doi: 10.21783/rei.v7i1.476

SUBMETIDO: 07 MAR. 2020
ACEITO: 23 ABR. 2021

PROTEÇÃO DE DADOS E ANONIMIZAÇÃO: PERSPECTIVAS À LUZ DA LEI Nº 13.709/2018

DATA PROTECTION AND ANONYMIZATION: PERSPECTIVES IN LIGHT OF LAW NO. 13.709/2018

JOSÉ LUIZ DE MOURA FALEIROS JÚNIOR [1]
GUILHERME MAGALHÃES MARTINS [2]

RESUMO: A anonimização de dados é tema de peculiar relevância na Lei Geral de Proteção de Dados Pessoais brasileira (Lei nº 13.709/2018), uma vez que afasta a incidência normativa e deixa de acarretar todas as consequências legalmente previstas para a tutela de situações que envolvam dados pessoais. O que se problematiza neste breve estudo, contudo, é exatamente a nebulosidade quanto aos processos de anonimização existentes e à insegurança jurídica causada por sua incompleta e pouco clara compreensão. A hipótese de pesquisa parte da "entropia" como filtro adequado à delimitação de processos de anonimização, reversíveis ou não, e para viabilizar a aplicabilidade dos institutos de proteção contidos na lei de regência do tema aos referidos processos. A pesquisa utilizará o método de abordagem dedutivo, partindo de um conceito amplo de dados pessoais para um conceito mais específico de dados anonimizados para, então, ponderar os impactos da reversão e a adequação da hipótese de pesquisa para a solução da insegurança jurídica problematizada. Além disso, o trabalho promoverá a análise bibliográfico-doutrinária para, logo em seguida, apresentar as considerações finais, das quais se procurará extrair uma compreensão mais assertiva quanto à problemática explicitada.

PALAVRAS-CHAVE: Anonimização; dados pessoais; entropia; Lei Geral de Proteção de Dados Pessoais.

[1] Mestre em Direito pela Universidade Federal de Uberlândia. Especialista em Direito Processual Civil, Direito Digital e *Compliance*. Graduado em Direito pela Universidade Federal de Uberlândia. Advogado.

[2] Professor associado de Direito Civil da Faculdade Nacional de Direito da UFRJ. Professor permanente do Doutorado em Direito, Instituições e Negócios da UFF. Doutor e Mestre em Direito Civil, e Bacharel pela Faculdade de Direito da Universidade do Estado do Rio de Janeiro - UERJ. Pós-doutorando em Direito Comercial pela Faculdade de Direito da USP. Foi professor visitante do Mestrado e Doutorado em Direito e da Graduação em Direito da UERJ (2009-2010). É Membro Honorário do Instituto dos Advogados Brasileiros - IAB NACIONAL, junto à Comissão de Direito do Consumidor. Leciona Direito Civil, Direito do Consumidor e temas ligados ao Direito Digital e aos novos direitos. Diretor do BRASILCON, Diretor institucional do IBERC e associado do IBDFAM. Promotor de Justiça – MPRJ.

ABSTRACT: Data anonymization is a topic of particular relevance in the General Law for the Protection of Personal Data in Brazil (Law No. 13.709/2018), as it averts the normative incidence and hinders the legal consequences for the protection of situations involving personal data in light of the aforementioned framework. What is problematized in this brief study, however, is exactly the nebulosity regarding the existing anonymization processes and the legal insecurity caused by their incomplete and unclear understanding. The research hypothesis claims that "entropy" might be a suitable filter for the delimitation of anonymization processes, reversible or not, and to enable the applicability of the protective legal institutes governing the subject to the referred processes. The research will use the deductive approach method, starting from a broad concept of personal data to a more specific concept of anonymized data to then consider the impacts of the reversal procedures and the adequacy of the research hypothesis in order to solve the problematized legal insecurity. In addition, the work will promote a bibliographic-doctrinal analysis to, soon after, present the final considerations, from which a more assertive understanding regarding the explicit problem will attempted be extracted.

KEYWORDS: Anonymization; personal data; entropy; General Personal Data Protection Law.

INTRODUÇÃO

A sociedade da informação chega a seu ápice em um ambiente no qual a coleta e o tratamento massivo de dados – que dão origem ao chamado *Big Data* – passam a revelar um movimento irrefreável, com impactos políticos, econômicos e sociais. Cabe ao Direito o dever de regulamentar tais situações para ressignificar os impactos dos novos meios de comunicação sobre os domínios sociais.

Sendo inegavelmente crescente a utilização de dados pessoais para as mais variadas atividades, como identificação, classificação, autorização e diversas outras, torna-se imperiosa a implementação de marcos regulatórios para a adequada proteção de direitos, para a promoção da pessoa humana e para o fortalecimento e a renovação da unidade do ordenamento jurídico.

Nesse contexto, a promulgação da Lei Geral de Proteção de Dados Pessoais (LGPD) brasileira – Lei nº 13.709, de 14 de agosto de 2018 – representou inegável avanço, mas seu longo período de *vacatio legis*, além de sinalizar a complexidade de adaptação a seus rigores, incitou revisões críticas que já culminaram em alterações legislativas, tendo sido inicialmente editada a Medida Provisória nº 869, de 27 de dezembro de 2018, que reformulou densamente o texto original da norma, e que foi consolidada, após diversas audiências públicas e intensos debates, pela Lei nº 13.853, de 08 de julho de 2019, que manteve alguns dos ajustes realizados, efetivou outros e recompôs o texto original em certos pontos.

Um dos temas mais efervescentes da lei é a anonimização de dados, conceituada no artigo 5º, III e XI, e definida como direito do titular de dados nos artigos 11, II, b, 16, II e IV, e 18, IV, mas essencialmente peculiar por afastar a incidência da LGPD em razão de expressa disposição contida no artigo 12, §§1º a 3º.

Para além da preocupação com a proteção de dados em contextos econômicos, sociais e políticos, tem-se que ter em conta uma preocupação com a proteção da pessoa, haja vista a inerência de tais informações à própria configuração da personalidade e de seus atributos qualitativos, cuja identificação e eventual difusão configuram inegável ruptura à privacidade.

Não se olvida da facilidade de acesso, coleta e armazenamento de dados pessoais em praticamente qualquer contexto, tamanho o impacto que a Internet acarreta na difusão dos mais variados conteúdos e, como não poderia deixar de ser, a utilização de tais dados para inúmeros objetivos torna as informações pessoais elementos essenciais para a realização e para a exploração de atividades que envolvem risco, especialmente quando voltadas ao mercado de consumo e propulsionadas por predições algorítmicas.

O problema desta pesquisa parte dessa questão. Isso porque o intuito de retirar dos dados pessoais coletados quaisquer elementos que permitam identificar e, evidentemente, expor a pessoa à qual dizem respeito é a força motriz da anonimização, e alguns processos técnicos prometem viabilizar o uso desses dados sem risco de que se descubra os dados pessoais originais. Ocorre que a confiabilidade de tais mecanismos não podem ser atestada de forma cabal no atual estado da arte das estruturas tecnológicas da ciência da computação, o que causa reflexos no que diz respeito aos modais de tutela jurídica adequados ao tema. E, nesse sentido, a chamada "entropia" surge como uma perspectiva conceitualmente adequada à compreensão dos dispositivos da LGPD que versam sobre o assunto.

