

Prefeitura da Cidade do Rio de Janeiro
Conselho Municipal de Proteção de Dados Pessoais e da Privacidade

Boas Práticas em Anonimização e Pseudonimização para o Poder Público Carioca

Prefeitura da Cidade do Rio de Janeiro
Conselho Municipal de Proteção de Dados Pessoais e da Privacidade

Título do Grupo de Trabalho: GT 4 - Boas Práticas em Anonimização e
Pseudonimização para o Poder Público Carioca

Composição do Grupo de Trabalho:

Theonácio Carvalho
Rafael Moraes
Antônio Sérgio
William Rocha

Relator:
Fernando Felipe Bourguy de Medeiros

Sumário

Introdução _____	03
Apresentação do problema _____	04
Benckmarck / Boas Práticas _____	06
Recomendações _____	29
Conclusão _____	31

1. Introdução

No contexto dos trabalhos do Conselho Municipal de Privacidade e Proteção de Dados Municipais do Rio de Janeiro, o Grupo de Trabalho 4 apresenta seu relatório de Boas Práticas em Anonimização e Pseudonimização para o Poder Público Carioca. Este trabalho buscará orientações nacionais e internacionais acerca de práticas de de-identificação.

No segundo capítulo, será feita a apresentação do problema. Nela, o relatório identificará dificuldades legais e práticas no manejo de dados pessoais, que justifiquem a busca por soluções recomendadas por entidades renomadas.

O terceiro capítulo analisará as práticas aconselhadas por diversas entidades, notadamente o Comitê Europeu para Proteção de Dados, o National Institute for Standards and Technology dos Estados Unidos da América, e as autoridades nacionais de proteção de dados de Singapura, Reino Unido e Brasil.

Finalmente, baseado nas práticas das entidades acima, o quarto capítulo apresentará uma lista de recomendações estratégicas e operacionais, a serem adotadas por órgãos municipais de diferentes hierarquias.

Com este relatório, espera-se que o Poder Público carioca adquira conhecimentos suficientes à elaboração de planejamentos estratégicos na condução de processos de anonimização em suas diferentes atividades. Dessa forma, tende-se a aumentar o grau de segurança dos órgãos municipais em relação ao cumprimento da Lei Federal n. 13.709/18 e, por conseguinte, a confiança da população carioca nos serviços públicos por eles prestados.

2. Apresentação do problema

O Município depende de grandes quantidades de dados em suas atividades, documentados em diferentes suportes (e.g. documentos físicos, bases de dados digitais etc.). Parte relevante desses *conjuntos de dados* contém *dados mistos*, isto é, dados não-pessoais e dados pessoais. Levando em conta que apenas estes estão sujeitos à Lei Federal n. 13.709/18 – LGPD (BRASIL, 2018), o Município deve ser capaz de manusear ambas as espécies de forma diferente, para otimizar os recursos disponíveis. Nesse sentido, a anonimização e a pseudonimização podem ser técnicas úteis.

Entre suas obrigações, a LGPD exige que o tratamento de dados pessoais seja realizado apenas onde sua finalidade for adequada a hipóteses pré-estabelecidas (arts. 6º, I, 7º e 11). Onde a finalidade não existir ou se exaurir, os dados pessoais devem ser eliminados (art. 15), salvo exceções (art. 16). Dentre elas, destaca-se a anonimização dos dados pessoais. O art. 12 da LGPD prevê que a lei não incide sobre dados anonimizados, pois a anonimização os torna impessoais¹, sob requisitos rígidos. Nesse sentido, a anonimização se torna uma alternativa útil à eliminação aos órgãos municipais, pois logra manter a utilidade de conjuntos de dados cujo manuseio seria ilícito.

Não obstante, a anonimização é uma alternativa tecnicamente complexa. Para alcançá-la, a *reidentificação*, ou seja, a reversão da anonimização, deve ser inviável, utilizando-se meios próprios ou os esforços razoáveis disponíveis, considerando ainda o tempo e as tecnologias à mão no momento da anonimização. Para cumprir esses requisitos, não basta apagar um nome de uma lista. É necessária uma análise planejada e detida sobre a degradação da precisão do conjunto de dados *vis-à-vis* a utilidade que se quer extrair do conjunto de dados anonimizado. Uma anonimização muito forte poderia inviabilizar o uso dos dados, porém, uma muito fraca poderia aumentar o risco de reidentificação. Não bastasse isso, há casos em que os requisitos se tornam até impossíveis, e.g. se no Município há um único caso de certa doença rara, jamais será possível impossibilitar a reidentificação do doente no grau exigido pela LGPD. Consequentemente, os órgãos municipais, como qualquer agente de tratamento, dependem de orientações sobre como realizar anonimizações eficazes.

Embora o art. 12, §3º da LGPD preveja que tais orientações serão editadas pela Agência Nacional de Proteção de Dados – ANPD, até o momento, ela não dispôs sobre padrões e técnicas aplicáveis. Nesse cenário, o Município

¹ A previsão acompanha a diretriz do Considerando 26 do General Data Protection Regulation da União Europeia, que considera dados anonimizados como não-pessoais e, portanto, fora do escopo da norma. V. <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>. O mesmo entendimento expressou o Information Commissioner's Office britânico ("ICO"), em seu Guidance on Anonymization. V. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/>.

permanece num vácuo normativo imobilizador, que lhe impõe tratar dados não pessoais com cuidado e gastos desproporcionais ou tratar dados pessoais sem o cuidado previsto pela LGPD, ambas hipóteses ilegais. Diante do impasse, faz-se necessária a busca por boas práticas aptas a suprir a omissão da ANPD e capacitar os órgãos municipais ao saneamento seguro das bases de dados municipais, abrindo o leque de alternativas de tratamento ao Poder Público.

Ademais, ainda que a ANPD edite regulação a esse respeito, a anonimização é reconhecida como um processo aberto, que admite diferentes soluções conforme o caso concreto.² Nesse contexto, conclui-se a partir do princípio da prestação de contas (LGPD, art. 6º, X) que, para evitar responsabilização administrativa e civil, os referenciais de anonimização e pseudonimização adotados devem ser documentados, a fim de produzir trilha de auditabilidade; bem como eles devem estar consonantes às políticas de segurança da informação existentes, a fim de reforçar a governança respectiva. Por conseguinte, convém ao órgão municipal conhecer boas práticas a serem adotadas em cada caso, para garantir sua eficácia.

Por outro lado, enquanto nem sempre convenha anonimizar dados pessoais, deve-se sempre prezar pela sua minimização e segurança (LGPD, art. 6º, III e VII). Isso atrai a necessidade de buscarem-se meios de evitar expor mais dados pessoais do que os necessários em cada atividade dos órgãos municipais. Daí convir buscar soluções para minimizar o tratamento ao mesmo tempo em que se preserva sua utilidade, o que pode envolver técnicas de pseudonimização.

Em suma, se o Município pretende aprimorar a prestação de seus serviços, na medida em que a população adota inovações tecnológicas, deve buscar métodos eficientes para planejar e manter processos de anonimização eficazes. Portanto, a pesquisa de melhores práticas de anonimização e pseudonimização podem habilitar o Poder Público carioca a manejar suas bases de dados de maneira simplificada, econômica e responsável.

² “Several anonymization techniques may be envisaged, there is no prescriptive standard in EU legislation. – Importance should be attached to contextual elements (...)” WP29. Opinion 05/2014 on Anonymisation Techniques, 10.04.2014, p. 6, in https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

3. Benckmarck / Boas Práticas

3.1 Definições

A uniformização da terminologia parece dever ser o primeiro passo para o sucesso da adoção de novos processos em qualquer organização, mormente uma tão complexa quanto o Município do Rio de Janeiro. Ao equalizar-se a comunicação, os *stakeholders* responsáveis pela deliberação, adoção e execução das medidas recomendadas adiante poderão se entender mutuamente.

