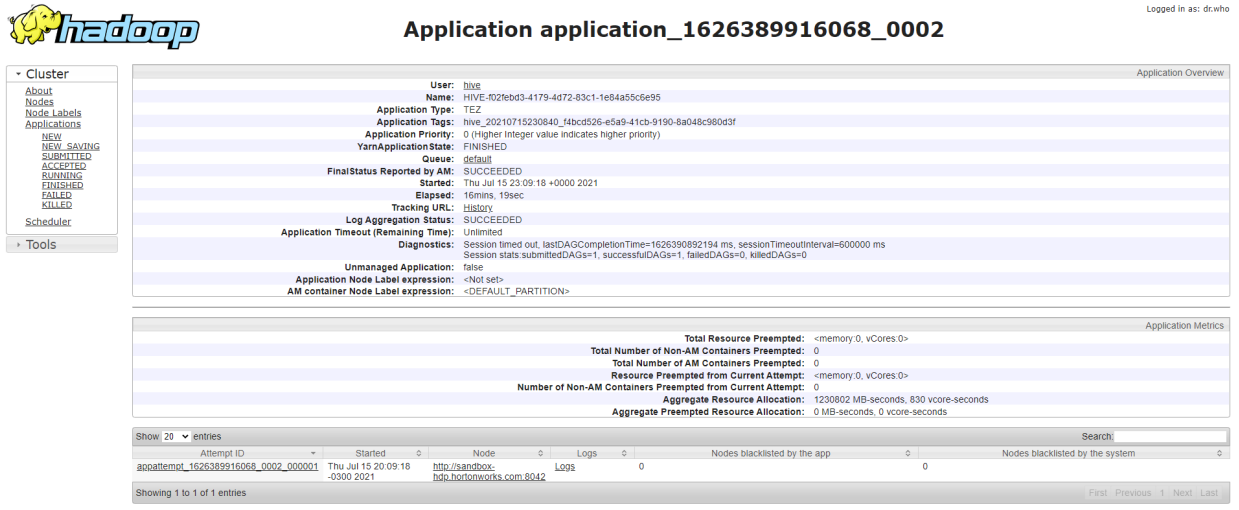

Figura 4.19: Métrica de *Performance* (carga TB_INDICIOS_TRANSFERENCIA) do Projeto 1. (Fonte: Autor)

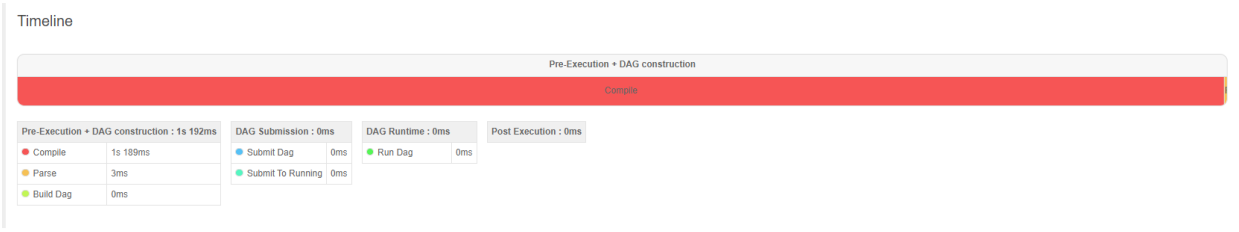

Figura 4.20: Métrica de *Performance* (consulta AX_CLIENTE) do Projeto 1. (Fonte: Autor)

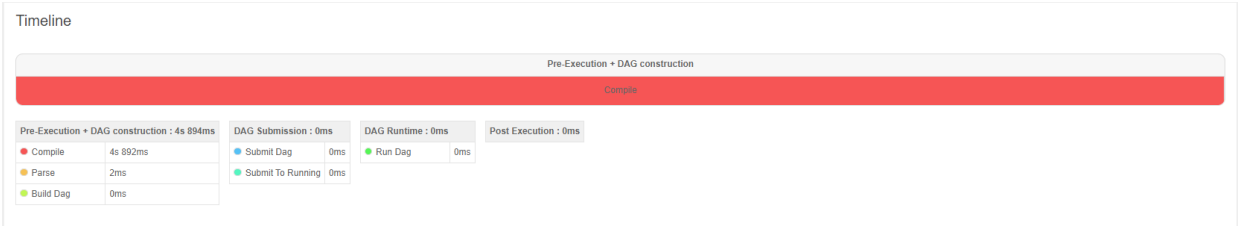

Figura 4.21: Métrica de *Performance* (consulta TB_INDICIOS_TRANSFERENCIA) do Projeto 1. (Fonte: Autor)

As tabelas de dados levaram 4 minutos e 2 segundos e 16 minutos e 19 segundos para serem criadas e carregadas, totalizando um período de **20 minutos e 21 segundos** de processamento total de carga do Projeto 1. Com relação a consulta, a tabela de cliente concluiu uma consulta completa (*"full table scan"*) em 1 segundo e 192 milésimos de segundo, enquanto a tabela de indícios performou a consulta em 4 segundos e 894 milésimos de segundo totalizando **6 segundos e 86 milésimos de segundos** de consultas.

A última métrica aferida foi a métrica de privacidade, calculada de acordo com as fórmulas 3.1, 3.2 e 3.3. Para calcular a privacidade, e considerando que a métrica proposta em essência é um cálculo de correlação entre duas variáveis, consideramos as variáveis, isto é, os atributos "AX_CLIENTE.NR_CONTA", "TB_INDICIOS_TRANSFERENCIA.NR_CONTA_DEBITO" e "TB_INDICIOS_TRANSFERENCIA.NR_CONTA_CREDITO". A relação entre estes três atributos foi selecionada para o cálculo de privacidade pois estes campos são a ligação entre os dados pessoais (representados pela tabela "AX_CLIENTE") e os dados sensíveis (representados pela tabela "TB_INDICIOS_TRANSFERENCIA").

Como explicado no Capítulo 3 (Metodologia), a métrica selecionada para mensurar a privacidade de dados foi uma adaptação da métrica de confidencialidade proposta por Domingo-Ferrer et. al. [51], baseada no cálculo de correlações canônicas entre as informações originais e informações anonimizadas. Para o contexto dessa pesquisa, consideramos o cálculo o coeficiente de *Pearson* que relaciona duas variáveis qualitativas e se baseia no cálculo da distribuição chi quadrado. Adaptamos esse coeficiente como a relação entre informações identificadoras de indivíduos e as demais informações da base (assumindo o papel do maior coeficiente de correlações canônicas de Domingo-Ferrer et. al. [51]) e o subtraindo de 1. Esta equação resulta em um índice, de 0 a 1, da correlação entre informações identificadoras (*"PII"*) e as demais informações, indicando a facilidade de identificar um indivíduo na base a partir de informações não identificadoras. A Tabela 4.7 apresenta o resultado desta aferição.

Tabela 4.7: Cálculo da métrica de privacidade CM1 (Projeto 1). (Fonte: Autor)

Relação	**CM1**
AX_CLIENTE x TB_INDICIOS_TRANSFERENCIA	7,90449494903335749426602560484 E-11

É interessante perceber o baixíssimo valor da métrica CM1 de privacidade (Tabela 4.7), indicando que as variáveis estão fortemente ligadas. Isto ocorre porque ambos os campos possuem a mesma regra de formação, isto é, são campos que armazenam números de contas bancárias. Outro fator agravante para esta forte correlação é que, geralmente, estes campos realmente são relacionados, isto é, clientes de uma conta costumam realizar transferências para um conjunto finito de outras contas, que geralmente não coincidem com o conjunto de contas envolvidas nas transferências de outros clientes. Uma fórmula matemática pode, portanto, relacionar os dois grupos de forma unívoca com uma certa facilidade. Por fim, este fenômeno pode ter raízes na presença de dados de má qualidade e em grande quantidade. Isso pode configurar relações espúrias entres os dois conjuntos de dados, relações estas que apesar de não aproveitáveis, contribuem para a diminuição da métrica. Para finalizar, foi preenchida a tabela listagem dos critérios tratamento de dados para prevenção à fraude da LGPD conforme a Tabela 4.8.

Tabela 4.8: Lista de checagem dos critérios de tratamento de dados para prevenção à fraude, em conformidade com a LGPD. (Projeto 1). (Fonte: Autor)

	Questão	**X**	**V**
CK1	O propósito do tratamento do dado foi concebido?		X
CK2	O propósito do tratamento do dado foi documentado?	X	
CK3	O propósito do tratamento do dado está acessível?	X	
CK4	A linhagem (origem e destino) do dado é documentada?		X
CK5	O dado é classificado (dado pessoal, dado sensível)?	X	
CK6	A classificação do dado está acessível?	X	

Após a execução do primeiro projeto e da tomada das métricas, o processo preconizado pelo *framework* proposto foi aplicado em uma nova ingestão dos mesmos dados. Esta nova ingestão, além do preenchimento dos artefatos indicados, envolveu atividades de um processo de qualidade, onde o arquivo intermediário gerado foi submetido a uma minimização de dados, com o intuito de melhorar a integridade referencial dos dados.

Neste processo de minimização de dados, as informações que não atendiam a integridade referencial entre os campos “AX_CLIENTE.NR_CONTA”, “TB_INDICIOS_TRANSFERENCIA. NR_CONTA_DEBITO” e “TB_INDICIOS_TRANSFERENCIA.NR_CONTA_CREDITO” foram descartadas. Por este motivo, as métricas de qualidade sobre as tabelas para o segundo projeto foram máximas, conforme apresentado na Tabela 4.9.

Tabela 4.9: Métricas de Qualidade sobre Tabelas - Projeto 2. (Fonte: Autor)

Tabela	**Métrica de Unicidade**	**Métrica de Integridade**
AX_CLIENTE	100%	100%
TB_INDICIOS_TRANSFERENCIA	100%	100%

Além disso, a aplicação de um processo de qualidade de dados, mesmo que inicial, teve um impacto positivos nas métricas de qualidade das colunas, conforme apresentado nas Tabelas 4.10 e 4.11.

Tabela 4.10: Métricas de Qualidade sobre Colunas - Projeto 2 - parte 1. (Fonte: Autor)

Tabela/Coluna	**Métrica de Completude**	**Métrica de Razoabilidade**	**Métrica de Validade**
AX_CLIENTE.CD_CLIENTE	100%	100%	100%
AX_CLIENTE.NR_CONTA	100%	100%	100%
AX_CLIENTE.NR_ORDEM	100%	100%	100%
AX_CLIENTE.CD_ORIGEM_CONTA	100%	100%	100%
AX_CLIENTE.DS_ORIGEM_CONTA	100%	100%	100%
AX_CLIENTE.CD_EMPRESA	100%	100%	100%
AX_CLIENTE.CD_DEPENDENCIA	100%	100%	100%
AX_CLIENTE.NM_DEPENDENCIA	100%	100%	100%
AX_CLIENTE.NR_MATRICULA_GERENTE	92,8571428571%	92,8571428571%	100%
AX_CLIENTE.NM_GERENTE	92,8571428571%	92,8571428571%	100%
AX_CLIENTE.ST_CONTRATO_UNICO	100%	100%	100%
AX_CLIENTE.ST_KIT_SERVICO	100%	100%	100%

Tabela 4.11: Métricas de Qualidade sobre Colunas - Projeto 2 - parte 2. (Fonte: Autor)

Tabela/Coluna	**Métrica de Completude**	**Métrica de Razoabilidade**	**Métrica de Validade**
TB_INDICIOS_TRANSFERENCIA.DT_OCORRENCIA	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_IP	68,4210526316%	65,7894736842%	100%
TB_INDICIOS_TRANSFERENCIA.NR_CPF	81,5789473684%	81,5789473684%	100%
TB_INDICIOS_TRANSFERENCIA.NR_CONTA_DEBITO	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.NO_CLIENTE	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.NR_CONTA_CREDITO	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.NO_CORRESPONDENTE	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_MOTIVO	84,2105263158%	100%	100%
TB_INDICIOS_TRANSFERENCIA.VL_TRANSACAO	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.CD_CANAL	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.ST_INDICIO	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_MATRICULA_ANALISTA	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DT_ANALISE	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.ST_CORRESPONDENTE	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.SQ_EVTSEQ	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.SQ_INDICIO	100%	100%	100%
TB_INDICIOS_TRANSFERENCIA.DS_AUTENTICACAO	0%	0%	100%

É possível perceber que absolutamente todas as métricas colunares representaram melhorias se comparadas às obtidas no Projeto 1. Este efeito positivo indica um impacto positivo nas análises

de fraude, uma vez que dados de melhor qualidade dão melhores insumos para a tomada de decisão. Entretanto, a minimização de dados acabou tendo um efeito reverso nas métricas de performance, conforme apresentado nas Figuras 4.22, 4.23, 4.24 e 4.25.