Assim, pelo método dedutivo, a pesquisa explorará o conceito de dado pessoal – mais amplo – para que se possa chegar a uma delimitação mais direta do que se entende por dado anonimizado. E, no contexto normativo inaugurado pela legislação brasileira, os dispositivos que cuidam do tema serão explorados e interpretados à luz da hipótese apresentada no afã de se estabelecer uma abordagem coerente e hígida ao problema, frente à enorme gama de mecanismos de proteção contemplados pelo ordenamento.

2. A PRIVACIDADE E OS DADOS ANONIMIZADOS

Samuel Warren e Louis Brandeis foram pioneiros ao analisar o direito à privacidade, em famoso estudo publicado em 1890.[3] Fato é que a evolução dessa

[3] Os autores, já no final do século XIX, descreveram o seguinte: "*The general object in view is to protect the privacy of private life, and to whatever degree and in whatever connection a man's life has ceased to be private, before the publication under consideration has been made, to that extent the protection is to be withdrawn. Since, then, the propriety of publishing the very same facts may depend wholly upon the person concerning whom they are published, no fixed formula can be used to prohibit obnoxious publications. Any rule of liability adopted must have in it an elasticity which shall take account of the varying circumstances of*

proteção sempre foi desafiada pela compreensão do controle exercido pela técnica sob diversos vieses, a demandar verdadeira atuação estatal voltada à regulação dos usos da Internet no alvorecer da sociedade da informação. É por essa constatação que propostas regulatórias como a da criação de uma *Lex Informatica*, tal como conceituada por Joel Reidenberg (1998, p. 582), ou de um *Code*, como sugerido por Lawrence Lessig (2006, p. 60), ganharam contornos mais fortes na última década do século XX. (CORREIA, 2018, p. 52-56)

As próprias perspectivas vislumbradas com o advento da Internet passaram a justificar a preocupação com o controle da técnica[4] e com o favorecimento de determinados fatores de predição de resultados na tomada de decisões lastreada em vasto repertório informacional. (WIENER, 2004, p. 485). Em verdade, uma nova economia (baseada em dados) se formou e passou a ditar os caminhos do capitalismo informacional feroz do século XXI (MAYER-SCHÖNBERGER; RAMGE, 2018 p. 7), alavancado, ainda, por fenômenos como a inclusão digital e a Internet das Coisas.[5]

As relações jurídicas passaram a ocorrer no ciberespaço, que, segundo Lorenzetti (2004, p. 31), "[p]roduz um efeito de 'desterritorialização' e 'descentralização', porque não há uma relação centro-periferia, não conhece ordens e hierarquias e, sobretudo, não há uma autoridade central." Este é o cenário perfeito para a consideração instrumental da imperiosa proteção à privacidade, que "não poderia estar limitada por esta, ao mesmo tempo em que faz referências a todo leque de garantias fundamentais que se encontram no ordenamento brasileiro". (DONEDA, 2006, p. 358-359)

Segundo Yuval Harari (2018, p. 83):

each case, - a necessity which unfortunately renders such a doctrine not only more difficult of application, but also to a certain extent uncertain in its operation and easily rendered abortive. Besides, it is only the more flagrant breaches of decency and propriety that could in practice be reached, and it is not perhaps desirable even to attempt to repress everything which the nicest taste and keenest sense of the respect due to private life would condemn." (WARREN; BRANDEIS, 1890, p. 215-216)

[4] Diz: "*The Internet as it originally was gave everyone a "Ring of Gyges", the ring which, as Plato reports in The Republic, made Gyges the shepherd invisible. The dilemma for regulation in such a world is precisely the fear Plato had about this ring: With such a ring, "no man can be imagined to be of such an iron nature that he would stand fast in justice." (...) But these gaps in the Internet's original design are not necessary. We can imagine networks that interact seamlessly with the Internet but which don't have these "imperfections." (...) Commerce acting alone has not yet eliminated these threats, to both commerce and civil life. For reasons I explore later in this book, it's not even clear commerce could. But commerce is not the only actor here. Government is also an important ally, and the framework of regulability that commerce has built could be built on again by government. Government can, in other words, help commerce and help itself. How it does so is the subject of the chapter that follows.*" (LESSIG, p. 59-60)

[5] Samuel Greengard sintetiza a preocupação que passou a permear a sociedade da informação do novo milênio: "*Within this emerging IoT framework, a dizzying array of issues, questions, and challenges arise. One of the biggest questions revolves around living in a world where almost everything is monitored, recorded, and analyzed. While this has huge privacy implications, it also influences politics, social structures, and laws.*" (GREENGARD, 2015, p. 58)

> Quando a autoridade passa de humanos para algoritmos, não podemos mais ver o mundo como o campo de ação de indivíduos autônomos esforçando-se por fazer as escolhas certas. Em vez disso, vamos perceber o universo inteiro como um fluxo de dados, considerar organismos pouco mais que algoritmos bioquímicos e acreditar que a vocação cósmica da humanidade é criar um sistema universal de processamento de dados – e depois fundir-se a ele. Já estamos nos tornando, hoje em dia, minúsculos *chips* dentro de um gigantesco sistema de processamento de dados que ninguém compreende a fundo. Todo dia eu absorvo incontáveis *bits* de dados através de *e-mails*, tuítes e artigos. Na verdade, não sei onde me encaixo nesse grande esquema de coisas, e como meus *bits* de dados se conectam com os *bits* produzidos por bilhões de outros humanos e computadores. Não tenho tempo para descobrir, porque eu também estou ocupado, respondendo a *e-mails*.

Na nebulosidade decorrente de certo grau de inovação tecnológica (HIDALGO, 2015, *passim*), recai sobre o legislador a tarefa de definir marcos para a tutela jurídica de determinadas contingências, ainda que nem todos os desdobramentos da tecnologia aventada possam ser sentidos de imediato ao redor do mundo (SOLOVE, 2008, p. 4) – e, com a proteção de dados pessoais, é exatamente esse o caso, uma vez que, conforme alerta Richard Warner (2011, p. 1084), compreender alguns alertas sobre riscos relacionados à privacidade demanda um comportamento que a maioria dos indivíduos simplesmente não vislumbra como viável, tendo em vista que os benefícios percebidos são, simplesmente, baixos demais.[6]

Se Herbert Marcuse (1979, p. 31-32) já indicava a existência de aspectos totalitários latentes nas sociedades industriais mais avançadas, até mesmo

