www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

O conceito de privacidade diferencial em relação à reidentificação de dados pessoais

DOI: https://doi.org/10.22409/pragmatizes.v10i19.41180

Agenor Alexsander C. Costa[1]

Maurício C. S. Filó[2]

Resumo: O presente artigo tem por objetivo pesquisar uma das lacunas encontradas na Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), que — se mal implementada em seu programa de conformidade — pode expor os dados de clientes e usuários. O problema de pesquisa se resume ao seguinte questionamento: há fragilidade na proteção de dados do cidadão, em razão dos mecanismos adotados pela legislação Brasileira? Num primeiro momento, tratou-se, brevemente, da anonimização de dados pessoais face à reidenticação, a fim de se poder tratar, posteriormente, da aplicação da privacidade diferencial. O método utilizado na abordagem foi o descritivo-sistemático. O método de interpretação jurídica foi o tópico sistemático. Verificou-se que é imperioso repensar os conceitos sobre segurança da informação de forma a transcender a mera obrigação legal, motivando de igual forma a inovação, a criatividade e a responsabilidade no tratamento de dados pessoais. Conclui-se, em linhas gerais, que a noção de segurança e de sigilo deve permear todas as atividades do tratamento de dados pessoais, desde a concepção de um produto ou serviço.

Palavras chave: Cidadania; Inovação; Governança Digital; PrivacybyDesing; Privacidade Diferencial.

El concepto de privacidad diferencial enrelaciónconlareidentificación de datospersonales

Resumen: Este artículo tiene como objetivo investigar una de las lagunas encontradas enlaLey General de Protección de Datos de Carácter Personal (Ley No 13.709/2018), que, si está mal implementada ensu programa de cumplimiento, puedeexponerlosdatos de clientes y usuarios. El problema de lainvestigación se reduce a lasiguiente pregunta: ¿existe fragilidadenlaprotección de losdatos de losciudadanos, debido a los mecanismos adoptados por lalegislaciónbrasileña? Al principio, se trataba brevemente de laanonimización de losdatospersonales frente a larehidratación, conelfin de poder tratar, más tarde, conlaaplicación de laprivacidad diferencial. El método utilizado enel enfoque fuedescriptivo-sistemático. El método de interpretación jurídica fueel tema sistemático. Se encontró que es imperativo repensar los conceptos de seguridad de lainformación para trascenderla mera obligación legal, motivando tambiénlainnovación, lacreatividad y

[1]Agenor Alexsander Carvalho Costa. Pós-graduando em Advocacia no Direito Digital e Proteção de Dados pela UNA/EBRADI, Pós-graduado em Direito e Processo do Trabalho pela Escola ESA-OAB/FUMEC. Advogado, pesquisador em Direito e Tecnologia do ITS Rio, Brasil. E-mail: alexsander.carvalho@itsrio.org - https://orcid.org/0000-0003-1440-0016

[2]Maurício da Cunha Savino Filó. Doutor em Direito pelo Programa de Pós-Graduação em Direito (PPGD) da Universidade Federal de Santa Catarina, professor da Universidade do Extremo Sul Catarinense (UNESC), Brasil. E-mail: mauriciosavino@hotmail.com - http://orcid.org/0000-0002-7436-1664

Recebido em 26/03/2020, aceito para publicação em 26/04/2020, disponibilizado online em 01/09/2020

pragMATIZES
Revista Latino Americana de Estudos em Cultura
www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

laresponsabilidadeneltratamiento de datospersonales. Se concluye, en general, que lanoción de seguridad y confidencialidaddebe impregnar todas lasactividadesdeltratamiento de datospersonales, desde laconcepción de unproducto o servicio.

Palabras clave: Ciudadanía; Innovación; Gobernanza digital; Privacidad por Desing; Privacidad diferencial.

The concept of differential privacy in relation to the reidentification of personal data

Abstract: The purpose of this article is to research one of the gaps found in the General Personal Data Protection Law (Law No. 13,709/2018), which - if poorly implemented in its compliance program - can expose customer and user data. The research problem boils down to the following question: is there a weakness in protecting citizens' data, due to the mechanisms adopted by Brazilian legislation? At first, it was briefly about the anonymization of personal data in view of their re-identification, in order to be able to deal, later, with the application of differential privacy. The method used in the approach was descriptive-systematic. The method of legal interpretation was the systematic topic. It was found that it is imperative to rethink the concepts of information security in order to transcend the mere legal obligation, equally motivating innovation, creativity and responsibility in the processing of personal data. It is concluded, in general, that the notion of security and confidentiality must permeate all activities of the processing of personal data, from the conception of a product or service.

Keywords: Citizenship; Digital Governance; Innovation; Privacy by Desing; Differential Privacy.

O conceito de privacidade diferencial em relação à reidentificação de dados pessoais

Introdução

Esta pesquisa possui como objetivo geral identificar possíveis falhas na proteção de dados pessoais, que violaria a intimidade e a vida privada, consagradas no inciso X, do art. 5º, da Constituição da República de 1988. A relevância deste artigo alberga-se na crescente preocupação global sobre a fragilidade dos mecanismos de proteção aos dados do cidadão.

Na Ordem Jurídica Brasileira, a Lei nº 13.709, de 14 de agosto de 2018, surgiu para dispor sobre a proteção de dados pessoais, sendo alterada pela Lei 13.853, de 08 de julho de 2019, que criou a Autoridade Nacional de Proteção de Dados. Destarte, configurou-se a Lei Geral de Proteção de Dados Pessoais (LGPD).

Apesar dos avanços legislativos, sabe-se que os mecanismos de proteção podem ser mal interpretados, tornando equivocada a implementação de seu programa de conformidade, expondo dados de clientes e usuários.