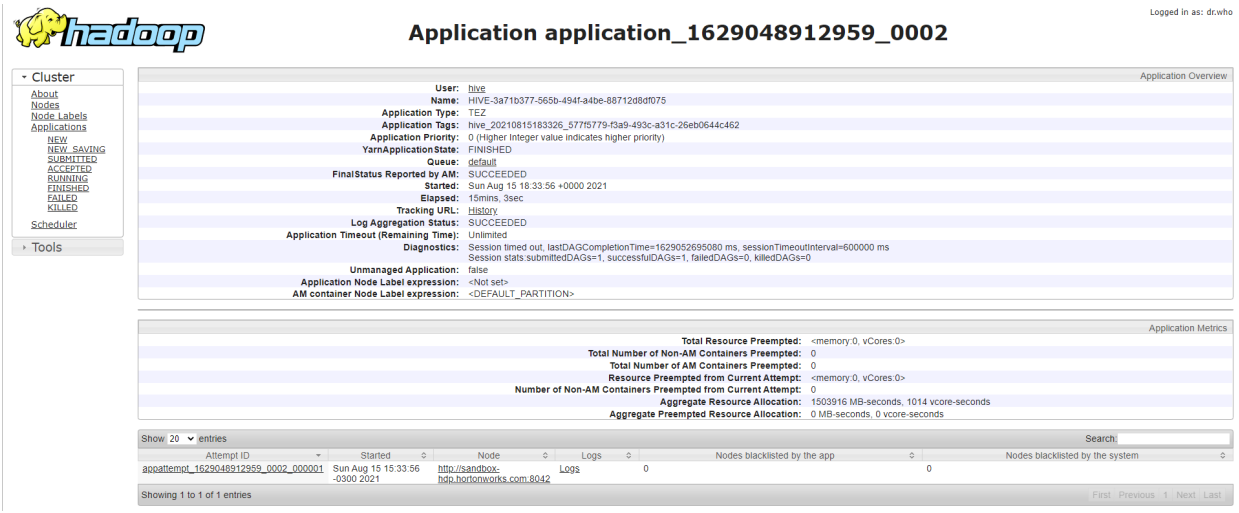

Figura 4.22: Métrica de *Performance* (carga AX_CLIENTE) do Projeto 2. (Fonte: Autor)

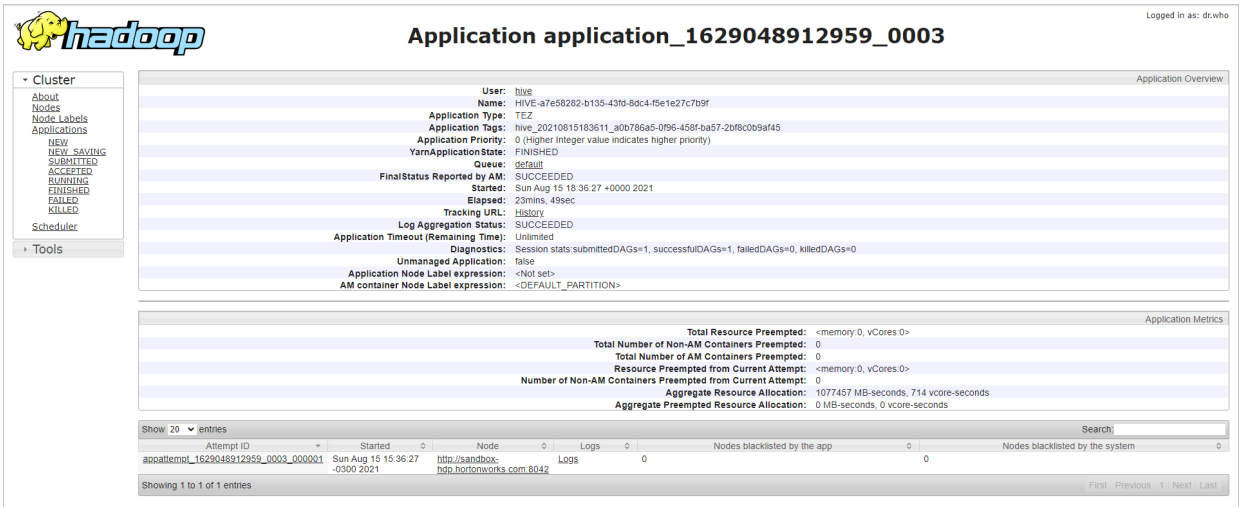

Figura 4.23: Métrica de *Performance* (carga TB_INDICIOS_TRANSFERENCIA) do Projeto 2. (Fonte: Autor)

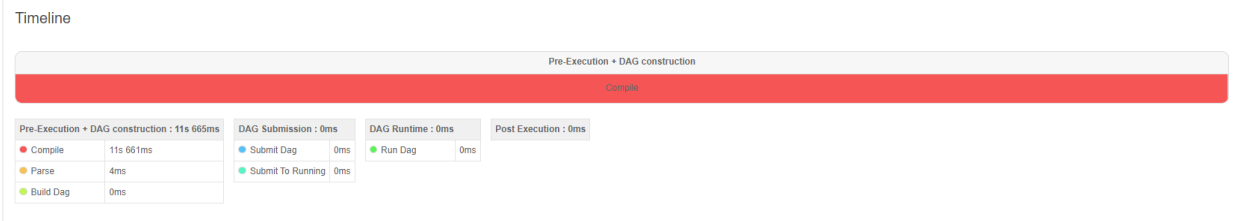

Figura 4.24: Métrica de *Performance* (consulta AX_CLIENTE) do Projeto 2. (Fonte: Autor)

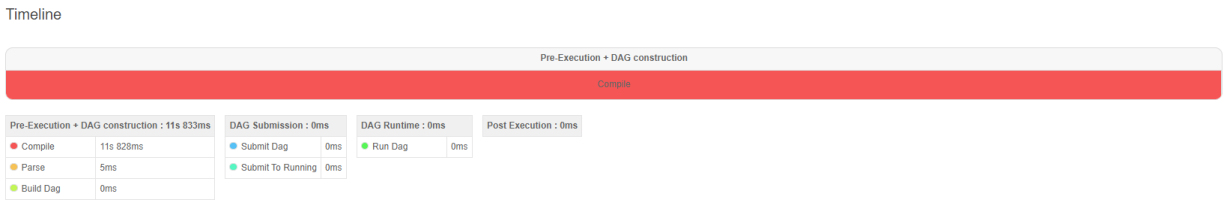

Figura 4.25: Métrica de *Performance* (consulta TB_INDICIOS_TRANSFERENCIA) do Projeto 2. (Fonte: Autor)

As Tabelas deste projeto 2 levaram 15 minutos e 3 segundos e 23 minutos e 49 segundos para serem criadas e carregadas, totalizando um período de **38 minutos e 52 segundos** de processamento total de carga. Com relação a consulta, a tabela de cliente concluiu uma consulta completa (*"full table scan"*) em 11 segundos e 661 milésimos de segundo, enquanto a tabela de indícios performou a consulta em 11 segundos e 828 milésimos de segundo, totalizando **23 segundos e 489 milésimos de segundos** de consultas.

Uma possível explicação para este efeito na performance seria que a minimização de dados pode ter gerado nos motores de busca do *Big Data* (preparados para paralelização e consulta de grandes volumes de dados) um superprocessamento desnecessário. Isso ocorre porque a paralelização possui um valor mínimo de volume de informações consultadas para gerar melhoria da *performance* de consulta, o que provavelmente foi violado ao realizar a citada redução da base. A última métrica aferida foi a de privacidade, conforme já calculada no primeiro projeto, a qual representou uma melhoria considerável (mais de 16.104.096%) conforme apresentado na Tabela 4.12.

Tabela 4.12: Cálculo da métrica de privacidade CM1 (Projeto 2). (Fonte: Autor)

Relação	**CM1**
AX_CLIENTE x TB_INDICIOS_TRANSFERENCIA	12,883277505797474877608863694923 E-4

A melhoria da métrica CM1 no Projeto 2 indica que o *framework* (principalmente pelo processo de qualidade de dados) tornou a relação entre os dados pessoais e os demais dados menos trivial, o que contribui para a proteção da privacidade. Por fim, o *checklist* proposto como critério de adequação à LGPD foi preenchido conforme a Tabela 4.13.

Tabela 4.13: Lista de checagem dos critérios de tratamento de dados para prevenção à fraude, em conformidade com a LGPD. (Projeto 2). (Fonte: Autor)

	Questão	**X**	**V**
CK1	O propósito do tratamento do dado foi concebido?		X
CK2	O propósito do tratamento do dado foi documentado?		X
CK3	O propósito do tratamento do dado está acessível?		X
CK4	A linhagem (origem e destino) do dado é documentada?		X
CK5	O dado é classificado (dado pessoal, dado sensível)?		X
CK6	A classificação do dado está acessível?		X

É importante perceber que, se comparado ao *checklist* resultante do Projeto 1, o Projeto 2 teve melhor desempenho (todos os itens de checagem foram validados). Isto porque no primeiro projeto, apesar de o propósito do tratamento ter sido concebido, não havia artefatos de formalização e comunicação deste propósito, uma vez que as atividades da disciplina de governança de dados e privacidade não foram consideradas. De semelhante forma, a linhagem e a classificação dos dados, apesar de definidas, não são formalizadas no Projeto 1, de forma inversa ao que ocorreu no Projeto 2. Isto comprova a importância das disciplinas supracitadas (Governança de Dados e Privacidade) para a gestão e a conformidade dos dados aos padrões e normas.

De forma comparativa, percebemos que todas as métricas calculadas (com exceção das métricas de *performance*) favoreceram o uso do *framework* que, através do método escolhido de validação, comprovou trazer ganhos tanto na prevenção a fraude através da melhoria da qualidade de dados e do processo de governança de dados em geral, como na privacidade do dado, explicitada na melhoria significativa da métrica CM1 proposta.

Esta última validação responde a subquestão RQ.5, a qual questiona se o *framework* consolidado a partir dessas boas práticas recomendadas pode otimizar os aspectos de proteção à privacidade e prevenção a fraude de processos de ingestão de dados em *Big Data*, uma vez que o *framework* proposto, de fato, pôde contribuir com projetos de ingestão de dados em ambientes *Big Data*. Os melhores resultados obtidos com as métricas de qualidade de dados indicam um uso otimizado dos dados quanto à prevenção a fraude, enquanto a métrica de privacidade e o *checklist* de critérios de conformidade à LGPD apontam para uma melhor proteção da privacidade. Entretanto, este processo de validação ainda carece de aplicação ampla, em outros cenários e contextos, a fim de que seja possível generalizar a obtenção dos benefícios verificados na utilização *framework* a qualquer instituição e projeto de dados em *Big Data*.

Com tais resultados, podemos responder à questão problema desta pesquisa, sendo ela: "Quais as melhores práticas relacionadas à Governança de Dados, Privacidade de Dados e Segurança da Informação, recomendadas pelos profissionais de tecnologias da informação (que lidam com os dados pessoais) e pela literatura pertinente, podem ser elencadas e consolidadas para o tratamento de dados pessoais destinado à otimização da garantia da prevenção a fraude, no contexto de *Big Data*"?

Entendemos então que, as atividades que compõem o *framework*, dispostas como o sugerido incluindo os seus artefatos (obviamente adaptando-se as sugestões à realidade de cada projeto), configuram este conjunto de melhores práticas, nas disciplinas basilares (Governança de Dados, Privacidade de Dados e Segurança da Informação), tendo indicação tanto dos profissionais que lidam com dados pessoais (através dos questionários aplicados), como da literatura pertinente (através da revisão de literatura descrita). Tais atividades, consolidadas na forma do *framework*, quando aplicadas, indicaram melhorias tanto no aspecto de prevenção a fraude, como na proteção da privacidade de dados (verificadas através das métricas aplicadas no estudo de caso).

4.5 AMEAÇAS PARA VALIDAÇÃO

O primeiro aspecto que pode representar uma ameaça à validação deste trabalho é o de que as atividades levantadas aqui fizeram parte de uma revisão bibliográfica. Isto significa que, apesar de ter sido extensa a pesquisa das atividades referenciadas pela literatura, não houve a pretensão de se levantar todas as atividades disponíveis na vasta literatura, através de uma revisão sistemática. Isto ocorreu, pois, o objetivo era o levantamento das principais (não todas) boas práticas, uma vez que elas seriam testadas quanto a aplicação prática através dos questionários. Este risco, entretanto, é minimizado pelo fato de que, na opinião dos entrevistados através dos questionários, todas as atividades foram abarcadas, ou seja, não houve sugestões significativas de atividades fora do rol apresentado.