A terminologia abaixo foi extraída e traduzida a partir do Guia de Técnicas Básicas de Anonimização da autoridade nacional de Singapura (PDPC, 2024)³, da Opinião 05/2014 sobre Técnicas de Anonimização do WP29, adotado pelo Comitê Europeu para Proteção de Dados (WP29, 2014)⁴ e de outras fontes, além de propostas por esta relatoria.

- i) Adversário: parte que tenta reidentificar um titular a partir de um conjunto de dados anonimizado.
- ii) Anonimização: Conversão do dado pessoal em dado anonimizado, pela aplicação de técnicas de anonimização.
- iii) Atributo: Trata-se de uma informação contida no registro de um titular. Também pode ser referida como “campo”, “coluna”, “variável” ou “categoria de dado”, e.g. nome, gênero ou endereço.
- iv) Classe de Equivalência: registros num conjunto de dados que compartilham os mesmos valores em um mesmo atributo, normalmente, em identificadores indiretos. E.g., num conjunto de dados sobre moradores dos bairros do Rio de Janeiro, todas as pessoas que moram em Madureira pertencem a mesma classe de equivalência.
- v) Conjunto de Dados: Um agrupamento de dados, e.g., um banco de dados documentado em planilha, com registros (linhas) e atributos (colunas), físico ou digital.
- vi) Conjunto de Dados Anonimizado: O conjunto de dados resultante das técnicas de anonimização.

³ V. [guide-to-basic-anonymisation-\(updated-24-july-2024\).pdf](#)

⁴ V. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

- vii) Conjunto de Dados Original: o conjunto de dados anterior à aplicação das técnicas de anonimização.
- viii) De-Identificação: Técnica pela qual o dado pessoal perde a capacidade de ser vinculado a uma pessoa identificada. A de-identificação é gênero, do qual a anonimização e a pseudonimização são espécies.⁵
- ix) Identificabilidade vs. Reidentificabilidade: grau mediante o qual uma pessoa pode ser identificada por meio de um ou mais conjuntos de dados que contêm identificadores diretos e indiretos, contra o grau pelo qual uma pessoa pode ser identificada a partir de um conjunto anonimizado.
- x) Identificador Direto: um atributo que, isoladamente, identifica um titular (ex.: impressão digital) ou que foi atribuído exclusivamente a um titular (ex.: CPF).⁶ Também pode ser referido como “identificador único”.
- xi) Identificador Indireto: Também conhecidos como “quase-identificadores”. Trata-se de um atributo que, por si só, não identifica um titular, mas pode identificá-lo quando combinado com outros dados. Enquanto um identificador indireto ou não-identificadores estiverem vinculado a um identificador direto, todos eles serão dados pessoais. Um exemplo de identificador indireto é o atributo “Prefeito do Rio de Janeiro”. Por si só, não identifica o titular, mas unido ao ano de mandato, é possível identificá-lo. Outro exemplo é o endereço de IP. Sozinho, ele não identifica o usuário da Internet, mas unido a outros atributos, ele pode revelá-lo.
- xii) Inferência: informação obtida indiretamente a partir da análise de outros atributos. Em outras palavras, é a dedução do valor de um atributo a partir dos valores de um conjunto de atributos.
- xiii) Não-identificador: Atributos que não correspondem a identificadores diretos nem indiretos. Tais atributos não dependem de anonimização.
- xiv) Processo de Anonimização: Planejamento operacional do projeto de anonimização sobre um conjunto de dados, dividido nas etapas pré-definidas expostas no “Metodologia para um Projeto de Anonimização”.

⁵ V. <https://www.lexology.com/library/detail.aspx?g=a882094c-2175-46d0-9c67-07c8ed7deee7>

⁶ Entendemos que o nome é um exemplo de identificador único, desde que não seja tão comum que informações adicionais sejam necessárias para identificar seu titular, como decidido no caso *Efiom Edem v. The Information Commissioner and the Financial Services Authority*, cf. <http://www.bailii.org/ew/cases/EWCA/Civ/2014/92.html>.

- xv) Projeto de Anonimização: Planejamento estratégico com diretrizes sobre a metodologia aplicável a diferentes hipóteses de anonimização.
- xvi) Pseudonimização: técnica que substitui um identificador por um atributo alheio, mas ainda único, singularizando o titular, e.g. substituição do nome por um número único.
- xvii) Registro: grupo de atributos associados a um titular ou um assunto, e.g. a “linha” de planilhas.
- xviii) Reidentificação: Identificação de um titular a partir de um conjunto de dados anonimizado. Reidentificação espontânea se refere à reidentificação não intencional devido a conhecimento especial a respeito de titulares.
- xix) Singularizar: ato de discriminar um titular dentre outros. A singularização é um risco a ser considerado num projeto de anonimização.
- xx) Vinculabilidade: capacidade de vincular, ao menos, dois registros a respeito do mesmo titular ou um grupo de titulares, em um banco de dados ou em diferentes bancos de dados.

3.2. Anonimização

3.2.1. Metodologia para um Projeto de Anonimização

O processo de anonimização é complexo e exige planejamento sob metodologia eficiente. Sua complexidade decorre da proporcionalidade inversa existente entre a probabilidade de reidentificação do conjunto de dados e a sua usabilidade pelo órgão municipal após a anonimização. Quanto mais profundamente se anonimiza um conjunto de dados, menos preciso ele se torna e, conseqüentemente, menos útil para o agente de tratamento.

A seguir, veja-se a metodologia recomendada pelo Guia de Técnicas Básicas de Anonimização do PDPC de Singapura.

3.2.1.1. Determinação do Modelo de Lançamento

Modelo de lançamento é o formato do conjunto de dados após a anonimização. Quanto mais profunda for a anonimização, menos útil será o conjunto de dados. Portanto, o órgão municipal deve identificar, em primeiro lugar:

- i) quem será o público-alvo do conjunto de dados anonimizado; e
- ii) qual será a finalidade para a qual o público-alvo o manuseará.

Com base nessas definições, torna-se possível decidir quais técnicas de anonimização serão necessárias para atingir o grau de usabilidade almejado. Para orientações quanto a escolha das técnicas de anonimização, veja o capítulo 3.2.3. *Crítérios para Definição da Técnica de Anonimização Aplicável*.

3.2.1.2. Determinação do Limite de Risco de Reidentificação Aceitável e de Usabilidade

A partir do art. 12 da LGPD, verifica-se que a anonimização pode ser considerada bem-sucedida, quando a reidentificação é *improvável*. Note-se que não se fala de impossibilidade, mas sim de improbabilidade baseada em certos critérios, quais sejam o grau de esforço e as ferramentas que seriam necessários para reverter a anonimização, ao longo do tempo. Isso significa que pode ser considerado anonimizado – hoje – um conjunto de dados que exigiria esforços e recursos descomunais, não normalmente disponíveis ao público – hoje.

Entretanto, na medida em que a tecnologia avança com o tempo, podem surgir ferramentas aptas a reidentificar conjuntos de dados anonimizados no passado com esforços razoáveis. Ademais, a anonimização é um *processo discreto*, segundo a ANPD.⁷ Em outras palavras, não representa dois estados opostos separados, o não-anonimizado e o anonimizado, mas sim um espectro, que pode variar segundo as finalidades, tempo e recursos disponíveis. Consequentemente, nenhuma anonimização é definitiva nem pode ser entendida como um ato. A anonimização deve ser considerada um processo operacional, a ser controlado perenemente.