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

O problema de pesquisa está na seguinte pergunta: há fragilidade na proteção de dados do cidadão, em razão dos mecanismos adotados pela legislação Brasileira? Parte-se da premissa de que se devem repensar os conceitos adotados no Brasil sobre segurança da informação, indo além do mero legalismo jurídico, a fim de inovar no tratamento de dados pessoais.

Para poder responder ao problema, o artigo será desenvolvido em dois momentos distintos. Na primeira seção de desenvolvimento explicar-se-á, de forma sucinta, o que é a anonimização e reidentificação de dados pessoais. Estes conceitos são fundamentais para se realizar a segunda seção, que é verificar como ocorre a aplicação da privacidade diferencial.

O método eleito para esta abordagem será o descritivo-sistemático. O método de interpretação jurídica será o tópico sistemático.

1 - O conceito da anonimização e reidentificação de dados pessoais

A Lei Geral de Proteção de Dados inova ao criar um conjunto de novos conceitos jurídicos que estabelecem as condições nas quais os dados pessoais podem ser tratados, define um conjunto de direitos para os titulares dos dados, e gera novas obrigações específicas para os controladores dos dados ao criar uma série de procedimentos e normas para que haja maior cuidado no tratamento de dados pessoais e compartilhamento com terceiros.

No que diz respeito ao compartilhamento de dados pessoais com terceiros, a primeira recomendação a se fazer quando se pensa nas novas exigências trazidas pela legislação seria que: “bancos de dados que compartilham informações de consumidores devem informá-los previamente acerca da utilização desses dados, sob pena de terem que pagar indenização por danos morais”, conforme orienta a Terceira Turma do

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

Superior Tribunal de Justiça (STJ)[3].

Neste sentido, é preciso entender que será necessário coletar o consentimento inequívoco de todos os clientes já cadastrados em seu sistema de banco de dados, bem como revisar seus contratos atuais, a fim de coletar o consentimento de futuros clientes. Conforme expresso no art. 5º, XII, da Lei nº 13.709/2018 *"consentimento: manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma finalidade determinada"*.

No que diz respeito à forma de coleta deste consentimento, a lei não impõe uma forma em específico.À vista disto, poderá a coleta ser feita por intermédio do envio de e-mail com

[3] RECURSO ESPECIAL. FUNDAMENTO NÃO IMPUGNADO. SÚM. 283/STF. AÇÃO DE COMPENSAÇÃO DE DANO MORAL. BANCO DE DADOS. COMPARTILHAMENTO DE INFORMAÇÕES PESSOAIS. DEVER DE INFORMAÇÃO. VIOLAÇÃO. DANO MORAL IN RE IPSA. JULGAMENTO: CPC/15. 2. O propósito recursal é dizer sobre: (i) a ocorrência de inovação recursal nas razões da apelação interposta pelo recorrido; (ii) a caracterização do dano moral em decorrência da disponibilização/comercialização de dados pessoais do recorrido em banco de dados mantido pela recorrente. [...] 5. A gestão do banco de dados impõe a estrita observância das exigências contidas nas respectivas normas de regência – CDC e Lei 12.414/2011 – dentre as quais se **destaca o dever de informação**, que tem como uma de suas vertentes o dever de comunicar por escrito ao consumidor a abertura de cadastro, ficha, registro e dados pessoais e de consumo, quando não solicitada por ele. 6. **O consumidor tem o direito de tomar conhecimento de que informações a seu respeito estão sendo arquivadas/comercializadas por terceiro**, sem a sua autorização, porque desse direito decorrem outros dois que lhe são assegurados pelo ordenamento jurídico: o direito de acesso aos dados armazenados e o direito à retificação das informações incorretas. 7. A inobservância dos deveres associados ao tratamento (que inclui a coleta, o armazenamento e a transferência a terceiros) dos dados do consumidor – dentre os quais se inclui o dever de informar – faz nascer para este a pretensão de indenização pelos danos causados e a de fazer cessar, imediatamente, a ofensa aos direitos da personalidade. 8. Em se tratando de **compartilhamento das informações** do consumidor pelos bancos de dados, prática essa autorizada pela Lei 12.414/2011 em seus arts. 4º, III, e 9º, deve ser observado o disposto no art. 5º, V, da Lei 12.414/2011, o qual prevê o direito do cadastrado ser informado previamente sobre a identidade do gestor e sobre o armazenamento e o objetivo do tratamento dos dados pessoais. 9. O fato, por si só, de se tratarem de dados usualmente fornecidos pelos próprios consumidores quando da realização de qualquer compra no comércio, não afasta a responsabilidade do gestor do banco de dados, na medida em que, quando o consumidor o faz não está, implícita e automaticamente, autorizando o comerciante a divulgá-los no mercado; está apenas cumprindo as condições necessárias à concretização do respectivo negócio jurídico entabulado apenas entre as duas partes, confiando ao fornecedor a proteção de suas informações pessoais. 10. Do mesmo modo, o fato de alguém publicar em rede social uma informação de caráter pessoal não implica o consentimento, aos usuários que acessam o conteúdo, de utilização de seus dados para qualquer outra finalidade, ainda mais com fins lucrativos. 11. Hipótese em que se configura o dano moral *in reipsa*. (STJ - REsp nº 1758799 MG 2017/0006521-9, RELATOR(A): Min. NANCY ANDRIGHI - TERCEIRA TURMA, Data de Julgamento 11/12/2019 (12:20), T3 - TERCEIRA TURMA, Data de Publicação: DJe 19/11/2019 - grifos nossos)

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

objetivo da obtenção de assinaturas digitais face aos contratos já firmados, e a inclusão de cláusula específica de tratamento de dados para futuros contratos.