Um segundo ponto de ameaça à validação deste trabalho, comum a trabalho empíricos envolvendo questionários, é representado pelo possível enviesamento das respostas obtidas na aplicação dos questionários, que foi voluntária e distribuída por meio eletrônico (redes sociais, canal de comunicação interno da empresa estudo de caso, fóruns e outros). Estes questionários eram longos, e podem ter resultado em um desinteresse de resposta por parte dos entrevistados, uma vez que apresentava uma série de atividades e ferramentas talvez desconhecidas ou nunca aplicadas por estes profissionais.

Como já discutido, apesar de as respostas na empresa estudo de caso representarem 10% (aproximadamente) do efetivo do quadro de profissionais da tecnologia da informação na empresa estudo de caso, houve representatividade da maioria das unidades organizacionais e de diversas equipes de trabalho, podendo-se afirmar que a amostra apesar de pequena esteve bem estratificada e representa, ao menos em diversidade de equipes, o total da população. Isto posto, a validação do *framework* apresentada fica restrita a instituições com o perfil da empresa estudo de caso, isto é, empresas do setor financeiro e público, que envolvem em suas atividades dados pessoais e de prevenção a fraude. Como exposto nos resultados obtidos no questionário, a percepção e o contexto cultural dos profissionais da empresa estudo de caso não representam em sua totalidade a opinião e contexto cultural dos demais profissionais da área.

5 CONCLUSÃO

O problema da privacidade de dados e da fraude de sistemas de informação é latente e precisa ser uma preocupação constante dos profissionais de tecnologias da informação que lidam com dados pessoais. Além disso, o grande volume de dados envolvido em projetos de *Big Data*, ainda em expansão, contribui para o aumento da entropia do ambiente, tornando a tarefa de gerenciar os dados destes projetos um verdadeiro desafio. As normas e legislações, não apenas no Brasil, têm depositado em algumas ferramentas, tais como a anonimização, uma expectativa irreal de garantia da proteção da privacidade de dados. Da mesma forma, a utilização de métodos mais rígidos, como a completa exclusão de dados, corrobora para a perda de utilidade de dados, causando o aumento do custo, e sem contribuir com a extração de valor dos mesmos. Portanto, é necessária a consolidação de um guia, o qual seja construído com o propósito de conciliar todas essas preocupações concernentes a esse contexto, que se apresenta aparentemente caótico. Principalmente no contexto normativo que vivemos, onde a Lei Geral de Proteção a Dados Pessoais está em vigor sem que seja observada de maneira extensiva a mesma evolução das disciplinas de privacidade de dados, na opinião dos profissionais de tecnologias da informação e do mercado em geral.

Este guia, que compila as principais boas práticas das disciplinas de Governança, Privacidade e Segurança da Informação, pode ser uma referência não só para o *compliance* a LGPD, mas um auxílio metodológico para projetos de ingestão de dados em *Big Data*, a fim de evitar a ingerência de dados e o surgimento de *data swamps*, conforme apresentado no capítulo 2 (Referencial Teórico). Neste trabalho, exploramos este auxílio para a análise de fraude, embora haja a expectativa de aplicabilidade em muitos outros contextos.

O guia proposto, além de consolidar uma série de preocupações observadas no mercado, foi construído sobre uma base prática já consolidada em outros contextos, o que tem o objetivo de tornar intuitiva a aplicação do mesmo, reaproveitando conhecimentos já consolidados como boas práticas em outras disciplinas. Enfatizamos o caráter prático, como sugestões de atividades replicáveis, a fim de elencar sugestões a níveis operacional, tático e estratégico para as organizações.

Como observado através dos métodos escolhidos para a validação deste trabalho, a aplicação das atividades sugeridas em nosso *framework* pode contribuir de maneira geral para a gerência de dados, economia de recursos, proteção da privacidade, *compliance* a LGPD, além de fortalecer a extração do valor de dados no processo de ingestão, uma vez que foram observadas melhorias nítidas de qualidade de dados.

Dada então a questão central desta pesquisa, qual seja: “quais as melhores práticas relacionadas à Governança de Dados, Privacidade de Dados e Segurança da Informação, recomendadas pelos profissionais de tecnologias da informação (que lidam com os dados pessoais) e pela literatura pertinente, podem ser elencadas e consolidadas para o tratamento de dados pessoais destinado à otimização da garantia da prevenção a fraude, no contexto de *Big Data*”, entendemos que o *fra-*

mework proposto, desenvolvido sobre as disciplinas basilares (Governança de Dados, Privacidade de Dados e Segurança da Informação), baseado no levantamento de uma revisão de literatura validada pelos profissionais de TI, consolida as melhores práticas para a prevenção a fraude e proteção da privacidade de dados, contribuindo para o *compliance* à LGPD.

O estudo de caso aplicado também corrobora para a confirmação da melhoria dos aspectos de prevenção a fraude e proteção da privacidade de dados, comprovando a eficácia do *framework* para projetos de ingestão de dados em *Big Data* no contexto de instituições financeiras de caráter público.

Entretanto, como esta pesquisa se configura como um trabalho inicial neste campo de pesquisa, principalmente no contexto brasileiro de proteção à privacidade, é possível vislumbrar algumas melhorias e contribuições a este trabalho, discutidas a seguir.

5.1 TRABALHOS FUTUROS

Como discutido, a validação da aplicabilidade das atividades levantadas no *framework* proposto ficou restrita às empresas que se enquadram no contexto da empresa estudo de caso, a saber, empresas públicas e do setor financeiro. Há uma expectativa que tais boas práticas possam ser aplicáveis fora deste contexto, o que demanda validação por parte da academia.

Também entendemos que um conjunto maior de profissionais da área de tecnologia pode refinar o entendimento sobre quais atividades devem ser incluídas ou excluídas do rol de práticas constantes do *framework*, o que poderia ser averiguado em um momento futuro.

Um estudo aprofundado da métrica aqui utilizada para validação da privacidade da informação (adaptação da métrica proposta por Domingo-Ferrer [51]) pode fortalecer ainda mais o arsenal disponível para a proteção da privacidade de dados. Por fim, um estudo específico e aprofundado debruçado sobre cada atividade proposta individualmente poderia ser desenvolvido, a fim de elencar quais atividades têm mais impacto nos aspectos de proteção da privacidade e prevenção a fraude.

REFERÊNCIAS BIBLIOGRÁFICAS

1 AKEEL, F., FATHABADI, A. S., PACI, F., GRAVELL, A., AND WILLS, G. Formal modelling of data integration systems security policies. *Data Science and Engineering 1*, 3 (2016), 139–148.

2 ÅKERLUND, A., AND GROSSE, C. Integration of data envelopment analysis in business process models: A novel approach to measure information security. In *International Conference on Information Systems Security and Privacy (ICISSP), February 25-27, 2020, in Valletta, Malta* (2020), SciTePress, pp. 281–288.

3 AL HAMID, H. A., RAHMAN, S. M. M., HOSSAIN, M. S., ALMOGREN, A., AND ALAMRI, A. A security model for preserving the privacy of medical big data in a healthcare cloud using a fog computing facility with pairing-based cryptography. *IEEE Access 5* (2017), 22313–22328.

4 ALASHOOR, T., HAN, S., AND JOSEPH, R. C. Familiarity with big data, privacy concerns, and self-disclosure accuracy in social networking websites: An apco model. *Communications of the Association for Information Systems 41*, 1 (2017), 4.

5 ALBAUM, G. The likert scale revisited. *Market Research Society. Journal. 39*, 2 (1997), 1–21.

6 ALI-ELDIN, A., ZUIDERWIJK, A., AND JANSSEN, M. A privacy risk assessment model for open data. In *International Symposium on Business Modeling and Software Design* (2017), Springer, pp. 186–201.

7 ALMASI, M. M., SIDDIQUI, T. R., MOHAMMED, N., AND HEMMATI, H. The risk-utility tradeoff for data privacy models. In *2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (2016), IEEE, pp. 1–5.

8 ALSHAMMARI, M., AND SIMPSON, A. A uml profile for privacy-aware data lifecycle models. In *Computer Security*. Springer, 2017, pp. 189–209.

9 ANTIGNAC, T., SCANDARIATO, R., AND SCHNEIDER, G. A privacy-aware conceptual model for handling personal data. In *International Symposium on Leveraging Applications of Formal Methods* (2016), Springer, pp. 942–957.

10 ASIJA, R., AND NALLUSAMY, R. Healthcare saas based on a data model with built-in security and privacy. *International Journal of Cloud Applications and Computing (IJCAC) 6*, 3 (2016), 1–14.

11 ASKI, V. J., GUPTA, S., AND SARKAR, B. An authentication-centric multi-layered security model for data security in iot-enabled biomedical applications. In *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)* (2019), IEEE, pp. 957–960.

12 AYALA-RIVERA, V., AND PASQUALE, L. The grace period has ended: An approach to operationalize gdpr requirements. In *2018 IEEE 26th International Requirements Engineering Conference (RE)* (2018), IEEE, pp. 136–146.

13 BALOGH, Z., AND TURČÁNI, M. Modeling of data security in cloud computing. In *2016 Annual IEEE Systems Conference (SysCon)* (2016), IEEE, pp. 1–6.

14 BARBIERI, C. *BI2–Business intelligence: Modelagem & Qualidade*. Elsevier Editora, 2011.

15 BARBIERI, C. *Governança de Dados: Práticas, conceitos e novos caminhos*. Alta Books, 2020.

16 BECKER, R. A., VOLINSKY, C., AND WILKS, A. R. Fraud detection in telecommunications: History and lessons learned. *Technometrics 52*, 1 (2010), 20–33.

17 BEIERLE, F., TRAN, V. T., ALLEMAND, M., NEFF, P., SCHLEE, W., PROBST, T., PRYSS, R., AND ZIMMERMANN, J. Context data categories and privacy model for mobile data collection apps. *Procedia computer science 134* (2018), 18–25.

18 BEWONG, M., LIU, J., LIU, L., LI, J., AND CHOO, K.-K. R. A relative privacy model for effective privacy preservation in transactional data. *Concurrency and Computation: Practice and Experience 31*, 23 (2019), e4923.

19 BHALADHARE, P. R., AND JINWALA, D. C. Novel approaches for privacy preserving data mining in k-anonymity model. *J. Inf. Sci. Eng. 32*, 1 (2016), 63–78.

20 BLANCO, C., GARCÍA-SAIZ, D., PERAL, J., MATÉ, A., OLIVER, A., AND FERNÁNDEZ-MEDINA, E. How the conceptual modelling improves the security on document databases. In *International Conference on Conceptual Modeling* (2018), Springer, pp. 497–504.

21 BOLTON, R. J., HAND, D. J., ET AL. Statistical fraud detection: A review. *Statistical science 17*, 3 (2002), 235–255.

22 BRACKENBURY, W., LIU, R., MONDAL, M., ELMORE, A. J., UR, B., CHARD, K., AND FRANKLIN, M. J. Draining the data swamp: A similarity-based approach. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics* (2018), pp. 1–7.

23 BRACKETT, M., AND EARLEY, P. S. The DAMA guide to the data management body of knowledge (DAMA-DMBOK guide). *Estados Unidos: Technics Publications* (2009).

24 BRASHER, E. A. Addressing the failure of anonymization: Guidance from the european union's general data protection regulation. *Colum. Bus. L. Rev.* (2018), 209.

25 BURMEISTER, F., DREWS, P., AND SCHIRMER, I. A privacy-driven enterprise architecture meta-model for supporting compliance with the general data protection regulation. *The Hawai'i International Conference on System Sciences* (2019).

26 CAO, Y. Rigorous and flexible privacy models for utilizing personal spatiotemporal data. In *PhD@ VLDB* (2016).

27 CARMICHAEL, P., AND MORISSET, C. Learning decision trees from synthetic data models for human security behaviour. In *International Conference on Software Engineering and Formal Methods* (2017), Springer, pp. 56–71.

28 CARVALHO, A. P., CANEDO, E. D., CARVALHO, F. P., AND CARVALHO, P. H. P. Anonymisation and compliance to protection data: Impacts and challenges into big data. In *ICEIS (1)* (2020), pp. 31–41.

29 CASANOVAS, P., DE KOKER, L., MENDELSON, D., AND WATTS, D. Regulation of big data: Perspectives on strategy, policy, law and privacy. *Health and Technology 7*, 4 (2017), 335–349.

30 CHAKRABORTY, S., AND TRIPATHY, B. Privacy preservation in relational data through l-diversity and recursive (c, l) diversity anonymisation. *International Journal of Mathematical Modelling and Numerical Optimisation 7*, 3-4 (2016), 338–362.

31 CHEN, A., LU, G., XING, H., XIE, Y., AND YUAN, S. Dynamic and semantic-aware access-control model for privacy preservation in multiple data center environments. *International Journal of Distributed Sensor Networks 16*, 5 (2020), 1550147720921778.

32 CHEN, X. A security integration model for private data of intelligent mobile communication based on edge computing. *Computer Communications 162* (2020), 204–211.

