Compreendendo o conceito de anonimização e dado anonimizado

Bruno Ricardo Bioni[1]
Professor

1. Dados anonimizados como a antítese de dados pessoais: o filtro da razoabilidade

A antítese do conceito de dado pessoal seria um dado anônimo, ou seja, aquele que é incapaz de revelar a identidade de uma pessoa. Diante do próprio significado do termo, anônimo seria aquele que não tem nome nem rosto (HOUAISS; VILLAR, 2009, p. 140).

Essa inaptidão pode ser fruto de um processo pelo qual é quebrado o vínculo entre o(s) dado(s) e seu(s) respectivo(s) titular(es), o que é chamado de anonimização (DONEDA, 2006, p. 44). Esse processo pode se valer de diferentes técnicas que buscam eliminar tais elementos identificadores de uma base de dados (COUNCIL OF EUROPE, 2018), variando entre: a) supressão; b) generalização; c) randomização e; d) pseudoanonimização[2].

Com maior ou menor grau de intensidade – *e.g.*, supressão ou generalização – nota-se um método cujo mote é gerenciar circunstancialmente a *identificabilidade* de uma base de dados. As características de cada dado e a percepção de eles estarem inseridos em uma gama de informações devem orientar tal análise.

Por isso, não há um único método ou uma combinação perfeita *ex ante* para parametrizar o processo de anonimização, devendo-se analisar contextualmente como este deve ser empreendido para que os titulares dos dados anonimizados não sejam reidentificados, nem mesmo por quem procedeu à sua anonimização.

Amarrar o conceito teórico de dados anônimos a uma *análise contextual*, com os olhos voltados para a irreversibilidade do processo de anonimização, joga luz diretamente sobre o fator problemático dessa proposição: o seu caráter elusivo ou mesmo a sua impossibilidade teórica (TEIXEIRA, 2015).

Torna-se cada vez mais recorrente a publicação de estudos que demonstram ser o processo de anonimização algo falível. A representação simbólica de que os vínculos de identificação de uma base de dados poderiam ser completamente eliminados, garantindo-se, com 100% (cem por cento) de eficiência, o anonimato das pessoas, é um mito (NARAYANAN; SHMATIKOV, 2010, p. 24).

Por essa lógica, qualquer dado pessoal anonimizado detém o *risco inerente* de se transmudar em um dado pessoal (TENE, 2013, p. 1.242). A agregação de diversos "pedaços" de informação (dados) pode revelar (identificar) a imagem (sujeito) do quebra-cabeça, a qual era até então desfigurada (anônimo) – o chamado efeito mosaico.

---

[1] Doutorando em Direito Comercial e mestre em Direito Civil pela Universidade de São Paulo. Foi pesquisador visitante do Centro de Tecnologia, Sociedade, Direito e Internet da Universidade de Ottawa e do Departamento de Proteção de Dados Pessoais do Conselho da Europa. É Professor e Fundador do Data Privacy Brasil.

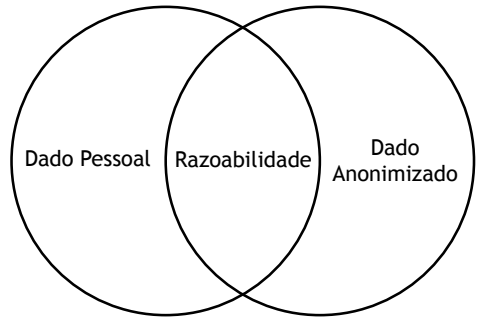
[2] Para muitos, a pseudoanonimização não é considerada uma técnica de anonimização. Isso porque se substituem, apenas, os identificadores diretos – *e.g.*, nome, CPF etc. – por pseudônimos – *e.g.*, números aleatórios, de modo que a pessoa permanece sendo identificável em razão de tais pseudônimos serem um retrato detalhado indireto delas (WP 29, 2014, p. 20).

Por isso, leis que adotam o conceito expansionista[3] de dados pessoais e, ao mesmo tempo, estabelecem uma dicotomia deste com dados anônimos correriam o risco de serem tautológicas. Isso porque haveria uma *redundância normativa*, já que dados anônimos seriam, em última análise, potencial e provavelmente, dados relacionados a uma pessoa identificável.

Para não gerar tal incoerência, a única saída foi a adoção de um "filtro" que delimitasse a *elasticidade* desse conceito expansionista – neste caso o termo identificável –, sob pena de a fronteira entre dados pessoais e dados anônimos ser sempre transponível.

E, nesse sentido, o direito comunitário europeu[4] e a LGPD[5] valeram-se do critério da razoabilidade para delimitar o espectro do conceito expansionista de dados pessoais. Não basta a mera possibilidade de que um dado seja atrelado a uma pessoa para atrair o termo identificável (WP, 2007, p. 1.749). Essa vinculação deve ser objeto de um "esforço razoável"[6], sendo esse o perímetro de elasticidade do conceito de dado pessoal como aquele relacionado a uma pessoa identificável.

