mythos
MYTHOS. FACULDADES INTEGRADAS DE CATAGUASES
Journal homepage:
https://periodicos.unis.edu.br/index.php/mythos
ISSN 1994-0098 http:// 10.36674/mythos.v20i2.804

LEI GERAL DE PROTEÇÃO DE DADOS: MÉTODOS DE ANONIMIZAÇÃO E PSEUDONIMIZAÇÃO.

GENERAL DATA PROTECTION LAW: METHODS OF ANONYMIZATION AND PSEUDONYMIZATION.

Salvador Márcio Rodrigues da Silva[1*], Ana Carolina Nogueira Gomes[2], Tiago Bittencourt Nazaré[3]

[1] Graduando do Curso de Sistemas de Informação, Rede de Ensino Doctum, Cataguases, MG, Brasil, salvador.analista@gmail.com
[2] Graduanda do Curso de Sistemas de Informação, Rede de Ensino Doctum, Cataguases, MG, Brasil, carolngomes99@gmail.com
[3] Mestre em Gestão de Sistemas de Engenharia/UCP, Rede de Ensino Doctum, Cataguases, MG, Brasil, tiago_bit@yahoo.com.br

Resumo

O presente estudo tem por objetivo refletir acerca da relevância que os dados pessoais vêm adquirindo diante do rápido crescimento da rede mundial de computadores e do consequente aumento das transações eletrônicas e dos incidentes de segurança da informação, bem como realizar, sem a pretensão de esgotar as possibilidades, um levantamento bibliográfico de métodos de anonimização e pseudonimização para proteção desses dados, considerando ainda o risco residual de reidentificação do seu titular. Será abordado brevemente o entrelaçamento principiológico existente entre a normatização europeia, intitulada *General Data Protection Regulation* (GDPR), e a Lei Geral de Proteção de Dados Pessoais (LGPD), em vigor no Brasil, normas nas quais se identifica a preocupação com o estabelecimento de um nível mínimo de proteção aos dados pessoais. Serão tratados dos elementos da segurança da informação, fundamentada na tríade CIA ("*Confidentiality*", "*Integrity*" e "*Availability*"), assim como de conceitos inerentes ao estudo dos incidentes de segurança da informação, temas indispensáveis à compreensão da proteção de dados, culminando na apresentação de conceitos e de um rol não exaustivo de métodos de anonimização e pseudonimização. A metodologia adotada foi a pesquisa bibliográfica em fontes acadêmicas, artigos científicos, livros e documentações oficiais, notadamente a manifestação técnica da Autoridade Nacional de Proteção de Dados – ANPD, partindo, fundamentalmente, da legislação específica em vigor, que reflete os últimos avanços na proteção de dados.

Palavras-chave: dados pessoais, LGPD, anonimização, pseudonimização.

Abstract

The present study aims to reflect on the relevance that personal data has been acquiring in view of the rapid growth of the world wide web and the consequent increase in electronic transactions and information security incidents, as well as carrying out, without the intention of exhausting the possibilities, a bibliographical survey of anonymization and pseudonymization methods to protect this data, also considering the residual risk of re-identification of its holder. The principled intertwining between European standards, entitled General Data Protection Regulation (GDPR), and the General Data Protection Law (LGPD), in force in Brazil, will be briefly discussed, standards in which the concern with the establishment of a minimum level of protection for personal data. The elements of information security will be covered, based on the CIA triad ("Confidentiality", "Integrity" and "Availability"), as well as concepts inherent to the study of information security incidents, topics essential to understanding data protection, culminating in the presentation of concepts and a non-exhaustive list of anonymization and pseudonymization methods. The methodology adopted was bibliographical research in academic sources, scientific articles, books and official documentation, notably the technical statement from the National Data Protection Authority – ANPD, fundamentally based on the specific legislation in force, which reflects the latest advances in data protection.

Keywords: personal data, LGPD, anonymization, pseudonymization.

Journal homepage:
https://periodicos.unis.edu.br/index.php/mythos
ISSN 1994-0098 http:// 10.36674/mythos.v20i2.804

1 INTRODUÇÃO

O surgimento da rede mundial de computadores vem induzindo uma série de mudanças nas relações humanas. A percepção do tempo e do espaço vem sendo profundamente alterada, tendo em vista a rapidez com que se tem acesso a dados localizados em qualquer parte do globo terrestre. Sejam relações comerciais ou afetivas, todas estão sofrendo os impactos do encurtamento virtual da distância entre as pessoas que, nessa rede, se apresentam através de seus dados pessoais, os quais, a fim de se viabilizar tais relações, transitam freneticamente pelas conexões e computadores que compõem a internet.

Conforme pesquisa feita pela empresa de carteiras digitais PayPal Brasil e pela consultoria de pesquisas BigData Corp[1], o mercado online aumentou em 40,7% desde que a pandemia da Covid 19 teve início, notadamente entre pequenos empreendedores que buscaram nas vendes online uma alternativa às vendas em lojas físicas, que, em grande parte, estavam fechadas. De outro lado, a facilidade de realização de compras e pagamentos no mercado virtual estimulou os consumidores a fornecerem seus dados pessoais para a concretização desses negócios, sem maiores cuidados no que se refere à destinação e utilização dessas informações.

No mesmo sentido, o Comitê Gestor da Internet no Brasil (CGI.br) [2]divulgou que o número de usuários da internet no Brasil chegou a 74%, o que, inevitavelmente, fez aumentar o tráfego e o armazenamento dos dados pessoais desses internautas. Conforme relatório da Unisys[3], publicado no ano de 2021, a preocupação no Brasil com assuntos relacionados à segurança da informação alcançou 197 pontos, em uma escala que chega até máximo de 300, um aumento de 7 pontos em relação ao ano anterior.

Como nos explica TEIXEIRA (2020), antes de tais mudanças, as relações eram concretizadas pessoalmente, tendo, hodiernamente, migrado para o formato virtual, o que transformou a maneira como nos relacionamos. Para ele, nessa nova configuração social, o dado se tornou o cerne de um "sistema econômico virtual gigantesco", tendo em vista a troca constante de dados. O autor faz

[1] PAYPAL; BIGDATACORP. Pesquisa: **Perfil do E-commerce Brasileiro 2020**. Disponível em: https://newsroom.br.paypal-corp.com/pesquisa-perfil-do-e-commerce-brasileiro-2020-ritmo-de-expansao-do-total-de-lojas-online-no-brasil-e-superior-a-40-porcento-ao-ano. Acesso em: 17 de setembro 2023.

[2] CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO. Pesquisa: **TIC Domicílios 2019**. Disponível em: https://cetic.br/media/analises/tic_domicilios_2019_coletiva_imprensa.pdf. Acesso em: 15 de setembro 2023.

[3] UNISYS. Pesquisa: **Índice de Segurança Unisys 2021**. Disponível em: https://www.unisys.com/unisys-security-index/ Acesso em: 16 de setembro 2023.

menção às ferramentas denominadas *cookies* que, armazenadas nas máquinas dos usuários, rastreiam suas navegações e pesquisas, fazendo com que o usuário seja vigiado diariamente, e seus dados coletados e armazenados.

Os casos de uso indevido dos dados pessoais, bem como de vazamento destes, não tardaram a aparecer.

O blog de notícias G1 divulgou que, em 2011, a Playstation network sofreu uma invasão cibernética que resultou no vazamento de dados, inclusive aqueles relacionados a cartões de créditos, de cerca de 2,2 milhões de usuários, evidenciando a necessidade de proteção dos dados sensíveis dos usuários da internet (SILVA, 2020).

No ano de 2016, a Red Cross Blood Service, entidade que presta serviço de captação de sangue humano para doação, sofreu uma invasão no seu sistema, que guardava dados de mais de 500 mil doadores, por causa de transferência de dados em ambiente desprotegido. Naquele episódio, foram divulgados dados relacionados à identificação, bem como outros sigilosos, tais como comportamento sexual e avaliação se o doador era ou não de risco, além de um questionário que, de acordo com as respostas, determinava uma relação com os doadores (TEIXEIRA, 2020).

De acordo com relatório divulgado pela Identity Theft Resource Center, em 2017 ocorreram inúmeros casos de vazamento de dados, como aqueles envolvendo o Pentágono norte-americano e o birô de crédito Equifax, período aquele em que se verificou um aumento de 44,7% de ocorrências de vazamento de dados em relação ao ano anterior, perfazendo um total de 1.579 incidentes dessa natureza (MACHADO e DONEDA, 2018).

Na esteira dessas modificações sociais, em 27 de abril de 2016, o Parlamento Europeu aprovou o novo Regulamento Geral de Proteção de Dados (RGPD), implementado a partir de 25 de maio de 2018.