Bem como observar os mesmos cuidados no informar sobre o compartilhamento destes dados:

> Art. 7º, § 5º, da Lei nº 13.709/2018 - O controlador que obteve o consentimento referido no inciso I do caput deste artigo que necessitar comunicar ou **compartilhardados pessoais com outros controladores** deverá obter consentimento específico do titular para esse fim, ressalvadas as hipóteses de dispensa do consentimento previstas nesta Lei; (grifos nossos)

Entretanto, ainda é preciso se ponderar que o compartilhamento desses dados pessoais sempre deverá ser realizado de forma anonimizada. E aqui encontra-se a lacuna para o vazamento de dados, que, mesmo anonimizados podem ser recuperados face ao cruzamento de dados, realizados por diferentes plataformas.

Conforme Bioni(2019, p.71),

> Esse processo pode se valer de diferentes técnicas que buscam eliminar tais elementos identificadores de uma base de dados, variando entre: a) supressão; b) generalização; c) randomização e; d) pseudoanonimização.

Todavia, Bioni (2019, p.73) prossegue afirmando que “Torna-se cada vez mais recorrente a publicação de estudos que demonstram se o processo de anonimização algo falível”.

A recente pesquisa realizada por Kurtz (2020) ao IRIS - Instituto de Referência em Internet e Sociedade-, revela que ainda será preciso lidar com o conceito de reidentificação, pois, “dada uma série de informações aparentemente desconexas que, juntas, revelam situações e características, nos bancos de dados é possível combinar informações para descobrir algo sobre indivíduos singulares”.

Em igual sentido, Wodinsky (2020) alerta que Dasha Metropolitansky e Kian Attari, ambos pesquisadores de Harvard, já deram prova disso ao desenvolver uma ferramenta que faz uma varredura a partir de conjuntos de dados de consumidores que foram vazados na web.

> Metropolitansky e Attari explicaram à *Motherboard*[4] que o programa foi criado para juntar e ligar informações “não tão anônimas” – como e-mails e nomes de usuário – a dados “anônimos” que foram encontrados em bases de dados vazadas de praticamente mil domínios diferentes, indo desde a Adobe ao YouPorn. E

[4] Motherboard é uma revista online e canal de vídeo dedicado à tecnologia, ciência e seres humanos. (VICE, 2009)

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

> apesar desses conjuntos de dados serem "anonimizados", identificar alguém em um determinado vazamento **não é nada difícil**, segundo os pesquisadores. (grifos nossos)

Assim sendo, "o fato de os repositórios operarem com dados anonimizados, ou com aqueles que passaram por processos de pseudonimização, não é garantia nenhuma para a aplicação da LGPD" (SIMÕES, 2019).

Ante o exposto, além das formas de anonimização já apresentadas por Bioni, será preciso acrescentar a privacidade diferencial a este rol,visando com isso se evitar a possibilidade de compreensão destes dados em um futuro cruzamento de dados entre *Big Datas*[5], ou mesmo em razão de um vazamento de um *Data Broker*[6], mantendo a sua utilidade da estatística.

2 - O conceito da privacidade diferencial

Resumidamente, a privacidade diferencial é uma tecnologia que coloca diversos ruídos nas informações a ponto de tornar inviável o rastreamento da fonte por meio do cruzamento de informação entre bancos de dados relacionais. As informações são embaralhadas, como alterar a idade ou sexo da pessoa, mas sem prejudicar as informações que são importantes.

Segundo Chen (2020), "privacidade diferencial é uma técnica matemática que torna esse processo rigoroso, medindo o quanto a privacidade aumenta quando o ruído é adicionado". E mais, "o método já é usado pela Apple e pelo Facebook para coletar dados agregados sem identificar usuários específicos".

A seguir, será verificado como o conceito da privacidade diferencial vem sendo adotado por estas empresas em seus produtos e serviços, e o seu significado.

2.1 - Privacidade diferencial da Apple

Para a Apple, a privacidade de seus usuários sempre foi muito importante e, durante a WWDC 2016- *Apple Worldwide Developers*

[5] *Big Data* é a área do conhecimento que estuda como tratar, analisar e obter informações a partir de conjuntos de dados grandes demais para serem analisados por sistemas tradicionais.(Significados, 2020)

[6] Uma pessoa ou empresa cujo negócio é vender informações sobre empresas, mercados etc.(Cambridge Dicionary, tradução nossa)

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

Conference, o seu vice-presidente sênior de engenharia veio a público explicar abertamente como o método chamado de privacidade diferencialé usado pela empresa.

"Acreditamos que você deve ter ótimos recursos e muita privacidade[7]", disse Craig Federighi (*apud* GREENBERG, 2016, tradução nossa) à multidão de desenvolvedores presentes no evento.

> A privacidade diferencial é um tópico de pesquisa nas áreas de estatística e análise de dados que usa *hash*[8], sub-amostragem e injeção de ruído para permitir o aprendizado por meio de *crowdsourcing*[9], mantendo os dados de usuários individuais completamente privados. A Apple está realizando um trabalho super importante nesta área para permitir que a privacidade diferencial seja implantada em escala[10]. (FEDERIGHI, *apud* GREENBERG, 2016, tradução nossa)

Em uma explicação mais detalhada sobre o assunto:

> A privacidade diferencial, traduzida da linguagem Apple, é a ciência estatística de tentar aprender o máximo possível sobre um grupo enquanto aprende o mínimo possível sobre qualquer indivíduo nele. Com privacidade diferenciada, a Apple pode coletar e armazenar os dados de seus usuários em um formato que permite coletar noções úteis sobre o que as pessoas fazem, dizem, gostam e querem. Mas não pode extrair nada sobre uma única e específica daquelas pessoas que podem representar uma violação de privacidade. E nem, em teoria, hackers ou agências de inteligência[11].(GREENBERG, 2016, tradução nossa)

Aaron Roth, professor de ciência da computação da Universidade da Pensilvânia, a quem Federighi(*apud* GREENBERG, 2016, tradução nossa) citou em seu discurso como "escrevendo o livro sobre privacidade diferencial" afirma que,

> A privacidade diferencial permite que você obtenha insights[12] de grandes conjuntos de dados, mas com uma

[7] No original: "We believe you should have great features and great privacy".