[6] Anos antes da virada do milênio, Alvin Toffler (1990, p. 111) já diagnosticava o problema: "*In theory, the customers' reward for providing data will be lower prices deriving from greater efficiency in the system. (...) Since data originating with the customer are increasingly needed for the design and production (as well as distribution) of goods and services, the customer is in fact becoming a contributor to, if not an actual part of, the production process. The consumer, in a sense, is a co-producer of his or her own purchases. But does the customer in fact "own" this information? Or does it acquire value only after it is collected and processed? We lack the vocabulary, let alone laws and economic concepts, with which to deal with these unfamiliar questions arising from the information wars. But the issues involve the transfer of billions of dollars—and a subtle shift of economic and social bargaining power. What does a customer give away free to the store, the manufacturer, or his or her credit card company? Take the simplest of cases: A mother, home from work, in haste to make dinner, discovers she is out of margarine. Dashing into the nearest store, she snatches a pound of Fleischmann's sweet unsalted margarine made by Nabisco off the shelf. Hurrying to the checkout counter, she waits her turn, grabbing a copy of TV Guide from the rack near the register, and hands her purchases to the clerk, who passes them over the scanner. In principle, she has communicated the following to the store computer: (1) a type of product she uses; (2) its brand; (3) its size or amount; (4) the fact that she preferred unsalted margarine to the regular; (5) the time of the purchase; (6) what other items, brands, sizes, etc., she bought at the same time; (7) the size of her total bill; (8) the kind of magazine in which an advertiser might reach her; (9) information about where additional shelf space is now available; and much more besides.*"

sinalizando uma racionalidade tecnológica de dominação e de opressão em massa que se relacionava ao controle da consciência humana, não há dúvidas de que o advento de uma pretensa pós-modernidade acaba sendo marcado pela incredulidade frente ao metadiscurso filosófico. (LYOTARD, 1988, p. viii; BAUMAN, 2001, p. 38-40)

Nesse contexto em que a hiperexposição se manifesta na coleta e no tratamento massivo de dados, a privacidade se transforma em verdadeira *commodity*, e sua proteção adquire contornos de imperatividade. Por isso, a anonimização não pode ser confundida com a ideia de anonimato (MARX, 1999, p. 109; NISSENBAUM, 1999, p. 143-144), pois a transcende, sendo importante até mesmo para compreender elementos de impacto na formação da identidade virtual para além do nome ou de outros dados que permitam a precisa identificação do indivíduo.[7] Noutros termos, a preocupação com a proteção de dados pessoais não cuida apenas da dificultação do acesso indevido, por exemplo, ao nome, à filiação, à data de nascimento ou mesmo ao número do Cadastro de Pessoas Físicas de um indivíduo; vai além, e passa a contemplar hábitos de consumo, históricos de navegação, registros de acesso e muitas outras informações que são, por vezes, coletadas sem que o usuário sequer perceba.[8]

Segundo Rolf Weber e Ulrike Heinrich (2012, p. 73, tradução livre):

> No início da era da Internet, em parte prevaleceu a percepção de um novo mundo, sem fronteiras legais, uma vez que a Internet de médio porte é de natureza virtual e parecia difícil obter as possibilidades de monitoramento técnico pelas autoridades de Estado. Enquanto isso, essa avaliação mudou fundamentalmente; a Internet é tão regulamentada quanto o mundo tradicional e o exercício das liberdades não se tornou muito mais fácil. Em vista desses desenvolvimentos, a existência de um direito de agir anonimamente na Internet se valendo da anonimização ainda é muito debatida, (...). A proteção legalmente consolidada da vida privada, do lar e da correspondência dos participantes da Internet indica a existência de um direito de não ser totalmente monitorado;

[7] Analisando a inter-relação entre anonimato e Internet, diz a doutrina: "*With the advent of computer-based communications networks, there has been a resurgence of interest in the nature and value of certain types of anonymity. The range of techniques by which individuals are able to operate incognito creates a virtual laboratory for experimenting with the social construction of identity. However, as the Internet's surveillance potential becomes better understood and exploited, the measure of anonymity shifts from its historical focus on names to a broader investigation of a range of personal identifiers that can be linked to an individual.*" (STAPLES, 2007, v. 1, p. 17)

[8] Com efeito, "(...) os modelos de negócio assentes na *Big Data* possibilitam um controle permanente sobre os consumidores, com dados obtidos e tratados em tempo real, conservados por tempo indeterminado, com vista a obter informações ainda não evidentes, muitas delas nem sequer previstas no momento de recolha e tratamento inicial dos dados. O que tende a conduzir ao estabelecimento de perfis detalhados para cada cliente, depois usados para prever e avaliar os respectivos comportamentos." (MASSENO, 2019, p. 8.)

> de fato, os Estados têm a obrigação de criar um ambiente livre de vigilância, melhorando os quadros legislativos existentes. No entanto, o direito de confiar no anonimato não pode ficar sem limites, uma vez que os interesses do Estado existem, justificando a intervenção governamental na esfera dos indivíduos.[9]

Eis o papel da anonimização (artigo 5º, IX, da LGPD): trata-se do processo técnico que nada mais representa do que a dissociação entre determinado dado pessoal e o seu respectivo titular. Para seu implemento, inúmeros procedimentos específicos podem ser utilizados, quase sempre a partir da eliminação de determinados elementos identificadores que constam de uma base de dados, por meio de supressão do dado, generalização, randomização ou pseudonimização. (MARTINS; FALEIROS JÚNIOR, 2019, p. 61)

Sobre o tema, a doutrina europeia se reporta ao termo "pseudonimização" para descrever o mesmo processo. A despeito de qualquer predileção semântica, porém, insta anotar que o dado anonimizado não se confunde com o dado "anônimo"; se situa, em verdade, em posição mediana em um especto que varia entre o dado pessoal e o anônimo, ou, nos dizeres de Doneda e Machado, em um *continuum* descritivo que demanda investigações mais profundas do que a puramente semântica:

> Em meio a essa discussão, com uma abordagem orientada pelo risco, há o surgimento de propostas que visualizam entre o dado pessoal e o dado anônimo um gradiente de cores ou um *continuum* com categorias que superam a lógica binária dado pessoal/dado anônimo, informações a que não se aplicam o regime de proteção de dados pessoais. É nesse contexto que se coloca a ideia de dado pseudonimizado. (DONEDA; MACHADO, 2019, p. 149)

Em razão dessa dicotomia conceitual, muitas dúvidas ainda pairam em relação aos impactos da anonimização e de sua (potencial) reversão. Fala-se em re-identificação e repersonalização como sinônimos do processo de reversão da anonimização de dados, mas ainda é preciso compreender com maior

[9] No original: "*At the beginning of the Internet era partly the perception of a new world without legal borders prevailed since the medium Internet is of a virtual nature and technical monitoring possibilities by State authorities seemed to be difficult to obtain. Meanwhile, this assessment has fundamentally changed; the World Wide Web is as regulated as the traditional world and the exercise of freedoms did not become much easier. In view of these developments the existence of a right to act anonymously on the Internet by using anonymization is still hotly debated, (...). The legally consolidated protection of private life, home and correspondence of Internet participants pleads for the existence of a right of not being totally monitored; in fact, States have the obligation to create an environment free of surveillance by improving the existing legislative frameworks. However, a right to rely on anonymity cannot be without limits since State interests do exist, justifying governmental intervention into the sphere of individuals.*"

profundidade, a partir de parâmetros objetivos, o modo como isso ocorre e as consequências que traz.

