ESTUDO PRELIMINAR

Anonimização e Pseudonimização para a proteção de dados pessoais

Autoridade Nacional de Proteção de Dados

Diretor-Presidente
Waldemar Gonçalves Ortunho Junior

Diretores
Arthur Pereira Sabbat

Joacil Basilio Rael

Miriam Wimmer

Equipe de elaboração

Albert França Josuá Costa

Diego Carvalho Machado

Fabíola de Gabriel Soares Pinto

Jeferson Dias Barbosa

Katia Adriana Cardoso de Oliveira

Mariana Talouki

Paulo Cesar dos Santos

Rodrigo Santana dos Santos

Versão 1.0	Dezembro/ 2023

Sumário

1. APRESENTAÇÃO

1. A Lei nº 13.709, de 14 de agosto de 2018 - Lei Geral de Proteção de Dados Pessoais (LGPD), com o objetivo de definir fundamentos e promover a cultura de proteção de dados no Brasil, faz menção a processos que, mediante diferentes técnicas, possibilitam de algum modo afetar a vinculação do dado pessoal, de forma direta ou indireta, com o indivíduo, como as utilizadas em processos de anonimização e de pseudonimização.
2. A Autoridade Nacional de Proteção de Dados (ANPD), na perspectiva de estabelecer um ambiente normativo e orientativo para a proteção de dados, recebeu a autorização legal do § 3º do art. 12, da LGPD para dispor sobre essas técnicas, na forma de orientação aos agentes de tratamento de dados pessoais no Brasil.
3. A ANPD, em sua missão central de salvaguardar a privacidade e a proteção dos dados pessoais, com base em estudos técnicos desenvolvidos internamente[1], elaborou orientações e esclarecimentos sobre o tema, por entender que um melhor conhecimento sobre o processo e as técnicas de anonimização e pseudonimização é importante para que os agentes de tratamento adotem abordagens mais robustas de proteção de dados.
4. Alinhada a esse entendimento, a ANPD oferece este estudo preliminar com o intuito de disseminar os processos e as práticas de anonimização e pseudonimização, não só entre os agentes de tratamento, como também entre os titulares de dados pessoais, reforçando o seu compromisso em ser um parceiro ativo na construção de uma cultura de proteção de dados pessoais sólida e responsável no Brasil.
5. Quanto à **estrutura**, o estudo preliminar está organizado com a seguinte estrutura:
    - **Conceitos básicos** | Apresentação dos conceitos basilares, a partir de um glossário, e uma introdução geral ao regramento da anonimização e pseudonimização de dados de acordo com a disciplina normativa da LGPD.
    - **Os processos de anonimização e pseudonimização de dados na LGPD** | Análise dos processos de anonimização e pseudonimização de dados e seus aspectos jurídicos e técnicos, ressaltando a importância da avaliação contextual, o tipo de tratamento a ser realizado, o volume dos dados pessoais tratados e os riscos de reidentificação envolvidos para tomar a decisão de qual ou quais técnicas devem ser adotadas.
    - **Considerações finais** | Apontamento dos aspectos conclusivos e recomendações sobre os processos de anonimização e pseudonimização de dados à luz da LGPD.

[1] Os estudos técnicos sobre a anonimização e pseudonimização de dados realizados pela ANPD serão publicados em momento oportuno no sítio eletrônico da Autoridade.

- **Apêndices** | Elementos complementares compostos de síntese geral, caderno com técnicas mais relevantes, suas caraterísticas, aplicações e estudos de caso.

6. Devido ao surgimento de novas técnicas e padrões, esta primeira versão tratará do tema observando, nesse contexto, a possibilidade de atualizações com base na evolução tecnológica.
7. Assim, a ANPD observará a evolução sobre o tema com o objetivo de atualização deste estudo preliminar, à medida que novas técnicas e novos entendimentos forem estabelecidos. Ademais, sugestões também podem ser enviadas para a Ouvidoria da ANPD, por meio da Plataforma Fala.BR (https://falabr.cgu.gov.br/).

2. CONCEITOS BÁSICOS

8. Para que seja possível melhor compreender as orientações que se pretende passar, alguns termos são esclarecidos de forma a padronizar e entender o seu significado e sua utilização ao longo deste estudo preliminar.

2.1. GLOSSÁRIO

- **Agente de tratamento:** O controlador e o operador.
- **Anonimização:** 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.
- **Banco de dados:** Conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico.
- **Conjunto de dados:** *Vide* Banco de dados.
- **Controlador:** Pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais.
- **Dado anonimizado:** Dado relativo ao titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento.
- **Dado auxiliar**: identificador adicional empregado para vincular um dado pessoal, que passou por um processo de pseudonimização, e que é capaz de permitir a reidentificação da pessoa natural.
- **Dado pseudonimizado:** Dado que 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.
- **Dado em fluxo**: Dado gerado continuamente a uma alta taxa de velocidade, com tamanho potencialmente infinito e necessidade de processamento imediato.
- **Dado pessoal:** Informação relacionada a pessoa natural identificada ou identificável.

- **Dado pessoal sensível:** Dado pessoal 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.
- **Equivalência de classe:** Subconjunto de um conjunto que contém todos os elementos com algum valor de atributo igual a todos os elementos.
- **Identificador direto:** Dado que, por si só, permite identificar unicamente uma pessoa natural.
- **Identificador indireto:** Dado que, por si só, não tem a capacidade de identificar uma pessoa natural, mas pode ser agregado ou vinculado a dados auxiliares para identificar uma pessoa natural.
- **Métrica base:** Valor definido para mensurar o risco de reidentificação calculado unicamente com base no próprio conjunto de dados, como, por exemplo, a Equivalência de Classe.
- **Métrica contextual:** Métrica derivada de uma métrica base, com a incorporação de elementos particulares
- **Operador:** Pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador.
- **Titular:** Pessoa natural a quem se referem os dados pessoais que são objeto de tratamento.
- **Tratamento:** Toda a operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração.
- **Variável dependente do contexto:** Característica interna do agente de tratamento que pode afetar o cálculo do risco de reidentificação.

2.2. ANONIMIZAÇÃO E PSEUDONIMIZAÇÃO DE DADOS NA LGPD

9. A LGPD tratou, em seu art. 5º, incisos III e XI, sobre a anonimização como um processo em que um agente de tratamento utiliza determinadas técnicas para desvincular, de forma direta ou indireta, o dado pessoal do seu titular por meio do uso de técnicas de processamento de dados.
10. A anonimização, conforme definido no art. 5º, XI, da Lei, é o processo por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, tornando-se, portanto, anonimizado.
11. Em consequência, o dado anonimizado surge, no estágio atual da tecnologia, como o resultado da implementação de processo de anonimização por agente de tratamento, em que são empregados meios técnicos razoáveis e disponíveis na ocasião do tratamento.

12. O dado anonimizado, conforme disposto no art. 5º, III, da LGPD, é aquele dado inicialmente vinculado à pessoa natural, mas que foi posteriormente submetido a processo de anonimização a partir de técnicas ou paradigmas, como generalização e privacidade diferencial. Em razão da remoção dos identificadores diretos e indiretos, os dados perdem, a princípio, o caráter pessoal.

13. Os conjuntos de dados podem conter identificadores que possibilitam a associação, direta ou indireta, a um indivíduo, nos termos do art. 5º, XI e art. 12, § 4º, da LGPD. Daí se dizer que os identificadores podem ser diretos ou indiretos.

14. **O Identificador direto** é o dado que por si só permite identificar unicamente uma pessoa natural, sem a necessidade de combiná-lo com dados de outras fontes. O típico identificador direto de um titular de dados é o seu nome completo. Outro exemplo é o número de inscrição no Cadastro de Pessoas Físicas (CPF), que é considerado número único e suficiente para identificação do cidadão nos bancos de dados de serviços públicos, nos termos da Lei nº 14.534/2023.[2]

15. Já o **identificador indireto**, por sua vez, é considerado o dado que por si só não tem a capacidade de identificar alguém, mas pode ser agregado e vinculado a dados auxiliares para identificar uma pessoa natural, a exemplo da nacionalidade, da idade, da raça, do CEP da residência, das características fenotípicas, ou do endereço de IP que podem ser necessários para distinguir alguém. Também conhecidos como "quase-identificadores", os identificadores indiretos se relacionam ao "fenômeno das 'combinações únicas"[3], isto é, tendo em vista que os atributos dos quase-identificadores variam de pessoa a pessoa, a combinação pode se tornar suficientemente singular a um único indivíduo. Por exemplo, em um estudo publicado no ano 2000, demonstrou-se que 87% da população dos Estados Unidos da América[4] possuía características provavelmente únicas com base apenas no CEP de cinco dígitos (*5-digit ZIP code*), gênero e data de nascimento.[5]

16. 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. As circunstâncias podem mudar com o tempo e novos desenvolvimentos

[2] Art. 1º, *caput*, da referida lei. BRASIL. **Lei nº 14.534, de 11 de janeiro de 2023.** Altera as Leis nº 7.116, de 29 de agosto de 1983, nº 9.454, de 7 de abril de 1997, nº 13.444, de 11 de maio de 2017, e nº 13.460, de 26 de junho de 2017, para adotar número único para os documentos que especifica e para estabelecer o Cadastro de Pessoas Físicas (CPF) como número suficiente para identificação do cidadão nos bancos de dados de serviços públicos. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/l14534.htm. Acesso em: 09 mai. 2023.

[3] GRUPO DE TRABALHO DE PROTEÇÃO DE DADOS DO ARTIGO 29. **Parecer 4/2007 sobre o conceito de dados pessoais**. Bruxelas: [s. n.], 2007. p. 13. Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2004/wp89_en.pdf Acesso em: 12 mai. 2023.

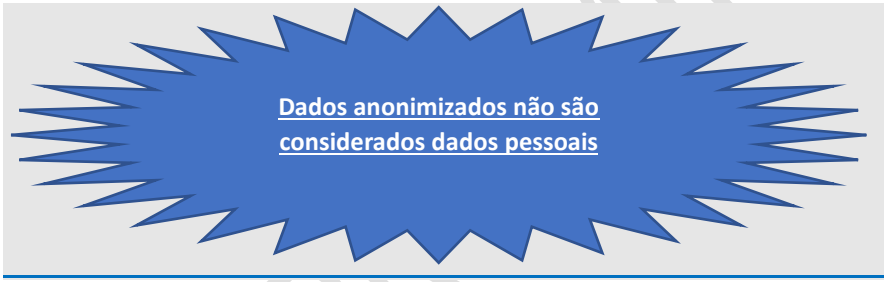
[4] Segundo os números da época, seriam 216 milhões de indivíduos de uma população total de 248 milhões.

[5] SWEENEY, Latanya. **Simple Demographics Often Identify People Uniquely**: Data Privacy Working Paper. Pittsburgh: [s.n.], 2000. Disponível em: https://dataprivacylab.org/projects/identifiability/paper1.pdf Acesso em: 01 mai. 2023.

tecnológicos e a disponibilidade de informações adicionais podem comprometer os processos de anonimização anteriores.

A anonimização não reduz a probabilidade de reidentificação de um conjunto de dados a zero, isto é, a anonimização não elimina todo e qualquer risco de reidentificação de um conjunto de dados; o processo de anonimização e a forma como é implementado terão influência direta na probabilidade de reidentificação.

17. A reidentificação é o processo de tentar discernir os identificadores que foram removidos dos dados desidentificados, inclusive a partir de técnicas de anonimização de dados.[6] Assim, a reidentificação pode transformar dados anonimizados em dados pessoais por meio do uso, por exemplo, de correspondência de dados ou técnicas semelhantes.



18. Os **dados anonimizados não são considerados dados pessoais**, por isso não estão sujeitos à proteção da LGPD, salvo quando o processo de anonimização a que foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.
19. Já o termo **pseudonimização** não é o mesmo que **anonimização**, conforme define a LGPD no § 4º do seu art. 13:

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 (art. 13, § 4º, da LGPD).

20. Ou seja, a anonimização consiste na conversão de dados pessoais em dados que não podem ser usados para identificar qualquer indivíduo. Já no processo de

[6] GARFINKEL, Simson L. **De-Identification of Personal Information**. [S.l.]: National Institute of Standards and Technology, 2015. p. 9.

pseudonimização, é necessário que o dado pessoal seja substituído por identificador ou informação adicional que permita fazer a vinculação entre o dado pseudonimizado e o dado pessoal do seu titular, observando que:

a) essas informações adicionais devem ser mantidas separadamente dos dados pseudonimizados; e

b) devem ser tomadas medidas técnicas e organizacionais de segurança da base de identificadores ou informações adicionais, para garantir que os dados pessoais não sejam atribuídos a um indivíduo.