Nesse contexto, na segunda fase, o órgão municipal deve fixar os riscos de reidentificação que está disposto a assumir, levando em consideração a

⁷ V. https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/estudo_tecnico_sobre_anonimizacao_de_dados_na_lgpd_uma_visao_de_processo_baseado_em_ri_sco_e_tecnicas_computacionais.pdf

criticidade dos dados pessoais em questão, os controles à disposição e a finalidade do uso que se dará ao conjunto de dados anonimizado. Com base no limite de risco aceitável planejado será possível, após a aplicação das técnicas de anonimização e para o futuro, avaliar se o risco real de reidentificação se encontra dentro do limite planejado.

3.2.1.3. Classificação dos Atributos

Nesta fase, o órgão municipal deve avaliar os atributos do conjunto de dados, para classificá-los entre “identificadores diretos”, “identificadores indiretos” e “não-identificadores”. Com base nesta qualificação, será possível delimitar o escopo da anonimização, dentro da usabilidade e limite de risco definidos.

3.2.1.4. Remoção de Atributos Inúteis

Dentro do conceito de minimização (LGPD, art. 6º, III), o órgão municipal deve eliminar, do conjunto de dados, os atributos desnecessários à(s) finalidade(s) definida(s) previamente.⁸

3.2.1.5. Anonimização de Identificadores Diretos e Indiretos

Após a remoção de atributos inúteis, o órgão municipal deverá decidir quais técnicas de anonimização são necessárias e executá-las.

O órgão municipal pode adotar diversas técnicas de anonimização, inclusive em conjunto, conforme o caso concreto demande. A técnica a ser adotada poderá variar conforme a natureza do conjunto de dados e a finalidade para a qual ele será empregado após a anonimização.

Em algumas situações, pode convir a combinação de mais de uma técnica de anonimização no processo. Mesmo após aplicação de mais de uma técnica de anonimização, alguns registros podem apresentar alta singularidade, i.e., contêm características únicas a permitir a identificação do titular⁹. Neste caso, recomenda-se que o órgão municipal avalie o registro singular, a fim de entender se a anonimização é de fato possível ou se o dado permanecerá pessoal, sob pseudonimização.

⁸ A manutenção de atributos inúteis, além de violar o princípio da necessidade também aumenta o risco de re-identificação, por permitir maior oportunidade de cruzamento com outros conjuntos de dados (maior superfície de ataque).

⁹ Como no exemplo da doença rara de ocorrência única no Município.

Por fim, o órgão municipal também deve levar em consideração que a anonimização deve ser realizada de *maneira tecnologicamente neutra*.¹⁰ Isso significa que o projeto deve ser pensado levando em conta que as tecnologias de reidentificação evoluirão com o tempo, de modo que a anonimização deve ser o mais resiliente possível contra novas tecnologias. Para garantias de manutenção da anonimização, veja o capítulo “*Determinação de Necessidade de Controles*” adiante.

3.2.1.6. *Determinação do Risco Real e Comparação com o Limite de Risco Definido*

O órgão municipal pode determinar o risco real sob diversas metodologias, como *l-diversidade*, *t-proximidade*, *privacidade diferencial* e *k-anonimato*. Neste relatório, adotamos esta última, conforme recomendado pela PDPC de Singapura.

O k-anonimato é um cálculo, cujo resultado forma um índice que revela a probabilidade de reidentificação, especialmente contra riscos de ataques de vinculação (*linking attacks*)¹¹. O k-anonimato pode ser utilizando antes do processo de anonimização, para planejá-lo com base no risco, e após, para verificar sua incidência. Recomenda-se que o órgão municipal siga instruções técnicas para o k-anonimato¹².

Como visto acima, a comparação do risco real com o limite de risco aceitável planejado no início do processo de anonimização é um ciclo contínuo. A ANPD bem definiu o processo de aferição de riscos no fluxograma abaixo¹³:

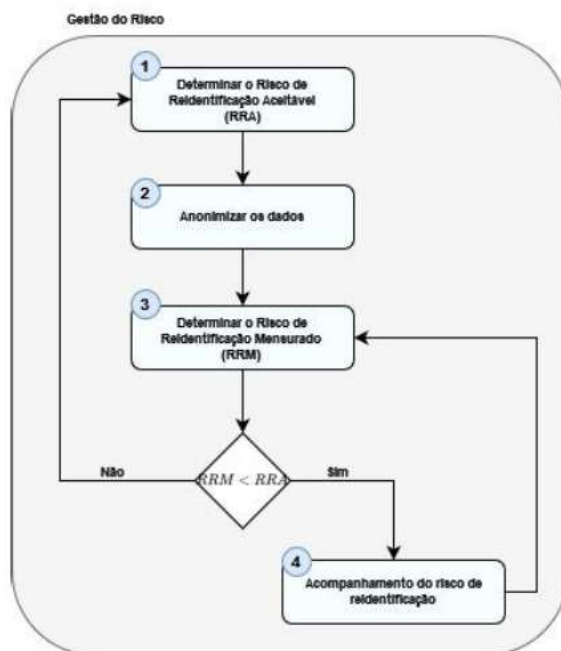
¹⁰ WP92, op. cit., p. 9.

¹¹ Trata-se de ataque onde um terceiro adversário tenta vincular um conjunto de dados anonimizado a outro, com o objetivo de viabilizar a re-identificação dos registros, cf. <https://www.sciencedirect.com/topics/computer-science/linking-attack>

¹² V. <https://programming-dp.com/chapter2.html>

¹³ V. https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/estudo_tecnico_sobre_anonimizacao_de_dados_na_lgpd_uma_visao_de_processo_baseado_em_ri_sco_e_tecnicas_computacionais.pdf

Figura 2: Etapas essenciais para a gestão do risco de reidentificação.



O WP29 acrescenta três riscos a serem considerados.¹⁴ Em primeiro lugar, confundir anonimização com pseudonimização pode ocasionar tratamentos ilícitos. Dados pseudonimizados continuam sendo pessoais e sujeitos às normas legais, de modo que não podem ser manuseados como os anonimizados. Em segundo lugar, não avaliar se existem outras normas aplicáveis sobre o conjunto de dados anonimizados pode ocasionar infrações. Ainda que os dados estejam anonimizados, é possível que alguma outra norma recaia sobre eles, como por exemplo, os registros obrigatórios do Marco Civil da Internet, ou normas de confidencialidade da Administração Pública. Em terceiro lugar, um mal uso do conjunto de dados anonimizado ainda pode causar dano a pessoas¹⁵.

O WP29 também recomenda a adoção de um teste de avaliação da anonimização em três etapas¹⁶. Segundo o WP29, o órgão municipal poderia avaliar a eficácia da anonimização, se puder responder negativamente às seguintes perguntas:

- i) **Singularização:** Ainda é possível singularizar uma pessoa a partir do conjunto de dados?

¹⁴ WP29, op. cit., p. 10/11.

¹⁵ No caso *Havasupai Tribe v. Arizona State University Board of Regents*, dados pessoais genéticos coletados sob consentimento para pesquisa sobre diabetes foram posteriormente, sem aviso, usados para pesquisar incidência de esquizofrenia e incesto na tribo, com forte potencial de consequências discriminatória contra seus membros. Disto se conclui que, mesmo que o resultado do processamento seja um conjunto de dados anonimizados, ele pode gerar um impacto discriminatório negativo. Cf. <https://law.stanford.edu/2010/04/23/the-havasupai-case-and-how-to-make-consent-forms-better/>.

¹⁶ WP29, op. cit., p. 3.

