Universidade de Brasília – UnB
Faculdade de Direito- FD
Programa de Pós-Graduação em Direito

Fernanda Potiguara Carvalho

O SER ATRÁS DO DADO:
LIMITES E DESAFIOS DA ANONIMIZAÇÃO E SEUS REFLEXOS NOS REQUISITOS ESTABELECIDOS PELA LGPD

Brasília – DF
2021

FERNANDA POTIGUARA CARVALHO

O SER ATRÁS DO DADO: LIMITES E DESAFIOS DA ANONIMIZAÇÃO E SEUS REFLEXOS NOS REQUISITOS ESTABELECIDOS PELA LGPD

Dissertação apresentada ao Programa de Pós-Graduação da Faculdade de Direito da Universidade de Brasília, como requisito parcial à obtenção do título de Mestra em "Direito, Estado e Constituição", na linha de pesquisa "Transformações da Ordem Social e Econômica e Regulação".

Orientadora: Professora Doutora Ana Frazão

Brasília – DF
2021

Fernanda Potiguara Carvalho

O SER ATRÁS DO DADO:
LIMITES E DESAFIOS DA ANONIMIZAÇÃO E SEUS REFLEXOS NOS REQUISITOS ESTABELECIDOS PELA LGPD

Dissertação apresentada como requisito parcial à obtenção do título de Mestra, no Programa de Pós-Graduação da Faculdade de Direito da Universidade de Brasília, mestrado em "Direito, Estado e Constituição", linha de pesquisa "Transformações da Ordem Social e Econômica e Regulação".

BANCA EXAMINADORA

_____

Professora Doutora Ana de Oliveira Frazao Vieira de Mello
Faculdade de Direito da Universidade de Brasília (UnB) – Orientadora

_____

Professor Doutor Danilo Cesar Maganhoto Doneda
Faculdade de Direito do Instituto Brasiliense de Direito Público (IDP)– Membro Externo

_____

Professor Doutor Fabiano Hartmann Peixoto
Faculdade de Direito da Universidade de Brasília (UnB) – Membro

_____

Professora Doutora Edna Dias Canedo
Departamento de Ciência da Computação da Universidade de Brasília (UnB) – Suplente

AGRADECIMENTOS

A Deus, todo o meu entendimento, todo o trabalho de minhas mãos, com uma imensa gratidão por tanto. Aos meus pais Denise e Sávio, por serem meu suporte diário e meu refúgio em todo o momento. Aos meus irmãos, que me abriram o novo horizonte dos dados, e tiveram toda a paciência com as inúmeras perguntas, às vezes repetidas dezenas de vezes. À minha querida orientadora, que, muito mais do que minha referência acadêmica de excelência, foi uma grande incentivadora na exploração de áreas antes nunca trilhadas por mim no Direito e em todos os aspectos desta pesquisa tão desafiadora. Eu sempre serei grata por todo amparo, carinho, paciência e pela liberdade que recebi de suas mãos. Sem eles esse trabalho não teria o mesmo valor acadêmico e emocional para mim. Aos professores Veronese, Edna e Fabiano Hartmann por me acolherem tão docemente em seus grupos de estudos e pesquisas, e seguirem comigo essa trajetória de tanta aprendizagem. Ao Thiago Sampaio, por todo o incentivo e por tudo o que tem me ensinado sobre a vida e o amor. Aos meus amigos, companheiros de pesquisa, Amanda Espiñera, Cláudio Marcelo, Luana Lund, Mariana Moutinho, Taynara Tiemi, Thales Alessandro. A todos os colegas e professores da Universidade de Brasília, minha segunda casa.

SOLI DEO GLORIA

RESUMO

A sociedade informacional que vivemos atribui valor aos dados, não apenas como instrumentos de poder decisório, mas como verdadeiros ativos da chamada *Data Driven Economy*. No cenário em que dados ganham valor de mercado, eclodem também reivindicações por maior proteção desses ativos, principalmente no que tange aos dados pessoais tendo em vista os riscos que sua utilização pelos mercados causa à privacidade dos indivíduos. Nesse contexto, a anonimização surge como uma importante ferramenta de conciliação de interesses divergentes, em prol da garantia de maior segurança e privacidade na extração de valor dos dados. Este trabalho busca analisar se a Lei Geral de Proteção de Dados Brasileira, a LGPD, e os diversos atores que lidam com dados no país têm depositado excessiva confiança nas técnicas de anonimização, sem atenção às limitações apresentadas por esse tipo de tratamento. Para viabilizar essa análise dividimos o trabalho em três principais partes. A **primeira parte** introduz conceitos como dados, *Big Data*, Anonimização. Também apresentamos um breve relato sobre como as legislações em âmbito internacional e nacional vem lidando com os dados, em especial com dados pessoais, ressaltando a confiança que é depositada atualmente nas técnicas de anonimização. A **segunda parte** expõe as principais limitações da ferramenta, com base nas críticas erigidas por Paul Ohm em seu artigo "*Broken promises of privacy: responding to the surprising failure of anonymization*". Realizamos uma classificação dos limites da anonimização separando-os em intrínsecos e extrínsecos à técnica além dos limites externos à anonimização, que são limites não inseridos no processamento da técnica, mas que afetam o uso ético da anonimização. Por fim, na **terceira parte** do trabalho elaboramos um *framework* autoral a partir da Lei Geral de Proteção de Dados- LGPD, apresentando os requisitos exigidos pela lei para se considerar um dado como anonimizado. Comparamos os limites levantados na segunda parte do trabalho com os requisitos legais apresentados no *framework* a fim de constatar se a legislação acolheu as principais limitações da técnica estabelecendo diretrizes jurídicas a fim de mitigar riscos. A pesquisa conclui que a LGPD levou em consideração poucos aspectos relacionados aos limites da técnica. Falta à lei uma maior especificação sobre seus conceitos e procedimentos para conferir segurança jurídica aos desenvolvedores e maior proteção à anonimização. A lei também é lacunosa no tocante ao estabelecimento de parâmetros mais objetivos de aferição dos requisitos legais e no estabelecimento de padrões de governança e padrões éticos ao uso dos dados anonimizados. Esses aspectos determinam a consolidação de uma estrutura normativa mais frágil para o tratamento de anonimização.

Palavras-Chave: Anonimização. Big Data. LGPD. Elicitação de Requisitos. Framework Legal. Governança de Dados

ABSTRACT

We live in the information society in which the value of data changes from instruments of decision power to being also active in the so-called Data-Driven Economy. In the scenario in which data gains market value, demands for greater protection of these assets also arise, mainly concerning personal data, given the risks that their use by the markets cause to the privacy of individuals. In this context, anonymization emerges as an essential tool to reconcile divergent interests to guarantee greater security and privacy in extracting value from data. This paper seeks to analyze whether the Brazilian General Data Protection Law, the LGPD, and the various stakeholders, have placed excessive trust in anonymization techniques without regard to the limitations presented by this type of treatment. To make this analysis feasible, we divided the work into three main parts. The **first part** introduces concepts such as data, Big Data and Anonymization. We also present a history of national and international legislation that deals with data, especially personal data, highlighting the confidence currently placed in anonymization techniques. The **second part** exposes the main limitations of the tool, based on reviews raised by Paul Ohm in his article "Broken promises of privacy: responding to the surprising failure of anonymization." Finally, we built the classification of the limits of anonymization, separating them into intrinsic and extrinsic to the technique and into the external limits to anonymization, which are limits not inserted in the technique's processing but affect the ethical use of anonymity. Finally, in the **third part** of the work, we elaborated an authorial framework based on the Brazilian General Data Protection Law - LGPD, presenting the requirements demanded by the law to consider data as anonymized. We compared the limits raised in the second part of the work with the legal requirements presented in the framework to verify if the legislation accepted the main limitations of the technique, establishing legal guidelines to mitigate risks.The research concludes that the LGPD took into account a few aspects related to the limits of the technique. The law lacks further specification on some of its concepts and procedures to provide developers with greater legal certainty and greater protection from anonymity. The law is also lacking in establishing more objective parameters for assessing legal requirements and establishing governance standards and ethical standards for the use of anonymized data. These aspects determine the consolidation of a more fragile normative structure for the treatment of anonymization.

Keywords: Anonymization. Big Data. LGPD. Requirements Elicitation. Legal Framework. Data Governance.

SUMÁRIO

1. Introdução

Vivemos um novo paradigma social, onde as informações correm sem se deter em fronteiras geográficas, e ganham cada vez mais relevância, já que moldam não só aspectos da vida cultural, mas diversas outras esferas da vida social, em especial da economia. Castells defende que, pela importância que as informações assumem nesse paradigma, nós poderíamos classificar nossa sociedade como informacional (CASTELLS, 1999, p.119).

Ressaltando também a singularidade desse novo paradigma, Schwab aponta que vivenciamos uma "quarta revolução industrial" (SCHWAB, 2016)[1]. Schwab aponta três importantes características dessa quarta revolução que impactaram drasticamente a sociedade contemporânea. São elas: 1) a alta velocidade das mudanças; 2) as alterações sistêmicas de várias esferas da sociedade causadas pela alta velocidade; 3) a mudança no paradigma social[2], já que ele passa a ser informacional, como aponta Castells (SCHWAB, 2016, p. 13).

Portanto, a sociedade informacional é caracterizada pela alta velocidade de mudanças e pela constante necessidade de conexão. Temos a nosso dispor uma gama variada de serviços personalizados, disponíveis 24 horas por dia, o que seria inimaginável há poucos anos atrás. Isso não seria possível se não houvesse a ressignificação e revaloração dos dados e de seus fluxos nessa sociedade, fazendo com que mais informações se tornassem disponíveis.

Danilo Doneda diferencia os conceitos de dados e informações, ressaltando que "(...) o dado estaria associado a uma espécie de "pré-informação", anterior à interpretação e ao processo de elaboração. A informação, por sua vez, alude a algo além da representação contida no dado, chegando ao limiar da cognição, e mesmo nos efeitos que esta pode apresentar para o seu receptor" (DONEDA, 2011, pg. 94). [3]

Segundo aponta um estudo realizado pela Mckinsey em 2016, os fluxos globais de dados, medidos em *terabits* por segundo, vinham se elevando por um fator de 45 a cada década desde 2005, e, no final de 2016, o volume de fluxos chegou a cerca de 400 *terabits* por segundo

---

[1] Para o autor, as revoluções seriam marcadas pelos seguintes fatores: 1) a primeira revolução industrial, pelo maquinário movido à vapor; 2) a segunda, pela eletricidade e produção em massa; 3) a terceira, denominada revolução digital, pelos computadores pessoais e pela internet; 4) a quarta, pela universalização da internet, pelo trânsito intenso de dados e pelo profundo enraizamento de tecnologias no cotidiano, com impactos na sociedade e na economia global (SCHWAB, 2016, p. 15).

2 Segundo Schwab, "a revolução não está modificando apenas 'o que' e o 'como' fazemos as coisas, mas também 'quem' somos" (SCHWAB, 2016, p. 13).

3 Por sua vez, a Lei nº 12.527/2011 (Lei de Acesso à informação) prevê, em seu art. 4º, que: Para os efeitos desta Lei, considera-se: I - informação: dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato.

(BUGHIN & LUND, 2017). Enquanto isso, os fluxos tradicionais de valor de bens e serviços físicos mal conseguiram crescer no ritmo do PIB nominal mundial (BUGHIN & LUND, 2017).

Entre os anos de 2016 a 2020, houve uma diminuição do fator de crescimento, que ainda assim continuaram expressivos já que estudos indicaram uma taxa de aumento anual de 30% entre 2016 e 2020 (BRODSKY, 2020).

No entanto, mais recentemente, o estudo DHL Global Connectedness Index 2020, apontou uma excepcional retomada do quadro inicial. Em 2020, houve um aumento significativo dos fluxos de dados internacional associado a uma redução dos fluxos de bens e serviços tradicionais devido às limitações trazidas pelas medidas de contenção do Covid-19 (ALTMAN & BASTIAN, 2020, p. 40).

Para se ter uma ideia do impacto da pandemia nestes fluxos, de acordo com a pesquisa Telegeográfica analisada por Paul Brodsky, o tráfego internacional da Internet aumentou 48% em meados de 2019 até meados de 2020 (BRODSKY, 2020). Na data do estudo, ou seja, em meados de 2020, os fluxos chegavam a 618 *terabits* por segundo.

É possível, portanto, se vislumbrar a importância dos fluxos de dados nessa sociedade informacional, que é capaz de extrair valor desses ativos.

O acesso à internet é um dos fatores que tornou possível a transmissão de dados e informações, circulando notícias em tempo quase real e dando maior força ao movimento de globalização. Também viabilizou o acesso e armazenamento de grandes quantidades de dados, que possibilitam a extração das mais diversas informações.

Nesse contexto, os dados podem servir aos mais variados motivos. Podem viabilizar a transparência, tornar mais efetivas as políticas públicas, proporcionar o fornecimento de produtos personalizados, identificar fragilidades sociais, aprimorar ferramentas de gestão, possibilitar a criação de ferramentas de orientação da população, até mesmo auxiliar no diagnóstico de doenças e no seu tratamento.

Os dados ganham, portanto, a importância evidente para a sociedade informacional e as pessoas vivem diariamente o dilema da exposição de seus dados. Isso porque, se por um lado a disponibilização desses ativos torna a vida das pessoas muito mais fácil, com o acesso à informação, a serviços e produtos cada vez mais personalizados, por outro cria riscos para os cidadãos à medida que expõe informações sobre os indivíduos de forma nem sempre desejável. O risco é particularmente iminente com as sofisticadas ferramentas de extração de informações diversas das bases de dados, como trataremos posteriormente.

O consumidor do século XXI conhece as vantagens inúmeras do acesso a informações e serviços sem pausas e de baixo custo; mas começa a perceber também os riscos

do tráfego de seus próprios dados nas redes, e os reflexos que essa disponibilização pode causar em sua privacidade e em sua realidade presente e futura.

Na tentativa de conciliar esses dois cenários em contraposição, ou seja, a liberdade no uso e nos fluxos desses ativos tão importantes para a sociedade informacional e a privacidade dos titulares, as legislações se apropriam de ferramentas de proteção de dados como a anonimização, que será objeto de estudo neste trabalho.

Dessa forma, voltamos as atenções nesta exposição para os riscos na utilização dos dados dos cidadãos e em uma das mais eminentes soluções que os mercados e governos têm investido confiança: as técnicas de anonimização.

O enfoque é a anonimização como tratamento de bancos de dados, principalmente em contextos de *Big Data*, expondo os limites da ferramenta em contraposição à confiança ainda nela depositada pelos diversos agentes.

Temos por problema de pesquisa a seguinte questão:

Os requisitos definidos pela Lei Geral de Proteção de Dados Brasileira para anonimização, elencados no framework, levam em consideração os principais limites da técnica?

Para a resolução deste problema de pesquisa, esse trabalho foi dividido em quatro partes distintas, que buscam dar o embasamento contextual e teórico da questão.

A **primeira parte** pretende esclarecer ao leitor a realidade acerca dos bancos de dados formados na sociedade em rede, fornecendo conceitos como o que é o *Big Data*, como foi possível o armazenamento em massa de dados, como as legislações e o Direito lidaram ao longo do tempo com os dados, até chegarmos na anoninimização e sua importância nesse contexto. Desta forma, buscamos apresentar como a legislação internacional tem se voltado à proteção dos dados a partir de ondas legislativas que foram construindo a arquitetura jurídica dos dados. Além disso, apresentamos também a evolução da legislação brasileira quanto à matéria, demonstrando como foi possível, pela política de transparência, a publicização de dados governamentais. Ressaltamos ainda as dificuldades e emblemas na unificação das bases de dados estatais no Brasil e o despertar nacional para a proteção de dados. Também introduzimos o conceito de anonimização e apresentamos a importância da ferramenta e a confiança que tem sido depositada nesta técnica pelos diversos atores.

A **segunda parte** do trabalho se dedica a apresentar o que a literatura aponta como principais limites técnicos da ferramenta de anonimização. Para isso, utilizamos como marco teórico o interessante artigo produzido pelo professor Paul Ohm "*Broken promises of privacy: responding to the surprising failure of anonymization*", que ainda em 2010 já apresentava

pontos importantes de preocupação acerca da confiabilidade depositada neste tipo de tratamento de dados. A partir da abordagem de Ohm, apresentamos a classificação de limites intrínsecos e extrínsecos às técnicas de anonimização. Além dos aspectos puramente técnicos, também apresentamos as situações desafiadoras do ponto de vista ético na utilização desses dados, o que denominamos limites externos à técnica.

Por fim, na **terceira parte** do trabalho expomos um framework autoral desenhado a partir da Lei Geral de Proteção de Dados- LGPD, com o objetivo de descrever e esquematizar o passo a passo estabelecido pela Lei para considerar como regular o tratamento de anonimização. Neste tópico, abordamos os requisitos legais que devem preexistir ao tratamento, coexistir e proceder o tratamento. Apontamos ainda a peculiaridade da anonimização como um tratamento de dados, já que, de acordo com nossa legislação, a anonimização se apresenta como um tratamento em constante adaptação, sujeito a periódicas atualizações de acordo com a evolução da técnica. Realizamos ainda a análise do *framework* proposto, a partir dos limites da anonimização apresentados na segunda parte da pesquisa.

Nesta última parte do trabalho, o problema de pesquisa é efetivamente analisado, já que são dispostos em comparação as etapas estabelecidas pela Lei e esquematizadas no *framework* por um lado, e os obstáculos apresentados pela academia e classificados como limites intrínsecos, extrínsecos ou externos à técnica.

Buscamos analisar se a LGPD dispôs a anonimização de maneira exaustiva, observando os limites da técnica e contornando-os com a proteção jurídica adequada. O objetivo é verificar se existem lacunas na lei, e, se existirem, quais seriam essas lacunas e como poderíamos mitigá-las através da regulação. Dessa forma, buscamos notar se temos depositado a confiança adequada a esta ferramenta da anonimização ou se, ao contrário, estamos incorrendo em excesso.

2. *Big Data*, Dados Pessoais e Direito: A anonimização como possível solução para as demandas de privacidade e dos mercados na economia informacional.

Os dados foram ressignificados na sociedade informacional, tornando-se ativos importantíssimos. Não sem motivo, os dados são classificados como o novo petróleo, "um combustível que move a economia da informação" (MAYER-SCHÖNBERGER & CUKIER, 2013. p. 27-30), uma vez que eles possibilitam a extração de informações diversas, através do seu tratamento.