[8]Hash é a transformação de uma grande quantidade de dados em uma pequena quantidade de informações.(Wikipédia, março 2020)

[9]Crowdsourcing, em português, significa contribuição colaborativa ou colaboração coletiva.(Wikipédia, janeiro 2020)

[10] No original*: "Differential privacy is a research topic in the areas of statistics and data analytics that uses hashing, subsampling and noise injection to enable...crowdsourced learning while keeping the data of individual users completely private. Apple has been doing some super-important work in this area to enable differential privacy to be deployed at scale."*

[11] No original: *"Differential privacy, translated from Apple-speak, is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any individual in it. With differential privacy, Apple can collect and store its users' data in a format that lets it glean useful notions about what people do, say, like and want. But it can't extract anything about a single, specific one of those people that might represent a privacy violation. Andneither, in theory, could hackers orintelligence agencies."*

[12]*Insight* é o entendimento de uma causa e efeito específicos dentro de um contexto específico.(Significados, 2014).

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

> prova matemática de que ninguém pode aprender sobre um único indivíduo[13].

Nesta esteira,

> Como Roth observa, quando se refere a uma "prova matemática", a privacidade diferencial não tenta apenas ofuscar ou "anonimizar" os dados dos usuários. Essa abordagem de anonimato, ele argumenta, tende a falhar. Em 2007, por exemplo, a Netflix lançou uma grande coleção de classificações de filmes de seus telespectadores como parte de uma competição para otimizar suas recomendações, removendo o nome das pessoas e outros detalhes de identificação e publicando apenas as classificações da Netflix. Mas os pesquisadores logo fizeram uma referência cruzada dos dados da Netflix com os dados públicos de revisão no IMDB[14] para comparar padrões semelhantes de recomendações entre os sites e adicionar nomes novamente ao banco de dados supostamente anônimo da Netflix[15].(GREENBERG, 2016, tradução nossa)

Significa dizer que, "Se você começar a remover o nome das pessoas dos dados, isso não impedirá que as pessoas façam referências cruzadas inteligentes[16]", diz Aaron Roth (*apud* GREENBERG, 2016, tradução nossa). Concluindo que, "Esse é o tipo de coisa que é comprovadamente impedida pela privacidade diferencial"[17].

Em contrapartida, Greenberg (2017, tradução nossa) nos revela que "a privacidade diferencial não é uma simples alternância entre privacidade total e invasão sem restrições"[18], e prossegue "um novo estudo, que investiga profundamente como a Apple realmente implementa a técnica, sugere que a empresa aumentou ainda mais essa discrepância em direção à mineração de dados agressiva do que suas promessas públicas implicam"[19].

> Pesquisadores da Universidade do Sul da Califórnia, da Universidade de Indiana e da Universidade de Tsinghua da China adotaram o código dos sistemas operacionais Mac OS e iOS da Apple para fazer

[13] No original: "*Differential privacy lets you gain insights from large datasets, but with a mathematical proof that no one can learn about a single individual."*

[14] O IMB é um website onde as pessoas compartilham suas impressões sobre filmes (BIONI, 2019, p.73).

[15] No original: *"As Roth notes when he refers to a "mathematical proof," differential privacy doesn't merely try to obfuscate or "anonymize" users' data. That anonymization approach, he argues, tends to fail. In 2007, for instance, Netflix released a large collection of its viewers' film ratings as part of a competition to optimize its recommendations, removing people's names and other identifying details and publishing only their Netflix ratings."*

[16] No original: *"If you start to remove people's names from data, it doesn't stop people from doing clever cross-referencing."*

[17] No original: *"That's the kind of thing that's provably prevented by differential privacy"*

[18] No original: *"[...] differential privacy isn't a simple toggle switch between total privacy and no-holds-barred invasiveness."*

[19] No original: *"[...] a new study, which delves deeply into how Apple actually implements the technique, suggests the company has ratcheted that dial further toward aggressive data-mining than its public promises imply".*

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

> engenharia reversa da maneira como os dispositivos da empresa implementam privacidade diferencial na prática. Eles examinaram como o software da Apple injeta ruídos aleatórios em informações pessoais - desde o uso de emojis até o histórico de navegação, dados do Health Kit e consultas de pesquisa - antes que o iPhone ou o MacBook carregue esses dados nos servidores da Apple.
> Idealmente, essa ofuscação ajuda a proteger seus dados privados de qualquer hacker ou agência governamental que acesse os bancos de dados da Apple, anunciantes que a Apple algum dia os venderão, ou até a própria equipe da Apple. Porém, a eficácia da privacidade diferencial depende de uma variável conhecida como "parâmetro de perda de privacidade", ou "epsilon", que determina quanta especificidade um coletor de dados está disposto a sacrificar para proteger os segredos de seus usuários.
> Ao desmontar o software da Apple para determinar o epsilon escolhido pela empresa, os pesquisadores descobriram que o MacOS carrega dados significativamente mais específicos do que o típico pesquisador diferencial de privacidade pode considerar privado.[20] (GREENBERG, 2017, tradução nossa)

Destarte, a privacidade diferencial é uma das 10 tecnologias mais promissoras de 2020, segundo lista do MIT - Massachusetts InstituteofTechnology.(KURTZ, 2020)

2.2 - Privacidade diferencial do Google e sua biblioteca open-source

A privacidade diferencial não é uma invenção exclusiva da Apple. Conforme Fagundes (2018), foram os pesquisadores da Berkeley - Universidade da Califórnia, que "desenvolveram a ferramenta de código aberto que limita o quanto os funcionários podem aprender sobre os clientes individuais, analisando os dados dos usuários". Seus estudos sobre a privacidade diferencial já ocorrem há anos e é resultado do trabalho de vários especialistas em segurança e privacidade.