No Brasil, foi aprovada a Lei 13.709/2018 (Lei Geral de Proteção de Dados Pessoais – LGPD), inspirada na legislação europeia. O direito brasileiro seguiu a orientação da RGPD e adotou o conceito amplo de dado pessoal, qual seja dado pessoal é "informação relacionada a pessoa natural identificada ou identificável" (artigo 5º, I, da Lei 13.709/2018), ou seja, se os dados não se referem a pessoa identificada ou identificável, são considerados anônimos ou anonimizados, não estando, portanto, nesse caso, sob a proteção do manto jurídico da LGPD.

A própria Lei Geral de Proteção de Dados Pessoais (LGPD) prevê os métodos de anonimização e pseudonimização de dados, suscitando de usuários e organizações uma maior compreensão desses conceitos, sobretudo por causa das graves consequências que os incidentes de segurança envolvendo dados pessoais podem acarretar para instituições públicas e privadas.

Nesse contexto, tanto os métodos de anonimização quanto os de pseudonimização adquirem relevância, justificando o presente estudo, posto que, despersonalizados os dados por alguns desses métodos, as organizações que os manipulam podem se beneficiar do eventual afastamento da aplicação, em casos concretos, da legislação específica vigente, uma vez que, se o objetivo da lei é a proteção de direitos de liberdade e privacidade, inerentes à pessoa, em não havendo uma pessoa titular dos dados sob análise, também não existirá razão para a proteção dos referidos direitos.

Quanto aos objetivos deste trabalho, a proposta busca, essencialmente, com base na legislação específica, apresentar e comparar conceitos legais e técnicos, bem como realizar, sem a pretensão de esgotar as possibilidades, um levantamento bibliográfico dos métodos de anonimização e pseudonimização denominados k-anonimato, Privacidade Diferencial (*Differential Privacy*), *Fully Homomorphic Encryption* (FHE), *Property Preserving Encryption* (PPE), *Oblivious*

Random Access Memory (ORAM), *Randomization*, *Faking, Advanced Counterfeiting*, Codificação parcial, Generalização, Tokenização, *Data Encryption Standard* (DES), *Advanced Encryption Standard* (AES), *Rivest-Shamir-Adleman* (RSA) e *Elliptic-curve Cryptography* (ECC), considerando a possibilidade de risco residual de reidentificação do titular dos dados, contribuindo, assim, com possíveis soluções para o cumprimento das exigências legais e para a segurança da informação.

Para tanto, primeiramente, foram apresentados os conceitos de dados e dados pessoais, partindo-se, em seguida, para uma análise comparativa da *General Data Protection Regulation* (GDPR) com a Lei Geral de Proteção de Dados Pessoais (LGPD), abordando, sobretudo, as convergências principiológicas. Posteriormente, o trabalho se concentrou em conceitos e exemplos referentes a segurança da informação e incidentes de segurança da informação, para culminar em conceituar e apresentar métodos de anonimização e pseudonimização de dados, com a respaldo na legislação, literatura especializada e manifestação da Autoridade Nacional de Proteção de Dados – ANPD. Por fim, foram apresentados os resultados e as conclusões.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Dados

Há uma relação entre os conceitos de dados, informação e conhecimento, para a qual tem sido proposta uma representação de pirâmide ou hierárquica, na qual os dados podem ser definidos como o resultado da observação de objetos existentes no mundo. Podem ser números, textos, imagens, áudios etc. Isoladamente, considera-se que os dados não apresentam um significado. Informação, por seu turno, é entendida como o fruto da organização e da análise das relações que existem entre os dados, gerando valor e utilidade para estes, enquanto conhecimento, é compreendido como a "síntese de informações" oriundas de fonte única ou variada (CARVALHO e LORENA, 2017).

Segundo Carvalho e Lorena (2017, p. 49), "Não existe um consenso sobre quando e quem propôs essa representação do relacionamento".

Para o presente estudo, se faz imprescindível a apresentação do conceito de dados pessoais, para o qual será utilizado o que dispõe o art. 5º, I da Lei Geral de Proteção de Dados (LGPD), onde se lê que dado pessoal é "informação relacionada a pessoa natural identificada ou identificável".

2.2 General Data Protection Regulation (GDPR) e sua influência sobre a legislação brasileira

A proteção de dados, atualmente, encontra referência no modelo europeu, consubstanciado no Regulamento Geral sobre Proteção de Dados (RGPD), em inglês, *General Data Protection Regulation* (GDPR). Esse modelo se funda na sistematização de disciplinas anteriores, referentes ao tema, passíveis de inserção na legislação própria de cada estado membro, e que, com a regulamentação, passam a se concentrar basicamente no referido regulamento (DONEDA, 2020).

O Regulamento Geral sobre Proteção de Dados (RGPD), doravante denominado simplesmente de GDPR, é um projeto cujo objetivo é a proteção dos dados e da identidade dos cidadãos da União Europeia, idealizado a partir de 2012, com primeira publicação em 2016, mas que entrou em vigência em 2018, composto por três pilares, que são os fundamentos da GDPR: governança de dados, gestão de dados e transparência (CASTRO, 2021).

A governança de dados é o sistema de gestão e monitoramento que engloba todos os níveis de uma organização, por meio do qual seus princípios e valores são convertidos em recomendações objetivas, realizando o alinhamento de seus interesses com a finalidade de otimizar o seu valor econômico de longo prazo (CASTRO, 2021).

A gestão de dados, por sua vez, é baseada em operações de processamento, transações de dados e da exclusão, tendo, como principais objetivos estratégicos: 1. Compreender as necessidades de informação da empresa e de todos os envolvidos; 2. Coletar, armazenar, proteger, e assegurar a integridade dos dados ativos; 3. Melhorar continuamente a qualidade dos dados e informações, o que inclui: precisão, utilidade, integridade e integração dos dados; 4. Garantir privacidade e confidenciabilidade, bem como a prevenção do uso não autorizado/inapropriado das informações; 5. Maximizar o uso efetivo e o valor dos dados ativos e informações (CASTRO, 2021).

A transparência de dados é constituída pelos processos de consentimento, portabilidade de dados e políticas de privacidade, o que pode ser traduzido na obrigação das empresas de registrar todo o procedimento envolvido na manipulação dos dados, a fim de que seja possível disponibilizá-los aos clientes, quando solicitados, além de fornecer os métodos escolhidos para garantir a proteção e a privacidade dos dados (CASTRO, 2021).

Oliveira (2021) explica que, em Portugal, paralelamente à legislação europeia, a proteção de dados pessoais tem sede na Constituição da República Portuguesa e demais legislações, sendo considerada um direito fundamental, uma vez que sua violação é passível de punição.

Nesse sentido, se de fato o titular de dados pessoais sofre danos materiais ou imateriais devido a uma violação definida no GDPR, nascerá para ele o direito a receber uma indenização, do responsável pelo tratamento ou do subcontratante, pelos prejuízos sofridos (OLIVEIRA, 2021).

Além disto, o regime do RGPD favorece o lesado no que se refere ao ônus da prova, uma vez que permite a sua inversão, bastando ao titular dos dados apenas provar a ocorrência do prejuízo causado pelo incidente, enquanto ao responsável pelo tratamento caberia provar que os fatos não poderiam lhe ser imputados (OLIVEIRA, 2021).

De acordo com o site[4] da Secretaria-Geral da Presidência do Conselho de Ministros (SGPCM) da República Portuguesa, o Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, que é o novo Regulamento Geral sobre a Proteção de Dados (RGPD) da União Europeia (UE), estabelece:

> [...] as regras relativas ao tratamento, por uma pessoa, uma empresa ou uma organização, de dados pessoais relativos a pessoas na UE.
> Para além do reforço da proteção jurídica dos direitos dos titulares dos dados, o RGPD define novas regras e procedimentos do ponto de vista tecnológico. O RGPD é aplicável obrigatoriamente a partir do dia 25 de maio de 2018.

Merecem destaque as seguintes considerações contidas na GDPR:

> (107) A Comissão pode reconhecer que um país terceiro, um território ou um setor específico de um país terceiro, ou uma organização internacional, deixou de

[4] **Regulamento Geral de Proteção de Dados (RGPD).** 2016. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral de Proteção de Dados). União Europeia. Disponível em: https://www.sg.pcm.gov.pt/sobre-nos/regulamento-geral-de-prote%C3%A7%C3%A3o-de-dados.aspx, Acesso em: 20 de agosto de 2023.

> assegurar um nível adequado de proteção de dados. Por conseguinte, deverá ser proibida a transferência de dados pessoais para esse país terceiro ou organização internacional, a menos que sejam cumpridos os requisitos constantes do presente regulamento relativos a transferências sujeitas a garantias adequadas, incluindo regras vinculativas aplicáveis às empresas, e derrogações para situações específicas. Nesse caso, deverão ser tomadas medidas que visem uma consulta entre a Comissão e esse país terceiro ou organização internacional. A Comissão deverá, em tempo útil, informar o país terceiro ou a organização internacional das razões da proibição e iniciar consultas com o país ou organização em causa, a fim de corrigir a situação.
>
> (169) A Comissão deverá adotar atos de execução imediatamente aplicáveis quando haja elementos que comprovem que um país terceiro, um território ou um setor específico nesse país terceiro ou uma organização internacional não assegura um nível de proteção adequado, e imperativos urgentes assim o exigirem.