- ii) *Vinculabilidade*: Ainda é possível cruzar os dados do conjunto de dados anonimizado com outro conjunto de dados?
- iii) *Inferência*: É possível inferir informações sobre um titular identificável a partir dos dados?

3.2.1.7. Combinação de Outra Técnica de Anonimização

Se o risco aferido for superior ao limite planejado, o órgão municipal deverá reforçar a anonimização com outra técnica até que o risco seja reduzido. Em seguida, o órgão municipal deverá avaliar novamente o risco, repetindo os passos anteriores até que o risco esteja abaixo do limite aceitável.

3.2.1.8. Avaliação do Resultado

Finda a execução das técnicas de anonimização, dentro do risco previsto, o órgão municipal deverá avaliar se o conjunto de dados anonimizado atende à finalidade planejada. Caso a finalidade não possa ser alcançada, recomenda-se replanejar a anonimização ou cancelá-la e tratar o dado como pessoal, sob as regras da LGPD.

3.2.1.9. Determinação de Necessidade de Controles

Viu-se no capítulo *Determinação do Limite de Risco de Reidentificação Aceitável e de Usabilidade* que nenhuma anonimização é definitiva. Diante disso, o órgão municipal deve estabelecer controles técnicos e não-técnicos para garantir a manutenção da anonimização.

A depender do processo de anonimização realizado e do risco planejado, o órgão municipal pode aplicar diferentes controles. A seguir, indicam-se alternativas de controles recomendados pelo PDPC¹⁷:

- i) *Acesso Revogável*: O acesso ao conjunto de dados anonimizado deve ser limitado aos agentes públicos conforme a necessidade do manuseio (*need to know basis*). Uma vez que a necessidade de manuseio pelo agente público termine, a autorização de acesso respectiva deve ser revogada, de modo que o funcionário que não possa mais acessá-lo.

¹⁷ PDPC, op. cit., p. 33.

- ii) **Apenas Consultas (*query only*):** trata-se de controle mais eficaz que o anterior, onde o agente público não terá acesso à íntegra do conjunto de dados, mas pode obter as informações de que precisar por meio de *consulta* a um sistema. Dessa forma, impede-se que o funcionário tenha contato com mais informações do que as necessárias. Numa versão mais rígida, o órgão municipal pode nomear um supervisor que autorize ou não cada consulta, avaliando a real necessidade de o agente público ter o acesso almejado.
- iii) **Limitação de Recipientes (*limit recipients*):** este controle prevê autenticação prévia da identidade do agente público que deverá manusear o conjunto de dados anonimizado. Um exemplo de autenticação é o uso de credenciais, como senhas.
- iv) **Controle de gestão de direitos digitais (digital rights management – DRM):** trata-se de medida voltada para tratamentos digitais, na qual impede-se o agente público de salvar ou imprimir o conjunto de dados anonimizado.
- v) **Acesso *in loco*:** exigir que o agente público se apresente pessoalmente para manusear o conjunto de dados fisicamente é um método tradicional de controle. Conforme inovações tecnológicas surjam, convém acrescentar novos controles, como proibição do uso de celulares, de fotos, área sem sinal de internet ou telefone e vigilância por circuito fechado de televisão (CFTV).
- vi) **Acesso a subconjunto:** este controle visa evitar o acesso integral ao conjunto de dados, liberando o acesso a apenas parte dele.
- vii) **Medidas físicas:** em complementação às medidas digitais vistas acima, medidas físicas podem ser utilizadas de diferentes maneiras para evitar abusos de acesso. Como exemplo, podemos citar a instalação de tokens ou dispositivos físicos, presos a uma determinada sala segura, portas com senha, salas-cofre com segurança etc.

Ainda segundo o PDPC, a manutenção da anonimização também exige adoção de regras de governança. Após a anonimização do conjunto de dados, seu manuseio continua exigindo cuidados, donde se recomenda:

- i) **Manter trilha de auditoria sobre o manuseio dos conjuntos de dados,** inclusive sobre os agentes públicos que o manusearam e a forma de acesso, com carimbo de data e hora (*time stamp*). Isso se faz necessário, pois o risco de reidentificação pode aumentar quando um mesmo funcionário tem acesso a dois conjuntos de dados

anonimizados, mas complementares. Daí, a necessidade de registrar quem acessou o que e quando.

- ii) Gerenciar bem as chaves de acesso e de encriptação. Em regra, conjuntos de dados digitais são protegidos por encriptação e senhas. Contudo, sabe-se que, se as senhas não forem fortes nem atualizadas periodicamente, podem ser descobertas por adversários. Portanto, convém ao Município elaborar um protocolo de gestão de chaves de acesso para execução pelos órgãos municipais, para evitar vulnerabilidades.
- iii) Reavaliar o risco de reidentificação periodicamente. O Município deve prever, em seu Projeto Municipal de Anonimização, a revisão periódica do grau de risco aceitável e real, para cumprimento pelos órgãos municipais.
- iv) Realizar auditorias sobre os recipientes dos dados. De pouco adianta elaborar normas; deve-se ter a certeza de que todos as estão cumprindo. Nesse sentido, convém planejar auditorias periódicas para coletar evidências de que os recipientes não estão burlando regras contra re-identificação.

3.2.1.10. Documentação do Processo de Anonimização

Com base no princípio da prestação de contas (LGPD, art. 6º, X), toda atividade envolvendo a proteção de dados pessoais deve ser documentada para produzir evidências. Nesse sentido, o órgão municipal deve documentar o processo de anonimização e a execução de suas fases, com detalhes sobre análises feitas e as decisões tomadas, inclusive para consulta futura no desiderato de preservação da memória das atividades dos órgãos municipais em contraposição a rotatividade de seus servidores.

Ademais, a documentação viabiliza a revisão, a manutenção e a auditoria sobre os processos de anonimização.

Portanto, recomenda-se a documentação do projeto e a guarda segura respectiva, tendo em vista que a revelação das decisões tomadas pode facilitar a engenharia reversa por um adversário.

3.2.2. Técnicas de Anonimização

Viu-se que um projeto de anonimização deve levar em consideração a diversidade de casos concretos possíveis. Cada técnica terá vantagens e desvantagens quanto ao risco de reidentificação e a usabilidade desejada, a recomendar planejamento criterioso. A seguir, vejam-se as técnicas recomendadas em diferentes guias internacionais e cogitadas por esta relatoria, bem como, em casos relevantes, as garantias de proteção e os erros comuns associados.

3.2.2.1. Supressão de Atributo

Esta técnica consiste na deleção pura e simples de atributos (colunas) de um conjunto de dados. Feita corretamente, ela é a forma mais forte de anonimização, pois efetivamente destrói o atributo pessoal, mas onde a supressão prejudicar a usabilidade de um atributo, ela deve ser substituída por outra técnica.

A supressão de atributo pode ser usada quando algum atributo não é necessário ou não pode ser anonimizado. Ela deve ser a primeira técnica cogitada sobre o conjunto de dados, pois corresponde à execução do princípio da necessidade (LGPD, art. 6, III).

Para garantir a supressão segura do atributo, faz-se necessário eliminá-lo definitivamente do conjunto de dados, com *sobrescrita*. Não basta apenas ocultá-lo ou excluí-lo, pois nesses casos, um adversário pode recuperar a informação.

- Exemplo:

A Secretaria de Educação precisa conhecer as notas dos alunos sob diferentes professores, mas o nome do aluno é desnecessário.