Outro exemplo de como a privacidade diferencial não é uma exclusividade da Apple, seria citarmos o Google Fotos, cujo aplicativo permite

[20] No original: *"Researchers at the University of Southern California, Indiana University, and China's Tsinghua University have dug into the code of Apple's MacOS and iOS operating systems to reverse-engineer just how the company's devices implement differential privacy in practice. They've examined how Apple's software injects random noise into personal information—ranging from emoji usage to your browsing history to HealthKit data to search queries—before your iPhone or MacBook upload that data to Apple's servers. Ideally, that obfuscation helps protect your private data from any hacker or government agency that accesses Apple's databases, advertisers Apple might someday sell it to, or even Apple's own staff. But differential privacy's effectiveness depends on a variable known as the "privacy loss parameter," or "epsilon," which determines just how much specificity a data collector is willing to sacrifice for the sake of protecting its users' secrets. By taking apart Apple's software to determine the epsilon the company chose, the researchers found that MacOS uploads significantly more specific data than the typical differential privacy researcher might consider private."*

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

que os usuários armazenem na nuvem do Google um número ilimitado de fotos de celular. Encontraremos maiores informações sobre como a empresa utiliza a privacidade diferencial neste tipo de armazenamento nos Termos de uso do Google (2017), onde informa suas duas técnicas para proteger tais dados: "*Generalização dos dados"* e a "*Adição de ruídos aos dados"*.

Temos ainda o navegador Chrome, que, segundo Santana (2017)

> [...] utiliza um código aberto de privacidade diferencial — denominado RAPPOR (RandomizedAggregatablePrivacy-Preserving Ordinal Response) para anonimizar os dados e declara em alto e bom som que o seu sistema tem um coeficiente épsilon 2, com limite 8 ou 9 considerando toda a vida do usuário.

Lado outro, o mais interessante é que o Google tornou *open-source*[21]sua biblioteca de privacidade diferencial: "Nossa biblioteca de código aberto foi projetada para atender às necessidades dos desenvolvedores. Além de ser acessível gratuitamente, queríamos que ela fosse fácil de implantar e útil [...]" (GOOGLE, 2019).

2.2.1 - Vulnerabilidades e possíveis soluções para se mitigar riscos

Tudo isso pode trazer muitos benefícios significativos aos serviços já existentes, que só têm a se beneficiar com a biblioteca de código aberto desenvolvida pelo Google.

Por outro lado, segundo Santana (2017), a privacidade diferencial poderá não ser tão diferencial assim.É sempre bom ficarmos atentos, pois o Google poderia rastrear movimentos de pessoas por meio de seus presentes em celulares pelo mundo: "Quando o Google está ganhando de alguém em algum quesito relacionado à privacidade, sabemos que algo não está exatamente certo".

No mesmo sentido de Santana, Fagundes (2018) recorda que:

> Mesmo gigantes da computação, como o Facebook, entregaram dados de usuários que adequadamente manipulados podem ter influenciado as eleições americanas. Além do roubo de informações, algoritmos de *Machine Learning*[22] podem

[21]O software de código aberto é fabricado por muitas pessoas e distribuído sob uma licença compatível com OSD que concede todos os direitos de uso, estudo, alteração e compartilhamento do software na forma modificada e não modificada. A liberdade de software é essencial para permitir o desenvolvimento comunitário de software de código aberto.(Open SourceInitiative – tradução nossa)

[22]Machinelearning é uma área da ciência da computação que significa ¨aprendizado da máquina¨.(Significados, 2018)

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

> estabelecer relações entre dados e definir comportamentos de usuários, mesmo usando dados anonimizados.

Para além do exposto, Snowden (2019) durante a abertura do Web Summit 2019, realizado em Lisboa, afirmou que, “Dados não são inofensivos ou abstratos quando se trata do ser humano. Não são dados que estão sendo explorados. São pessoas que estão sendo exploradas”. Mais adiante, recorda a pergunta que se faz até hoje: “O que pode ser feito quando as instituições mais poderosas da sociedade se tornaram as menos responsáveis?”.

À vista disso, segundo Pinheiro (2018, p. 95),

> O grande paradigma não está no conceito ético ou mesmo filosófico se a privacidade deve ou não ser protegida (claro que deve ser), mas sim no modelo de negócios estabelecido, visto que a informação virou não apenas a riqueza do século XXI como também a moeda de pagamento.

Outrossim, segundo Fagundes (2018) "o uso combinado de Blockchain[23] e a inteligência artificial podem, se não resolver por completo, minimizar o risco de vazamento de informações privadas".

Portanto, a fim de mitigar riscos, estar sempre atento à inovação e às novas tecnologias será mais que uma exigência, mas também um diferencial competitivo. Afinal, Atheniense (2019, p. 22) recorda que “a LGPD quer estimular o desenvolvimento de uma cultura de proteção de dados”, pois esta lei, de forma alguma, põe fim aos dilemas atuais ou se esgota em si mesma; porém, ela serve de estímulo à inovação.