Por causa da importância dos dados para os mercados, Castells aponta que estamos em uma *Data Driven Economy,* pelo fato de que, dentro da sociedade informacional, temos uma economia movida a dados e direcionada pelas informações que se extraem ou que se especulam deles.

Castells afirma que, na *Data Driven Economy*, os quesitos da produtividade e lucratividade, antes suficientes para movimentarem os mercados, são agora reorientados a partir da maior financeirização dos mercados (CASTELLS, 1999, p. 201).

Até mesmo a produtividade é reorientada, já que se utiliza do processamento da informação, de tecnologias de geração de conhecimentos e da comunicação de símbolos, como aponta Castells (CASTELLS, 1999, p. 53-54).

Houve, portanto, um movimento em que os dados, e as informações dele decorrentes, deixaram de ser ferramentas úteis nos processos decisórios econômicos, para se tornarem elementos próprios e indispensáveis à economia (CASTELLS, 1999, p. 140). Isso se traduz numa verdadeira ruptura, caracterizada pelo desacoplamento entre a produção material e a geração de valor (MAZUCATTO, 2018).

Nesse sentido, a expectativa se torna um outro importante vetor econômico, deixando de lado até mesmo a lucratividade imediata, conferindo um caráter subjetivo ao que denominamos "valor" (MAZUCATTO, 2018). Segundo Castells, é a partir da confiança que se tem em determinados atores nos mercados, que se atribui valor ou não a determinados serviços e produtos (CASTELLS, 1999, p. 200)[4].

Isso se torna claro quando percebemos que muitas das empresas de tecnologias e plataformas de serviços, apesar de terem suas ações em altas cotações na bolsa de valores, na verdade não geram lucro imediato, como é o caso da Amazon, além do uber dentre outras (KHAN, 2017) (CELLAN-JONES, 2015).

Várias plataformas oferecem serviços a custo zero, ou próximo de zero, como aponta Schwab:

> Produtos e serviços inovadores criados na quarta revolução industrial possuem, de forma significativa, maior funcionalidade e qualidade, mas são entregues a mercados que são fundamentalmente diferentes daqueles que estamos tradicionalmente acostumados a mensurar. Muitos dos novos produtos e serviços são "não rivais", possuem custos marginais zero e/ou canalizam mercados altamente competitivos através de plataformas digitais;

4 Nas palavras do autor: "Parece que há dois fatores essenciais no processo de valorização: confiança e expectativas. Se não houver confiança no ambiente institucional no qual opera a geração de valor, nenhum desempenho em lucros, tecnologia ou valor de uso (por exemplo, recursos energéticos) se traduzirão em valor financeiro. Por outro lado, se houver confiança nas instituições subjacente ao mercado, então as expectativas do possível valor futuro de uma futura ação aumentarão seu valor" (CASTELLS, 1999, p. 200).

> isso tudo resulta em preços mais baixos. Nessas condições, as nossas estatísticas tradicionais talvez não consigam capturar os aumentos reais em termos de valores, pois o excedente do consumidor ainda não foi traduzido em vendas realizadas ou lucros mais elevados (SCHWAB, 2016, p.39).

A competitividade dessas empresas passa a estar ligada à capacidade de processamento de dados de suas plataformas, gerando um ciclo crescente de captação de dados, extração de informação, geração de serviços e captura de clientes.

Um exemplo disso é a curiosa frase em que Tom Goodwin apresenta o paradigma da sociedade informacional, ao afirmar que o Uber, mesmo sendo a maior empresa de táxis do mundo, não possui veículos; o Facebook, o mais popular proprietário de mídia do mundo, não cria conteúdos; o Alibaba, sendo o varejista mais valioso do mundo, não possui nenhum estoque; e o Airbnb, maior provedor de hospedagem, não possui nenhum imóvel (GOODWIN, 2015).

São todas plataformas que ganham em competitividade por orientar dados, tornando possível que os serviços sejam disponibilizados, sem que o serviço em si seja o seu negócio. O ativo passa a ser os fluxos de dados e as informações que deles se pode extrair.

Assim, a competitividade e o próprio valor de mercado dessas plataformas, atualmente, têm forte relação não só com os bens e serviços prestados, mas também com a capacidade de armazenar dados, e a capacidade de processamento de informações. Ou seja, tem forte relação com a capacidade de gerar respostas rápidas às demandas, criando expectativas, ajustando o fornecimento de serviços em plataformas cada vez mais personalizadas que consigam disputar a atenção do consumidor nos mercados da atenção (WU, 2016).

Nesse sentido, os dados pessoais se tornam verdadeira estratégia de negócio, principalmente por viabilizar a oferta de serviços extremamente personalizados, a partir da formação de perfis de preferência e de sugestões obtidas por inferência.

Dentro principalmente das relações de consumo, a personalização torna-se um grande atrativo, diferenciando a experiência do cliente com as empresas. Para os mercados, a utilização dos dados como frações de conteúdo que possibilitam a obtenção de informações foi essencial para o aprimoramento de serviços personalizados, como afirma Schwab:

> A capacidade de utilizar várias fontes de dados — desde as pessoais até as industriais, das fontes sobre estilos de vida às fontes comportamentais - oferece conhecimento granular sobre a caminhada de compras do cliente; algo impensável até recentemente. Hoje, dados e métricas (índices) oferecem informações cruciais em tempo quase real sobre as necessidades e comportamentos dos clientes que dirigem as decisões de marketing e vendas (SCHWAB, 2016, p. 59).

Nesse sentido, informações sobre o histórico de compras do cliente, seu perfil de preferência e a análise de possíveis interesses conexos são possíveis através do tratamento de dados pessoais e são extremamente valiosas para os mercados. Esse é fenômeno o chamado de "comodificação" descrito por Stefano Rodotà, uma vez que os dados, em especial os dados pessoais, passam a ser objeto de comércio, como uma mercadoria no novo mundo digital (RODOTÀ, 2008).

De fato, através dos seus dados pessoais, há uma verdadeira monetarização de parcelas do sujeito e de sua personalidade. Isto torna possível aos mercados, dentro do capitalismo informacional, a captação dos mais íntimos interesses de seus clientes e a construção de plataformas de produtos e serviços quase irresistíveis, já que são desenhadas para atender especificamente cada consumidor

O conceito de dados pessoais pode ser definido de forma restrita ou abrangente. Pelo conceito restrito, dado pessoal é aquele relacionado à pessoa identificada, através de indicadores diretos, como o nome, ou indiretos, como o telefone, o endereço, etc. Já pelo conceito amplo, dado pessoal é aquele relacionado à pessoa identificável, ou seja, o dado que possibilite identificar uma pessoa, ainda que por meio de tratamento, é considerado dado pessoal (DONEDA & MACHADO, 2018, p. 99-128). O ordenamento brasileiro adota o conceito amplo de dado pessoal[5].[6]

Há uma grande quantidade de dados pessoais, dentre os dados que trafegam pelas redes. Isso devido a muitos fatores. Primeiro porque, apenas recentemente, depois das notícias mais expressivas de vazamento de dados é que os cidadãos têm tomado consciência dos riscos que envolvem a exposição dos dados pessoais, como iremos tratar em tópico específico. Mas ainda hoje (e até pouco tempo, ainda mais inadvertidamente), as pessoas disponibilizam voluntariamente os seus dados para ter acesso a aplicativos gratuitos, downloads de conteúdo, ou mesmo para utilizar aplicativos que são dependentes dos dados para proporcionar serviços.

---

5 Conforme se depreende do art. 5º, I, da Lei Geral de Proteção de Dados- LGPD (Lei nº 13.709/ 2018): Art. 5º Para os fins desta Lei, considera-se: I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável. No mesmo sentido, o art. 4º da Lei nº 12.527/2011 (Lei de Acesso à Informação- LAI): Art. 4º Para os efeitos desta Lei, considera-se: IV - informação pessoal: aquela relacionada à pessoa natural identificada ou identificável.

6 Para os autores Diego Machado e Danilo Doneda há ainda uma terceira classificação do conceito de dado pessoal, que seria o conceito absoluto, onde a identificabilidade da pessoa natural seria auferível não apenas pelos esforços do responsável pelo tratamento mas também por qualquer outra pessoa com acesso à base. Segundo os autores, essa classificação foi adotada tanto pelo Grupo de Trabalho de Proteção de Dados do Artigo 29, no Parecer 04/2007, quanto pelo Tribunal de Justiça da União Europeia no caso Breyer. Cf. SPINDLER, Gerald; SCHMECHEL, Philipp. Nesse trabalho, no entanto, não utilizarem essa classificação, portanto abordaremos sob a mesma nomenclatura de conceito amplo de dado pessoal, o dado que possua identificabilidade tanto pelos agentes de tratamento quanto por terceiros.

Além dos dados fornecidos espontaneamente, temos também os dados disponibilizados pelos governos, em seus programas de transparência. Há ainda diversas notícias de compartilhamento e comercialização de dados pessoais entre instituições do próprio governo, como é o caso de notícias envolvendo o SERPRO (Serviço Federal de Processamento de Dados) (CRUZ, 2018), além de transações entre o governo e instituições privadas, uma vez que, como mencionado, os dados são importantes ativos para os mercados[7].

Muitos desses dados, apesar de não serem capazes de identificar o indivíduo por indicadores diretos, podem ser considerados pessoais, ao viabilizar essa identificação, quando em conjunto com outros dados, uma vez que, pelo conceito amplo, o dado que puder identificar uma pessoa também é considerado pessoal.

Além dessas formas de disponibilização de dados pessoais, o surgimento da denominada "internet das coisas" tornou ainda mais massiva e praticamente imperceptível essa remessa de dados. O Decreto nº 9.854/2019, em seu art. 2º, I, define internet das coisas como "a infraestrutura que integra a prestação de serviços de valor adicionado com capacidades de conexão física ou virtual de coisas com dispositivos baseados em tecnologias da informação e comunicação existentes e nas suas evoluções, com interoperabilidade".

Segundo Cirani, Ferrari, Picone e Veltri, a internet das coisas (IoT) "encapsula uma visão de um mundo no qual bilhões de objetos com inteligência incorporada, meios de comunicação e capacidades de detecção e atuação se conectarão por redes de IP (Internet Protocol)" (CIRANI et al., 2019, p. 1). Os autores afirmam que cada vez mais objetos da vida cotidiana têm sido incorporados à IoT, através de microcontroladores, transceptores ópticos e / ou de rádio, sensores, atuadores e pilhas de protocolo, reunindo dados do ambiente e formulando uma interface com o mundo físico (CIRANI et al., 2019, p. 1).

Para IoT, o armazenamento de dados, seu processamento e análise são requisitos essenciais para transformar dados brutos em informações úteis (CIRANI et al., 2019, p. 1).

O termo "internet das coisas" foi primeiramente utilizado pelo cofundador e diretor executivo do Auto-ID Center, Kevin Ashton, e se torna mais usual com a criação de sensores cada vez menores, capazes de captar informações dos ambientes e encaminhá-las para as redes.

Normalmente os dados repassados em sistemas de IOT são os denominados metadados, que em tese teriam menor potencialidade de fornecer informações sensíveis.

7 O compartilhamento de dados entre os setores público e privado ganhou destaque no Acordo de Cooperação Técnica 7/2013 do TSE, que envolveu a tentativa de acordo para compartilhamento de dados entre o Tribunal Superior Eleitoral e o Serasa EXPERIAN, o qual foi interrompido por decisão judicial (HADAIR, 2013).

"Metadados" significa basicamente "dados" sobre "dados". São dados utilizados para identificar outros dados, a fim de se organizar ou gerir essas informações a fim de se descrever pacotes de dados, sem que eles sejam violados. (DIAMANTINI et. al., 2019, p.305).[8]

São, portanto, dados aparentemente inofensivos, posto que referentes somente ao modo de transmissão ou ao arquivo em que os dados pessoais, por exemplo, estão contidos. Entretanto, mesmo os metadados podem revelar muito mais do que o que se imagina. Isso porque quando somados a outros dados, diversas informações podem ser extraídas.

A internet das coisas se utiliza dos metadados para trazer inúmeras facilidades nas tarefas cotidianas, em tempo quase real, tornando-nos cada vez mais dependentes delas. É a própria concretização da crítica de Jonathan Crary ao capitalismo informacional, que não dorme e nem permite que seus cidadãos "durmam" (CRARY, 2016, p.12-38), uma vez que, mesmo dormindo, permanecemos indiretamente conectados transmitindo nossos dados através dos aparelhos eletrônicos em rede.

Com esses aparelhos, a conectividade segue ativa 24 horas por dia, 7 dias por semana, principalmente quando se trata de tecnologias ligadas às rotinas domésticas. Dessa forma, em todo o tempo, os dados da casa são enviados pelas redes para análise e aprimoramento dos serviços.

A hiperconectividade, traduzida na conexão 24h, permitiu que uma quantidade enorme de dados fosse gerada e transmitida. Como vimos, a partir desses dados, acumulados sem aparente e atual finalidade, (SIEWERT, 2013) se extraem informações valiosas para os mercados, em nível de estratégia de negócio, gerando valor a partir das expectativas conferidas.

Concomitantemente à geração cada vez mais frequente de dados nas redes, culminando nas chamadas plataformas *Big Data,* houve também o aprimoramento de maneiras de capturar e utilizar esses dados. Converter esse grande volume de dados em serviços só foi possível através das ferramentas de análise massiva de dados, como o *Big Data* Analytics, que se tornaram extremamente importantes para a subsistência das empresas nesses mercados.

É o que discorreremos a seguir.

2.1. *Big Data* e *Big Data Analytics*- A realidade atual dos dados.

O tratamento de dados massivos (o chamado *Big Data)* se traduz na utilização de novos métodos de domínio da informação para produção de bens ou serviços (MAYER-

8 Uma das formas de classificação de metadados é o Dubin Core, que recebeu o padrão ISO 15836:2009. Disponível em: http://dublincore.org/metadata-basics/

SCHÖNBERGER e CUKIER, 2013, p. 14). O termo é caracterizado pelo volume de dados, velocidade de captura, suas variadas fontes, e ainda pela veracidade desses dados (TAURION, 2013) (BARBIERI, 2020).

Tanto o armazenamento desses dados quanto o seu processamento através de *Big Data Analytics* foi se tornando uma realidade nos mercados através da diminuição dos custos e da sofisticação dos processadores. Por meio dessas ferramentas se extraem informações variadas dos dados acumulados que podem servir aos mais diversos fins, a depender da demanda projetada em seus algoritmos.

As técnicas de *Big Data Analytics* permitem a obtenção de informações diversas com base em dados por vezes, triviais, como o que acontece com os já mencionados metadados, tornando particularmente arriscada a disponibilidade de dados nas redes.

Com a exposição desses dados, não é apenas a privacidade que se encontra ameaçada. A má utilização de dados pessoais pode levar a práticas discriminatórias, ao aprofundamento da marginalização e de desigualdades sociais, pode dificultar o acesso a produtos ou serviços e mesmo proporcionar grandes desafios à democracia. Existem casos documentados sobre isso, como as práticas discriminatórias do site Decolar.com aos consumidores brasileiros (FRAZÃO, 2018).

O site ofertava as mesmas acomodações com sobrepreço para as pessoas que realizavam as pesquisas em território brasileiro (o denominado *geopricing*). Além disso, algumas acomodações disponíveis não eram sequer ofertadas aos consumidores de determinadas localizações (*geoblocking*) (FRAZÃO, 2018). Após denúncia pela concorrente *Booking*, a Decolar foi julgada pelo Departamento de Proteção e Defesa do Consumidor da Secretaria Nacional de Relações de Consumo do Ministério da Justiça e condenada a pagar multa de sete milhões e meio de reais (FRAZÃO, 2018).

Dessa forma, o que se percebe é que mesmo um metadado, como a informação sobre o país de origem do IP de conexão, pode ser mal utilizado, resultando no preço diferenciado de uma transação, ou mesmo na indisponibilidade do serviço por meio de práticas discriminatórias.

E, para além da discriminação, os dados podem servir à restrição de liberdades individuais, aumentando o controle não apenas do Estado, através do chamado *dataveillance*[9], mas também das demais instituições de poder que permeiam a vida social, como no capitalismo de vigilância relatado pela professora Shoshana Zuboff (ZUBOFF, 2021).

9 O termo, derivado da língua inglesa, remete ao uso de dados para se alcançar monitoramento e investigação. A expressão é atribuída ao cientista da computação Roger Clarke (MENEZES NETO, 2016, p. 121).

A autora adverte que a apreensão de dados pessoais de forma cotidiana e generalizada se traduz em uma forma de capitalismo que se alimenta da vigilância dos indivíduos. Nas palavras da autora:

> "O capitalismo de vigilância reivindica de maneira unilateral a experiência humana como matéria-prima gratuita para a tradução em dados comportamentais. Embora alguns desses dados sejam aplicados para o aprimoramento de produtos e serviços, o restante é declarado como superávit comportamental do proprietário, alimentando avançados processos de fabricação conhecidos como "inteligência de máquina" e manufaturado em produtos de predição que antecipam o que um determinado indivíduo faria agora, daqui a pouco e mais tarde" (ZUBOFF, 2021, p. 22).

Isso resulta no que Shoshana Zuboff chama de mercados de comportamento futuro, ambiente no qual essas informações são comercializadas e utilizadas não apenas para capturar consumidores, mas também para moldar novos comportamentos, orientar escolhas e até mesmo controlar emoções.

O uso dos dados para esses fins vem alertando cidadãos e instituições públicas, na tentativa de impor limites legais e éticos à utilização dos dados pessoais.

As preocupações sobre os riscos da exposição desses dados culminaram em propostas legislativas para regulação do uso de dados e de sua proteção. Apesar da inegável importância das legislações anteriores, a proteção de dados ganhou espaço no debate internacional nos anos 90 com as iniciativas de normatização do uso dos dados no contexto da União Europeia, através da Lei 2/94 e da Diretiva 95/46 do Parlamento Europeu. Já em 2016 houve a promulgação do Regulamento Geral sobre a Proteção de Dados (RGPD) 679 de 2016, que, sobrepondo-se às diretrizes de cada Estado-Nação, unificou as regras de proteção de dados na Comunidade Europeia.

No entanto, o debate sobre proteção de dados ganhou ainda maior repercussão internacional depois do marco internacional traduzido nas revelações de Edward Snowden (BBC NEWS, 2014) e com o escândalo da Cambridge Analytica (THE NEW YORK TIMES, 2018). Com o episódio, as iniciativas legislativas europeias já existentes de proteção de dados ganharam enorme visibilidade, enquanto a discussão sobre a necessidade de proteção ganhou corpo nos Estados Unidos, repercutindo em diversos outros países. É o que veremos a seguir.

