ARTIGO

Dados não pessoais: a retórica da anonimização no enfrentamento à covid-19 e o *privacywashing*

Sérgio Marcos Carvalho de Ávila Negri
sergio.negri@ufjf.br

Professor da Faculdade de Direito e do PPGD em Direito e Inovação da Universidade Federal de Juiz de Fora. Doutor em Direito Civil (UERJ).

Carolina Fiorini Ramos Giovanini
cfrgiovanini@gmail.com

Graduanda em Direito pela Universidade Federal de Juiz de Fora (UFJF).

V. 1 / N. 2 / DEZEMBRO DE 2020
PÁGINAS 126 A 149

SÉRGIO MARCOS C. DE ÁVILA N,
CAROLINA F. R. GIOVANINI

Dados não pessoais: a retórica da anonimização no enfrentamento à covid-19 e o *privacywashing*

Palavras-chave
anonimização
privacywashing
dados não pessoais
proteção de dados
COVID-19

Resumo
A Lei Geral de Proteção de Dados (Lei nº 13.709/18) define o dado pessoal como informação relacionada à pessoa natural identificada ou identificável. Apesar da promessa das técnicas de anonimização no sentido de desvincular pessoa e informação, o campo da Ciência da Computação vem demonstrando que estas técnicas não são livres de falhas, sendo impossível atingir o anonimato absoluto e irreversível, logo, dados anonimizados podem ser reidentificados. A partir do cenário de utilização de dados pessoais no enfrentamento à pandemia de COVID-19, o presente artigo, a partir de uma abordagem exploratória, procura demonstrar que a anonimização pode ser utilizada como estratégia discursiva para criar uma falsa imagem de proteção de dados (*privacywashing*). Pretende-se destacar a importância do conceito de *privacywashing* e de como a retórica da anonimização pode ser utilizada para neutralizar conflitos, atenuando o impacto de potenciais violações na utilização de dados pessoais.

V. 1 / N. 2 / DEZEMBRO DE 2020
PÁGINAS 126 A 149

SÉRGIO MARCOS C. DE ÁVILA N,
CAROLINA F. R. GIOVANINI

Non-personal data: the rhetoric of anonimization in the fight against covid-19 and the *privacywashing*

Keywords
anonymization
privacywashing
non-personal data
data protection
COVID-19

Abstract
The Brazilian General Data Protection Law defines personal data as information regarding an identified or identifiable natural person. Despite the promise of anonymization techniques to untie people and information, the field of Computer Science has been demonstrating that these techniques are not free from failures, being impossible to achieve absolute and irreversible anonymity, therefore, anonymized data can be reidentified. Based on the scenario of using personal data to deal with the COVID-19 pandemic, this article seeks to demonstrate that anonymization can be used as a discursive strategy to create a false image of data protection (*privacywashing*). It is intended to highlight the importance of the concept of *privacywashing* and how the rhetoric of anonymization can be used to neutralize conflicts, mitigating the impact of potential violations in the use of personal data.

1. Introdução

No final de 2019, foi descoberto um novo agente do Coronavírus – família de vírus que causam infecções respiratórias – após uma série de casos registrados na China. A proliferação da doença desencadeou uma pandemia e, consequentemente, uma grave crise de saúde. Na tentativa de controlar o surto da doença, muitas iniciativas utilizam dados para monitorar o aumento do número total de casos e a localização de pessoas supostamente contaminadas. A utilização de dados na formulação de ações de saúde pública em situações emergenciais não é um movimento que surgiu com a emergência causada pela COVID-19. Entre 2014 e 2016, durante a epidemia de Ebola na África Ocidental, os dados pessoais já eram, por exemplo, utilizados para monitorar a expansão da doença e para realizar projeções, auxiliando, desse modo, na formulação de políticas de combate[1].

O uso de tecnologia e de dados para o enfrentamento à pandemia de COVID-19 se faz presente na gestão da infraestrutura hospitalar, no uso de Inteligência Artificial para auxílio de diagnóstico, na identificação de surtos, bem como na atuação preditiva com projeções de disseminação. Nesse contexto, surgem modelos que utilizam dados de localização coletados a partir do GPS dos celulares, da triangulação de antenas de telefonia e de tecnologias Bluetooth Low Energy (BLE). O aumento do fluxo de informações suscita críticas e preocupações com o direito à privacidade e à proteção dos dados pessoais, notadamente relacionadas à possibilidade desses dados serem utilizados, posteriormente, em ações de controle e vigilância de cidadãos.

A proteção dos dados pessoais não é uma barreira às políticas públicas de saúde e contenção de doenças. Nesse sentido, muitas iniciativas que coletam e processam dados sob a justificativa de monitoramento apontam que seus modelos utilizam somente dados anonimizados, isto é, dados que não possuem associação com pessoa identificada ou identificável. O problema da narrativa da anonimização é que, por vezes, os aplicativos se limitam a informar que utilizam dados anonimizados, mas não há qualquer informação sobre qual técnica de anonimização foi utilizada e os seus potenciais riscos. Quando a anonimização é utilizada apenas como um instrumento retórico capaz de afastar a aplicação de um regime de tutela mais rigoroso e criar uma falsa imagem de proteção de dados, pode-se falar de uma conduta de *privacywashing*.

O presente artigo pretende analisar de que maneira as técnicas de anonimização podem ser utilizadas de forma a legitimar a coleta e o tratamento de dados pessoais, sendo responsáveis pela construção da imagem de modelo "*privacy friendly*". A abordagem exploratória é utilizada com o objetivo de oferecer uma visão geral do problema′. De acordo com Gil (2008), as pesquisas exploratórias têm como objetivo oferecer proximidade com o objeto de modo a torná-lo evidente e compreensível.

A adoção de uma abordagem exploratória se justifica em razão da falta de referências sobre o tema do *privacywashing*. A maioria das pesquisas exploratórias vale-se do levantamento bibliográfico. Como existe significativa literatura sobre as técnicas da anonimização, o presente artigo procura desenvolver o tema principal a partir da análise de trabalhos que apontaram os riscos na utilização dessas variadas técnicas, destacando, para tanto, o desenvolvimento da ciência da reidentificação. Com essa estratégia metodológica, procurar-se demonstrar que o caminho aberto pelos autores que demonstraram as falhas da anonimização pode auxiliar na densificação do conceito de *privacywashing*. Analisa-se também a utilização de dados pessoais no combate à pandemia de COVID-19, tendo em vista que as pessoas supostamente diagnosticadas se encontram, em sua maioria,

em situação de vulnerabilidade. Busca-se tornar o problema mais explícito para, ao final, construir hipóteses que podem ser desenvolvidas em trabalhos futuros.

As disposições da Lei Geral de Proteção de Dados (LGPD) e do Regulamento Geral de Proteção de Dados da União Europeia (RGPD) serão analisadas para se estabelecer a moldura normativa da anonimização e da pseudonimização. Esta análise objetiva investigar as classificações e conceitos delineados pelas normas e verificar se são suficientes para incorporar os desafios técnicos da anonimização.Por fim, investiga-se o *privacywashing* a partir do papel da anonimização na construção do falso marketing *privacy friendly*, tendo em vista a utilização de dados pessoais no enfrentamento à pandemia de COVID-19.

2. Limites da tutela jurídica: conceituação de dados pessoais.

De acordo com Stefano Rodotà (2008), o direito à privacidade não mais se estrutura em torno do eixo "pessoa-informação-segredo", no paradigma da *zero-relationship*, mas sim no eixo "pessoa-circulação-controle. Para Rodotà (2008), a privacidade se apresenta como noção fortemente dinâmica e em constante contato com as mudanças promovidas pelas tecnologias da informação. A privacidade, em uma dimensão informacional[2], é definida como o direito de manter controle sobre as próprias informações.

Diante do cenário de crescente manipulação de dados dá-se, cada vez mais, importância à tutela jurídica de dados pessoais. O Regulamento Geral de Proteção dos Dados Pessoais da União Europeia (RGPD) entrou em vigor em 25 de maio de 2018. No ordenamento jurídico brasileiro foi publicada a Lei Geral de Proteção de Dados (Lei n° 13.709/2018), abreviada como LGPD.

No que diz respeito à definição do termo "dados pessoais", o estudo da conceituação é de extrema importância, pois a partir do conceito adotado são definidos os limites da tutela jurídica. Bruno Bioni (2019) destaca que o vocabulário utilizado na conceituação pode ser responsável por retrair ou expandir a moldura normativa de uma lei de proteção de dados pessoais. Desse modo, existem dois modelos de conceituação: reducionista e expansionista.