Observa-se nos textos acima a preocupação da GDPR com o estabelecimento de um nível mínimo de proteção aos dados pessoais, presente em todo o espaço da UE" (DONEDA, 2020, p. 230). Doneda (2020, p. 230) afirma ainda, acerca da influência da regulamentação europeia sobre o fluxo internacional de dados pessoais, que:

> A normativa europeia também acaba tendo uma marcante influência internacional. Entre os motivos para tanto, um é que o crescente fluxo internacional de dados pessoais gera uma demanda por padrões normativos que o legitimem, e as normas europeias são certamente o modelo mais desenvolvido nesse sentido. Outro motivo é a existência de uma cláusula de vedação da transferência de dados para países fora do espaço comunitário que não apresentem nível "adequado" de tutela. O legislador comunitário optou por trazer a discussão sobre esse importantíssimo ponto para o plano normativo – assim evitando a insinuação de uma disputa comercial que corria o risco de ser decidida em outras instâncias; nesse mecanismo percebem-se na Diretiva a força dos princípios que a regem, e impõem a observância da proteção da pessoa.

A lei brasileira de proteção de dados pessoais, por sua vez, foi inspirada na GDPR, circunstância celebrada por Blum e Maldonado (2019, p. 52):

> Nesse contexto, é demasiadamente positivo a LGPD basear-se no GDPR, pois agora o Brasil pode ser reconhecido mundialmente por ter uma legislação robusta, equivalente à norma da UE, facilitando explicações de segurança jurídica para empresas internacionais que buscam investir no País, bem como pela possibilidade da análise, pela Comissão Europeia, do livre fluxo de dados com o Brasil, com base em uma decisão de adequação, assim como Argentina e Uruguai, na América Latina, já estão chancelados.

Vale ressaltar que, refletindo os mesmos objetivos da GDPR, o art. 33 da LGPD dispõe que a "transferência internacional de dados pessoais somente é permitida nos seguintes casos: I - para países ou organismos internacionais que proporcionem grau de proteção de dados pessoais adequado ao previsto nesta Lei".

Outro ponto de contato entre os referidos diplomas normativos é que, de forma semelhante, tanto LGPD como GDPR abordam exclusivamente os dados pessoais, ou seja, aqueles dados ligados a uma pessoa natural identificada ou identificável (BLUM e MALDONADO, 2019). Este conceito é de

suma importância para o entendimento da relevância dos métodos de anonimização e pseudonimização previstos na normatização pátria.

Quanto ao grau de influência da GDPR sobre a legislação pátria, Blum e Maldonado (2019, p. 326) ainda esclarecem que:

> O próprio parecer da Comissão Especial da Câmara dos Deputados – que foi constituída à época da tramitação do Projeto de Lei 4.060/2012 para análise do tema – deixa claro que a intenção do legislador foi, justamente, a de adotar regras muito similares àquelas do direito europeu, inclusive para que o Brasil passasse a apresentar cenário mais atrativo do ponto de vista comercial-regulatório ao setor da TIC (Tecnologia da Informação e Comunicações). Nesse sentido, referido documento ainda relembra que a Argentina, que possui Lei de Proteção de Dados desde 2000, foi o primeiro país latino-americano a conseguir o reconhecimento da União Europeia como "país de nível adequado" para a transferência de dados provenientes do aludido território, o que indica certa atenção do legislador brasileiro em também ambicionar tal acreditação.
>
> De fato, o reconhecimento da União Europeia quanto ao quesito adequação em proteção de dados é cada vez mais buscado pelos países de relevante economia. Na América Latina, além da já comentada Argentina, o Uruguai figura como "porto seguro" para o recebimento de dados oriundos da União Europeia. Os Estados Unidos da América, que não contam com tal chancela europeia, tiveram de editar acordo com a União Europeia para desburocratizar o fluxo de dados UE-EUA, garantindo, assim, maior agilidade e eficácia nas relações comerciais entre os dois territórios.

Percebe-se, assim, um nítido entrelaçamento principiológico da norma brasileira com a GDPR, visando formar uma espécie de "rota" segura, entre países, por onde os dados pessoais possam trafegar.

2.3 Lei Geral de Proteção de Dados (LGPD)

Claramente influenciada pela regulamentação europeia – a *General Data Protection Regulation* (GDPR) – a Lei Federal nº 13.709/2018, Lei Geral de Proteção de Dados Pessoais (LGPD), foi sancionada em 14 de agosto de 2018 e entrou em vigor no dia 18 de setembro de 2020.

A Lei Geral de Proteção de Dados Pessoais (LGPD), doravante denominada simplesmente LGPD, não visa proteger os dados das empresas (pessoas jurídicas), mas os dados das pessoas físicas, manipulados pelas empresas. A lei se funda na concepção de que os indivíduos devem ter conhecimento e controle sobre o processo de coleta e a manipulação de seus dados, a fim de que os seus direitos fundamentais de liberdade, privacidade e livre formação sejam protegidos, o que acarreta inúmeros deveres às organizações sobre toda o ciclo de vida dos dados. A lei brasileira foi inspirada na GDPR, no que tange ao estabelecimento de princípios, direitos dos titulares, definição de controladores operadores e encarregados do tratamento de dados pessoais, bem como à criação da Autoridade Nacional de Proteção de Dados (CASTRO, 2021).

Quanto à Autoridade Nacional de Proteção de Dados – ANPD, acima referida, ela é, nos termos do art. 55-A e art. 55-J da LGPD, uma autarquia de natureza especial, dotada de autonomia técnica e decisória, a qual compete, entre outras, as funções de zelar pela proteção dos dados pessoais, nos termos da legislação; elaborar diretrizes para a Política Nacional de Proteção de Dados Pessoais e da Privacidade; fiscalizar e aplicar sanções em caso de tratamento de dados realizado em descumprimento à legislação, mediante processo administrativo que assegure o contraditório, a

ampla defesa e o direito de recurso; apreciar petições de titular contra controlador após comprovada pelo titular a apresentação de reclamação ao controlador não solucionada no prazo estabelecido em regulamentação; promover na população o conhecimento das normas e das políticas públicas sobre proteção de dados pessoais e das medidas de segurança; e promover e elaborar estudos sobre as práticas nacionais e internacionais de proteção de dados pessoais e privacidade.

Cabe à Autoridade Nacional de Proteção de Dados – ANPD, portanto, manifestar-se acerca do devido cumprimento das normas da LGPD nos casos concretos levados ao seu conhecimento. As notas técnicas, fruto dessas manifestações, adentram os conceitos da lei e servem para nortear as instituições quanto a sua aplicação.

Quanto ao conceito de dado pessoal, o art. 5º, I da LGPD o define como "informação relacionada a pessoa natural identificada ou identificável". *Contrario sensu*, quando a pessoa natural, titular do dado, não for identificada ou identificável não se estará diante de um dado pessoal, mas sim de um dado anônimo.

A capacidade de tornar anônimo um dado possibilita a geração do dado anonimizado que, conforme art. 5º, III da LGPD, é o "dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento". Por isso, é de suma importância a compreensão do conceito de anonimização, pois é a partir dele que se definem os contornos de aplicação da LGPD.

Os conceitos de anonimização e de dado anonimizado, previstos na LGPD, possuem importância crucial na salvaguarda das informações pessoais, visando proteger a iniciativa privada e sustentar modelos de negócios inovadores. Isso porque a LGPD não abrange danos decorrentes de dados anonimizados, ou seja, informações ligadas a um titular que não podem ser identificadas, sendo, portanto, considerados dados não pessoais. Esse entendimento resulta na inaplicabilidade da legislação em análise a esse tipo específico de informação, pois se o propósito da lei é resguardar os direitos essenciais de liberdade e privacidade, bem como o pleno desenvolvimento da identidade de um indivíduo, então, quando os dados não têm capacidade de identificar ou tornar uma pessoa determinada identificável, não existe razão para que sejam protegidos segundo as diretrizes da Lei Geral de Proteção de Dados Pessoais (BLUM e MALDONADO, 2019).

Enfim, em regra, será dentro dos limites da anonimização que as instituições públicas e privadas realizarão o tratamento de dados permitido pela legislação. Para além desses limites, a LGPD se torna aplicável para fins de responsabilização pelo uso indevido de dados pessoais.

Faz-se mister ressaltar que tais conceitos não devem ser confundidos com a pseudonimização de dados, que ocorre quando um dado perde a capacidade de ser diretamente ou indiretamente associado a um indivíduo, a menos que seja utilizado um conjunto adicional de informações mantidas de forma separada pelo responsável, em um ambiente controlado e seguro. Nesse cenário, o dado ainda manteria seu caráter pessoal (BLUM e MALDONADO, 2019).