21. Em diferentes disposições da LGPD há indicações para a aplicação de um dos processos de anonimização ou de pseudonimização. Durante e depois do tratamento dos dados, em situações específicas, no tratamento e utilização de dados pessoais, é aplicável um desses processos para garantir ao titular a proteção contra o uso indevido ou abusivo dos seus dados pessoais.

22. Há recomendação para uso da anonimização e da pseudonimização quando do tratamento de dados pessoais para realização de estudos por órgãos de pesquisa (art. 7º, IV) e no campo da saúde pública (art. 13, *caput*), em casos em que o controlador deseja conservar os dados para uso posterior e como um direito que o titular de dados possui, respectivamente, podendo requerer do controlador a anonimização de seus dados pessoais, quando esta é viável.

Situações e Aplicação das técnicas na LGPD	**Processo**
Condicionante para o tratamento nas hipóteses do uso dos dados pessoais e dados pessoais sensíveis em pesquisas – art. 7º, inciso IV; art. 11, alínea "c" do inciso II;	Anonimização
Reversão do processo de anonimização – art. 12, *caput* e §§ 1º e 3º;	Anonimização
Tratamento de dados sensíveis – estudos e pesquisas em saúde pública – art. 13, *caput* e § 4º.	Pseudonimização
Conservação dos dados após o término do tratamento – *caput* no art. 16, incisos II e IV;	Anonimização
Direito dos titulares no art. 18, inciso IV; **compartilhamento e da portabilidade de dados** - § 6º e 7º do art. 18.	Anonimização

3. OS PROCESSOS DE ANONIMIZAÇÃO E PSEUDONIMIZAÇÃO DE DADOS

23. Os dados pessoais, quando submetidos a processos de anonimização e pseudonimização, passam por alterações que visam a impedir sua associação direta ou indireta a um indivíduo específico. A distinção crucial entre dados anonimizados e pseudonimizados reside na reversibilidade do processo e na

capacidade de reestabelecer a associação com a identidade original do indivíduo.

24. No caso do processo de anonimização, os dados são modificados de tal forma que se reduz substancialmente o risco de vinculá-los novamente a pessoa natural identificada ou identificável, mesmo com o uso de dados auxiliares. A remoção dos identificadores mediante esse processo torna tais dados como não pessoais para qualquer entidade, inclusive para o controlador dos dados.
25. Já na pseudonimização, embora a associação direta seja inicialmente obscurecida, existe a possibilidade de reverter esse processo mediante o uso de informações adicionais mantidas separadamente pelo controlador em um ambiente controlado e seguro. Essas informações adicionais, sob controle estrito, são essenciais para reestabelecer a ligação entre os dados pseudonimizados e a identidade do titular de dados.
26. Ambos os processos buscam atender aos preceitos de proteção da privacidade e de proteção dos dados pessoais. Contudo, a pseudonimização, por permitir a reversibilidade do processo pelo controlador, demanda uma gestão cuidadosa das informações adicionais utilizadas para essa finalidade. É crucial que essas informações sejam mantidas em um ambiente seguro e controlado, evitando qualquer possibilidade de acesso não autorizado que possa comprometer a privacidade dos titulares de dados. Dessa forma, a escolha entre anonimização e pseudonimização dependerá da necessidade de preservação da privacidade e da reversibilidade dos dados no contexto específico de tratamento, considerando a finalidade, a utilidade dos dados e os riscos envolvidos no processo.

3.1. ASPECTOS JURÍDICOS RELEVANTES

3.1.1 Anonimização e os princípios de proteção de dados pessoais

27. A partir da análise do art. 12, *caput*[7], da LGPD, compreende-se que a utilização de meios técnicos na anonimização de dados consiste, na verdade, em um conjunto de atos ou medidas entre si relacionadas que fazem parte de um **processo**. Dessa forma, a anonimização se desenvolve em uma série de etapas que se inicia com o processamento de dados pessoais e tem o objetivo de, com a aplicação de técnicas variadas, desassociar identificadores do dado em seu estado originário ou bruto.[8]

[7] "Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido."

[8] A concepção da anonimização como **processo** tem sido adotada por várias autoridades de proteção de dados, havendo aquelas com estudos já publicados sobre o tema. Para referência, vide: Agência Espanhola de Proteção de Dados. Orientações e garantias nos procedimentos de anonimização de dados pessoais.

28. O objetivo da anonimização é afetar os **identificadores** presentes em um dado, ou conjunto de dados, porque esses são os elementos informativos que "mantém relação particularmente privilegiada e próxima com certo indivíduo." Os identificadores podem ser diretos ou indiretos, como já mencionado anteriormente.

29. Na análise sobre a anonimização, é importante considerar uma premissa adotada pelo regime de proteção de dados pessoais brasileiro[9]: consistindo a anonimização de dados em um processo de remoção de identificadores diretos e indiretos, os dados pessoais submetidos ao processo de anonimização devem ser, na origem, objeto de legítimo tratamento pelo agente responsável.[10]

30. Tal afirmação possui desdobramentos relevantes. Primeiramente, fica evidenciado que o ato inicial do processo de anonimização configura operação de tratamento de dado pessoal, atraindo, assim, a aplicação de princípios e regras da LGPD. O segundo desdobramento é o de que a anonimização não é capaz de *per se* legitimar atividade de tratamento originalmente ilícita por falta de hipótese legal que lhe dê fundamento.

31. Em outras palavras, se todo tratamento de dado pessoal deve ser legitimado por ter suporte normativo em hipótese legal estabelecida previamente, como as previstas nos artigos 7º e 11 da LGPD, a anonimização pressupõe tratamento lícito, pois não é processo capaz de transformar em legítima a irregular atividade de tratamento de dados sem fundamentação legal.

Por exemplo, num contexto de emergência sanitária, um controlador que fornece aplicativo móvel de edição de imagem e texto começa a coletar dados de geolocalização dos dispositivos de seus usuários sem qualquer hipótese legal que legitime sua atividade. Não será eventual anonimização de dados que removerá a ilicitude do tratamento; tais dados deverão, portanto, ser eliminados e o tratamento, interrompido.[14]

32. Sendo assim, por se tratar o ato inicial do processo de anonimização uma operação de tratamento de dado pessoal, deve-se levar em consideração os princípios e regras de proteção de dados pessoais aplicáveis, **em especial os princípios da finalidade, adequação e necessidade.**

2016. p. 5. Disponível em: https://datos.gob.es/es/documentacion/orientaciones-y-garantias-en-los-procedimientos-de-anonimizacion-de-datos-personales Acesso em: 26 jan. 2024.

[9] O art. 3°º, *caput*, da LGPD informa que esta será aplicável "a qualquer operação de tratamento". Neste sentido, a análise sobre a aplicação da LGPD e seus desdobramentos deve ser realizada para cada operação de tratamento.

[10] Nessa direção: GRUPO DE TRABALHO DE PROTEÇÃO DE DADOS DO ARTIGO 29. **Parecer 05/2014 sobre técnicas de anonimização** Bruxelas: [s. n.], 2014. p. 09. Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf Acesso em: 26 jan. 2024.

33. O **princípio da finalidade** estabelece que o tratamento de dados pessoais deverá ser realizado em consonância com propósitos legítimos, explícitos, específicos e informados ao titular quando da operação de tratamento de dados pessoais.[11]

34. É o que prescreve o art. 6º, I, da LGPD. Isso significa que, para a realização da anonimização de acordo com o regime geral de proteção de dados, deve o controlador informar com clareza que uma das finalidades da coleta dos dados pessoais é a futura anonimização[12].

35. Entretanto, se a finalidade de anonimização não houver sido informada originalmente, a sua realização importará "tratamento posterior"[13] ou uso secundário, que, necessariamente, deverá ser compatível com a finalidade inicialmente informada aos titulares dos dados[14].

36. Nessa linha, deve a anonimização, como tratamento posterior, observar o **princípio da adequação**[15], que, por sua vez, determina que a licitude da operação de tratamento depende da sua compatibilidade com a(s) finalidade(s) legítima(s), específica(s) e explicitamente informada(s) ao titular dos dados, levando-se em consideração o contexto em que se realiza o tratamento.

37. De maneira semelhante ao que já foi objeto de recomendação no "Guia Orientativo – Tratamento de dados pessoais pelo Poder Público", a avaliação da compatibilidade da anonimização de dados com a(s) finalidade(s) originária(s) deve ter em consideração, por exemplo:

    I. o contexto da atividade de tratamento de dado pessoais, riscos envolvidos e outras circunstâncias relevantes do caso concreto;

---

[11] Tal como já apontado pela ANPD anteriormente, a finalidade deve ser: "(i) **legítima**, isto é, lícita e compatível com o ordenamento jurídico, além de amparada em uma base legal, que autorize o tratamento; (ii) **específica**, de maneira que a partir da finalidade seja possível delimitar o escopo do tratamento e estabelecer as garantias necessárias para a proteção dos dados pessoais; (iii) **explícita**, isto é, expressa de uma maneira clara e precisa; e (iv) **informada**, isto é, disponibilizada em linguagem simples e de fácil compreensão e acesso ao titular dos dados". AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Guia orientativo – Tratamento de dados pessoais pelo Poder Público.** [S.l.]: ANPD, 2022. p. 13. Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/guia-poder-publico-anpd-versao-final.pdf Acesso em: 17 mar. 2023.

[12] Na mesma direção: Comissão de Proteção de Dados. **Guia sobre Anonimização e Pseudonimização.** [S.l.]: DPC, 2019. p. 13. Disponível em: https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf Acesso em: 26 jan. 2024.

[13] De acordo com o art. 6º, I, da LGPD: "Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de **tratamento posterior** de forma incompatível com essas finalidades [...]" (grifou-se).

[14] Cf. GRUPO DE TRABALHO DE PROTEÇÃO DE DADOS DO ARTIGO 29. **Parecer 05/2014 sobre técnicas de anonimização** Bruxelas: [s. n.], 2014. p. 7-8; DATA PROTECTION COMMISSION. **Guidance on Anonymisation and Pseudonymisation.** [S.l.]: DPC, 2019. p. 13.

[15] LGPD, art. 6º, II. Na tradição do direito de proteção de dados da União Europeia (UE), as noções de "adequação" e "uso compatível" são compreendidas como elementos estruturantes do princípio da finalidade ou da limitação dos propósitos (*purpose limitation principle*) de diversas normativas. Cf. ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 3/2013 on purpose limitation**. Bruxelas: [s. n.], 2013. Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf. Acesso em 17 mar. 2023.

II. a existência de conexão fática ou jurídica entre a finalidade original e os objetivos do processo de anonimização; e
III. as expectativas legítimas dos titulares e os possíveis impactos do tratamento posterior sobre seus direitos.

38. O **princípio da necessidade** é outra norma de alta relevância para a anonimização de dados. De acordo com o art. 6º, III, o tratamento de dados pessoais deverá ser limitado ao mínimo necessário para a realização de suas finalidades, abrangendo apenas os "dados pertinentes, proporcionais e não excessivos em relação às finalidades" especificadas.

39. A necessidade do tratamento da informação exige uma avaliação preliminar direcionada a verificar se o propósito especificado pode ser alcançado com o uso mínimo de dados pessoais ou com métodos idôneos a reduzir ou eliminar seus identificadores. Dessa forma, uma vez cumprida a finalidade para a qual certos dados pessoais foram coletados, a retenção dos dados para exclusivo uso do controlador será possível desde que, à luz do princípio da necessidade, os dados sejam anonimizados[16].

40. Ainda nesse sentido, importa chamar atenção ao fato de que *"a anonimização não é uma medida de segurança impositiva, que deve ser adotada em todo e qualquer tratamento de dados pessoais"*.[17] A pertinência da adoção do processo de anonimização decorre de um juízo de necessidade à luz da(s) finalidade(s) especificada(s) para o tratamento de dados na situação concreta.