O autor aponta que um modelo expansionista define os dados pessoais como informações relacionadas à pessoa identificável, indeterminada e com vínculo mediato, indireto, impreciso ou inexato. O modelo reducionista, por outro lado, trata de pessoa identificada, específica/determinada e vínculo imediato, direto, preciso ou exato. O artigo 5º, inciso I, da LGPD define que o dado pessoal é uma informação relacionada à pessoa natural identificada ou identificável, desse modo, adota o modelo expansionista (sem rol exemplificativo). Ressalta-se que, entre autores norte-americanos, utiliza-se, geralmente, o termo "informação pessoalmente identificável" – *personally identifiable information*, *PII* (Schawrtz & Solove, 2011).

No âmbito do direito comunitário europeu, o Regulamento Geral de Proteção de Dados [(Regulamento (UE) 2016/679] é aplicado ao tratamento de dados pessoais (Art. 2.º n.º 1) não só de uma pessoa singular identificada, mas também identificável, vale dizer, que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrônica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, econômica, cultural ou social dessa pessoa singular (Art. 4.º 1).

Na definição do RGPD incluem-se os quase-identificadores e os metadados, pois, conforme o item 30 da exposição de motivos, as pessoas singulares podem ser associadas a identificadores por via eletrônica tais como endereços IP (protocolo de internet), testemunhos de conexão (cookies) ou outros identificadores, como as etiquetas de identificação por radiofrequência. O Grupo de Trabalho do Art 29 (2007) enfatizou que as informações se referem a um indivíduo quando tratarem da identidade, das características, do comportamento de uma pessoa ou quando essas informações forem usadas para determinar ou influenciar a maneira como essa pessoa é tratada ou avaliada. Sendo assim, uma pessoa natural é considerada "identificada" dentro de um determinado grupo de pessoas porque se distingue de todos os outros membros do grupo.

O dado pessoal sensível é espécie de dado pessoal que possui tipologia distinta devido ao seu conteúdo. É definido pelo Art. 5º, inciso II, da LGPD como dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural. Observa-se que o dispositivo não estabelece um conceito, mas sim determinados tipos de dados que são sensíveis.

No sentido oposto, quando um dado não pode ter associação com pessoa identificada ou identificável de forma permanente e irreversível, desde a origem ou após tratamento, é denominado dado não pessoal. Fink e Pallas (2020) apontam que, atualmente, os dados não pessoais também possuem alto valor econômico, por isso, a delimitação do conceito não é somente um interesse teórico, pelo contrário, a conceituação dos dados não pessoais possui importância prática para qualquer processamento que venha a ser realizado.

Em 2018, a União Europeia publicou o Regulamento (EU) 2018/1807, que trata do fluxo livre de dados não pessoais, com o objetivo de garantir segurança jurídica para as empresas processarem dados em qualquer parte do território da União Europeia, incentivando a adoção de tecnologias em nuvem. O regulamento aplica-se ao tratamento de dados eletrônicos que não sejam dados pessoais (Art. 2º, nº 1).

O item 9 da exposição de motivos do regulamento de fluxo livre de dados não pessoais aponta que dados agregados e anônimos, usados para análises de big data, são exemplos específicos de dados não pessoais. Nesse ponto, cabe destacar também a distinção entre dados agregados e dados desagregados. Os dados agregados são usados frequentemente em trabalhos estatísticos, representando um indicador único. Os dados desagregados são aqueles que estão disponíveis sob a forma em que foram coletados, ou seja, não passaram por processo de agregação.

A orientação da Comissão Europeia (2019) - *Guidence on the Regulation on a framework for the free flow of non-personal data in the European Union* - sobre o regulamento fluxo livre de dados não pessoais classifica os dados não pessoais por origem: (i) dados que originalmente não estavam relacionados a uma pessoa natural identificada ou identificável e (ii) dados que inicialmente eram dados pessoais, mas que posteriormente foram anonimizados. O regulamento já é, contudo, objeto de críticas, como, por exemplo, a dificuldade de implementação prática da distinção entre dados pessoais e dados não pessoais.

Graef, Gellert e Husovec (2018) argumentam que os bancos de dados geralmente são mistos, isto é, contêm dados pessoais e não pessoais. Os autores entendem que as fronteiras dos dados pessoais são muito fluidas para atuar como âncora regulatória. A existência de dois regimes separados pode levar a um comportamento empresarial estratégico voltado para explorar esses conceitos, isto é, as corporações podem

passar a fundamentar suas políticas de privacidade nessa distinção para limitar ou eliminar as obrigações impostas pelo ordenamento. As técnicas de anonimização podem ser utilizadas discursivamente na exploração desta distinção, facilitando as práticas de *privacywashing*.

Vicenzo Zeno-Zencovich (2020) também critica esta distinção argumentando que a classificação em dados pessoais e dados não pessoais é irrealista na medida em que os dados descrevem situações complexas que possuem elos em comum, por exemplo, nos dados relativos à venda de um objeto que deve ser entregue, há o nome do comprador, seu endereço e seu cartão de crédito, mas também há dados referentes ao objeto vendido, a transportadora, a rota etc.

O processo de "anonimização" promove a retirada do vínculo da informação com a pessoa a qual se refere, porém, os dados não pessoais podem ser reidentificados se forem realizados esforços suficientes. Há também a pseudonimização, definida pelo Grupo de Trabalho do Artigo 29º (2014) como a substituição de um atributo, normalmente único, em um registro por outro atributo. Dessa forma, as informações não podem ser ligadas a um indivíduo específico sem que se recorra a informações suplementares, mantidas em bases externas separadas.

O Grupo de Trabalho do Artigo 29º (2014) também considera a utilização das técnicas criptográficas como forma de pseudonimização, pois a cifragem de informações faz com que os dados deixem de poder se conectar a um titular de dados específico sem que se recorra a informações suplementares, no caso, à chave de decriptação. Por isso, Doneda e Machado (2019) entendem que é mais adequado pensar o dado cifrado/criptografado como informação pessoal *prima facie*.

Recentemente, o Supremo Tribunal Federal iniciou o julgamento da ADPF 403 e da ADI 5.527. As duas ações constitucionais sobre o uso da criptografia de ponta-a-ponta[3], que é utilizada em aplicativos de mensagens para que apenas remetente e destinatário possuam as chaves para decodificar a informação[4]. Nesse caso, terceiros não conseguem ter acesso à informação trocada entre duas pessoas. As ações discutem a interpretação dos Artigos 10 e 12 do Marco Civil da Internet e, embora o julgamento ainda esteja em curso, já foi pontuado que a proibição de criptografia de ponta a ponta é inconstitucional, pois garante o sigilo das informações e o direito à privacidade dos cidadãos[5].

Fink e Pallas (2020) entendem que o teste jurídico para conceituar um dado como pessoal ou como não pessoal está estabelecido no item 26 da exposição de motivos do RGPD: para determinar se uma pessoa natural é identificável, deve se levar em consideração (i) a utilização de todos os meios razoavelmente prováveis de serem utilizados, (ii) fatores objetivos, como os custos e a quantidade de tempo necessária para a identificação, observando a tecnologia disponível no momento.

Em síntese, a anonimização dos conjuntos de dados tem sido a principal justificativa para legitimar o compartilhamento e processamento de dados porque seria uma técnica que protege os dados pessoais do usuário. Isso é observável, inclusive, nas legislações de proteção de dados, que consideram os dados anonimizados não mais como dados pessoais e, consequentemente, estes estão fora do escopo de tutela jurídica das leis de proteção de dados.

3. A anonimização e a pseudonimização na Lei Geral de Proteção de Dados (Lei 13.709/2018) e no Regulamento Geral de Proteção de Dados da União Europeia (RGPD)

A Lei Geral de Proteção de Dados utiliza a anonimização como uma referência técnica que garante a segurança de dados pessoais. O dado anonimizado é definido como aquele relativo a titular que não possa ser identificado, tendo em vista a utilização de "*meios técnicos razoáveis e disponíveis na ocasião de seu tratamento*" (Art. 5º, III). Seguindo essa linha, a técnica de anonimização é definida como a "*utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo*" (Art. 5º, XI).

Uma das bases legais para o tratamento de dados pessoais é a realização de estudos por órgão de pesquisa, "*garantida, sempre que possível, a anonimização dos dados pessoais*" (Art.º 7.º, IV). No caso de tratamento de dados sensíveis, este poderá ocorrer sem fornecimento de consentimento do titular quando for indispensável para estudos realizados por órgão de pesquisa, novamente, a lei estabelece que sempre que possível será garantida a anonimização (Art.º 11.º, II, c). A anonimização também surge na hipótese de utilização de dados pessoais para realização de estudos em saúde pública, neste caso, a lei também se refere à pseudonimização (Art.º 13.º, caput).