3. REVERSÃO DA ANONIMIZAÇÃO E SEUS IMPACTOS

O "grau de identificabilidade" de um dado pessoal é o parâmetro utilizado na aferição dos processos de anonimização, que visa à prevenção da reversão para que não seja possível, por quaisquer métodos, recuperar a informação originalmente anonimizada. Para isso, são utilizados procedimentos de supressão, ocultação ou fragmentação da informação, como exemplifica Bruno Bioni (2019, p. 71-72):

> *a) supressão do CPF:* por ser um identificador capaz de diferenciar até mesmo pessoas homônimas, sendo um identificador único; logo, a sua disponibilização, ainda que parcial – *e.g.*, cinco primeiros dígitos –, não seria prudente;
> *b) generalização do nome completo:* constaria apenas o prenome, desde que fosse observado que os nomes da base de dados não são comuns. O objetivo é evitar que um nome possa ser atribuído a um indivíduo em específico;
> *c) generalização da localização geográfica:* em vez de disponibilizar o número completo do CEP, seriam divulgados apenas os seus primeiros dígitos. Assim, haveria uma localização menos detalhada, a fim de quebrar o vínculo de identificação desta informação com um sujeito;
> *d) generalização da idade:* em vez de divulgar a idade exata, seria divulgada a faixa etária para viabilizar a categorização dos indivíduos como jovens, adultos ou idosos (coluna "E") e, por outro lado, inviabilizar a sua individualização, dado o universo de pessoas que se enquadram naquela mesma faixa etária.

Evidentemente, quando se questiona a confiabilidade dos processos de anonimização[10], nota-se, por parte da doutrina estrangeira mais especializada, sólido esforço na realização de investigações técnicas sobre a possibilidade e a facilidade de se concretizar um procedimento de anonimização confiável e de difícil reversão.

Segundo Doneda e Machado (2019, p. 149):

> No exame da robustez e do nível de garantia oferecidos por técnicas e práticas de anonimização de dados, sugere-se que três

[10] Com efeito: "*In the coming years, analytics will offer an enormous opportunity to generate economic and social value from data. But much of the success of data analytics will depend on the ability to ensure that individuals' privacy is respected. One of the most effective ways in which to do this is through strong "de-identification" of the data — in essence, storing and sharing the data without revealing the identity of the individuals involved*". (CAVOUKIAN; CASTRO, 2014, p. 1)

> tipos de riscos principais sejam levados em consideração: distinção (*singling out*), possibilidade de ligação e inferência. O primeiro versa sobre a possibilidade de se isolar alguns ou todos os registros que destacam uma pessoa em uma base de dados; o segundo é a capacidade de se estabelecer uma conexão entre pelo menos dois registros relativos ao mesmo indivíduo ou mesmo grupo de pessoas; e o terceiro, por fim, diz com a possibilidade de deduzir, com uma significativa probabilidade, o valor de um atributo a partir dos valores de um conjunto de outros atributos.

Paul Ohm é o autor que mais tem se dedicado a esse estudo sob uma perspectiva jurídica. Desde 2010, seus estudos têm sido assertivos em dizer que a anonimização se tornou ubíquita:

> Algo importante mudou. Por décadas, os tecnólogos acreditavam que poderiam proteger de forma robusta a privacidade das pessoas fazendo pequenas alterações em seus dados, usando as técnicas pesquisadas abaixo. Eu chamo isso de suposição robusta de anonimização. Abraçando essa suposição, os reguladores e tecnólogos prometeram privacidade aos usuários e, por sua vez, a privacidade é o que os usuários esperam. Hoje, a anonimização é onipresente. (...) Ao minar a robusta suposição de anonimização, a fácil reidentificação derrubará os edifícios de promessa e expectativa que construímos sobre a anonimização. O resultado fácil da reidentificação também causará estragos em nossos sistemas jurídicos, porque nossa fé na anonimização robusta se infiltrou completamente em nossas leis e regulamentos de privacidade, (...).[11] (OHM, 2010, p. 1706-1707, tradução livre)

A Lei Geral de Proteção de Dados brasileira situou em polaridades opostas os dados pessoais e os dados anonimizados, deles cuidando em seu artigo 12, de forma expressa, mas com ressalvas contidas nos parágrafos subsequentes ao *caput*:

> Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de

[11] No original: "*Something important has changed. For decades, technologists have believed that they could robustly protect people's privacy by making small changes to their data, using techniques surveyed below. I call this the robust anonymization assumption. Embracing this assumption, regulators and technologists have promised privacy to users, and in turn, privacy is what users have come to expect. Today, anonymization is ubiquitous. (...) By undermining the robust anonymization assumption, easy reidentification will topple the edifices of promise and expectation we have built upon anonymization. The easy reidentification result will also wreak havoc on our legal systems because our faith in robust anonymization has thoroughly infiltrated our privacy laws and regulations, (...).*"

> anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.
> § 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.
> § 2º. Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.
> § 3º. A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.

A leitura do *caput* e dos parágrafos desse dispositivo revela algumas nuances cruciais: (i) se realizada a reversão, o dado anonimizado passa a ser considerado dado pessoal para os fins definidos na lei; (ii) procurou-se estabelecer um critério de 'razoabilidade objetiva' para que os desdobramentos da reversão gerem consequências jurídicas; (iii) estabeleceu-se, no §2º, uma conceituação específica para os dados pessoais utilizados em aferição algorítmica de perfil comportamental; (iv) atribuiu-se competência à Agência Nacional de Proteção de Dados para dispor sobre parâmetros técnicos de higidez da anonimização. (MARTINS; FALEIROS JÚNIOR, 2019, p. 65-66)

Nesse contexto, Paul Ohm (2010, p. 1711-1712) descreve a técnica denominada *release-and-forget*, tida como o procedimento mais adequado para que se proceda à anonimização de um dado. Para o autor, este processo se dá a partir da revelação pontual a terceira pessoa (ou internamente, dentro de sua própria organização) de determinado dado pessoal, e, imediatamente após, ela o esquece, o que significa dizer que tal pessoa não praticará nenhuma tentativa de rastrear o que vier a acontecer com os registros após o lançamento. Entretanto, em vez de colocar os seus dados em risco inadvertidamente, antes da liberação, a parte reveladora modifica algumas das informações. Tal sistemática é descrita por Ohm como *easy anonymization*, ao passo que técnicas mais apuradas são categorizadas como *robust anonymization* e envolvem os já citados instrumentais de supressão do dado, generalização, randomização ou pseudonimização.

A reversão, porém, pode ocorrer a partir de algumas situações específicas, também listadas por Paul Ohm e assim sintetizadas por Martins e Faleiros Júnior (2019, p. 65-66):

> a) *O Adversário*: os cientistas da computação modelam o anonimato e a reidentificação como um jogo no qual a anonimização é,

> simplesmente, um movimento de deflagração. (...) A característica definidora do adversário parece ser que ele ou ela é – o que não surpreende – o adversário - motivado para fazer algo com os dados que o administrador não deseja que aconteça. São citados como exemplos de potenciais "adversários", em alusão aos estudos de Narayanan e Shmatikov [12], *stalkers*, investigadores, colegas curiosos, empregadores e vizinhos.
>
> b) *Informação Externa*: quando um adversário encontra uma 'impressão digital' de dados única, ele pode vincular esses dados para informações externas, às vezes chamadas de informações auxiliares. Diversas técnicas poderiam ser úteis nesse contexto, se apenas o adversário não soubesse mais sobre a natureza humana, pois, na realidade, o mundo está inundado de dados sobre pessoas, com novos bancos de dados criados todos os dias. Adversários combinam dados anonimizados com informações externas para extrair detalhamentos sobre determinadas identidades.
>
> c) *Princípio Básico das Mãos Cruzadas e Junções Internas*: Segundo o autor, um especialista em segurança de computadores lhe resumiu todo o campo de possibilidades de reversão da anonimização ao simplesmente juntar as mãos, intercalando seus dedos, como se pretendesse orar. Esta simples imagem mental bem resume a operação básica de reidentificação. Ao se imaginar que a mão esquerda é composta por dados anônimos, a direita será composta pelas informações externas, e o entrelaçamento entre ambas representa, a partir dos dedos intercalados, lugares onde, em havendo correção do cruzamento de dados, a reidentificação é alcançada.