Ainda segundo Atheniense (2019, p.12), o "caminho da conformidade à LGPD não tem uma reta de chegada. É um processo continuado, portanto não espere encerrar a questão do dia para noite. Tão importante quanto estar em conforme é se manter conforme". Neste sentido, verifica-se por pontual:

> nenhuma empresa deve prescindir de algum nível de proteção envolvendo uso de softwares que propiciem maior controle sobre os riscos de segurança da informação para repelir invasões, antivírus e manutenção periódica de equipamentos e atualização de softwares. (ATHENIENSE, 2019, p.21)

Nada obstante, a regulamentação, aqui, não deve se mostrar como uma mera burocracia

[23] A blockchain é uma tecnologia de registro distribuído que visa a descentralização como medida de segurança.(Wikipédia, fevereiro 2020)

estatal, e sim como um meio efetivo de impedir que as grandes corporações tecnológicas continuem se valendo das "pegadas digitais" para guiarem o mundo a seu alvedrio.

Diante deste cenário, adverte Cabella(2018, p.75) que uma empresa brasileira, dependendo de sua atividade, "[...] pode ter um website[24] ou aplicativo para dispositivo móvel que é passível de ser acessado por uma pessoa natural localizada fisicamente em algum país membro da União Europeia".

Fato este que, por si só, já representa grande risco. Visto que, conforme previsão do "artigo 30 do GDPR, o valor da multa por infração pode chegar a 20 milhões de euros (aproximadamente 78 milhões de reais) ou 4% do faturamento global anual da empresa no ano anterior, o que for maior" (CABELLA, 2018, p. 77).

Já a LGPD, tem previsto no seu artigo 52 a aplicação de multa simples, de até 2% (dois por cento) do faturamento da pessoa jurídica de direito privado, grupo ou conglomerado no Brasil no seu último exercício, excluídos os tributos, limitada, no total, a R$ 50.000.000,00 (cinquenta milhões de reais) por infração, entre várias outras implicações.

O que faz da prevenção na adoção de uma maior segurança da informação e incentivos à inovação um grande negócio. Segundo Pinheiro (2016a, p. 98), legislar sobre a privacidade e sobre a proteção de dados impacta a economia digital, sob o risco de afetar a forma de como a própria internet vem se desenvolvendo: "A segurança da informação sempre encontrou uma barreira natural na privacidade." Sendo "evidente que o direito à privacidade constitui um limite natural ao direito à informação". (PINHEIRO, 2016b, p. 481)

3 - Considerações finais

Como se observa, no que tange à parte técnica, o compartilhamento seguro e responsável de dados pessoais — mesmo que anonimizados — será uma tarefa mais complexa que o mero consentimento previsto em Lei Geral de Proteção de Dados Pessoais.

Dito isso, para se pensar em uma efetiva conformidade em privacidade e a proteção de dados do

[24] Website é uma palavra que resulta da justaposição das palavras inglesas web (rede) e site (sítio, lugar).(Significados, 2011)

cidadão, e para além do que agora está previsto em lei, será preciso investir em novas tecnologias, inovação e constante atualização dos seus colaboradores.

Simplesmente anonimizar dados pessoais na busca por isenção de responsabilidade, apenas por que a lei assim o permite não é estar em *compliance*[25], visto que, conforme Wodinsky (2020) a Lei Geral de Proteção de Dados Pessoais deixa em aberto esta lacuna àqueles "que recorrem a essa linha de pensamento para manter a consciência limpa, sabendo que a coleta de dados estaria de acordo com as regras".

Neste contexto, a observância à Governança Digital apresenta-se como essencial para gerenciar os riscos que a Tecnologia da Informação oferece ao negócio e garantir segurança para a empresa. Pois ela atua como um mecanismo de controle que assegura a proteção das informações, diminui custos, otimiza recursos e dá suporte à tomada de decisões.

Ao se pensar na privacidade devemos estar sempre atentos à metodologia Privacy by Desing, que, segundo Bruno Bioni (2019, p.176) é "a ideia de que a proteção de dados pessoais deve orientar a concepção de um produto ou serviços, devendo eles ser embarcados com tecnologias que facilitem o controle e a proteção das informações pessoais". De mais a mais, conforme Souza (2019, p. 418), "[...] segurança e sigilo aparecem como dois elementos incontornáveis para que se possa afirmar que dados (pessoais) são protegidos. Dito de outra forma, não existe proteção de dados sem segurança e sigilo dos mesmos".

Antecipa a lei:

> Art. 12, da Lei nº 13.709/2018 - Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, **salvo quando o processo de anonimização ao qual foram submetidos for revertido**, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido. (grifos nossos)

Para além da Lei Geral de Proteção de Dados Pessoais, devemos igualmente nos ater ainda ao Marco Civil da Internet (Lei n° 12.965/2014), onde se encontra prevista, nos artigos 13 e 15, a determinação expressa de que o regime de guarda dos dados armazenados pelos provedores de conexão e pelos provedores de

[25]Compliance é o dever de estar em conformidade com atos, normas e leis, para seu efetivo cumprimento. (BOBSIN, 2019)

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

aplicação deverá atender a parâmetros de controle e segurança:

> Art. 13 - Na provisão de conexão à internet, cabe ao administrador de sistema autônomo respectivo o dever de manter os registros de conexão, **sob sigilo**, em **ambiente controlado e de segurança**, pelo prazo de 1 (um) ano, nos termos do regulamento.
> Art. 15 - O provedor de aplicações de internet constituído na forma de pessoa jurídica e que exerça essa atividade de forma organizada, profissionalmente e com fins econômicos deverá manter os respectivos registros de acesso a aplicações de internet, **sob sigilo**, em **ambiente controlado e de segurança**, pelo prazo de 6 (seis) meses, nos termos do regulamento.
> (grifos nossos)

Como se observa, a segurança no armazenamento de dados não é algo que possa simplesmente ser esquivado por conta meramente do cumprimento de determinação legal, visto que a própria lei deixa em aberto quais os métodos a serem adotados para a segurança, mas é pontual no que determina que tais dados devem ser armazenados em ambiente controlado e de segurança.