A técnica também justifica a conservação de dados após o término do seu tratamento para finalidades de estudo por órgão de pesquisa (Art.º 16.º, II) e para uso exclusivo do controlador (Art. 16, IV). Observa-se que a anonimização é utilizada como forma de legitimação, seja para autorizar um tratamento, seja para autorizar armazenamento mesmo após o término de um tratamento. Por fim, a LGPD também determina que fica excluída a portabilidade dos dados anonimizados (Art. 18, §7º).

A LGPD estabelece que a determinação da razoabilidade da técnica de anonimização adotada 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 (Art. 12, §1º), seguindo a linha do item 26 da exposição de motivos do RGPD. De acordo com Bruno Bioni (2020), o primeiro eixo de análise é o estado da arte da tecnologia, trata-se de fator objetivo que se decompõe em outros dois: custo e tempo.

É necessário avaliar o gasto de recursos financeiros e humanos para realizar a reversão de um processo de anonimização, considerando a tecnologia disponível no momento. Por isso, Doneda e Machado (2019) entendem que o critério depende de aspectos contextuais e, como consequência, a caracterização de um dado como dado pessoal torna-se um estado dinâmico. Os autores argumentam que a dinamicidade do conceito contribui para que a legislação de proteção de dados acompanhe as transformações tecnológicas e socioeconômicas, mas apontam que essa maleabilidade conceitual pode gerar insegurança, uma vez que surgem novas técnicas de reidentificação, cada vez mais eficientes.

Conforme Bioni (2020), o segundo eixo de análise do filtro da razoabilidade é subjetivo, trata-se de analisar quem é o agente de tratamento de dados e se ele possui "meios próprios" para realizar a reversão do processo de anonimização. É necessário analisar o fluxo de dados dentro da organização, por exemplo, quando o próprio agente tem informações adicionais, ainda que mantidas separadamente, para reverter uma pseudonomização.

O autor mostra que também é necessário considerar o fluxo de dados para fora da organização, analisando-se como terceiros possuem "meios próprios" para reverter a anonimização. Bioni (2020) entende que os critérios objetivos e subjetivos mencionados fazem parte de uma matriz de risco, ou seja, a resiliência do processo de anonimização será essencial para determinar uma possível intersecção entre dados anonimizados e dados pessoais.

Desse modo, para determinar se os dados são pessoais ou não pessoais, é necessário avaliar a possibilidade de identificação não só da perspectiva do agente de tratamento, mas também de qualquer outro terceiro. Fink e Pallas (2020) esclarecem que existem abordagens relativistas e absolutistas. A abordagem absoluta entende que pouco importa de qual perspectiva (controlador ou terceiro) os dados se qualifiquem como dados pessoais, uma vez que qualquer sujeito deve proteger os direitos do titular dos dados.

Por outro lado, os defensores da perspectiva relativa argumentam que a abordagem absoluta elimina a necessidade de qualquer gerenciamento de riscos por parte dos agentes de tratamento, que serão forçados a fazer suposições sobre os piores cenários possíveis, mesmo que estas hipóteses não sejam relevantes para aquele contexto específico. A adoção de uma abordagem absoluta poderia levar ao descarte da existência de dados anônimos, já que sempre haverá alguém com o interesse de identificar um conjunto de dados que foi anonimizado.

Fink e Pallas (2020) sinalizam para a existência de uma abordagem intermediária, como a formulada pelo *Information Commissioner's Office* – ICO, baseada na figura do "intruso motivado", que seria um terceiro "razoavelmente competente" e com acesso a recursos como Internet, bibliotecas, documentos públicos. Contudo, não é necessário que tenha conhecimento especializado, habilidades *hackers* ou acesso a equipamentos especiais.

Ao lado da anonimização, há a pseudonimização, que surge na LGPD somente como alternativa à anonimização, na hipótese de tratamento de dados para a realização de estudos em saúde pública por órgãos de pesquisa (Art. 13, *caput*). É definida como o "*tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro*" (Art. 13, §4º).

Por outro lado, a pseudonomização é fortemente sugerida pelo Regulamento Geral de Proteção de Dados (Regulamento (UE) 2016/679), inclusive, o item 27 da exposição de motivos pontua que esta técnica pode reduzir os riscos para os titulares de auxiliar os responsáveis pelo tratamento e os seus subcontratantes a cumprir as suas obrigações estabelecidas pelo regulamento. O item 29 cria incentivos para aplicar a pseudonomização e item 26 informa que dados pseudonimizados deverão ser considerados como informações sobre uma pessoa natural identificável.

O item 78 da exposição de motivos e o Art. 25º, Item 1, apresentam a pseudonimização como uma das medidas que respeitam os princípios da proteção de dados, em especial os princípios de *Privacy by Design*. Por outro lado, o item 85 da exposição de motivos ressalta que se as medidas não são adotadas de forma adequada e, no caso da pseudonomização, venha a ocorrer uma inversão não autorizada, essa violação pode vir a causar danos aos titulares, devendo ser notificada à autoridade.

No que diz respeito à segurança do tratamento de dados pessoais, a pseudonomização surge como medida técnica apta a assegurar um nível de segurança adequado ao risco (Art. 32, Item 1, c). O RGPD também sugere que esta técnica esteja presente em códigos de conduta elaborados pelos responsáveis pelo tratamento ou por subcontratantes (Art. 40º, Item 2, d). O Art, 4º, Item 5 do RGPD define a pseudonimização

como uma forma de tratamento na qual os dados pessoais deixem de poder ser atribuídos a uma pessoa natural específica sem recorrer a informações suplementares, desde que estas sejam mantidas separadamente e sujeitas a medidas técnicas e organizativas que impeçam a identificação.

Existem dois tipos de pseudonimização que permitem mascarar os dados de uma pessoa natural, com base nas chaves usadas: simétrica e assimétrica. Na pseudonimização simétrica, a mesma chave é usada para criptografar ou mascarar os dados e torná-los legíveis novamente, porém, há o problema de como compartilhar a chave sem que ela seja descoberta. Na pseudonimização assimétrica, duas chaves distintas são usadas: a primeira para criptografar os dados, a segunda para decifrá-los. Nesse caso, a chave é usada para criptografar é visível para qualquer pessoa, e a chave para decifrar somente o destinatário sabe, tornando desnecessário o compartilhamento.

As técnicas de anonimização e pseudonimização surgem nos diplomas normativos como medidas que justificam, legitimam ou asseguram determinados tipos de tratamento de dados pessoais, sendo consideradas mitigadoras de riscos.

4. Técnicas de anonimização e reidentificação

Em Massachusetts (Estados Unidos), uma agência governamental chamada *Group Insurance Commission* (GIC) contratou um seguro de saúde para os funcionários do Estado. Posteriormente, a GIC decidiu liberar, sem nenhum custo e para qualquer pesquisador que as solicitasse, os registros que compilavam as visitas hospitalares de todos os funcionários estaduais. A agência removeu os campos que continham identificadores como nome, endereço e número de previdência social, considerando que, dessa forma, protegeria a privacidade do paciente, apesar de ainda estarem incluídos o CEP, a data de nascimento e o sexo.

Quando a agência governamental divulgou os referidos dados, William Weld, então governador de Massachusetts, garantiu ao público que a GIC havia protegido a privacidade dos pacientes ao excluir identificadores. Em resposta, Latanya Sweeney (2000) começou a procurar os registros hospitalares do governador nos dados do GIC. Ela sabia que o governador Weld residia em Cambridge, Massachusetts, uma cidade de 54 mil habitantes e sete códigos postais. Por cerca de vinte dólares, ela comprou os boletins completos de eleitores da cidade de Cambridge – um banco de dados contendo, entre outras coisas, nome, endereço, CEP, data de nascimento e sexo de cada eleitor.

Sweeney combinou os dados eleitorais com os registros divulgados pela GIC e, então, conseguiu encontrar o Governador Weld com facilidade. Apenas seis pessoas em Cambridge compartilhavam sua data de nascimento; apenas três eram homens e, dos três, apenas ele morava em seu CEP. Em resposta à fala de William Weld, Sweeney enviou todos os registros de saúde do governador (incluindo diagnósticos e prescrições) a seu escritório.