Outras abordagens – mais voltadas à análise matemática e computacional – provaram que a reversão da anonimização (ou reidentificação) pode ser eficientemente alcançada sem grande complexidade técnica. (MONTJOYE *et al*, 2013, p. 1-5) Para tanto, a problemática concernente à anonimização envolve os chamados "quasi-identificadores", variáveis que podem não identificar indivíduos diretamente, mas que circunstancialmente mantêm uma correlação substancial com elementos identificadores únicos e que ainda podem ser usadas para a reidentificação indireta (reversão), o que torna o método clássico de *release-and-forget* naturalmente obsoleto para quaisquer fins jurídicos.

A combinação de dados pessoais com informações externas foi o objeto de um estudo conduzido por Latanya Sweeney (2000), no qual se descobriu que 87% (216 milhões de 248 milhões) da população dos Estados Unidos da América tem características relatadas que provavelmente as tornam únicas com base somente em

[12] Sobre os pioneiros estudos dos autores, confira-se: NARAYANAN; SHMATIKOV, 2010, p. 24-26.

um código ZIP de 5 dígitos [equivalente ao CEP brasileiro], gênero e data de nascimento. Também se constatou que aproximadamente 53% dos cidadãos norte-americanos (132 milhões de 248 milhões) devem ser unicamente identificáveis somente por local, gênero e data de nascimento, onde o local é basicamente o município em que a pessoa reside.

Esses quasi-identificadores são úteis tanto em si mesmos quanto nas hipóteses em que venham a ser combinados com outras informações disponíveis para identificar unicamente indivíduos. Os estudos de Paul Ohm (2010), Arvind Narayanan e Vitaly Shmatikov (2010) revelam preocupações que vão além das políticas de integridade e da *privacy by design*.[13] O filtro de razoabilidade depende de uma "régua" que viabilize a imputação de responsabilidade civil em caso de reversão.

4. ALGORITMOS E ENTROPIA DE DADOS: UM CAMINHO NECESSÁRIO

Os algoritmos empregados nos processos de reidentificação são considerados agnósticos quanto à semântica dos elementos de dados, ou seja, não há qualquer variável subjetiva em situação de proponderância subjacente à delimitação contextual dos dados analisados e que se pretenda reverter; em simples termos, os algoritmos são frios.

Há que se considerar, todavia, um amplo espectro de características humanas que permitem a reidentificação, tais como as preferências de consumo, o histórico de transações comerciais, os registros de navegação *web* e de busca, dentre vários outros que se conectam àquilo que, algoritmicamente, Tim Wu (2016, p. 267) descreve como uma "engenharia de demanda", o que significa dizer que, dentro de um contexto no qual a privacidade é tomada em sua acepção conceitual, a formação de bases de dados anonimizadas deve ser encarada sob suas duas propriedades-chave: (i) estabilização razoável através do tempo e de contextos; e que (ii) os atributos de dados correspondentes sejam, de fato, suficientemente numerosos e detalhados para que seja muito improvável encontrar duas pessoas similares.

Com efeito:

[13] A *privacy by design* é veementemente defendida por Ann Cavoukian (2011), Information and Privacy Commissioner da província de Ontario, no Canadá, que delineou os 7 princípios fundamentais de sua implementação. Ademais, consta do artigo 25 da *General Data Protection Regulation* europeia, impõe a qualquer ação empreendida por uma empresa que atue no processamento de dados pessoais, os imperativos da proteção de dados e da privacidade em todas as suas etapas. Isso inclui projetos internos, desenvolvimento de produtos, desenvolvimento de *software*, sistemas de TI e muito mais. Na prática, significa que o departamento de TI, ou qualquer departamento que processe dados pessoais, deve garantir que a privacidade seja incorporada a um sistema durante todo o ciclo de vida do sistema ou processo. Não se confunde com a *privacy by default* (privacidade por padrão ou por defeito), que está relacionada aos produtos e serviços liberados ao público, cujas configurações de privacidade mais rígidas devem ser aplicadas por padrão, sem nenhuma entrada manual do usuário final. Além disso, quaisquer dados pessoais fornecidos pelo usuário para permitir o uso ideal de um produto devem ser mantidos somente durante o tempo necessário para fornecer o produto ou serviço. Se mais informações do que o necessário para fornecer o serviço forem divulgadas, haverá violação da "privacidade por padrão". Para maiores detalhes, consulte-se: ROST; BOCK, 2011, *passim*.

> (...) em princípio, eventual dicotomia entre dados pessoais e dados anônimos só guardaria coerência junto ao conceito reducionista de dados pessoais. Isso porque dados anônimos não são dados relacionados a uma pessoa identificada, demandando a reversão do processo de anonimização para se chegar aos respectivos titulares, sendo a sua identificabilidade remota (identificável) e não imediata (identificada). Dessa forma, leis que adotam o conceito expansionista de dados pessoais e, ao mesmo tempo, estabelecem uma dicotomia deste com dados anônimos correriam o risco de ser tautológicas. Isso porque haveria uma redundância normativa, já que dados anônimos seriam, em última análise, potencial e provavelmente, dados relacionados a uma pessoa identificável. (BIONI, 2019, p. 75)

Por um lado, determinados atributos podem ser identificadores únicos por si mesmos, não demandando qualquer intervenção algorítmica para reverter a anonimização; por outro, qualquer atributo pode ser um identificador em combinação com outros dados ou mesmo com outras bases de dados, em cruzamentos informacionais. (*Cf.* HARTZOG, 2013) A mineração de dados, emanada de hipóteses como as do mapeamento dos livros que uma pessoa já leu ou até mesmo do acervo de roupas existente em seu armário não revelam, isoladamente, nenhum único elemento capaz de, em perfunctória aferição, ser considerado um "quasi-identificador" ou um subconjunto suficientemente grande para a identificação de um indivíduo [14]; porém, a partir do intercâmbio informacional e da heurística – aqui considerada como um repositório de critérios e parâmetros que, a partir de processos, permite localizar problemas de usabilidade em interfaces computador-usuário (ROSA; VERAS, 2013, p. 140) –, isto se torna plausível.