A contrario sensu, se para a correlação entre um dado e uma pessoa demanda-se um esforço fora do razoável, não há que se falar em dados pessoais. Nessa situação, o dado é considerado como anônimo, uma vez que o "filtro da razoabilidade" barra o seu enquadramento como aquele relacionado a uma pessoa identificável[7].



Com isso, há coerência em se estabelecer conceitos diferentes para tais espécies de dados, sobretudo sob o ponto de vista de uma dicotomia mutualmente excludente entre eles, que é delimitada pelo fator da razoabilidade[8]. Do contrário, repita-se, haveria uma redundância normativa, na medida em que dados anônimos – sem o critério da razoabilidade – seriam sempre enquadrados dentro do conceito de dado pessoal, como aquele relacionado a uma pessoa identificável.

---

[3] A definição do conceito de dados pessoais pode seguir uma orientação expansionista (a partir da delimitação de "pessoa identificável") ou reducionista ("pessoa identificada"), respectivamente alargando ou restringindo o escopo de aplicação da lei (BIONI, 2019).

[4] A Diretiva 95/46 e a sua proposta de regulamentação adotam os conceitos de razoabilidade, respectivamente, nas considerandas 26 (vinte e seis) e 23 (vinte e três).

[5] Na definição de dados anônimos, de anonimização, bem como no dispositivo que prevê em quais hipóteses um dado anonimizado pode ser considerado como dado pessoal, a LGPD faz alusão ao termo razoável(is) – respectivamente, arts. 5º, II e III, e 18.

[6] Essa é exatamente a terminologia utilizada pelo art. 12, *caput*, da LGPD: "Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido".

[7] Ibidem, p. 21.

[8] Sobre as disputas interpretativas em torno do conceito jurídico indeterminado de razoabilidade, ver: BIONI, Bruno Ricardo. *Xeque-mate*: o tripé de proteção de dados pessoais no xadrez das iniciativas legislativas no Brasil. São Paulo: GPoPAI-USP, 2016, p. 34-35.

2. Calibrando o filtro da razoabilidade: critérios objetivos e subjetivos como fatores de uma análise de risco e os incentivos implícitos à pseudoanonimização

O legislador brasileiro procurou talhar uma norma neutra tecnológica[9]. Ao contrário de apontar para uma tecnologia em específico que poderia se tornar obsoleta ao longo do tempo, utilizou-se de um conceito indeterminado — razoabilidade — a ser significado e atualizado pelo próprio desenvolvimento científico. Simultaneamente, contudo, prescreveu balizas para reduzir a discricionariedade de tal exercício interpretativo e, com isso, alcançar um mínimo de previsibilidade quando tal norma viesse a ser colocada em movimento.

O primeiro eixo de análise é objetivo, sendo composto por uma matriz e dois elementos fatoriais respectivamente[10]: a) estado da arte da tecnologia; a.1) custo e; a.2) tempo[11]. Deve-se analisar o quão custoso e moroso seria reverter um processo de anonimização, de acordo com as tecnologias disponíveis para tanto. Trata-se, portanto, de uma análise dinâmica[12], a ser demarcada pelo próprio progresso tecnológico, que aponta qual deve ser o grau de investimento financeiro e temporal para se reidentificar uma base de dados anonimizada.

Por exemplo, há muito tempo se fala e se espera a chegada da computação quântica[13]. Quando isso acontecer, testemunhar-se-á um verdadeiro progresso acerca da capacidade, em termos quantitativos e qualitativos, de processamento de dados. Consequentemente, atualizar-se-á, por completo, o custo e o tempo quanto ao emprego das técnicas de anonimização, mas, também, por outro lado, das suas respectivas contratecnologias.

Em síntese, o primeiro eixo de análise propõe uma análise acerca do grau de *resiliência* de um processo de anonimização frente aos *padrões sociais*. Uma investigação de ordem objetiva cujo marcador é verificar como o estado da técnica calibra a escala de recursos (custo e tempo) para transmudar um dado anonimizado em dado pessoal.

---

[9] O conceito de "technology-neutral regulation" tem sido evocado para se discutir o desenho de modelos regulatórios capazes de estimular e acompanhar o desenvolvimento tecnológico, sem engessá-lo nem ser permissivo a riscos. Sobre isso: KOOPS, Bert-Jaap. Should ICT regulation be technology-neutral? *In*: KOOPS, Bert-Jaap; LIPS, Miriam; PRINS, Corien; SCHELLEKENS, Maurice (ed.). *Starting points for ICT Regulation*: Deconstructing prevalent policy one-liners. The Hague: TMC Asser Press, 2006. v. 9, p. 77-108. (IT & Law Series). ISBN 90-6704-216-1. REED, Chris. Taking sides on technology neutrality. *SCRIPT-ed*, Edimburgo, v. 4, n. 3, 2007; MOSES, Lyria Bennett. How to think about law, regulation and technology: problems with 'technology' as a regulatory target. *Law, Innovation and Technology*, Abingdon-on-Thames, v. 5, n. 1, p. 1-20, 2013. Para a discussão no cenário nacional, ver: BAPTISTA, Patrícia; KELLER, Clara. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. *Revista de Direito Administrativo*, v. 273, p. 123-163, 2016.