3.1.2 Riscos de reidentificação de dados anonimizados

41. O processo de anonimização se desenvolve por meio da utilização de técnicas diversificadas (ver Apêndices II e III) cuja pertinência é justificada de acordo com as características e outros aspectos contextuais do banco de dados que o agente de tratamento pretende anonimizar. Isso porque, além de a LGPD não impor o uso de técnicas de anonimização específicas, não há qualquer metodologia universalmente aplicável.

42. De acordo com o atual estado da arte, pode-se afirmar a existência de um consenso científico sobre a impraticabilidade de um cenário de ausência de risco de reidentificação[18] nas situações de tratamento de dados anonimizados. Tendo em vista o enorme volume de dados auxiliares disponibilizados publicamente na internet e o desenvolvimento da capacidade de processamento e análise de

[16] LGPD, art. 16, IV.
[17] AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Nota Técnica nº 46/2022/CGF/ANPD.** Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/sei_00261-000730_2022_53-nt-46.pdf. Acesso em: 07 ago. 2023; AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Guia Orientativo – Tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas**. Brasília: ANPD, 2023. p. 41.
[18] Denonima-se "risco de reidentificação" o risco de identificação incidente sobre dados anonimizados.

algoritmos de reidentificação, é fundada a afirmativa de que sempre haverá fatores de risco de reidentificação.

43. Nesse sentido, a adoção de **modelo baseado em riscos** relacionado à identificabilidade de dados, a partir dos meios e esforços suscetíveis de serem razoavelmente utilizados, também se mostra pertinente na avaliação da robustez do processo de anonimização. Tal avaliação não pode ser episódica ou pontual, mas sim iterativa e contínua, visto que novos riscos podem advir ao longo do tempo na medida dos avanços tecnológicos e da quantidade de dados auxiliares disponíveis, por exemplo.

44. Os riscos de reidentificação de dados anonimizados são expressos, em linguagem técnica, como possíveis **ataques de reidentificação**. O termo "ataque" é tomado por empréstimo da literatura especializada em segurança computacional, em que a avaliação do nível de segurança de determinado sistema computacional ou algoritmo de cifragem ocorre a partir do uso da figura de um hipotético "atacante" que possui certas habilidades, conhecimento ou acesso.[19] "Uma avaliação de risco envolve a catalogação da variedade de potenciais atacantes, e, para cada um, a probabilidade de sucesso".[20]

45. Cumpre ressaltar que essa noção de "atacante" não se confunde com aqueles sujeitos que praticam crimes ou atos antijurídicos. Basta considerar o exemplo de pesquisadores que avaliam a robustez de base de dados anonimizada compartilhada publicamente frente a certos algoritmos de reidentificação com o uso de dados auxiliares disponibilizados em bases de acesso público.

46. Alguns exemplos de ataques ou riscos de reidentificação que podem ser mencionados são:

    I. a distinção;
    II. a possibilidade de ligação; e
    III. a inferência.

    A distinção consiste na possibilidade de se isolar alguns ou todos os registros que destacam um indivíduo numa base de dados. A possibilidade de ligação é definida pela capacidade de se estabelecer uma conexão entre pelo menos dois registros relativos ao mesmo indivíduo ou ao mesmo grupo de pessoas. Já o risco de inferência diz respeito à possibilidade de inferir, com uma significativa probabilidade, o valor de um atributo a partir dos valores de um conjunto de outros atributos.

[19] A figura do "atacante" muito se aproxima do "intruso" (*intruder*) a que a autoridade de proteção de dados da Irlanda se refere: COMISSÃO DE PROTEÇÃO DE DADOS. **Guia sobre Anonimização e Pseudonimização.** [S.l.]: DPC, 2019. p. 8-10. Disponível em: https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf Acesso em 26 jan. 2026.

[20] GARFINKEL, Simson L. **De-Identification of Personal Information**. [S.l.]: National Institute of Standards and Technology, 2015. p. 9.

3.1.3 As noções de “esforços razoáveis” e “meios próprios”

47. A compreensão do processo de anonimização e dos critérios a serem considerados para avaliar os riscos de reidentificação, requer, necessariamente, a interpretação de dois termos previstos no artigo 12 da LGPD: “esforços razoáveis” e “meios próprios”.[21]

48. O primeiro configura um **conceito jurídico indeterminado** normativo, ou seja, um conceito em larga medida incerto em seu conteúdo e extensão, dependente de preenchimento valorativo pelo aplicador do Direito. Em termos práticos, isso significa que a ANPD, como intérprete e aplicadora da LGPD, deve preencher, com elementos e critérios pertinentes com o caso concreto, a noção de “esforços razoáveis”, dentro do sentido literal possível e em coesão com o contexto significativo da lei, que, aliás, prevê no § 1º do art. 12, relevantes parâmetros interpretativos.

49. A LGPD estabelece no art. 12, § 1º, um **rol exemplificativo** de aspectos objetivos que devem ser avaliados pelo intérprete ao preencher (ou determinar), nas situações concretas, o conteúdo do que é esforço razoável, isto é, dos meios suscetíveis de ser razoavelmente utilizados. Conforme o texto da lei*, “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”.*

50. Na análise dos fatores **custo e tempo** necessários para a possibilidade de reidentificação dos titulares e reversão do processo de anonimização, deve se considerar, por exemplo, os encargos derivados da força de trabalho e recursos humanos, custos econômicos e tempo de dedicação exigidos para se alcançar a reidentificação. Neste sentido, a ANPD já teve a oportunidade de se manifestar em Nota Técnica elaborada no caso envolvendo o tratamento de microdados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP):

> “*A avaliação relativa à eventual reversão dos dados e aos seus impactos deve se basear em evidências e em cenários que considerem aspectos objetivos da realidade. Afastam-se, assim, análises meramente especulativas, baseadas em cenários irreais, de difícil ou improvável ocorrência ou, ainda, que desconsiderem limitações práticas, decorrentes de custos muito elevados ou de meios técnicos de disponibilidade restrita.*” [22]

51. Outros fatores objetivos importantes para a compreensão dos esforços razoáveis para reidentificação ou reversibilidade do processo de anonimização são **as**

[21] “Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido”.

[22] AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Nota Técnica nº 46/2022/CGF/ANPD**. Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/sei_00261-000730_2022_53-nt-46.pdf. Acesso em: 07 ago. 2023.

tecnologias e técnicas disponíveis ao tempo das operações de tratamento e a **licitude** dos meios utilizados. Este último fator implica dizer que a prática de crimes cibernéticos ou o uso de meios proibidos por lei configuram meios e esforços irrazoáveis para a reidentificação ou reversão do processo de anonimização.

52. Diferentemente da noção de "esforços razoáveis", o conceito de **meios próprios** tem conteúdo mais delimitado, podendo-se afirmar que são **meios próprios** as habilidades, os dados, instrumentos e técnicas disponíveis ao próprio agente de tratamento responsável pela anonimização. Sendo assim, importa ressaltar que, a partir do texto normativo do art. 12, *caput*, da LGPD, compreende-se que a avaliação da possibilidade de reidentificação de dados e a reversão do processo de anonimização devem ter em consideração não apenas o uso de meios próprios do agente de tratamento responsável pela anonimização, mas também a atuação de outras pessoas ou entidades que, com meios e esforços razoáveis, podem reidentificar conjunto de dados anonimizados.

3.2. O PROCESSO DE ANONIMIZAÇÃO

53. Os dados pessoais podem ser tratados em diversos formatos, tais como tabular, imagem, áudio e vídeo. Cada um desses formatos apresenta diferentes características que devem ser abordadas por técnicas de anonimização distintas. Por esse motivo, o agente de tratamento não deve considerar a anonimização de forma restrita às técnicas, mas considerar uma abordagem mais ampla baseada em processo, em que as técnicas de anonimização são elementos que compõem o todo.

54. Convém ressaltar que os dados que tenham sido tornados irreversivelmente anonimizados deixam de ser considerados "dados pessoais" e o processamento desses dados não exige conformidade com a legislação de proteção de dados. Isso implica que as organizações podem utilizá-los para finalidades, desde que compatíveis, que vão além daquelas para as quais foram originalmente coletados e esses dados podem ser mantidos indefinidamente.

55. O processo de anonimização, orientado por uma abordagem baseada em riscos, tem como objetivo fornecer um conjunto mínimo de etapas que podem servir de guia de boas práticas aos agentes de tratamentos de dados. Essas etapas sugerem que o agente identifique e compreenda os riscos envolvidos em sua atividade, bem como adote medidas para mitigá-los.

56. A discussão do processo de anonimização é iniciada com a apresentação do conflito entre a utilidade e o grau de anonimização do dado pessoal, seguida por uma importante ponderação sobre a gestão do risco de reidentificação de dados anonimizados por meio de um processo de anonimização baseado em risco.

57. Dentro do âmbito das técnicas de anonimização, este documento apresenta no Apêndice II um caderno com o objetivo de elucidar em quais cenários, contextos

e para qual formato de dado cada técnica abordada se mostra mais adequada. Adicionalmente, são apresentadas suas aplicações e limitações, fornecendo aos responsáveis pela anonimização informações para uma análise criteriosa com o objetivo de identificar a melhor abordagem de acordo com as características específicas e considerações de segurança e privacidade aplicáveis.

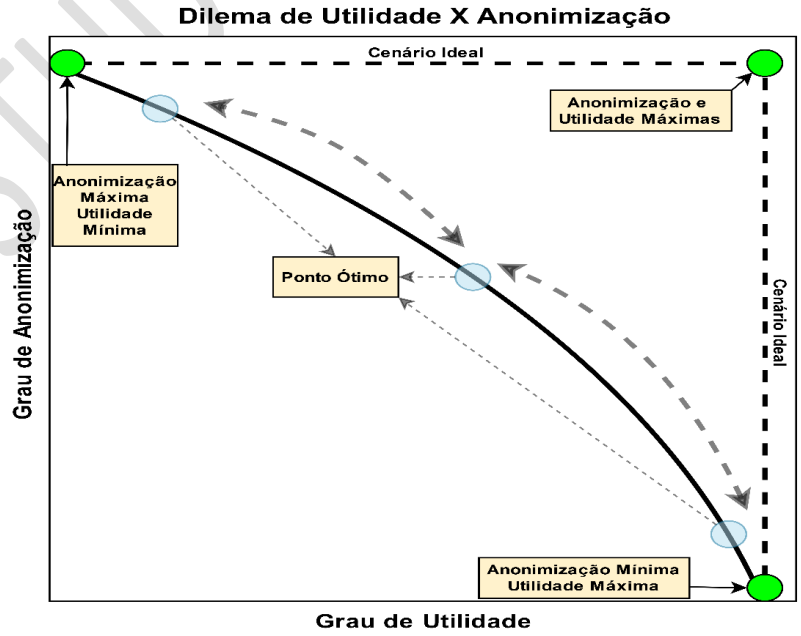
3.2.1 Utilidade dos dados pessoais derivada da finalidade da operação de tratamento

58. A LGPD determina que os dados pessoais devem ser tratados para propósitos legítimos, específicos, explícitos e informados ao titular. Partindo desse enunciado, é possível observar que a atividade de tratamento de dados pessoais precisa estar atrelada a uma finalidade específica, de tal forma que compete ao agente de tratamento identificar o grau de utilidade do dado pessoal para alcançar a finalidade especificada, e em consequência estabelecer o grau necessário de anonimização dos dados.

> **Anonimização não torna os dados inúteis** - um processo adequado de anonimização mantém os dados funcionais para um determinado propósito de tratamento e finalidades específicas.

59. Em termos teóricos, existe um ponto ótimo em que o grau de utilidade do dado pessoal e o grau de anonimização são simultaneamente máximos. Entretanto, em termos práticos esse ponto ótimo não é fácil de ser alcançado, pois depende de um ajuste fino entre duas variáveis conflitantes. Conforme exposto na Figura 1, o ponto ótimo encontra-se em um ponto entre os dois extremos do dilema.

Figura 1: Dilema Utilidade x Anonimização.



