tradução não-oficial (1.ª versão)

APPA Technology Working Group

Guia de

Introdução à Anonimização

Publicado em Junho de 2025

Introdução

Este guia apresenta uma visão geral dos conceitos básicos de anonimização e dos passos práticos que as organizações podem pôr em prática para iniciar a sua trajectória de anonimização, a partir de conjuntos de dados estruturados, textuais e pouco complexos. Pretende-se que este guia permita aos leitores adquirir uma compreensão fundamental sobre anonimização antes de consultarem recursos mais avançados ou orientações mais específicas das respectivas jurisdições.

A anonimização adequada requer bom conhecimento do contexto dos dados e competência técnica no processo de anonimização. Quando a anonimização é considerada complexa ou quando o responsável pelo tratamento dos dados não possui as competências necessárias, deve ponderar a contratação de um especialista para efectuar a anonimização. Importa notar que as técnicas de anonimização e de reidentificação constituem áreas de investigação em constante evolução, com novas metodologias publicadas periodicamente. Os responsáveis pelo tratamento de dados que planeiam implementar a anonimização devem acompanhar os mais recentes desenvolvimentos ou procurar aconselhamento especializado.

Recomenda-se igualmente a consulta da norma ISO intitulada "*Information Security, Cybersecurity and Privacy Protection – Privacy Enhancing Data De-identification Framework*" (*ISO/IEC 27559:2022*[1]). Esta norma reconhece que a anonimização abrange não apenas os dados em si, mas também o contexto em que são partilhados e utilizados, bem como as práticas de governação aplicadas. A adopção destas boas práticas pode auxiliar as organizações a gerir eficazmente o risco de reidentificação, preservando ao mesmo tempo a utilidade dos dados anonimizados.

Declaração: Este guia foi elaborado apenas para fins informativos. Aborda a anonimização sob uma perspectiva técnica, e não sob uma perspectiva jurídica ou de políticas públicas. Não contém aconselhamento jurídico, nem as opiniões nele expressas reflectem necessariamente a política ou posição oficial das "*Asia Pacific Privacy Authorities" (APPA)* ou dos membros do "*APPA Technology Working Group" (TWG)*. Este guia não esgota a abordagem de todas as questões relacionadas com a anonimização. Não substitui as orientações regulamentares emitidas pelas jurisdições competentes. As organizações devem consultar essas orientações regulamentares, sempre que existam, para assegurar a conformidade com as normas aplicáveis de protecção de dados.

[1] *Importa salientar que a norma ISO/IEC 27559 foi redigida partindo do pressuposto de uma distinção estritamente binária entre informação sujeita às leis de privacidade (i.e. informação «pessoal») e informação fora do âmbito dessas leis (i.e. informação «anonimizada»). Enquanto as primeiras leis de privacidade funcionavam sob este mesmo pressuposto, as leis mais modernas tendem a adoptar uma distinção tripla mais flexível entre informação pessoal, «desidentificada» (ou «pseudonimizada») e informação anonimizada, sendo que a informação desidentificada (ou pseudonimizada) permanece sujeita às leis de privacidade. Face a esta discrepância, não é claro se, e em que circunstâncias, a informação «anonimizada» de acordo com a norma ISO/IEC 27559 deve ser considerada como desidentificada (ou pseudonimizada) no quadro das actuais normas de privacidade.*

O que é a Anonimização?

A nível técnico, a anonimização[2] é o processo de converter dados pessoais em dados que já não podem ser usados para identificar um indivíduo[3], isoladamente ou em combinação com outras informações, tomando medidas razoáveis que tenham em conta o estado actual da técnica.

Porquê Anonimizar Dados?

Em regra, os dados que passaram por anonimização deixam de ser considerados dados pessoais. Contudo, é importante reconhecer que o processo de anonimização exige uma avaliação rigorosa, gestão de riscos e governação contínua, conforme definido na *ISO/IEC 27559 – "Privacy Enhancing Data De-identification Framework"*. Tal processo inclui avaliações de contexto, avaliação de dados, avaliação de identificabilidade e a implementação de medidas robustas de mitigação para assegurar que o risco de reidentificação permaneça abaixo de um nível de tolerância predefinido. A monitorização contínua e a observância das práticas estabelecidas de anonimização são essenciais para manter o estado não pessoal dos dados anonimizados.