[10] A GDPR, em sua consideranda 26, também utilizada esses três fatores objetivos como delimitação à razoabilidade.

[11] Artigo 12 da LGPD. Art. 12, § 1º: "A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios".

[12] A LGPD, em seu art. 5º, III e XI, define dado anonimizado a partir do emprego dos meios técnicos razoáveis disponíveis na ocasião (III) e no momento (XI) de seu tratamento. Esse tipo de avaliação torna-se, assim, contextual. Se, por um lado, essa análise contextual incentiva estudos sobre o tema, por outro, traz complicações à avaliação de seu cumprimento tendo em vista, por exemplo, diferenças quanto ao acesso à informação e recursos econômicos disponíveis entre os diferentes atores.

[13] Em 08.01.2019 foi lançado o primeiro computador quântico de uso comercial do mundo. Contudo, estima-se um período entre cinco e dez anos para que a computação quântica passe a ser adotada nos negócios. Assim, apesar de existente, essa tecnologia não compreenderia o estado da arte da tecnologia (ou meio técnico razoável disponível, nos termos da LGPD), tornando um encargo demasiado excessivo a expectativa de sua adoção. Disponível em: https://glo.bo/36eoGLd. Acesso em: 20 jan. 2020.

O segundo eixo de análise é subjetivo. Deve-se levar em consideração quem é o agente de tratamento de dados e se ele dispõe de "meios próprios"[14] para reverter o processo de anonimização. Ao invés de considerar quais são os padrões sociais acerca da reversibilidade de um dado anonimizado, foca-se em analisar qual é a capacidade individual de engenharia reversa de quem processa tais dados. Abre-se, com isso, dois vetores importantes de análise.

Em primeiro lugar, sob o ponto de vista do fluxo de dados dentro de uma organização. É cada vez mais comum que organizações segmentem as suas bases de dados de acordo com suas respectivas áreas de negócio e, até mesmo em alguns casos, empreguem práticas de anonimização para a geração de *business intelligence/BIA*.

Por exemplo, é o caso de uma grande rede de lojas varejistas que decide utilizar a sua base de dados de programa de fidelidade para melhorar o seu sistema de distribuição logística. Uma nova finalidade foi atribuída a um conjunto de dados, não sendo necessário saber quem são seus respectivos consumidores de forma individualizada, mas, tão somente, quais produtos têm mais ou menos entrada e saída de acordo com o perfil de vendas de cada um dos seus estabelecimentos geograficamente espalhados. Dessa forma, é factível a estruturação de uma nova base de dados sem que haja a associação direta ou indireta a indivíduos, podendo ser mantida, inclusive, em separado da outra base de dados (programa de fidelidade) que lhe deu origem.

Nesse cenário, o próprio agente tem informações adicionais, ainda que mantidas separadamente, para reverter o processo de anonimização. Ou seja, ele possui meios próprios para transmudar um dado aparentemente anonimizado em um dado pessoal, o que é revelado com base em uma análise subjetiva focada na sua própria capacidade de entropia de informação[15].

O cenário acima descrito é o que se convencionou chamar de pseudoanonimização, ou seja, uma falsa, superficial, técnica de anonimização que é quebrável em especial pela própria organização que a empregou.

A primeira reflexão que pode seguir a esse respeito é: por que a organização deveria empregar todo o esforço acima mencionado, se toda a carga regulatória da legislação de proteção de dados ainda assim recairá sobre ela (o dado não deixará de ser pessoal)?

Diferentemente da GDPR, a legislação de proteção de dados pessoais brasileira não sistematizou adequadamente a figura da pseudoanonimização, muito menos desenhou normativamente incentivos expressos para a sua adoção por parte dos agentes de tratamento de dados. Enquanto o regulamento europeu previu até mesmo o relaxamento de algumas obrigações legais[16], a lei geral brasileira de proteção de dados pessoais apenas citou pseudoanonimização de forma assistemática[17].

---

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

[15] Entropia da informação é o uso de uma informação auxiliar para a reversão do processo de anonimização. No caso em análise, as informações adicionais em posse do agente de tratamento.

[16] O artigo 11 da GDPR estabelece que, se o propósito do tratamento dos dados pessoais não exige (ou não exige mais) que o agente seja capaz de identificar o titular, o agente não será obrigado a manter informações adicionais para identificá-lo. E, por não sê-lo, estará escusado de garantir os direitos de acesso, retificação, exclusão e portabilidade do titular – a menos que o próprio titular, buscando exercer esses direitos, forneça as informações adicionais para sua identificação. Disponível em: http://bit.ly/3arPYRT. Acesso em: 20 jan. 2020.