Existem diversos casos que relatam situações semelhantes: a combinação das tabelas de um banco de dados marcado pelo "anonimato" e das tabelas de um banco de dados externos conecta linhas de uma às linhas da outra, combinando informações compartilhadas e respondendo à pergunta "qual indivíduo estes dados descrevem?".

Paul Ohm (2010) aponta que a razão para realizar a anonimização de dados pessoais é a proteção da privacidade dos titulares quando houver armazenamento ou divulgação destes dados. Os dados podem ser divulgados para terceiros,

para o público (como no caso da GIC) e para pessoas dentro da organização em que os dados foram coletados ou tratados.

Em um processo de anonimização, é necessário definir quais conjuntos de dados (atributos) serão anonimizados e qual técnica de anonimização será aplicada a cada um deles. Por isso, Camenisch *et al*. (2011, como citado em Silva, 2019, p.28) apresentam uma classificação de atributos a partir da sensibilidade da informação que aquele conjunto representa se for divulgado ou compartilhado: (i) atributos identificadores, que identificam os indivíduos (por exemplo, nome, CPF, RG); (ii) atributos semi-identificadores, que, se combinados com informações externas, expõem indivíduos ou aumentam a certeza sobre suas identidades (por exemplo, data de nascimento, CEP, cargo, tipo sanguíneo); e (iii) atributos sensíveis, que se referem a condições específicas dos indivíduos (por exemplo, salário, exames médicos).

Ainda no que diz respeito aos riscos, o Grupo de Trabalho do Artigo 29º (2014) aponta que é necessário observar três riscos que são fundamentais para a anonimização: (i) identificação: possibilidade de isolar alguns ou todos os registros que identifiquem uma pessoa num conjunto de dados; (ii) possibilidade de ligação: capacidade de ligar pelo menos dois registros sobre a mesma pessoa; e (iii) inferência: possibilidade de deduzir, com uma probabilidade significativa, o valor de um atributo a partir dos valores de um conjunto de outros atributos.

A anonimização de dados pode ser realizada a partir de diferentes técnicas, mas existem dois tipos gerais de abordagem: aleatorização e generalização. Este trabalho se concentra em apresentar brevemente o funcionamento de algumas dessas técnicas, como supressão, generalização, adição de ruídos, permutação, privacidade diferencial, k-anonimato e l-diversidade.

A aleatorização é uma família de técnicas que altera a veracidade dos dados a fim de eliminar a estreita ligação entre os dados e a pessoa, ou seja, os dados tornam-se imprecisos. São técnicas desse tipo de abordagem: adição de ruído, permutação e privacidade diferencial. A adição de ruídos modifica atributos no conjunto de dados de modo a torná-los menos precisos, enquanto se mantém a distribuição global. A permutação mistura aleatoriamente os valores dos atributos numa tabela, de modo a que alguns destes sejam ligados artificialmente a titulares de dados diferentes. Por fim, a privacidade diferencial ocorre quando o responsável pelo tratamento de dados gera visualizações anonimizadas de um conjunto de dados, conservando uma cópia dos dados originais.

A generalização, segunda abordagem possível, consiste em generalizar ou diluir, os atributos dos titulares dos dados através da alteração da escala ou ordem de grandeza (isto é, uma região em vez de uma cidade, um mês em vez de uma semana). São técnicas desse tipo de abordagem: agregação, k-anonimato, l-diversidade e t-proximidade. Paul Ohm (2010) diferencia a generalização da supressão, na medida em que esta é a remoção completa de identificadores, enquanto aquela é a alteração dos valores de um identificador. Por exemplo, imagine um banco de dados que contém o nome completo, o CPF e o CEP de indivíduos, nesse caso, o CPF poderia ser suprimido e o nome e o CEP poderiam ser generalizados, constando apenas o prenome e os três primeiros dígitos do CEP.

Em um conjunto de dados k-anonimizado, cada registro é indistinguível de pelo menos $(k-1)$ outros registros com relação a certos atributos de "identificação". Narayanan e Shmatikov (2010) esclarecem que, para aplicar o k-anonimato, conjunto dos atributos de *quase-identificador* deve ser previamente fixado e considerado o mesmo para todos os usuários (geralmente, CEP, data de nascimento, sexo etc.) e o restante dos atributos é considerado não identificador.

Machanavajjhala, Kifer, Gehrke e Venkitasubramaniam (2007) mostraram que um

invasor pode descobrir os valores de atributos sensíveis quando há pouca diversidade nesses atributos, além disso, os invasores geralmente possuem "conhecimento de segundo plano", que ao ser combinado com os dados anonimizados permite a reidentificação. Narayanan e Shmatikov (2010) também apontam que o k-anonimato falha completamente em conjuntos de dados de alta dimensão.

A técnica l-diversidade busca garantir que em cada classe de equivalência determinado atributo tenha, pelo menos, l (um) valor diferentes. Conforme o parecer do Article 29 Data Protection Working Party, o objetivo é limitar a ocorrência de classes de equivalência com fraca variabilidade do atributo, para que o intruso que tenha conhecimentos de base sobre um titular de dados específico permaneça sempre com um grau significativo de incerteza.

A t-proximidade surge a partir da l-diversidade. O objetivo é a criação de classes equivalentes semelhantes à distribuição inicial de atributos na tabela, para isso, insere-se uma nova restrição: além da existência de, ao menos, l valor diferentes em cada classe, também é necessário que cada valor seja representado quantas vezes forem necessárias para refletir a distribuição inicial de cada atributo.

O parecer do Grupo de Trabalho do Artigo 29º (2014) concluiu que nenhuma das técnicas apresentadas está imune aos três riscos que foram apresentados inicialmente (identificação, possibilidade de ligação e inferência). É, nesse sentido, que Paul Ohm (2010) afirma que os dados podem ser úteis ou perfeitamente anônimos, mas nunca os dois.

A anonimização é vista como uma forma de proteção dos dados pessoais, mas a ciência da reidentificação rompe com essa visão, ampliando os danos à privacidade, pois cada reidentificação bem-sucedida, mesmo que revele dados aparentemente sem sentido, possibilita uma reidentificação futura. O uso de termos como "anonimização", "dado anônimo" ou "dado anonimizado" tem sido criticado, pois cria uma ilusão de que aqueles dados seriam, permanentemente, dados não pessoais. É nesse processo que se insere a importância do conceito de *privacywashing*.

A pessoa que promove a reidentificação dos dados é denominada, geralmente, "adversário". Quando um adversário encontra uma base de dados anonimizada é possível que ele vincule esses dados à uma base de dados externa, chamada de informações auxiliares, e desse modo, poderá extrair as identidades. A combinação das duas bases de dados, de modo a conectar linhas de uma tabela às linhas da outra é denominada "junção interna" (*Inner Joins*).

Atualmente, novos bancos de dados são criados diariamente, sendo impossível prever o tipo e a quantidade de informações externas que o adversário pode acessar, por isso, Paul Ohm (2010) argumenta que assumir que o adversário não conseguirá encontrar um banco de dados externo com a parte específica de dados necessária para desbloquear dados anonimizados é uma ingenuidade.

Ohm (2010) apresenta ainda o "Mito do Superusuário", isto é, o entendimento de que poucas pessoas seriam capazes de combinar os bancos de dados, logo o adversário seria uma espécie rara de usuário. Porém, estudos como os de Narayanan e Shmatikov[6] têm demonstrado que há certa facilidade na reidentificação, embora o usuário médio de computador não possa realizar uma junção interna, a maioria das pessoas que fez um curso em gerenciamento de banco de dados ou trabalhou em TI provavelmente pode replicar as técnicas de reidentificação com um software amplamente disponível, como o Microsoft Excel.

Há, ainda, uma relação desproporcional entre utilidade e privacidade, pois apesar de estarem vinculadas, à medida que a utilidade dos dados aumenta um pouco, a privacidade diminui. Para que os dados sejam úteis, também devem ser imperfeitamente anônimos.

Isso ocorre porque a forma mais usual de impedir ataques de ligação é a alteração de atributos comuns (como data de nascimento, CEP e sexo) e outros quase identificadores. Porém, os quase-identificadores transmitem informações relevantes que podem ser úteis para análises posteriores.

Brickell e Shmatikov (2008) compararam um banco de dados completamente limpo, contendo apenas o campo único de informações em estudo (por exemplo, em um estudo sobre trabalho este campo seria o valor dos salários), com várias técnicas de anonimização amplamente utilizadas. Os pesquisadores demonstraram que a relação utilidade-privacidade é oposta, ou seja, pequenos aumentos na utilidade são comparados a reduções maiores na privacidade, e pequenos aumentos na privacidade causam grandes reduções na utilidade.