Conforme Castro (2021), os princípios que norteiam a LGPD servem de guia para as organizações agirem de acordo com as medidas nela previstas, além de fundamentarem as decisões da ANPD, nos casos de violação às normas legais. O autor assim explica os referidos princípios:

> 1. Finalidade: todos os dados, desde o momento de coleta, devem possuir propósitos legítimos, específicos, explícitos ao titular, que justifiquem o recolhimento dos dados;
> 2. Transparência: o indivíduo tem direito a saber exatamente o que, porque e para que seus dados estão sendo coletados. As empresas devem garantir, informações

> claras, precisas e acessíveis aos titulares dos dados, assim como a realização do tratamento e os respectivos agentes atuantes no processo;
> 3. Segurança: qualquer dado que for tratado deverá ser garantido a segurança e confidencialidade por meio medidas técnicas e administrativas de segurança para proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas;
> 4. Prevenção: é uns dos pontos mais importantes sobre segurança, pois é a adoção de medidas preventivas contra quaisquer eventuais problemas no processo de tratamento de dados pessoais;
> 5. Adequação: o tratamento dos dados deve condizer com as finalidades informadas ao titular, conforme o contexto do tratamento;
> 6. Necessidade: restringe o tratamento dos dados ao mínimo necessário para alcançar suas finalidades;
> 7. Livre Acesso: é um complemento da transparência, já que garante ao titular o livre acesso para consultar informações sobre o tratamento de forma facilitada e gratuita, assim como a integridades de seus dados pessoais;
> 8. Qualidade dos Dados: as informações sobre os titulares, devem garantir que as informações sejam verdadeiras, precisas e atualizadas, de acordo com a necessidade e a finalidade;
> 9. Não Discriminação: o nome é auto-explicativo, ou seja, o tratamento é impossibilitado de ser feito para fins de discriminação ou gerar qualquer abuso contra o titular;
> 10. Responsabilização e Prestação de Contas: de forma simples, é a obrigação de prestar contas do procedimento de manipulação das informações. Determina que os agentes responsáveis devem ser capazes de demonstrar todas as medidas adotadas que comprovem o cumprimento da LGPD (CASTRO, 2021, p. 10-11).

Para Costa (2022), o princípio fundamental da LGPD é o da boa-fé. Sousa et al. (2020), por sua vez, aduz que, tal como a GDPR, a LGPD inclui a segurança dentre os seus dez princípios gerais (art. 6º, VII), determinando que os agentes de tratamento devam adotar medidas técnicas, administrativas e de segurança que tenham aptidão para proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas, tais como destruição, perda, alteração, comunicação, difusão ou qualquer outra forma de tratamento inadequado ou ilícito (art. 46).

Se, de um lado, a lei brasileira não dispõe especificamente acerca de quais métodos de anonimização e pseudonimização devam ser adotados, tendo em vista os constantes avanços tecnológicos na área da segurança da informação, de outro, é expressa no sentido de que, em caso de incidente de segurança que possa acarretar risco ou dano relevante aos titulares, será avaliada eventual existência de medidas técnicas adequadas para tornar ininteligíveis os dados pessoais afetados (art. 48, § 3º). Com efeito, técnicas tais como o k-anonimato e a criptografia, dentre outras, visam exatamente isso: tornar o dado dissociado de seu titular ou ininteligível para aqueles que não detenham capacidade para proceder a sua reidentificação ou não possuam as chaves criptográficas, ou seja, a qualquer pessoa que não esteja autorizada a acessá-lo.

Ao final, se pode concluir que a LGPD, através de seus princípios, notadamente o da segurança, visa garantir a segurança da informação, sobretudo aquela formada por dados pessoais utilizados pelas organizações.

2.4 Segurança da informação

A segurança da informação pode ser traduzida como a proteção das informações importantes de uma organização - arquivos e dados digitais, documentos em papel, mídia física e até mesmo fala humana - contra acessos, divulgação, uso ou alteração não autorizados. A segurança da informação digital também é chamada de segurança de dados.

A prática de segurança da informação é fundamentada na tríade CIA ("*Confidentiality*", "*Integrity*" e "*Availability*"), que corresponde à sigla CID em português (Confidencialidade, Integridade e Disponibilidade), e se destina a orientar a escolha de tecnologias, políticas e práticas das organizações para proteger seus sistemas de informação (*hardware*, *software* bem como as pessoas envolvidas na produção, armazenamento, uso e troca de dados dentro da tecnologia da informação da empresa (TI) a infraestrutura). Serão tratados, a seguir, cada um dos elementos da tríade.

Confidencialidade: garante que pessoas não autorizadas não possam acessar dados que não estão autorizados a acessar. A confidencialidade define os níveis de autorização, desde pessoas com acesso privilegiado até usuários externos autorizados a ver tão somente determinadas informações.

A preservação da confidencialidade pode ser assegurada através da aplicação de criptografia aos dados durante o armazenamento e transmissão, empregando a técnica de preenchimento de tráfego na rede (*traffic padding*), impondo um rigoroso controle de acesso, categorizando os dados e instruindo os colaboradores nos procedimentos adequados (HINTZBERGEN et al., 2018).

Hintzbergen et al. (2018, p. 34) explica o conceito de *traffic padding*:

> As camadas de rede são criptografadas, reduzindo a oportunidade de análise do tráfego. Ainda é possível, nessas condições, um atacante acessar o volume de tráfego na rede e observar o que entra e o que sai de cada sistema final. Uma contramedida para esse tipo de ataque é o preenchimento de tráfego (*traffic padding*).
> O preenchimento de tráfego produz continuamente texto cifrado, mesmo na ausência de texto simples. Um fluxo contínuo de dados aleatórios é gerado. Quando um texto simples está disponível, ele é criptografado e transmitido. Quando não há um texto simples na entrada, dados aleatórios são criptografados e transmitidos.
> Isso torna impossível para um atacante distinguir entre um fluxo de dados verdadeiro e um preenchimento de dados e, portanto, deduzir o volume de tráfego.
> O preenchimento de tráfego é essencialmente uma função de criptografia de enlace. Se apenas a criptografia fim-a-fim for empregada, então as medidas disponíveis para o defensor são mais limitadas. Se a criptografia for empregada na camada de aplicação, então o oponente pode determinar a camada de transporte, o endereço da camada de rede e os padrões de tráfego, os quais permanecerão todos acessíveis.

Hintzbergen et al. (2018, p. 34) fornece exemplos de medidas de confidencialidade:

> ● O acesso à informação é concedido com base na “necessidade de conhecer”. Não é necessário, por exemplo, que um funcionário do departamento financeiro seja capaz de ver relatórios de discussões com clientes.
> ● Os funcionários tomam medidas para garantir que a informação não vá para pessoas que não necessitem dela. Eles asseguram, por exemplo, que nenhum documento confidencial seja deixado sobre suas mesas enquanto estão ausentes (política da mesa limpa).

> ● O gerenciamento de acesso lógico assegura que pessoas ou processos não autorizados não tenham acesso a sistemas automatizados, base de dados e programas. Um usuário, por exemplo, não tem o direito de alterar as configurações do PC.
> ● É criada uma separação de funções entre a organização de desenvolvimento do sistema, a organização de processamento e a organização do usuário. O desenvolvedor não pode, por exemplo, fazer qualquer modificação nos salários.
> ● São criadas separações estritas entre o ambiente de desenvolvimento, o ambiente de teste e aceitação, e o ambiente de produção.

Integridade: garante que todas as informações contidas nos bancos de dados da empresa não sejam alteradas de maneira não autorizada.

Por exemplo, é uma expectativa que os dados armazenados em discos permaneçam estáveis, sem alterações aleatórias devido a problemas com os controladores de disco. De maneira semelhante, o que se espera é que os programas de aplicação registrem as informações de maneira precisa, sem introduzir valores diferentes dos pretendidos (HINTZBERGEN et al., 2018).

A preservação da integridade dos dados também pode ser amplamente assegurada por meio da implementação de técnicas de criptografia, que salvaguardam as informações contra acesso não autorizado ou alterações não autorizadas. Os princípios de políticas e administração relacionados à criptografia podem ser estabelecidos em um documento de políticas distinto (HINTZBERGEN et al., 2018).

Hintzbergen et al. (2018, p. 35) apresenta exemplos de medidas de integridade:

> ● Mudanças em sistemas e dados são autorizadas. Por exemplo, um membro da equipe atribui um novo preço a um artigo no website e outro verifica a validade desse preço antes de ser publicado.
> ● Onde possível, são criados mecanismos que forcem as pessoas a usar o termo correto. Por exemplo, um cliente é sempre chamado de “cliente”; o termo “freguês” não pode ser inserido na base de dados.
> ● As ações dos usuários são gravadas (*logged*) de forma que possa ser determinado quem modificou a informação.
> ● Ações vitais para o sistema, como, por exemplo, a instalação de novo software, não podem ser conduzidas por uma só pessoa. Ao segregar funções, posições e autoridades, ao menos duas pessoas serão necessárias para realizar mudanças que tenham graves consequências.

Disponibilidade: garante que os usuários, a qualquer momento, possam acessar as informações para as quais têm acesso autorizado.