2.2. Dados e Legislação- Breve relato do desenvolvimento da legislação internacional de proteção de dados

Legislar sobre dados sempre foi uma tarefa complexa. Principalmente porque o Direito vem aos poucos construindo suas diretrizes sobre o assunto. A temática segue com problemas essenciais como: Qual a natureza jurídica dos dados pessoais? Quais desdobramentos dos dados são relevantes para o Direito? Até que ponto as legislações nacionais serão de fato eficientes na regulamentação do uso de dados? Dados seriam considerados coisa? Se sim, seria possível atribuir propriedade a esses dados? E de quem seria a propriedade? O assunto é ainda mais delicado quando aplicamos esses conceitos aos dados anonimizados.

Danilo Doneda aponta que, ao longo do tempo, houve ondas legislativas que buscavam respostas a essas perguntas de forma bem diversas. O autor afirma que uma das estruturas de proteção de dados pessoais seria por meio do direito privado, reconhecendo a qualidade de bem jurídico à informação. Dessa forma, seriam disponibilizados instrumentos jurídicos como o sistema de propriedade intelectual, ou reconhecido o direito de propriedade sobre dados, formalizando a existência de um mercado sobre esses bens (DONEDA, 2006, p.165). O autor faz crítica, no entanto, à adequação em se considerar a informação como bem jurídico e estender-lhe tutela puramente patrimonial. Segundo ele, os dados podem envolver uma multiplicidade de situações e questões, em que o interesse patrimonial é o menor deles.

Há ainda a possibilidade de objetivação em relação aos dados pessoais "que os considera elementos objetivos da abordagem que a matéria vem recebendo, no entanto sem consistir em sua patrimonialização" (DONEDA, 2006, p. 168). Aqui prevalece a instrumentalidade dos dados, que seriam então tutelados por si, sem que se concentre em seu sujeito, mas sim no seu uso e nas plataformas que os detém.

Por fim, uma outra forma de tutela é o enfoque no sujeito. Nesse caso, a informação é considerada representação direta do indivíduo e, portanto, extensão de sua personalidade (DONEDA, 2006, p. 168).

Essas três possibilidades foram desenvolvidas na prática através das gerações legislativas voltadas à proteção de dados. Doneda faz uma explanação dessa evolução legislativa (DONEDA, 2006, p. 206-214).

Uma **primeira geração**, no início dos anos 70, estaria com suas preocupações voltadas aos bancos de dados, de forma a evitar a aglomeração de informações facilitando o processamento. Ela foi marcada pelos casos do *National Data Center (NDC*) e Safari, respectivamente nos Estados Unidos e na França.

Por volta de 1965 houve a proposta de compor uma base de dados unificada dos dados pessoais dos americanos no *National Data Center*. A ideia era reunir os dados do Censo,

dos registros trabalhistas, do fisco e da previdência social (GARFINKEL, 2000, p. 13). A intenção era melhorar a eficiência, ou seja, era voltada a questões técnicas, mas não se preocupou com as repercussões na privacidade dos cidadãos. A questão de eficiência era realmente um fator, já que a unificação traria mais facilidade de identificação de incongruências nos dados, menor custo de armazenagem, além de se evitar a duplicação de informações, causando dispêndio de recursos.

O congresso americano realizou debates sobre o tema, abarcando a opinião pública, e considerou que essa unificação se refletiria em mais poder centralizado nos governos. Isso só poderia ser feito se fosse garantida, o máximo possível, a proteção da privacidade dos cidadãos. Uma das conclusões do Congresso foi pela opção da permanência de bases de dados separadas como uma das opções viáveis pró-privacidade, mesmo que em detrimento da eficiência administrativa. Àquela época, já foi possível observar nos debates discussões acerca da relação entre os dados pessoais, a dignidade e a proteção da personalidade. Ao final, o projeto do *National Data Center* foi arquivado.

De forma semelhante ao que ocorreu nos EUA, a França, nos anos 70, buscou facilitar a troca de dados entre os órgãos da administração pública para sistemas informatizados, através do *Instituit National de la Statistique*, por meio do projeto SAFARI (*Système Automatisé pour les Fichiers Administratifs et le Répertoire des Individus*). Cada cidadão seria identificado pelo Estado por um número único (*Sécurité Sociale*), válido por toda a vida. O projeto também visava eficiência administrativa, mas não contou com o apoio popular e foi vetado pelo primeiro-ministro francês em 1974. O caso deu ensejo a discussões sobre a privacidade de dados, culminando na lei francesa de proteção em 1978 (*Loi Informatique, Fichiers et Libertés*) (DONEDA, 2066, p. 183-192).

Os casos mencionados refletem os esforços da primeira geração legislativa em se evitar a unificação de bases de dados como forma de se preservar a privacidade, numa **abordagem de instrumentalidade dos dados**, com foco nos dados em si.

Interessante mencionar que, com a evolução das técnicas de processamento, ainda que as bases não estivessem unificadas, o tratamento dos dados se tornou uma realidade por meio, por exemplo, do mencionado *Big Data Analytics*, tornando possível a vinculação de dados e inferência de informações.

Por sua vez, nos anos 70, houve uma **segunda geração de leis**. A Alemanha foi pioneira na promulgação de uma lei específica para proteção de dados através do Estado Alemão de Hesse, em 1970. Nesta lei foi cunhada pela primeira vez o termo “proteção de

dados" (*Datenschutz*), ao invés dos tradicionais termos "*Datensicherung"* ou "*Datensicherheit"*, voltados à segurança da informação (DONEDA, 2021).

Porém, o marco dessa segunda geração foi a mencionada *Informatique et Libertées*, lei francesa de 1978 (DONEDA, 2006, p. 208-209), que considerou a privacidade e a proteção de dados como parte da liberdade negativa. Dessa forma, o foco deixa de ser os dados e passa a ser o cidadão titular dos dados particularmente sensíveis. Além disso, foram traçadas diretrizes sobre o que seria o uso indevido dos dados pessoais, através do princípio da finalidade.

Nessa geração iniciaram-se as discussões sobre se atribuir um **enfoque patrimonial** aos dados pessoais[10]. A aplicação dos princípios do direito de propriedade aos dados pessoais surgiu como uma alternativa à falta de uma tutela mais específica a esses direitos, e já era suscitado ainda em 1890 por autores como Samuel Warren e Louis Brandeis (MAURO, 2019), principalmente nas questões envolvendo a corte Norte-Americana. A quarta emenda da Constituição Americana com sua disposição sobre a inviolabilidade de domicílio embasou a proteção conferida à privacidade e ao impedimento de sua violação (MAURO, 2019).

Mas a questão de se conferir tutela patrimonial por meio dos direitos de propriedade perdurou mesmo após o estabelecimento de legislações voltadas à privacidade. Isso porque, em um primeiro momento, o instrumental jurídico utilizado em relação aos dados se relacionava com a indenização posterior, por danos causados (OHM, 2010).[11] O dano à privacidade, seria então um dano psíquico e moral; trazendo um caráter físico não corporal do dano.

A atribuição de um caráter de patrimonialidade a esses dados poderia conferir uma proteção não apenas repressiva, mas preventiva, fazendo com que não fosse necessário aguardar a ocorrência do dano comprovado para suscitar direitos frente ao uso indevido de dados pessoais.

O próprio Lawrence Lessig apontava a possibilidade de uma vertente patrimonial dos dados, já que eles seriam objetos implícitos ou explícitos de negócios jurídicos firmados na

10 Ressalta-se que essa não é uma questão pacificada atualmente. A exemplo da declaração de Brittany Kaiser, da Cambridge Analytica, no congresso estadunidense em defesa da classificação do dado pessoal como um ativo precificável, e, portanto, de que seu uso deveria remunerar o titular. Algumas iniciativas nesse sentido têm sido realizadas por sites especializados, como no caso da Ocean Protocol (www.oceanprotocol.com), plataforma para a troca descentralizada de informações, onde o titular controla e libera os seus dados para análise, eventualmente, em troca de pagamento.

11 Nas palavras de Ohm: "Before deciding how to respond to the rise of easy reidentification, we must recognize three themes from this history of privacy law. First, while privacy torts focus solely on compensating injured victims of privacy harms, more recent privacy statutes shift the focus from post hoc redress to problem prevention. Second, this shift has led to the hunt for PII through quasi-scientific exercises in information categorization. Third, legislatures have tried to inject balance into privacy statutes, often by relying on robust anonymization" (OHM, 2010, p. 1732).

rede. Desta forma, os dados pessoais seriam a moeda de troca para o acesso aos diversos serviços gratuitos disponíveis nas redes mediante cadastro (LESSIG, 1999).

Essa discussão sobre conferir a natureza de propriedade a esses dados ganhou repercussão no direito de diversas formas. Uma delas foi a tentativa de equiparação dos dados pessoais com a propriedade intelectual. Foi o que aconteceu nos EUA, no caso *Moore x Regents of the University of California*. O julgado envolveu a discussão sobre a propriedade do material genético colhido de Moore para um tratamento no hospital de Los Angeles. As células do baço de Moore tinham propriedades peculiares e foram analisadas pela Universidade da Califórnia. A pesquisa resultou na produção de um fármaco, sobre o qual Moore postulava o direito de parcela da patente. O processo não chegou à suprema corte dos EUA, mas foi até a suprema corte do Estado da Califórnia, que negou o pedido de Moore (BOYLE, 1992).

A corte mencionou quatro motivos para o indeferimento: 1) o caráter de abandono das células extraídas do corpo; 2) a existência de legislação que previa a destituição do caráter de propriedade quanto ao material genético usado para pesquisa; 3) a não compatibilidade imediata entre a violação de privacidade alegada com o pedido de aferição dos lucros de patente; 4) os efeitos que um precedente atribuindo direito de propriedade a material genético poderia causar nas pesquisas científicas (BOYLE, 1992).

Boyle, analisando o caso Moore, critica a atribuição de propriedade aos dados pessoais, principalmente no que tange à equiparação com direitos autorais. Ele aponta que a teoria jurídica clássica não seria suficiente para resolver os problemas que envolvem a informação na sociedade em rede. De fato, apesar das tentativas de se atribuir um caráter estritamente patrimonial aos dados, várias lacunas foram identificadas na efetivação da proteção jurídica deste bem.

Nesse sentido, uma **terceira geração de leis** teve seu marco com a decisão do Tribunal Constitucional Alemão que definiu o princípio da autodeterminação informativa atribuindo às informações pessoais uma vertente do **direito de personalidade**.

Como mencionado, a primeira lei de proteção de dados foi de origem Alemã, em 1970 por um dos seus estados. A Lei federal alemã, por sua vez, foi promulgada em 1977- a Bundesdatenchutzgesetz (DONEDA, 2006, p. 193-197). Com essa tradição de proteção, alguns aspectos da lei que organizava o senso da República Federativa Alemã- RFA, de 1982, foram questionados. A lei submetia todos os cidadãos a responder a 160 perguntas, sob pena de multa e cuja resposta poderia ser usada para retificação de registros dos cidadãos e transmitida para os Länder. Havia ainda o favorecimento de denúncia dos cidadãos que se recusassem a responder ao censo.

Surgiram então preocupações sobre se os dados seriam utilizados para além de se promover estudos estatísticos, o que não era proibido pela lei de proteção em vigor. Assim, numa decisão em 1983, um juiz administrativo decidiu que as disposições da lei do censo prevaleceriam sobre as disposições da Lei de proteção de dados em caso de conflito.

Com isso, aumentaram as insatisfações acerca da lei do censo, que acabou sendo levada à Corte Constitucional. A sentença da Corte foi pela suspensão temporária do Censo e a declaração de inconstitucionalidade da lei que o instituiu, com base no chamado Direito Geral da Personalidade.

Dentre os motivos para a inconstitucionalidade estava a diversidade de finalidade do censo (que servia não só a fins estatísticos mas à retificação de registro civil). Portanto, era necessário observar o princípio da Finalidade, para informar ao cidadão sobre o uso efetivo de seu dado.

A decisão também ressaltou que mesmo os dados considerados inicialmente como irrelevantes para a privacidade podem adquirir um novo significado quando há o seu processamento, apontando a necessidade de proteção também para esses dados (DONEDA, 2006, pg. p. 193-197).

A sentença cunhou a expressão "autodeterminação informativa" para afirmar que os cidadãos teriam o direito de decidir sobre o uso de seus dados. Estabeleceu ainda o direito ao controle sobre suas próprias informações[12]. A terceira geração, que se originou das discussões que emergiram dessa sentença, continuava centrada no cidadão, como na geração anterior. Mas aqui a informação é considerada parte integrante do sujeito, como elemento de sua personalidade. Nesse sentido, seria importante que o cidadão tivesse consciência sobre quais de seus dados estão disponíveis, como eles serão utilizados, estabelecer ferramentas jurídicas para se evitar o abuso nessa utilização e para possibilitar a retificação e mesmo exclusão desses dados.

Além disso, o direito passou a se preocupar na atuação de forma preventiva, ou seja, visando a proteção anterior à ocorrência de danos, através da identificação prévia e proteção de dados particularmente sensíveis ou de dados que poderiam identificar o usuário, mas que normalmente recebiam um menor grau de atenção. (OHM, 2010). Assim, a atuação passa de reparadora para protetiva utilizando meios jurídicos diversos para este fim.

12 Iniciativas como a Solid, de Tim Berners-Lee, criador da World Wide Web (http://solid.inrupt.com/) vem promover justamente um maior domínio dos titulares sobre os seus dados, já que os eles podem controlar quais serviços podem acessar cada informação cadastrada. A proposta vem na mesma linha da criação da carteira virtual, defendida por Lessig (LESSIG, 2006).

Atualmente, as principais legislações sobre dados pessoais encontram-se alinhadas a essa terceira geração normativa. Interessante ressaltar ainda que, no Brasil, o entendimento de que os direitos pessoais são, na verdade, direitos de personalidade se refletiu na proposta de Emenda Constitucional nº 17, que segue em tramitação[13]. A emenda propõe a inclusão do direito aos dados como um direito fundamental pela Constituição. Há autores, no entanto, que já consideram esse direito como fundamental, pela previsão do art. 5º, XII da Carta Magna (MAURO, 2019), ou mesmo de forma implícita com a tutela concedida por meio do *habeas data* (SARLET, 2021).

Além disso, as discussões sobre a autodeterminação informativa ganharam visibilidade no contexto nacional, uma vez que o princípio foi expressamente mencionado em uma recente decisão do STF. Trata-se das ações diretas de inconstitucionalidade contra a medida provisória nº 954/2020, que possibilitava o compartilhamento de dados entre as empresas de telecomunicações e o IBGE, para fins de pesquisa demográfica em tempos de COVID-19.

Como direito da personalidade, portanto, os dados pessoais ganham um reforço simbólico em direção à proteção.

A autodeterminação informativa, no entanto, vem sendo objeto de discussões e críticas. Isso porque percebeu-se que, para se garantir essa autodeterminação de forma concreta, o enfoque deve estar na efetividade dessa liberdade em se fornecer ou não os dados. Isso porque, com a utilização cada vez mais frequente das tecnologias, o próprio acesso à vida social torna a troca de dados um requisito indispensável, o que faz com que a liberdade de decidir livremente seja cerceada. Nesse sentido, a autodeterminação informativa, ou seja, a centralização do controle dos dados no cidadão nem sempre seria suficiente para a proteção.

---

13 A PEC 17, originária do Senado Federal visa alterar os artigos 5º e 22 da Constituição, alterando-os com o seguinte teor:
Art. 1º O inciso XII do art. 5º da Constituição Federal passa a vigorar com a seguinte redação:
"Art. 5º .....
XII – é inviolável o sigilo da correspondência e das comunicações telegráficas, de dados e das comunicações telefônicas, salvo, no último caso, por ordem judicial, nas hipóteses e na forma que a lei estabelecer para fins de investigação criminal ou instrução processual penal, bem como é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais;
....." (NR)
Art. 2º O caput do art. 22 da Constituição Federal passa a vigorar acrescido do seguinte inciso XXX:
"Art. 22. .....
XXX – proteção e tratamento de dados pessoais.
....." (NR)
Dessa forma, a alteração definiria, portanto, a competência para legislar sobre o tema como sendo da União. Isso auxiliaria na uniformização da política de proteção de dados, que estaria centralizada neste Ente federativo.

Aparecem então as primeiras nuances de uma **quarta geração de leis** de proteção de dados, que almeja suprir as lacunas da proteção com enfoque individual, centrado na escolha individual, buscando-se elevar o padrão coletivo de proteção. (DONEDA, 2006, pg. 210-212).

A quarta geração deveria levar em conta o desequilíbrio entre o cidadão e as entidades que coletam seus dados; e as lacunas na liberdade em dispor de seus dados através de uma escolha individual, no que inicialmente se cogitava de autodeterminação informativa. (DONEDA, 2006, pg. 212). Dessa forma, as discussões sobre a natureza de direito de personalidade dos dados levam ainda a outras questões, como se eles poderiam ser objeto de livre disposição ou, ao contrário, se seria necessário a imposição de limites.

Podemos destacar que o consentimento livre e informado parece ser uma das saídas que mais caracterizam a terceira onda legislativa, apesar de existirem outros mecanismos legais que viabilizam o tratamento do uso de dados pessoais. Mas é possível se pensar em formas de restrições ao uso de dados construídas a partir de outros princípios legais, mesmo quando exista o consentimento do titular.

Além das questões relacionadas à liberdade de fato envolvida no consentimento, acrescentamos ainda a discussão sobre o impacto social dos dados pessoais disponibilizados.

Ainda que possa ser caracterizado como um direito individual mais flexivelmente disposto, entendemos que a proteção desses dados, para ser efetiva, deve levar em conta seu aspecto comunitário, incidindo o princípio da fraternidade. Isso porque o acesso aos dados pode se dar de formas variadas, inclusive sem a atuação do indivíduo, afinal, vivemos na realidade em comunidade e estamos vinculados em rede a diversos outros indivíduos que compõem nossos ciclos sociais. Portanto, de fato, os eventos sociais repercutem no mundo dos dados. Desta forma, o fato de dispor dos próprios dados pode atingir pessoas vinculadas direta ou indiretamente aos indivíduos.

Essas questões parecem não estar ainda amadurecidas nas legislações contemporâneas, apesar de permearem o debate. No entanto, é cada vez mais necessário que a regulação ganhe uma nuance para além do indivíduo.