[17] Artigo 13 da LGPD. Art. 13. "Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de

No entanto, ainda assim, é possível chegar à conclusão de que há sim incentivos, mesmo que indiretos, a serem burilados na lei geral de proteção de dados pessoais. Na medida em que pseudoanonimização é o "meio do caminho"[18] entre um dado pessoal e um dado anonimizado, seria possível correlacioná-la às diversas menções que a LGPD faz para que os agentes de tratamento "sempre que possível" anonimizem os dados[19]. Isto porque a lógica normativa em questão é encarar o processo de retirada dos identificadores de uma base de dados como algo que *minimiza os riscos* de uma atividade de tratamento de dados. Esse é exatamente o mote de técnicas de pseudoanonimização, ainda que não retire por completo o caráter pessoal de um dado.

Soma-se, ainda, o fato de que técnicas de pseudoanonimização podem compor o espectro de medidas, políticas e processos de um programa de governança que é referenciado pela LGPD[20]. E, ainda, por ser uma medida tradicional de segurança da informação que pode reduzir significativamente os impactos de um incidente de segurança, a partir da simples constatação de que: a) uma base de dados pseudoanonimizada pode não ser reversível por terceiros-atacantes[21] e; b) certamente, gera menos riscos em relação a uma base de dados comprometida que não tenha havido o emprego de tais medidas.

Por fim, ainda quanto ao eixo de análise subjetiva, deve-se considerar o fluxo de dados para fora da organização. Nesse caso, como terceiros deteriam "meios próprios" para reverter o processo de anonimização dos dados. Trata-se de uma questão particularmente importante no que diz respeito a eventuais parcerias que envolvam o uso compartilhado de dados[22], mesmo que não sejam dados pessoais *a priori*.

---

estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudoanonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas. […] § 4º Para os efeitos deste artigo, a pseudoanonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro". O tema foi uma das últimas inclusões na Lei, tendo sido inserido pela primeira vez em 24.05.2018, pelo relator, deputado Orlando Silva, no substitutivo 1 ao PL 4060/2012 apresentado à Câmara dos Deputados.

[18] Na pseudoanonimização, as informações adicionais que permitiriam a identificação do titular são mantidas em separado pelos agentes de tratamento, que podem, assim, reidentificar os dados se fizerem uso dessa informação. Contudo, caso excluam essas informações adicionais, os agentes não mais poderão efetuar a reidentificação "por meios próprios", caracterizando, assim, uma técnica de anonimização. É nesse sentido que a pseudoanonimização seria "o meio do caminho" para a anonimização.

[19] A LGPD estabelece a necessidade de que, sempre que possível, haja a anonimização dos dados utilizados em pesquisas (arts. 7º, IV, 11, II, "c", 13 e 16, II), assim como determina que, embora uma das exceções à eliminação dos dados após o término do tratamento seja o uso exclusivo do controlador, ela está condicionada à vedação do acesso aos dados por terceiro e à anonimização dos dados (art. 16, IV).

[20] Artigo 50 da LGPD. Art. 50: "Os controladores e operadores, no âmbito de suas competências, pelo tratamento de dados pessoais, individualmente ou por meio de associações, poderão formular regras de boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, incluindo reclamações e petições de titulares, as normas de segurança, os padrões técnicos, as obrigações específicas para os diversos envolvidos no tratamento, as ações educativas, os mecanismos internos de supervisão e de mitigação de riscos e outros aspectos relacionados ao tratamento de dados pessoais. […] § 2º Na aplicação dos princípios indicados nos incisos VII e VIII do caput do art. 6º desta Lei, o controlador, observados a estrutura, a escala e o volume de suas operações, bem como a sensibilidade dos dados tratados e a probabilidade e a gravidade dos danos para os titulares dos dados, poderá: I — implementar programa de governança em privacidade que, no mínimo: a) demonstre o comprometimento do controlador em adotar processos e políticas internas que assegurem o cumprimento, de forma abrangente, de normas e boas práticas relativas à proteção de dados pessoais".

[21] Como as informações adicionais que permitiriam a identificação do titular são mantidas separadamente e em posse dos agentes de tratamento, terceiros terão maior dificuldade em reverter a anonimização.

[22] Artigo 5º da LGPD. Art. 5º: "Para os fins desta Lei, considera-se: […] XVI — uso compartilhado de dados: comunicação, difusão, transferência internacional, interconexão de dados pessoais ou tratamento compartilhado de bancos de dados pessoais por órgãos e entidades públicos no cumprimento de suas competências legais, ou entre esses e entes privados, reciprocamente, com autorização específica, para uma ou mais modalidades de tratamento permitidas por esses entes públicos, ou entre entes privados".

Por exemplo, é muito comum que organizações se associem, mediante o compartilhamento e cruzamento de dados, para pesquisas científicas e outras atividades econômicas. Imagine o seguinte cenário:

a) uma pesquisa cujo objetivo é mensurar a eficácia de um determinado tratamento médico;
b) de antemão, reconhece-se ser necessário que a amostra de pessoas deve ser a mais ampla com objetivo de capturar pacientes com características distintas;
c) então, se faz necessária uma análise que envolva um conjunto de hospitais e clínicas médicas que trataram grupos de pacientes com diferentes perfis;
d) também se nota, desde logo, ser desnecessário o compartilhamento das bases de dados brutas (*raw data*), as quais identificariam diretamente cada um dos pacientes;
e) seria necessário apenas a indicação do perfil dos pacientes, os quais seriam agrupados de acordo com características semelhantes sem os tornar identificáveis *a priori*;
f) diversos testes de reidentificação foram executados, a fim de se assegurar e ser certificada a razoabilidade das técnicas de anonimização empregadas que correspondem ao estado atual da arte.