Antes da supressão:

Aluno	Professor	Nota
João	Jairo	8,0
Maria	Regina	9,0

Após a supressão do atributo desnecessário:

Professor	Nota
Jairo	8,0
Regina	9,0

3.2.2.2. Supressão de Registro

Ao contrário da supressão de atributo, que elimina uma coluna, a supressão de registro elimina uma linha, suprimindo múltiplos atributos de uma vez.

Essa técnica pode ser usada antes ou após outras técnicas. E.g., ela pode ser usada para eliminar registros muito singulares ou que não cumprem certos requisitos (como o k-anonimato, que se verá adiante). Registros excessivamente únicos admitem re-identificação, em tese, como o único voto de um vereador num bairro.

Assim como a supressão do atributo, a supressão de registro é operada ao eliminar-se definitivamente o registro, com sobrescrita.

Atente-se para o fato que a supressão de um registro pode impactar na utilidade do conjunto de dados. Por exemplo, estatísticas podem não ser verdadeiras por não considerarem os registros eliminados.

- Exemplo;

A Secretaria de Saúde está estudando a incidência de doenças nos hospitais, mas em um deles, surgiu caso de doença rara, que permitiria a reidentificação do paciente ao cruzarem-se outros dados.

Hospital	Doença	Casos
Pedro II	Influenza	500
Paulino Werneck	Covid	100
Miguel Couto	Hanseníase	1

Após a supressão do registro singular

Hospital	Doença	Casos
Pedro II	Influenza	500
Paulino Werneck	Covid	100

3.2.2.3. Mascaramento de Caracteres

Trata-se da substituição dos caracteres de um atributo por outro símbolo constante (e.g. “XXXXXX” ou “*****”). Normalmente, o mascaramento é feito apenas em parte dos caracteres do atributo, que seja suficiente à anonimização, e.g. parte de um telefone (ex.: 9****-6114).

O mascaramento de caracteres pode ser usado quando eliminar apenas parte do conteúdo do atributo é suficiente para impedir a reidentificação.

O comprimento do conteúdo pode guardar informação sobre o dado original para um adversário.¹⁸ Caso o comprimento do conteúdo não seja relevante para o uso do conjunto de dados, o mascaramento pode ser feito com uma quantidade diferentes de caracteres.

Alguns bancos de dados podem preservar uma *soma de verificação* (*checksum*) que considera o conteúdo original. Nesse caso, um adversário poderia reverter o processo de anonimização a partir da soma de verificação. Para garantir a anonimização, a reversão pela soma de verificação não deve ser possível.

Em alguns tratamentos, pode ser importante que apenas o titular reconheça sua identidade, como nos exemplos do telefone ou cartão de crédito acima. Embora tais situações não sejam tecnicamente anonimização, o mascaramento garante que o público em geral seja incapaz de re-identificar a informação.

- Exemplo:

O Município pretende publicar os aprovados em concurso à sua Procuradoria-eral, sem expor a identidade dos candidatos reprovados.

CPF	Nota Final	Situação
123.456.789-10	10	Aprovado
098.765.432-01	0	Reprovado

Após o mascaramento de caracteres:

CPF	Nota Final	Situação
..789-10	10	Aprovado
..432-01	0	Reprovado

¹⁸ Como se verá adiante, deve-se incluir no processo de anonimização alguém com expertise na atividade que aproveitará o conjunto de dados anonimizados, pois especialistas tem maior probabilidade de observar inferências que poderiam ser aproveitadas por adversários.

3.2.2.4. Generalização

A generalização é a redução voluntária da precisão do atributo, como a conversão da idade de uma pessoa a um a faixa (e.g. no lugar do atributo “30 anos” o atributo “20-50 anos”). Ela pode ser usada em tratamentos que não dependem de alta precisão.

Para tanto, faz-se necessário avaliar previamente qual o grau de precisão que a finalidade exige e quais serão os atributos e as faixas respectivas a serem submetidos à generalização.

Recomenda-se que as faixas não sejam estreitas demais, sob risco de permitir reidentificação. Considere ainda que, caso a técnica de k-anonimato seja utilizada conjuntamente, o valor k também afetará as faixas. Ademais, as faixas colocadas nos extremos, i.e. as mais baixas e mais altas, costumam conter menos indivíduos. Se esse for o caso, elas podem ser aplicadas para comportar mais pessoas, dificultando a reidentificação.

Outro exemplo é a supressão de parte do atributo que causa singularidade. Na geolocalização, por exemplo, dados muito minuciosos como “rua”, podem ser substituídos por outros mais abrangentes como “bairro”.

- Exemplo:

CPF	Data de Nascimento	Localização
123.456.789-10	01.01.1980	Av. Rio Branco, 110, 1110.
098.765.432-10	02.02.1990	Av. Pres. Vargas, 1, 500.

Após a Generalização:

CPF	Ano de Nascimento	Localização
*****	<1985	Centro
*****	1985-1995	Centro

3.2.2.5. Embaralhamento, Troca ou Permuta

Esta técnica rearranja atributos entre diferentes registros, embaralhando um conjunto. Assim, atributos verdadeiros continuam representados estatisticamente, porém deixam de corresponder à pessoa original.

O embaralhamento é conveniente quando o conjunto de dados será sujeito a análise como um todo, mas a pormenorização dos registros é irrelevante. Com efeito, a quantidade de ocorrências dos valores dos atributos permanecerá a mesma após o embaralhamento.

Para realizar o embaralhamento, é necessário identificar o atributo que será embaralhado entre os registros e rearranjá-los de maneira aleatória. Note que o método de embaralhamento não deve ser conhecido, sob risco de viabilizar reidentificação.

- Exemplo

Num estudo de incidência de pessoas com necessidades especiais, a Secretaria de Saúde pretende anonimizar uma lista de pessoas:

Pessoa	Possui Necessidades Especiais?
A	Sim
B	Não

Após o embaralhamento:

Pessoa	Possui Necessidades Especiais?
A	Não
B	Sim

- Garantias:

- Singularização: em se tratando de espécie de randomização, ainda é possível singularizar registros, mas sem precisão.
- Vinculabilidade: se o embaralhamento ocorrer sobre quase-identificadores, pode evitar uma vinculação correta entre atributos e registros.
- Inferência: inferências ainda podem ser extraídas do conjunto de dados, especialmente se os atributos forem correlacionados e tiverem uma forte relação lógica entre si. Contudo, ao desconhecer o modo do embaralhamento, o adversário só conseguiria extrair inferências probabilísticas.

- Erros Comuns:

- Escolher o atributo errado: embaralhar atributos irrelevantes para a identificação não resultará em aumento da eficácia da anonimização.

- ii) Embaralhar atributos aleatoriamente: Se dois atributos são fortemente correlacionados, embaralhá-los aleatoriamente pode não ser eficaz. Um exemplo de forte correlação se encontra entre o salário de um titular, seu cargo e sua idade. Um titular mais idoso, com cargo de liderança, provavelmente terá um salário maior. Se seus atributos forem embaralhados com os de um titular com atributos diametralmente opostos (jovem, em início de carreira e baixo salário), reidentificar o conjunto pode ser possível.
- iii) Presumir que o embaralhamento é suficiente: como espécie de randomização, a permuta não causa a anonimização por si e deve ser combinada com outras técnicas, como as de supressão.

3.2.2.6. Perturbação de Dados

Onde a exatidão dos atributos não é essencial, é possível efetuar pequenas alterações em seus valores para dificultar a reidentificação. A perturbação é conveniente para quase-identificadores, ou seja, números ou datas que potencialmente identifiquem o titular quando combinados com outros dados.

A forma de aplicação pode variar segundo a categoria de dado a ser anonimizada. São exemplos de perturbação o arredondamento de valores¹⁹ e a soma/subtração de valores de um atributo.