Souza (2019, p. 421) defende a importância de que a consciência das pessoas se desperte para a relevância dos seus dados pessoais, a fim de que a segurança e o sigilo acerca deles aumentem. Parece haver a necessidade de se saber que o tratamento de dados pessoais é do interesse de todos. Esta compreensão encontra guarida em Olivo (2016, p. 180-253), para quem o sentido contemporâneo de direito à privacidade passou a necessitar de uma perspectiva e proteção coletivas.

Conclusão

Verificou-se que um dos muitos paradoxos do futuro tecnológico envolve, cada vez mais, o campo da privacidade pessoal. Toda a pesquisa leva à conclusão de que os dados pessoais, por mais anônimos que as empresas digam que estejam, não estão protegidos corretamente.

Cabe à legislação a regulamentação adequada, com o fim de garantir condições para a implementação de uma eficaz Proteção de Dados Pessoais. Entretanto, os avanços legislativos são inócuos na proteção da intimidade e da privacidade das pessoas. Testemunham este entender as mais recentes pesquisas ao que denunciam que 85% das empresas nacionais não estão preparadas para a vigência da Lei Geral de Proteção de Dados Pessoais.

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

A atuação cidadã se inicia, primeiro, no reconhecimento do problema do tratamento dado aos dados pessoais, para – posteriormente – reivindicar e pensar em uma maior responsabilização de todos os envolvidos no manuseio destes dados.

E o mais preocupante é que tudo isso tem sido feito a partir de uma série de dados que podem parecer irrelevantes para o cidadão comum, e este é o ponto de maior vulnerabilidade: quando não há preocupação em ler e entender as diretivas de tratamento de dados pessoais no ato de contratação de um serviço, qual tecnologia será aplicada por determinada empresa ou mesmo qual aplicativo será utilizado para o referido serviço: por *default*, geralmente, é costume apensar aceitar os termos de uso sem os ter lido devidamente.

Revela-se importante um debate amplo, aberto, sincero e democrático a respeito do avanço da tecnologia. Não se mostra viável, de modo algum, que as pessoas continuem sem respostas sobre o que é feito com seus dados pessoais, tendo seus dados utilizados como moeda de troca nas plataformas digitais, comsua intimidade violada de maneira brutal, sem que nada possam fazer em relação a isso.

A solução para a maioria desses problemas enfrentados na sociedade da informação passa pela Educação Digital, que possui o viés de atingir um maior número de pessoas, visto serpelo estudo que se leva ao conhecimento de si próprio como sujeito de direitos.

Referências bibliográficas:

ATHENIENSE, Alexandre Rodrigues. *Lei Geral de Proteção de Dados nos escritórios de advocacia.* São Paulo: Deepcontent, 2019. p.12, 21, 22

BIONI, Bruno Ricardo. Dados 'anônimos' como a antítese de dados pessoais: filtro da razoabilidade. In: *Proteção de Dados Pessoais* - A Função e os Limites do Consentimento. Rio de Janeiro: Editora Forense, 2019. p.71, 73, 176.

BOBSIN, Arthur. 2019. *Entenda o que é compliance e como colocar em prática.* (2019). Disponível em: https://www.aurum.com.br/blog/o-que-e-compliance/ Acesso: 30 mar. 2020.

BRASIL. *Marco Civil da Internet*(2014). 25ª ed. São Paulo: Vade Mecum Tradicional; Editora Saraiva, 2018. p.2217-2218.

BRASIL. *Lei Geral de Proteção de Dados Pessoais (LGPD).* Capítulo VIII, Da Fiscalização Seção, Das Sanções Administrativas. Disponível em: www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso: 20 mar. 2020.

CABELLA, Daniela Motta Monte Serrat. A GDPR pode puxar o tapete da sua empresa: saiba como se prevenir**.** In: *Direito Digital Aplicado 3.0*, São Paulo: Thomson Reuters, 2018. p.57,77.

CABELLA, Daniela Motta Monte Serrat. *Cambridge Dicionary*. Data broker**.** Disponível em: https://dictionary.cambridge.org/pt/dicionario/ingles/data-broker. Acesso: 30 mar. 2020.

CHEN, Angela. Differentialprivacy. In:*10 Breakthrough Technologies 2020*. Disponível em: https://www.technologyreview.com/lists/technologies/2020/ Acesso: 15 mar. 2020.

EUROPA. *General Data Protection Regulation (GDPR).* Capítulo IV. Controllerand processor. Disponível em: https://gdpr-info.eu/chapter-4/ Acesso: 20 mar. 2020.

FAGUNDES, Eduardo. *Como Blockchain e IA podem proteger sua privacidade.* Disponível em: https://efagundes.com/blog/uso-de-blockchain-e-inteligencia-artificial-para-proteger-dados-pessoais/. Acesso: 15 mar. 2020.

FAGUNDES, Eduardo. GOOGLE, 2017. *Privacidade & Termos: Como o Google Anonimiza os dados***.** Disponível em: https://policies.google.com/technologies/anonymization?hl=pt-BR. Acesso: 17 mar. 2020.

FAGUNDES, Eduardo. GOOGLE, 2019.*Permitindo que desenvolvedores e organizações usem privacidade diferencial***.** Disponível em: https://brasil.googleblog.com/2019/09/permitindo-que-desenvolvedores-e-organizacoes-usem-privacidade-diferencial.html. Acesso: 25 mar. 2020.