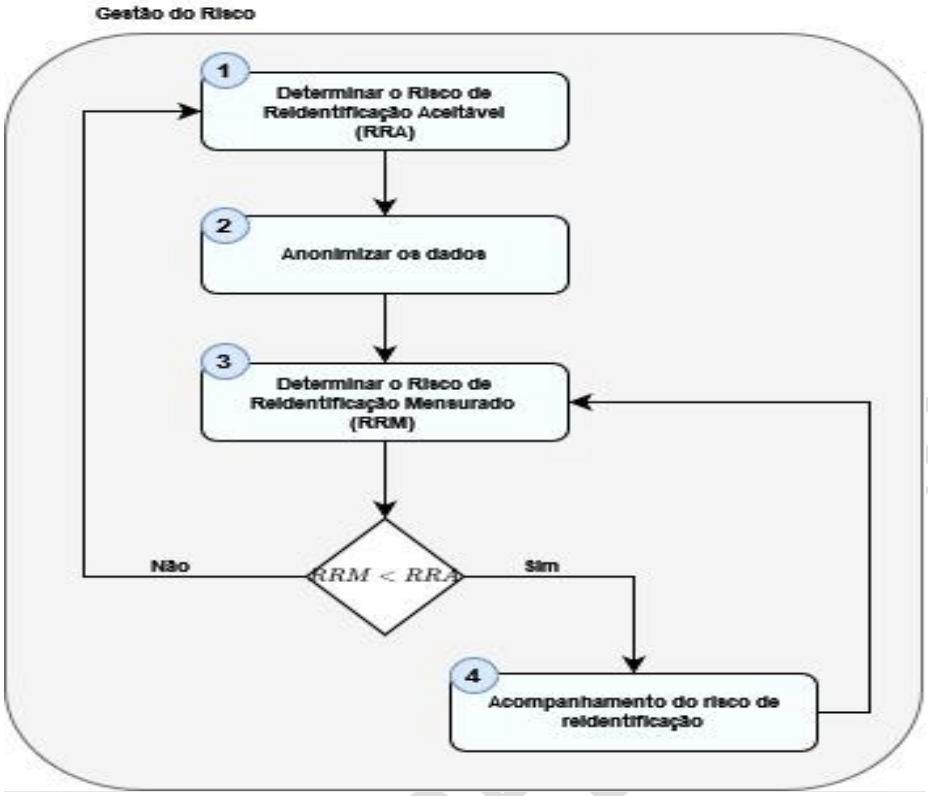
Fonte: Elaboração própria.

60. De tal forma, a abordagem da anonimização como um processo contínuo baseado em risco possibilita que o agente de tratamento defina, de acordo com seu contexto, o compromisso entre o grau de utilidade e o grau de anonimização que contemple a finalidade definida no tratamento e minimize o risco de reidentificação do titular.

3.2.2 Gestão do risco de reidentificação

61. O processo de anonimização de dados pessoais não deve ser entendido como um processo definitivo, em que os dados após a anonimização apresentam risco zero de reidentificação. Corroborando com essa ideia, não há técnica de anonimização com eficácia plena, tendo todas elas um risco de reidentificação associado, cabendo ao agente de tratamento gerenciar esse risco com a adoção de um processo de anonimização adequado.
62. O processo tem como objetivo minimizar os riscos de reidentificação mantendo a utilidade dos dados tratados. Para isso, a gestão do risco de reidentificação deve ser realizada de forma contínua durante todo o tratamento dos dados, permitindo que o agente de tratamento tenha evidências suficientes para a tomada de decisão relacionada à proteção de dados e à privacidade dos titulares.
63. O cenário de anonimização apresenta características que podem variar para a realidade de cada agente de tratamento. Por esse motivo, a anonimização não deve se restringir à discussão de técnicas, mas sim uma abordagem mais ampla baseada em processo. Nesse contexto, este estudo preliminar apresenta uma proposta de processo de anonimização baseado em risco, com 4 etapas, que pode ser adaptado às necessidades de cada agente de tratamento (Figura 2).

Figura 2: Processo de anonimização baseado em risco.



Fonte: Elaboração própria.

64. A **primeira etapa** consiste na determinação do **Risco de Reidentificação Aceitável** (RRA) para um certo conjunto de dados, e tem como objetivo estipular um limite superior para o risco. Um risco de reidentificação superior ao limite estabelecido descaracterizará o conjunto de dados como anonimizado.
65. Essa primeira etapa é de extrema importância e possui uma gama de variáveis dependentes do contexto que devem ser observadas pelo agente de tratamento. Desse modo, não é possível estabelecer uma metodologia padronizada a todos os casos, cabendo ao agente de tratamento definir o risco de reidentificação aceitável para os dados tratados.
66. Pode-se citar como exemplos de variáveis de contexto a existência de dados pessoais sensíveis ou dados financeiros que podem diminuir o limite do risco aceitável.
67. A **segunda etapa** consiste na aplicação do conjunto de técnicas de anonimização escolhido. O objetivo dessa etapa é produzir um conjunto de dados anonimizados que tenha um risco de reidentificação não superior ao limite do risco aceitável definido na etapa anterior. A escolha das técnicas de anonimização deve levar em consideração as características dos dados.
68. A **terceira etapa** consiste em definir o Risco de Reidentificação Mensurado (RRM) de um ataque de reidentificação ter sucesso no conjunto de dados, pós-

anonimização; o RRM preferencialmente deve assumir a forma de probabilidade[23].

69. De modo semelhante à primeira etapa, variáveis dependentes do contexto podem ser observadas pelo agente de tratamento, como exemplo tem-se a condição do conjunto de dados ser público, compartilhado ou privado. Essa condição pode afetar o risco real de reidentificação.
70. Considerando a diversidade de natureza, escopo, contexto e finalidade de cada tratamento realizado pelo agente de tratamento, não é possível definir uma métrica única para a mensuração do risco de reidentificação. Por tal razão, este Guia adota a expressão **Métrica Contextual** para se referir à métrica utilizada para mensurar o RRM de acordo com a realidade de cada agente de tratamento.
71. É importante destacar que para dados textuais estruturados há algumas métricas de mensuração de risco bem conhecidas, tais como a K-Anonimização[24], T-Proximidade[25] e L-Diversidade[26]. Essas métricas derivam de uma métrica base que utiliza o conceito de equivalência de classe da teoria dos conjuntos para determinar o risco de reidentificação. Inclusive, a K-Anonimização é exemplificada no Apêndice III.
72. A métrica de risco de reidentificação pode ser computada para cada um dos titulares pertencentes ao conjunto de dados, e os valores resultantes podem ser ponderados, por exemplo, com a média aritmética, para determinar o valor geral da métrica contextual. Por fim, o valor geral da métrica contextual pode ser então ponderado pelas variáveis contextuais, resultando no Risco de Reidentificação Mensurado.

$$\textit{Risco de Reidentificação Mensurado} = \boldsymbol{\theta} * \boldsymbol{V_C}$$

73. $\boldsymbol{\theta}$ representa o valor geral da métrica contextual e $\boldsymbol{V_C}$ representa um fator de ponderação das variáveis contextuais, quando existentes, caso não existam $\boldsymbol{V_C}$ pode assumir o valor de 1.
74. Por exemplo, no contexto de base de dados públicas ou compartilhadas, o risco deve ser majorado e, consequentemente, o valor de V_C deve ser definido de forma adequada a representar a majoração do risco.
75. O Risco de Reidentificação Mensurado (RRM) deve ser comparado ao Risco de Reidentificação Aceitável (RRA). Caso o RRM seja maior do que o RRA, o conjunto de dados não apresenta a condição de estar anonimizado, sendo necessário o reinício do processo de anonimização. Caso contrário, é necessário acompanhar o uso do conjunto de dados, especialmente quando operações realizadas sobre ele possam modificar o risco mensurado, tais como operações de inclusão,

[23] Uma probabilidade é um valor real no intervalo fechado de 0 a 1.

[24] SAMARATI, P.; SWEENEY. L. **Protecting privacy when disclosing information: k-anonymity and tis enforcement through generalization and suppression**. Technical Report, 1998

[25] LI N.; LI, T.; VENKATASUBRAMANIAN S. **T-Closeness: Privacy Beyond k-Anonymity and L-Diversity**. IEEE 23rd International Conference on Data Engineering, 2007.

[26] AGGARWAL, Charu C.; YU, Philip S. **A general Survey of Privacy-Preserving Data Mining Models and Algorithms**. Privacy-Preserving Data Mining. Advances in Database System. vol 34. Springer. 2008.

alteração ou deleção de dados; havendo essas operações é necessário atualizar o nível do risco mensurado.

3.3. O PROCESSO DE PSEUDONIMIZAÇÃO

76. A pseudonimização de dados pessoais significa substituir quaisquer características identificáveis dos dados por um pseudônimo, ou seja, um valor que não permite a identificação direta do titular dos dados. A LGPD define a pseudonimização como o tratamento de dados pessoais de forma que esses dados não possam mais ser atribuídos a um titular de dados específico sem o uso de informações adicionais, desde que:
    a) essas informações adicionais sejam mantidas separadamente; e
    b) estejam sujeitas a medidas técnicas e organizacionais para garantir que os dados pessoais não sejam atribuídos a um indivíduo identificado ou identificável.
77. Embora a pseudonimização tenha diversas utilidades, é importante distingui-la da anonimização, já que aquela oferece apenas uma proteção limitada à identidade dos titulares em muitos casos, permitindo ainda a identificação por meio de métodos indiretos. Quando se utiliza um pseudônimo, a depender da técnica utilizada, é possível identificar o titular por meio da análise dos dados subjacentes ou relacionados, o que deve ser tratado com atenção à luz dos princípios da LGPD.
78. Em certos casos, a natureza dos dados, o contexto em que são utilizados ou o propósito da coleta e retenção podem tornar a anonimização eficazmente impossível. Mesmo nessas circunstâncias, as organizações podem optar por empregar técnicas de anonimização ou pseudonimização com as seguintes finalidades, de acordo com a LGPD:
    a) Como parte de uma estratégia de “privacidade desde a concepção” (*privacy by design*) destinada a oferecer uma proteção adicional aos titulares dos dados;
    b) Como parte de uma estratégia de minimização de riscos ao compartilhar dados com operadores ou outros controladores de dados;
    c) Para evitar violações acidentais quando a equipe tem acesso a informações pessoais; e
    d) Como parte de uma estratégia de "minimização de dados" voltada a reduzir os riscos de violações de dados para os titulares dos dados.
79. Vale ressaltar que, mesmo após a anonimização, persistem alguns riscos inerentes e que a pseudonimização não é equivalente à anonimização, uma vez que as informações ainda mantêm sua característica de dados pessoais.
80. A Lei Geral de Proteção de Dados Pessoais não indica técnicas de pseudonimização específicas, mas estabelece princípios e requisitos gerais para o tratamento de dados pessoais.

81. No entanto, a pseudonimização é um conceito amplamente reconhecido na LGPD e é encorajada como uma medida de proteção de dados. Algumas técnicas de pseudonimização em conformidade com a LGPD podem ser:

a) **Substituição de Dados:** nesta técnica, dados pessoais são substituídos por pseudônimos ou códigos, tornando-os menos identificáveis. Por exemplo, um número de CPF pode ser substituído por um código alfanumérico único.

b) **Ofuscação de Dados:** envolve a transformação de dados pessoais de forma que sejam mais difíceis de identificar. Isso pode incluir o embaralhamento de informações ou a substituição de valores de dados por outros valores semelhantes.

c) **Tokenização:** envolve a substituição de dados pessoais por tokens ou códigos que não têm significado fora do contexto do sistema. Esses tokens podem ser usados para fins de identificação, mas não revelam as informações reais dos titulares de dados.

d) **Cifração:** técnica que converte dados em um formato criptografado que só pode ser decifrado com uma chave. Isso torna os dados pessoais ilegíveis para qualquer pessoa sem acesso à chave correspondente.

e) **Mascaramento de Dados:** envolve a ocultação parcial de informações, revelando apenas uma parte dos dados e ocultando o restante. Por exemplo, um número de telefone pode ser mascarado como "(XX) XXXX-1234", mostrando apenas os últimos dígitos.

f) **Salting:** técnica é comumente usada na criptografia de senhas. Um valor aleatório (chamado de "salt") é adicionado aos dados antes da encriptação, tornando os pseudônimos únicos e mais seguros contra-ataques de força bruta.