Recomenda-se que o grau de perturbação seja proporcional ao valor do atributo e à abrangência de valores no conjunto de dados. Quanto menor o valor, menor deve ser a perturbação. Paralelamente, se a quantidade de registros é pequena, é recomendável uma perturbação maior para dificultar a distinção entre eles.

Em casos de alta singularidade de um registro, recomenda-se a adoção da perturbação, aqui também chamada de *barnardização*, onde o agente de tratamento pode somar um ou subtrair um, aleatoriamente, impedindo a reidentificação do registro único.

- Exemplo

A Secretaria de Saúde quer planejar programas de prevenção a doenças a partir de características corporais de pessoas, mas sua identidade é irrelevante para a alocação dos recursos. Considerando que altura e peso

¹⁹ O National Institute of Standards and Technology dos Estados Unidos da América considera o “arredondamento” (*averaging*) como uma técnica autônoma. Contudo, entendemos que ela é uma espécie de perturbação, que segue a regra de levar cada atributo a um valor mediano. Cf. <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf>, p. 29.

podem ser cruzados com outras bases e viabilizar a identificação da pessoa, deve ser aplicada a perturbação.

Pessoa	Altura (m)	Peso (kg)	Fumante
A	1,62	51	Sim
B	1,84	89	Não
C	1,73	81	Sim
D	2,43	150	Não

Após a perturbação:

Pessoa	Altura (m)	Peso (kg)	Fumante
A	1,60	50	Sim
B	1,85	90	Não
C	1,70	80	Sim
D	2	120	Não

- Garantias

i) Singularização: Ainda é possível singularizar um titular, embora de maneira não identificável.

ii) Vinculabilidade: Ainda é possível vincular registros sobre um mesmo indivíduo, mas eles são menos precisos.

iii) Inferência: Ataques de inferência podem ser possíveis, mas a probabilidade de sucesso será menor com falsos positivos e negativos.

- Erros Comuns

i) Acrescentar ruído inconsistente: Se o ruído não for semanticamente viável, ou seja, se o novo valor não for harmônico com o valor real ou a natureza do atributo, um adversário poderia filtrar o ruído e identificar os valores reais.

ii) Presumir que a perturbação é suficiente: a perturbação deve ser sempre uma medida complementar.

3.2.2.7. Agregação de Dados

A agregação resume o conjunto de dados em valores condensados. Em outras palavras, ela sintetiza atributos de diferentes registros, sob um atributo

único, resumido, que compreende os originais dentro de uma faixa ou espectro. Para realizá-la o órgão deve converter os registros (linhas) por estatísticas.

Esta técnica pode ser usada quando o órgão municipal necessitar da análise condensada dos dados, mas não dos registros pessoais individualizados.

Em alguns casos, é possível substituir a supressão por uma técnica de agregação. E.g. em registros de uso de transporte público pode não ser interessante manter o horário de entrada ou saída do usuário, de modo que esses atributos deveriam ser eliminados. Por outro lado, para evitar a perda completa da inteligência, seria possível substituí-los por uma informação derivada, como o tempo de uso do transporte (que consideraria a entrada e saída, embora ocultando-os e preservando a privacidade do usuário).

- Exemplo

Usuário	Embarque - ônibus	Desembarque - ônibus
A	05:00	07:00
B	07:00	07:30
C	08:00	10:00

Após a agregação:

Qtde. de usuários	Tempo de uso - ônibus
2	02:00
1	00:30

- Garantias

- I) Singularização: Visto que um atributo tem por valor uma faixa genérica de valor, e que dentro deste atributo, um número “K” de registros maior que 1 é agregado, a singularização não é possível.
- II) Vinculabilidade: é limitada, mas ainda é possível vincular registros por grupos de K titulares e, dentro dos grupos, a probabilidade de dois registros corresponderem aos mesmos identificadores pode ser maior que a esperada.
- III) Inferência: não impede ataques de inferência. Se todos os K titulares estiverem num mesmo grupo, então conhecer o grupo ao qual pertence um titular permite conhecer alguma informação sobre ele, como no exemplo da Tribo Havasupai acima.

- Erros Comuns

- I) Desprezar quase-identificadores: Quanto maior o número K de titulares sob um mesmo registro, mais forte é a anonimização. Mas aumentar o valor de K artificialmente pela redução do conjunto de quase-identificadores torna possível construir “blocos” de K-titulares, mediante a força da identificação de outros atributos. Exemplo disso, são os atributos com alta singularidade. Portanto, não considerar todos os quase-identificadores pode viabilizar a re-identificação.
- II) Valor de K pequeno: Se K foi muito reduzido, a relevância de cada titular dentro do grupo será grande, de modo a aumentar a probabilidade de sucesso de ataques de inferência.
- III) Não agrupar titulares de mesmo peso: agrupar titulares sob uma distribuição desequilibrada pode ser prejudicial à anonimização. Registros com atributos muito singulares aumentam a probabilidade de re-identificação. Portanto, é importante que o número K de titulares seja grande o suficiente para anular a singularização desses titulares.

3.2.2.8. Técnicas para Conjuntos de Dados Físicos

Parte relevante do acervo informacional de órgãos públicos é físico. Por sua natureza, a anonimização do registro em papel ou outro suporte físico é diferente da de conjuntos de dados digitais.

Quanto a isso, a solução mais recomendável seria a digitalização. Por meio dela, simplifica-se o manuseio da informação e se reduzem os gastos de tempo, espaço e recursos em geral. Atualmente, é possível encontrar diversas soluções no mercado para digitalizar documentos e fornecer *softwares* para Gestão Eletrônica de Documentos – GED. Uma boa ferramenta deste tipo poderia auxiliar o Município em diversas questões, entre as quais se destacam o controle de acessos, trilhas de auditoria, anonimizações e gestão de prazos de retenção e eliminação.

Não obstante, a anonimização de documentos físicos também é possível. Neste caso, o processo envolve técnicas mecânicas ou manuais, para tornar o dado pessoal ilegível. Como exemplo, podemos citar o uso de tarjas pretas indelévels sobre dados pessoais irrelevantes ou o uso de carimbos de ruído (*identity protection roller stamps*), que imprimem formas e caracteres sobre o dado no papel e impedem o leitor de distinguir os caracteres originais.

3.2.2.9. Técnicas para Imagens e Áudio

Outra situação excepcional que exige técnicas diferentes de anonimização é a dos registros de imagens ou de áudio, sejam fotos, áudios isolados ou filmagens. Com efeito, o Município possui em seus arquivos múltiplos formatos de áudio e vídeo como fotos de crachás e gravações de CFTV ou telefonemas para centrais de atendimento. Cada um deles representa um conjunto complexo de dados e metadados cuja vinculação a uma pessoa física pode não ter finalidade clara, a exigir eliminação ou anonimização.

Diante disso, recomenda-se a adoção de ferramentas digitais capazes de alterar vozes para frequências que tornem impossível a identificação do titular; bem como de mascarar definitivamente o rosto ou até a silhueta de titulares irrelevantes para filmagens.