Apesar de a situação hipotética descrever um cenário no qual a pesquisa rodaria em cima de uma base de dados anonimizada (critério objetivo – item "e"), isso por si só não encerraria a discussão acerca dos riscos de reidentificação. Deve-se verificar, ainda, se algum hospital ou clínica participante poderia lançar mão de "meios próprios" capazes de reverter o processo de anonimização da base como um todo. Mais uma vez, entra em cena uma análise subjetiva que é focada na capacidade de um agente em específico. Pense, por exemplo, que um dos hospitais deteria uma alta capacidade de entropia de informação, em razão de: **a)** deter uma série de informações adicionais por conta da sua capilaridade no setor com atendimento a grande parte da população representada no estudo; **b)** possuir tecnologias de processamento de dados disruptivas, que superam os padrões praticados até então no setor.

Dessa forma, também é relevante observar a capacidade subjetiva de terceiros que ingressem no fluxo informacional de uma organização. Especialmente, quando se tem em vista atividades de enriquecimento de dados que envolvam agentes externos para viabilizar uma atividade de tratamento de dados.

Em síntese, o legislador brasileiro adotou uma estratégia normativa alinhada à premissa de que os dados anonimizados seriam sempre passíveis de reversão. Os dois eixos de análise acima descritos – objetivo e subjetivo – compõem uma matriz de risco[23] em torno de possíveis engenharias reversas de um processo de anonimização. A *resiliência* de tal processo é o que determinará se haverá algum tipo de intersecção entre dados anonimizados e dados pessoais, cujos elementos de análise são de ordem objetiva (razoabilidade) e subjetiva (meios próprios).

[23] Sobre a estratégia regulatória baseada no risco e, em particular, relacionada ao conceito de dado pessoal e dado anonimizado, veja-se: RUBINSTEIN, Ira; HARTZOG, Woodrow. Anonymization and risk. *91 Washington Law Review*, Washington, DF, v. 703, 2016; *NYU School of Law, Public Law Research Paper*, Nova York, n. 15-36. Disponível em: http://bit.ly/2TKMmo8. Acesso em: 20 jan. 2020.

3. Exemplificando alguns fatores de risco: os enigmáticos termos "no momento" e "ocasião" do tratamento

Ao invés de considerar anonimização como algo cujo resultado (*output*) é infalível, foca-se em uma abordagem que considera a aplicação sistemática de técnicas de anonimização com o objetivo de agregar consistência ao processo como um todo[24]. Por essa razão, a análise acerca de se um dado deve ser, de fato, considerado como anonimizado é eminentemente *circunstancial*. Os dois critérios de análise – objetivo e subjetivo – acima mencionados, ganharão vida somente a partir do contexto no qual está inserida uma atividade de tratamento de dados, sobre a qual se busca retirar, ao máximo, seus respectivos identificadores.

Aliás, não é por outra razão que a LGPD amarra o conceito de dado anonimizado e anonimização, respectivamente, à "ocasião" e ao "momento" no qual se dá uma atividade de tratamento de dados pessoais. Na medida em que a definição de atividade de tratamento de dados engloba nada mais do que 20 (vinte) ações,[25] tudo o que é feito com um dado, o processo de anonimização deve representar um conjunto de ações contínuo e logicamente ordenado que abrace toda a extensão do ciclo de vida de um dado – da coleta ao descarte.

A título de exemplo, listam-se ao menos 06 (seis) fatores de risco (RUBINSTEIN; HARTZOG, 2015) e algumas medidas de mitigação:

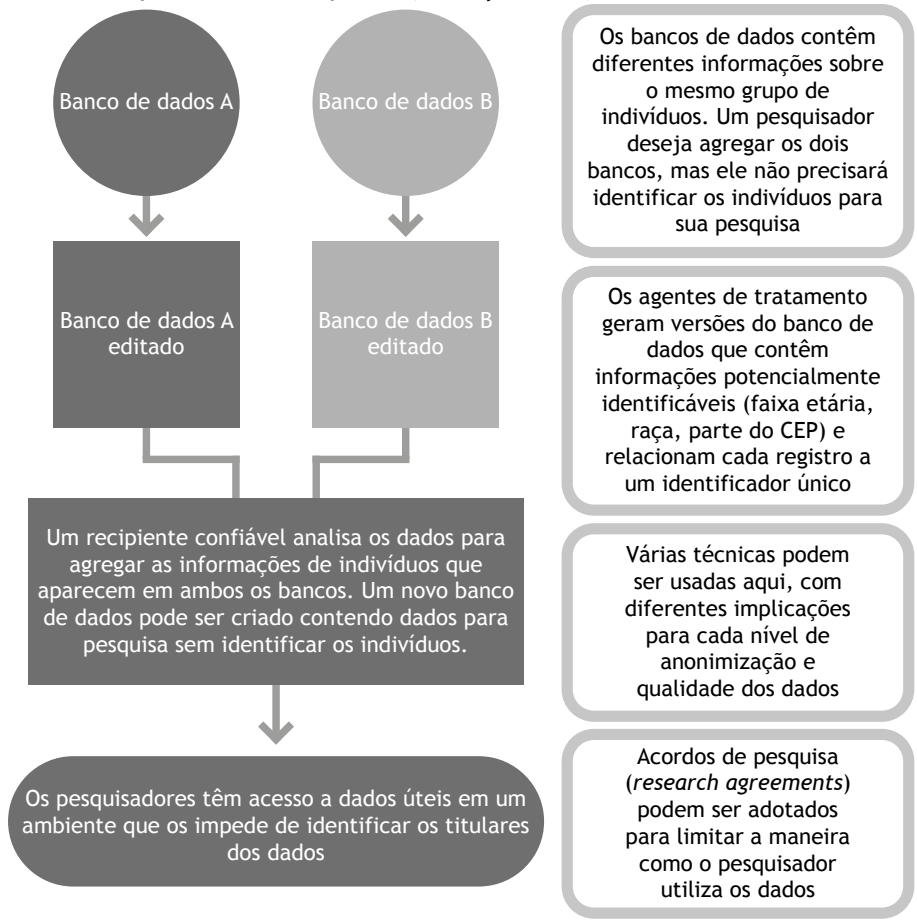
a) **Volume dos dados**: quanto maior for a quantidade de dados, maiores são as chances de alguém fazer o caminho inverso de um processo de anonimização. Desta forma, modelos de negócios, produtos ou serviços e, até mesmo, políticas públicas (incluindo de dados abertos) que envolvam grandes massas de dados devem proporcionalmente apresentar técnicas de anonimização correspondentes aos altos riscos de reidentificação em jogo;
b) **Natureza dos dados:** a natureza do dado (e.g., saúde, financeiro, geolocalização etc.) é determinante sobre o quão valiosas são eventuais informações que dele podem ser extraídas. Com isso, o apetite de terceiros e o quão recompensador seria reverter um processo de anonimização impulsiona os seus respectivos riscos de reidentificação;
c) **Cadeia da atividade de tratamento de dados (recipientes, compartilhamento e uso compartilhado)**: em muitas situações há uma complexa cadeia de atores para viabilizar um modelo de negócio ou mesmo uma política pública. Em regra, quanto maior for o ingresso de entidades para a geração ou mesmo o uso de uma base de dados anonimizada, mais elevado será o risco de sua reidentificação. Isto porque, não se aumenta apenas o volume do fluxo informacional (item "a"), como, também, a população que dele participa. Por exemplo, no caso acima mencionado relacionado à pesquisa científica, é comum se utilizar dos chamados "recipientes confiáveis". Esses são terceiros no qual organizações,

[24] Ao se considerar todo o ciclo de vida dos dados em sua divulgação, a análise (e preocupação) se desloca *do dado* – i.e. seus atributos, qualidades e riscos em determinado momento – para *o processo* – i.e. a realização de um conjunto de ações voltado à proteção da informação durante toda o seu processamento. RUBINSTEIN, Ira S. e HARTZOG, Woodrow. Anonymization and risk. *New York University Public Law and Legal Theory Working Papers 530*, 2015.

[25] Artigo 5º da LGPD: "Art. 5º Para os fins desta Lei, considera-se: […] X – tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração".

que desejam gerar uma nova base de dados (anonimizado) a partir dos seus bancos de informações, confiam a sua execução. Uma espécie de filtro com relação a quem deteria poder informacional para reverter o processo de anonimização. Nesse caso, o ingresso do terceiro no fluxo informacional se dá justamente para tornar mais resiliente o processo de anonimização.

Figura 1 — Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012. p. 42. (tradução livre)



d) **Gerenciamento de identidades e segmentação**: tão importante quanto colocar em prática processos de pseudoanonimização, é, também, controlar quem acessa as informações adicionais capazes de revertê-los. Por isso, é o caso de não só segregar fisicamente, mas, também, logicamente as bases de dados de uma organização (vide: exemplo supramencionado sobre uma rede varejista). Dessa forma, os riscos (internos) de reidentificação também passam a ser menores, na medida em que se reduz o número de atores que teria capacidade de juntar as peças do quebra-cabeça para formar a imagem dos titulares da informação. Nesse sentido, é importante destacar que o Decreto do Marco Civil da Internet (Decreto 8.771/2016) já determina a adoção de mecanismos

de gerenciamento de identidade a uma base de dados, inclusive com a previsão de sistemas de autenticação dupla e a individualização do respectivo usuário[26].