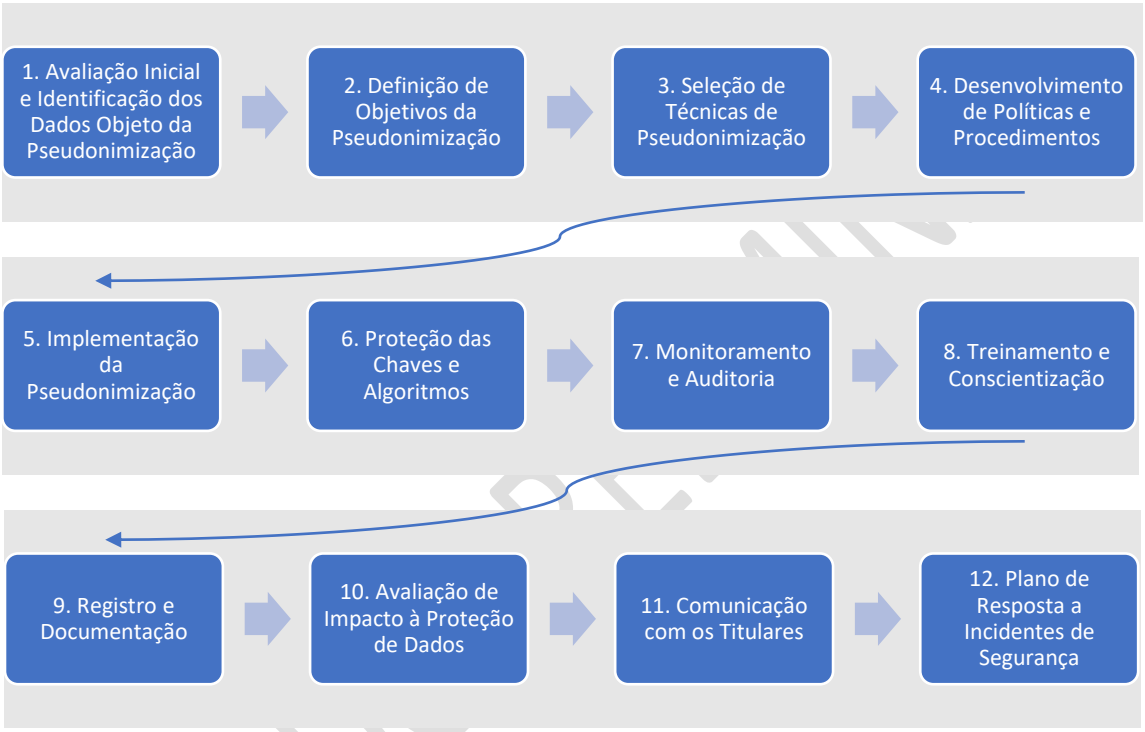
Criptografia típica não é anonimização – Criptografia é uma técnica de pseudonimização. Como a informação original precisa estar acessível, as transformações aplicadas pelos algoritmos criptográficos são projetadas para serem reversíveis, no que é conhecido como descriptografia. Entretanto, vários algoritmos criptográficos (simétricos, assimétricos e de hash) podem realizar processamentos unidirecionais. Nesses casos atendem os requisitos da anonimização, desde que os dados cifrados sejam úteis.

82. É importante observar que a LGPD enfatiza que, para que a pseudonimização seja eficaz, as informações adicionais que permitem a reversão da pseudonimização (por exemplo, as chaves criptográficas) devem ser mantidas separadamente e protegidas por medidas técnicas e organizacionais adequadas. Além disso, a LGPD enfatiza a importância de garantir a privacidade e a segurança dos dados pessoais em todas as etapas do tratamento. Portanto, a escolha da técnica de

pseudonimização deve ser feita com cuidado, levando em consideração o contexto, os riscos associados e a sensibilidade dos dados.

83. Desenvolver uma metodologia eficaz de pseudonimização de dados pessoais, alinhada com as melhores práticas de mercado e em conformidade com os princípios da LGPD é fundamental para garantir a privacidade e a segurança das informações pessoais.

Figura 3: Metodologia Eficaz de Pseudonimização.



Fonte: Elaboração própria.

84. Conforme ilustração acima (Figura 3), para o desenvolvimento dessa metodologia algumas etapas devem ser observadas:

1. Avaliação Inicial e Identificação dos Dados Objeto da Pseudonimização: inicie com uma avaliação abrangente de quais dados pessoais serão coletados e tratados. Identifique quais dados pessoais serão objeto da pseudonimização, considerando os riscos e o tratamento realizado, dando ênfase a dados considerados sensíveis, como por exemplo, dados de saúde, origem racial ou étnica, convicção religiosa, opinião política, entre outros.

2. Definição de Objetivos da Pseudonimização: estabeleça claramente os objetivos da pseudonimização, incluindo a proteção da privacidade do titular dos dados, a redução do risco de violações de dados e o cumprimento da LGPD.

3. Seleção de Técnicas de Pseudonimização: escolha as técnicas de pseudonimização apropriadas com base na natureza dos dados. Isso pode incluir

o mascaramento de informações pessoais, o uso de tokenização, o embaralhamento de dados ou a criptografia, dentre outros. A escolha dependerá das características específicas dos dados e dos riscos associados.

4. Desenvolvimento de Políticas e Procedimentos: crie políticas e procedimentos claros para garantir a pseudonimização adequada. Isso inclui diretrizes sobre como realizar a pseudonimização, armazenar chaves criptográficas de forma segura e garantir a rastreabilidade e o acesso somente a pessoal autorizado.

5. Implementação da Pseudonimização: implemente as técnicas de pseudonimização de acordo com as políticas e procedimentos estabelecidos. Certifique-se de que todos os dados pessoais sejam adequadamente pseudonimizados antes de serem armazenados ou processados. Em alguns casos, técnicas diferentes podem ser aplicadas, concomitantemente, para produzir uma pseudonimização eficiente.

6. Proteção das Chaves e Algoritmos: garanta que as chaves e algoritmos utilizados no processo de pseudonimização, como por exemplo, chaves criptográficas, senhas de acesso a sistemas ou a arquivos, códigos-fonte, dentre outros, sejam armazenadas de forma segura e acessíveis apenas a pessoal autorizado. Os registros de auditoria devem ser mantidos, documentando quando as chaves foram usadas, quem as utilizou e com que finalidade. Isso é valioso para conformidade regulatória, registro de operações e investigações de segurança. É fundamental garantir que os dados possam ser revertidos quando necessário de forma segura, pelo controlador.

Como uma boa prática para o gerenciamento de chaves, técnicas como a implementação de logs de eventos e sistemas de monitoramento podem ser empregados para facilitar a rastreabilidade no uso das chaves, e ainda, as chaves podem ser armazenadas de forma segura usando práticas como a criptografia de chaves mestras e Módulos de Segurançaem Hardware (HSMs).

7. Monitoramento e Auditoria: implemente sistemas de monitoramento e auditoria para verificar continuamente a eficácia da pseudonimização e garantir o cumprimento das políticas e procedimentos. Realize revisões e auditorias regulares a fim de acompanhar as mudanças regulatórias, tecnológicas e melhores práticas de mercado relacionadas à pseudonimização e ajuste sua metodologia conforme necessário.

8. Treinamento e Conscientização: forneça treinamento regular aos colaboradores que lidam com dados pessoais para garantir que compreendam a importância da pseudonimização e saibam como aplicá-la corretamente.

9. Registro e Documentação: mantenha registros detalhados de todas as atividades de pseudonimização, incluindo datas, técnicas utilizadas, responsáveis e propósitos. Isso é importante para fins de prestação de contas, rastreabilidade e registro de operações.

10. Avaliação de Impacto à Proteção de Dados: realize a avaliação de impacto sobre a proteção de dados, elaborando o Relatório de Impacto à Proteção de Dados Pessoais (RIPD) quando apropriado, a fim de avaliar os riscos associados à pseudonimização e garantir a conformidade com a LGPD. Considere a elaboração do RIPD sempre que o tratamento envolver alto risco.

11. Comunicação com os Titulares: esteja preparado para informar de forma transparente e acessível aos titulares sobre a pseudonimização e os direitos de acesso e correção de suas informações pessoais, conforme exigido pela LGPD.

12. Plano de resposta a Incidentes de Segurança: desenvolva um plano de resposta a incidentes de segurança com dados pessoais que inclua procedimentos para lidar, entre outras situações, com acessos não autorizados e tratamentos inadequados ou ilícitos, incluindo as ações de mitigação apropriadas para reverter ou mitigar os efeitos dos prejuízos gerados.

4. CONSIDERAÇÕES FINAIS

85. Com a publicação do presente estudo preliminar, a ANPD pretende manter sua postura estratégica de promover na sociedade brasileira maior efetividade do regime de proteção de dados pessoais, fornecendo esclarecimentos e orientações em linha com o atual contexto socioeconômico e tecnológico do país.
86. Busca-se orientar os agentes de tratamento sobre a anonimização e pseudonimização de dados como processos contínuos com base em abordagem de riscos, e não somente limitar-se à indicação de aplicação de técnicas. Assim, esclarece-se que, dada a velocidade do progresso tecnológico, disponibilização de dados auxiliares e sofisticação de possíveis ataques, torna-se necessário manter a segurança e o cuidado com os processos.
87. Nos processos de anonimização e pseudonimização não existe uma solução única que se adeque a todas as organizações. Na maioria dos casos, não é possível fornecer recomendações mínimas de parâmetros a serem usados e cada organização deve, portanto, utilizar os mecanismos e técnicas que sejam apropriadas para as suas circunstâncias.
88. O presente documento não objetiva esgotar o tema da anonimização e pseudonimização no contexto da proteção de dados. Ao contrário, lança as bases para a expansão das orientações da ANPD para fortalecer a cultura e a proteção de dados pessoais. Sendo assim, contamos com as colaborações e sugestões sobre questões importantes que, porventura, não tenham sido tratadas ou que precisem de mais esclarecimentos.

5. REFERÊNCIAS

AGGARWAL, Charu C.; YU, Philip S. A general Survey of Privacy-Preserving Data Mining Models and Algorithms. Privacy-Preserving Data Mining. Advances in Database System. vol 34. Springer. 2008.

AEPD-EDPS, AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS – *European Data Protection Supervisor. Misunderstandings Related to Anonymization*. Disponível em: <https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en

AEPD-EDPS, AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS - European Data Protection Supervisor. *Orientaciones y garantias en los procedimentos de anonimización de datos personales.* [S.l.]. AEPD, 2016.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

BRASIL. Lei nº 9.784, de 29 de janeiro de 1999, que regula o processo administrativo no âmbito da Administração Pública Federal. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9784.htm.

BRASIL. Lei nº 14.534, de 11 de janeiro de 2023. Altera a Lei nº 7.116, de 29 de agosto de 1983, 9.454, de 7 de abril de 1997, a Lei nº 13.444, de 11 de maio de 2017, e Lei nº 13.460, de 26 de junho de 2017, para adotar número único para os documentos que especifica e para estabelecer o Cadastro de Pessoas Físicas (CPF) como número suficiente para identificação do cidadão nos bancos de dados de serviços públicos. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/l14534.htm

_____ AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). Guia Orientativo para Definições dos Agentes de Tratamento de Dados Pessoais e do Encarregado. Brasília: ANPD, versão 2.0., abr. 2022. Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/Segunda_Versao_do_Guia_de_Agentes_de_Tratamento_retificada.pdf

_____AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). Portaria nº 1, de 8 de março de 2021. Disponível em: https://www.in.gov.br/en/web/dou/-/portaria-n-1-de-8-de-marco-de-2021-307463618

_____AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). Resolução CD/ANPD nº 2, de 27 de janeiro de 2022; Disponível em: https://www.in.gov.br/en/web/dou/-/resolucao-cd/anpd-n-2-de-27-de-janeiro-de-2022-376562019#wrapper

_____AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. Guia orientativo – Tratamento de dados pessoais pelo Poder Público. [S.l.]: ANPD, 2022. Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/guia-poder-publico-anpd-versao-final.pdf

_____AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. Nota Técnica nº 46/2022/CGF/ANPD. Disponível em: https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/sei_00261-000730_2022_53-nt-46.pdf

_____AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). Resolução CD/ANPD nº 4, de 24 de fevereiro de 2023. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-regulamento-de-dosimetria/Resolucaon4CDANPD24.02.2023.pdf

_____AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS (ANPD). Guia Orientativo – Tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas. Brasília: ANPD, 2023.

DATA PROTECTION COMMISSION. *Guidance on Anonymisation and Pseudonymisation* [S.l.]: DPC, 2019. Disponível em: https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf

GARFINKEL, Simson L. *De-Identification of Personal Information*. [S.l.]: National Institute of Standards and Technology, 2015.