Protocolos de contingência são implementados para assegurar a rápida recuperação das operações após ocorrer uma interrupção de grande magnitude (HINTZBERGEN et al., 2018).

São características da disponibilidade: I) Oportunidade: a informação deve estar disponível sempre que necessário; II) Continuidade: em caso de indisponibilidade dos dados, é possível seguir trabalhando; e III) Robustez: se refere à capacidade adequada para permitir que toda a equipe opere no Sistema (HINTZBERGEN et al., 2018).

Hintzbergen et al. (2018, p. 36) traz exemplos de medidas de disponibilidade:

> ● A gestão (e o armazenamento) de dados é tal que o risco de perder informações seja mínimo.

> ● O dado é, por exemplo, armazenado em um disco de rede, e não no disco rígido do PC.
> ● Os procedimentos de *backup* são estabelecidos. Os requisitos legais de quanto tempo os dados devem ser armazenados são levados em conta. A localização do *backup* é separada fisicamente do negócio, a fim de garantir a disponibilidade nos casos de emergência.
> ● Os requisitos legais sobre quanto tempo os dados devem ser mantidos armazenados variam de país para país na União Europeia, nos EUA e em outros lugares. É importante checar as agências reguladoras individuais do governo para requisitos específicos.

O processo continuado de obtenção e manutenção da confidencialidade, integridade e disponibilidade de dados dentro de um sistema de informação é denominado "garantia da informação".

Nas palavras de Hintzbergen et al. (2018) Segurança da Informação pode ser traduzida na "Preservação da confidencialidade, integridade e disponibilidade da Informação", acrescentando os autores que, "Adicionalmente, outras propriedades, tais como autenticidade, responsabilidade, não repúdio e confiabilidade, também podem ser incluídas".

Segundo Sousa et al. (2020), segurança não é a mesma coisa que privacidade, uma vez que, inobstante ser difícil garantir a privacidade em sistemas não seguros, é possível não ter privacidade em sistemas com boas práticas de segurança.

Como se pode depreender dos ensinamentos de Hintzbergen, Hintzbergen, Smulders e Baars, a criptografia, que na LGPD é classificada como método de pseudonimização, vem sendo considerada como técnica apta a cumprir as exigências da tríade CIA.

2.5 Incidente de segurança da informação

Se, de um lado, a Segurança da Informação é descrita como a preservação dos elementos da tríade CIA em um sistema, de outro, o incidente de segurança da informação é caracterizado por um ou vários eventos, indesejáveis ou inesperados, com potencial para comprometer a segurança da informação (HINTZBERGEN et al., 2018). Dito de outra forma, o incidente de segurança da informação desafia exatamente os elementos da tríade CIA.

Nessa seara, importa conhecer os conceitos de risco, ameaça e vulnerabilidade. O risco envolve a chance de um agente ameaçador aproveitar-se de uma vulnerabilidade, resultando em um impacto nos negócios. Uma ameaça, que se concretiza através de um agente ameaçador, é um possível causador de um incidente que poderia acarretar danos ao sistema ou à organização. Uma vulnerabilidade representa uma fragilidade no sistema passível de exploração por uma ou várias ameaças (HINTZBERGEN et al., 2018).

Dessa forma, considerando que o art. 48, §3º da LGPD dispõe que, na análise da gravidade do incidente de segurança, será avaliada "eventual comprovação de que foram adotadas medidas técnicas adequadas que tornem os dados pessoais afetados ininteligíveis, no âmbito e nos limites técnicos de seus serviços, para terceiros não autorizados a acessá-los", é fácil observar a importância dada pela lei à capacidade, do responsável pela guarda dos dados, de torná-los ininteligíveis a todos aqueles que não ostentam autorização para utilizá-los. Eis o teor do art. 48, § 3º da Lei Geral de Proteção de Dados (LGPD), *ipsis literis*:

> No juízo de gravidade do incidente, será avaliada eventual comprovação de que foram adotadas medidas técnicas adequadas que tornem os dados pessoais

> afetados ininteligíveis, no âmbito e nos limites técnicos de seus serviços, para terceiros não autorizados a acessá-los.

O incidente de segurança da informação é um evento que tem o potencial de comprometer elementos da tríade CIA. Nesse contexto, ganha relevância a adequação das medidas técnicas que tornem ininteligíveis os dados pessoais ante os ditames da LGPD, uma vez que será considerada para fins de definição da gravidade do incidente de segurança da informação.

Para Blum e Maldonado (2019), o teor do art. 48, § 3º da LGPD faz crer que o legislador se referiu a criptografia como tecnologia específica com capacidade para tornar os dados pessoais incompreensíveis para usuários não autorizados.

2.6 Anonimização e Pseudonimização

Segundo o art. 5º, XI da LGPD, anonimização é a "utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo". Já o art. 13, §4º do mesmo diploma normativo dispõe que "pseudonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e Seguro".

Algumas formas de proteção da privacidade de indivíduos em bases de dados são a anonimização e a pseudonimização, as quais consistem, basicamente, na modificação ou destruição de informações em bases de dados, não tornando mais possível a identificação dos indivíduos, o que, geralmente, é realizado de três formas: supressão (alguns tipos de dados sensíveis são eliminados da base de dados), substituição (dados sensíveis são substituídos por outros dados não sensíveis ou falsos) e generalização (dados específicos são substituídos por categorias mais genéricas como, por exemplo, quando a idade de indivíduos é substituída por intervalos como "entre 20 e 30 anos") (SOUSA et al., 2020).

Enquanto a pseudonimização permite reidentificar e recuperar os dados originais, a anonimização, em tese, não. Neste contexto, uma das técnicas de anonimização seria a denominada k-anonimato, que busca garantir que a informação para cada pessoa seja indistinguível de pelo menos k − 1 outros indivíduos, cujas informações estejam nos dados disponibilizados, enquanto um método de pseudonimização seria a Tokenização, que consiste na substituição de dados por informações aleatórias denominadas como tokens (SOUSA et al., 2020).

Sousa et al. (2020) ainda apresenta as seguintes técnicas:

- Privacidade Diferencial (*Differential Privacy*): adiciona ruído aos dados de maneira que seja impossível ter certeza da informação particular de determinado indivíduo.
- *Fully Homomorphic Encryption* (FHE): permite realizar operações no texto em claro, utilizando-se somente os seus cifrados correspondentes sem a necessidade de decifração.
- *Property Preserving Encryption* (PPE): cifra valores numéricos de forma que a ordem dos dados em claro seja preservada em seus cifrados correspondentes.
- *Oblivious Random Access Memory* (ORAM): utilizada para esconder padrões de acesso em bases de dados, e, para isso, os blocos de dados que são lidos precisam se mover, e não ficarem estáticos, o que acarretaria vazamento de informações de frequência sobre os dados armazenados e, para impedir tal incidente, cada vez que um bloco for lido ele precisará ser realocado em outro local de forma aleatória.

Para Silva (2020), a LGPD dispõe que os dados anonimizados, em princípio, estão vinculados a uma pessoa, mas passaram por etapas que garantiram a sua desvinculação do usuário. Ensina o mesmo autor que os bancos de dados disponibilizam funções para propiciar a anonimização dos dados, impedindo a identificação do usuário através do dado tratado, e nos apresenta técnicas de anonimização dos dados, de acordo com PostgreSQL Anonymizer (2018):

- Adição de ruídos: Implementação de variações de valores desejadas, como por exemplo, a aplicação de um acréscimo de +10% ou uma redução de -10% em uma coluna de salários, ou até mesmo a inserção de uma discrepância de dois anos em um campo de datas.
- *Randomization* (Randomização): Há diversas funções de aleatorização disponíveis para gerar dados completamente aleatórios.
- *Faking* (Falsificação): Substituição de dados sensíveis por valores aleatórios, eliminando qualquer identificação do registro de dados, mas ainda mantendo a utilidade para testes, análises e processamento de dados.
- *Advanced Counterfeiting* (Falsificação Avançada): Criação de dados fictícios na base de dados usando a biblioteca Python Faker3, requerendo a instalação de outros softwares.
- Pseudonimização: Similar à falsificação, a pseudonimização gera valores realistas, mas com a diferença principal de que esses valores sempre permanecerão fictícios. Ela incorpora um número de "salt", que é opcional, para aumentar a complexidade e prevenir ataques de força bruta.
- Codificação parcial: Substituição de partes específicas dos dados. Por exemplo, um e-mail como "maria123@gmail.com" poderia ser transformado em "*****123@gm***.com".
- Generalização: Substituição do valor original por um intervalo que contém tais valores. Por exemplo, de "Felipe tem 24 anos", a generalização permitiria afirmar que "Paulo tem entre 20 e 40 anos".