A validade das reidentificações de dados anonimizados é, por vezes, contestada por empresas e por pesquisadores de controle estatístico de divulgação sob o argumento de que como os conjuntos de dados estão sempre incompletos, nunca se poderia ter certeza de ter reidentificado a pessoa certa, mesmo que uma correspondência tenha sido encontrada.

Contudo, Rocher *et al*. (2019) mostram que a probabilidade de um indivíduo específico ter sido reidentificado corretamente pode ser estimada com alta precisão, mesmo quando o conjunto de dados anonimizado está muito incompleto. Os autores propuseram um modelo gráfico generativo que pode ser treinado com precisão e eficiência em dados incompletos. Utilizando esse modelo, os pesquisadores descobriram que 99,98% dos americanos seriam identificados novamente em qualquer conjunto de dados usando 15 atributos demográficos.

Rubinstein e Hartzof (2015) questionam o porquê de os debates sobre as falhas na anonimização não estarem refletidos nas leis de proteção de dados. Para os autores, um dos motivos é que, frequentemente, o debate se concentra em casos famosos nos quais um pesquisador desenvolve e aplica um método para reidentificar indivíduos em um conjunto de dados não identificados ou demonstra a viabilidade de um ataque publicando uma prova de conceito e, posteriormente, a mídia transforma esses resultados de pesquisa em anedotas que comprovam o fracasso do anonimato. Rubinstein e Hartzof consideram (2015) que o problema dessa narrativa é o foco em um único método (desidentificação) à custa de outros métodos.

Silva (2019) apresenta uma abordagem, baseada em anonimização, para plataformas de análise de dado, tal abordagem divide-se em duas etapas. A primeira é menos restritiva e deve ser aplicada durante o processo de Extração, Transformação e Carga (ETL) e a segunda, mais restritiva, está presente antes da divulgação dos dados para usuários externos. O autor mostra que a abordagem tem baixo impacto nos resultados de desempenho e utilidade na plataforma de análise de dados, superando a relação oposta entre utilidade dos dados e privacidade dos indivíduos.

Contudo, Paul Ohm (2010) aponta que não devemos apostar em avanços técnicos que supostamente serão responsáveis por resolver todos os problemas e riscos da anonimização. O autor entende que não há como substituir a necessidade de uma resposta regulatória baseada em uma avaliação mais realista dos riscos de reidentificação. Narayanan e Shmatikov (2010) demonstraram a partir de reidentificações que os riscos das técnicas de anonimização não podem ser considerados como meramente hipotéticos.

Independentemente da tecnologia utilizada, a governança de dados não pessoais é indispensável, uma vez que até as técnicas de anonimização mais desenvolvidas são passíveis de falhas. Contudo, mesmo diante dos riscos da anonimização, a simbologia dos dados anonimizados continua sendo adotada, isso pode ser

observado nas diversas iniciativas que coletam dados sob a justificativa de combate à pandemia de COVID-19.

5. Anonimização e *privacywashing* no enfrentamento à pandemia de COVID-*19*

No final de 2019, a China registrou diversos casos envolvendo um novo agente da família de vírus denominada Coronavírus, o SARS-CoV-2. Desde então, a doença atingiu uma série de países, resultando em uma pandemia e dando início a uma emergência global de saúde. Buscando frear o desenvolvimento da doença, muitas iniciativas públicas e privadas passaram a utilizar dados visando o monitoramento da localização de novos casos, bem como o aumento do número de infectados.

No enfrentamento da doença, iniciativas públicas e privadas passam a coletar e tratar dados de localização, que podem ser obtidos a partir de diferentes tecnologias, como a utilização de GPS dos celulares, triangulação de antenas de telefonia e tecnologias Bluetooth Low Energy (BLE). Apple e Google anunciaram uma parceria para desenvolvimento de uma tecnologia que usa BLE[7], na qual indivíduo que aceita participar passa a ter um código de identificação e uma chave de rastreamento únicos. O código de identificação é uma sequência de números aleatórios que muda a cada período de 10 a 20 minutos. Este código é transmitido pelo celular de um indivíduo, por meio do Bluetooth, e poderá ser recebido por qualquer outro aparelho que esteja próximo.

Os dados de GPS dos celulares são coletados a partir do uso de diferentes aplicativos provedores de serviços, como aplicativos de transporte, mapas e rotas, trânsito etc. Já a triangulação entre antenas de telefonia móvel monitora as conexões que os aparelhos celulares fazem com as diferentes antenas que existem em uma localidade, sinalizando uma movimentação.

No que tange ao uso de dados de localização, Goldsmith e Wu (2006) apontam que a segmentação geográfica na Internet oferece uma maneira econômica de combinar informações e indivíduos específicos. Desse modo, embora a rede tenha componentes extraterritoriais, a localização geográfica dos indivíduos é importante, principalmente para que empresas consigam desenvolver estratégias de marketing individual refinado.

Para que esses modelos recebam confiança e aceitação social, sendo de fato utilizados pela maior parte dos cidadãos, as iniciativas precisam fornecer garantias de que há proteção à privacidade e, geralmente, informam que utilizam dados anonimizados. As rotas de mobilidade de um indivíduo podem ser vulneráveis a tentativas de reidentificação, pois são altamente correlacionadas. O trajeto que uma pessoa faz diariamente, os lugares que frequenta e a ordem em que essa pessoa esteve em cada local são informações únicas que podem identificar um indivíduo.

A Coreia do Sul, no combate à pandemia, utilizou um sistema de rastreamento dos contaminados pela COVID-19 baseado em dados de GPS, câmeras, cartão de crédito e entrevistas com pacientes diagnosticados e indivíduos que estiveram em contato com estes. A partir destes dados, era possível reconstruir o trajeto do indivíduo contaminado pelo vírus, monitorando o avanço da doença e enviando alertas por SMS para pessoas que passaram pelos mesmos locais onde esteve a pessoa diagnosticadas. Apesar de nenhum nome ou endereço ser divulgado, algumas pessoas conseguiram conectar os locais e as características do indivíduo diagnosticado e identificá-lo, por exemplo, as pessoas de uma determinada localidade chegaram a concluir, por meio das notificações, que

duas pessoas estavam tendo uma relação extraconjugal durante a quarentena[8].

Além do monitoramento individual da circulação do indivíduo, algumas iniciativas passaram a monitorar também a rede de pessoas que entram em contato com um indivíduo que testou positivo para o vírus, é o chamado *contact tracing* (rastreamento de contatos). A utilização de *contact tracing* não está necessariamente relacionada ao uso de tecnologias, pois pode ser feita a partir de entrevistas presenciais ou por ligação telefônica. Também não se trata de uma estratégia que surgiu com a pandemia de COVID-19. A diferença deste contexto é o surgimento de aplicativos nos quais a pessoa contaminada insere a informação de que possui o vírus e o serviço envia notificações para quem esteve em contato com ela.

O documento de Diretrizes 04/2020 sobre o uso de dados de localização e ferramentas de *contact tracing* no contexto da pandemia de COVID-19, elaboradas pelo *European Data Protection Board* (EDPB), estabeleceu que quando houver uso de dados de localização, a preferência deve ser o processamento de dados anonimizados. Apesar da anonimização ser vista como uma forma de proteção da privacidade, foi demonstrado que a reidentificação dos dados tornados anônimos não é meramente hipotética.

As orientações do EDPB ressaltam que dados de localização considerados anônimos podem, na verdade, não serem anônimos. O EDPB, apesar de sugerir a anonimização, aponta que um único padrão de rastreamento dos dados de localização de uma mesma pessoa durante significativo período de tempo não pode ser totalmente anonimizado. O mesmo vale para precisões de coordenadas geográficas que não são suficientemente reduzidas.

A Secretaria de Saúde do Estado de Santa Catarina divulga, periodicamente, um conjunto de dados com os casos confirmados de COVID-19 no âmbito do estado, informando que o conjunto é estruturado de forma anonimizada[9]. O SIMI - SP (Sistema de Monitoramento Inteligente de São Paulo), viabilizado por meio de acordo entre as operadoras de telefonia Vivo, Claro, Oi e TIM e Governo de São Paulo também afirma utilizar dados anonimizados "*sem desrespeitar a privacidade de cada usuário*", destacando que "*os dados de georreferenciamento servem para aprimorar as medidas de isolamento social para enfrentamento ao coronavírus*"[10]. A InLoco, empresa que vende serviços baseados em dados de localização, esclareceu que "*para fins da colaboração com o controle da pandemia, a In Loco está utilizando as mais avançadas técnicas de anonimização*"[11], assim como uma série de iniciativas que informaram estar utilizando somente dados anonimizados.