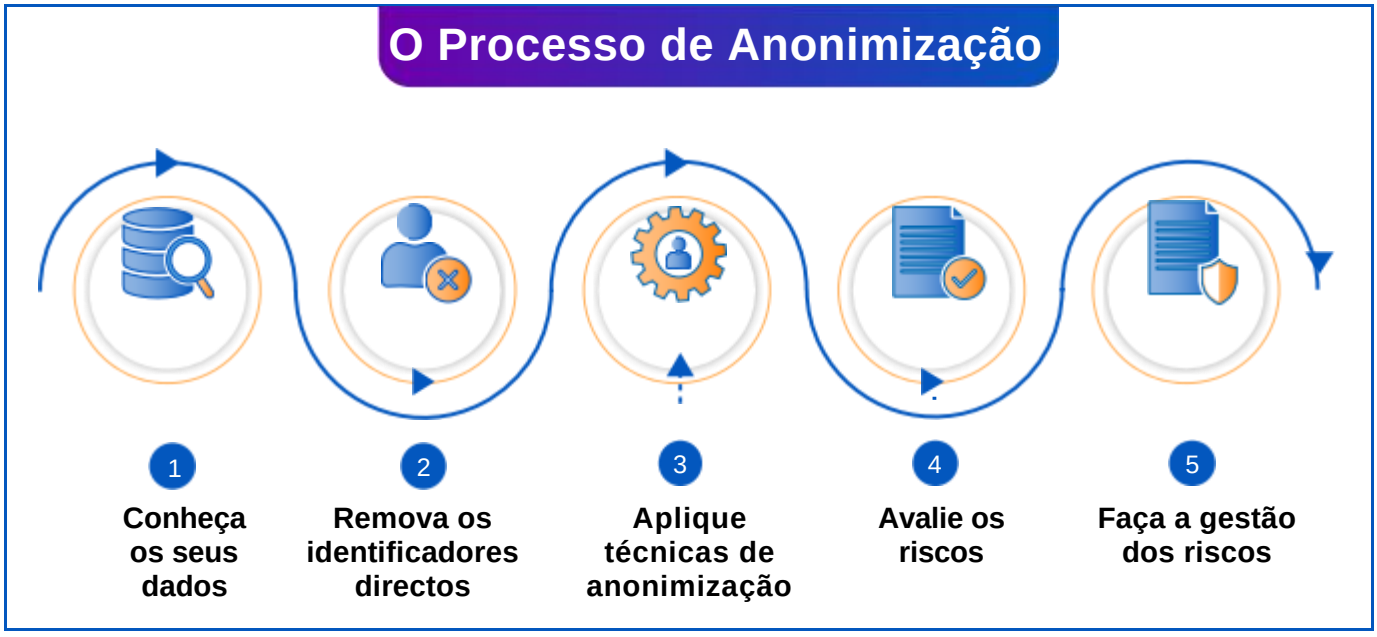
As boas práticas e as normas, como as descritas na ISO/IEC 27559, tratam a anonimização como um processo baseado em riscos, que engloba a aplicação de técnicas de anonimização aos dados, bem como a implementação de outras medidas de privacidade e de segurança destinadas a mitigar riscos de reidentificação. Este guia propõe passos práticos para apoiar as organizações na avaliação e redução desses riscos, garantindo simultaneamente a utilidade dos dados.

Ao anonimizar dados, por exemplo, no momento de partilha de dados com entidades externas, as organizações podem obter conhecimento útil a partir dos dados e, ao mesmo tempo, proteger os titulares dos mesmos. As leis de protecção de dados podem ainda obrigar as organizações a eliminar ou a anonimizar dados pessoais sempre que não exista fundamento válido para a sua conservação.

[2] Nas legislações de protecção de dados de alguns países, o termo anonimização é referido como desidentificação, enquanto certas definições podem considerar desidentificação como apenas a remoção de identificadores directos. Além disso, algumas jurisdições podem considerar que, se ainda existirem riscos a gerir, tais dados apenas podem ser designados como pseudonimizados e não como anonimizados. As normas legais sobre o que é considerado dado anonimizado ou desidentificado podem variar entre jurisdições de protecção de dados.

[3] O Passo 4 deste documento aborda a forma de avaliar os riscos de reidentificação.

[4] Os Passos 1 a 5 deste guia são aplicáveis às organizações abrangidas por tais jurisdições.



O diagrama acima apresenta uma visão simplificada[5] do processo de anonimização sugerido.

Passo 1: Conheça os seus dados

Um registo de dados pessoais é composto por atributos de dados que possuem diferentes graus de identificabilidade e sensibilidade relativamente a um indivíduo. A anonimização envolve, em regra, a remoção de identificadores directos e a modificação de identificadores indirectos. Os atributos-alvo permanecem, em geral, inalterados.

Identificadores directos são atributos de dados geralmente únicos para um indivíduo e podem ser usados como atributos-chave no registo de dados para voltar a identificar um indivíduo. Exemplos comuns de identificadores directos são nome e número de identificação nacional.

Identificadores indirectos são atributos de dados geralmente não únicos para um indivíduo, mas suas combinações podem ser únicas e, assim, permitir voltar a identificar um indivíduo no registo de dados quando combinadas com outras informações, incluindo identificadores directos. Exemplos comuns de identificadores indirectos são data de nascimento, género e código postal.