33 CHENTHARA, S., AHMED, K., AND WHITTAKER, F. Privacy-preserving data sharing using multi-layer access control model in electronic health environment. *EAI Endorsed Transactions on Scalable Information Systems 6*, 22 (2019).

34 CHOW, J., PFAFF, B., GARFINKEL, T., AND ROSENBLUM, M. Shredding your garbage: Reducing data lifetime through secure deallocation. In *USENIX Security Symposium* (2005), pp. 22–22.

35 CONTI, M., DI PIETRO, R., AND MARCONI, L. Privacy for lbss: On using a footprint model to face the enemy. In *Advanced Research in Data Privacy*. Springer, 2015, pp. 169–195.

36 CUPPENS, F., AND CUPPENS-BOULAHIA, N. Stratification based model for security policy with exceptions and contraries to duty. In *From Database to Cyber Security*. Springer, 2018, pp. 78–103.

37 CUZZOCREA, A., AND MASTROIANNI, C. A general overview of privacy-preserving big data management and analytics models, methods and techniques in specific domains: Static and dynamic distributed environments. In *2018 IEEE International Conference on Big Data (Big Data)* (2018), IEEE, pp. 5093–5100.

38 DA REPÚBLICA, P. Lei de crimes contra o sistema financeiro nacional (sfn). *Secretaria-Geral, Último acesso em 24 de Junho de 2021* (1986). <http://www.planalto.gov.br/ccivil_03/leis/l7492.htm>.

39 DA REPÚBLICA, P. Lei de crimes de "lavagem" ou ocultação de bens, direitos e valores; a prevenção da utilização do sistema financeiro para os ilícitos previstos nesta lei; cria o conselho de controle de atividades financeiras - coaf. *Secretaria-Geral, Último acesso em 24 de Junho de 2021* (1986). <http://www.planalto.gov.br/ccivil_03/leis/l9613.htm>.

40 DA REPÚBLICA, P. Lei geral de acesso a informação (lai). *Secretaria-Geral, Último acesso em 04 de Fevereiro de 2020* (2011). <http://www.mpf.mp.br/atuacao-tematica/sci/normas-e-legislacao/legislacao/legislacao-em-ingles/law-12.527>.

41 DA REPÚBLICA, P. Marco civil da internet. *Secretaria-Geral, Último acesso em 04 de Fevereiro de 2020* (2014). <https://www.cgi.br/pagina/marco-civil-law-of-the-internet-in-brazil/180>.

42 DA REPÚBLICA, P. Lei geral de proteção de dados pessoais (lgpd). *Secretaria-Geral, Último acesso em 19 de Novembro de 2019* (2018). <https://www.pnm.adv.br/wp-content/uploads/2018/08/Brazilian-General-Data-Protection-Law.pdf>.

43 DALLA FAVERA, R. B., AND DA SILVA, R. L. Cibersegurança na união europeia e no mercosul: Big data e surveillance versus privacidade e proteção de dados na internet. *Revista de Direito, Governança e Novas Tecnologias 2*, 2 (2016), 112–134.

44 DARWISH, S., NOURETDINOV, I., AND WOLTHUSEN, S. Modelling the privacy impact of external knowledge for sensor data in the industrial internet of things. In *Security and Privacy Trends in the Industrial Internet of Things*. Springer, 2019, pp. 223–243.

45 DASGUPTA, A., KOSARA, R., AND CHEN, M. Guess me if you can: A visual uncertainty model for transparent evaluation of disclosure risks in privacy-preserving data visualization. In *2019 IEEE Symposium on Visualization for Cyber Security (VizSec)* (2019), IEEE, pp. 1–10.

46 DEMING, W. E. The new economics for industry. *Government, Education, Massachusetts Institute of Technology, Cambridge, MA 1* (1993), 235.

47 DIEZ, F. P., VASU, A. C., TOUCEDA, D. S., AND CÁMARA, J. M. S. Modeling xacml security policies using graph databases. *IT Professional 19*, 6 (2017), 52–57.

48 DODERO, J. M., RODRIGUEZ-GARCIA, M., RUIZ-RUBE, I., AND PALOMO-DUARTE, M. Privacy-preserving reengineering of model-view-controller application architectures using linked data. *Journal of Web Engineering 18*, 7 (2019), 695–728.

49 DOMINGO-FERRER, J. Big data anonymization requirements vs privacy models. In *ICETE (2)* (2018), pp. 471–478.

50 DOMINGO-FERRER, J. Personal big data, gdpr and anonymization. In *International Conference on Flexible Query Answering Systems* (Am Thalbach 22, 4600 - Thalheim bei Wels - AUSTRIA, 2019), Springer, pp. 7–10.

51 DOMINGO-FERRER, J., MURALIDHAR, K., AND BRAS-AMORÓS, M. General confidentiality and utility metrics for privacy-preserving data publishing based on the permutation model. *IEEE Transactions on Dependable and Secure Computing* (2020).

52 DONG, X., GUO, Y., LI, F., DONG, L., AND KHAN, A. Combination model of heterogeneous data for security measurement. *JUCS-Journal of Universal Computer Science 25* (2019), 270.

53 DRESCH, A. Design science e design science research como artefatos metodológicos para engenharia de produção. *Brasil* (2013).

54 DRUCKER, P. F. O advento da nova organização. *Harvard Business Review (Org). Gestão do conhecimento 9* (2000), 9–26.

55 DRUGAN, M. M. A bayesian model for anomaly detection in sql databases for security systems. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (2016), IEEE, pp. 1–8.

56 DU, M., WANG, K., LIU, X., GUO, S., AND ZHANG, Y. A differential privacy-based query model for sustainable fog data centers. *IEEE Transactions on Sustainable computing 4*, 2 (2017), 145–155.

57 DU, M., WANG, K., XIA, Z., AND ZHANG, Y. Differential privacy preserving of training model in wireless big data with edge computing. *IEEE transactions on big data 6*, 2 (2018), 283–295.

58 DUAN, Y., SUN, X., CHE, H., CAO, C., LI, Z., AND YANG, X. Modeling data, information and knowledge for security protection of hybrid iot and edge resources. *IEEE Access 7* (2019), 99161–99176.

59 EDWARDS, B., HOFMEYR, S., FORREST, S., AND VAN EETEN, M. Analyzing and modeling longitudinal security data: Promise and pitfalls. In *Proceedings of the 31st Annual Computer Security Applications Conference* (2015), pp. 391–400.

60 ELSHEKEIL, S. A., AND LAOYOOKHONG, S. Gdpr privacy by design. *Ph. D. dissertation, Master's thesis* (2017), 198.

61 ERLINGSSON, U. Data-driven software security: Models and methods. In *2016 IEEE 29th Computer Security Foundations Symposium (CSF)* (2016), IEEE, pp. 9–15.

62 EXPERIAN, S. Pesquisa global de identidade e fraude 2021.

63 FERNANDES, A. A., AND DE ABREU, V. F. *Implantando a Governança de TI-: Da estratégia à Gestão de Processos e Serviços*. Brasport, 2014.

64 FOTHERGILL, D. B., KNIGHT, W., STAHL, B. C., AND ULNICANE, I. Responsible data governance of neuroscience big data. *Frontiers in neuroinformatics 13* (2019), 28.

65 FU, W. Mass internet of things data security exchange model under heterogeneous environment. *International Journal of Embedded Systems 12*, 4 (2020), 484–490.

66 GAO, C., AND IWANE, N. A social network model for big data privacy preserving and accountability assurance. In *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)* (2015), IEEE, pp. 19–22.

67 GAO, H., XU, H., ZHANG, L., AND ZHOU, X. A differential game model for data utility and privacy-preserving in mobile crowdsensing. *IEEE Access 7* (2019), 128526–128533.

68 GHAZINOUR, K., AND ALBALAWI, T. A usability study on the privacy policy visualization model. In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (2016), IEEE, pp. 578–585.

69 GIRGIS, A. M., DATA, D., DIGGAVI, S., KAIROUZ, P., AND SURESH, A. T. Shuffled model of federated learning: Privacy, communication and accuracy trade-offs. *arXiv preprint arXiv:2008.07180* (2020).

70 GOODRICH, M. T., AND TAMASSIA, R. *Introdução à segurança de computadores*. Bookman, 2013.

71 GOPE, P., AND AMIN, R. A novel reference security model with the situation based access policy for accessing ephr data. *Journal of medical systems 40*, 11 (2016), 1–14.

72 GRACE, P., BURNS, D., NEUMANN, G., PICKERING, B., MELAS, P., AND SURRIDGE, M. Identifying privacy risks in distributed data services: A model-driven approach. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)* (2018), IEEE, pp. 1513–1518.

73 GRUSHO, A., TIMONINA, E., AND SHORGIN, S. Security models based on stochastic meta data. In *International Conference on Analytical and Computational Methods in Probability Theory* (2017), Springer, pp. 388–400.

74 GUCLU, M., BAKIR, C., AND HAKKOYMAZ, V. A new scalable and expandable access control model for distributed database systems in data security. *Scientific Programming 2020* (2020).

75 GUO, Z., AND XU, L. Research of security structure model for web application systems based on the relational database. *International Journal of Security and Networks 10*, 4 (2015), 207–213.

76 HAHN, S.-J., AND LEE, J. Privacy-preserving federated bayesian learning of a generative model for imbalanced classification of clinical data. *arXiv preprint arXiv:1910.08489* (2019).

77 HAJDER, M., KOLBUSZ, J., HAJDER, P., NYCZ, M., AND LIPUT, M. Data security platform model in networked medical it systems based on statistical classifiers and ann. *Procedia Computer Science 176* (2020), 3682–3691.

78 HAMADEH, H., AND TYAGI, A. Privacy preserving data provenance model based on puf for secure internet of things. In *2019 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)* (2019), IEEE, pp. 189–194.

79 HINTZE, M., AND EL EMAM, K. Comparing the benefits of pseudonymisation and anonymisation under the gdpr. *Journal of Data Protection & Privacy 2*, 2 (2018), 145–158.

80 HIRANI, M., HALGAMUGE, M. N., AND HANG, P. D. T. Data security models developed by blockchain technology for different business domains. In *2019 11th International Conference on Knowledge and Systems Engineering (KSE)* (2019), IEEE, pp. 1–10.

81 HU, X., JIANG, R., SHI, M., AND SHANG, J. A privacy protection model for health care big data based on trust evaluation access control in cloud service environment. *Journal of Intelligent & Fuzzy Systems 38*, 3 (2020), 3167–3178.

82 INSTITUTE, T. T. D. W. Tdwi advanced analytics maturity model guide. *TDWI research* (2019). Disponível em: <https://tdwi.org/pages/assessment/adv-all-tdwi-advanced-analytics-maturity-model.aspx>, Último acesso em 04 de Fevereiro de 2020.

83 ISACA. *Cobit 5 Enabling Information*. ISACA, 2013.

84 ISO CENTRAL SECRETARY. Information technology — security techniques — code of practice for information security controls. Standard ISO/IEC TR 27002:2013, International Organization for Standardization, Geneva, CH, 2013.

85 ISO CENTRAL SECRETARY. Information technology — security techniques — information security management systems — requirements. Standard ISO/IEC TR 27001:2013, International Organization for Standardization, Geneva, CH, 2013.

86 ISO CENTRAL SECRETARY. Security techniques — extension to iso/iec 27001 and iso/iec 27002 for privacy information management — requirements and guidelines. Standard ISO/IEC TR 27701:2019, International Organization for Standardization, Geneva, CH, 2019.

87 IVANOVA, M., GROSSECK, G., AND HOLOTESCU, C. Researching data privacy models in elearning. In *2015 International Conference on Information Technology Based Higher Education and Training (ITHET)* (2015), IEEE, pp. 1–6.

88 JAIN, P., GYANCHANDANI, M., AND KHARE, N. Big data security and privacy: New proposed model of big data with secured mr layer. In *Advanced Computing and Systems for Security*. Springer, 2019, pp. 31–53.

89 JAIN, S. K., KESSWANI, N., AND AGARWAL, B. Security, privacy and trust: privacy preserving model for internet of things. *International Journal of Intelligent Information and Database Systems 13*, 2-4 (2020), 249–277.

90 JEGADEESWARI, S., DINADAYALAN, P., AND GNANAMBIGAI, N. A neural data security model: Ensure high confidentiality and security in cloud datastorage environment. In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2015), IEEE, pp. 400–406.

91 JYOTHEESWARI, P., AND JEYANTHI, N. Hybrid encryption model for managing the data security in medical internet of things. *International Journal of Internet Protocol Technology 13*, 1 (2020), 25–31.

92 KAKANAKOV, N., AND SHOPOV, M. Adaptive models for security and data protection in iot with cloud technologies. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2017), IEEE, pp. 1001–1004.

93 KANG, R., DABBISH, L., FRUCHTER, N., AND KIESLER, S. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)* (2015), pp. 39–52.