3.2.3. Critérios para Definição da Técnica de Anonimização Aplicável

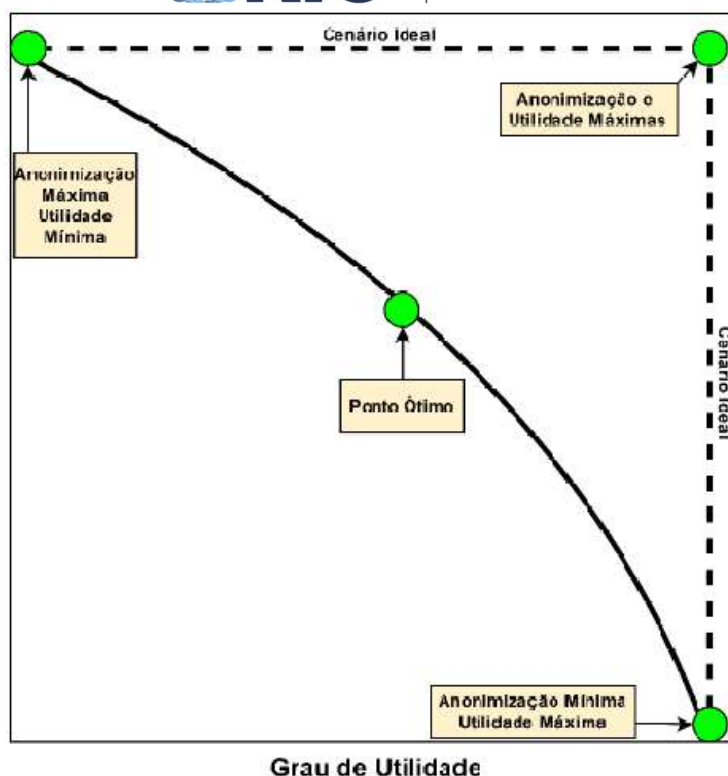
Como os conjuntos de dados e os usos aos quais são empregados podem variar, também variarão as técnicas de anonimização. Com efeito, o resultado de uma técnica pode não ser útil ou até prejudicar o uso para o qual o órgão municipal pretende tratar o conjunto de dados. Diante disso, na escolha da técnica apropriada, o órgão municipal deve levar em conta:

- i) A natureza e tipo de dado a ser tratado;
- ii) A gestão de risco no órgão municipal, como forma de complementar a anonimização; e
- iii) A utilidade pretendida do conjunto de dados anonimizado.

Com base nos critérios acima, o órgão municipal deve avaliar os aspectos abaixo.

- i) *Finalidade da anonimização e utilidade*: a finalidade da anonimização deve ser bem definida, pois a anonimização reduz a quantidade de informação no conjunto de dados original. Como orientado pela ANPD²⁰, deve-se atingir um ponto ótimo entre os dois vetores:

²⁰ Estudo Técnico Sobre Anonimização de Dados na LGPD: Uma Visão de Processo Baseado em Risco e Técnicas Computacionais, cf. https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/estudo_tecnico_sobre_anonimizacao_de_dados_na_lgpd_uma_visao_de_processo_baseado_em_risco_e_tecnicas_computacionais.pdf



Por isso, o órgão municipal deve ter em mente a utilidade que pretende dar ao conjunto de dados anonimizado em proporção ao grau de de-identificação que deve ser aplicado, sob risco de anonimizações profundas o tornarem inútil ao fim que se destina.

- ii) *Características da técnica de anonimização*: certas técnicas são mais adequadas a certos usos do que outras. Por exemplo, o mascaramento de caracteres normalmente é aplicado sobre identificadores diretos, enquanto outras como agregação servem para identificadores indiretos.
- iii) *Inferências*: Em alguns casos, é possível inferir informações a partir de dados anonimizados. Por exemplo, se apenas uma única pessoa votou em um vereador num bairro, pode-se inferir quem ela é.
- iv) *Expertise sobre o tratamento de dados*: Especialistas na atividade onde ocorre o tratamento podem extrair inferências. Por isso, faz-se necessário que especialistas participem do planejamento da anonimização. E.g. a anonimização de um conjunto de dados sobre questões de educação, deve ser acompanhada por um profissional da educação, pois apenas sua expertise seria capaz de enxergar inferências e potenciais re-identificações.

- v) *Competência em processos e técnicas de anonimização*: as técnicas de anonimização e sua execução exigem conhecimentos não comuns a qualquer profissão. Nesse sentido, o órgão municipal deve empregar especialistas em anonimização onde pretender adotar esta medida.
- vi) *O recipiente*: Assim como o órgão municipal deve envolver alguém com expertise na atividade subjacente, ele também deve considerar a expertise do público que terá acesso ao conjunto de dados anonimizado. Este também pode ter conhecimentos prévios que viabilizem inferências.
- vii) *Ferramentas*: Atualmente, há softwares capazes de aplicar técnicas de anonimização automaticamente em conjuntos de dados originais. Não obstante, recomenda-se o envolvimento de equipe apta a operá-los.

3.3. Pseudonimização

A pseudonimização não é uma técnica de anonimização e não afasta a aplicação da LGPD. Apesar de serem comumente confundidas, a pseudonimização é meramente uma medida técnica de segurança sobre os dados pessoais, que dificulta – mas não impossibilita – a reidentificação. Ela pode ser usada onde a ciência da identidade do titular não é necessária, mas a distinção entre cada titular continua sendo-o.

A pseudonimização se caracteriza pela substituição de identificadores por outros valores, de modo a preservar a singularização do registro dentre os demais. Trata-se, pois, de espécie de encriptação, onde, com base em uma chave particular, o valor de um atributo num conjunto de dados é substituído por outro, mas pode ser revertido por quem detém a chave.

Onde houver mais de um conjunto de dados relevante, o mesmo pseudônimo pode ser aplicado para registros referentes ao mesmo titular (*pseudônimo persistente*), ou não, a depender da finalidade do tratamento dos conjuntos de dados. *Pseudônimos persistentes* permitem construir coerência entre diferentes conjuntos de dados, conforme isto seja necessário.

Para realizá-la, é possível utilizar ferramentas para gerar valores aleatórios em substituição aos verdadeiros no conjunto de dados. Por óbvio, cada registro deve ter seu próprio pseudônimo, sob pena de perda de precisão. Ademais, o pseudônimo não pode ter forte relação com o valor original, sob risco de aumentar a probabilidade de reversão.

Em caso de pseudônimos reversíveis, os destinatários do conjunto de dados pseudonimizados não devem ter acesso ao conjunto de dados original nem à chave, sob risco de reidentificação. O conjunto de dados original e a chave de encriptação devem ser mantidos em local seguro e só devem ser utilizados em casos excepcionais. Um vazamento desses elementos tornaria todos os registros pseudonimizados reidentificáveis.

Chaves de encriptação se tornam obsoletas com o tempo. Na medida em que o poder de processamento dos computadores aumenta, a capacidade de um adversário testar todas as combinações de códigos possíveis para descriptar um conjunto de dados se torna mais provável. Diante disso, se a pseudonimização for feita por chave, é recomendável atualizar a chave periodicamente (*rotação periódica de pseudônimos*).

Certos sistemas de processamentos de dados só funcionam se os atributos possuírem uma quantidade fixa de caracteres. Por exemplo, um sistema que lê CPFs, normalmente, processa 11 caracteres e não funcionaria com uma quantidade diferente, prejudicando o processamento do conjunto de dados. Nessas situações, o pseudônimo deve possuir o mesmo número de caracteres que o atributo real, o que pode ser realizado por meio de geradores de pseudônimos com encriptação preservadora de formato (*format preserving encryption*).

- Exemplo:

A Guarda Municipal pretende contratar um fornecedor para processar as multas impostas e descobrir se há infratores contumazes, com o fim de realizar fiscalização mais eficiente. Entretanto, só a Guarda Municipal precisa conhecer o cidadão, de modo que o fornecedor não precisa saber a identidade da pessoa. Nesse caso, a pseudonimização pode ser usada para que fornecedor processe o conjunto de dados pseudonimizado e encontre o resultado esperado pela Guarda Municipal, sem conhecer a identidade do motorista. Uma vez devolvido resultado do tratamento, a Guarda Municipal pode reverter os pseudônimos para identificar os titulares de interesse.