[5] Os passos não têm carácter prescritivo, servem apenas de orientação geral sobre o processo de anonimização. As organizações podem ajustá-los de acordo com o seu contexto e processos internos. Por exemplo, a avaliação de riscos pode ser realizada antes da aplicação das técnicas de anonimização.

Atributos-alvo são os atributos de dados remanescentes que contêm a principal utilidade do conjunto de dados. No contexto da avaliação da adequação da anonimização, este atributo de dados pode ter natureza sensível e pode resultar num elevado potencial de efeitos adversos para um indivíduo, caso seja divulgado. Estes atributos de dados não se encontram geralmente disponíveis ao público nem são facilmente acessíveis. Um exemplo de atributo-alvo pode ser o diagnóstico de saúde de um indivíduo. Deve ser dada atenção e cautela adicionais sempre que tais atributos-alvo possam ser facilmente acessíveis ou, de outro modo, disponíveis, devendo considerar-se a sua categorização como atributos de dados indirectos.

Passo 2: Remova os identificadores directos

Remover todos os identificadores directos. Caso seja necessário, atribuir um pseudónimo a cada registo. Os pseudónimos devem ser únicos para cada identificador directo. A atribuição de pseudónimos deve ser igualmente robusta, significando que não devem conter informação identificável nem ser reversíveis por partes não autorizadas, seja por tentativa ou por cálculo dos valores originais dos identificadores directos a partir dos pseudónimos.

Passo 3: Aplique técnicas de anonimização

Neste passo, aplicar técnicas de anonimização aos identificadores indirectos, de modo que não possam ser combinados com outros conjuntos de dados que possam conter informação adicional para reidentificar indivíduos. Deve notar-se que a aplicação destas técnicas modificará os valores dos dados e poderá afectar a utilidade dos dados anonimizados. As técnicas de anonimização incluem supressão de dados, mascaramento, generalização, adição de ruído aos dados, amostragem e permutação de dados. A norma *ISO/IEC 20889*, intitulada "*Privacy Enhancing Data De-identification Terminology and Classification of Techniques"*, fornece uma lista abrangente de possíveis técnicas de anonimização.

É importante escolher a(s) técnica(s) adequada(s) ao conjunto de dados e à forma como os dados serão utilizados. É da responsabilidade das organizações assegurar que seleccionam a(s) técnica(s) apropriada(s) às circunstâncias. Diferentes técnicas apresentam vantagens e desvantagens, tanto para a privacidade como para a utilidade, e as organizações devem estar cientes destas para tomar decisões informadas e deliberadas sobre qual(is) técnica(s) utilizar.

Passo 4: Avalie os riscos de reidentificação

É útil calcular os riscos de reidentificação dos dados anonimizados. Métodos como o anonimato-*k*[6] podem ser utilizados para esse fim. O anonimato-*k* pode não ser adequado para todos os conjuntos de dados. Outras abordagens ou ferramentas, como a "*Special Unique Detection Algorithms*" (*SUDA*) e o *µ-Argus*, também podem ser consideradas pelas organizações para avaliar o risco de reidentificação de conjuntos de dados partilhados.

O anonimato-*k* é um método simples para calcular o nível de risco de reidentificação de um conjunto de dados, desde que este não seja complexo e não contenha um grande número de atributos. Essencialmente, refere-se ao menor número de registos idênticos que podem ser agrupados no conjunto de dados. Esse número é geralmente tomado como representando o pior cenário possível na avaliação do risco geral de reidentificação do conjunto de dados. Normalmente, apenas identificadores indirectos são considerados no cálculo do anonimato-*k*. Um valor de anonimato-*k* mais elevado significa um risco menor de reidentificação, enquanto um valor mais baixo implica um risco maior. A Figura 1 abaixo apresenta um exemplo simples de um conjunto de dados com três grupos de registos idênticos, cujos valores de *k* variam entre 2 e 4. Globalmente, o valor de anonimato-*k* do conjunto é 2, reflectindo o menor valor (maior risco)[7] em todo o conjunto de dados.