94 KANWAL, T., SHAUKAT, S. A. A., ANJUM, A., CHOO, K.-K. R., KHAN, A., AHMAD, N., AHMAD, M., KHAN, S. U., ET AL. Privacy-preserving model and generalization correlation attacks for 1: M data with multiple sensitive attributes. *Information Sciences 488* (2019), 238–256.

95 KHALOUFI, H., ABOUELMEHDI, K., BENI-HSSANE, A., AND SAADI, M. Security model for big healthcare data lifecycle. *Procedia Computer Science 141* (2018), 294–301.

96 KHAMAISEH, S., AND XU, D. Software security testing via misuse case modeling. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (2017), IEEE, pp. 534–541.

97 KHOKHAR, R. H., FUNG, B. C., IQBAL, F., ALHADIDI, D., AND BENTAHAR, J. Privacy-preserving data mashup model for trading person-specific information. *Electronic Commerce Research and Applications 17* (2016), 19–37.

98 KIM, S.-H., JUNG, C., AND LEE, Y.-J. An entropy-based analytic model for the privacy-preserving in open data. In *2016 IEEE International Conference on Big Data (Big Data)* (2016), IEEE, pp. 3676–3684.

99 KOOHANG, A., ANDERSON, J., NORD, J. H., AND PALISZKIEWICZ, J. Building an awareness-centered information security policy compliance model. *Industrial Management & Data Systems* (2019).

100 KUMAR, A., LIGATTI, J., AND TU, Y.-C. Query monitoring and analysis for database privacy-a security automata model approach. In *International Conference on Web Information Systems Engineering* (2015), Springer, pp. 458–472.

101 KUMAR, M., KUMAR, S., BUDHIRAJA, R., DAS, M., AND SINGH, S. Lightweight data security model for iot applications: a dynamic key approach. In *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (2016), IEEE, pp. 424–428.

102 KWAKYE, M. M., AND BARKER, K. Privacy-preservation in the integration and querying of multidimensional data models. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)* (2016), IEEE, pp. 255–263.

103 LEE, M. C., MITRA, R., LAZARIDIS, E., LAI, A. C., GOH, Y. K., AND YAP, W.-S. Statistical disclosure control for data privacy using sequence of generalised linear models. In *Australasian Conference on Information Security and Privacy* (2016), Springer, pp. 77–93.

104 LEE, M. C., MITRA, R., LAZARIDIS, E., LAI, A.-C., GOH, Y. K., AND YAP, W.-S. Data privacy preserving scheme using generalised linear models. *Computers & Security 69* (2017), 142–154.

105 LEE, S. U., ZHU, L., AND JEFFERY, R. Design choices for data governance in platform ecosystems: A contingency model. *arXiv preprint arXiv:1706.07560* (2017).

106 LI, J. A synthetic research on the multimedia data encryption based mobile computing security enhancement model and multi-channel mobile human computer interaction framework. *Multimedia Tools and Applications 76*, 16 (2017), 16963–16987.

107 LI, L. Predicting online invitation responses with a competing risk model using privacy-friendly social event data. *European Journal of Operational Research 270*, 2 (2018), 698–708.

108 LI, Q.-L., MA, F.-Q., AND MA, J.-Y. A stochastic model for file lifetime and security in data center networks. In *International Conference on Computational Social Networks* (2018), Springer, pp. 298–309.

109 LI, S., SHEN, H., AND SANG, Y. An efficient model and algorithm for privacy-preserving trajectory data publishing. In *International Conference on Parallel and Distributed Computing: Applications and Technologies* (2018), Springer, pp. 240–249.

110 LI, Y., OUYANG, J., MAO, B., MA, K., AND GUO, S. Data flow analysis on android platform with fragment lifecycle modeling and callbacks. *EAI Endorsed Transactions on Security and Safety 4*, 11 (2017), e2.

111 LI, Y., SONG, L., AND ZENG, Y. Research on information security and privacy protection model based on consumer behavior in big data environment. *Concurrency and Computation: Practice and Experience 31*, 10 (2019), e4881.

112 LIANG, L. Abnormal detection of electric security data based on scenario modeling. *Procedia computer science 139* (2018), 578–582.

113 LIN, C., WANG, P., SONG, H., ZHOU, Y., LIU, Q., AND WU, G. A differential privacy protection scheme for sensitive big data in body sensor networks. *Annals of Telecommunications 71*, 9-10 (2016), 465–475.

114 LIU, G., YANG, G., WANG, H., DAI, H., AND ZHOU, Q. Qsdb: An encrypted database model for privacy-preserving in cloud computing. *KSII Transactions on Internet and Information Systems (TIIS) 12*, 7 (2018), 3375–3400.

115 LIU, L., JU, J., AND FENG, Y. An extensible framework for collaborative e-governance platform workflow modeling using data flow analysis. *Information Technology for Development 23*, 3 (2017), 415–437.

116 LIU, S., YOU, S., YIN, H., LIN, Z., LIU, Y., YAO, W., AND SUNDARESH, L. Model-free data authentication for cyber security in power systems. *IEEE Transactions on Smart Grid 11*, 5 (2020), 4565–4568.

117 LIU, X., XIE, Q., AND WANG, L. Personalized extended (α, k)-anonymity model for privacy-preserving data publishing. *Concurrency and Computation: Practice and Experience 29*, 6 (2017), e3886.

118 LU, Y., ET AL. From data flows to privacy issues: a user-centric semantic model for representing and discovering privacy issues. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (2020), University of Hawaii.

119 MADAN, S., AND GOSWAMI, P. k-ddd measure and mapreduce based anonymity model for secured privacy-preserving big data publishing. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 27*, 02 (2019), 177–199.

120 MADAN, S., AND GOSWAMI, P. A privacy preservation model for big data in map-reduced framework based on k-anonymisation and swarm-based algorithms. *International Journal of Intelligent Engineering Informatics 8*, 1 (2020), 38–53.

121 MAZAK, A., WOLNY, S., AND WIMMER, M. On the need for data-based model-driven engineering. In *Security and Quality in Cyber-Physical Systems Engineering*. Springer, 2019, pp. 103–127.

122 MEHMOOD, A., NATGUNANATHAN, I., XIANG, Y., HUA, G., AND GUO, S. Protection of big data privacy. *IEEE access 4* (2016), 1821–1834.

123 MENDES, V. Empresas afirmam que anonimização de dados pode ser solucao mais rapida para adequacao a lgpd, 2019. (Último acesso em 20 de Abril de 2020).

124 MICROSTRATEGY. 2020 global state of enterprise analytics, 2020. (Date last accessed 20-April-2020).

125 MIRAMIRKHANI, N., STAROV, O., AND NIKIFORAKIS, N. Dial one for scam: A large-scale analysis of technical support scams. *arXiv preprint arXiv:1607.06891* (2016).

126 MOON, S. K., AND RAUT, R. D. Innovative data security model using forensic audio video steganography for improving hidden data security and robustness. *International Journal of Information and Computer Security 10*, 4 (2018), 374–397.

127 MORGADO, C., BAIOCO, G. B., BASSO, T., AND MORAES, R. A security model for access control in graph-oriented databases. In *2018 IEEE International Conference on Software Quality, Reliability and Security (QRS)* (2018), IEEE, pp. 135–142.

128 NOGUEIRA, D. M., MACIEL, C., VITERBO, J., AND VECCHIATO, D. A privacy-driven data management model for smart personal assistants. In *International Conference on Human Aspects of Information Security, Privacy, and Trust* (2017), Springer, pp. 722–738.

129 OH, H., PARK, S., LEE, G. M., CHOI, J. K., AND NOH, S. Competitive data trading model with privacy valuation for multiple stakeholders in iot data markets. *IEEE Internet of Things Journal 7*, 4 (2020), 3623–3639.

130 OLIFER, D., GORANIN, N., JANULEVICIUS, J., KACENIAUSKAS, A., AND CENYS, A. Improvement of security costs evaluation process by using data automatically captured from bpmn and epc models. In *International Conference on Business Process Management* (2017), Springer, pp. 698–709.

131 ORORBIA II, A. G., LINDER, F., AND SNOKE, J. Privacy protection for natural language: Neural generative models for synthetic text data. *arXiv preprint arXiv:1606.01151* (2016).

132 OTHMANE, L. B., CHEHRAZI, G., BODDEN, E., TSALOVSKI, P., AND BRUCKER, A. D. Time for addressing software security issues: Prediction models and impacting factors. *Data Science and Engineering 2*, 2 (2017), 107–124.

133 PAJA, E., DALPIAZ, F., AND GIORGINI, P. Modelling and reasoning about security requirements in socio-technical systems. *Data & Knowledge Engineering 98* (2015), 123–143.

134 PARFENOV, D., AND BOLODURINA, I. Investigation of the neural network model for security and quality of service for a multi-cloud system in virtual data center. In *2018 41st International Conference on Telecommunications and Signal Processing (TSP)* (2018), IEEE, pp. 1–5.

135 PFITZMANN, A., AND HANSEN, M. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. *Dresden, Alemanha* (2010).

136 PIAO, C., SHI, Y., ZHANG, Y., AND JIANG, X. Research on government data publishing based on differential privacy model. In *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)* (2017), IEEE, pp. 76–83.

137 PIRAS, L., AL-OBEIDALLAH, M. G., PRAITANO, A., TSOHOU, A., MOURATIDIS, H., CRESPO, B. G.-N., BERNARD, J. B., FIORANI, M., MAGKOS, E., SANZ, A. C., ET AL. Defend architecture: A privacy by design platform for gdpr compliance. In *International Conference on Trust and Privacy in Digital Business* (2019), Springer, pp. 78–93.

138 POLTAVTSEVA, M. A., AND KALININ, M. O. Conceptual data modeling using aggregates to ensure large-scale distributed data management systems security. In *International Symposium on Intelligent and Distributed Computing* (2019), Springer, pp. 41–47.

139 POLTAVTSEVA, M. A., AND KALININ, M. O. Modeling big data management systems in information security. *Automatic Control and Computer Sciences 53*, 8 (2019), 895–902.

140 POLTAVTSEVA, M. A., ZEGZHDA, D. P., AND KALININ, M. O. Big data management system security threat model. *Automatic control and computer sciences 53*, 8 (2019), 903–913.

141 POMARES-QUIMBAYA, A., SIERRA-MÚNERA, A., MENDOZA-MENDOZA, J., MALAVER-MORENO, J., CARVAJAL, H., AND MONCAYO, V. Anonylitics: From a small data to a big data anonymization system for analytical projects. In *21st International Conference on Enterprise Information Systems* (Avenida de S. Francisco Xavier, Lote 7 Cv. C 2900-616 Setubal - Portugal 38.524098, -8.905325, 2019), Springer, pp. 61–71.

142 POPOVICH, C., JEANSON, F., BEHAN, B., LEFAIVRE, S., AND SHUKLA, A. Big data, big responsibility! building best-practice privacy strategies into a large-scale neuroinformatics platform, 2017.

143 POTIGUARA CARVALHO, A., POTIGUARA CARVALHO, F., DIAS CANEDO, E., AND POTIGUARA CARVALHO, P. H. Big data, anonymisation and governance to personal data protection. In *The 21st Annual International Conference on Digital Government Research* (2020), pp. 185–195.

144 PRIEBE, T., AND MARKUS, S. Business information modeling: A methodology for data-intensive projects, data science and big data governance. In *2015 IEEE International Conference on Big Data (Big Data)* (2015), IEEE, pp. 2056–2065.

145 PULLONEN, P., TOM, J., MATULEVIČIUS, R., AND TOOTS, A. Privacy-enhanced bpmn: Enabling data privacy analysis in business processes models. *Software and Systems Modeling 18*, 6 (2019), 3235–3264.

146 RAMADAN, Q., STRÜBER, D., SALNITRI, M., JÜRJENS, J., RIEDIGER, V., AND STAAB, S. A semi-automated bpmn-based framework for detecting conflicts between security, data-minimization, and fairness requirements. *Software and Systems Modeling 19*, 5 (2020), 1191–1227.

147 RAMADAN, Q., STRÜBER, D., SALNITRI, M., RIEDIGER, V., AND JÜRJENS, J. Detecting conflicts between data-minimization and security requirements in business process models. In *European Conference on Modelling Foundations and Applications* (2018), Springer, pp. 179–198.

148 RATHI, N., GHOSH, S., IYENGAR, A., AND NAEIMI, H. Data privacy in non-volatile cache: Challenges, attack models and solutions. In *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)* (2016), IEEE, pp. 348–353.

149 RAZAQUE, A., AND RIZVI, S. S. Triangular data privacy-preserving model for authenticating all key stakeholders in a cloud environment. *Computers & Security 62* (2016), 328–347.

150 RÊGO, B. L. *Gestão e Governança de Dados: Promovendo dados como ativo de valor nas empresas*. Brasport, Rua Pardal Mallet, 23 - Tijuca, 20270-280 Rio de Janeiro-RJ, 2013.