Nielsen e Molich (1990) propõem um rol de dez parâmetros de heurística computacional cabíveis em uma análise de segurança da informação: (i) visibilidade do estado do sistema; (ii) correspondência entre o sistema e o mundo real (ou seja, o sistema deve "falar a linguagem do usuário"); (iii) liberdade e

[14] Sobre o assunto, veja-se: "*Data mining is a process where critical business data are analyzed to gain new insights about customers, businesses, and markets. This new knowledge gained can be used to improve customer relationships and to produce better-quality products and services, which can result in higher revenues and profits. These data are generally in a relational or multidimensional format and stored in companies' central data warehouses. But with the evolution of the enterprise, a diverse set of data structures have come to be used: graph data, which could feed from social network sites; time series data; longitudinal data; semistructured data, such as XML; unstructured data; and big data. There is a need for different data repositories to store all these diverse data. Analytics is carried out on the data in the repositories. Access to these data repositories is strictly controlled by access control rights. Strict security measures are employed to secure the data as they are very sensitive and contain customer-identifying information. In addition to all security measures, companies ensure that the data are anonymized before being used for analytics/mining. More often, companies share their data with specialized analytics firms, and the data need to be protected before sharing*." (VENKATARAMANAN; SHRIRAM, 2017, p. 99)

controle por parte do usuário; (iv) adoção de consistência e padrões; (v) prevenção de erros; (vi) reconhecimento preferível à memorização, evitando-se que o usuário tenha que se lembrar de grande número de informações para usar o sistema; (vii) flexibilidade e eficiência de uso; (viii) *design* estético e minimalista; (ix) ajuda sistêmica aos usuários para o diagnóstico e a recuperação dos erros; (x) suporte e documentação.

Surge, neste âmbito, o problema do acréscimo ou crescimento (*accretion problem*, na expressão em inglês, também definida por Paul Ohm): "A partir do momento em que um adversário *linkou* dois bancos de dados anonimizados, ele pode adicionar os novos dados *linkados* para sua coleção de informações externas e usá-los para destravar outros bancos de dados anonimizados."[15] Segundo Bruno Bioni (2019, p. 76), "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".

A dificuldade reside na aferição do que é ou não um "esforço razoável" nesse exercício de definição da potencialidade de que um dado seja vinculado a uma pessoa. Nesse aspecto é que reside a dificuldade de aferição do que é ou não passível de enquadramento nos marcos regulatórios de proteção de dados.

O problema do crescimento (*accretion*) contribui fortemente para a reidentificação (ou reversão da anonimização) e tudo funciona como se o problema do crescimento conectasse todos os bancos de dados do mundo em um único, gigante, acervo de extremo interesse para pessoas mal-intencionadas, o que realça o papel dos reguladores.

A Lei Geral de Proteção de Dados descreve a almejada "razoabilidade" em seu artigo 12, §1º, segundo o qual "[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", revelando-se "uma estratégia normativa alinhada à premissa de que os dados anônimos seriam sempre passíveis de reversão. Daí por que o critério da razoabilidade nada mais é do que uma diretriz acerca do que venha a ser um *risco aceitável-tolerável*". (BIONI, 2019, p. 76)

Percebe-se uma correlação inescapável entre heurística e usabilidade sistêmica, mas alguns dos princípios elencados denotam a relevância do incremento das referidas exigências para que se aprimore a fidúcia que deve ser inerente à relação entre o homem e o computador, inclusive para a prevenção de riscos.

É nesse contexto que se defende o conceito de *entropia* de dados: o termo é utilizado na física para, em um sistema termodinâmico bem definido, medir seu grau de irreversibilidade; por isso, parece definir com exatidão o espírito da "razoabilidade" insculpida na lei. Seu uso é atribuído a Paul Ohm (2010, p. 1760) e

[15] No original: "*Once an adversary has linked two anonymized databases together, he can add the newly linked data to his collection of outside information and use it to help unlock other anonymized databases*." (OHM, 2010, p. 1746).

"surge como um parâmetro de reforço. Para além da razoabilidade que a lei já prevê, seria possível, a depender da heurística aplicada na aferição de riscos de determinado procedimento de reidentificação, inferir falibilidades e, consequentemente, responsabilidades." (MARTINS; FALEIROS JÚNIOR, 2019, p. 74)

Paul Ohm ainda descreve:

> Grandes redutores de entropia são entidades que acumulam bancos de dados massivos que contém tantas associações entre tantos tipos diferentes de informação que elas representam uma parte significativa do banco de dados da ruína, mesmo que eles removam de seus bancos de dados todas as informações particularmente sensíveis e diretamente associáveis. (...)
> Quem são os grandes redutores de entropia? Na metáfora do corredor, são as pessoas segurando muitas chaves; imagine o chaveiro do lendário zelador, tilintando com dezenas de chaves diferentes. Na prática, essa categoria inclui grandes agências de crédito como Experian, TransUnion, e Equifax; corretores comerciais de dados como ChoicePoint, Acxiom e LexisNexis; e provedores de busca na Internet como Google, Microsoft e Yahoo.[16] (OHM, 2010, p. 1760)

A "metáfora do corredor" citada no excerto acima é didaticamente explicada pelo autor imaginando-se que determinada pessoa se posiciona em um dos lados de um longo corredor dedicado somente a ela e, do lado oposto, está situado o fato arruinador daquela pessoa, ou seja, algum segredo ou detalhe que uma pessoa inimiga poderia usar para prejudicá-la; porém, ao longo deste corredor, tem-se uma série de portas fechadas e trancadas, e cada cadeado requer uma chave diferente, que representa os atributos de banco de dados que devem ser reconectados ou os *links* na cadeia de inferências que devem ser estabelecidos para conectar a pessoa ao fato.

Ohm (2010, p. 1760-1761) explica que muitas outras pessoas são as guardiãs das chaves para algumas das portas, ou seja, controlam, cada qual, um banco de dados, e cada chave representa as inferências que ela pode fazer, usando os dados que possui. A presença dessas entidades, se delas for exigido o rigor no zelo pelas "chaves", incrementará a segurança do sistema, levando à almejada entropia.

[16] No original: "*Large entropy reducers are entities that amass massive databases containing so many links between so many disparate kinds of information that they represent a significant part of the database of ruin, even if they delete from their databases all particularly sensitive and directly linkable information. (...) Who are large entropy reducers? In the hallway metaphor, they are the people clutching many keys; imagine the mythical janitor's keyring, jangling with dozens of different keys. In practice, this category includes large credit agencies like Experian, TransUnion, and Equifax; commercial data brokers like ChoicePoint, Acxiom, and LexisNexis; and internet search providers like Google, Microsoft, and Yahoo*".

Basicamente, o que se extrai desse raciocínio é que danos legalmente reconhecíveis e capazes de arruinar reputações surgem de inúmeras situações específicas, na medida em que praticamente todas as pessoas podem ser interligadas por ao menos um fato que esteja inserido em um banco de dados de computador e porque este fato pode vir a ser utilizado por terceiros para fins de chantagem, discriminação, assédio, fraudes de identidade ou mesmo para causar constrangimento ou inconveniência.

A primazia da proteção de dados perpassa pela exigência da entropia, ou seja, para muito além do que se possa considerar "razoavelmente reversível", é necessário que se indique elementos mínimos para a confiabilidade da anonimização, não se podendo deixar de ponderar que a definição de dados anonimizados leva em conta todo o conjunto de meios suscetíveis para a razoável reidentificação do usuário, e tais meios estão em constante desenvolvimento, dependendo fundamentalmente do acesso a outros bancos de dados, de uma forma praticamente impossível de ser mensurada do ponto de vista da segurança.