Na Nota Técnica nº 46/2022/CGF/ANPD, a Autoridade Nacional de Proteção de Dados – ANPD cita métodos de anonimização aventados pelos autores supramencionados:

> 5.24. Como parte da metodologia de anonimização, utilizam-se medidas para garantir que o limite de risco de reidentificação não seja ultrapassado, como por exemplo o k-anonimato. O modelo k-anonimato é usado como uma diretriz antes e após as técnicas de anonimização terem sido aplicadas, para garantir que os identificadores diretos e/ou indiretos (os "quase identificadores") de qualquer registro sejam compartilhados por pelo menos k-1 outros registros. “K” é a proteção de chave fornecida contra ataques de vinculação, onde k registros são idênticos nos atributos de identificação e, portanto, criam uma "classe de equivalência" com k membros, não sendo possível vincular ou isolar o registro de um único indivíduo sempre que há "k" atributos idênticos.
>
> 5.25. O k-anonimato, porém, não é a única medida disponível, mas é relativamente bem compreendida e fácil de aplicar. Métodos alternativos, como a "privacidade diferencial", podem ser mais adequados. A privacidade diferencial envolve vários conceitos, tais como adicionar ruído aleatório aos registros individuais protegidos, fornecer garantias matemáticas de que o "nível de privacidade" predefinido não seja excedido. Um conjunto de dados anonimizados pode ter diferentes níveis de k-anonimato para diferentes conjuntos de identificadores indiretos (quase identificadores), mas para obter a proteção contra vinculação, o k não poderá ser um número baixo que comprometa a anonimização […].

É imperioso ressaltar que dados anteriormente considerados pessoais, mas que foram submetidos a um processo de "anonimização" — o qual envolve a aplicação de métodos técnicos razoáveis e disponíveis no momento do processamento que tornem um dado incapaz de ser diretamente ou indiretamente associado a um indivíduo — deixam de estar sujeitos ao escopo de aplicação da LGPD (BLUM e MALDONADO, 2019).

No entanto, não obstante os métodos de anonimização promoverem a dissociação entre os dados e o seu titular, remanesce o risco residual de uma reidentificação, que nada mais é que o caminho inverso da anonimização, qual seja a capacidade de se vincular os dados, antes anônimos, a um único titular. A Autoridade Nacional de Proteção de Dados – ANPD, na Nota Técnica nº 46/2022/CGF/ANPD, se manifestou acerca do processo de reidentificação:

> 5.20. A reidentificação, por outro lado, é a possibilidade de identificar um único indivíduo, transformando dados anônimos em dados pessoais por meio do uso de correspondência de dados ou técnicas semelhantes. O processo de anonimização e a forma como é implementado terão uma influência direta na probabilidade de riscos de reidentificação. Considerando que, para se anonimizar um dado pessoal, serão utilizados meios técnicos razoáveis e disponíveis no momento desse processo, existe o risco de que alguns processos de anonimização possam ser revertidos no futuro. Convém reconhecer que as circunstâncias podem mudar com o tempo e novos desenvolvimentos tecnológicos e a disponibilidade de informações adicionais pode comprometer os processos de anonimização anteriores.
> 5.21. Importante registrar que a anonimização não reduz a probabilidade de reidentificação de um conjunto de dados a zero. Embora a anonimização total seja o objetivo desejável do ponto de vista da proteção de dados pessoais, em alguns casos, isso não é possível e deve ser considerado um risco residual de reidentificação […].

No §4º do art. 13 da LGPD é apresentado o conceito de pseudonimização, cuja aplicação, nos estritos termos de uma interpretação literal da lei, estaria limitada aos parâmetros estabelecidos no caput deste artigo. O referido dispositivo legal assim dispõe:

> Art. 13. Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas.
> § 1º A divulgação dos resultados ou de qualquer excerto do estudo ou da pesquisa de que trata o caput deste artigo em nenhuma hipótese poderá revelar dados pessoais.
> § 2º O órgão de pesquisa será o responsável pela segurança da informação prevista no caput deste artigo, não permitida, em circunstância alguma, a transferência dos dados a terceiro.
> § 3º O acesso aos dados de que trata este artigo será objeto de regulamentação por parte da autoridade nacional e das autoridades da área de saúde e sanitárias, no âmbito de suas competências.
> § 4º Para os efeitos deste artigo, a pseudonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro.

A LGPD não escolheu adotar um conceito abrangente de pseudonimização, optando por uma abordagem mais restrita e específica. Isso porque, na LGPD, a previsão da criptografia é encontrada unicamente nos dispositivos que tratam de estudos em saúde pública. Não obstante tal restrição à utilização da pseudonimização na LGPD, não há impedimento para que os responsáveis pelo tratamento de dados façam uso abrangente desse procedimento. Isso pode ser adotado como uma estratégia para atenuar os riscos de segurança de maneira geral, uma vez que, sem dúvida, o risco de manipulação dos dados dessa forma será inferior ao risco associado a eles de outra maneira (BLUM e MALDONADO, 2019). Nesse mesmo sentido é o teor da Nota Técnica nº 46/2022/CGF/ANPD, da Autoridade Nacional de Proteção de Dados – ANPD:

> 5.30. De fato, a anonimização é uma das possíveis medidas de segurança que podem ser adotadas visando à proteção de dados pessoais. Ainda no âmbito das medidas técnicas, pode ser mencionada a pseudonimização, que é definida pela LGPD como "o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro" (art. 13, § 4º). Embora a única menção ao termo seja efetuada no art. 13, que trata dos estudos em saúde pública, a pseudonimização de dados pessoais, como, por exemplo, por meio de criptografia, pode ser utilizada nos mais diversos contextos, sempre considerando os objetivos de prevenção e segurança identificados no caso concreto.

A criptografia envolve a codificação de informações compreensíveis por meio de algoritmos, que transformam um texto inicial em um conteúdo totalmente ilegível. É viável realizar o processo oposto, a descriptografia, para recuperar essas informações (SILVA, 2020).

Do ponto de vista técnico, de maneira geral, a pseudonimização é atingida substituindo uma informação pessoal significativa que identifica o titular, o que resulta na perda da capacidade de associação direta ou indireta a um indivíduo. Essa associação só se torna possível com o uso de informações extras, mantidas de forma restrita pelo responsável pelo tratamento. Entretanto, essa simples modificação não é suficiente para impedir a identificação do titular dos dados, especialmente diante de eventual tentativa de "força bruta", que envolve testar todas as possibilidades concebíveis para reidentificar o titular (BLUM e MALDONADO, 2019).

Poder-se-á citar, a título de exemplo, o uso de técnicas como criptografia ou *hash*. No entanto, operar com o *hash* do CPF (Cadastro de Pessoas Físicas), ao invés do próprio dado, por exemplo, não é o bastante para considerar que os dados foram anonimizados. Isso ocorre porque, se o algoritmo do CPF for conhecido, é viável calcular o *hash* de todas as combinações possíveis e, consequentemente, reverter o processo para identificar qual *hash* está associado a qual CPF. No melhor dos casos, tais técnicas podem apenas tornar os dados pseudonimizados (BLUM e MALDONADO, 2019).

Em um sistema criptográfico é necessário que tanto o remetente quanto o destinatário empreguem o mesmo sistema. Uma qualidade fundamental de um sistema criptográfico eficaz é que o algoritmo em si seja de conhecimento público. De maneira geral, existem três categorias de algoritmos criptográficos: criptografia simétrica, assimétrica e unidirecional (HINTZBERGEN et al., 2018).

No sistema simétrico, o remetente e o destinatário compartilham um algoritmo e uma chave secreta. Por isso, se faz essencial que tal chave seja protegida e que seja trocada sempre antes da

comunicação entre remetente e destinatário, pois é utilizada por ambos. Quanto maior o número de remetentes/destinatários trocando as mensagens, maior o risco de comprometimento da chave, uma vez que pode ser interceptada por um agente ameaçador, principalmente se ela não estiver adequadamente protegida (HINTZBERGEN et al., 2018).

O método de encriptação simétrica mais utilizado em todos os tempos é o denominado *Data Encryption Standard* (DES). Ele consiste em uma série de etapas repetidas chamadas de "rounds" (rodadas). Cada rodada aplica uma série de operações, incluindo substituição de bits (chamada de S-boxes) e permutações. Uma chave de 56 bits é usada para gerar subchaves para cada rodada. O tamanho da chave do DES eventualmente se tornou insuficiente para proteger contra os ataques de força bruta, nos quais um adversário tenta todas as chaves possíveis para encontrar a correta (STALLINGS, 2015).

Já o *Advanced Encryption Standard* (AES), algoritmo de criptografia simétrica considerado altamente seguro e eficiente, também utiliza chaves de tamanho fixo para criptografar e descriptografar informações. O AES substituiu o antigo *Data Encryption Standard* (DES) devido à sua maior capacidade de resistir a ataques de criptoanálise. Com tamanhos de chave de 128, 192 ou 256 bits, o AES oferece diferentes níveis de segurança, tornando-o adequado para uma variedade de aplicações (STALLINGS, 2015)

No sistema assimétrico, também conhecido como criptografia de chave pública, são utilizadas chaves diferentes para cifrar e decifrar as mensagens, afastando as vulnerabilidades inerentes às chaves secretas compartilhadas, porquanto remetente e destinatário não necessitam ter a mesma chave. O sistema é baseado no conceito de pares de chaves, em que uma chave pública realiza a criptografia da mensagem, enquanto apenas a chave privada, do respectivo par de chaves, consegue descriptografá-la. Dessa forma, todos podem conhecer a chave pública, desde que a chave privada permaneça secreta (HINTZBERGEN et al., 2018).