151 REGULATION, G. D. P. Eu data protection rules. *European Commission, Último acesso em 9 de Outubro de 2019* (2018). <https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en>.

152 REN, X.-Y., ZHANG, P., AND ZHOU, Y.-Q. Distinct model on privacy protection of dynamic data publication. *Cluster Computing 22*, 6 (2019), 15127–15136.

153 RIVERA, S., LOARTE, N., RAYMUNDO, C., AND DOMÍNGUEZ-MATEOS, F. Data governance maturity model for micro financial organizations in peru. In *ICEIS (3)* (2017), pp. 203–214.

154 ROY, A., MISRA, A., AND BANERJEE, S. Discrete model for cloud computing: Analysis of data security and data loss. *arXiv preprint arXiv:1812.05445* (2018).

155 RUAN, N., WEI, Z., AND LIU, J. Cooperative fraud detection model with privacy-preserving in real cdr datasets. *IEEE Access 7* (2019), 115261–115272.

156 RYAN, M., AND BRINKLEY, M. Navigating privacy in a sea of change: new data protection regulations require thoughtful analysis and incorporation into the organization's governance model. *Internal Auditor 74*, 3 (2017), 61–63.

157 SANTOSO, I., SIAHAAN, I. S. R., ET AL. Privacy modelling of sensitive data in universal healthcare coverage in indonesia. In *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)* (2016), IEEE, pp. 1–6.

158 SAXENA, M. A., UBNARE, G., AND DUBEY, A. Virtual public cloud model in honeypot for data security: A new technique. In *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence* (2019), pp. 66–71.

159 SCHMIEDERS, E., METZGER, A., AND POHL, K. Runtime model-based privacy checks of big data cloud services. In *International Conference on Service-Oriented Computing* (2015), Springer, pp. 71–86.

160 SENARATH, A., AND ARACHCHILAGE, N. A. G. A data minimization model for embedding privacy into software systems. *Computers & Security 87* (2019), 101605.

161 SENARATH, A., GROBLER, M., AND ARACHCHILAGE, N. A. G. A model for system developers to measure the privacy risk of data. *arXiv preprint arXiv:1809.10884* (2018).

162 SENGE, P. M. *The fifth discipline: The art and practice of the learning organization*. Broadway Business, 2006.

163 SION, L., VAN LANDUYT, D., WUYTS, K., AND JOOSEN, W. Privacy risk assessment for data subject-aware threat modeling. In *2019 IEEE Security and Privacy Workshops (SPW)* (2019), IEEE, pp. 64–71.

164 SION, L., YSKOUT, K., VAN LANDUYT, D., AND JOOSEN, W. Solution-aware data flow diagrams for security threat modeling. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (2018), pp. 1425–1432.

165 SOHAIL, S. A., KRABBE, J., DE ALENCAR SILVA, P., AND BUKHSH, F. A. Privacy value modeling: A gateway to ethical big data handling. In *14th International Workshop on Value Modelling and Business Ontologies, VMBO 2020* (2020), CEUR, pp. 5–15.

166 SORIA-COMAS, J., AND DOMINGO-FERRER, J. Big data privacy: challenges to privacy principles and models. *Data Science and Engineering 1*, 1 (2016), 21–28.

167 STEINKE, M., AND HOMMEL, W. A data model for federated network and security management information exchange in inter-organizational it service infrastructures. In *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium* (2018), IEEE, pp. 1–2.

168 STOLL, M. A data privacy governance model: The integration of the general data protection regulation into standard based management systems. *International Journal of IT/Business Alignment and Governance (IJITBAG) 10*, 1 (2019), 74–93.

169 STUPNIKOV, S., MILOSLAVSKAYA, N., AND BUDZKO, V. Unification of graph data models for heterogeneous security information resources' integration. In *2015 3rd International Conference on Future Internet of Things and Cloud* (2015), IEEE, pp. 457–464.

170 SURRIDGE, M., MEACHAM, K., PAPAY, J., PHILLIPS, S. C., PICKERING, J. B., SHAFIEE, A., AND WILKINSON, T. Modelling compliance threats and security analysis of cross border health data exchange. In *International Conference on Model and Data Engineering* (2019), Springer, pp. 180–189.

171 SWART, I., IRWIN, B. V., AND GROBLER, M. M. Adaptation of the jdl model for multi-sensor national cyber security data fusion. In *National Security: Breakthroughs in Research and Practice*. IGI Global, 2019, pp. 92–107.

172 SYED, D., REFAAT, S. S., AND BOUHALI, O. Privacy preservation of data-driven models in smart grids using homomorphic encryption. *Information 11*, 7 (2020), 357.

173 TADHG NAGLE, T. C. R., AND SAMMON, D. Only 3% of companies' data meets basic quality standards, 2020. (Date last accessed 20-April-2020).

174 TAKAGI, S., TAKAHASHI, T., CAO, Y., AND YOSHIKAWA, M. P3gm: Private high-dimensional data release via privacy preserving phased generative model. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)* (2021), IEEE, pp. 169–180.

175 TAO, Y., LEI, Z., AND RUXIANG, P. Fine-grained big data security method based on zero trust model. In *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)* (2018), IEEE, pp. 1040–1045.

176 TEOH, T., NGUWI, Y., ELOVICI, Y., CHEUNG, N.-M., AND NG, W. Analyst intuition based hidden markov model on high speed, temporal cyber security big data. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (2017), IEEE, pp. 2080–2083.

177 THIRUMALAI, C. S., AND VISWANATHAN, P. Modelling a side channel resistant chan-pkc cryptomata for medical data security. *Multimedia Tools and Applications 78*, 18 (2019), 25977–25997.

178 THOMAS, G. How to use the dgi data governance framework to configure your program. *Data Governance Institute 17* (2009). Disponível em: <http://www.datagovernance.com/wp-content/uploads/2020/07/wp_how_to_use_the_dgi_data_governance_framework.pdf>. Ultimo acesso em 14-Abril-2021.

179 TRAN, N. H., NGUYEN-NGOC, T.-A., LE-KHAC, N.-A., KECHADI, M., ET AL. A security-aware access model for data-driven ehr system. *arXiv preprint arXiv:1908.10229* (2019).

180 TU, H., DOUPÉ, A., ZHAO, Z., AND AHN, G.-J. Sok: Everyone hates robocalls: A survey of techniques against telephone spam. In *2016 IEEE Symposium on Security and Privacy (SP)* (2016), IEEE, pp. 320–338.

181 VAISHNAVI, V., KUECHLER, W., AND PETTER, S. Design science research in information systems. *June 20* (2019), 30. Disponível em: <http://www.desrist.org/design-research-in-information-systems>, Último acesso em 24 de Agosto de 2020.

182 VENTURA, M., AND COELI, C. M. Para além da privacidade: direito à informação na saúde, proteção de dados pessoais e governança. *Cadernos de Saúde Pública 34* (2018), e00106818.

183 VERGINADIS, Y., PATINIOTAKIS, I., MENTZAS, G., VELOUDIS, S., AND PARASKAKIS, I. Data distribution and encryption modelling for paas-enabled cloud security. In *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)* (2016), IEEE, pp. 497–502.

184 VOLOCH, N., LEVY, P., ELMAKIES, M., AND GUDES, E. An access control model for data security in online social networks based on role and user credibility. In *International Symposium on Cyber Security Cryptography and Machine Learning* (2019), Springer, pp. 156–168.

185 WANG, D., ZHANG, H., GABOARDI, M., AND XU, J. Estimating smooth glm in non-interactive local differential privacy model with public unlabeled data. In *Algorithmic Learning Theory* (2021), PMLR, pp. 1207–1213.

186 WEI, H., HU, G.-Y., HAN, X., QIAO, P., ZHOU, Z., FENG, Z.-C., AND YIN, X.-J. A new brb model for cloud security-state prediction based on the large-scale monitoring data. *IEEE Access 6* (2017), 11907–11920.

187 WINTER, J. S., AND DAVIDSON, E. Investigating values in personal health data governance models. In *Americas Conference on Information Systems* (2017), vol. 2017.

188 XIAO, Y., SHEN, Y., LIU, J., XIONG, L., JIN, H., AND XU, X. Dphmm: customizable data release with differential privacy via hidden markov model. *arXiv preprint arXiv:1609.09172* (2016).

189 XIE, P., FAN, H.-J., FENG, T., YAN, Y., MA, G., AND HAN, X.-M. Adaptive access control model of vehicular network big data based on xacml and security risk. *Int. J. Netw. Secur. 22*, 2 (2020), 347–357.

190 YANG, J., AND XING, C. Personal data market optimization pricing model based on privacy level. *Information 10*, 4 (2019), 123.

191 YIN, R. K. *Estudo de Caso-: Planejamento e métodos*. Bookman editora, 2015.

192 ZHANG, C., LIANG, Y., YUAN, X., AND CHENG, L. Fdnas: Improving data privacy and model diversity in automl. *arXiv preprint arXiv:2011.03372* (2020).

APÊNDICES

I.1 GOVERNANÇA - IDENTIFICAÇÃO DA APLICABILIDADE DE ATIVIDADES DE PRIVACIDADE, SEGURANÇA E GOVERNANÇA DE DADOS NA PREVENÇÃO À FRAUDE (NOS TERMOS DA LGPD)

Identificação da aplicabilidade de atividades de Privacidade, Segurança e Governança de Dados na Prevenção à Fraude (nos termos da LGPD)

Bem vindo ao formulário para identificação das atividades e modelos listados na literatura pertinente e aplicáveis no mercado de desenvolvimento de softwares para prevenção à fraude e conformidade com a Lei Geral de Proteção de Dados Pessoais (Lei Federal nº 13.709 - LGPD). Abaixo, serão relacionados alguns modelos e atividades identificados em publicações acadêmicas e de mercado. Será solicitada sua opinião quanto ao conhecimento destes modelos e a contribuição efetiva deles à proteção da privacidade, segurança e governabilidade de dados pessoais. Não se preocupe se alguns dos modelos forem desconhecidos ou não tiver opinião formada sobre sua aplicabilidade, basta marcar o item identificado com essa opção (Não concordo nem discordo) ou deixá-lo em branco. A estimativa de tempo para preencher o questionário é de 30 minutos. Muito Obrigado!

Esta pesquisa é parte de um trabalho de mestrado do Programa de Pós-Graduação Profissional em Engenharia Elétrica da Universidade de Brasília - PPEE/UnB.

Contatos dos pesquisadores: ednacanedo@unb.br, artur.carvalho@aluno.unb.br, artur.carvalho@brb.com.br

DECLARAÇÃO DE CONSENTIMENTO INFORMADO

Ao responder a esta pesquisa, você permite que os pesquisadores obtenham, usem e divulguem as informações anônimas fornecidas conforme descrito abaixo.

CONDIÇÕES E ESTIPULAÇÕES

1. Eu entendo que todas as informações são confidenciais. Não serei identificado pessoalmente. Concordo em preencher o questionário para fins de pesquisa e que os dados derivados desta pesquisa anônima podem ser publicados em periódicos, conferências e postagens de blog.

2. Eu entendo que minha participação nesta pesquisa é totalmente voluntária e que a recusa em participar não implicará em nenhuma penalidade ou perda de benefícios. Se eu quiser, posso cancelar minha participação a qualquer momento. Também entendo que, se decidir participar, posso recusar-me a responder a qualquer pergunta para a qual não me sinta confortável em responder.

3. Entendo que posso entrar em contato com os pesquisadores se tiver alguma dúvida sobre a pesquisa. Estou ciente de que meu consentimento não me beneficiará diretamente. Também estou ciente de que o autor manterá os dados coletados em perpetuidade e poderá utilizar os dados para trabalhos acadêmicos

futuros.

4. Ao clicar no botão abaixo, eu livremente dou consentimento e reconheço meus direitos como um participante voluntário de pesquisa conforme descrito acima e dou consentimento aos pesquisadores para usarem minhas respostas na condução de pesquisas nas áreas mencionadas acima.

*Obrigatório

Informações Gerais

1) Qual o nome da instituição à qual você é vinculado(a)?

---

2) Como você classificaria a empresa onde trabalha? *

Marcar apenas uma oval.

◯ Somos uma micro-empresa, empresa familiar, pequeno negócio.

◯ Somos uma empresa de porte médio, com alguns poucos funcionários, mas sem funcionários específicos de desenvolvimento de software

◯ Somos uma empresa de porte médio e nossa empresa possui quadro de pessoal específico de desenvolvimento de sofware. Somos uma empresa de grande porte mas não possuímos participações em outras empresas (como filiais ou conglomerados)

◯ Somos uma empresa de grande porte e possuímos filiais.

◯ Somos um conglomerado sob uma marca.

◯ Minha empresa não se encaixa nas descrições acima.