2.3. Dados e Legislação - a Evolução legislativa Brasileira quanto aos dados e seu tratamento.

Como reflexo dessas iniciativas internacionais na proteção de dados, vários países têm se empenhado na regulamentação do uso de dados em seus territórios, até para assegurar a

viabilidade de transações comerciais e financeiras pelas redes com os países que já estabeleceram seus critérios de compartilhamento de dados.

Os esforços brasileiros para a regulamentação da proteção de dados culminaram na promulgação da Lei 13.709/2018, denominada Lei Geral de Proteção de Dados Pessoais- LGPD. Destacaremos um tópico próprio para o estudo da mencionada legislação. De qualquer forma, para entender os desafios que a efetividade da mencionada lei enfrenta e enfrentará para atingir o objetivo da proteção de dados especialmente no que tange aos dados anonimizados, é necessário compreender os passos que antecederam a sua proposta.

Assim como no contexto internacional, o Brasil passou por fases legislativas quanto ao uso e armazenamento de dados. Partimos de uma primeira fase, centrada na unificação de bases de dados estatais, seguindo por um movimento para transparência dos dados públicos e, por fim, transitamos na fase onde o debate acerca da necessidade de proteção aos dados pessoais ganha ênfase.

A **primeira fase legislativa** teve por contexto a onda de desburocratização dos anos 60-70. Nesse período, houve reformas administrativas que promoveram a criação de programas nacionais de desburocratização e desestatização (COSTA, 2008, pg. 853). Com base em modelos gerenciais[14], a administração pública buscava uma maior eficiência[15] na gestão dos serviços públicos.[16]

Seguindo essa lógica, o Brasil vivenciou (e vivencia) a busca de uma política administrativa de eficiência na proposição de bases de dados públicas cada vez mais unificadas, visando o aprimoramento da gestão, incremento da qualidade de dados e melhoria de serviços públicos prestados.

O projeto Registro Nacional de Pessoas Naturais- RENAPE ou RNPN - foi um dos primeiros movimentos nesse sentido de unificação de dados sob uma mesma base. A proposta era confidencial e se deu durante os anos 70 (VIANNA, 2014, p. 1448-1471). A ideia era criar uma base de dados cadastrais unificada de forma a atribuir uma numeração única para cada cidadão. Propostas anteriores de identificação em um número único, como o PL nº 352, de

14 A atuação gerencial buscava trazer boas práticas de administração de empresas privadas, para reparação da burocratização exacerbada, na busca de uma maior efetividade e eficiência da atuação estatal. As políticas públicas, portanto, foram fortemente influenciadas pelo modelo gerencial de gestão pública, visando, além da eficiência, à contenção de gastos, seguindo uma política fiscal restritiva.

15 Apesar do princípio da eficiência ter sido inserido na Constituição através da Emenda Constitucional 19 apenas em 1998.

16 Nos anos 90 houve uma terceira grande reforma administrativa que, apesar de um relativo retorno a práticas mais burocráticas pela constituição de 1988 (COSTA, 2008, pg.855), promoveu, no geral, uma progressiva ampliação do poder decisório, tradicionalmente atribuído ao Estado, à participação de outros atores sociais.

03/06/1955, tinham recebido críticas, ante a impopularidade de identificar as pessoas por número e não por nome.

Por causa disso, em 1971 o ministro da Justiça à época, Alfredo Buzaid, instituiu Comissão Interministerial para discutir o tema sem que o assunto fosse publicizado, com o objetivo de se criar uma base unificada que possibilitasse o registro único, cadastrada pelo Serpro e gerida por um novo órgão, ligado ao Ministério da Justiça[17]. Após a divulgação de dados sobre o projeto RENAPE pela mídia e, como resposta às preocupações que surgiram acerca dele, surgiram em 1977 dois projetos de lei que buscavam regulamentar alguma proteção aos dados em uma possível unificação das bases. Os PLs organizavam a fiscalização para criação e manutenção de bases de dados pela CAPRE- Comissão de Coordenação de Atividades de Processamento Eletrônico. Entretanto, não foram bem sucedidos.

Percebe-se, portanto, à semelhança da primeira onda legislativa internacional, a preocupação brasileira na existência de grandes bases de dados unificadas, que permitiriam assim o acesso centralizado a dados de qualquer cidadão (DONEDA, 2021).

De qualquer forma, o próprio Projeto RENAPE não pôde atingir seus objetivos. Primeiro, porque começou a sofrer oposição por especialistas ligados ao Serpro. Segundo, devido aos problemas técnicos relacionados à magnitude do projeto. A integração dos números de cadastros disponíveis (certidão de nascimento, identidade, título de eleitor...) era um grande desafio, que se agravava com as inconsistências das próprias bases, às vezes baixa qualidade dos dados e incompatibilidades entre as bases. Além disso, os softwares de base de dados dos anos 70 apresentavam limites técnicos para a eficiência desses sistemas. Por conta disso, o projeto foi arquivado em 1978.

As iniciativas para unificação dos dados tiveram prosseguimento com a instituição do número único de Registro de Identidade Civil- RIC, pela Lei 9.454, de 1997, que foi regulamentada apenas em 2010, com o Decreto nº 7.166/2010, mas sem que, na prática, a unificação das bases realmente se efetivasse.

A tentativa mais recente de uniformização tem se desdobrado com o Decreto 10.046 de 09/10/2019 que instituiu o Cadastro Base do Cidadão, além de estabelecer critérios para o compartilhamento de dados no âmbito da administração pública federal. O mencionado decreto

17 É possível fazer uma breve cronologia sobre o Projeto: Em 1972, a comissão interministerial concluiu seus trabalhos, e alguns dos avanços do Serpro foram divulgados na imprensa, o que repercutiu numa reportagem crítica da Revista Realidade. Em 1977 houve também uma matéria contundente contra o RENAPE pelo Estado de São Paulo, o que gerou críticas pelo presidente do Conselho Federal da OAB.

revogou o Decreto Federal nº 8.789/2016, e estabeleceu regras ainda mais flexíveis de compartilhamento de dados (FRAGOSO & MASSARO, 2019).

O projeto Cadastro Base do Cidadão se propõe a viabilizar a criação de um meio unificado de identificação do cidadão[18], facilitando o compartilhamento de informações cadastrais entre os órgãos da administração[19], promovendo a interoperabilidade entre essa base e as bases de dados de determinadas políticas públicas (bases temáticas)[20]. Dessa forma, com exceção dos atributos genéticos[21], todos os outros dados, ou seja, atributos biográficos, biométricos e cadastrais, podem compor a base integradora do Cadastro Base do Cidadão[22], o que inclui informações como endereço, impressão digital, retina ou íris, formato da face, voz, além de informações comportamentais como a maneira de andar. O Cadastro Base do Cidadão "servirá como referência de informações sobre os cidadãos para os órgãos e entidades do Poder Executivo Federal"[23]. Portanto, temos aí o despontar mais recente do projeto de uma grande base unificada contendo dados importantes de todos os cidadãos.

Apesar de existirem algumas vozes que se insurgiram contra essas movimentações, pode-se afirmar que o tema não ganhou a mesma repercussão negativa que vivenciou a França e os EUA nos casos do *National Data Center* (NDC), e Safari. De qualquer forma, o Brasil vivencia uma realidade própria, já que a divisão federativa entre União, Estados e Municípios têm causado maiores dificuldades na troca de dados entre os entes públicos e entre os entes e os cidadãos, principalmente no que tange à garantia da confiabilidade e integridade dessas informações. Podemos afirmar que o Brasil ainda caminha para uma gestão digital dos dados e que as falhas nos bancos de dados dos entes dificultam a obtenção de informações até mesmo pelo próprio Estado, e, consequentemente, o seu repasse aos cidadãos.

Mas para além das bases de dados estatais, a crescente demanda por transparência dos dados governamentais, para *accountability* também culminou em diversas inovações legislativas, que podemos classificar como uma **segunda fase**.

A lei Complementar 131/09, por exemplo, também conhecida como "lei da transparência", tornou acessível informações sobre a execução orçamentária e financeira do governo, lançando à rede uma grande quantidade de dados que estariam restritos aos órgãos públicos.

---

18 Conforme art. 16, III.
19 Conforme art. 16, V.
20 Conforme art. 17.
21 Conforme art. 18, §6º.
22 Conforme art. 18, § 2º.
23 Conforme art. 17.

O acesso a informações pessoais também foi garantido pela via constitucional, através do *Habeas Data,* que foi posteriormente regulamentado pela a Lei 9.507/1997.

Já a Lei nº 12.527/2011, conhecida como Lei de acesso à Informação (LAI), foi um marco na disponibilização em larga escala aos cidadãos de dados armazenados pelo Poder público e por entidades privadas que recebem verba pública. A LAI trouxe uma ampliação de mecanismos de acesso, o que possibilitou não só o acesso pelos cidadãos, mas também o compartilhamento de dados entre os setores públicos e entre os setores público-privado. A lei prevê a obtenção de informação como direito fundamental, tendo como diretrizes princípios como o da transparência, e estabelecendo que a regra seria a publicidade, enquanto que o sigilo seria excepcional.

Importante ressaltar que, mesmo quando há restrições de acesso decorrentes do sigilo, se houver a necessidade das informações para preservação de direitos fundamentais, o acesso teria respaldo legal, conforme art. 21 da LAI.[24]

Aponta, portanto, o direito à informação não como um direito negativo do Estado frente à privacidade, mas como direito positivo de prestar contas e de visibilidade.[25] Nesse sentido, o art. 5º:

> "Art. 5º É dever do Estado garantir o direito de acesso à informação, que será franqueada, mediante procedimentos objetivos e ágeis, de forma transparente, clara e em linguagem de fácil compreensão
> Parágrafo único. As informações ou documentos que versem sobre condutas que impliquem violação dos direitos humanos praticada por agentes públicos ou a mando de autoridades públicas não poderão ser objeto de restrição de acesso.".[26]

A Lei dispôs sobre o acesso tanto de dados pessoais como de dados não pessoais em poder de entes públicos ou entidades com recursos públicos. Apesar de o enfoque da lei ser a disponibilização e não as formas sobre como esses dados seriam utilizados pelo Poder Público e pelo Cidadão que peticionou o acesso, a LAI traz algumas diretrizes no tratamento dessas

---

24 "Art. 21. Não poderá ser negado acesso à informação necessária à tutela judicial ou administrativa de direitos fundamentais".

25 Conforme teor do art. 3º: "Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes:
I - observância da publicidade como preceito geral e do sigilo como exceção;
II - divulgação de informações de interesse público, independentemente de solicitações;
III - utilização de meios de comunicação viabilizados pela tecnologia da informação;
IV - fomento ao desenvolvimento da cultura de transparência na administração pública;
V - desenvolvimento do controle social da administração pública."

26 No mesmo sentido, o art. 8º da LAI: "É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas".

informações. Vale mencionar que a lei conceituou como tratamento de informação, o "conjunto de ações referentes à produção, recepção, classificação, utilização, acesso, reprodução, transporte, transmissão, distribuição, arquivamento, armazenamento, eliminação, avaliação, destinação ou controle da informação".[27]

Como exemplo de diretriz, a LAI determina que o tratamento de dados sigilosos será definido por regulamento[28] (o que foi posteriormente definido no Decreto 7.845/2012), além de definir em que ocasiões o dado pode ser considerado sigiloso[29], quando então a restrição de acesso se torna imprescindível à segurança da sociedade ou do Estado.

Com relação às informações pessoais, a LAI prevê também a incidência do princípio da transparência, afirmando que as informações pessoais deveriam ser manipuladas, "com respeito à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais"[30]. Dessa forma, a lei prevê uma proteção especial para algumas informações pessoais, quais sejam, aquelas relativas à intimidade, vida privada, honra e imagem, que terão seu acesso restrito, ante a proteção de inviolabilidade do art. 5º, X da Constituição, conforme teor do art. 31, § 1º:

> Art. 31. § 1º As informações pessoais, a que se refere este artigo, relativas à intimidade, vida privada, honra e imagem:
> I - terão seu acesso restrito, independentemente de classificação de sigilo e pelo prazo máximo de 100 (cem) anos a contar da sua data de produção, a agentes públicos legalmente autorizados e à pessoa a que elas se referirem; e
> II - poderão ter autorizada sua divulgação ou acesso por terceiros diante de previsão legal ou consentimento expresso da pessoa a que elas se referirem.

Percebe-se que o consentimento já era aventado como uma das formas válidas de disposição dos próprios dados, na linha seguida pela terceira geração das legislações de proteção de dados. No entanto, a própria Lei já ressalva situações em que o consentimento não seria necessário, como descrito no art. 31 §3º:

> § 3º O consentimento referido no inciso II do § 1º não será exigido quando as informações forem necessárias:

27 Conforme art. 4º, V.
28 Conforme teor: "Art. 25. É dever do Estado controlar o acesso e a divulgação de informações sigilosas produzidas por seus órgãos e entidades, assegurando a sua proteção. § 1º O acesso, a divulgação e o tratamento de informação classificada como sigilosa ficarão restritos a pessoas que tenham necessidade de conhecê-la e que sejam devidamente credenciadas na forma do regulamento, sem prejuízo das atribuições dos agentes públicos autorizados por lei. § 2º O acesso à informação classificada como sigilosa cria a obrigação para aquele que a obteve de resguardar o sigilo. § 3º Regulamento disporá sobre procedimentos e medidas a serem adotados para o tratamento de informação sigilosa, de modo a protegê-la contra perda, alteração indevida, acesso, transmissão e divulgação não autorizados".
29 Conforme art. 23 da LAI.
30 Conforme art. 31, caput.

> I - à prevenção e diagnóstico médico, quando a pessoa estiver física ou legalmente incapaz, e para utilização única e exclusivamente para o tratamento médico;
> II - à realização de estatísticas e pesquisas científicas de evidente interesse público ou geral, previstos em lei, sendo vedada a identificação da pessoa a que as informações se referirem;
> III - ao cumprimento de ordem judicial;
> IV - à defesa de direitos humanos; ou
> V - à proteção do interesse público e geral preponderante.

A LAI determina ainda as sanções aplicáveis a quem tenha acesso a informação sigilosa ou pessoal e a submeta a tratamento indevido[31], e previu a existência de um Regulamento sobre os procedimentos para tratamento de informação pessoal. A regulamentação, no entanto, para o tratamento pelo setor privado viria apenas anos depois sob a forma da Lei de Proteção de Dados Pessoais, como falaremos mais adiante.

Os efeitos da LAI são concretizados nas políticas de "dados abertos governamentais",[32] que tem favorecido cada vez mais o acesso de dados públicos pelas redes, fazendo com que a informação esteja disponível de forma mais amplamente acessível.

O Decreto nº 8.777/16 dispõe que o "dado acessível ao público" é qualquer dado gerado ou acumulado pelo Governo, excepcionando o núcleo de dados pessoais referente à intimidade, vida privada, honra e imagem, protegido pela Lei nº 12.527/11 e considerado não público. Por sua vez, os "dados abertos" são conceituados pelo Decreto nº 8.777/16 como "dados acessíveis ao público, representados em meio digital, estruturados em formato aberto".

Os dados abertos são disponibilizados pelo setor público "em sua forma primária, com o maior grau de granularidade possível", ou referenciando "as bases primárias, quando disponibilizadas de forma agregada", respeitando as diretrizes de completude e interoperabilidade.[33] Isso torna possível a livre utilização desses dados, tanto pelos Poderes Públicos quanto pelos mais diversos atores sociais, conforme disposto no art. 4º do mencionado Decreto. Dessa forma, é possível alimentar bases de dados privadas com os dados públicos obtidos.

---

31 Conforme art. 34: "Os órgãos e entidades públicas respondem diretamente pelos danos causados em decorrência da divulgação não autorizada ou utilização indevida de informações sigilosas ou informações pessoais, cabendo a apuração de responsabilidade funcional nos casos de dolo ou culpa, assegurado o respectivo direito de regresso. Parágrafo único. O disposto neste artigo aplica-se à pessoa física ou entidade privada que, em virtude de vínculo de qualquer natureza com órgãos ou entidades, tenha acesso a informação sigilosa ou pessoal e a submeta a tratamento indevido."

32 O termo é conceituado pelo escritório do W3C no Brasil como "a publicação e disseminação das informações do setor público na Web, (...) de modo a permitir sua reutilização em aplicações digitais desenvolvidas pela sociedade" (W3C BRASIL, 2018). Acrescenta-se que são dados denominados "não-discriminatórios", uma vez que não exigem qualquer tipo de cadastro ou requerimento para o seu acesso (THE ANNOTATED 8 PRINCIPLES OF OPEN GOVERNMENT DATA, 2018).

33 Conforme art. 3º, V, do Decreto nº 8.777/16.

Todas essas iniciativas legislativas tiveram por resultado um acesso muito mais simplificado aos dados, o que repercute de forma direta nos sistemas de anonimização, como veremos adiante. Podemos introduzir, no entanto, que quanto mais dados estão disponíveis nas redes, maiores as dificuldades a serem enfrentadas nesse tipo de tratamento.

Esse cenário se agrava quando pensamos, para além dos dados disponibilizados publicamente, nas iniciativas de compartilhamento de dados entre entidades públicas entre si e entre entidades públicas e privadas.[34] Nesse sentido, alguns atores, através desses acordos, conseguem informações privilegiadas para os seus próprios negócios, através do tratamento desses dados públicos obtidos.

Por outro lado, ressaltamos que o debate sobre a proteção de dados também tem sido fomentado no país, o que apontamos como uma **terceira fase** de maior destaque atualmente. No âmbito privado, as primeiras discussões sobre proteção de dados ocorreram com a judicialização de demandas envolvendo a privacidade de dados. Houve largo debate sobre a aplicabilidade das leis consumeristas aos cadastros negativos de crédito, para fins de retificação de informações que prejudicavam clientes. Começou a se cogitar a privacidade não mais como o direito de ser deixado só, mas como o consentimento para divulgação da informação pessoal (CUEVAS, 2019).

No âmbito do STJ, em 2001, a Ministra Eliana Calmon reconheceu o direito à privacidade dos dados aos contribuintes ou titulares de contas bancárias em relação aos seus dados pessoais (REsp 306.570), e em 2010 o Ministro Luis Felipe Salomão asseverou que o consentimento seria um requisito para a divulgação da informação pessoal, como respeito a sua privacidade e ao direito do indivíduo de dispor de suas próprias informações. (REsp 1.168.547/RJ) (CUEVAS, 2019).