As iniciativas que coletam e processam dados sob a justificativa de monitoramento do avanço da pandemia realizam a coleta destes dados por meio de aplicativos para celular, afirmando que utilizam técnicas de anonimização. A proteção dos dados pessoais não é um obstáculo à formulação de políticas públicas. Contudo, quando aplicativos e serviços apenas informam que utilizam dados anonimizados, simplesmente reproduzem trechos da lei e não fornecem qualquer informação acerca da técnica de anonimização utilizada e seus possíveis riscos, observa-se o uso discursivo da anonimização.

Os riscos da anonimização não são explicados ao titular dos dados pessoais, que se encontra em posição vulnerável durante a pandemia de COVID-19. O não esclarecimento acerca da possibilidade técnica de reidentificação gera no usuário uma falsa sensação de que seus dados estão seguros. A inexistência de agentes externos que demonstrem as possíveis falhas da técnica adotada e trabalhem na identificação dos riscos envolvidos também contribuem para a retórica da anonimização.

A ausência de fiscalização sobre as práticas descritas acima pode abrir espaço para a prática de *privacywashing*. *Privacy* é o termo em

inglês para privacidade e *whitewash* é uma tinta branca de baixo custo, aplicada em fachadas. O *privacywashing* é expressão que caracteriza o falso marketing social de empresas que desejam construir uma imagem de responsabilidade com a privacidade do usuário, mas que não realizam ações concretas para garantir a proteção de dados.

A maquiagem da privacidade se faz presente em termos de uso que meramente citam dispositivos da LGPD ou tão-somente apontam que os dados são anonimizados. Essas iniciativas se apresentam para a sociedade como "*privacy friendly*", isto é, constroem uma imagem de que o produto é desenvolvido de modo a respeitar a privacidade do usuário, reforçando uma sensação de segurança que é, na verdade, falsa.

No Chile, um jornal divulgou o banco de dados georreferenciados do Ministério da Saúde na forma de mapas[12]. De acordo com o jornal, cada ponto que indica um caso testado positivo para o coronavírus foi movido entre 50 e 100 metros de seu local de origem. Apesar de a adição de ruído ser uma técnica de anonimização, era possível visualizar blocos, prédios e casas específicas nas quais havia casos confirmados. Consequentemente, também era possível identificar os indivíduos a partir da associação de seus endereços com os pontos apresentados nos mapas.

A relação de interdependência entre a atividade corporativa e a opinião pública, proporciona, por um lado, ganhos em maior eficiência comercial e possibilidade de expansão de mercado das corporações. Por outro lado, faz com que as empresas estejam submetidas à cobrança, pressão e fiscalização do público. Esse panorama faz com que as empresas se voltem para a gestão do risco social, forjando uma imagem de credibilidade de modo a neutralizar conflitos, atenuando o impacto de potenciais violações.

Transpondo o cenário exposto acima para o contexto da sociedade de informação, marcada pela extração de dados pessoais e pela de monetização destes dados, que passam a integrar transações e surgem como objeto de um novo mercado, verifica-se que as empresas de tecnologia se beneficiam das novas formas de interação online para expandir suas áreas de atuação e auferir lucro. Contudo, são estas mesmas formas de interação online (novas mídias sociais) que promovem um debate sobre a proteção dos dados pessoais e o direito à privacidade, colocando as empresas em uma encruzilhada: a opinião pública é responsável, ao mesmo tempo, por garantir ganhos comerciais e por gerar grandes crises de reputação institucional causadas por violações ao direito à privacidade (vazamentos, manipulação e venda de dados, por exemplo).

O *privacywashing* também pode ser observado, em certa medida, na atuação de governos durante a pandemia. Por exemplo, no Brasil, a Prefeitura da cidade de Miranda, no Mato Grosso do Sul, divulgou um boletim epidemiológico acerca do avanço da doença que, apesar de suprimir o nome e demais identificações diretas, segregava os dados de indígenas e não-indígenas, dando notícia de que o primeiro caso havia sido confirmado no último dia 16 de julho na comunidade indígena da Aldeia Moreira[13]. Posteriormente, foram registrados ataques e ameaças contra a população indígena, acusada de deflagrar a disseminação do coronavírus na região[14].

O debate sobre a proteção de dados pessoais é, cada vez mais, presente na sociedade. Para que as pessoas passem a entender que estes dados são uma projeção de sua própria *persona*, conscientizando-se de que possuem o direito de manter controle sobre as informações que lhe digam respeito, é fundamental também a fiscalização de uma dimensão simbólica, revelando práticas existentes que contrariam uma imagem socialmente construída. Com o objetivo de enfrentar a opinião pública que levanta a questão da privacidade e da proteção de

dados pessoais, as empresas direcionam esforços para a construção de uma imagem "*privacy friendly*". Por vezes, a retórica da anonimização é utilizada para construção de uma reputação institucional que indica ao público que aquela empresa se preocupa com a privacidade do usuário e direciona esforços para que seus dados pessoas estejam protegidos, gerando uma falsa sensação de segurança.

Lawrence Lessig (2006) aponta que os benefícios gerados por iniciativas de monitoramento podem ser alcançados enquanto se protege a privacidade. É claro que o planejamento exige mais tempo para criar codificações e rotinas essenciais à governança dos dados. Arquitetar as proteções de privacidade desde o início é, contudo, menos custoso do que atualizá-las posteriormente.

As iniciativas, criadas pelo Poder Público e por empresas, que informam utilizar dados anonimizados precisam esclarecer quais técnicas de anonimização estão sendo utilizadas. Trata-se de medida de transparência necessária para o exercício de controle social. A anonimização não deve ser apresentada ao usuário como técnica totalmente efetiva, pelo contrário, o titular dos dados pessoais precisa ser informado dos riscos de ser identificado mesmo em um conjunto de dados anonimizado.

Esse cenário se agrava ainda mais com a ausência da Autoridade Nacional de Proteção de Dados (ANPD), órgão que já foi criado, mas ainda está sendo estruturado e organizado pela Administração Pública Federal[15]. A LGPD estabelece que a ANPD poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações de segurança destes (Art. 12, §3º). A presença de um órgão implementado de forma independente e autônoma é essencial para complementar a autorregulação que já vem sendo feita.

Na Europa, o processo de utilização de dados para o combate à pandemia tem sido mediado pela atuação das agências ou autoridades de proteção de dados, o que, por sua vez, confere maior legitimidade às iniciativas. Para as agências europeias, o enfrentamento a pandemia não depende da suspensão ou modificação dos marcos institucionais de proteção de dados. A própria legislação já apresentaria regras para as situações emergências, as quais devem ser observadas. Ao mesmo tempo, somente poderiam ser tratados os dados essenciais para a finalidade previamente determinada, sem que ocorra a divulgação ao público de dados como identidade, idade e residência. Há também uma preocupação com a transparência das medidas e a segurança dos dados, que somente devem ser tratados por pessoas autorizadas e previamente indicadas.

Ainda que se admita que o rastreamento de dados de geolocalização possa contribuir com o combate a pandemia quando associado a outras medidas, principalmente a testagem em massa, as autoridades europeias de proteção de dados têm trabalhado com a ideia de que a utilização de tecnologias nos celulares para o rastreamento de contaminados somente poderia ser admitida em presença do consentimento livre e informado dos interessados.

Além de outras garantias, nota-se a preocupação com limites temporais, isto é, com o fim da emergência os dados deverão ser apagados, permitindo-se apenas a utilização de dados agregados, não identificados, para fins de pesquisa. A princípio, a legislação de proteção de dados pessoais não se aplica aos dados que foram devidamente anonimizados. Como procuramos demonstrar, isso não significa que não existam riscos nesse tipo de utilização. Esse tipo de iniciativa revela uma preocupação com as rotinas e procedimentos, isto é, com a governança dos dados coletados para o combate à pandemia.

Nesse processo, é imprescindível fiscalizar também a utilização discursiva dessas garantias, voltadas apenas para criar uma falsa reputação que não se confirma na prática das rotinas

efetivamente implementadas. A retórica da anonimização deve ser observada com atenção porque constrói a imagem de modelo de negócio "*privacy friendly*". A simples referência aos dispositivos da lei, sem uma concreta implementação da cultura de transparência e de proteção da privacidade também acarretam a falsa sensação de segurança no usuário. Morley, Cowls, Taddeo e Floridi (2020) apontam a necessidade de se justificar eticamente o desenvolvimento de sistemas de monitoramento. Mesmo diante de uma crise global, é necessário considerar riscos éticos e implementar medidas adequadas para impedir que dados coletados sob a justificativa de contenção da pandemia sejam utilizados de forma potencialmente antiética, impactando a privacidade dos cidadãos de modo irreversível.