GRUPO DE TRABALHO DE PROTEÇÃO DE DADOS DO ARTIGO 29. *Opinion 4/2007 on the concept of personal data.* Bruxelas: [s. n.], 2007. Disponível em: https://www.clinicalstudydatarequest.com/Documents/Privacy-European-guidance.pdf

GRUPO DE TRABALHO DE PROTEÇÃO DE DADOS DO ARTIGO 29. *Opinion 3/2013 on purpose limitation.* Bruxelas: [s. n.], 2013. Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

GRUPO DE TRABALHO DE PROTEÇÃO DE DADOS DO ARTIGO 29. *Opinion 5/2014 on Anonymisation Techniques.* Bruxelas: [s. n.], 2014. Disponível em: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

LI N.; LI, T.; VENKATASUBRAMANIAN S. **T-Closeness: Privacy Beyond k-Anonymity and L-Diversity**. IEEE 23rd International Conference on Data Engineering, 2007.

SAMARATI, P.; SWEENEY. L. **Protecting privacy when disclosing information: k-anonymity and tis enforcement through generalization and suppression**. Technical Report, 1998.

SWEENEY, Latanya. *Simple Demographics Often Identify People Uniquely: Data Privacy Working Paper*. Pittsburgh: [s.n.], 2000. Disponível em: https://dataprivacylab.org/projects/identifiability/paper1.pdf

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 sobre a Proteção de Dados – RGPD). Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679

6. APÊNDICES

APÊNDICE I – PRINCIPAIS ESCLARECIMENTOS [27]

a) **Dados anonimizados não são considerados dados pessoais**, por isso não estão sujeitos à proteção da LGPD, salvo quando o processo de anonimização a que foram submetidos for revertido, por algum meio eficaz.

b) A determinação do que seja **esforço razoável deve levar em consideração fatores objetivos, tais como o custo e o tempo necessários para reverter o processo** de anonimização, de acordo com as tecnologias disponíveis no momento da anonimização, e a utilização exclusiva de recursos tecnológicos próprios do agente de tratamento.

c) A anonimização de dados não é perpétua e indeterminada. **Existem riscos de que o processo de anonimização possa ser revertido no futuro**. As circunstâncias podem mudar com o tempo e os **novos desenvolvimentos tecnológicos e a disponibilidade de informações adicionais podem comprometer** os processos de anonimização anteriores.

d) Visando minimizar os possíveis impactos de um incidente de segurança, recomenda-se a **adoção de técnicas de pseudonimização e, quando cabível, a anonimização que busque a irreversibilidade, ou criem maior dificuldade de reidentificação**, desincentivando a tentativa de sua reversão.

e) **Pseudonimização não é o mesmo que anonimização -** 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 (art. 13, § 4º, da LGPD).

f) **Criptografia típica não é anonimização** - criptografia é uma técnica de pseudonimização. Como a informação original precisa estar acessível, as transformações aplicadas pelos algoritmos criptográficos são projetadas para serem reversíveis, no que é conhecido como descriptografia. Entretanto, vários algoritmos criptográficos (simétricos, assimétricos e de hash) podem realizar processamentos unidirecionais. Nesses casos atendem os requisitos da anonimização, desde que os dados cifrados sejam úteis.

g) **A anonimização dos dados pessoais nem sempre será possível** - nem sempre é possível reduzir o risco de reidentificação abaixo de um limite previamente definido, mantendo um conjunto de dados útil para um processamento ou finalidade específica.

h) **A anonimização não é para sempre** - existe o risco de que alguns processos de anonimização possam ser revertidos no futuro. As circunstâncias podem mudar com o tempo e novos desenvolvimentos tecnológicos e a disponibilidade de informações adicionais podem comprometer os processos de anonimização anteriores.

i) **A anonimização, geralmente, não reduz a probabilidade de reidentificação de um conjunto de dados a zero** - a anonimização não impossibilita a reidentificação de um

[27] Inspirado na publicação "10 Misunderstanding Related to Anonymisation" da AEPD/EDPS. Disponível em: https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf

conjunto de dados; o processo de anonimização e a forma como é implementado terão influência direta na probabilidade de riscos de reidentificação.

j) **A anonimização não é um conceito binário e dependendo do processo utilizado pode ser medida** - é possível analisar e medir o grau de anonimização, por meio de técnicas usadas para garantir que o limite de risco de reidentificação não seja ultrapassado, como parte da metodologia de anonimização.

k) **A anonimização não deve ser totalmente automatizada -** ferramentas automatizadas podem ser usadas durante o processo de anonimização, no entanto, dada a importância do contexto e a avaliação geral do processo, poderá ser necessária a intervenção de um especialista humano.

Anonimização não torna os dados inúteis - um processo adequado de anonimização mantém os dados funcionais para um determinado propósito de tratamento e finalidades específicas.

l) **Seguir um processo de anonimização que outros usaram com sucesso nem sempre levará a organização a resultados equivalentes** - seguir um caso de sucesso é um excelente ponto de partida, mas não é garantia de sucesso quando aplicado a outros casos.

m) Os **processos de anonimização precisam ser adaptados à natureza, escopo, contexto e finalidade do processamento -** bem como aos riscos e às variáveis probabilísticas e gravidade para os direitos e liberdades individuais.

n) **Existem riscos e interesses em reindentificar os dados anonimizados -** os dados pessoais têm um valor em si, para os próprios indivíduos e para terceiros. A reidentificação de um indivíduo pode ter um sério impacto sobre seus direitos e liberdades, assim a possibilidade de se reidentificar uma pessoa em um conjunto de dados, seja por curiosidade, por acaso ou motivado por um interesse real, como por exemplo, para pesquisa científica, fins jornalísticos ou atividade criminosa, não pode ser desconsiderada.

APÊNDICE II. CADERNO DE TÉCNICAS PARA ANONIMIZAÇÃO E PSEUDONIMIZAÇÃO

01. Os dados pessoais anonimizados e pseudonimizados não possuem uma associação, direta ou indireta, a um indivíduo não sendo possível a sua identificação. A diferença primordial entre dados anonimizados e pseudonimizados se remete ao fato de que na pseudonimização essa associação pode ser reestabelecida pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro. Neste sentido, algumas técnicas que serão apresentadas podem ser empregadas tanto para a anonimização quanto para pseudonimização de dados pessoais. O que difere neste caso é a possibilidade de reversibilidade do processo pelo controlador com o uso de meios próprios e informações adicionais, mantidos sob o seu controle, na ocasião do tratamento. A seguir serão apresentadas algumas técnicas, exemplificativas e não exaustivas, para anonimização e pseudonimização de dados pessoais, textuais estruturados.

TÉCNICAS PARA ANONIMIZAR DADOS TEXTUAIS ESTRUTURADOS

Técnica de Adição de Ruído

Descrição

A técnica consiste em realizar pequenas modificações nos dados originais adicionando ruído nos dados. Normalmente utilizada em dados numéricos.

Exemplo

Dado original

Nome Completo	Idade	Altura	Peso	Qtd. Filhos
FM	30	1,55	53	2
AFB	36	1,60	67	2
LB	20	1,80	92	0
MTL	22	1,68	52	1
CGG	44	1,71	61	3
RJ	27	1,75	72	1

Dado anonimizado por meio da adição de ruído.
Ao valor original é somando 1 desvio-padrão do intervalo de valores.
Para as colunas Idade e Quantidade filhos o valor do ruído foi trucando.

Nome Completo	Idade	Altura	Peso	Qtd. Filhos
FM	37	1,65	71,63	2
AFB	43	1,70	85,63	2
LB	27	1,90	110,63	0
MTL	29	1,78	70,63	1
CGG	37	1,62	71,63	2
RJ	43	1,70	85,63	2

Aplicação

Apropriada para dados numéricos em cenários em que a precisão dos dados não é essencial para o alcance da finalidade pretendida. A adição de ruído pode fazer com que o conjunto de dados perda suas propriedades estatísticas e o invalide para a finalidade pretendida.

Limites

- Aplicável preferencialmente em dados numéricos.
- Perda da precisão dos dados, a adição de ruído pode descaracterizar os dados com a perda de sua utilidade.
- Não deve ser utilizada quando a precisão dos dados é essencial.

Técnica de Generalização

Descrição

A técnica agrupa os dados com características em comum em um nível de granularidade maior. Os valores dos atributos são substituídos pelos valores do grupo.

Exemplo

Dado original

Nome Completo	Cidade de Nascimento	Idade
FM	Rio Branco	79
AFB	Macaíba	63
LB	Natal	91
MTL	Xapuri	85
CGG	Macaé	34
RJ	Rio de Janeiro	66

Dado anonimizado por meio da generalização

Nome Completo	Estado de Nascimento	Faixa Etária
FM	Acre	70-79
AFB	Rio Grande do Norte	60-69
LB	Rio Grande do Norte	90-99
MTL	Acre	80-89
CGG	Rio de Janeiro	30-39
RJ	Rio de Janeiro	60-69

Aplicação

Apropriada quando os dados possuem características em comum que permitem sua representação de forma generalizada, sem perda da utilidade.

Limites

- Aplicável somente a dados textuais estruturados.
- Aplicável somente em dados que compartilham características em comum.
- Não aplicável quando houver dados com valores únicos.
- Perda da precisão dos dados.
- Custo computacional elevado para aplicação em dados textos estruturados em fluxo

Técnica de Mascaramento

Descrição

A técnica consiste em substituir uma parte dos caracteres dos dados por um caractere símbolo (por exemplo * ou x).

Exemplo

Dado original

Nome Completo	CPF	Idade
FM	111.111.111-11	79
AFB	222.222.222-22	63
LB	333.333.333-33	91
MTL	444.444.444-44	85
CGG	555.555.555-66	34
RJ	666.666.666-66	66

Dado anonimizado por meio do mascaramento

Nome Completo	CPF	Idade
FM	***.111.111-**	79
AFB	***.222.222-**	63
LB	***.333.333-**	91
MTL	***.444.444-**	85
CGG	***.555.555-**	34
RJ	***.666.666-**	66

Aplicação

Apropriada quando a substituição de parte dos dados por um caractere simbólico fornece o nível desejado de anonimização, sem perda da utilidade.
No caso do CPF o mascaramento tem sido bastante aplicado, mas não deve ser empregado para qualquer finalidade ou qualquer contexto. Quando são substituídos cinco dígitos, a mascará obtida poderá ser válida para 100000 titulares de dados.

Limites

- Aplicável somente a dados textuais.
- Perda da precisão dos dados.
- Custo computacional elevado para aplicação em dados textuais estruturados em fluxo.

- Escolha do padrão adequado que permita desvincular o titular do dado anonimizado.
- Em casos de dados públicos ou compartilhados, como não há um padrão para o mascaramento, é possível que partes distintas dos dados estejam visíveis e por consequência os dados originais sejam reconstruídos.

Técnica de Permutação

Descrição

A técnica consiste em reorganizar os valores dos dados dentro do conjunto de dados, de tal forma que os valores originais ainda são representados, mas geralmente não mais associado ao seu titular.

Exemplo

Dado original

Nome Completo	Profissão	Tempo de Profissão
FM	Professor	20
AFB	Vendedor	12
LB	Advogado	15
MTL	Engenheiro de Software	8
CGG	Enfermeiro	25
RJ	Médico Veterinário	12

Dado anonimizado por meio da permutação

Nome Completo	Profissão	Tempo de Profissão
FM	Advogado	15
AFB	Professor	20
LB	Enfermeiro	25
MTL	Vendedor	12
CGG	Médico Veterinário	12
RJ	Engenheiro de Software	8

Aplicação

Apropriada somente quando a análise dos dados precisa ser feita de forma agregada, pois a técnica elimina a possibilidade de analisar os dados ao nível do titular.

Limites

- Aplicável somente a dados textuais estruturados.
- Aplicável somente para análise agregada.
- Perda da precisão dos dados
- Custo computacional elevado para aplicação em dados textos estruturados em fluxo.

Técnica de Supressão

Descrição

A técnica consiste em excluir registros ou partes deles. A exclusão pode ser realizada em identificadores ou em partes dos registros.

Exemplo

Dado original

Nome Completo	Profissão	Tempo de Profissão
FM	Professor	20
AFB	Professor	12
LB	Professor	15
MTL	Professor	8
CGG	Enfermeiro	25
RJ	Professor	12

Dado anonimizado por meio da supressão.