e) **Cláusulas contratuais**[27]: na medida em que fluxo informacional envolva cada vez mais agentes, em particular quando há o compartilhamento de dados para extração de informações, com ou sem os chamados "recipientes confiáveis", é cada vez mais comum cláusulas que: **a)** proíbam as partes de reverterem o processo de anonimização; **b)** delimitem o papel de cada um dos agentes de tratamento de dados de acordo com o objeto da atividade de tratamento de dados e, adicionalmente, vedando ou condicionando o repasse a terceiros que executariam tal atividade em nome de uma das partes; **c)** a destruição dos dados tão logo seja concluída a atividade de tratamento de dados ou caso haja a resolução de alguma condição pactuada;

f) **Atualização contínua**: anonimização é algo inacabado e fluido tal como é a própria definição da atividade de tratamento de dados, a qual procura capturar os dados em todos os seus movimentos. Ao expressamente correlacionar o conceito de dado anonimizado e anonimização ao "momento" e de acordo com a "ocasião" na qual um dado está sendo processado, a LGPD procurou deixar claro que as técnicas de anonimização devem considerar toda a jornada de um dado e, sobretudo, ser constantemente atualizadas. Por exemplo, não adiantará um contrato de processamento de dados, que especificou todas as técnicas de anonimização e inclusive a forma pela qual os "recipientes confiáveis" as colocariam em prática, se esse contrato foi firmado há bastante tempo e tais medidas já se encontram defasadas. O *continuum* de uma atividade de tratamento de dados, espelhado por nada mais do que 20 (vinte) ações diferentes, também deve nortear a *dinamicidade* com a qual se empregam técnicas de anonimização[28].

Com isso, o legislador convida os agentes de tratamento de dados a conceberem e aplicarem as melhores técnicas de anonimização de acordo com as particularidades das suas respectivas atividades. É uma empreitada multifacetada, de ordem técnica, organizacional e, inclusive, contratual com o objetivo de controlar os riscos associados à reidentificação de uma determinada atividade de tratamento de dados.

---

[26] Artigo 13 do Decreto 8.771/2016. Art. 13: "Os provedores de conexão e de aplicações devem, na guarda, armazenamento e tratamento de dados pessoais e comunicações privadas, observar as seguintes diretrizes sobre padrões de segurança: I – o estabelecimento de controle estrito sobre o acesso aos dados mediante a definição de responsabilidades das pessoas que terão possibilidade de acesso e de privilégios de acesso exclusivo para determinados usuários; II – a previsão de mecanismos de autenticação de acesso aos registros, usando, por exemplo, sistemas de autenticação dupla para assegurar a individualização do responsável pelo tratamento dos registros; III – a criação de inventário detalhado dos acessos aos registros de conexão e de acesso a aplicações, contendo o momento, a duração, a identidade do funcionário ou do responsável pelo acesso designado pela empresa e o arquivo acessado, inclusive para cumprimento do disposto no art. 11, § 3º, da Lei nº 12.965, de 2014 ; e IV – o uso de soluções de gestão dos registros por meio de técnicas que garantam a inviolabilidade dos dados, como encriptação ou medidas de proteção equivalentes".

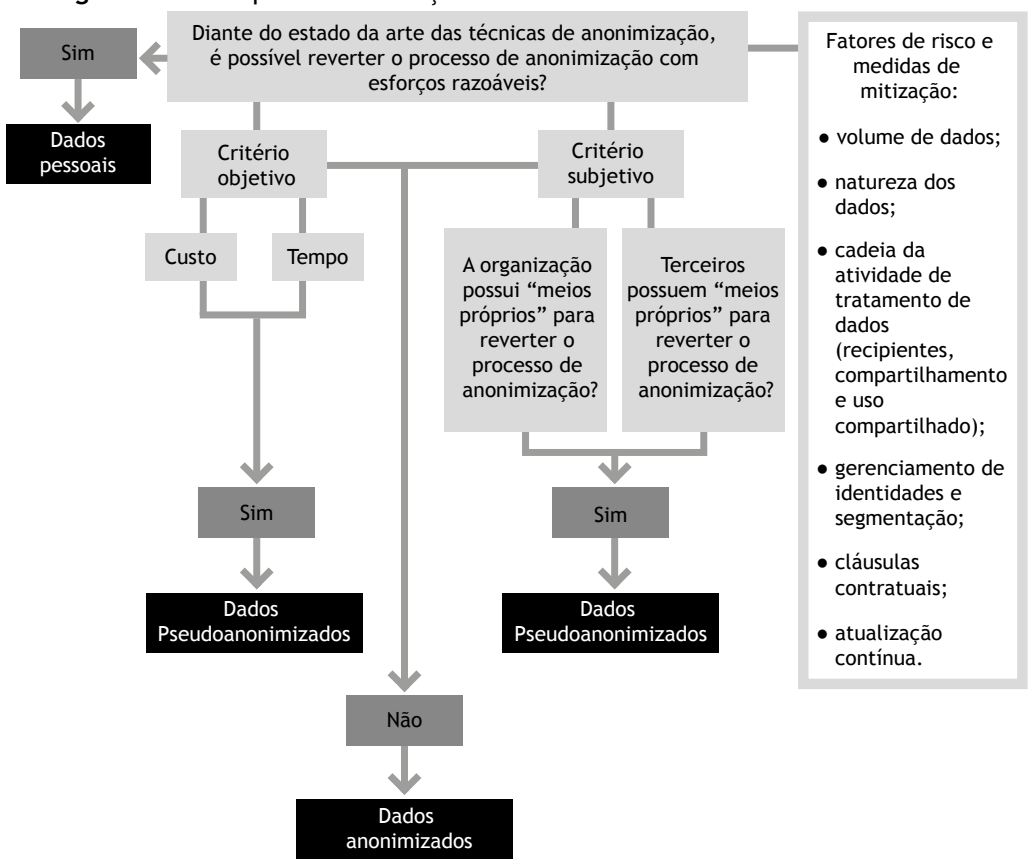
[27] FTC. *Protecting consumer privacy in an era of rapid change*: recommendations for businesses and policymakers, 2012; ROSENFELD, Dana B.; HUTNIK, Alysa Zeltzer. Data security contract clauses for service provider arrangements (pro-customer). *Practical Law Company*, 2011.