A Lei Federal nº 12.414/2011 surgiu garantindo ainda maiores direitos sobre os próprios dados que o Código de Defesa do Consumidor, inclusive os critérios legais para análise dos riscos. A lei previu a criação e a consulta a bancos de dados com informações sobre o adimplemento de pessoas naturais ou jurídicas, para formação de histórico de crédito. No entanto, a norma estabelecia como vedada a inclusão de informações excessivas, ou seja, não vinculadas à análise de risco de crédito ao consumidor; e as informações sensíveis, que são

34 Um caso de compartilhamento deu ensejo à investigação pelo MPDFT (Ministério Público do Distrito Federal e Territórios). Trata-se do suposto pagamento de quantias vultosas pela administração à empresa pública Serpro (Serviço Federal de Processamento de Dados) a fim de adquirir dados, inclusive pessoais, de forma organizada (CRUZ, 2018).Por sua vez, o compartilhamento de dados entre os setores público e privado ganhou destaque no Acordo de Cooperação Técnica 7/2013 do TSE, que envolveu a tentativa de acordo para compartilhamento de dados entre o Tribunal Superior Eleitoral e o Serasa EXPERIAN, o qual foi interrompido por decisão judicial (HAIDAR, 2013).

aquelas relacionadas à origem social e étnica, à saúde, à informação genética, à orientação sexual e às convicções políticas, religiosas e filosóficas, conforme art. 3º, §3º, I, II.

Outra legislação muito importante para fixar as diretrizes quanto a utilização dos dados e os seus limites no setor privado se deu com a instituição do Marco Civil da Internet, Lei nº 12.965/14, que se debruçou sobre as implicações jurídicas do novo meio digital.

O Marco Civil da Internet, Lei 12.965/2014, buscou estipular parâmetros de direitos e deveres no uso da internet, assegurando o direito à privacidade e à liberdade de expressão como condições ao pleno exercício de acesso à internet[35].

A lei estipula a proibição do fornecimento a terceiros de dados pessoais dos usuários da internet, exceto se houvesse o consentimento livre, expresso e informado, ou nas hipóteses previstas em lei[36]. Além disso, a Lei já prevê algumas diretrizes na utilização de dados pessoais, já que exige, além do consentimento expresso,[37] a clareza de informação sobre a coleta, uso, armazenamento, tratamento e proteção de dados pessoais. Os dados somente poderiam ser utilizados para finalidades que justificassem sua coleta[38]; que não fossem vedadas pela legislação e estivessem especificadas em contrato ou termo de uso[39]. Também prevê a exclusão dos dados pessoais após o término da relação das partes[40]. Importante mencionar ainda a previsão de sanções aos descumprimentos das regras sobre dados pessoais estabelecidas na lei, o que inclui multa, suspensão de atividades ou proibição do exercício de atividades no Brasil.

No entanto, mesmo com as previsões mais amplas do Marco Civil da Internet, o crescente uso dos dados e o valor econômico, social e político a eles conferidos na *Data-Driven Society*, a necessidade de uma legislação específica de proteção de dados se tornou cada vez mais urgente. Nesse sentido, surgem os esforços para a recente Lei Geral de Proteção de Dados.

2.4. A Lei Geral de Proteção de Dados Brasileira- LGPD

A Lei Geral de Proteção de Dados- LGPD, Lei nº 13.709/18, aprovada em 14 de agosto de 2018, contou com participação popular no anteprojeto de lei, promovida pela Secretaria de Assuntos Legislativos do Ministério da Justiça e também com audiências públicas durante sua tramitação.

35 Conforme art. 8º.
36 Conforme art. 7ª, VII.
37 Conforme art. 7º, IX.
38 Conforme art. 16, II.
39 Conforme art. 7º, VIII.
40 Conforme art. 7ª, X.

A lei dispõe sobre as especificidades dos dados pessoais, já que até então a legislação nacional previa uma proteção mais acentuada apenas para dados governamentais considerados sigilosos. Assim, a LGPD, visa à proteção dos indivíduos por trás dos dados. Ressalte-se que a LGPD é considerada de interesse nacional e deve ser observada pela União, Estados, DF e Municípios.[41]

Em muitos dispositivos da lei é possível se observar a influência do Regulamento Europeu de 2016 (GDPR), conforme o quadro comparativo que estabelecemos quanto a alguns conceitos que serão abordados nesta pesquisa:

Princípios/ Institutos	**Previsão Legal (LGPD)**	**Previsão Legal (GDPR)**
Conceito do Princípio da Necessidade ou "*Data minimisation*"	Art. 6º, III.	Art. 5(1)(c)
Conceito do Princípio da Finalidade	Art. 6º, I	Art. 5º (1)(b)
Conceito do princípio da responsabilização e prestação de contas/ princípio da Responsabilidade	Art. 6º, X	Art. 5º (2)
Conceito de Legítimo Interesse	Arts. 7º, IX, 10	Art. 6 (1)(f); "Considerandos" precedentes ao texto do GDPR, ponto 47
Conceito de Anonimização	Art. 5, XI.	"Considerandos" precedentes ao texto do GDPR, ponto 26; Art. 4(5)(pseudonimização)
Conceito de Dado Pessoal	Art. 5, I	Art. 4(1)
Exclusão dos dados anônimos do conceito de dado pessoal	Art. 12	"Considerandos" precedentes ao texto do GDPR, ponto 26.
Conceito de Tratamento	Art. 5, X.	Art. 4(2)
Dados pseudonimizados como dados pessoais	Art. 14, § 4º	"Considerandos" precedentes ao texto do GDPR, ponto 26.

Figura 1- Tabela comparativa LGPD- GDPR

Voltaremos a esta tabela comparativa no decorrer da explanação à medida que os temas de interesse sejam destacados.

41 Conforme art. 1º, Parágrafo Único, da LGPD.

De maneira geral, podemos afirmar que a LGPD, assim como o GDPR, seguindo as tendências da terceira onda legislativa, se preocupa com os dados pessoais sobre o enfoque do indivíduo e sua personalidade, conforme se observa do art. 1º caput da Lei:

> Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

Há, inclusive, expressa previsão de que um dos fundamentos da proteção de dados pessoais é a autodeterminação informativa (art. 2º, II, da LGPD), aliada a outros fundamentos, como o respeito à privacidade; a liberdade de expressão; a inviolabilidade da intimidade, da honra e da imagem; os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais dentre outros.

A LGPD conceitua, em seu art. 5º, o que seriam dados pessoais, como sendo a "informação relacionada a pessoa natural identificada ou identificável" (art. 5º, I), de forma semelhante ao que prevê o GDPR[42], seguindo o conceito amplo de dados pessoais que já mencionamos. Além disso, a lei prevê uma proteção ainda maior para dados pessoais sensíveis, que são dados pessoais "sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural" (art. 5º, II).

A LGPD protege dados pessoais em quaisquer das formas de sua utilização, como se observa da própria descrição de tratamento de dados, no art. 5º, X, da LGPD, que define tratamento como:

> toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;

A LAI, conforme mencionamos, prevê um conceito semelhantemente amplo de tratamento de informações, quanto aos dados utilizados pelo poder público[43]. O conceito de tratamento também é semelhante à previsão no GDPR[44].

---

42 Conforme art. 4(1), apontado na Tabela comparativa (Figura 1).
43 Conforme art. 4º, V da LAI.
44 Conforme art. 4(2) do GDPR descrito na Tabela comparativa (Figura 1).

A LGPD define ainda os agentes de tratamento[45], quais sejam, o controlador e o operador, como, respectivamente, a pessoa a quem competem as decisões referentes ao tratamento de dados pessoais; e a pessoa que realiza o tratamento de dados pessoais em nome do controlador.[46]

Os agentes de tratamento se comprometem, sob pena de responsabilidade, a adotar medidas de "segurança, técnicas e administrativas para proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito"[47].

Somando-se a esses dois agentes, há a figura do encarregado, que é a pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD)[48].

A lei ainda traz um capítulo específico sobre o uso de dados pessoais pelo poder público, em observância a já mencionada lei de acesso à informação (LAI). Nesses casos, a LGPD prevê que os dados devem ser mantidos em formato interoperável e estruturado, para viabilizar o uso compartilhado para fins de políticas públicas, serviços públicos, descentralização da atividade pública e disseminação e acesso das informações.[49] O poder público pode compartilhar esses dados já estruturados para entidades privadas, quando esses dados forem acessíveis publicamente[50], o que facilita bastante a formação e o fomento de bases de dados privadas. Os dados também poderão ser compartilhados mediante contratos, convênios ou instrumentos congêneres, desde que comunicados à autoridade nacional. [51]

Por fim, sem a pretensão de esgotar os assuntos tratados pela LGPD, enfatizamos que a lei prevê algumas exceções às hipóteses taxativas de tratamento de dados pessoais, expressas no art. 7º, caput.

A primeira delas é o tratamento de dados pessoais públicos, ou tornados públicos pelo titular. Quanto aos dados pessoais públicos, a lei permite, no art. 7º, § 3º, que o tratamento seja realizado mesmo que não preenchidos os requisitos ordinários de tratamento, mas desde que considere o princípio da finalidade, a boa-fé e o interesse público que justificaram a disponibilização do dado. Já os dados tornados públicos pelo titular podem ser tratados com

45 Conforme art. 5º, IX.
46 Conforme art. 5º, VI, VII.
47 Conforme art. 46.
48 Conforme art. 5º, VIII.
49 Conforme art. 25.
50 Conforme art. 26, §1º, III.
51 Conforme art. 26, §1º, V, e § 2º.

dispensa de consentimento, desde que respeitados os princípios gerais dispostos na lei e os direitos do titular [52].

Para ambos os casos, a lei flexibilizou o princípio da finalidade, que é o princípio que vincula o tratamento do dado à finalidade previamente estabelecida pelos agentes de tratamento. Essa flexibilização se dá uma vez que a lei definiu como possível o tratamento posterior desses dados para novos objetivos (que não aqueles previamente estabelecidos), sendo que nesses casos a lei exige apenas que sejam "observados os propósitos legítimos e específicos para o novo tratamento", além dos direitos do titular, os fundamentos e princípios da lei[53].

Ressaltamos ainda, que a lei prevê esses limites mais fluidos para o tratamento de dados pessoais públicos, mas não há nenhuma restrição expressa na LGPD ao tratamento de dados públicos não pessoais. Ou seja, quando há o tratamento de dados pessoais públicos, há um abrandamento do rigor quanto aos limites do tratamento, mas para os dados públicos não pessoais não há nenhuma restrição legal específica.

A lei também não estabelece entraves ao tratamento de dados anonimizados, uma vez que não os considera como dados pessoais. De fato, assim como ocorre no GDPR[54], a lei prevê que uma vez anonimizados esses dados perdem a característica de dados pessoais. É o que afirma o art. 12 da LGPD:

> Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.

A LGPD excepciona apenas os casos em que a anonimização for revertida, ou quando, "com esforços razoáveis", puder ser revertida.

A lei não determina exatamente o que seria um esforço razoável, mas prevê que ele deverá ser mensurado através de fatores objetivos, como custo e tempo necessários para a reversão, as técnicas disponíveis ao tempo da análise e a utilização exclusiva de meios próprios para a reversão[55]. Veremos com maior profundidade esses requisitos no capítulo que propomos o framework legal da anonimização contido na LGPD.

Dessa forma, para ambos, dados públicos e anonimizados, o tratamento dispensa, pela ótica da LGPD, o estrito cumprimento dos rígidos requisitos impostos pela lei para

52 Conforme art. 7º, §4º.
53 Conforme art. 7º, §7º.
54 Conforme os "Considerandos" precedentes ao texto do GDPR, ponto 26, descrito na Tabela comparativa (Figura 1).
55 Conforme art. 12, § 1º.

compliance do tratamento. Essa saída legal acabou por abarcar algumas das expectativas dos mercados, principalmente no que tange ao tratamento de dados massivos. Isso porque, principalmente quanto ao *Big Data*, o *compliance* restrito à lei por vezes acarretaria grandes dificuldades ou mesmo inviabilizaria o negócio.

Principalmente a anonimização foi considerada por esses *players* como uma possível alternativa a seguir para manutenção de suas atividades sem que houvesse o descumprimento normativo.

Ressaltamos que o uso das técnicas de anonimização para proteção de dados pessoais é o enfoque deste trabalho, especialmente no contexto de plataformas de *Big Data*, e, por isso, debruçar-nos-emos nas especificidades deste tipo de tratamento e na confiança depositada nesse tipo de técnica.

2.5. Natureza Jurídica dos dados pessoais e dos dados anonimizados (Dados Pessoais entre propriedade, personalidade e fraternidade)

Uma das questões que seguem em debate, nacional e internacionalmente, é a discussão acerca de que natureza jurídica dos dados pessoais teria prevalecido na atual fase legislativa. A insuficiência de se conferir apenas um caráter patrimonial aos dados através da tutela por meio do direito de propriedade tem sido levantad por diversos autores. Isso porque seria necessário se alcançar através das legislações o caráter de direitos pessoais desses ativos, reconhecendo-os como parcela da personalidade dos indivíduos.

Por outro lado, a tutela por meio dos direitos de propriedade tem destaque histórico na proteção de bens, sendo a forma de maior proteção das relações jurídicas. Dessa forma, a necessidade de se conferir uma melhor proteção a esses ativos vem ainda hoje dividindo opiniões acerca da melhor forma de tutela desses bens.

Roberta Mauro explica esse debate, afirmando que as teorias que fomentaram o enquadramento dos dados pessoais como propriedade tiveram como finalidade o estabelecimento da relação de pertencimento do objeto para com o sujeito de direito além da proteção patrimonial desses ativos, uma vez que sobre os direitos reais incide uma maior proteção jurídica. Os direitos reais, incluindo o direito de propriedade, são direitos opostos *erga omnes* pelo sujeito, ou seja, os efeitos do direito real recaem sobre toda a coletividade, que não pode interferir na relação entre o sujeito de direitos e o seu objeto. Disso decorre, por exemplo, o direito de sequela, ou seja, o direito de recuperar a coisa das mãos de quem injustamente a

possua. (MAIA, 2019). Assim, normalmente, os direitos reais concentram-se na tutela patrimonial dos bens.

Já os direitos de personalidade, por vezes, se concentram na tutela extrapatrimonial desses bens, fazendo com que a proteção jurídica resulte na indenização por perdas e danos em caráter repressivo, e, portanto, não possuam eficácia direta e imediata.

Entretanto, segundo a autora, no contexto brasileiro, haveria uma superação prática da discussão entre direito de propriedade e direito de personalidade, ao menos no que tange à tutela patrimonial dos dados.

No direito brasileiro, os direitos de personalidade ganham caráter de direitos fundamentais e recebem tutela específica com oposição *erga omnes*, à semelhança dos direitos reais. Dessa forma, segundo a autora, é garantido aos direitos de personalidade a tutela patrimonial e extrapatrimonial desses bens.

A tutela patrimonial estaria vinculada ao "lucro da intervenção", ou seja, à proteção contra o aumento patrimonial obtido indevidamente pelo sujeito que não é titular dos dados e nem tem o aval do titular para exploração desses bens. Por sua vez, a tutela extrapatrimonial continuaria vinculada à reparação por perdas e danos.

Nesse sentido, a autora entende como correta a posição do legislador que se absteve de tratar os dados pessoais como direito de propriedade, mas os caracterizou como parcela de personalidade, tratando os sujeitos como titulares desse bem jurídico. É o que descreve o art. 17 da LGPD, ao prever que "toda pessoa natural tem assegurada a titularidade de seus dados pessoais e garantidos os direitos fundamentais de liberdade, de intimidade e de privacidade, nos termos desta Lei".

A autora defende então, com base nas lições de Pietro Perlingieri, que a titularidade seria um conceito jurídico mais amplo do que a propriedade, e que a propriedade seria uma das espécies possíveis de titularidade. Por sua vez, a titularidade alcançaria outras nuances de proteção da relação de pertencimento entre o sujeito de direitos e o bem, para além da esfera patrimonial conferida pelo direito de propriedade.

Nas palavras da autora:

> O conceito de titularidade exprime, portanto, não apenas a ideia de poder de controle sobre um bem jurídico mas, também e consequentemente, o sentido de atribuição do mesmo, com regras claras disponíveis acerca de seus modos de utilização e disposição. Se dados pessoais são hoje bem jurídico – daí a inequívoca necessidade de tutelá-los –, precisava o legislador determinar a quem pertencem, fosse acerca de seus aspectos extrapatrimoniais – principal justificativa da crítica dirigida por Daniel Solove aos que enxergavam a privacidade dos dados pessoais apenas como o objeto de um direito de

> propriedade –, fosse relativamente a seus aspectos patrimoniais, decorrentes do valor econômico que lhes foi atribuído pela sociedade digital.

Em concordância com Solove (SOLOVE, 2008, p. 26-27)[56], a autora entende que a proteção puramente patrimonial que traria o direito de propriedade seria insuficiente, já que não abarcaria a complexidade desses bens jurídicos. No mesmo sentido se posiciona Danilo Doneda, como mencionamos anteriormente (DONEDA, 2006, p.165-168). Portanto, para a autora, a escolha do legislador em enquadrar esses bens como pertencentes aos direitos de personalidade, envoltos sobre o caráter de titularidade conferiria uma maior e melhor proteção aos direitos sobre dados pessoais.

Como visto, a terceira onda legislativa fez com que prevalecesse a tese de que os dados pessoais seriam bens jurídicos ligados aos direitos da personalidade, se distanciando das teorias puramente patrimonialistas.

Se considerada essa perspectiva de direitos de personalidade, os dados anonimizados parecem deslocados desse tipo de proteção jurídica. Isso porque os dados anonimizados, segundo o conceito legal, não estariam mais vinculados ao seu titular. Portanto, uma vez desvinculados de seu titular, seria aparentemente desafiador defender a proteção jurídica voltada aos direitos de personalidade para esses dados. Afinal de contas, de quem seria a personalidade a ser protegida se, pelo conceito legal, o dado anonimizado é justamente aquele desvinculado de seu titular?

Entretanto, como veremos, a técnica possui limites a essa completa e total desvinculação. Além disso, é importantíssimo ressaltar que o dado anonimizado mantém seu valor de mercado justamente pela veracidade do dado e pela possibilidade de sua utilização. Portanto, ainda que a ligação à pessoa natural não seja possível de forma imediata, ainda assim o dado só é valioso quando se refere a um hábito, uma característica ou um acontecimento vinculado a uma pessoa real, não se tratando de dado fictício ou meramente aleatório.

Nesse sentido, ainda que a manutenção da proteção à personalidade pareça contraintuitiva para os dados anônimos, para nós, ela continua sendo uma das possibilidades de atribuição de natureza para esses bens jurídicos, ainda que ressalvadas suas peculiaridades.