Profissão	Tempo de Profissão
Professor	20
Professor	12
Professor	15
Professor	8
Professor	25

Aplicação

Apropriada somente quando a exclusão dos dados não afete a qualidade do tratamento a ser realizado ou o impossibilite. Registros com características únicas, isto é, com alto grau de unicidade, podem ser excluídos sem afetar a qualidade do conjunto de dados.

Limites

- A exclusão de registros pode afetar o conjunto de dados.
- Perda da informação excluída.

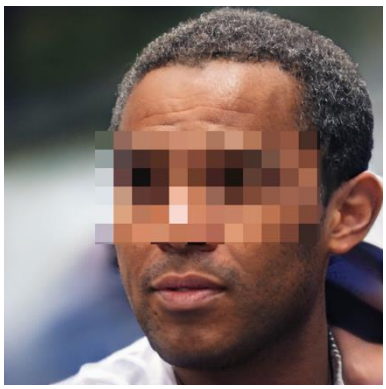
TÉCNICAS PARA ANONIMIZAR IMAGENS

Técnica de Desfoque Gaussiano (*blur*)

Descrição

A técnica consiste em aplicar um filtro de convolução nos *pixels* com o objetivo de desfocar uma área de interesse na imagem.

Dado Original	Dado Anonimizado com a técnica de desfoque gaussiano.
	
Aplicação Apropriada para dados de imagem ou vídeo em que se deseja desfocar regiões de interesse, em geral faces, para minimizar o risco de identificação do titular.	
Limites • Aplicável em dados de imagens ou frame de vídeo. • Dificuldade de definir o limite dos parâmetros do desfoque para garantir a utilidade do dado e ao mesmo tempo preservar a privacidade. • Dificuldade em identificar quais regiões da imagem o desfoque deve ser aplicado para garantir a utilidade do dado e ao mesmo tempo preservar a privacidade.	

Técnica de *Pixelização*	
Descrição A técnica consiste diminuir a resolução da imagem, ou em uma área de interesse dessa para reduzir a nitidez da imagem.	
Dado Original	Dado Anonimizado com a técnica de *pixelização*.
	

Aplicação

Apropriada para dados de imagem ou vídeo em que se deseja *pixelização* regiões de interesse, em geral faces, para minimizar o risco de identificação do titular.

Limites

- Aplicável em dados de imagens ou frame de vídeo.
- Dificuldade de definir o limite dos parâmetros da *pixelização* para garantir a utilidade do dado e ao mesmo tempo preservar a privacidade.
- Dificuldade em identificar quais regiões da imagem a *pixelização* deve ser aplicada para garantir a utilidade do dado e ao mesmo tempo preservar a privacidade.

TÉCNICAS PARA PSEUDONIMIZAR DADOS TEXTUAIS ESTRUTURADOS

Técnica de Substituição por Contador

Descrição

A técnica consiste em substituir os identificadores por códigos únicos. É imprescindível que os códigos utilizados não se repitam para evitar ambiguidades e que os códigos não tenham relação com o identificador.

Exemplo

Dado original

Nome Completo	CPF	Idade
FM	111.111.111-11	79
AFB	222.222.222-22	63
LB	333.333.333-33	91
MTL	444.444.444-44	85
CGG	555.555.555-66	34
RJ	666.666.666-66	66

Dado pseudonimizado por meio de substituição por contador.

Nome Completo	CPF	Idade
FM	9000	79
AFB	9001	63
LB	9002	91
MTL	9003	85
CGG	9004	34
RJ	9005	66

Aplicação

Apropriada para conjunto de dados simples e possui fácil implementação.

Limites

- A escalabilidade da técnica para grandes conjuntos de dados é limitada.
- Necessidade de armazenar uma tabela auxiliar de mapeamento entre o identificador e o valor pseudonimizado.

Técnica de Função *Hash*

Descrição

A técnica consiste em aplicar uma função matemática que recebe como entrada o dado pessoal em sua forma original e o mapeia para um valor de saída na forma de um dado pseudonimizado. O dado de entrada pode ter um tamanho arbitrário e a função de mapeamento deve objetivar ser irreversível e livre de colisões.

Exemplo

Dado original

Nome Completo	Tempo de Profissão
FM	20
AFB	12
LB	15
MTL	8
CGG	25
RJ	12

Dado pseudonimizado por meio de função criptográfica

Nome Completo	Tempo de Profissão
965940fc76dd718d0000c10f964a31ab	20
5353c821d9b43ee6c394ab8cbdf007c4	12
fc1df2e704dbf009cb911d3a645fa6f0	15
b96ff5b76a82dc8e188964e2b66c1c93	8
c7aff31b290c0c21ac13a61f92cf7298	25
126a0eb057128bfc8b342507c1311aa4	12

Aplicação

A técnica contribui significativamente para a integridade dos dados, porém é considerada uma técnica sujeita à ataques de força bruta e de dicionário. Ambos os ataques consistem em testar as entradas possíveis para a função *hash* e verificar qual delas produz o valor *hash* (pseudonimizado) correspondente.

O ataque de força bruta testa todas as entradas possíveis, para *hashs* grandes desconsidera as colisões, tendo em vista a baixíssima probabilidade de ocorrência. Portanto, a força bruta somente é aplicável quando se conhece os candidatos aos dados em claro.

Por sua vez, o ataque de dicionário testa somente entradas pré-selecionadas, extraídas de um dicionário que contém as entradas com maior probabilidade de acerto.

Limites

- Sensível ao ataque de força bruta.
- Sensível ao ataque de dicionário.
- Necessidade de armazenar uma tabela auxiliar de mapeamento entre o identificador e o valor pseudonimizado.

Técnica de Encriptação

Descrição

A técnica consiste em converter os dados pessoais em um formato criptografado que só pode ser decifrado com a chave utilizada para criptografar, nos algoritmos simétricos, ou com a chave correspondente nos algoritmos assimétricos. A encriptação torna os dados pessoais ilegíveis sem o conhecimento da chave correspondente.

Exemplo

Dado original

Nome Completo	Tempo de Profissão
FM	20
AFB	12
LB	15
MTL	8
CGG	25
RJ	12

Dado pseudonimizado por meio de encriptação com chave de 128 bits.

Nome Completo	Tempo de Profissão
SzVEcrP6uX8yy1MaWSYC8Q==	20
lLAUoq0cWfkzTDN2+rddSQ==	12
EkLLOZImXFL2lzLzHCb9YQ==	15
Nvo1OFYnDKuZmQ56NNK5CQ==	8
NCXsMdEJUrIZmRVHszmiCQ==	25
iEl9JafMW0spZEFHRaxu7A==	12

Aplicação

A técnica de encriptação é uma técnica de pseudonimização considerada robusta e pode ser utilizada em bases de dados extensas, pois não depende de uma tabela auxiliar para o mapeamento entre o identificador e o valor pseudonimizado.

Limites

- Robustez depende do sigilo da chave utilizada para a encriptação e da chave correspondente para decriptação.

APÊNDICE III – TÉCNICAS DE MENSURAÇÃO DE RISCO PARA DADOS TEXTUAIS ESTRUTURADOS

K-ANONIMIZAÇÃO

01. A K-Anonimização é uma métrica de mensuração de risco de reidentificação para dados textuais estruturados derivada do conceito de equivalência de classe da teoria dos conjuntos. Apesar das limitações, é uma métrica de fácil compreensão e implementação.
02. Como parte do processo de anonimização, a K-Anonimização analisa se cada registro compartilha dados anonimizados com ao menos K-1 outros registros. Caso essa condição não seja alcançada, o conjunto de dados deve ser novamente anonimizado.
03. A Tabela 1 apresenta o conjunto de dados antes da anonimização que contém informações sobre funcionários de uma empresa. As técnicas de anonimização utilizadas são supressão, generalização e permutação. O valor definido para K é igual 2.

Tabela 1: Exemplo de conjunto de dados antes da anonimização.

Registro Profissional	Tempo de Serviço	Profissão	Gênero
3693	11	Engenheiro Civil	M
7807	28	Programador	M
6026	15	Encanador	M
0872	24	Eletricista	F
3164	20	Artífice	F
5190	19	Engenheiro Elétrico	F
4845	28	Programador	M
5867	22	Engenheiro de Software	M
9881	15	Encanador	M
3528	13	Engenheiro Civil	F
4442	21	Ajudante	F

04. A Tabela 2 apresenta o conjunto de dados após a anonimização considerando a métrica K-Anonimização com K = 2. Ao identificador Registro Profissional foi aplicada a técnica de supressão e para os identificadores Tempo de Serviço e Profissão foi aplicado a técnica de generalização. Por último, para o identificador Gênero foi aplicada a técnica de permutação
05. Os subconjuntos dos registros que compartilham os dados com ao menos um outro registro estão identificados por cores na Tabela 2.

Tabela 2 Exemplo de conjunto de dados após anonimização com valor de K = 2 para a K-Anonimização.

Tempo de Serviço	Profissão	Gênero
10 a 20	Obras	M
21 a 30	Tecnologia da Informação	F
10 a 20	Obras	M
21 a 30	Obras	F
10 a 20	Obras	M
10 a 20	Obras	M
21 a 30	Tecnologia da Informação	F
21 a 30	Tecnologia da Informação	F
10 a 20	Obras	M
10 a 20	Obras	M
21 a 30	Obras	F

APÊNDICE IV. ESTUDO DE CASOS

Caso 1: Dados agregados de localização – Supressão

01. A fim de adotar decisões informadas para combater emergência sanitária causada por epidemia de certa doença infectocontagiosa, a Secretaria Estadual de Saúde de um Estado-membro necessita de dados confiáveis a respeito da localização dos seus cidadãos. Os dados de localização são relevantes para que as autoridades estaduais do sistema de saúde pública possam identificar aglomerações de pessoas e, assim, orientarem-se para a implementação de medidas de prevenção, controle e fiscalização sanitárias de forma mais eficaz. A restrição ao ajuntamento das pessoas é medida importante para conter e diminuir o índice de contágio, além de identificar tendências de movimentação.
02. Com base na legislação instituidora de política pública de vigilância epidemiológica, discussões internas e consulta a especialistas, o governo estadual firmou acordo com os provedores de serviço de telefonia móvel A, B e C, para ter acesso a dados agregados de localização dos celulares dos respectivos usuários, com limites fixados à circunscrição territorial do Estado-membro e à duração da emergência sanitária.
03. Sendo assim, os provedores ou operadoras de telefonia móvel A, B e C compartilharam dados dos aparelhos de telefonia móvel conectados às Estações Rádio Base – ERBs. Cada aparelho de telefonia móvel envia para a ERB a que estão conectados a Identidade Internacional do Assinante Móvel (*Internation Mobile Subscriber Indentify* – IMSI) e a Identidade Internacional do Equipamento Móvel (*Internation Mobile Equipament Identify* – IMEI). Esses dados permitem que essas operadoras consigam identificar quais usuários estão conectados em quais ERBs em um determinado momento. Entretanto, para alcançar o objetivo da Secretaria Estadual de Saúde é necessário conhecer somente o quantitativo de usuários conectados em cada ERB em certo marco temporal.
04. Para resguardar a privacidade dos titulares de linhas móveis e atender ao interesse público, as operadoras de telefonia móvel, ao compartilharem os dados com a Secretaria de Saúde, aplicaram a técnica de supressão dos dados IMSI e IMEI, além de realizar a agregação do quantitativo de telefones móveis a fim de permitir o cálculo do índice de isolamento ou mapas de calor. Para tanto, consideraram-se: (i) o total de 21.641.000, o número de celulares somados os clientes das operadoras A, B e C no território do Estado; e (ii) a localização a partir das antenas (Estações Rádio Base – ERBs) às quais os dispositivos móveis estavam conectados.

Caso 2: Dados clínicos para pesquisa acadêmica – Supressão e Pseudonimização

01. Em estudo de dados clínicos de pacientes conduzido por grupo de pesquisadores de determinado Hospital das Clínicas de uma universidade federal, os dados relacionados à pressão arterial de 100 pacientes foram coletados nos atendimentos realizados com intervalo de 7 (sete) dias. Os dados coletados estão dispostos na Tabela 3.