Código postal	Idade	Programa favorito		
22xxxx	21 a 25	*Emily in Paris*	*k*=2	
22xxxx	21 a 25	*Emily in Paris*		
10xxxx	41 a 45	*Brooklyn Nine-Nine*	*k*=4	**No geral k=2**
10xxxx	41 a 45	*Brooklyn Nine-Nine*		
10xxxx	41 a 45	*Brooklyn Nine-Nine*		
10xxxx	41 a 45	*Brooklyn Nine-Nine*		
58xxxx	56 a 60	*Attenborough's Life in Colour*	*k*=3	
58xxxx	56 a 60	*Attenborough's Life in Colour*		
58xxxx	56 a 60	*Attenborough's Life in Colour*		

Figura 1: Ilustração do anonimato-*k*

[6] *Em geral, a suficiência do processo de anonimização é avaliada caso a caso. A "Singapore's Personal Data Protection Commission" recomenda um valor de k de, pelo menos, 5 para que os dados sejam considerados suficientemente anonimizados, juntamente com as salvaguardas relevantes. Além disso, a técnica de anonimização por agrupamento (anteriormente designada anonimato-k) é usada principalmente para proteger contra-ataques de ligação e de individualização. Extensões desta técnica, como a diversidade-l e a proximidade-t, podem ser consideradas para proteger contra outros tipos de ataques, como ataques de inferência.*

[7] *Focar no maior risco é a abordagem mais conservadora, pois considera o risco máximo de reidentificação dentro de um conjunto de dados. Existem também outras abordagens, como o risco médio e o risco médio estrito.*

Um teste[8] de "*motivated intruder*" pode ser conduzido para avaliar o risco residual de reidentificação depois de aplicadas as técnicas de anonimização. Este teste considera se os indivíduos podem ser reidentificados a partir de dados anonimizados por alguém que esteja motivado, razoavelmente competente, tenha acesso a dados públicos ou privados disponíveis (por exemplo, a Internet, bases de dados comerciais e informação publicada em diretórios públicos) e utilize técnicas investigativas padrão (como inferências e ligação de dados).

As organizações devem usar as suas avaliações de risco de reidentificação para garantir que os seus dados estão suficientemente anonimizados. Isto pode exigir que as organizações ajustem a(s) sua(s) técnica(s), apliquem medidas adicionais, revejam o âmbito do conjunto de dados, remover valores atípicos, etc. Assim, pode ser necessário regressar ao Passo 3 ou até a passos anteriores e repetir este ciclo até que os resultados do Passo 4 sejam satisfatórios, ou seja, até que os dados estejam suficientemente anonimizados.

Passo 5: Faça a gestão dos riscos de reidentificação

As medidas de mitigação servem para gerir qualquer risco residual após a aplicação das técnicas de anonimização. Geralmente, serão necessárias medidas de mitigação mais fortes para conjuntos de dados anonimizados com maior risco residual (por exemplo, valores baixos de *k* na anonimização por agrupamento). As organizações também devem considerar o potencial grau de dano aos indivíduos no caso de reidentificação para determinar as medidas de mitigação a serem implementadas.

As medidas de mitigação incluem geralmente medidas de segurança para garantir que os dados anonimizados partilhados só possam ser acedidos por utilizadores autorizados, bem como medidas legais e/ou de governação para assegurar que os dados são usados apenas para os fins previstos.

Devem também ser implementadas medidas de mitigação para proteger o mapeamento da identidade ou a ligação de informações dos dados anonimizados a dados pessoalmente identificáveis, caso os controladores dos dados optem por reter tal informação.

Embora a aplicação de medidas de mitigação possa ser vista como parte do processo de anonimização, também é possível entendê-la como uma extensão das obrigações legais de protecção de dados. Se, após o Passo 4, o risco de reidentificação não for suficientemente baixo para que os dados sejam considerados anonimizados, então os dados devem continuar a ser considerados identificáveis. Isso significa que as leis de protecção de dados continuarão a ser aplicáveis. Em determinadas jurisdições, a aplicação de medidas de mitigação pode ser vista como uma condição para considerar que o processo de anonimização foi concluído com sucesso, enquanto noutras jurisdições pode ser necessária uma interpretação baseada em avaliação de riscos. As organizações devem consultar os regulamentos das jurisdições específicas para essa interpretação.

[8] *Destacado no "Information Commissioner's Office (UK)'s Code of Practice, Anonymisation: Managing Data Protection Risk Code of Practice".*

Considerações úteis da ISO/IEC 27559

Além do processo de 5 passos descritos acima, as organizações podem também consultar a ISO/IEC 27559, que fornece um quadro útil e considerações para avaliação de contexto, avaliação de dados, avaliação e mitigação da identificabilidade, e governação.

Avaliação de Contexto: Consiste em avaliar o ambiente e as circunstâncias em que os dados são disponibilizados aos destinatários. Inclui a modelagem de ameaças, a avaliação das práticas de segurança e privacidade, bem como a avaliação das motivações e da capacidade de reidentificação.

Avaliação de Dados: Foca-se na compreensão das características dos dados e na modelagem de potenciais ataques que poderiam explorar vulnerabilidades. Abrange as características dos dados, a modelagem de ataques e a selecção de modelos de privacidade de dados para quantificar a identificabilidade.