FAGUNDES, Eduardo. GOOGLE, 2020.*Differential-privacy***.** Disponível em: https://github.com/google/differential-privacy/tree/master/differential_privacy. Acesso: 17 mar. 2020.

GREENBERG, Andy. *Apple's 'Differential Privacy' Is About Collecting Your Data—But Not Your Data***.** **(**,2016). Disponível em: https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/. Acesso: 16 mar. 2020.

GREENBERG, Andy. *How One of Apple's Key Privacy Safeguards Falls Short*(2017).Disponível em: https://www.wired.com/story/apple-differential-privacy-shortcomings/. Acesso: 26 mar. 2020.

KURTZ, Lahis. *Privacidade diferencial: o ruído anonimizador.* Disponível em: http://irisbh.com.br/privacidade-diferencial-o-ruido-anonimizador/. Acesso: 2 mar. 2020.

OLIVO, Mikhail Vieira Cancelier de. *Infinito particular: privacidade no século XXI e a manutenção do direito de estar só***.** (Doutorado em Direito). Universidade Federal de Santa Catarina, 2016. Disponível em: https://repositorio.ufsc.br/xmlui/handle/123456789/174424. Acesso: 20 mar. 2020.

OLIVO, Mikhail Vieira Cancelier de. *Open Source Initiative*. To promote and protect open source software and communities...Disponível em: https://opensource.org/ aberto. Acesso: 30 mar. 2020.

PINHEIRO, Patrícia Peck. Privacidade e anonimato**.** In: *Direito Digital*. 6. ed. rev., atual e ampl. – São Paulo: Saraiva, 2016a. p. 95, 98.

PINHEIRO, Patrícia Peck. Proteção de dados pessoais. In: *Direito Digital.* 6.

ed. rev., atual e ampl. – São Paulo: Saraiva, 2016b. p.481.

SANTANA, Bruno. *A privacidade diferencial da Apple pode não ser tão diferencial assim, como mostra este estudo.* Disponível em: https://macmagazine.uol.com.br/post/2017/09/18/a-privacidade-diferencial-da-apple-pode-nao-ser-tao-diferencial-assim-como-mostra-este-estudo/. Acesso: 17 mar. 2020.

SANTANA, Bruno. Big data. In: *Significados*, (março 2020). Disponível em: https://www.significados.com.br/big-data/. Acesso: 17 mar. 2020.

SANTANA, Bruno. Insight. In: *Significados.* (2014)..Disponível em: https://www.significados.com.br/insight/. Acesso: 17 mar. 2020.

SANTANA, Bruno. Machinelearning. In: *Significados.* (2018).. Disponível em: https://www.significados.com.br/machine-learning/. Acesso: 15 mar. 2020.

SANTANA, Bruno. Website. In: *Significados*. (2011). Disponível em: https://www.significados.com.br/website/. Acesso: 16 mar. 2020.

SIMÕES, Moisés.*Anonimização e pseudonimização são o suficiente?* Disponível em: https://www.serpro.gov.br/lgpd/noticias/2019/anonimizacao-pseudonimizacao-dados-suficientes-adequar-lgpd. Acesso: 10 fev. 2020.

SNOWDEN, Edward. *Edward Snowden: 'Não são dados sendo explorados, são pessoas'.* Disponível em: https://epocanegocios.globo.com/Web-Summit/noticia/2019/11/edward-snowden-nao-sao-dados-sendo-explorados-sao-pessoas.html. Acesso: 13 out. 2019.

SOUZA, Carlos Affonso Pereira de. Blockchain. In: *Wikipédia (*fevereiro 2020). Disponível em: https://pt.wikipedia.org/wiki/Blockchain. Acesso: 16 mar. 2020.

SOUZA, Carlos Affonso Pereira de. Crowdsourcing. In: *Wikipédia (*janeiro 2020). Disponível em: https://pt.wikipedia.org/wiki/Crowdsourcing. Acesso: 26 mar. 2020.

SOUZA, Carlos Affonso Pereira de. Hash. In: *Wikipédia. (*março 2020). Disponível em: https://pt.wikipedia.org/wiki/Fun%C3%A7%C3%A3o_hash. Acesso: 26 mar. 2020.

SOUZA, Carlos Affonso Pereira de. Segurança e Sigilo dos Dados Pessoais: primeiras impressões à luz da Lei 13.709/2018**.** In: *Lei Geral de Proteção de dados Pessoais e suas Repercussões no Direito Brasileiro*. São Paulo: Thomson Reuters Brasil, 2019. p.418, 421.

SOUZA, Carlos Affonso Pereira de. STJ. Recurso Especial REsp Nº 1758799 / MG (2017/0006521-9) Relator(a): Min. Nancy Andrighi - Terceira Turma, Data de Julgamento 11/12/2019. Disponível em: https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1888267&num_registro=201700065219&data=20191119&formato=PDF. Acesso: 25 mar. 2020.

SOUZA, Carlos Affonso Pereira de.. Compartilhamento de informações de banco de dados exige notificação prévia ao consumidor**.***STJ. Notícias*. Disponível em: http://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/Compartilhamento-de-informacoes-de-banco-de-dados-exige-notificacao-previa-ao-consumidor.aspx Acesso: 04/03/2020

www.periodicos.uff.br/pragmatizes - ISSN 2237-1508
(Dossiê Cultura, Tecnologia e Sociedade)

TANG, J.; KOROLOVA, A.; BAI, X.; WANG, X.; WANG, X. *Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12***.** Disponível em: https://arxiv.org/pdf/1709.02753.pdf. Acesso: 26 mar. 2020.

WODINSKY, Shoshana. *Dados anônimos não ajudam a realmente proteger identidades***.** Disponível em: https://gizmodo.uol.com.br/dados-anonimos-nao-protegem-identidades/. Acesso: 10 fev. 2020.