---

56 O autor retrata as dificuldades em se conferir uma tutela puramente patrimonial ou puramente extrapatrimonial aos dados pessoais: “At other times, the privacy problem at issue is misconstrued. For example, identification is often understood as a harm created by revealing one’s name, but the essence of the problem is being linked to a stream of data, not only a name. Insecurity is often not adequately addressed by the law because a materialized harm has not yet occurred. But insecurity remains a problem even where there has been no actual disclosure or leakage of embarrassing details. Appropriation is understood primarily as a harm to property interests, and its dignitary dimensions are thus frequently ignored by courts. Further complicating matters is the fact that privacy problems are inconsistently recognized across different areas of the law. For example, tort law readily recognizes and redresses breach of confidentiality, but Fourth Amendment law ignores it” (SOLOVE, 2008, p. 188).

Destaque-se ainda que a Teoria Personalista do Direito Civil, a qual trata da relação apenas de direitos pessoais (relações intersubjetivas) e de direitos reais, teve suas diretrizes flexibilizadas para a atribuição de personalidade também às pessoas jurídicas. Portanto, não seria a primeira vez que o próprio direito flexibiliza a necessidade de que atributos de personalidade estejam necessariamente relacionados a uma pessoa física.

Feita a crítica quanto à pretensa impossibilidade de aplicação da natureza de personalidade para os dados anônimos, destacamos, no entanto, que não tem prevalecido a atribuição de características de personalidade para esses dados.

Uma vez enfraquecida a vertente da personalidade, as demais teses acerca da natureza jurídica desses bens, seja da patrimonialização, seja da objetificação desses dados, ganham destaque. Surge, por exemplo, a possibilidade de se considerar o dado uma vez anonimizado como *res develicta*, ou seja, como coisa que pode ser apropriada, por ser "abandonada", uma vez presente a perda da atribuição de titularidade.

Tanto a patrimonialização quanto a objetificação fortalecem os argumentos de livre utilização dos dados anônimos, o que aumenta os riscos não apenas da reidentificação, mas de abusos no uso desses dados. A utilização irresponsável desses dados, como *res* qualquer, pode trazer graves implicações de justiça social e aprofundamentos de desigualdades.

Por fim, destacamos ainda que um campo pouco explorado no que tange aos dados anônimos é a possibilidade de atribuição de natureza difusa ou coletiva, a depender do caso, para esses bens. Justamente em consonância com as discussões contemporâneas da quarta onda legislativa, temos cada vez mais clareza na percepção de que os dados pessoais não envolvem apenas um indivíduo, mas por vezes atribuem características por inferência a toda uma coletividade. Nesse sentido, o dado anônimo, de forma mais abstrata do que o pessoal, fornece informações verídicas, que podem pertencer, se não a um indivíduo específico, a uma categoria analisada.

Todas essas questões revelam o quanto a discussão sobre a natureza jurídica desses dados ainda necessita aprofundamento, e principalmente, que pensar em formas alternativas de classificação pode, por fim, viabilizar ou inviabilizar a proteção jurídica desses dados.

2.6. O que é anonimização?

As legislações de proteção de dados, de maneira geral, surgiram como um verdadeiro desafio para adequação de práticas até então lícitas, mas que agora, por exigência legal, tinham que cumprir uma série de requisitos, ou simplesmente eram vedadas.

Muitos dos dispositivos legais iam na contramão da política estratégica de diversas instituições, que viam na captação desenfreada de dados (inclusive dados pessoais) uma forma de negócio. Por vezes, o ímpeto pela captação da maior quantidade de dados possível fez surgir bases de dados massivas, desordenadas, o que foi chamado pela literatura de "*data swamp"* ou "pântano de dados" (BARBIERI, 2020). Afinal de contas, os dados armazenados, ainda que não tivessem proveito imediato, poderiam ser utilizados em estratégias futuras, por meio de ferramentas de *Big Data Analytics.*

Com a regulamentação do uso de dados pessoais essas práticas se tornaram ilegais, já que a utilização desses dados deveria estar vinculada a uma finalidade específica e expressa, por um tempo determinado, e por vezes requerendo o consentimento do titular.

Nesse sentido, até mesmo a manutenção de grandes bancos de dados contendo dados pessoais passou a sofrer entraves. Primeiro, porque o tratamento jurídico diferencia, como vimos, dados pessoais dos dados não considerados pessoais, fazendo com que a gestão desses dados ganhe contornos muito distintos. Segundo, porque a própria obtenção e armazenamento dos dados pessoais passou a ter que obedecer a uma série de diretrizes legais, conforme mencionado. Dessa forma, particularmente as plataformas de *Big Data* viram seu funcionamento sensivelmente ameaçado.

Por outro lado, a confiança depositada tanto pelo GDPR quanto pela LGPD nas ferramentas de anonimização abrem espaço para a continuidade das plataformas de *Big Data* a partir da anonimização dos dados. Nesse sentido, a anonimização surge como uma possível saída para a manutenção dessas bases sem que haja uma interferência legal severa em seu funcionamento.

Conceitualmente, pela LGPD, dados anonimizados são aqueles relativos ao "titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento" (art. 5º, III). Como mencionamos, os dados anonimizados não são considerados como dados pessoais pela LGPD e, por isso, estão liberados das regras restritivas impostas pela lei de proteção.

O regulamento europeu descreve de forma similar que os dados anonimizados não são considerados pessoais e, portanto, não estão sujeitos às restrições estipuladas pelo regulamento, conforme descrito no ponto 26 dos "considerandos" precedentes ao texto do GDPR:

> Os princípios da proteção de dados não deverão, pois, aplicar-se às informações anónimas, ou seja, às informações que não digam respeito a uma pessoa singular identificada ou identificável nem a dados pessoais tornados de

> tal modo anônimos que o seu titular não seja ou já não possa ser identificado. O presente regulamento não diz, por isso, respeito ao tratamento dessas informações anônimas, inclusive para fins estatísticos ou de investigação.

Conforme se depreende dos dispositivos, na anonimização o tratamento é voltado para a dissociação entre dados e seus titulares, ainda que esse processo se dê de formas diversas. Nesse sentido, a anonimização pode ter sua expressão de forma *ex ante*, pelo usuário ou pela arquitetura no momento da própria formação e captura dos dados, ou *ex post*, situação em que o dado será dissociado de seu titular após a sua captura, quando já estruturado em bancos de dados implicando níveis de abstração diferentes.

A abstração é um conceito muito utilizado no âmbito da programação orientada a objetos (DALL'OGLIO, 2007) e apresenta que a formação de modelos e mesmo a definição do objeto implica em uma redução da realidade. Nas palavras do autor:

> No paradigma de orientação a objetos se prega o conceito da 'abstração'. De acordo com o dicionário *Priberam*, 'abstrair' é separar mentalmente, considerar isoladamente, simplificar, alhear-se. Para construir um sistema orientado a objetos, não devemos projetar o sistema como se fosse uma grande peça monolítica; devemos separá-lo em partes, concentrando-nos nas peças mais importantes e ignorando os detalhes (em um primeiro momento) para podermos construir peças bem definidas que possam ser reaproveitadas depois, formando uma estrutura hierárquica. (DALL'OGLIO, 2007, p. 103)

A abstração é normalmente aliada à modularidade, ou seja, à repartição da realidade em módulos menores, tornando a abstração mais organizada e permitindo a interação entre os diversos módulos. Dessa forma, a abstração cria um gap semântico (FALBO & SOUZA, 2006) entre o mundo real e o objeto de estudo, que se apresentará de forma simplificada para permitir a formação de modelos.

A anonimização *ex post* pode ser associada a um maior nível de abstração, já que é realizada em ambientes controlados, como no caso de bancos de dados, facilitando a criação de modelos e o gerenciamento dos dados. Por esse motivo, esse tipo de técnica tem um maior gap semântico frente à realidade, já que o objeto se restringe a um recorte específico.

Já no caso da anonimização *ex ante*, temos a aplicação da técnica de anonimização durante a produção de dados no ambiente digital. A abstração nesses casos é menor, já que há maior interação entre os dados gerados pelo titular anônimo e a rede, não restrita a um banco de dados específico. Nesses casos, o objeto se aproxima mais da realidade, diminuindo o gap semântico. Nada impede, no entanto, que dados gerados por usuários anônimos, ou seja, anonimizados *ex ante*, sejam armazenados em bancos de dados estruturados pelas plataformas.

Isso tem sido cada vez mais comum, como aponta Ramakrishnan e Gehrke[57] (RAMAKRISHNAN, R.; GEHRKE, 2003).

Portanto, ressaltamos que o enfoque aqui é a anonimização *ex post*, ou seja, relacionada aos bancos de dados já formados, como forma de *compliance* dos sistemas já construídos às legislações protetivas. Trataremos então da anonimização *ex ante* apenas para fins exemplificativos no capítulo acerca da ambiguidade da ferramenta, para melhor compreensão do dilema abordado, mas o objeto deste trabalho serão os dados já estruturados em bancos de dados.

A LGPD conceitua anonimização como a "utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo", conforme art. 5º, XI.

Para nós, ao menos para fins desta pesquisa, a anonimização pode ser conceituada como uma técnica de tratamento de dados pessoais, na modalidade "processamento de dados", cujo objetivo é a dissociação de dados (dispostos em banco de dados, públicos ou privados, organizados ou desorganizados) dos seus respectivos titulares, considerando meios legais de razoabilidade. Da forma semelhante, para fins dessa pesquisa, o dado anônimo pode ser definido como o dado pessoal que em um determinado momento e em um determinado contexto pode ser considerado dissociado de seu titular, deixando sua "pessoalidade" para ganhar o *status* de anonimizado enquanto seguir as regras legais que lhe conferem o mencionado *status*, considerando a razoabilidade.

É possível afirmar que a anonimização é uma das técnicas de *Privacy by design*, ou seja, técnicas que buscam assegurar a confidencialidade da informação em sua própria estrutura (BARBIERI, 2020).

Essas técnicas vêm ao encontro de uma regulação direcionada e mais específica às tecnologias, já que, como afirma Lessig, o Direito tradicionalmente considerado teria dificuldade em estabelecer diretrizes para os códigos. Assim, através do próprio código e de sua estrutura (LESSIG, 2006), seria possível se regular o uso das informações, por exemplo, em *Big Data Analytics*, ou, de forma mais geral, os fluxos e usos dos dados pela rede.

O enfoque das técnicas de anonimização é a confidencialidade, que é um dos três princípios básicos da segurança da informação (DATE, 2004). A confidencialidade tem por principal função garantir que o dado só possa ser acessado por quem estiver habilitado para tal

57 Nas palavras dos autores: "While the firt generation of Internet sites were collections of HTML files, most major sites today store a large part (if not all) of their data in database systems, over the internet. This is especially trie of site for electronic commerce and other busuness aplications" (RAMAKRISHNAN, R.; GEHRKE, 2003).

fim. Os outros dois princípios básicos da segurança da informação são a Integridade: o qual garante que o dado não será alterado indevidamente (sem autorização) ou danificado; e a Disponibilidade: cuja função é garantir que uma informação esteja disponível toda a vez que for demandada a sua utilização.

A anonimização garante não apenas a confidencialidade, mas permite ainda a maior disponibilidade dos dados ante a ausência das restrições de tratamento decorrente das leis de proteção. Dessa forma, a progressiva disponibilização de dados nas redes, favorecida pelas ondas legislativas nacionais, é ainda mais fomentada com as técnicas de anonimização.

Portanto, é possível se afirmar que, tanto a LGPD quanto o GDPR, entendem que a anonimização seria uma boa técnica para a manutenção da privacidade, ao ponto de considerarem que a aplicação dessas técnicas transmuda a própria natureza dos dados pessoais[58].

Ressalte-se que esse tratamento legal não é concedido para outras técnicas utilizadas em prol da confidencialidade. Além da anonimização, existem ainda outras técnicas de *privacy by design*, como as modalidades mais simples de de-identificação, a pseudonimização, e as técnicas de criptografia, que diferenciaremos para melhor compreensão.

A de-identificação é a técnica que permite a remoção ou camuflagem das informações de identificação pessoal da base de dados, também chamadas PII- Personally

---

58 Interessante mencionar ainda que mesmo antes da LGPD a anonimização já era considerada uma técnica confiável de proteção de identidade pelo poder público, sendo inclusive utilizada como forma de proporcionar concomitantemente privacidade e transparência. Um caso prático em que os dados anonimizados viabilizaram uma atuação do poder público é o acórdão 1391/2016 do Tribunal de Contas da União.
O TCU solicitou dados da Receita Federal em sede de auditoria, para verificação de práticas aduaneiras no Canal Verde. O pedido de acesso ao sistema integrado de comércio exterior (Siscomex) foi negado pela Receita, sob o argumento de violação de dados protegidos pelo sigilo fiscal.
O TCU, então, formulou representação pela sonegação de informações, acolhido pelo plenário através do acórdão n. 1835/2007. Por sua vez, a Receita Federal ingressou com o MS 27.091/DF, e o Supremo Tribunal Federal, analisando o caso, concedeu a segurança para o impetrante, desobrigando-o a apresentar os dados.
Posteriormente, em nova auditoria, o TCU solicitou novamente os mesmos dados, pedindo que eles fossem repassados de forma anonimizada para o órgão. Novamente, a Receita Federal negou o pedido, alegando violação do art. 198 do Código Tributário Nacional caso cumprisse a requisição. Isso ensejou a nova representação por sonegação, processo administrativo 017.090/2015-6, que considerou inválida a negativa. Por meio do Acórdão n. 1958/2015 foi deliberado pelo plenário que, em 15 dias, a Receita deveria fornecer os dados.
Contra a decisão, a Receita ingressou com recurso, alegando que a anonimização seria ineficaz para garantir a privacidade, o que foi negado em decisão posterior (acórdão n. 1391/2016- Plenário) que manteve a obrigatoriedade de apresentação dos dados no prazo estipulado.
Interessante observar que o Min. Luís Roberto Barroso, relator do MS 27.901/DF, afirmou em sua decisão que o novo pedido do TCU para que a Receita fornecesse os dados anonimizados distinguia a requisição daquela inicial, que estava sendo objeto da segurança, fazendo com que, naquele caso, os dados pudessem ser compartilhados. Nas palavras do relator: "Por fim, ressalto que, em julgado recente (Acórdão nº 1.391/2016), o TCU alterou sua perspectiva a respeito do tema, requisitando da Secretaria da Receita Federal do Brasil o compartilhamento de dados "anonimizados", isto é, com ocultação da identidade dos sujeitos passivos. Essa técnica, numa primeira análise, parece viabilizar a concordância prática entre a garantia de sigilo fiscal e a necessidade de controle da administração tributária." (STF. MS 27.091/DF. Min. Luís Roberto Barroso. Data de Julgamento: 03.04.2017. Data de Publicação: DJ 04.04.2017.

Identifiable Information. Além das informações claramente pessoais, a técnica também modifica os "*quasi-identifiers*" - ou quase-identificadores - dos dados, ou seja, aquelas informações que combinadas a outras poderiam identificar o titular, como a data de nascimento de uma pessoa, por exemplo[59]. Desta forma, a de-identificação generaliza ou acrescenta fatores de incertezas aos quase-identificadores. Não é considerado, necessariamente, um processo irreversível, uma vez que há, em muitos casos, a preservação de uma tabela de mapeamento que permite a re-identificação dos dados, alcançando as informações originais (PINHO, 2017 p. 29-30).

Por sua vez, a anonimização é considerada um mecanismo aprimorado de de-identificação, já que a técnica pressupõe a "impossibilidade convencionada" de se reverter o processo. A irreversibilidade é convencionada pois leva em conta os meios considerados razoáveis no momento da anonimização, ou seja, o estado da técnica, o custo e o conhecimento necessário para que seja possível identificar o titular do dado. Portanto, nem toda técnica de de-identificação se adequa ao que a legislação aponta como anonimização e, nesses casos de não adequação, o dado permanecerá sendo tratado como dado pessoal.

Já na chamada pseudonimização, os identificadores pessoais são substituídos por pseudônimos (PINHO, 2017, p. 29-30), criados aleatoriamente ou a partir do mascaramento dos originais. Na pseudonimização todos os atributos da base de dados são mantidos. A técnica pode ou não incidir sobre os "*quasi-identifiers*" e, a depender dessa incidência, a técnica se torna mais forte ou mais fraca quanto à identificação do titular. De qualquer forma, há grandes dificuldades em se aferir os riscos e a taxa de sucesso do processo de pseudonimização (PINHO, 2017, p. 33).

A pseudonimização é prevista na LGPD em seu art. 13, abordando a possibilidade de seu uso na realização de estudos em saúde pública. O §4º deste artigo conceitua a pseudonimização como "o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro". É uma técnica que atinge principalmente os indicadores diretos de identidade, conferindo-lhes um mascaramento ou uma forma de disfarce, sem que haja sua completa dissociação.

59 Sobre os quase-identificadores, Brasher (BRASHER, 2018) descreve: "dados não facilmente identificáveis que podem ser vinculados a informações auxiliares para reidentificar os sujeitos de dados". Mehmood et al. (MEHMOOD et al., 2016) complementa: "Os atributos que não podem identificar exclusivamente um registro por si mesmos, mas se vinculados a algum conjunto de dados externo, podem ser capazes de re-identificar os registros."

A LGPD não exclui os dados pseudonimizados da regulamentação legal para dados pessoais, apesar de destacar a técnica como mecanismo de segurança dos dados, no mencionado art 13. O GDPR trata a pseudonimização de forma semelhante, já que não deixa de considerar o dado pseudonimizado como dado pessoal, conforme expresso no ponto 26 do Considerando que antecede a lei[60]:

> Os princípios da proteção de dados deverão aplicar-se a qualquer informação relativa a uma pessoa singular identificada ou identificável. Os dados pessoais que tenham sido pseudonimizados, que possam ser atribuídos a uma pessoa singular mediante a utilização de informações suplementares, deverão ser considerados informações sobre uma pessoa singular identificável. (...)

O Parecer nº 5 do Grupo de Trabalho de Proteção de Dados do Artigo 29 do GDPR, reitera que a pseudonimização não é um método de anonimização de dados pessoais, mas uma técnica para se dificultar a possibilidade de correspondência de um conjunto de dados à identidade original de seu titular. Dessa forma, apesar de reconhecer a utilidade da técnica como medida de segurança, atribui-lhe menor confiabilidade do que às técnicas de anonimização.