Avaliação e Mitigação da Identificabilidade: Consiste em avaliar a probabilidade de um ataque e a probabilidade de identificação bem-sucedida de um indivíduo. A estrutura descreve métodos para quantificar a identificabilidade e sugerir medidas de mitigação, como reconfigurar o ambiente ou transformar os dados para reduzir a identificabilidade.

Governação da Desidentificação/Anonimização: A governação inclui o estabelecimento de princípios, políticas e procedimentos para gerir as actividades de processamento de dados e garantir a conformidade com as normas de segurança e privacidade de dados. Também abrange papéis e responsabilidades, monitoração de riscos e gestão de divulgações não intencionais.

Outras práticas recomendadas

Devem ser realizadas revisões periódicas para garantir que o risco de reidentificação permaneça baixo ao longo do tempo. As organizações precisam assegurar que nenhuma nova tecnologia ou técnica tenha evoluído, ou que nenhum novo conjunto de dados tenha sido disponibilizado ou tornado público, permitindo a reidentificação dos dados anonimizados. As revisões também garantirão que as medidas de mitigação de risco sejam eficazes, estejam funcionando conforme o previsto e continuem apropriadas caso as circunstâncias organizacionais mudem.

Se as autoridades de protecção de dados consideram ou não um conjunto de dados como anonimizado geralmente depende da probabilidade de reidentificação. Dependendo da jurisdição, as autoridades de protecção de dados também podem considerar as salvaguardas (incluindo medidas técnicas, de governação e contratuais) e o processo de anonimização utilizado. Por isso, é útil manter registos de tais informações quando a anonimização é realizada.

Anexo A: Recursos sobre Anonimização

ISO/IEC

A ISO/IEC publicou, entre outros, os seguintes padrões relacionados à anonimização.

Recurso	Descrição
ISO/IEC 20889:2018 – "Privacy enhancing data de-identification terminology and classification of techniques"	Este padrão especifica a terminologia, uma classificação de técnicas de desidentificação de dados de acordo com suas características e sua aplicabilidade para reduzir o risco de reidentificação. Disponível em (apenas em inglês) https://www.iso.org/standard/69373.html
ISO/IEC 27559:2022 – "Information security, cybersecurity, and privacy protection – Privacy enhancing data de-identification framework"	Este padrão fornece uma estrutura para identificar e mitigar riscos de reidentificação e riscos associados ao ciclo de vida dos dados desidentificados. Pode ajudar as organizações a determinar como um processo de desidentificação é implementado na prática. É um padrão baseado em elementos comuns para ampla adopção e melhores práticas. Adota uma abordagem baseada em riscos. Disponível em (apenas em inglês) https://www.iso.org/standard/71677.html

Austrália (*Commonwealth*)

O "*Office of the Australian Information Commissioner*" (*OAIC*) fornece orientações relacionadas à desidentificação de dados sob as leis nacionais de privacidade da Austrália. Garantir que a informação tenha sido "desidentificada" para fins da Lei de Privacidade de 1988 (Cth) exige que as entidades adoptem uma abordagem semelhante à "anonimização".

O "*De-identification Decision-Making Framework*", publicado em conjunto pelo OAIC e pela Data61 da *Commonwealth Scientific and Industrial Research Organization (CSIRO),* auxilia as organizações a desidentificar seus dados de forma eficaz. O *Framework* é um recurso prático e acessível para organizações australianas que lidam com informações pessoais e estão a considerar compartilhar ou divulgar tais informações para cumprir responsabilidades éticas e obrigações legais.

O OAIC também fornece orientações sobre a "*De-identification and the Privacy Act*", que incluem conselhos gerais sobre desidentificação e protecção da privacidade para maximizar a utilidade e o valor dos dados enquanto se salvaguarda a privacidade.

Recurso	Disponível em (apenas em inglês)
"*De-identification Decision-Making Framework*"	https://www.oaic.gov.au/privacy/privacy-guidance for-organisations-and-government-agencies/ handling-personal-information/de-identification decision-making-framework
"*De-identification and the Privacy Act*"	www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/ handling-personal-information/de-identification-and-the-privacy-act

Victoria, Austrália

O "*Office of the Victorian Information Commissioner*" (OVIC) disponibiliza um conjunto de recursos gratuitos no seu *website*, abrangendo:

- uma introdução elementar à desidentificação – o que significa, como funciona, assim como os riscos e desafios envolvidos,
- dicas práticas sobre o que considerar aquando da desidentificação de informações, e como gerir os riscos de reidentificação e
- orientações mais aprofundadas acerca de técnicas de desidentificação e as limitações destes métodos na protecção de informações pessoais ao nível registo unitário.

Na jurisdição do OVIC, o termo "desidentificação" é usado para descrever "anonimização", conforme definido neste guia.