[28] Nesse sentido, uma das práticas previstas para se avaliar condutas pelo Modelo de Maturidade de Privacidade (*Privacy Maturity Model*), criado pelo Instituto Americano dos Contadores Públicos Certificados e pelo Instituto Canadense de Contadores (AICPA/CICA), é a otimização, i.e. "a revisão e a avaliação periódicas são utilizadas para garantir a melhoria contínua de determinado processo". Disponível em: http://bit.ly/2RA0FZP. Acesso em: 20 jan. 2020. A aplicação desse modelo de análise (e a conformidade especificamente a essa prática) foi observada no tratamento de dados pessoais efetuado pela municipalidade de Seattle. Ver: Future of Privacy. City of Seattle: Open data risk assessment, 2018. Disponível em: http://bit.ly/37cEuPP. Acesso em: 20 jan. 2020.

4. Conclusão: modelo analítico acerca do processo de anonimização de um dado

A análise acerca de se a natureza de um dado pessoal, submetido a um processo de anonimização, pode ser transmudada envolve uma série de elementos. O teste da Figura 2 agrupa logicamente os 07 (sete) critérios normativos prescritos pela própria LGPD e, ainda, lista, paralelamente, uma série de fatores, com base na literatura revisada, que ajudam na identificação do quão tolerável (razoável) são os riscos de reversão das técnicas de anonimização aplicadas.

Figura 2 — Entropia de Informação



5. Bibliografia

ARTICLE 29 WORKING PARTY. *Opinion 04/2007 on the concept of personal data*. Disponível em: http://bit.ly/2G99aFY. Acesso em: 20 jan. 2020.

BAPTISTA, Patrícia; KELLER, Clara. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. *Revista de Direito Administrativo*, Rio de Janeiro, v. 273, p. 123-163, 2016.

BIONI, Bruno Ricardo. *Xeque-mate*: o tripé de proteção de dados pessoais no xadrez das iniciativas legislativas no Brasil. São Paulo: GPoPAI-USP, 2016.

BIONI, Bruno Ricardo. *Proteção de dados pessoais*: a função e os limites do consentimento. Rio de Janeiro: Forense, 2019.

COUNCIL OF EUROPE. *Handbook on European data protection law*. Luxemburgo: Publications Office of the Europe Union, 2018. Disponível em: http://bit.ly/30OT26d. Acesso em: 20 jan. 2020.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.

HOUAISS, Antônio; VILLAR, Mauro de Salles. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

KOOPS, Bert-Jaap. Should ICT regulation be technology-neutral? *In*: KOOPS, Bert-Jaap; LIPS, Miriam; PRINS, Corien; SCHELLEKENS, Maurice (ed.). *Starting points for ICT regulation*: deconstructing prevalent policy one-liners. The Hague: TMC Asser Press 2006. v. 9, p. 77-108. (IT & Law Series). ISBN 90-6704-216-1.

MOSES, Lyria Bennett. How to think about law, regulation and technology: problems with 'technology' as a regulatory target. *Law, Innovation and Technology*, Abingdon-on-Thames, v. 5, n. 1, p. 1-20, 2013.

NARAYANAN, Arvind; SHMATIKOV, Vitaly. Myths and fallacies of "personally identifiable information". *Communications of the ACM*, Nova York, v. 53, n. 6, p. 24-26, 2010. Disponível em: http://bit.ly/30G9CVq. Acesso em: 20 jan. 2020.

REED, Chris. Taking sides on technology neutrality. *SCRIPT-ed*, Edimburgo, v. 4, n. 3, p. 263-284, 2007.

RUBINSTEIN, Ira S.; HARTZOG, Woodrow. Anonymization and risk. *New York University Public Law and Legal Theory Working Papers 530*, Nova York, 2015.

TEIXEIRA, Lucas. Teoricamente impossível: problemas com a anonimização de dados pessoais. Disponível em: http://bit.ly/367kuwQ. Acesso em: 20 jan. 2020.

TENE, Omer. Privacy law's midlife crisis: a critical assessment of the second wave of global privacy laws. *Ohio State Journal*, Columbus, v. 74, n. 6, p. 127-1262, 2013.