Para a hipótese ora apreciada, clama-se pela governança de dados, que deve contemplar, também, os processos de anonimização (e os riscos de sua reversão), inclusive na eventual lacuna regulatória atribuída pela LGPD brasileira à Agência Nacional de Proteção de Dados (artigo 12, §3º), em verdadeira manifestação da chamada 'regulação responsiva' descrita por Ayres e Braithwaite (1992).[17]

A LGPD caminhou no sentido do fenômeno que David Jackman (2015, p. 200) denomina de 'revolução do *compliance'*, especialmente em seu artigo 50, que impõe a observância dos parâmetros de governança, ainda que como uma faculdade do agente de dados – e não um dever –, tema sobre o qual discorrem Martins e Faleiros Júnior (2020, p. 364-365):

> Infelizmente, o uso do verbo "poderá" acaba por esvaziar, em certa medida, o inegável potencial da norma, deixando para as grandes corporações – que, costumeiramente já se valem de programas de integridade desta estirpe – a continuidade do implemento de boas práticas.
>
> Considerando-se que a norma é uma lei "geral", ou seja, aplicável a todo aquele que efetue operações de coleta, tratamento e armazenagem de dados, a imposição do *compliance*, mesmo para pequenas e médias empresas, representaria inegável avanço rumo

[17] Comentam: "*The optimal structuring of responsive regulation will have two components: (1) choosing the appropriate form of delegation, and (2) choosing the right kind of escalating (nondelegated) regulation. Choosing the appropriate form of delegation focuses on what the base of the pyramid should be. The choice of specific forms of contingently escalating responses is the issue in the design of an appropriate enforcement pyramid. Delegation is only credible if it is backed up with various forms of more traditional regulatory fiat. Where delegation is successful, the upper reaches of the escalating interventions will seldom if ever be activated. However, the very success of the regulatory delegation will turn on the existence of a credible government response if the delegation fails to engender compliance and problem-solving.*" (AYRES; BRAITHWAITE, 1992, p. 161).

> à inserção de parâmetros ético-preventivos voltados à minimização de danos em operações tão arriscadas como as que envolvem dados pessoais (especialmente os sensíveis).
> Vale dizer que, apesar da vastidão de elementos que um bom programa de governança de T.I. demanda, não é inviável sua implementação em pequena ou média escala, e sua imposição certamente representaria valioso avanço no contexto da propagação de uma cultura de prevenção.

Volta à tona a questão da entropia de dados, que Jane Yakowitz (2011, p. 44) assim descreve:

> Sob a ótica do meu paradigma, uma entidade que produz dados é obrigada a fazer somente duas coisas para converter dados que identificam pessoas em dados anonimizados: (1) retirar todos os identificadores diretos, e (2) ou checar por tamanhos mínimos de subgrupos em uma lista pré-estabelecida de identificadores diretos comuns – como etnia, sexo, indicadores geográficos e outros identificadores indiretos comumente encontrados em registro públicos – ou usar um quadro de amostragem aleatório eficaz.

A dicotomia conceitual entre dado pessoal e dado anonimizado – que os coloca em polaridades opostas do ponto de vista da tutela normativa – desvela uma *ratio* cognoscível, pois, se assim não o fosse, qualquer dado (mesmo o anônimo) poderia vir a ser um dado pessoal. A questão que se coloca diz respeito aos limites da filtragem que mantém a polarização conceitual e, nesse particular, o termo *entropia* parece definir com exatidão o espírito da "razoabilidade" insculpida na lei.

5. CONCLUSÃO

Nesta brevíssima exposição, procurou-se destacar o papel da regulação de determinadas práticas relacionadas ao trato jurídico da informação para o contingenciamento de situações hodiernas como a coleta massiva de dados pessoais para as mais variadas atividades, envolvendo sua identificação, classificação e aplicação a partir de algoritmos.

Destacou-se, no curso da exposição, que a sociedade da informação impõe a necessidade de controles contra a atividade exploratória e abusiva de informações, ampliando progressivamente o espaço para a proteção de dados pessoais e demandando novas emanações para o direito fundamental à privacidade.

Na Europa, foi promulgado o Regulamento Geral de Proteção de Dados (RGPD), cuja implementação se deu em 25 de maio de 2018, e, no Brasil, a Lei nº 13.709, de 14 de agosto de 2018 (a Lei Geral de Proteção de Dados, ou LGPD), com vigência a partir de agosto de 2020. A partir dessas legislações, um tema chama a

atenção: com o intuito de retirar dos dados pessoais coletados quaisquer informações que permitam identificar e, evidentemente, expor a pessoa à qual dizem respeito, alguns processos técnicos prometem viabilizar o uso desses dados sem risco de violações. Trata-se da anonimização de dados, que cumpre esse papel de retirar certos elementos dos dados pessoais, tornando-os dados anonimizados.

Questionando a confiabilidade de tais mecanismos, destacou-se que a heurística computacional surge como um passo inicial a ser exigido dos agentes de dados. Anotou-se, ainda, que o artigo 12, §1º, da LGPD brasileira se limitou a trabalhar com um filtro de razoabilidade que, embora conceituado no corpo do texto normativo, depende de atualizações constantes à luz das melhores técnicas da matemática e da ciência da computação.

Avançou-se, em seguida, à hipótese versada: a denominada *entropia* de dados, originalmente sugerida por Paul Ohm, que aparece como conceito adequado à expectativa não apenas de que se tenha boas práticas no implemento de técnicas de anonimização de dados para fins de proteção da privacidade, mas também como uma "régua" que conduza à aferição, por parâmetros objetivos, do potencial de reversão de um determinado método empregado para anonimizar dados.

As conclusões fundamentais extraídas dessa análise, evidentemente, impõem considerar a diferença semântica entre "dados anônimos" e "dados anonimizados". Os primeiros sequer foram considerados pela lei; os segundos são o objeto deste estudo e revelam a importância do tema, na medida em que a disponibilidade cada vez mais crescente de grandes bancos de dados (*Big Data*) e o aprimoramento das técnicas de cruzamento de dados e reidentificação/repersonalização fazem com que até mesmo bancos de dados que estejam assegurados por técnicas de anonimização possam ter tais processos revertidos no futuro.

Outrossim, o conceito legal de razoabilidade (artigo 12, §1º), se conjugado com a almejada entropia e, exigindo clareza quanto às técnicas empregadas no procedimento de anonimização para a coleta desse tipo de dado, pode contribuir para manter a utilidade desses dados para pesquisas e estatísticas internas sem enfraquecer a legislação. A adoção da heurística, compreendida como o rol de boas práticas aplicáveis aos processos computacionais, finalmente, deve perpassar por fortes e constantes atualizações para a anonimização, levando em conta os últimos avanços da matemática e da ciência da computação sobre o tema para, em um segundo momento, assegurar a almejada proteção a partir do respeito irrestrito à entropia de dados.

REFERÊNCIAS

AYRES, Ian; BRAITHWAITE, John. **Responsive regulation:** transcending the deregulation debate. Nova York: Oxford University Press, 1992.