Nome	Placa
João José	ABC1234
José João	DEF1234
Maria do Carmo	GHI1234
Maria das Dores	JKL1234

Após a pseudonimização:

Código	Multas
9876	1
8465	3

4983	2
7653	20

Onde a chave na posse da Guarda Municipal é:

Nome	Placa	Código
João José	ABC1234	9876
José João	DEF1234	8465
Maria do Carmo	GHI1234	4983
Maria das Dores	JKL1234	7653

- Garantias:
 - i) Singularização: É possível singularizar titulares, pois continuam identificáveis, por meio do código que reverte o atributo pseudonimizado.
 - ii) Vinculabilidade: É possível vincular diferentes conjuntos de dados onde houver um pseudônimo persistente ou onde as chaves para a pseudonimização de cada conjunto forem conhecida.
 - iii) Inferência: A depender da força da pseudonimização (e.g. grau de encriptação e sua extensão entre diferentes atributos), há probabilidade de inferência da real identidade do pseudônimo.
- Erros Comuns:
 - i) Crer que o dado pseudonimizado é anonimizado: É comum que agentes de tratamento confundam as duas práticas e superestimem a segurança decorrente da pseudonimização. Pseudonimizar não afasta a aplicação de normas de proteção de dados pessoais, pois a reidentificação é possível mediante cruzamento com outros conjuntos de dados ou a partir da descoberta da chave, inclusive por inferência.
 - ii) Manter a chave da pseudonimização junto do conjunto de dados pseudonimizado

4. Recomendações

Adotar terminologia uniforme (Item 3.1.).

Elaborar de Projeto Municipal de Anonimização, com adoção da metodologia proposta (item 3.2.1.), formalização das técnicas disponíveis (item 3.2.2.) e formalização dos critérios aplicáveis para decisão quanto às técnicas a aplicar (3.2.3.), para observância obrigatória no planejamento de processos de anonimização por órgãos municipais.

Prever no Projeto Municipal de Anonimização, que todo processo de anonimização deva explicitar a(s) finalidade(s) para a(s) qual(is) o conjunto de dados anonimizado será utilizado, a natureza dos dados, o tamanho do conjunto, atratividade a adversários e a potencial existência de outros conjuntos de dados vinculados.

Prever no Projeto Municipal de Anonimização que os RoPAs dos órgãos municipais contenham informação acerca de eventual anonimização/pseudonimização realizada.

Elaborar programa de treinamento periódico dos gestores municipais a respeito do Projeto Municipal de Anonimização.

Elaborar cronograma de conscientização dos gestores municipais a respeito do Projeto Municipal de Anonimização, inclusive reforçando o papel da Encarregada na orientação geral em caso de dúvidas.

Estabelecer a obrigatoriedade de o Encarregado do órgão municipal ser envolvido no projeto de anonimização respectivo.

Estabelecer a obrigatoriedade de um especialista na atividade onde o tratamento ocorre ser envolvido no projeto de anonimização respectivo (item 3.2.3.).

Elaborar cartilha com exemplos de aplicação de cada técnica de anonimização.

Aprofundar do estudo sobre de-identificação para definição de riscos toleráveis e intoleráveis, no Projeto Municipal de Anonimização, e gestão contínua respectiva pelos órgãos competentes em seus processos de anonimização (itens 3.2.1.2. e 3.2.1.6.).

Auditar e inventariar os controles contra reidentificação atualmente existentes nas atividades dos órgãos municipais e reavaliar, em conjunto com os respectivos gestores, a necessidade de adoção de novos controles, proporcional ao limite de risco adotado (item 3.2.1.9.).

Aprofundar estudo sobre categorias de ferramentas de anonimização disponíveis no mercado, especialmente sistema de Inteligência Artificial *open source*, para viabilizar licitações respectivas.

Elaborar regulamento de governança de processos de anonimização, contendo obrigações voltadas à manutenção da anonimização, em especial, o estabelecimento

de trilha de auditoria, protocolo de gestão de chaves de acesso e o órgão responsável pela fiscalização do cumprimento respectivo (item 3.2.1.9.).

Estabelecer método para guarda segura da documentação dos processos de anonimização dos órgãos municipais (3.2.1.10).

Digitalização do acervo documental, onde possível, e contratação de solução de Gestão Eletrônica de Documentos.

Harmonização, com base no Projeto Municipal de Anonimização, de normas administrativas municipais associadas à proteção de dados pessoais, em especial aquelas conteúdo de política de privacidade, de descarte e de retenção, bem como de normas associadas à Lei de Acesso à Informação e Open Data.

Adotar a pseudonimização como medida de segurança, sempre que a anonimização não for possível e a identidade do titular não for necessária (Item 3.3.).

Divulgar práticas de anonimização à população, para aumento de confiança nos serviços municipais.

Conclusão

Este trabalho buscou orientações nacionais e internacionais acerca de práticas de de-identificação, notadamente, anonimização e pseudonimização.

Na apresentação do problema, conclui-se que a grande quantidade de dados pessoais e da complexidade de categorias de suportes documentais, tanto físicos quanto digitais, torna o cumprimento da Lei Federal n. 13.709/18 uma tarefa de alta dificuldade, em se considerando os recursos disponíveis aos órgãos públicos municipais. Nesse contexto, a anonimização e a pseudonimização, tornam-se medidas indispensáveis para uma gestão segura e legal dos conjuntos de dados pessoais entre diversos órgãos municipais e sua interação com entidades privadas, inclusive para fins de monetização de bases de dados. Diante desse desafio, é indispensável que se pesquisem boas práticas nacionais e internacionais de anonimização e pseudonimização, com orientações estratégicas e operacionais para cumprimento pelos diferentes níveis do Poder Público municipal.

O terceiro capítulo apresentou práticas aconselhadas por entidades como a autoridade de proteção de dados pessoais de Singapura (PDPC), o National Institute of Standards and Technology dos Estados Unidos da América, o Comitê Europeu para Proteção de Dados e a Agência Nacional de Proteção de Dados (ANPD). Sintetizando aquelas adequadas à LGPD, o relatório

apresentou metodologia a ser formalizada em um Projeto Municipal de Anonimização e aplicada pelos órgãos municipais em processos de anonimização. Em especial, a metodologia se baseia na necessidade de um planejamento formal da anonimização, com definição de finalidade, de público-alvo de limite de riscos aceitável, bem como a necessidade de documentação formal dos projetos e adoção de medidas de controle. Por fim, apresentou-se também técnicas de anonimização e de pseudonimização recomendadas por essas entidades e os critérios de escolhas entre elas.

Por fim, o quarto capítulo apresentou lista de recomendações estratégicas e operacionais, a serem adotadas por órgãos municipais de diferentes hierarquias, dentre as quais destacam-se a necessidade de elaboração de uma norma estratégica, o Projeto Municipal de Anonimização, que estabeleça as diretrizes necessárias aos órgãos municipais, bem como o treinamento respectivo dos gestores aplicáveis.

Este relatório apresentou medidas aptas a serem adotadas pelo Poder Público carioca, no sentido do planejamento estratégico e dos processos de anonimização dos órgãos municipais. Não obstante o vácuo normativo atual, as orientações das entidades supramencionadas demonstram ser possível aumentar a proteção de dados pessoais por meio da anonimização e pseudonimização, com redução de risco de reidentificações e incidentes de Segurança da Informação. Espera-se que, com base nessas recomendações, a Administração Pública municipal eleve sua maturidade no cumprimento da LGPD a níveis internacionais, angariando a confiança da população carioca e estrangeiro.