Recurso	Disponível em (apenas em inglês)
"*An introduction to de identification*"	https://ovic.vic.gov.au/privacy/resources-for-organisations/an-introduction-to-de-identification/
"*De-identification: An exercise in risk management*"	https://vimeo.com/722443647
"*Limitations of de-identification*"	https://ovic.vic.gov.au/privacy/resources-for-organisations/the-limitations-of-de-identification-protecting-unit-record-level-personal-information/

Coreia do Sul

A "*Personal Information Protection Commission*" (*PIPC*) da Coreia do Sul explica a governação da informação pseudonimizada através das suas "*Pseudonymised Information Processing Guidelines*", alteradas em 2022.

A "*Personal Information Protection Act*" (*PIPA*) da Coreia do Sul estipula que a informação pseudonimizada é considerada informação pessoal. A *PIPC* publicou directrizes para uma melhor compreensão do processamento, combinação, exportação e medidas de segurança aplicáveis à informação pseudonimizada.

As directrizes especificam precauções para cada processo de tratamento de informação pseudonimizada e como proteger os direitos dos titulares dos dados através de medidas técnicas, administrativas e físicas de protecção.

Adicionalmente, em resposta à crescente procura de utilização de dados não estruturados, tais como imagens e gravações de vídeo, que se tornaram disponíveis com o avanço da tecnologia de Inteligência Artificial, as Directrizes publicadas (Fevereiro de 2024) incluem casos e cenários detalhados para diferentes áreas, incluindo cuidados de saúde, tráfego e serviços de *chatbot*.

Recurso	Disponível em
Brief Overview of Pseudonymization	※apenas em inglês https://www.pipc.go.kr/eng/user/lgp/bnp/pseudonymization.do
*Guidelines for Pseudonymizing Unstructured Data*Formato PDF. Versão resumida em inglês)	※apenas em inglês https://www.pipc.go.kr/eng/user/lgp/law/ordinancesDetail.do?bbsId=BBSMSTR_000000000005&nttId=2699#none
Guidelines for Pseudonymizing (versão de 2022)	※apenas em coreano https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS217&mCode=G010030000&nttId=8000

Japão

A "*Personal Information Protection Commission*" (*PPC*) publicou directrizes sobre o tratamento de informações pessoais anonimizadas ao abrigo da "*Act on the Protection of Personal Information*" (*APPI*), para apoiar a implementação adequada e eficaz de medidas de protecção dessas informações por organizações. A PPC publicou ainda um relatório com mais detalhes sobre o tratamento de informações pessoais anonimizadas, a fim de facilitar os esforços de autorregulação das organizações. Estes materiais também abordam informações pessoais pseudonimizadas, que a APPI estipula terem natureza diferente das informações pessoais anonimizadas.

Recurso	Disponível em
Directrizes	※ apenas em japonês https://www.ppc.go.jp/personalinfo/legal/#anc_Guide
Relatório do "*Personal Information Protection Commission Secretariat*"	(Edição sobre o Sistema Jurídico) ※ apenas em japonês https://www.ppc.go.jp/files/pdf/report_office_seido2205.pdf (Edição com Estudo de Caso) ※ apenas em japonês https://www.ppc.go.jp/files/pdf/report_office_zirei2205.pdf

Singapura

A "*Personal Data Protection Commission*" (*PDPC*) publicou directrizes consultivas para explicar como a anonimização é definida ao abrigo da "*Personal Data Protection Act*" (*PDPA*) e quais são os requisitos para que os dados sejam considerados anonimizados. A PDPC publicou também um guia técnico sobre anonimização, que fornece sugestões quanto a aspectos de implementação. Por fim, a ferramenta de anonimização de dados é gratuita e complementa o guia técnico, sendo útil para os leitores aprenderem sobre anonimização e para anonimizar conjuntos de dados simples.

Recurso	Disponível em
"*Advisory Guidelines on the Personal Data Protection Act for Selected Topics*" (Capítulo 3 – Anonimização)	※apenas em inglês https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/ag-on-selected-topics/advisory-guidelines-on-the-pdpa-for-selected-topics-(revised-may-2024).pdf
• " *Guide to Basic Anonymisation*" • *Ferramenta de anonimização de dados*	※Versão inglesa: https://www.pdpc.gov.sg/Help-and-Resources/2018/01/Basic-Anonymisation ※Versão espanhola (publicada pela AEPD, autoridade de protecção de dados da Espanha): https://www.aepd.es/es/documento/guia-basica anonimizacion.pdf https://www.aepd.es/es/descargas/herramienta anonimizacion-pdpc
[vídeo] *Introduction to PDPC's Data Anonymisation Tool*	※apenas em inglês www.youtube.com/watch?v=qlnYRl5VwQQ

Anexo B: Estudo de Caso

Esta secção apresenta um exemplo hipotético para ilustrar os passos descritos anteriormente.