Não se trata de liberar indiscriminadamente a utilização dos dados, nem de criar um espaço em que os direitos sejam completamente anulados. Na verdade, busca-se estabelecer parâmetros, rotinas e regras para que o uso de dados se mostre legítimo mesmo nessas circunstâncias excepcionais. Como condição para a aceitação social das novas medidas, é sempre importante pensar em garantias e modelos mais transparentes, capazes de gerar maior confiança nas medidas de monitoramento que se mostrem necessárias. A utilização de dados pessoais para o combate à pandemia não representa uma panaceia. Deve sempre ser vista como uma medida extrema, que somente pode ser utilizada quando for baseada em garantias jurídicas, e desde que associada a outras medidas sanitárias amparadas cientificamente.

A utilização de tecnologias digitais deve ser conciliada com a proteção dos dados pessoais dos cidadãos e com a garantia de direitos e liberdades civis. O cenário de enfrentamento à pandemia não pode ser uma justificativa para coleta e processamento abusivos. Não há como construir uma cultura de proteção de dados sem combater discursos que se apropriam indevidamente de um imaginário que não se traduz em rotinas e práticas efetivamente transparentes. Nesse contexto, a governança dos dados não pessoais, quando associada a uma fiscalização efetiva, assume extrema importância, pois cria parâmetros técnicos, jurídicos e éticos que devem ser seguidos para evitar danos irreversíveis.

6. Conclusão

É fato que as relações entre indivíduos evoluíram e se tornaram cada vez mais complexas e dinâmicas juntamente com o desenvolvimento tecnológico, que gerou um grande fluxo de informações. Na sociedade de informação, a proteção de dados se afastou do próprio discurso abstrato da privacidade.As técnicas de anonimização surgem como medidas de proteção à privacidade, de modo a legitimar a coleta e processamento de uma série de dados pessoais. A anonimização não é uma técnica livre de falhas, pelo contrário, traz em si riscos técnicos, jurídicos e éticos que precisam ser esclarecidos ao usuário.

A partir do cenário de utilização de dados pessoais no enfrentamento à pandemia de COVID-19, procurou-se demonstrar que a "anonimização" pode ser utilizada de forma retórica, dissociada de práticas e rotinas transparentes de proteção de dados, criando uma falsa sensação de segurança. Frequentemente, nota-se a referência à utilização de dados anomimizados sem a identificação da técnica utilizada e dos riscos que lhe são inerentes. A mera referência aos dispositivos da lei, sem ações concretas de proteção à privacidade também auxiliam a construção da imagem "privacy friendly".

A construção de uma cultura de proteção de dados não pode ignorar as dimensões simbólicas e as relações de poder já existentes. O titular dos dados precisa ser informado acerca da

técnica utilizada e dos reais esforços despendidos na garantia da proteção de seus dados. Não se trata de banir o uso de técnicas de anonimização, mas de se evitar a sua utilização dissociada de rotinas e procedimentos transparentes, capazes de aferir os riscos, que deverão ser informados devidamente ao cidadão, sempre deixando claro que não existem modelos livres de falhas e suficientes para a completa proteção da privacidade.

Referências

Abreu, J. S. Passado, presente e futuro da criptografia forte: desenvolvimento tecnológico e regulação. Revista Brasileira de Políticas Públicas, Brasília, v. 7, n. 3, p. 25-43, dez. 2017. Disponível em: https://www.publicacoesacademicas.uniceub.br/RBPP/article/view/4869. Acesso em: 6 out. 2020.

Article 29 Data Protection Working Party. Opinion 4/2007 on the concept o personal data. Bruxelas: [s. n.], 2007. Disponível em: <http://ec.europa. eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_ en.pdf>. Acesso em: 30/05/2020.

Article 29 Data Protection Working Party. Opinion 5/2014 on Anonymisation techniques. Bruxelas: [s. n.], 2014. Disponível em: <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 30/05/2020.

BBC News. Coronavirus privacy: Are South Korea's alerts too revealing?. 30 de maio de 2020. Disponível em: https://www.bbc.com/news/world-asia-51733145

Bioni, B. R. Proteção de Dados pessoais: a função e os limites do consentimento. Rio de Janeiro. Forense, 2019.

Bioni, B. R. Compreendendo o conceito de anonimização e dado anonimizado. Cadernos Jurídicos, São Paulo, v. 19, n. 53, p. 191-202, jan./mar. 2020. Bimestral.

Brickell, J., & Shmatikov, V. The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, Nevada, Estados Unidos. DOI: 10.1145/1401890.1401904.

Doneda, D. Da privacidade à proteção de dados pessoais: elementos da lei geral de proteção de dados. 2. ed. São Paulo: Thomson Reuters Brasil, 2019. 352 p.

Doneda, D., & MENDES, L. S. Data protection in Brazil: new developments and current challenges. In: GUTWIRTH, Serge; LEENES, Ronald; HERT, Paul De. (Eds.) Reloading data protection: multidisciplinar insights and contemporary challenges. London: Springer, 2014, p. 15.

European Data Protection Board. Guidelines 04/2020 on the use of location data and contact tracing tools in the contexto of the COVID-19 outbreak. Bruxelas: [s.n.], 2020. Disponível em: https://edpb.europa.eu/our-work-tools/our-documents/usmernenia/guidelines-042020-use-location-data-and-contact-tracing_en. Acesso em: 30 de maio de 2020.

European Commission. Communication from the Commission to the European Parliament and the Council: Guidence on the Regulation on a framework for the free flow of non-personal data in the European Union. Bruxelas: [s.n.], 2019. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:250:FIN. Acesso em: 30 de maio de 2020.

Finck, M., & Pallas, F. , F. They who must not be identified: distinguishing personal from non-personal data under the GDPR. International Data Privacy Law, Oxford, v. 10, n. 1, p. 11-36, 10 mar. 2020. Disponível em: https://doi.org/10.1093/idpl/ipz026. Acesso em: 30 maio 2020.

Goldsmith, J. L.; Wu, T. Who controls the Internet? illusions of a borderless world. New York: Oxford University Press, 2006.

Graef, I.; Gellert, R.; Husovec, M. Towards a Holistic Regulatory Approach for the European Data Economy: why the illusive notion of non-personal data is counterproductive to data innovation. Tilec Discussion Paper, Tilburgo, v. 29, p. 1-18, 28 set. 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3256189##. Acesso em: 30 maio 2020.

Gil, A. C. Como elaborar projetos de pesquisa. 4. ed. São Paulo: Atlas, 2008,p. 41.

Lei nº 13.709, de 14 de agosto de 2018. (2018). Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília. Acesso em 30 de maio de 2020, disponível em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm

Lessig, Lawrence. Code and other laws of cyberspace: version 2.0. New York: Basic Books, 2006.

Machado, D., & Doneda, D. Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados. Revista dos Tribunais. vol. 998. Caderno Especial. p. 99-128. São Paulo: Ed. RT, dezembro 2018.

Machanavajjhala, A. et al. L-diversity: privacy beyond k-anonymity. Acm Transactions On Knowledge Discovery From Data, [s.l.], v. 1, n. 1, p. 3, mar. 2007. Association for Computing Machinery (ACM). http://dx.doi.org/10.1145/1217299.1217302.

Mischitelli, L. Tecniche di pseudonimizzazione e anonimizzazione: differenze e campi di applicazione Cybersecurity 360. 30 de maio de 2020. Disponível em: https://www.cybersecurity360.it/legal/privacy-dati-personali/tecniche-dipseudonimizzazione-e-anonimizzazione-differenze-e-campi-di-applicazione/

Morley, J.; Cowls, J.; Taddeo, M.; Floridi L. Ethical Guidelines for SARS-CoV-2 Digital Tracking and Tracing Systems. SSRN Electronic Journal, 2020. Disponível em: http://dx.doi.org/10.2139/ssrn.3582550.

Narayanan, A., & Shmatikov, V. Robust De-anonymization of Large Sparse Datasets. EEE Symposium on Security and Privacy. Oakland, Califórinia, Estados Unidos. DOI: 10.1109/SP.2008.33.