Apesar de manter o dado pseudonimizado sob as restrições da lei, o GDPR aponta a pseudonimização como uma das medidas técnicas e organizativas adequadas para aplicação eficaz dos princípios da proteção de dados, em seu art. 25 (1). Nesse sentido, o GDPR permite que o dado pseudonimizado seja utilizado para fins diversos do original e não expressamente abarcados pelo consentimento expresso do titular, desde que o propósito seja compatível com o consentimento dado, conforme art. 6(4)(e). A mesma previsão é válida para a cifragem, ou criptografia.

Criptografia, por fim, é também um conjunto de técnicas de aprimoramento da confidencialidade da informação. Há várias discussões sobre se as técnicas de criptografia seriam consideradas técnicas de anonimização, quando o dado criptografado é um dado pessoal.

Para alguns autores, a premissa para a resposta a essa pergunta seria o acesso ou não à chave criptográfica. Isso porque a criptografia pressupõe a preservação do conteúdo da mensagem original, cifrando-a através de uma escrita secreta (PAAR; PELZL, 2010, p. 3) (GOODRICH & TAMASSIA, 2012, p. 25), que poderá ser decifrada pelo destinatário, detentor da chave decriptográfica.

Dessa forma, para o destinatário final, ou para quem obtiver a chave, o conteúdo é plenamente acessível. Nesse caso, um dado pessoal que tivesse sido criptografado retomaria todos os seus atributos de dado pessoal.

---

60 Conforme Tabela Comparativa (figura 1).

Entretanto, para terceiros, o conteúdo permanece oculto, podendo ser caracterizado como anônimo, desde que respeitados os quesitos de não identificabilidade do dado impostos pela lei. Os autores dessa corrente apontam três aspectos necessários para se considerar um dado criptografado como anônimo: a força do algoritmo de cifragem; a extensão da chave de encriptação (*e. g.*, 128 ou 256 bits); e a segurança do gerenciamento de chaves (SPINDLER & SCHMECHEL, 2016, p. 171).

Já para Doneda e Machado, os dados criptografados se assemelhariam aos dados pseudonimizados, por manterem uma correspondência com os dados originais através da chave criptográfica. Os autores defendem que essa correspondência entre a criptografia e a pseudonimização foi chancelada pelo GDPR, já que a lei confere o mesmo tratamento dos dados pseudonimizados para os dados criptografados, conforme disposto no já mencionado art. 6(4)(e). Dessa forma, os dados seriam considerados pessoais, apesar de contarem com alguma flexibilização no seu tratamento por serem técnicas que aumentam a confidencialidade.

Para os autores, a criptografia seria semelhante a pseudonimização dos dados, o que os sujeitaria, de forma modulada, à proteção de dados pessoais. Desta forma, não seriam taxados de anonimizados, mas serviriam de resguardo para a segurança pública e para a viabilidade dos negócios, com menores riscos.

Portanto, de forma genérica, as três técnicas apresentadas se diferenciam da anonimização por não se caracterizarem como técnicas irreversíveis, já que mantêm uma ponte que permite a imediata e certa religação dos dados aos seus titulares originais. Assim, por causa da reversibilidade dos tratamentos mais leves de de-identificação e do armazenamento de informações adicionais capazes de associarem novamente os dados aos titulares, ou devido à existência da chave decriptográfica, essas técnicas são consideradas menos eficazes para dissociar o indivíduo de seus dados que a anonimização. Logo, em todos esses casos, os dados permanecem como dados pessoais.

Ao comparar a anonimização de dados com as demais técnicas que se preocupam com a segurança da informação, o que se percebe é que a lei optou por imputar maior confiabilidade às técnicas de anonimização. De fato, apenas os dados anonimizados deixam de ser considerados dados pessoais, mesmo que originalmente os dados fossem considerados sensíveis. Uma vez anonimizados, não é necessário o consentimento do titular para o tratamento desses dados, nem sua vinculação à finalidade originária. A LGPD permite que o controlador mantenha esses dados sem anuência do titular ainda que o titular tenha solicitado a sua

portabilidade[61]. Inclusive, a lei possibilita que, após o término do tratamento de dados pessoais, ao invés dos dados pessoais serem eliminados, eles sejam anonimizados, para uso exclusivo do controlador[62]. Até mesmo o compartilhamento desses dados entre diversos atores não encontra obstáculos na lei.

A fim de explorar melhor a confiança depositada pelo legislador na anonimização, abordaremos sobre o funcionamento das principais técnicas e suas especificidades na promoção da confidencialidade da informação.

2.7. Principais Técnicas de Anonimização

Existem diversas técnicas de anonimização que podem ser utilizadas, de forma isolada, em conjunto entre si, ou com outras técnicas. A anonimização normalmente abarca as informações de identificação pessoal ("PII") e também os já mencionados quase-identificadores. Na sua forma ideal, a anonimização deve alcançar também, caso existentes na base de dados, os dados auxiliares ("AD"), que são dados não necessariamente pessoais, que podem revelar os assuntos referenciados quando tratados junto aos dados “PII” ou aos quase-identificadores. Esses tipos de dados devem ser tratados separadamente por anonimização, de acordo com os riscos inerentes a cada um, e na modalidade que melhor se adeque à própria base de dados.

O trabalho de Brasher (BRASHER, 2018) apresenta as cinco técnicas mais comuns de anonimização: (1) Supressão, (2) Generalização, (3) Agregação, (4) Adição de Ruído e (5) Substituição, conforme mostrado na Figura 2.

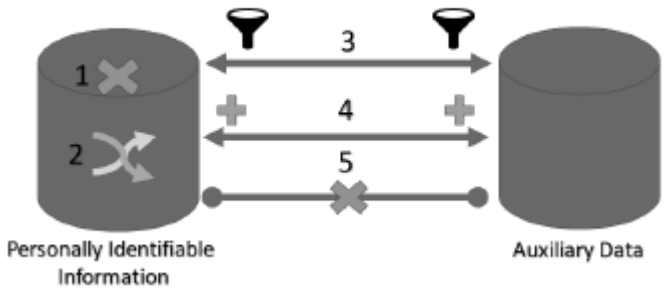

Figura 2: Técnicas de Anonimização.
Fonte: Adaptado de Brasher, 2018; Mehmood et al., 2016.

61 Conforme art. 18, caput, V c/c §7º da LGPD.
62 Conforme art. 16, IV.

A autora explica como seria o funcionamento de cada uma dessas técnicas, as quais fazemos referência por meio da numeração disposta na figura acima. Cabe ainda ressaltar que Brasher considera, para fins de classificação das técnicas de anonimização, os dados quase-identificadores como pertencentes ao PII. Por esse motivo, a figura acima apresenta apenas o que se refere ao PII (incluídos nestes também os quase-identificadores) e aos dados auxiliares (AD). São as técnicas de anonimização:

1) Supressão: É o processo que exclui qualquer PII da base. Dessa forma, são suprimidos os nomes dos titulares, seu número de CPF, RG, para citar alguns exemplos.

2) Generalização: É o processo que embaralha os identificadores PII dentro da base, sem excluir nenhuma informação, mas reduzindo e confundindo sua real vinculação. Dessa forma, a informação permanece completa, mas não necessariamente pertence ao titular apontado.

3) Agregação: Nessa modalidade de anonimização os dois tipos de dados (PII e AD) passam por algum tratamento de agregação de características, como o próprio nome sugere, reduzindo a especificidade dos dados, mas mantendo algumas de suas propriedades. É o caso, por exemplo, da realização da média ou distribuição estatística. Além disso, é possível se estabelecer a reunião de dados em patamares, como, por exemplo, a apresentação de intervalos de idade ou de região ao invés de apresentar idades específicas ou o endereço completo do titular.

4) Adição de Ruído: É o processo que, como o nome aponta, adiciona informações ruidosas, compostas por dados improdutivos, com a finalidade de confundir o vínculo entre PII / AD e seus sujeitos.

5) Substituição: É uma técnica que também embaralha os dados, de forma semelhante à generalização. A diferença se dá porque, enquanto a generalização mescla os identificadores (PII), a substituição mistura o valor dos dados em si, substituindo o conjunto de dados original por outros parâmetros. Esse processo pode ser aplicado a PII e AD (BRASHER, 2018).

Mehmood et al. dispõe sobre a classificação de forma muito semelhante à posteriormente proposta por Brasher, dividindo a anonimização em cinco operações com denominações distintas: (1) Supressão, (2) Generalização, (3) Permutação, (4) Perturbação e (5) Anatomização (MEHMOOD et al., 2016), todas correspondentes às estratégias apresentadas por Brasher, conforme a Figura 2.

De forma geral, a estratégia considerada mais agressiva é a supressão, já que, nesse caso, os dados correspondentes ao PII e quase-identificadores seriam completamente descartados, mantendo-se apenas dados auxiliares.

As demais técnicas são capazes de tornar as informações que compõem as bases mais imprecisas, afastando os dados da vinculação direta com seus titulares. Outras formas de anonimização podem ser extraídas da utilização dessas técnicas, de forma conjunta com outras técnicas ou mesmo a partir de alguns padrões de utilização.

Podemos citar como exemplo a denominada privacidade diferencial, que é uma modalidade que utiliza tanto a "adição de ruído" como a "agregação".

Nessa técnica, os dados pessoais especialmente divergentes são tratados por meio de distribuições estatísticas, para evitar a reidentificação de indivíduos com marcas excepcionais, frente ao banco de dados como um todo. Dessa forma, informações que poderiam ser identificadoras são tratadas com a adição de ruídos para evitar o rastreamento dos indivíduos, distribuindo as informações em curvas probabilísticas (BYRNE, 2015).

Um exemplo da aplicação da privacidade diferencial é o seu uso em estatísticas. Nesse caso, o resultado é a liberação não de uma estatística precisa, mas de dados com uma quantidade "cuidadosamente calculada de ruído aleatório na resposta, garantindo matematicamente que mesmo o reidentificador mais sofisticado não possa usar a resposta para descobrir informações sobre as pessoas no banco de dados". (OHM, 2009, p. 1756).

Outro exemplo bastante conhecido é a técnica chamada K-anonimato, que consiste no uso da supressão de alguns dados e também na generalização do indivíduo em grupos, de forma que cada registro no conjunto de dados publicados seja indistinguível de k - 1 outros registros com base nos quase-identificadores disponíveis (COMAS, 2013).

Desta forma, um atacante poderia no máximo determinar um conjunto de registros k no conjunto de dados publicado que poderiam conter o indivíduo alvo, mas não qual daqueles dados corresponde exatamente ao indivíduo (COMAS, 2013).

Há ainda autores que defendem a melhoria de utilidade dos dados quando aliados os padrões de privacidade diferencial com técnicas do K-anonimato, como defendido na tese de doutorado de Jordi Comas (COMAS, 2013). Mas de toda a forma, cada uma das técnicas apresenta limitações intrínsecas à própria anonimização, que discorreremos nos próximos capítulos.

É importante ressaltar que uma das características centrais da anonimização é a já mencionada irreversibilidade, que é medida a partir dos elementos contextuais que identificam a razoabilidade.

Nesse mesmo sentido, o Parecer nº 5/2014 do Grupo de Trabalho sobre o art. 29 da Diretiva 95/46/CE do Parlamento Europeu aponta que a anonimização se destaca por quatro características fundamentais, quais sejam: 1) a irreversibilidade de identificação do titular dos dados; 2) a viabilidade de se utilizar várias técnicas possíveis, inclusive de forma simultânea, no tratamento; 3) a inserção de elementos contextuais para se identificar o que é tido por razoável no estado da técnica para caracterização da irreversibilidade do processo de anonimização, ou seja, são levados em consideração o estado da técnica e outras características conjunturais; e por fim, 4) há sempre um fator de risco inerente na anonimização.

O reconhecimento do fator de risco se dá, justamente, na identificação de que a irreversibilidade é, na verdade, um elemento contextual convencionado. Dessa forma, cientes de que esse fator de risco é um elemento presente, discorreremos quais são esses fatores de risco intrínsecos ou extrínsecos à técnica de forma mais detalhada no capítulo 3 deste trabalho. Por hora é importante destacar que, através da anonimização, é possível se vislumbrar formas diversas de tratamento que não precisam seguir as prescrições mais estritas da LGPD.

Essas nuances da anonimização, não de forma eventual, serviram para a compatibilização das leis protetivas de dados pessoais com as mais diversas demandas de uma sociedade movida a dados, especialmente com as demandas das plataformas de *Big Data*. A anonimização desponta como balizadora entre os interesses dos cidadãos e dos mercados, prometendo atender ambas as demandas por vezes até mesmo conflitantes.

2.8. Dados anônimos como possível resposta para conciliação entre a Privacidade e o fomento à Inovação em *Big Data*.

Conforme mencionamos, os sistemas de *Big Data* são bases massivas que permitem tratamento de dados através de técnicas de *Big Data Analytics* que se aprimoram à medida em que são alimentadas de mais dados.

Com o despontar das leis de proteção de dados, essas plataformas se viram essencialmente ameaçadas. Apesar de a LGPD não mencionar expressamente essas bases de dados massivas, ela prevê diretrizes que podem interferir profundamente na utilização dessas ferramentas.

Por um lado, não há uma proibição legal quanto à captação e armazenamento de grandes massas de dados. Ao contrário, a lei normatiza mecanismos de tratamento de dados até mais flexíveis ao superar os entraves que a exigência de livre consentimento do titular poderia trazer, se necessária em todos os casos. Nesse sentido, o princípio da autodeterminação

informativa, que se remete ao consentimento expresso e inequívoco do usuário na transmissão de seus dados, foi relativamente mitigado na lei.

Esse tipo de mitigação atende em parte às demandas em *Big Data*, já que a exigência da concordância do titular, por vezes, não seria nem mesmo viável. É o que ocorre, por exemplo, quando se trata de internet das coisas, onde, ainda que não se esteja conectado à internet, dados são repassados máquina a máquina para orientar o funcionamento do aparelho, para manutenção futura, para *back up*, etc. São dados que não passam pela esfera do conhecimento do usuário, muito menos pela sua concordância direta, sendo que se essa autorização fosse necessária a cada transmissão de dado poderia inclusive inviabilizar o funcionamento da tecnologia, na prática. Nesse sentido, a lei traz uma série de alternativas para a aquisição de dados (art. 7º).

Por outro lado, ao mesmo tempo que a lei viabiliza a captura por outros meios que não o consentimento, também prevê limites ao uso de dados pessoais, como o já mencionado princípio da necessidade, que restringe o tratamento desses dados ao "mínimo necessário para a realização de suas finalidades" (art. 6º).

A lei lança as diretrizes de que, mesmo em grandes massas de dados, como ocorre em *Big Data*, os dados pessoais não poderão mais ser armazenados de forma irrefletida, sem que haja uma relação entre o dado e as finalidades a que ele se destina. Isso porque até mesmo o armazenamento de dados é considerado uma forma de tratamento pela lei, conforme o mencionado art. 5º, X da LGPD. A restrição, portanto, alcança dados pessoais dispostos em *Big Data*.

Surge então a necessidade de compatibilizar as técnicas de *Big Data*, tão importantes para os mercados e para extração de valor dos dados, e a proteção que merecem os dados como um todo, particularmente os dados pessoais.

Nesse contexto, a anonimização parece surgir como uma bala de prata, uma vez que promove uma alternativa às restrições legais, mantendo o dado anonimizado ao livre dispor dos gestores de banco de dados, já que não são considerados dados pessoais (art. 12 da LGPD). A confidencialidade assegurada na própria estrutura do dado traria a liberdade de manutenção das bases *Big Data*, além da continuidade de seu crescimento e aprimoramento dos algoritmos de tratamento, sem restrições.

Nesse sentido, o próprio Lessig, no seu livro "*Code: Version 2.0*" aponta a confiança nas tecnologias reforçadoras da privacidade ou *PETs – "Privacy Enhancing Technologies"*. Especificamente sobre a anonimização, o autor afirma que ela seria um resultado praticamente obrigatório para adequação dos quatro reguladores de comportamento

dos gestores. Para o autor, os quatro reguladores seriam: 1) normas e ética; 2) o mercado; 3) a arquitetura; e 4) a lei (LESSIG, 2006, p. 1708). Dessa forma, especialmente quanto à arquitetura, mas não somente, a anonimização seria a escolha padrão (LESSIG, 2006, p. 1709).

Não por outro motivo, diversas instituições brasileiras, públicas e privadas, têm lançado suas expectativas sobre essa técnica, confiando neste método relativamente mais simples de *compliance* à legislação de proteção de dados e suprindo a cada vez mais exigente demanda por proteção nas trocas comerciais internacionais.

Portanto, há uma grande fé em que a anonimização supriria as demandas de compartilhamento indiscriminado e estocagem perpétua de dados com a promessa de que a privacidade dos usuários estaria protegida (OHM, 2009). Entretanto, os avanços na reidentificação mostram que essa fé precisa ser revista, apontando os limites que essas ferramentas podem apresentar, principalmente quando se trata de ambientes de dados massivos.

3. **Riscos Inerentes à Anonimização- Levantamento dos limites**

A anonimização prometia fornecer o melhor dos dois mundos, ou seja, os benefícios do fluxo de informações e fortes garantias de privacidade.[63] Mas na verdade, por uma série de fatores, seja por fatores técnicos internos, como o inevitável trade-off entre utilidade dos dados e privacidade; seja por fatores externos, como a obtenção de informações adicionais em outros bancos de dados, a anonimização perfeita é simplesmente impossível.[64]

Paul Ohm, diretor do Centro de Privacidade e Tecnologia da Universidade de Georgetown, escreveu um artigo pioneiro e interdisciplinar sobre o assunto intitulado "*Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*". O artigo, publicado em 2010, foi impactante por mostrar as facetas preocupantes acerca da confiabilidade da anonimização, justamente em um período de grande euforia legislativa com a saída técnica.

O tom alarmista de Ohm ganhou repercussão na academia, tendo sido citado por mais de 1600 outros artigos[65]. A pergunta que se faz é se as denúncias e as preocupações acerca

63 Como prenunciava Paul Ohm: "In order to squeeze but not cut off valuable transfers of information, legislators have long relied on robust anonymization to deliver the best of both worlds: the benefits of information flow and strong assurances of privacy. The failure of anonymization has exposed this reliance as misguided, throwing carefully balanced statutes out of equilibrium" (OHM, 2010, p. 1732).

64 Conforme afirma Paul Ohm: "Thus, at least for useful databases, perfect anonymization is impossible. 255 Theorists call this the impossibility result. There is always some piece of outside information that could be combined with anonymized data to reveal private information about an individual" (OHM, 2010, p.1752).