3) Qual o ramo de atuação da sua instituição? *

Marcar apenas uma oval.

◯ Pública

◯ Privada

◯ Mista

4) Na sua opinião, sua empresa lida de alguma forma com dados suscetíveis a alguma análise de prevenção/combate à fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

5) Na sua opinião, o papel que você (individualmente) exerce na empresa envolve, de alguma forma, o tratamento de dados suscetíveis a alguma análise de prevenção/combate à fraude? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

6) Na sua opinião, sua empresa lida de alguma forma com dados pessoais (de clientes, empregados, fornecedores, contratados, entre outros)? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

7) Na sua opinião, o papel que você (individualmente) exerce na empresa envolve, de alguma forma, o tratamento de dados pessoais (de clientes, empregados, fornecedores, contratados, entre outros)? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

8) Na sua opinião, sua empresa tem desenvolvido atividades/plano para se adequar aos requisitos de tratamento de dados pessoais presentes na LGPD?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

9) Avalie a afirmação: "Eu conheço a LGPD e seus princípios". *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

Atividades relacionadas à Disciplina de Governança de Dados

Nesta seção iremos coletar suas impressões com relação a algumas atividades da disciplina de governança de dados encontradas na literatura da área e sua relação com a proteção à privacidade de dados e as atividades de prevenção à fraude. Não há resposta certa, apenas sua opinião, portanto sinta-se livre para responder de acordo com seus conhecimentos. Esta seção é composta de 66 questões múltipla escolha.

10) Quanto a "Escolha de arquitetura de governança (des)centralizada do controlador de dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

11) Quanto a "Escolha de arquitetura de governança (des)centralizada do controlador de dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

12) Quanto a "Definição do Acesso e Propriedade do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

13) Quanto a "Definição do Acesso e Propriedade do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

14) Quanto a "Análise do Ambiente Regulatório (sentido amplo)", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

15) Quanto a "Análise do Ambiente Regulatório (sentido amplo)", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

16) Quanto a "Análise de Utilidade do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

17) Quanto a "Análise de Utilidade do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

18) Quanto a "Análise do Custo-benefício (ROI) do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

19) Quanto a "Análise do Custo-benefício (ROI) do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

20) Quanto a "Aplicação de Casos de Uso de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

21) Quanto a "Aplicação de Casos de Uso de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

22) Quanto a "Análise de Conformidade Legal", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

23) Quanto a "Análise de Conformidade Legal", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

24) Quanto a "Monitoração de Processos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

25) Quanto a "Monitoração de Processos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

26) Quanto a "Desenho da Linhagem/Fluxo/Mapeamento de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

27) Quanto a "Desenho da Linhagem/Fluxo/Mapeamento de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

28) Quanto a "Processo de Qualidade de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

29) Quanto a "Processo de Qualidade de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

30) Quanto a "Política de Dados e Privacidade", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

31) Quanto a "Política de Dados e Privacidade", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

32) Quanto a "Desenho da Estratégia de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

33) Quanto a "Desenho da Estratégia de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

34) Quanto a "Planejamento do Ciclo de Vida do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

35) Quanto a "Planejamento do Ciclo de Vida do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

36) Quanto a "Gestão dos Metadados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

37) Quanto a "Gestão dos Metadados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo

◯ Discordo totalmente

38) Quanto a "Controle da Infraestrutura de Governança de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

39) Quanto a "Controle da Infraestrutura de Governança de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

40) Quanto a "Definição dos Princípios de Dados e Valor de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

41) Quanto a "Definição dos Princípios de Dados e Valor de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

42) Quanto a "Compreender Necessidades Corporativas Estratégicas de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

43) Quanto a "Compreender Necessidades Corporativas Estratégicas de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

44) Quanto a "Definição de Papéis e Responsabilidades", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

45) Quanto a "Definição de Papéis e Responsabilidades", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

46) Quanto a "Organização da Gestão e Governança de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

47) Quanto a "Organização da Gestão e Governança de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

48) Quanto a "Gestão de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

49) Quanto a "Gestão de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

50) Quanto a "Arquitetura de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

51) Quanto a "Arquitetura de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

52) Quanto a "Projetos e Serviços de Gestão de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

53) Quanto a "Projetos e Serviços de Gestão de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

54) Quanto a "Capacitação de Recursos Humanos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

55) Quanto a "Capacitação de Recursos Humanos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

56) Quanto a "Coordenar Atividades de Governança de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

57) Quanto a "Coordenar Atividades de Governança de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

58) Quanto a "Resolução de conflitos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

59) Quanto a "Resolução de conflitos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

60) Quanto a "Modelagem de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

61) Quanto a "Modelagem de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

62) Quanto a "Classificação de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

63) Quanto a "Classificação de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

64) Quanto a "Análise de Requisitos e Impacto", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

65) Quanto a "Análise de Requisitos e Impacto", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

66) Quanto a "Harmonização de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

67) Quanto a "Harmonização de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

68) Quanto a "Integração de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade

de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

69) Quanto a "Integração de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

70) Quanto a "Gestão de Dados Mestres", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

71) Quanto a "Gestão de Dados Mestres", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

72) Quanto a "Curadoria de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

73) Quanto a "Curadoria de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

74) Quanto a "Definição do Contexto da Organização", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

75) Quanto a "Definição do Contexto da Organização", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo

◯ Discordo totalmente

Encerramento

76) Houve alguma atividade da disciplina de governança de dados (e gestão de dados) que não foi impressa nesse questionário e que, em sua opinião, influencia nos aspectos pesquisados (a saber privacidade e prevenção a fraude). Deixe um comentário. Quaisquer sugestões também serão bem-vindas. Se quiser me contactar pessoalmente, sinta-se livre para usar minha correspondência eletrônica (arturpotiguaracarvalho@gmail.com, artur.carvalho@brb.com.br, artur.carvalho@unb.br).

I.2 PRIVACIDADE - IDENTIFICAÇÃO DA APLICABILIDADE DE ATIVIDADES DE PRIVACIDADE, SEGURANÇA E GOVERNANÇA DE DADOS NA PREVENÇÃO À FRAUDE (NOS TERMOS DA LGPD)

Identificação da aplicabilidade de atividades de Privacidade, Segurança e Governança de Dados na Prevenção à Fraude (nos termos da LGPD)

Bem vindo ao formulário para identificação das atividades e modelos listados na literatura pertinente e aplicáveis no mercado de desenvolvimento de softwares para prevenção à fraude e conformidade com a Lei Geral de Proteção de Dados Pessoais (Lei Federal nº 13.709 - LGPD). Abaixo, serão relacionados alguns modelos e atividades identificados em publicações acadêmicas e de mercado. Será solicitada sua opinião quanto ao conhecimento destes modelos e a contribuição efetiva deles à proteção da privacidade, segurança e governabilidade de dados pessoais. Não se preocupe se alguns dos modelos forem desconhecidos ou não tiver opinião formada sobre sua aplicabilidade, basta marcar o item identificado com essa opção (Não concordo nem discordo) ou deixá-lo em branco. A estimativa de tempo para preencher o questionário é de 30 minutos. Muito Obrigado!

Esta pesquisa é parte de um trabalho de mestrado do Programa de Pós-Graduação Profissional em Engenharia Elétrica da Universidade de Brasília - PPEE/UnB.

Contatos dos pesquisadores: ednacanedo@unb.br, artur.carvalho@aluno.unb.br, artur.carvalho@brb.com.br

DECLARAÇÃO DE CONSENTIMENTO INFORMADO

Ao responder a esta pesquisa, você permite que os pesquisadores obtenham, usem e divulguem as informações anônimas fornecidas conforme descrito abaixo.

CONDIÇÕES E ESTIPULAÇÕES

1. Eu entendo que todas as informações são confidenciais. Não serei identificado pessoalmente. Concordo em preencher o questionário para fins de pesquisa e que os dados derivados desta pesquisa anônima podem ser publicados em periódicos, conferências e postagens de blog.

2. Eu entendo que minha participação nesta pesquisa é totalmente voluntária e que a recusa em participar não implicará em nenhuma penalidade ou perda de benefícios. Se eu quiser, posso cancelar minha participação a qualquer momento. Também entendo que, se decidir participar, posso recusar-me a responder a qualquer pergunta para a qual não me sinta confortável em responder.

3. Entendo que posso entrar em contato com os pesquisadores se tiver alguma dúvida sobre a pesquisa. Estou ciente de que meu consentimento não me beneficiará diretamente. Também estou ciente de que o autor manterá os dados coletados em perpetuidade e poderá utilizar os dados para trabalhos acadêmicos futuros.

4. Ao clicar no botão abaixo, eu livremente dou consentimento e reconheço meus direitos como um participante voluntário de pesquisa conforme descrito acima e dou consentimento aos pesquisadores para usarem minhas respostas na condução de pesquisas nas áreas mencionadas acima.

*Obrigatório

Informações Gerais

1) Qual o nome da instituição à qual você é vinculado(a)?

---

2) Como você classificaria a empresa onde trabalha? *

Marcar apenas uma oval.

- ◯ Somos uma micro-empresa, empresa familiar, pequeno negócio.
- ◯ Somos uma empresa de porte médio, com alguns poucos funcionários, mas sem funcionários específicos de desenvolvimento de software
- ◯ Somos uma empresa de porte médio e nossa empresa possui quadro de pessoal específico de desenvolvimento de sofware. Somos uma empresa de grande porte mas não possuímos participações em outras empresas (como filiais ou conglomerados)
- ◯ Somos uma empresa de grande porte e possuímos filiais.
- ◯ Somos um conglomerado sob uma marca.
- ◯ Minha empresa não se encaixa nas descrições acima.

3) Qual o ramo de atuação da sua instituição? *

Marcar apenas uma oval.

- ◯ Pública
- ◯ Privada
- ◯ Mista

4) Na sua opinião, sua empresa lida de alguma forma com dados suscetíveis a alguma análise de prevenção/combate à fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo

◯ Discordo totalmente

5) Na sua opinião, o papel que você (individualmente) exerce na empresa envolve, de alguma forma, o tratamento de dados suscetíveis a alguma análise de prevenção/combate à fraude? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

6) Na sua opinião, sua empresa lida de alguma forma com dados pessoais (de clientes, empregados, fornecedores, contratados, entre outros)? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

7) Na sua opinião, o papel que você (individualmente) exerce na empresa envolve, de alguma forma, o tratamento de dados pessoais (de clientes, empregados, fornecedores, contratados, entre outros)? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

8) Na sua opinião, sua empresa tem desenvolvido atividades/plano para se adequar aos requisitos de tratamento de dados pessoais presentes na LGPD?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

9) Avalie a afirmação: "Eu conheço a LGPD e seus princípios". *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

Atividades relacionadas à Disciplina de Privacidade

Nesta seção iremos coletar suas impressões com relação a algumas atividades da disciplina de privacidade de dados encontradas na literatura da área e sua relação com as atividades de prevenção à fraude. Não há resposta certa, apenas sua opinião, portanto sinta-se livre para responder de acordo com seus conhecimentos. Esta seção é composta de 54 questões múltipla escolha.