BAUMAN, Zygmunt. **Modernidade líquida.** Tradução de Plínio Dentzien. Rio de Janeiro: Zahar, 2001.

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

CAVOUKIAN, Ann. Privacy by Design. The 7 foundational principles: implementation and mapping of fair information practices. **Internet Architecture Board,** 2011. Disponível em: https://iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf. Acesso em: 23 fev. 2020.

CAVOUKIAN, Ann; CASTRO, Daniel. Big Data and innovation, setting the record straight: de-identification does work. **The Information Technology & Innovation Foundation,** Ontario, p. 1-18, jun. 2014.

CORREIA, Victor. **Da privacidade:** significado e valor. Coimbra: Almedina, 2018.

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

DONEDA, Danilo; MACHADO, Diogo. Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados. *In:* DONEDA, Danilo; MACHADO, Diogo (Coords.). **A criptografia no direito brasileiro.** São Paulo: Revista dos Tribunais, 2019.

GREENGARD, Samuel. **The internet of things.** Cambridge: The MIT Press, 2015.

HARARI, Yuval Noah. **21 lições para o século 21.** Tradução de Paulo Geiger. São Paulo: Cia. das Letras, 2018.

HARTZOG, Woodrow. Social Data. **Ohio State Law Journal,** Columbus, v. 74, n. 6, p. 995-1028, 2013.

HIDALGO, César. **Why information grows:** the evolution of order, from atoms to economies. Nova York: Basic Books, 2015.

JACKMAN, David. **The compliance revolution:** how compliance needs to change to survive. Nova Jersey: John Wiley & Sons, 2015.

LESSIG, Lawrence. **Code, and other laws of cyberspace**. 2. ed. Nova York: Basic Books, 2006.

LORENZETTI, Ricardo Luis. **Comércio eletrônico.** Tradução de Fabiano Menke. São Paulo: Revista dos Tribunais, 2004.

LYOTARD, Jean-François. *O* **pós-moderno.** Tradução de Ricardo Corrêa. 3. ed. Rio de Janeiro: José Olympio, 1988.

MARCUSE, Herbert. **A ideologia da sociedade industrial**. Tradução de Giasone Rebuá. 5. ed. Rio de Janeiro: Jorge Zahar, 1979.

MARTINS, Guilherme Magalhães; FALEIROS JÚNIOR, José Luiz de Moura. A anonimização de dados pessoais: consequências jurídicas do processo de reversão, a importância da entropia e sua tutela à luz da Lei Geral de Proteção de Dados. *In:* DE LUCCA, Newton; SIMÃO FILHO, Adalberto; LIMA, Cíntia Rosa Pereira de; MACIEL, Renata Mota (Coord.). **Direito & Internet IV:** sistema de proteção de dados pessoais. São Paulo: Quartier Latin, 2019.

MARTINS, Guilherme Magalhães; FALEIROS JÚNIOR, José Luiz de Moura. Segurança, boas práticas, governança e compliance. *In:* LIMA, Cíntia Rosa Pereira de (Coord.). **Comentários à Lei Geral de Proteção de Dados:** Lei n. 13.709/2018, com alteração da Lei n. 13.853/2019. São Paulo: Almedina, 2020.

MARX, Gary T. What's in a name? Some reflections on the sociology of anonymity. **The Information Society,** Oxfordshire, v. 15, n. 2, p. 99-112, 1999.

MASSENO, Manuel David. Como a União Europeia procura proteger os cidadãos-consumidores em tempos de Big Data. **Revista Eletrônica do Curso de Direito da UFSM,** Santa Maria, v. 14, n. 3, p. 1-27, set./dez. 2019.

MAYER-SCHÖNBERGER, Viktor; RAMGE, Thomas. **Reinventing capitalism in the age of big data.** Nova York: Basic Books, 2018.

MONTJOYE, Yves-Alexandre de; HIDALGO, César A.; VERLEYSEN, Michel; BLONDEL, Vincent D. Unique in the Crowd: The privacy bounds of human mobility. **Scientific Reports,** Londres, v. 3, art. n. 1376, p. 1-5, mar. 2013.

NARAYANAN, Arvind; SHMATIKOV, Vitaly. Myths and fallacies of "Personally Identifiable Information". **Communications of the ACM,** Austin, v. 53, n. 6, p. 24-26, jun. 2010.

NIELSEN, Jakob; MOLICH, Rolf. Heuristic evaluation of user interfaces. *CHI'90* **Conf. Proceedings**, Seattle, EUA, 1-5 abril, p. 249-256, 1990.

NISSENBAUM, Helen. The meaning of anonymity in an Information **Age. The Information Society,** Oxfordshire, v. 15, n. 2, p. 141-144, 1999.

OHM, Paul. Broken promises of privacy: responding to the surprising failure of anonymization**. UCLA Law Review,** Los Angeles, v. 57, p. 1701-1777, 2010.

REIDENBERG, Joel R. Lex Informatica: the formulation of information policy rules through technology. **Texas Law Review**, Austin, v. 76, n. 3, p. 553-584, 1998.

ROSA, Juan Miguel; VERAS, Manoel. Avaliação heurística de usabilidade em jornais online: estudo de caso em dois sites. **Perspectivas em Ciência da Informação,** Belo Horizonte, v.18, n. 1, p.138-157, jan./mar. 2013

ROST, Martin; BOCK, Kirsten. Privacy by Design and the new protection goals: principles, goals, and requirements. **Datenschutz-Geschichte.de,** 2011. Disponível em: http://www.datenschutzgeschichte.de/pub/privacy/BockRost_PbD_DPG_en_v1f.pdf. Acesso em: 23 fev. 2020.

SOLOVE, Daniel J. **Understanding privacy.** Cambridge: Harvard University Press, 2008.

STAPLES, William G. **Encyclopedia of privacy.** Westport: Greenwood Press, 2007, v. 1.

SWEENEY, Latanya. Simple demographics often identify people uniquely. **Carnegie Mellon Universit***y*, Pittsburgh, Data Privacy Working Paper n. 3, p. 2-34, 2000.

TOFFLER, Alvin. **Powershift: k**nowledge, wealth, and violence at the edge of the 21st century. Nova York: Bantam Books, 1990.

VENKATARAMANAN, Nataraj; SHRIRAM, Ashwin. **Data privacy:** principles and practice. Boca Raton: CRC Press, 2017.

WARNER, Richard. Undermined norms: the corrosive effect of information processing technology on informational privacy. **St. Louis University Law Journal,** St. Louis, v. 55, p. 1047-1086, 2011.

WARREN, Samuel D.; BRANDEIS, Louis D. The right to privacy. **Harvard Law Review,** Cambridge, v. 4, n. 5, p. 193-220, dez. 1890.

WEBER, Rolf H.; HEINRICH, Ulrike I. **Anonymization.** Cham/Basileia: Springer, 2012.

WIENER, Jonathan B. The regulation of technology, and the technology of regulation. **Technology in Society**, Durham, n. 26, p. 483-500, 2004.

WU, Tim. **The attention merchants:** the epic scramble to get inside our heads. Nova York: Vintage, 2016.

YAKOWITZ, Jane. Tragedy of the data commons. **Harvard Journal of Law and Technology,** Cambridge, v. 25, n. 1, p. 1-67, mar. 2011.