A criptografia de chave pública traz uma mudança radical em relação a tudo que tinha sido feito anteriormente, apresentando as seguintes etapas essenciais: para garantir a segurança das mensagens, cada usuário cria um par de chaves, uma para encriptação e outra para decriptação; cada usuário, então, torna uma chave acessível ao público (inserindo em um arquivo acessível), enquanto mantém privada a outra chave, contendo, assim, uma lista de chaves públicas de outros usuários; para enviar uma mensagem de forma confidencial, o usuário remetente faz a encriptação desta mensagem se utilizando da chave pública do usuário destinatário; ao receber a mensagem criptografada, o usuário destinatário procede a sua descriptografia se utilizando da própria chave privada, uma vez que apenas ele conhece sua própria chave privada (STALLINGS, 2015).

Ao utilizar essa técnica, todos os remetentes/destinatários têm conhecimento das chaves públicas, enquanto as chaves privadas, por serem criadas localmente por cada um deles, não são conhecidas. Portanto, enquanto a chave privada for mantida secreta, a mensagem estará protegida. Um sistema pode modificar sua chave privada e substituir sua chave pública (STALLINGS, 2015).

O esquema *Rivest-Shamir-Adleman* (RSA) é considerado a técnica mais utilizada na encriptação assimétrica. A base matemática do RSA está fundamentada na dificuldade de fatorar grandes números inteiros em seus primos constituintes. Isso significa que o algoritmo explora o fato de que, enquanto é fácil multiplicar dois números primos grandes para obter um número grande, o processo inverso de determinar quais números primos foram usados na multiplicação é computacionalmente complexo e demorado. Entretanto, ele pode se tornar mais lento do que algoritmos simétricos quando a criptografia envolve grandes quantidades de dados (STALLINGS, 2015).

O *Elliptic-curve Cryptography* (ECC), por seu turno, é um tipo de criptografia assimétrica baseada na matemática das curvas elípticas. A ECC é conhecida por fornecer uma segurança equivalente a algoritmos tradicionais, como o *Rivest-Shamir-Adleman* (RSA), com chaves muito menores, tornando-a mais adequada para ambientes com recursos limitados, como dispositivos móveis (STALLINGS, 2015).

O sistema unidirecional também é denominado função *hash*. Nele, há a conversão da mensagem em um valor numérico, ou valor de *hash*, o qual, mediante o uso de um algoritmo, é analisado pelo destinatário para verificar se a mensagem apresenta idêntico valor de *hash*. Em resumo, se ambos os valores de *hash* forem iguais, presume-se que a mensagem não foi modificada (HINTZBERGEN et al., 2018).

A função *hash* é comumente usada em sistemas para autenticação de integridade de arquivo, como acontece com a assinatura eletrônica.

A escolha do método a ser adotado dependerá das necessidades específicas inseridas no contexto de aplicação, como nível de segurança desejado, tamanho de chave, desempenho e restrições de recursos.

3 METODOLOGIA

Para o presente estudo, a metodologia adotada foi a abordagem qualitativa de pesquisa, de natureza exploratória e bibliográfica, visando a compreensão dos conceitos legais e técnicos relacionados à anonimização e pseudonimização. Foram consultadas fontes acadêmicas, artigos científicos, livros e documentações oficiais, a fim de obter o embasamento teórico sobre o tema.

Inicialmente, a revisão concentrou-se na análise da Lei Geral de Proteção de Dados (LGPD), com uma breve análise comparativa com a regulamentação europeia, denominada Regulamento Geral sobre Proteção de Dados (RGPD). Foram apresentados os princípios e conceitos fundamentais ao desenvolvimento do presente estudo.

A segunda etapa da metodologia centrou-se na segurança da informação e nos incidentes de segurança. Foram analisados os conceitos e apresentados exemplos de integridade, confidencialidade e disponibilidade dos dados. Além disso, foram apresentados os conceitos de risco, ameaça e vulnerabilidade.

A fase final da revisão se debruçou sobre as técnicas de anonimização e pseudonimização, apresentando os conceitos subjacentes a essas técnicas, assim como sua distinção, juntamente com exemplos de sua aplicação.

4 RESULTADOS/DISCUSSÃO

A análise minuciosa dos textos proporcionou uma melhor compreensão sobre os conceitos essenciais de pseudonimização e anonimização no contexto da proteção de dados pessoais, especialmente em relação às disposições da Lei Geral de Proteção de Dados (LGPD) no cenário brasileiro. Os resultados principais obtidos podem ser resumidos da seguinte maneira.

Em primeiro lugar, foi possível identificar a profunda influência que a legislação europeia, denominada *General Data Protection Regulation* (GDPR), exerceu sobre a criação da Lei Geral de Proteção de Dados (LGPD), que é a norma pátria. Essa conexão com a normativa estrangeira tem sido recebida com entusiasmo pelos estudiosos brasileiros, porquanto é percebida como benéfica a abertura de mercados e a atração de investimentos entre diferentes países, os quais passam a

reconhecer no Brasil um "porto seguro", com a segurança jurídica necessária aos negócios que envolvem fluxo de dados.

Foi possível ainda identificar a preocupação, insculpida na *General Data Protection Regulation* (GDPR) e recepcionada na Lei Geral de Proteção de Dados (LGPD), de se criar um nível mínimo adequado de proteção de dados pessoais para fins de transferência internacional desses dados.

A adequação aos elementos da tríade CIA ("*Confidentiality*", "*Integrity*" e "*Availability*"), que são a base do conceito de segurança da informação, pode ser compreendida como objetivo implícito nos princípios da Lei Geral de Proteção de Dados (LGPD), mais especificamente no princípio da segurança, previsto no art. 6º, VII.

Essencialmente, foi possível fazer a diferenciação entre pseudonimização e anonimização, métodos previstos na Lei Geral de Proteção de Dados (LGPD). A pseudonimização envolve o tratamento dos dados de forma a remover qualquer possibilidade de associação direta ou indireta a um indivíduo, a menos que informações adicionais sejam utilizadas em um ambiente seguro. Por outro lado, a anonimização é uma técnica na qual os dados são modificados ou eliminados para garantir que não seja mais possível identificar uma pessoa específica. Ambas as abordagens visam a proteger a privacidade dos dados pessoais.

Ademais, foi demonstrado que a anonimização tem o potencial de afastamento de incidência das normas da Lei Geral de Proteção de Dados (LGPD) sobre eventual caso concreto de incidente de informação, bem como que a pseudonimização, apesar de prevista unicamente nos dispositivos da lei que tratam de estudos em saúde pública, pode ter aplicação abrangente nas mais variadas áreas em que ocorra a manipulação de dados, conforme doutrina especializada e manifestação técnica da Autoridade Nacional de Proteção de Dados - ANPD.

Além disso, foi apresentado um levantamento bibliográfico, em rol não exaustivo, das seguintes técnicas empregadas para alcançar a pseudonimização e anonimização: k-anonimato, Privacidade Diferencial (*Differential Privacy*), *Fully Homomorphic Encryption* (FHE), *Property Preserving Encryption* (PPE), *Oblivious Random Access Memory* (ORAM), *Randomization*, *Faking*, *Advanced Counterfeiting*, Codificação parcial, Generalização, Tokenização, *Data Encryption Standard* (DES), *Advanced Encryption Standard* (AES), *Rivest-Shamir-Adleman* (RSA) e *Elliptic-curve Cryptography* (ECC).

Uma questão crucial trazida foi a reidentificação dos dados. Embora a pseudonimização e anonimização reduzam a identificação direta dos dados, a possibilidade de reverter o processo e reidentificar os indivíduos ainda representa um risco residual. Isso destaca a importância de adotar técnicas robustas de pseudonimização e anonimização, levando em consideração os possíveis ataques de força bruta e o potencial impacto de avanços tecnológicos futuros, ressaltando os desafios contínuos e a importância de se manter atualizado em relação às técnicas emergentes de reidentificação.

A escolha da técnica adequada depende da sensibilidade dos dados, nível de segurança desejado, tamanho de chave, desempenho e restrições de recursos.

5 CONSIDERAÇÕES FINAIS

A partir dos resultados e das discussões elaboradas, é possível chegar a conclusões importantes.

A pseudonimização e anonimização emergem como estratégias cruciais para atender aos requisitos de privacidade e proteção de dados estabelecidos pela LGPD. Essas técnicas desempenham um duplo papel: no nível individual, são essenciais para a garantia da privacidade dos dados pessoais em um ambiente digital cada vez mais interconectado; no nível institucional, apresentam o potencial de afastar a incidência da norma sobre casos concretos de incidente da informação ocorridos dentro das organizações. Protegem, portanto, a ambos: pessoas naturais e jurídicas.