O cenário aqui exposto refere-se a um ginásio, Vivogym, que pretende partilhar dados anonimizados com um parceiro de *marketing*, The Pink Group, com o objectivo de traçar o perfil dos seus clientes e criar uma nova campanha de *marketing*.

A tabela abaixo apresenta um excerto dos dados originais dos clientes da base de dados do Vivogym. Todos os dados utilizados neste exemplo são fictícios. Neste guia, são apresentados 10 registos do conjunto de dados.

S/N	Nome	Data de Nascimento	Código Postal (Singapura)	Peso (kg)	Altura (m)	Actividade Mais Frequente nos Últimos 6 Meses
1	Demond Nix	9/12/1996	322607	52	1.60	Passadeira
2	Treyvon Coker	24/12/1998	335662	56	1.75	Pilates
3	Jarred Zielinski	3/10/1995	355895	72	1.65	Natação
4	Rolando Toth	10/12/1996	359383	79	1.67	Ciclismo estático
5	Benny Beckman	8/12/1996	316551	65	1.60	Musculação
6	Dakota Birch	5/9/1997	326125	66	1.75	Musculação
7	Jacques Colburn	4/9/1995	339035	72	1.68	Kickboxing
8	Kendyl Fletcher	25/10/1999	346214	79	1.72	Musculação
9	Keegan Knapp	26/10/1997	346204	59	1.62	Pilates
10	Yoselin Provost	4/9/1995	324946	61	1.75	Musculação
(existem muitos outros registos não listados neste documento por questões de espaço)						

Passo 1: Conheça os seus dados

Neste exemplo, o Vivogym classifica os atributos dos dados da seguinte forma:

Atributo de Dados	Nome	Data de Nascimento	Código Postal (Singapura)	Peso (kg)	Altura (m)	Mais Tempo Gasto nos Últimos 6 Meses
Classificação	Identificador directo	Identificador indirecto	Identificador indirecto	Identificador indirecto	Identificador indirecto	Atributo de destino

Passo 2: Remova os identificadores directos

Neste exemplo, o único identificador directo é o nome do cliente. Este dado é removido do conjunto de dados.

S/N	~~Nome~~	Data de Nascimento	Código Postal (Singapura)	Peso (kg)	Altura (m)	MaisTempo Gasto nos Últimos 6 Meses
1	~~Demond Nix~~	9/12/1996	322607	52	1.60	Passadeira
2	~~Treyvon Coker~~	24/12/1998	335662	56	1.75	Pilates
3	~~Jarred Zielinski~~	3/10/1995	355895	72	1.65	Natação
4	~~Rolando Toth~~	10/12/1996	359383	79	1.67	Ciclismo estático
5	~~Benny Beckman~~	8/12/1996	316551	65	1.60	Musculação
6	~~Dakota Birch~~	5/9/1997	326125	66	1.75	Musculação
7	~~Jacques Colburn~~	4/9/1995	339035	72	1.68	Kickboxing
8	~~Kendyl Fletcher~~	25/10/1999	346214	79	1.72	Musculação
9	~~Keegan Knapp~~	26/10/1997	346204	59	1.62	Pilates
10	~~Yoselin Provost~~	4/9/1995	324946	61	1.75	Musculação
(existem muitos outros registos não listados neste documento por questões de espaço)						

Passo 3: Aplique técnicas de anonimização

O ginásio decide como anonimizar os vários identificadores indirectos, conforme indicado na tabela seguinte:

Identificador Indirecto	Técnica de Anonimização
Data de Nascimento	Generalizar a partir de uma data exacta para apenas o ano de nascimento.
Código Postal (Singapura)	Ocultar os últimos 4 dígitos do código postal de 6 dígitos. Esta alteração evita que o valor aponte para uma unidade residencial específica, reduzindo a granularidade para um nível distrital menos preciso.
Peso (kg)	Combinar o peso e a altura num valor de Índice de Massa Corporal (IMC), generalizando de seguida para intervalos numéricos de 10 unidades.
Altura (m)	

Após a anonimização, o conjunto de dados fica assim:

S/N	Data de Nascimento	Código Postal (Singapura)	IMC	**Mais Tempo Gasto nos Últimos 6 Meses**
1	1996	32****	20 - 29	Passadeira
2	1998	33****	10 - 19	Pilates
3	1995	35****	20 - 29	Natação
4	1996	35****	20 - 29	Ciclismo estático
5	1996	31****	20 - 29	Musculação
6	1997	32****	20 - 29	Musculação
7	1995	33****	20 - 29	Kickboxing
8	1999	34****	20 - 29	Musculação
9	1997	34****	20 - 29	Pilates
10	1995	32****	10 - 19	Musculação
(existem muitos outros registos não listados neste documento por questões de espaço)				

Passo 4: Avalie os riscos de reidentificação

O ginásio agrupa, os registos semelhantes no conjunto de dados, isto é, registos que partilham a mesma data de nascimento, código postal e IMC. A similaridade dos registos é avaliada sem considerar o atributo "Mais Tempo Gasto nos Últimos 6 Meses", uma vez que este é classificado como atributo-alvo e, por esse motivo, não é utilizado no cálculo do anonimato-*k*.