Tabela 3. Dados coletados

Nome Completo	CPF	Endereço	Gênero	Idade	Peso	PD 1	PS1	PD 2	PS 2
Johanne Mendonça	111.111.111-11	Rua Norte, 372, Bairro A	M	51	113,30	13	9	15	7
Araci Coutinho Silva	222.222.222-22	Rua Leste, 122, Bairro A	F	46	48,50	10	6	12	9
Marcela Antunes	333.333.333-33	Rua dos Cocos, 7, Bairro B	F	37	97,44	10	7	11	7
Madruga Neves	444.444.444-44	Rua das Mangas, 22, Bairro B	M	41	59,28	14	7	11	8
Florinda Neves	555.555.555-55	Rua das Mangas, 22, Bairro B	F	58	54,30	11	7	11	7
Nilce Cavalcante	666.666.666-66	Rua Marte, 1, Bairro C	F	57	110,33	15	6	12	8
José Francisco	777.777.777-77	Rua Vênus, 36, Bairro C	M	73	58,55	18	10	17	10
Carmélia Andrade	888.888.888-88	Rua Vênus, 812, Bairro C	F	56	54,42	12	7	12	7
Andreia Priscila	999.999.999-99	Rua Sol, 12, Bairro C	F	35	109,38	17	10	16	9
...		...	...	...	...	...	...	...	...

02. Os pesquisadores submeteram esse conjunto de dados pessoais a processo de anonimização, tendo em vista que, conforme o desenho metodológico da pesquisa, a utilidade dos dados obtidos a partir da aplicação de certas técnicas de anonimização é preservada para os objetivos do estudo. Nesse sentido, foram aplicadas as técnicas expostas na Tabela 4.

03. Cumpre ressaltar, ainda, que os dados anonimizados serão mantidos em ambiente com controle de acesso e com pertinentes medidas de segurança previstas na política de segurança da informação do órgão de pesquisa.

Tabela 4. Técnicas utilizadas por Identificador.

Identificador	Técnica Utilizada	Descrição
Nome Completo	Supressão	Identificador direto é suprimido.
CPF	Pseudonimização	Substituição do CPF por um código único gerado.
Endereço	Supressão	O identificador é suprimido, pois não é útil para atender ao objetivo do tratamento.
Gênero		O processo de anonimização deve considerar a utilidade do dado para o tratamento desejado. No presente caso, os dados de gênero, peso, pressão diastólica 1, pressão sistólica 1, pressão diastólica 2 e pressão sistólica 2 estão correlacionados e essa correlação é útil para a finalidade da coleta de dados. Aplicação de técnicas de anonimização pode impactar na correlação dos dados e reduzir a utilidade deles.
Peso		O processo de anonimização deve considerar a utilidade do dado para o tratamento desejado. No presente caso, os dados de gênero, peso, pressão diastólica 1, pressão sistólica 1, pressão diastólica 2 e pressão sistólica 2 estão correlacionados e essa correlação é útil para a finalidade da coleta de dados. Aplicação de técnicas de anonimização pode impactar na correlação dos dados e reduzir a utilidade deles.
Pressão Diastólica 1		O processo de anonimização deve considerar a utilidade do dado para o tratamento desejado. No presente caso, os dados de gênero, peso, pressão diastólica 1, pressão sistólica 1, pressão diastólica 2 e pressão sistólica 2 estão correlacionados e essa correlação é útil para a finalidade da coleta de dados. Aplicação de técnicas de anonimização pode impactar na correlação dos dados e reduzir a utilidade deles.
Pressão Sistólica 1		O processo de anonimização deve considerar a utilidade do dado para o tratamento desejado. No presente caso, os dados de gênero, peso, pressão diastólica 1, pressão sistólica 1, pressão diastólica 2 e pressão sistólica 2 estão correlacionados e essa correlação é útil para a finalidade da coleta de dados. Aplicação de técnicas de anonimização pode impactar na correlação dos dados e reduzir a utilidade deles.
Pressão Diastólica 2		O processo de anonimização deve considerar a utilidade do dado para o tratamento desejado. No presente caso, os dados de gênero, peso, pressão diastólica 1, pressão sistólica 1, pressão diastólica 2 e pressão sistólica 2 estão correlacionados e essa correlação é útil para a finalidade da coleta de dados. Aplicação de técnicas de anonimização

		pode impactar na correlação dos dados e reduzir a utilidade deles.
Pressão Sistólica 2		O processo de anonimização deve considerar a utilidade do dado para o tratamento desejado. No presente caso, os dados de gênero, peso, pressão diastólica 1, pressão sistólica 1, pressão diastólica 2 e pressão sistólica 2 estão correlacionados e essa correlação é útil para a finalidade da coleta de dados. Aplicação de técnicas de anonimização pode impactar na correlação dos dados e reduzir a utilidade deles.

Caso 3: Compartilhamento de dados educacionais – Supressão, generalização, mascaramento, adição de ruídos e permutação[28]

01. A Secretaria Municipal de Educação da cidade de Privacinópolis precisa compartilhar os dados dos alunos matriculados com a Secretaria Municipal de Assistência Social com o objetivo da construção de relatórios sociais. Os dados estão dispostos na Tabela 5.

Tabela 5: Dados tratados

Nome Completo	Matrícula	Idade	Endereço	Gênero	Renda Familiar (R$)
Johanne Mendonça	2023010	7	Rua Norte, 372 – Bairro A	M	2188,44
Araci Coutinho Silva	2023011	10	Rua Leste, 122 – Bairro A	F	2195,82
Marcela Antunes	2023020	8	Rua dos Cocos, 7 - Bairro B	F	1947,20
Madruga Neves	2023021	9	Rua das Mangas, 22 - Bairro B	M	2014,38
Florinda Neves	2023022	11	Rua das Mangas, 22 - Bairro B	F	1942,34
Nilce Cavalcante	2023030	12	Rua Marte, 1 - Bairro C	F	1856,08
José Francisco	2023031	12	Rua Vênus, 36 - Bairro C	M	1835,86
Carmélia Andrade	2023032	8	Rua Vênus, 812 - Bairro C	F	1989,66
Andreia Priscila	2023033	10	Rua Sol, 12 - Bairro C	F	2082,96
Bruno da Costa	2023040	13	Rua Mercúrio, 36 - Bairro C	M	1911,34

[28] O presente estudo de caso visa exemplificar o processo de anonimização proposto na seção 3.2 deste estudo preliminar.

02. Para tanto, se faz necessário conhecer o resumo dos dados tratados (Tabela 6):

Tabela 6: Descrição dos dados

Dado	Tipo	Dado Pessoal	Dado Pessoal Sensível	Identificador Direto	Descrição Estatística
Nome Completo	Qualitativo	S	N	S	Não Aplicável
Matrícula	Qualitativo	S	N	S	Não Aplicável
Idade	Quantitativo	S	N	N	Média: 10 Mediana: 10 Desvio-Padrão: 1,89
Endereço	Qualitativo	S	N	N	Não Aplicável
Gênero	Qualitativo	S	N	N	Moda: F Frequência M: 4/10 Frequência F: 6/10
Renda Familiar	Quantitativo	N	N	N	Média: R$ 1996,41 Mediana: R$ 1968,43 Desvio-Padrão: R$ 119,34 Mínimo: R$ 1835,85 Máximo: R$ 2195,82

03. Considerando o processo proposto neste estudo preliminar (Seção 3.2), há 4 etapas essenciais para a gestão do risco de reidentificação.

Determinar o Risco de Reidentificação Aceitável (RRA): É importante observar que a mensuração do risco de reidentificação é uma etapa que deve ser executada e gerenciada pelo agente de tratamento de acordo com o caso concreto, conforme sugerido no documento de Estudo Técnico sobre Anonimização de Dados na LGPD: Processo de Anonimização Baseado em Risco e Técnicas de Anonimização – Uma Introdução Computacional. No presente estudo de caso, nenhum dos dados tratados é considerado como sendo dado pessoal sensível e o compartilhamento dos dados é feito com outro órgão público por meios próprios. Entretanto, os dados são de crianças e adolescentes. De tal forma, o Risco de Reidentificação Aceitável (RRA) é definido em 0,35.

Anonimizar os dados: A Tabela 7**Erro! Fonte de referência não encontrada.** apresenta as técnicas utilizadas em cada um dos dados tratados. Por sua vez, a Tabela 8**Erro! Fonte de referência não encontrada.** apresenta o conjunto de dados após a aplicação do conjunto de técnicas de anonimização.

Tabela 7: Técnicas utilizadas por Identificador.

Identificador	Técnica Utilizada	Descrição
Nome Completo	Supressão	Identificador direto que será suprimido, a matrícula será utilizada.

Matrícula	Mascaramento	Os dois primeiro e o último digito será substituído por *.
Idade	Generalização	Os dados serão agrupados por duas faixas etárias. 1ª_ ≤ 10 e 2ª >10
Endereço	Generalização	Os dados serão agrupados pelo bairro do endereço.
Gênero	Permutação	Os valores serão trocados entre os gêneros, porém mantendo a frequência de cada gênero e a moda do conjunto de dados.
Renda Familiar	Adição de Ruído e Generalização	Cada valor individual será deslocado um desvio-padrão à direita e posteriormente generalizado em duas faixas de renda: ≤ R$ 2.000,00 e > R$ 2.000,00

Tabela 8: Identificadores após aplicação do conjunto de técnicas de anonimização.

Matrícula	Idade	Endereço	Gênero	Renda Familiar (R$)
**2301*	≤ 10	Bairro A	F	> 2.000,00
**2301*	≤ 10	Bairro A	M	> 2.000,00
**2302*	≤ 10	Bairro B	F	> 2.000,00
**2302*	≤ 10	Bairro B	F	> 2.000,00
**2302*	> 10	Bairro B	M	> 2.000,00
**2303*	> 10	Bairro C	F	≤ 2.000,00
**2303*	> 10	Bairro C	M	≤ 2.000,00
**2303*	≤ 10	Bairro C	F	> 2.000,00
**2303*	≤ 10	Bairro C	M	> 2.000,00
**2304*	> 10	Bairro C	F	> 2.000,00

Risco de Reidentificação Mensurado (RRM): O processo indica que após a aplicação do conjunto de técnicas de anonimização é necessário mensurar o risco de reidentificação utilizando alguma métrica contextual.

04. Nesse estudo, optou-se por utilizar a K-Anonimização, métrica derivada da equivalência de classe. Conforme sugerido no processo, a métrica deve ser computada para cada um dos identificadores e os valores resultados ponderados para determinar o valor geral do risco mensurado de reidentificação (Tabela 9).

Tabela 9: Risco Mensurado de Reidentificação.

Identificador	K-Anonimização por Classe do Identificador	K-Anonimização do Identificador (Média da K-Anonimização por Classe do Identificador)

Matrícula	**2301* = $\frac{1}{2}$ =0,50 **2302* = $\frac{1}{3}$ = 0,33 **2303* = $\frac{1}{4}$ = 0,25 **2304* = $\frac{1}{1}$ = 1,00	0,52
Idade	≤ 10 = $\frac{1}{6}$ = 0,16 > 10 = $\frac{1}{4}$ = 0,25	0,20
Endereço	Bairro A = $\frac{1}{2}$ = 0,50 Bairro B = $\frac{1}{3}$ = 0,33 Bairro C = $\frac{1}{5}$ = 0,20	0,34
Gênero	F = $\frac{1}{6}$ = 0,16 M = $\frac{1}{4}$ = 0,24	0,20
Renda Familiar (R$)	> 2.000,00 = $\frac{1}{8}$ = 0,12 ≤ 2.000,00 = $\frac{1}{2}$ = 0,50	0,31
Métrica Contextual (Média da K-Anonimização do Identificador)		0,31

05. No caso em estudo, não foram identificadas variáveis contextuais que impactem significativamente no risco de reidentificação, sendo o fator de ponderação definido em 1,00. Conforme proposto no processo, o Risco Mensurado de Reidentificação é o valor resultante da ponderação entre as variáveis contextuais e a métrica contextual, no exemplo 1,00*0,31 = 0,31.
06. O Risco de Reidentificação Mensurado calculado é de 0,31, enquanto o Risco de Reidentificação Aceitável é de 0,35. De tal forma, o conjunto de dados após a aplicação do conjunto de técnicas de anonimização tem um risco de reidentificação menor do que o risco aceitável.
07. De acordo com o processo proposto, é necessário acompanhar o risco mensurado de reidentificação para que ele sempre se mantenha abaixo do risco de reidentificação aceitável.