10) Quanto a "Definição do Acesso e Propriedade do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

11) Quanto a "Definição do Acesso e Propriedade do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

12) Quanto a "Análise do Custo-benefício (ROI) do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

13) Quanto a "Análise do Custo-benefício (ROI) do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

14) Quanto a "Monitoração de Processos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

15) Quanto a "Monitoração de Processos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

16) Quanto a "Política de Dados e Privacidade", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

17) Quanto a "Política de Dados e Privacidade", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

18) Quanto a "Controle da Infraestrutura da Privacidade", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

19) Quanto a "Controle da Infraestrutura da Privacidade", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

20) Quanto a "Definição de Papéis e Responsabilidades", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

21) Quanto a "Definição de Papéis e Responsabilidades", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

22) Quanto a "Capacitação de Recursos Humanos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

23) Quanto a "Capacitação de Recursos Humanos", na sua opinião, esta atividade contribui para a preven-

ção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

24) Quanto a "Classificação de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

25) Quanto a "Classificação de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

26) Quanto a "Análise de Requisitos e Impacto", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

27) Quanto a "Análise de Requisitos e Impacto", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

28) Quanto a "Avaliação de Riscos de Segurança da Informação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

29) Quanto a "Avaliação de Riscos de Segurança da Informação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

30) Quanto a "Uso de Métodos Criptográficos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo

◯ Discordo totalmente

31) Quanto a "Uso de Métodos Criptográficos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

32) Quanto a "Auditoria", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

33) Quanto a "Auditoria", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

34) Quanto a "Controle da Transferência de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

Discordo

Discordo totalmente

35) Quanto a "Controle da Transferência de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

36) Quanto a "Política de Desenvolvimento Seguro", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

37) Quanto a "Política de Desenvolvimento Seguro", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

38) Quanto a "Política de Dados para Teste", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

39) Quanto a "Política de Dados para Teste", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

40) Quanto a "Definição do Contexto da Organização", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

41) Quanto a "Definição do Contexto da Organização", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

42) Quanto a "Coleta do Consentimento do Uso dos Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

43) Quanto a "Coleta do Consentimento do Uso dos Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

44) Quanto a "Definição do Propósito de tratamento", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

45) Quanto a "Definição do Propósito de tratamento", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

46) Quanto a "Mapeamento de Fontes Externas de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

47) Quanto a "Mapeamento de Fontes Externas de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

48) Quanto a "Definição de Métricas de Riscos à Privacidade", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

49) Quanto a "Definição de Métricas de Riscos à Privacidade", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo

◯ Discordo totalmente

50) Quanto a "Privacidade por Projeto (Privacy by Design) e Privacidade por padrão (Privacy by Default)", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

51) Quanto a "Privacidade por Projeto (Privacy by Design) e Privacidade por padrão (Privacy by Default)", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

52) Quanto a "Publicação de Dados (Statistical Disclosure Control Method)", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

53) Quanto a "Publicação de Dados (Statistical Disclosure Control Method)", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

54) Quanto a "Modelo Baseado em Estados para Privacidade de Usuário", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

55) Quanto a "Modelo Baseado em Estados para Privacidade de Usuário", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

56) Quanto a "Controle de Acesso", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

57) Quanto a "Controle de Acesso", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

58) Quanto a "Uso de Ferramentas de Mascaramento de Dados e Embaralhamento", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

59) Quanto a "Uso de Ferramentas de Mascaramento de Dados e Embaralhamento", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

60) Quanto a "Geração de Dados Sintéticos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

61) Quanto a "Geração de Dados Sintéticos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

62) Quanto a "Redes Neurais para Privacidade", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

63) Quanto a "Redes Neurais para Privacidade", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

Encerramento

64) Houve alguma atividade da disciplina de privacidade que não foi impressa nesse questionário e que, em sua opinião, influencia nos aspectos pesquisados (a saber privacidade e prevenção a fraude). Deixe um comentário. Quaisquer sugestões também serão bem-vindas. Se quiser me contactar pessoalmente, sinta-se livre para usar minha correspondência eletrônica (arturpotiguaracarvalho@gmail.com, artur.carvalho@brb.com.br, artur.carvalho@unb.br).

I.3 SEGURANÇA - IDENTIFICAÇÃO DA APLICABILIDADE DE ATIVIDADES DE PRIVACIDADE, SEGURANÇA E GOVERNANÇA DE DADOS NA PREVENÇÃO À FRAUDE (NOS TERMOS DA LGPD)

Identificação da aplicabilidade de atividades de Privacidade, Segurança e Governança de Dados na Prevenção à Fraude (nos termos da LGPD)

Bem vindo ao formulário para identificação das atividades e modelos listados na literatura pertinente e aplicáveis no mercado de desenvolvimento de softwares para prevenção à fraude e conformidade com a Lei Geral de Proteção de Dados Pessoais (Lei Federal nº 13.709 - LGPD). Abaixo, serão relacionados alguns modelos e atividades identificados em publicações acadêmicas e de mercado. Será solicitada sua opinião quanto ao conhecimento destes modelos e a contribuição efetiva deles à proteção da privacidade, segurança e governabilidade de dados pessoais. Não se preocupe se alguns dos modelos forem desconhecidos ou não tiver opinião formada sobre sua aplicabilidade, basta marcar o item identificado com essa opção (Não concordo nem discordo) ou deixá-lo em branco. A estimativa de tempo para preencher o questionário é de 30 minutos. Muito Obrigado!

Esta pesquisa é parte de um trabalho de mestrado do Programa de Pós-Graduação Profissional em Engenharia Elétrica da Universidade de Brasília - PPEE/UnB.

Contatos dos pesquisadores: ednacanedo@unb.br, artur.carvalho@aluno.unb.br, artur.carvalho@brb.com.br

DECLARAÇÃO DE CONSENTIMENTO INFORMADO

Ao responder a esta pesquisa, você permite que os pesquisadores obtenham, usem e divulguem as informações anônimas fornecidas conforme descrito abaixo.

CONDIÇÕES E ESTIPULAÇÕES

1. Eu entendo que todas as informações são confidenciais. Não serei identificado pessoalmente. Concordo em preencher o questionário para fins de pesquisa e que os dados derivados desta pesquisa anônima podem ser publicados em periódicos, conferências e postagens de blog.

2. Eu entendo que minha participação nesta pesquisa é totalmente voluntária e que a recusa em participar não implicará em nenhuma penalidade ou perda de benefícios. Se eu quiser, posso cancelar minha participação a qualquer momento. Também entendo que, se decidir participar, posso recusar-me a responder a qualquer pergunta para a qual não me sinta confortável em responder.

3. Entendo que posso entrar em contato com os pesquisadores se tiver alguma dúvida sobre a pesquisa. Estou ciente de que meu consentimento não me beneficiará diretamente. Também estou ciente de que o autor manterá os dados coletados em perpetuidade e poderá utilizar os dados para trabalhos acadêmicos futuros.

4. Ao clicar no botão abaixo, eu livremente dou consentimento e reconheço meus direitos como um participante voluntário de pesquisa conforme descrito acima e dou consentimento aos pesquisadores para usarem minhas respostas na condução de pesquisas nas áreas mencionadas acima.

*Obrigatório

Informações Gerais

1) Qual o nome da instituição à qual você é vinculado(a)?

---

2) Como você classificaria a empresa onde trabalha? *

Marcar apenas uma oval.

◯ Somos uma micro-empresa, empresa familiar, pequeno negócio.

◯ Somos uma empresa de porte médio, com alguns poucos funcionários, mas sem funcionários específicos de desenvolvimento de software

◯ Somos uma empresa de porte médio e nossa empresa possui quadro de pessoal específico de desenvolvimento de sofware. Somos uma empresa de grande porte mas não possuímos participações em outras empresas (como filiais ou conglomerados)

◯ Somos uma empresa de grande porte e possuímos filiais.

◯ Somos um conglomerado sob uma marca.

◯ Minha empresa não se encaixa nas descrições acima.

3) Qual o ramo de atuação da sua instituição? *

Marcar apenas uma oval.

◯ Pública

◯ Privada

◯ Mista

4) Na sua opinião, sua empresa lida de alguma forma com dados suscetíveis a alguma análise de prevenção/combate à fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

5) Na sua opinião, o papel que você (individualmente) exerce na empresa envolve, de alguma forma, o tratamento de dados suscetíveis a alguma análise de prevenção/combate à fraude? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

6) Na sua opinião, sua empresa lida de alguma forma com dados pessoais (de clientes, empregados, fornecedores, contratados, entre outros)? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

7) Na sua opinião, o papel que você (individualmente) exerce na empresa envolve, de alguma forma, o tratamento de dados pessoais (de clientes, empregados, fornecedores, contratados, entre outros)? *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

8) Na sua opinião, sua empresa tem desenvolvido atividades/plano para se adequar aos requisitos de tratamento de dados pessoais presentes na LGPD?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

9) Avalie a afirmação: "Eu conheço a LGPD e seus princípios". *

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

Atividades relacionadas à Disciplina de Segurança/Anonimização de Dados

Nesta seção iremos coletar suas impressões com relação a algumas atividades da disciplina de segurança/anonimização de dados encontradas na literatura da área e sua relação com a proteção à privacidade dedados e a atividades de prevenção à fraude. Não há resposta certa, apenas sua opinião, portanto sinta-se livre para responder de acordo com seus conhecimentos. Esta seção é composta de 72 questões múltipla escolha.

10) Quanto a "Definição do Acesso e Propriedade do Dado", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

11) Quanto a "Definição do Acesso e Propriedade do Dado", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

12) Quanto a "Política de Dados e Privacidade", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

13) Quanto a "Política de Dados e Privacidade", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

14) Quanto a "Controle da Infraestrutura de Segurança da Informação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

15) Quanto a "Controle da Infraestrutura de Segurança da Informação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

16) Quanto a "Arquitetura de Segurança da Informação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

17) Quanto a "Arquitetura de Segurança da Informação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

18) Quanto a "Princípios de Dados e Valor de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

19) Quanto a "Princípios de Dados e Valor de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

20) Quanto a "Resolução de conflitos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

21) Quanto a "Resolução de conflitos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

22) Quanto a "Modelagem de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

23) Quanto a "Modelagem de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

24) Quanto a "Classificação de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

25) Quanto a "Classificação de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

26) Quanto a "Avaliação de Riscos de Segurança da Informação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

27) Quanto a "Avaliação de Riscos de Segurança da Informação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

28) Quanto a "Uso de Métodos Criptográficos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

29) Quanto a "Uso de Métodos Criptográficos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

30) Quanto a "Definição de Estratégia de Cópias de Segurança", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

31) Quanto a "Definição de Estratégia de Cópias de Segurança", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

32) Quanto a "Auditoria", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

33) Quanto a "Auditoria", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

34) Quanto a "Controle da Transferência de Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

35) Quanto a "Controle da Transferência de Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

36) Quanto a "Política de Desenvolvimento Seguro e Dados para Teste", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

37) Quanto a "Política de Desenvolvimento Seguro e Dados Para Teste", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

38) Quanto a "Definição do Contexto da Organização", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo

◯ Discordo totalmente

39) Quanto a "Definição do Contexto da Organização", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

40) Quanto a "Coleta do Consentimento do Uso dos Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

41) Quanto a "Coleta do Consentimento do Uso dos Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

42) Quanto a "Definição dos Objetivos da Segurança da Informação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

43) Quanto a "Definição dos Objetivos da Segurança da Informação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

44) Quanto a "Definição dos Serviços de Autenticação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

45) Quanto a "Definição dos Serviços de Autenticação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

46) Quanto a "Deidentificação (Unlinkability)", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

47) Quanto a "Deidentificação (Unlinkability)", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

48) Quanto a "Desenvolvimento do Serviço de Proteção a Dados", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

49) Quanto a "Desenvolvimento do Serviço de Proteção a Dados", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

50) Quanto a "Uso de Ferramentas de Privacidade Diferencial e K-Anonimização", na sua opinião, esta

atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

51) Quanto a "Uso de Ferramentas de Privacidade Diferencial e K-Anonimização", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

52) Quanto a "Uso de Ferramentas de Inteligência Artificial Aplicadas e Aprendizado de Máquina à Segurança (tais como Modelo Oculto de Markov, Computação Bayesiana Aproximada e Redes Bayesianas, Aprendizado Federado, Redes Neurais, Block-Chain, Mineração de Dados, entre outras)", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

53) Quanto a "Uso de Ferramentas de Inteligência Artificial e Aprendizado de Máquina Aplicadas à Segurança (tais como Modelo Oculto de Markov, Computação Bayesiana Aproximada e Redes Bayesianas, Aprendizado Federado, Redes Neurais, Block-Chain, Mineração de Dados, entre outras)", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

54) Quanto a "Controle de Acesso", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

55) Quanto a "Controle de Acesso", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

56) Quanto a "Uso de Ferramentas de Criptografia Homomórfica", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

◯ Concordo totalmente

◯ Concordo

◯ Não concordo nem discordo

◯ Discordo

◯ Discordo totalmente

57) Quanto a "Uso de Ferramentas de Criptografia Homomórfica", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

◯ Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

58) Quanto a "Uso de Ferramentas de Mascaramento de Dados e Embaralhamento", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

59) Quanto a "Uso de Ferramentas de Mascaramento de Dados e Embaralhamento", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

60) Quanto a "Geração de Dados Sintéticos", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

Concordo totalmente

Concordo

Não concordo nem discordo

Discordo

Discordo totalmente

61) Quanto a "Geração de Dados Sintéticos", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

62) Quanto a "Definição da Política de Segurança da Informação", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

63) Quanto a "Definição da Política de Segurança da Informação", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

64) Quanto a "Definição dos Requisitos de Segurança", na sua opinião, esta atividade contribui para a proteção da privacidade de dados (nos moldes da LGPD)?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

65) Quanto a "Definição dos Requisitos de Segurança", na sua opinião, esta atividade contribui para a prevenção a fraude?

Marcar apenas uma oval.

- ◯ Concordo totalmente
- ◯ Concordo
- ◯ Não concordo nem discordo
- ◯ Discordo
- ◯ Discordo totalmente

Encerramento

66) Houve alguma atividade da disciplina de segurança e anonimização de dados que não foi impressa nesse questionário e que, em sua opinião, influencia nos aspectos pesquisados (a saber privacidade e prevenção a fraude). Deixe um comentário. Quaisquer sugestões também serão bem-vindas. Se quiser me contactar pessoalmente, sinta-se livre para usar minha correspondência eletrônica (arturpotiguaracarvalho@gmail.com, artur.carvalho@brb.com.br, artur.carvalho@unb.br).