S/N	Ano de Nascimento	Código Postal (Singapura)	IMC	Número de registos no conjunto de dados com os mesmos identificadores indirectos (data de nascimento, código postal e IMC), i.e. *k*	Mais Tempo Gasto nos Últimos 6 Meses
1	1996	32****	20 - 29	5	(valores inalterados)
2	1998	33****	10 - 19	6	(valores inalterados)
3	1995	35****	20 - 29	5	(valores inalterados)
4	1996	35****	20 - 29	5	(valores inalterados)
5	1996	31****	20 - 29	6	(valores inalterados)
6	1997	32****	20 - 29	7	(valores inalterados)
7	1995	33****	20 - 29	5	(valores inalterados)
8	1999	34****	20 - 29	4	(valores inalterados)
9	1997	34****	20 - 29	6	(valores inalterados)
10	1995	32****	10 - 19	5	(valores inalterados)
(existem muitos outros registos não listados neste documento por questões de espaço)					

O valor de anonimato-*k* pode ser calculado por ferramentas especializadas[9] ou utilizando *software* de folha de cálculo, calculando o número de registos que contêm os mesmos identificadores indirectos.

Neste exemplo, inicialmente o valor geral de anonimato-*k* do conjunto de dados é 4, reflectindo o grupo de registos com o menor valor de k (maior risco) dentro de todo o conjunto. Caso o ginásio decida melhorar o valor geral de *k* para 5, poderá considerar a remoção dos registos discrepantes (a vermelho, com valor de *k* inferior a 5) para aumentar o valor de anonimato-*k* de 4 para 5.

O ginásio avalia os riscos de reidentificação do registo fazendo o teste do "*motivated intruder*" para medir a probabilidade de reidentificação a partir dos dados anonimizados. O ginásio também leva em consideração quaisquer potenciais danos que possam surgir aos indivíduos caso ocorra a reidentificação, bem como as orientações da jurisdição em que actua. Com base nisso, o ginásio repetirá os passos 3 e 4 até alcançar um nível razoavelmente alto de anonimato-*k*, garantindo que o conjunto de dados permaneça útil à sua finalidade (perfilamento de clientes).

[9] *Estão disponíveis diversas ferramentas gratuitas e comerciais de anonimização.*

Passo 5: Faça a gestão dos riscos de reidentificação

O ginásio implementa as seguintes salvaguardas para assegurar que qualquer risco residual de reidentificação é mitigado ou razoavelmente minimizado:

- **Salvaguardas contratuais: no contrato com a empresa de *marketing*:**
  - Restringir o uso dos dados anonimizados unicamente aos fins previstos e ao pessoal autorizado,
  - Proibir tentativas deliberadas de reidentificação, e
  - Exigir a eliminação dos dados anonimizados após o cumprimento dos fins para os quais foram fornecidos ou quando os dados deixarem de ser necessários.

- **Salvaguardas técnicas:**
  - Implementar controlos de acesso na empresa de marketing para limitar o acesso ao conjunto de dados anonimizados apenas a empregados autorizados, e
  - Manter registos do conjunto de dados anonimizados e detalhes do processo de anonimização, bem como um registo das actividades de partilha de dados.

- **Salvaguardas de governação:**
  - Manter registos do conjunto de dados anonimizados e detalhes do processo de anonimização, bem como um registo das actividades de partilha de dados.

Agradecimentos

Este guia foi desenvolvido pelo "*Technology Working Group*" (TWG) das "*Asia Pacific Privacy Authorities*" (APPA), composto pelos seguintes membros:

- "*Office of the Victorian Information Commissioner*", Vitória, Austrália (*OVIC*)
- "*Office of the Privacy Commissioner*", Canadá (*OPC*)
- "*Office of the Information and Privacy Commissioner*", Colúmbia Britânica, Canadá (*OIPC*)
- "*Office of the Privacy Commissioner for Personal Data*", Hong Kong, China (*PCPD*)
- "*Personal Data Protection Bureau*", Macau, China (*PDPB*)
- "*The Personal Information Protection Commission*", Japão (*PPC*)
- "*Office of the Privacy Commissioner*", Nova Zelândia (*OPC*)
- "*Personal Data Protection Commission*", Singapura (*PDPC*)
- "*Korea Personal Information Protection Commission*", Coreia do Sul (*PIPC*)


APPA
Asia Pacific Privacy Authorities