No entanto, é fundamental reconhecer que a proteção de dados não é uma tarefa simples. A possibilidade de reidentificação permanece como um risco residual, destacando a necessidade contínua de implementar medidas de segurança robustas.

O equilíbrio entre a proteção da privacidade e a utilidade dos dados é um desafio constante. Diante das rápidas evoluções tecnológicas, a compreensão das técnicas de pseudonimização e anonimização torna-se fundamental.

A adequação das medidas técnicas adotadas deve levar em consideração o tempo e o "estado da arte" do desenvolvimento tecnológico, emergindo como opção a adoção de soluções menos por critérios fixos ou demasiadamente objetivos e mais por critérios flexíveis, com adesão a conceitos mais abertos que permitam acompanhar as inovações tecnológicas na área da segurança da informação. A evolução do conhecimento dentro da comunidade jurídica e dos profissionais de tecnologia, bem como a formação gradativa do entendimento técnico da Autoridade Nacional de Proteção de Dados – ANPD sugerem o melhor caminho para o desenvolvimento de critérios para adoção das medidas técnicas adequadas à proteção de dados, tema relevante que emerge desde já como sugestão de objeto para futuros estudos.

Em síntese, a pseudonimização e a anonimização não garantem uma proteção absoluta, mas representam pilares fundamentais na construção de um ambiente digital mais seguro, onde se deve conciliar a proteção da privacidade dos indivíduos com a capacidade das organizações de realizar análises e pesquisas valiosas.

REFERÊNCIAS

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS – ANPD. 2022. Nota Técnica nº 46/2022/CGF/ANPD. Autoridade Nacional de Proteção de Dados. Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/sei_00261-000730_2022_53-nt-46.pdf/view, Acesso em: 26 de agosto de 2023.

BRASIL. Lei Nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD).** Brasília, DF: Diário Oficial da União, [2022]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm, Acesso em: 20 de março de 2023.

CASTRO, Evandro Thalles Vale de. **Transações de Dados e Privacidade à luz da Lei Geral de Proteção de Dados Pessoais (LGPD).** Monografia (Especialização em Engenharia da Computação) - Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Ciência da Computação, Brasília, 2021. Disponível em: https://bdm.unb.br/bitstream/10483/30277/1/2021_EvandroThallesValeCastro_tcc.pdf, Acesso em: 20 de março de 2023.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO. Pesquisa: **TIC Domicílios 2019**. Disponível em: https://cetic.br/media/analises/tic_domicilios_2019_coletiva_imprensa.pdf. Acesso em: 15 de setembro 2023.

Comissão Europeia. Disponível em: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations_pt, Acesso em: 20/03/2023.

COSTA, Renato Brito. **A Lei Geral de Proteção de Dados Pessoais Aplicada à Internet das Coisas: Uma Revisão Sistemática.** Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Universidade Federal do Ceará, Sobral, 2022. Disponível em: https://repositorio.ufc.br/bitstream/riufc/66631/1/2022_tcc_rbcosta.pdf, Acesso em: 20 de março de 2023.

DONEDA, Danilo Cesar Maganhoto. **Da privacidade à proteção de dados pessoais: fundamentos da Lei Geral de Proteção de Dados.** 2. ed. São Paulo: Thomson Reuters Brasil, 2020.

International Business Machines Corporation (IBM). Disponível em: https://www.ibm.com/topics/information-security?mhsrc=ibmsearch_a&mhq=what%20is%20information%20security, Acesso em: 20 de março de 2023.

LORENA, Ana Carolina; CARVALHO, André Carlos Ponce de Leon Ferreira de. **Introdução à Computação: hardware, software e dados.** 1. ed. Rio de Janeiro: LTC, 2017.

MACHADO, Diego; DONEDA, Danilo. **Proteção de Dados Pessoais e Criptografia: Tecnologias Criptográficas Entre Anonimização e Pseudonimização de Dados**. 2018. Disponível em: https://d1wqtxts1xzle7.cloudfront.net/58203675/Protecao_de_dados_pessoais_e_criptografia-libre.pdf?1547734196=&response-content-disposition=inline%3B+filename%3DProtecao_de_dados_pessoais_e_criptografi.pdf&Expires=1694991246&Signature=VPNKTz8rhtw9SPIOFftc28F8g1KG01hvcgwc04x24o5OazFgliWr7FZKgs47jM-M4eLTWcPNyuMlVNGhysmXcETj02FUWObxDzTNQpgkE0zQMQUadnXJgmwIZ4Unhpx06Tx8RDJEXw286dUzCT~viZhUP3oi5NqS~XwRMNvQ5dK4PK2TTQSX0UvAjlQl31Yhra~xGJmXLwUxIRh-an0dghYDzxklDdazw3AaMCRAM1ytsz8jXdx8Q36cacKVmZ3tjB6IbBSDhagc1StfkYyHgLaeYMPABIcoWLqov2UZNFeiOuAYeYNhHhxoFTX~bSemzV2NXfu8EI23ji7ZKmA33w__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA, Acesso em: 17 de setembro de 2023.

MALDONADO, Viviane Nóbrega (Coord.); BLUM, Renato Opice (Coord.). **LGPD Lei Geral de Proteção de Dados Comentada.** 2019. Revista dos Tribunais. 2. ed.

OLIVEIRA, Sara Filipa Carneiro de. **Regime da Proteção de Dados Pessoais nas Empresas: Impacto e Adaptações à Nova Realidade.** Dissertação (Mestrado) - Universidade Lusíada do Porto, Porto, 2021. Disponível em: http://dspace.lis.ulusiada.pt/bitstream/11067/6440/1/FD_Disserta%C3%A7%C3%A3o%20de%20mestrado.pdf, Acesso em: 18 de maio de 2023.

PAYPAL; BIGDATACORP. Pesquisa: **Perfil do E-commerce Brasileiro 2020**. Disponível em: https://newsroom.br.paypal-corp.com/pesquisa-perfil-do-e-commerce-brasileiro-2020-ritmo-de-

expansao-do-total-de-lojas-online-no-brasil-e-superior-a-40-porcento-ao-ano. Acesso em: 17 de setembro 2023.

Portal de Informações sobre RGPD para Empresas (Disponibilizado pela Agência para a Competitividade e Inovação, I.P. - IAPMEI, I.P.). Disponível em: https://www.iapmei.pt/PRODUTOS-E-SERVICOS/Assistencia-Tecnica-e-Formacao/Regime-Geral-de-Protecao-de-Dados.aspx, Acesso em: 21 de março 2023.

Regulamento Geral de Proteção de Dados (RGPD). 2016. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral de Proteção de Dados). União Europeia. Disponível em: https://www.sg.pcm.gov.pt/sobre-nos/regulamento-geral-de-prote%C3%A7%C3%A3o-de-dados.aspx, Acesso em: 20 de agosto de 2023.

SILVA, Lucas Henrique de Moura e. **Escolha da Criptografia Ideal e Anonimação de Dados Sensíveis Citados a Lei Geral de Proteção de Dados.** Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) - Centro Universitário de Anápolis – UniEVANGÉLICA, Anápolis, 2020. Disponível em: http://45.4.96.19/bitstream/aee/17213/1/TCC2%20WiLucasHenrique_Final.pdf, Acesso em: 19 de junho de 2023.

SMULDERS, André; HINTZBERGEN, Kees; HINTZBERGEN, Jule; BAARS, Hans. **Fundamentos de Segurança da Informação: com Base na ISO 27001 e na ISO 27002.** 1. ed. São Paulo: Brasport, 2018.

SOUSA, Thiago do Rego; COUTINHO, Murilo; COUTINHO, Lilian; ALBUQUERQUE, Robson. **LGPD: Levantamento de Técnicas Criptográficas e de Anonimização para Proteção de Bases de Dados.** Sociedade Brasileira de Computação (SBC). 2020. Disponível em: https://sol.sbc.org.br/index.php/sbseg/article/view/19227/19056, Acesso em: 16 de abril de 2023.

STALLINGS, William. **Criptografia e Segurança de Redes: Princípios e Práticas.** tradução Daniel Vieira; revisão técnica: Paulo Sérgio Licciardi Messeder Barreto, Rafael Misoczki. – 6. ed. – São Paulo: Pearson Education do Brasil, 2015.

TEIXEIRA, Guilherme Cardoso. **O Papel Social da Lei Geral de Proteção de Dados no Brasil.** Trabalho de Conclusão de Curso (Bacharel em Direito) - Universidade do Sul de Santa Catarina, Araranguá, 2020. Disponível em: https://repositorio.animaeducacao.com.br/handle/ANIMA/7514, Acesso em: 11 de abril de 2023.

UNISYS. Pesquisa: **Índice de Segurança Unisys 2021**. Disponível em: https://www.unisys.com/unisys-security-index/ Acesso em: 16 de setembro 2023.