Narayanan, A., & Shmatikov, V. Myths and Fallacies of "Personally Identifiable Information". Ċommunications of the ACM, v. 53, n. 06, p. 24, Junho 2010. Disponível em: <www.cs.utexas.edu/~shmat/shmat_cacm10.pdf>.

Negri, S. M. C. De A., & Barbosa, L. V., 2018. "Green is the new black": o greenwashing e o controle empresarial do risco social. In: Seminários Internacionais de Direitos Humanos e Empresas. Juiz de Fora: pp.8-10. Disponível em: http://homacdhe.com/v-seminar/wpcontent/uploads/sites/8/2018/11/Anais.pdf. Acesso em: 30 de maio de 2020.

Negri, S. M. C., & Fernandes, E. R. Democracia e responsabilidade ambiental na mineração: uma relação conflituosa? In: GOMES, Ana Suelen Tossige; MATOS, Andityas Soares de Moura Costa (coord.). O Estado De Exceção Entre A Vida E O Direito. Belo Horizonte: Initia Via, 2019. p. 176-191.

Ohm, Paul. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Review, n. 57, p. 1701-1777, 2010.

Regulation (EU) 2016/679. (2016). Acesso em 30 de maio de 2020, disponível em https://eur-lex.europa.eu/eli/reg/2016/679/oj

Regulation (UE) 2018/1807. (2018). Acesso em 30 de maio de 2020, disponível em https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32018R1807

Rocher, L. et al. Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications 10, 3069 (2019). https://doi.org/10.1038/s41467-019-10933-3

Rodotà, S. A vida na sociedade da vigilância: a privacidade hoje. Organização, seleção e apresentação de Maria Celina Bodin de Moraes. Tradução de Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

Rubinstein, I. S., & Hartzog, W. Anonymization and risk. New York University Public Law and Legal Theory Working Papers 530, Nova York, 2015.

Schwartz, P. M., & Solove, D. J. The PII problem: privacy and a new concept of personally identifiable information. New York University Law Review, v. 86, p. 1827, dec. 2011

Silva, H. de O. Uma Abordagem Baseada em Anonimização para Privacidade de Dados em Plataformas Analíticas. 2019. 117 f. Dissertação (Mestrado) - Curso de Sistemas de Informação e Comunicação, Faculdade de Tecnologia, Universidade Estadual de Campinas, Limeira, 2019.

Sweeney, L. Uniqueness of Simple Demographics in the U.S. [s.n.] 2000. Population. Laboratory for Int'l Data Privacy, Working Paper LIDAP-WP4.

Sweeney, L. Achieving k-anonymity privacy protection using generalization and suppression. International J. of Uncertainty, Fuzziness and Knowledge-based Systems, 10(5), 2002.

Zeno-Zencovich, V. Free-Flow of Data. Is International Trade Law the Appropriate Answer?. In FABBRINI, F., CELESTE, E., QUINN, J. (eds.), Data Protection Imperialism and Data Sovereignty, Hart Pub. 2020 (Forthcoming). Available at SSRN: https://ssrn.com/abstract=3562986 or http://dx.doi.org/10.2139/ssrn.3562986

Notas finais

1 É possível citar, como exemplo, o HealthMap, que foi um banco de dados conduzido pela Organização das Nações Unidas (ONU), em conjunto com o Google e o governo dos EUA para criar relatórios e projeções sobre o surto de Ebola. Disponível em: https://www.healthmap.org/ebola/.

2 Destaca-se que Marcel Leonardi (2011) categoriza os diversos conceitos unitários de privacidade já produzidos pela doutrina e jurisprudência: (i) o direito a ser deixado só (the right to be let alone); (ii) o resguardo contra interferências alheias; (iii) segredo ou sigilo; (iv) controle sobre informações e dados pessoais. LEONARDI, Marcelo. Tutela e Privacidade na Internet. São Paulo: Saraiva, 2011, p.52.

3 As técnicas de criptografia buscam cifrar informações, ou seja, tornar confidencial o conteúdo das comunicações. Jacqueline de Souza Abreu aponta que, além de assegurar a confidencialidade de conteúdo de mensagens, dados e informações, a criptografia também garante a integridade (contra alterações do conteúdo) e autenticidade (das partes). ABREU, Jacqueline de Souza. Passado, presente e futuro da criptografia forte: desenvolvimento tecnológico e regulação. Revista Brasileira de Políticas Públicas, Brasília, v. 7, n. 3, p. 25-43, dez. 2017. Disponível em: https://www.publicacoesacademicas.uniceub.br/RBPP/article/view/4869. Acesso em: 6 out. 2020, p.27.

4 As informações são cifradas de modo que "apenas o destinatário da comunicação ou o detentor de chave criptográfica (simétrica ou assimétrica) pode acessar e compreender seu conteúdo informacional (plaintext)". MACHADO, D., & DONEDA, D. Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados. Revista dos Tribunais. vol. 998. Caderno Especial. p. 99-128. São Paulo: Ed. RT, dezembro 2018, p.101.

5 O julgamento ainda não foi concluído e, até o momento em que o presente artigo é escrito, está suspenso devido ao pedido de vista do ministro Alexandre de Moraes. Votaram os relatores das referidas ações: ministra Rosa Weber (ADI 5527), e ministro Edson Fachin (ADPF 403), que entendem que o sigilo das comunicações, inclusive pela internet, é uma garantia constitucional.

6 NARAYANAN, A., & SHMATIKOV, V. Robust De-anonymization of Large Sparse Datasets. EEE Symposium on Security and Privacy. Oakland, Califórinia, Estados Unidos. DOI: 10.1109/SP.2008.33. NARAYANAN, A., & SHMATIKOV, V. Myths and Fallacies of "Personally Identifiable Information". Ċommunications of the ACM, v. 53, n. 06, p. 24, Junho 2010. Disponível em: <www.cs.utexas.edu/~shmat/shmat_cacm10.pdf>.

7 APPLE. Apple e Google formam parceria para tecnologia de rastreamento de contato com COVID-19. 10 de abril de 2020. Disponível em: https://www.apple.com/br/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/#:~:text=Arquivo-,Apple%20e%20Google%20formam%20parceria%20para%20tecnologia%20de-%20rastreamento%20de,sociedade%20de%20volta%20%C3%A0%20normalidade.

8 BBC NEWS. Coronavirus privacy: Are South Korea's alerts too revealing?. 30 de maio de 2020. Disponível em: https://www.bbc.com/news/world-asia-51733145.

9 Informação disponível em: http://dados.sc.gov.br/dataset/covid-19-dados-anonimizados-de-casos-confirmados. Acesso em: 30/05/2020.

10 Informação disponível em: https://www.saopaulo.sp.gov.br/spnoticias/isolamento-social-em-sao-paulo-e-de-47-aponta-sistema-de-monitoramento-inteligente-10/. Acesso em: 30/05/2020.

11 Informação disponível em: https://content.inloco.com.br/hubfs/Estudos%20-%20Conte%C3%BAdo/Coronavirus/Meios%20de%20controle%20a%CC%80%20pandemia%20da%20COVID-19%20e%20a%20inviolabilidade%20da%20privacidade.pdf?hsCtaTracking=ad1577ba-e5bc-4ff3-afdd-54a896891088%7C07ab4d6b-53d3-4a06-9f43-fb43621df88f. Acesso em: 30/05/2020.

12 HERRERO, Victor A. Exclusivo: Estos son los mapas de contagio de Covid-19 que Mañalich mantiene en secreto. Interferencia. 2020 Mai 11. Disponível em: https://interferencia.cl/articulos/exclusivo-estos-son-los-mapas-de-contagio-de-covid-19-que-manalich-mantiene-en-secreto. Acesso em 30 jul 2020.

13 PREFEITURA MUNICIPAL DE MIRANDA. Miranda contabiliza seis casos positivos para a Covid-19. 2020 jul. 21. Disponível em: https://miranda.ms.gov.br/miranda-contabiliza-seis-casos-positivos-para-covid-19/.

14 MARQUES, H. Ameaças contra indígenas por 'levarem coronavírus' geram tensão em aldeias de MS. Jornal Midiamax 2020 jul. 24. Disponível em: https://www.midiamax.com.br/cotidiano/2020/gravacoes-com-acusacoes-e-ameacas-contra-indigenas-por-levar-coronavirus-a-cidade-de-ms-geram-tensao-em-aldeias.

15 Art. 55-A da Lei 13.709/18: "Fica criada, sem aumento de despesa, a Autoridade Nacional de Proteção de Dados (ANPD), órgão da administração pública federal, integrante da Presidência da República".