65 Segundo a ferramenta de busca do google, acessada em 12/03/2020.

da anonimização, tão bem explicitadas pelo artigo do professor, ganharam repercussão também nas práticas legislativas quanto ao assunto.

Neste trabalho, procuramos extrair as principais críticas do autor, promovendo uma classificação sobre os pontos levantados. As questões levantadas por Ohm como limites, foram ordenadas em dois grandes grupos: limites intrínsecos, ou seja, aqueles que se referem à própria aplicação da técnica puramente considerada; e limites extrínsecos, aqueles vinculados ao ambiente, à forma e à gestão dos dados anonimizados.

Além disso, promovemos um tópico de discussão sobre os limites da ferramenta para além dos aspectos técnicos, que denominamos "limites externos" ao uso da ferramenta. Nesse tópico, abordamos as implicações acerca dos limites éticos no uso dos dados anonimizados como pontos de tensão prática em como se dará a proteção desses dados. Essas questões foram suscitadas a partir de algumas das discussões que Ohm menciona discretamente em seu trabalho, ou que podemos inferir a partir de pontos levantados por Ohm e identificados no debate contemporâneo da quarta onda legislativa sobre proteção de dados.

O objetivo é apresentar esses limites (intrínsecos, extrínsecos e externos) e, posteriormente, mensurar se a lei de proteção de dados brasileira levou em consideração essas questões em suas previsões legais acerca da técnica.

3.1. Limites Intrínsecos:

3.1.1. Informação Teoricamente Segura X Segurança Perfeita

"Nenhuma base de dados útil pode ser perfeitamente anônima" (OHM, 2010, p. 1705). Essa é uma das fortes afirmações do professor Ohm, que encontra respaldo entre os principais teóricos da ciência da computação.

Como veremos no próximo capítulo, a anonimização implica sempre um certo grau de perda de utilidade do dado, uma vez que a técnica se dá justamente na desvinculação e fragmentação do próprio dado.

Nesse sentido, a anonimização perfeita faria o dado inútil, pois completamente dissociado da realidade, seja pela adição de ruídos aleatórios, seja pela exclusão de conteúdo relevante. Os teóricos chamam isso de resultado da impossibilidade (OHM, 2010, p. 1752). Por isso a afirmação de que, pelo menos para bancos de dados úteis, o anonimato perfeito é impossível.

E, de fato, vários estudos (SWEENEY, 2000) (NARAYANAN & SHMATIKOV, 2007) apontaram falhas em processos de anonimização considerados robustos, colocando em cheque a suposição de anonimização absoluta.

Na criptografia, os conceitos de segurança perfeita e informação teoricamente segura ilustram bem esse fenômeno, e pode ser equiparado ao que acontece na anonimização.

A informação teoricamente segura é aquela que se encontra dentro de um limite de segurança pré-estabelecido, onde, ainda que exista um poder teoricamente ilimitado de computação, a criptografia não pode ser quebrada na sua totalidade (MAURER, 1999).

Para a criptografia, a informação teoricamente segura se dá quando a mensagem criptografada ainda contém informações decifráveis, mas não em sua totalidade, ou seja, é possível extrair informações da mensagem criptografada, ainda que seu inteiro teor esteja oculto.

Já a segurança perfeita se caracteriza pela cifragem cuja decodificação seria impossível. Assim, nessa técnica, não é possível extrair informação alguma da mensagem criptografada sem a utilização da chave específica (MAURER, 1993). Entretanto, a obtenção da segurança perfeita teria um custo tão grande para se concretizar que seria inviável sua aplicação cotidiana.

De forma semelhante ao que ocorre com a criptografia, a anonimização perfeita, se possível, seria extremamente custosa, além de levar a uma perda total ou muito significativa da utilidade do dado. Além disso, ressalta-se que, mesmo na anonimização considerada robusta (o equivalente a informação teoricamente segura), há sempre a possibilidade de extração de informação do dado anonimizado. Isso porque, diferente do padrão classificatório da criptografia, a anonimização normalmente é considerada robusta baseada em critérios de razoabilidade e não com base em poderes computacionais ilimitados para reidentificação.

Por esse motivo, Spindler e Schmechel apontam a dificuldade prática em se considerar a anonimização por um aspecto absoluto, ou, o que chamamos aqui de segurança perfeita. Segundo os autores, dentro da abordagem absoluta, todas as possibilidades e todas as chances que o controlador pode ter para identificar uma pessoa como vinculada ao dado são levadas em consideração para se determinar se o dado é pessoal, independentemente dos custos ou demais recursos necessários à empreitada. Em última instância, é difícil imaginar qual dado não poderia ser de alguma forma relacionado a um indivíduo ou a um grupo de indivíduos. (SPINDLER; SCHMECHEL, 2016).

Já a abordagem relativa (informação teoricamente segura, em analogia à classificação de Maurer) considera o esforço requerido para que o controlador relacione o dado

a um sujeito, considerando o dado como anônimo se for exigido um esforço desproporcional (SPINDLER; SCHMECHEL, 2016). Dentro dessa abordagem devem ser estabelecidos critérios para se determinar o que se entende por "desproporcional", fazendo com que mais ou menos dados sejam considerados anonimizados à medida que os parâmetros estabelecidos sejam mais ou menos rígidos.

Por isso a anonimização envolve, intrinsecamente, um fator de risco, já que com elevado esforço, é possível a conversão de dados anônimos úteis em dados pessoais. Esse risco foi inclusive mencionado pelo Grupo de Trabalho sobre o art.29 do Comitê Europeu para Proteção de Dados, através do Parecer n. 5 de 2014, que assim se manifestou:

> Por último, os responsáveis pelo tratamento de dados devem ter em conta que um conjunto de dados anonimizados ainda é passível de apresentar riscos residuais para os titulares dos dados. (...) Assim, a anonimização de dados pessoais não deve ser considerada um exercício pontual e os riscos inerentes devem ser reavaliados regularmente por responsáveis pelo tratamento de dados (COMITÊ EUROPEU PARA PROTEÇÃO DE DADOS, 2014).

A verdade é que os dados, mesmo anonimizados, deixam rastros informacionais, que são como digitais, para usar o termo de Ohm. A singularidade de algumas das informações inferidas é semelhante a "impressões digitais de dados" que os vinculariam aos seus titulares, apesar do processo de anonimização (OHM, 2010, p. 1723).[66]

Em seu artigo de 2010, Paul Ohm faz duras críticas aos estudos que, apesar de observarem os limites da anonimização e os riscos a ela intrínsecos, continuam assegurando que a utilização de técnicas de anonimização garante a privacidade (OHM, 2010, p. 1710). A crítica do autor é extremamente válida, principalmente quando estamos diante de um cenário que parece abraçar uma fé de que a anonimização seria uma solução definitiva para a privacidade.

3.1.2. *Trade-off* Utilidade e anonimização.

A anonimização parecia prometer o melhor dos dois mundos (OHM, 2010. p. 1703): a garantia da privacidade dos indivíduos e, ao mesmo tempo, a manutenção do dado útil, ou seja, aquele dado apto a fornecer informações relevantes sem identificar o seu titular.

66 Nas palavras do autor: "Even though administrators had removed any data fields they thought might uniquely identify individuals, researchers in each of the three cases unlocked identity by discovering pockets of surprising uniqueness remaining in the data. Just as human fingerprints left at a crime scene can uniquely identify a single person and link that person with "anonymous" information, so too do data subjects generate "data fingerprints"—combinations of values of data shared by nobody else in their table" (OHM, 2010, p. 1723).

A manutenção de informações no processo de anonimização garantiria a utilidade do dado para os atores e para os mercados. Destaque-se que o dado dissociado da realidade, ou seja, o dado aleatório, falso ou irreal, não teria a capacidade de orientar corretamente os processos decisórios, processos de inferência e análises de forma geral.

A utilidade do dado, portanto, está diretamente ligada à fidelidade do dado com a realidade. Quanto maior a integridade, como veremos no capítulo seguinte, maior a confiabilidade do dado, interferindo, inclusive, no seu valor de mercado.

Entretanto, a técnica de anonimização é justamente a ferramenta de fragmentação, alteração ou redução de informações passíveis de se extrair dos dados, o que implica diretamente no distanciamento (maior ou menor) entre o dado e a realidade. Esse fato sobre os dados anônimos levou os pesquisadores a concluírem que o dado pode ser útil ou perfeitamente anônimo, mas nunca os dois (OHM, 2010, p. 1704).

Segundo Ohm, os pesquisadores apontaram que mesmo as técnicas de anonimização mais sofisticadas dificilmente eram melhores do que simplesmente jogar fora quase todos os dados, quando se trata de maior proteção à privacidade (OHM, 2010, p.1755).

Isso porque quanto maior privacidade se garante ao dado, ou seja, quanto maior garantia contra a identificação do titular, menor sua potencialidade de utilidade, conforme apontam Luk Arbuckle e Khaled El Emam. Toda anonimização implica em perda da potencialidade de se extrair informações, que se caracteriza aqui pela utilidade do dado (ARBUCKLE & EL EMAM, 2013, p. 19-21). Haveria assim, uma situação ótima intransponível, e não uma situação ideal em que se preservasse ao máximo os dois fatores, conforme a figura abaixo:

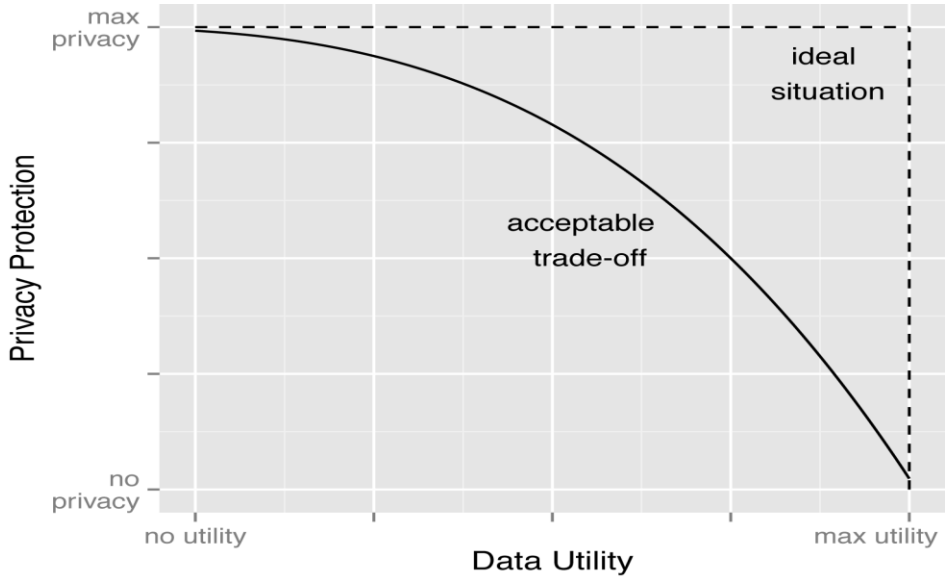

Figura 3 - Anonimização: trade-off entre privacidade e utilidade dos dados

Fonte: ARBUCKLE & EL EMAM, 2013, p. 20

A proteção eficiente da privacidade deve considerar o equilíbrio desses dois aspectos: perda de utilidade e ganho de privacidade de dados baseados em PII. Supostos ganhos de privacidade ocorrem às custas da perda de utilidade. Quando um dado é descartado, informações menos valiosas podem ser extraídas da base remanescente (DOMINGO-FERRER, 2019).

A utilidade fica ainda prejudicada uma vez que, a partir da anonimização, a correlação entre as características e o titular é suprimida, apesar de não serem obstadas as possibilidades de inferência. Isso pode gerar distorções, já que são os próprios algoritmos que realizarão as projeções de correção. Assim, quanto mais robusta a anonimização, maiores as chances de serem inferidas informações distorcidas da realidade, podendo ter consequências danosas, já que, nesses casos, a integridade dos dados pode estar comprometida. É o que veremos a seguir.

3.1.3. A integridade dos dados e as técnicas de anonimização

Esse é um desdobramento do *trade-off* da utilidade apontado por OHM. Discorremos no ítem "o que é anonimização" que a integridade dos dados é um dos pilares da segurança da informação.

Segundo a norma ISO/IEC 27001 (ISO/IEC 27001, 2005, p. 2) a integridade é a propriedade da segurança da informação que tem como objetivo garantir a precisão e completude da informação, ou seja, ela garante que a informação não sofrerá alterações fora do controle pela organização e, portanto, os dados salvos no banco de dados serão íntegros e completos. Com a importância que os dados ganharam na sociedade informacional, cada vez mais a qualidade com que os dados se apresentam passou a ser objeto de valor nesses mercados. Nesse sentido, a integridade, a confiabilidade e veracidade dos dados ganharam nova repercussão.

A integridade como um princípio da segurança da informação, não conseguiria assim esgotar a proteção almejada a essa característica dos dados pelos mercados. Isso porque a integridade, enquanto princípio da segurança da informação, mantém-se intacta quando o dado é alterado ou excluído, desde que a ação seja realizada por um agente institucionalmente autorizado.

A partir de sua ressignificação, a integridade passou a ser um valor do próprio dado e de demonstração de sua qualidade. É o que apontam autores como Barbieri e inclusive o próprio guia de gerenciamento de dados, o conhecido DAMA-DMBOK.

Segundo Barbieri, a integridade é uma das características da qualidade do dado e está "associada com a coerência das suas manifestações em diferentes fontes de dados" (BARBIERI, 2019, p. 21). Dessa forma, um dado é íntegro quando ele mantém correspondência com a realidade, uniformizando a informação nos mais diversos bancos de dados.

Segundo DAMA, a integridade é classificada como acuracidade, também uma das dimensões da qualidade do dado que se refere ao "grau em que um dado corretamente representa uma entidade na vida real seguindo o modelo definido" (DAMA-DMBOK, 2012, p. 314).

Entendendo, portanto, a integridade sob esse aspecto, para além da segurança da informação, podemos afirmar que a anonimização se apresenta como um grande desafio para a integridade e, por consequência, à qualidade dos dados.

As técnicas de anonimização tornam as bases mais imprecisas, como ocorre na adição de ruídos, na generalização, na substituição e na agregação. A adição de informações aleatórias, a mesclagem dos dados que compõem a base, sua agregação ou substituição gera dados diversos dos originais, distanciando-os da realidade. Todas essas técnicas adicionam distorções que, em certa medida, comprometem a integridade e a qualidade. Nesse sentido, essas técnicas podem gerar informações duvidosas e prejudiciais aos titulares, principalmente se não estiver clara a presença da anonimização no banco de dados em questão.

A crítica se torna ainda mais pertinente quando se trata da disponibilização de dados do poder público para a população, em programas de transparência. No acesso à informação, a transparência que justifica a disponibilização dos dados resta comprometida quando há, por exemplo, a adição de informação ruidosa, já que a informação disponibilizada não se encontra perfeitamente alinhada à realidade.

Claro que assegurar a veracidade dos dados é algo que fragiliza a privacidade, já que, a própria incerteza quanto à confiabilidade da informação é uma forma defensiva da exposição dos dados. Se o agente não sabe se a informação é confiável, terá ressalvas em utilizar esses dados para seus processos analíticos e decisórios. Isso faz com que esses dados percam poder de mercado, já que é justamente na confiabilidade das possíveis previsões que se encontra o valor desses ativos.

No entanto, não assegurar essa veracidade pode causar danos tão grandes ou maiores aos direitos dos titulares na sociedade da informação. Isso porque, no caso de uma reidentificação errônea da anonimização dessas bases, a associação de informações distorcidas

a perfis pessoais dificilmente seria retificada, causando um duplo prejuízo para o cidadão, ou seja, causando, além da quebra da anonimização, a associação de um dado falso ao seu perfil.

A inclusão de ruídos nos dados, principalmente com relação às bases públicas, se contrapõe frontalmente, portanto, ao princípio da transparência e da veracidade dos dados. Há uma quebra na confiabilidade dessas bases de dados públicas, que pretensamente trariam informações confiáveis. Isso também acontece em outras técnicas de anonimização, como na generalização, agregação e na substituição.

Já na técnica de supressão, existe mais o fator de perda de integralidade (visto que uma parcela da informação é deletada) do que de integridade, uma vez que os dados restantes permanecem íntegros, caso essa técnica seja utilizada de forma isolada. Por outro lado, a estratégia de supressão não está imune a críticas. Domingo-Ferrer aponta que mesmo esta técnica mais impactante não é suficiente em um contexto de dados massivos, uma vez que a vinculação dos identificadores excluídos se torna trivial caso dados externos sejam incluídos na análise (DOMINGO-FERRER, 2019).

Dessa forma, as inferências seriam não só possíveis, como até mesmo prováveis em plataformas de *Big Data*. Segundo o autor, as preocupações com o impacto social dessa proteção insuficiente são realmente relevantes nesses contextos. É o que analisaremos a seguir.

3.2. Limites Extrínsecos:

A anonimização de dados apresenta desafios internos à própria técnica, que nos fazem questionar até que ponto as técnicas de anonimização, mesmo aquelas consideradas robustas, garantem de fato a privacidade.

O desafio se torna ainda maior quando consideramos não apenas os limites intrínsecos à técnica, mas também a utilização de meios externos à técnica, ou à base anonimizada, para extração de informação dos dados.

Este subitem dedica-se a analisar alguns desses fatores extrínsecos à técnica que podem fragilizar a anonimização, fazendo uma breve explicação sobre como atuam, e como podem ser utilizados inclusive, em favor de uma técnica de anonimização mais confiável, se levados em consideração.

3.2.1. Linkabilidade e poder de inferência de bases de dados externas: Anonimização para além do PII e dos *quasi-identifiers*.

Nem sempre é binária a classificação se um dado é pessoal ou não. O Google foi um dos pioneiros desse debate com a discussão sobre se endereços de IP seriam considerados PII- Personally Identifiable Information- ou "*quasi-identifiers*" ou mesmo dados não pessoais. Nesse sentido, um endereço de IP poderia ser um identificador caso se tratasse de um computador pessoal, por exemplo, ao mesmo tempo em que poderia ser um identificador indireto ou mesmo não ser um identificador, caso se tratasse de um computador de lan house, por exemplo (OHM, 2010, p. 1771-1772).

Esse simples exemplo mostra que, na verdade, existe uma gradação, e não uma diferenciação binária, entre dado pessoal e dado anônimo, a depender do contexto e das informações a ela associadas.

Segundo Doneda, existem três fatores que contribuem com essa característica de gradação entre esses dados: 1) a distinção (*slingling out*); 2) a possibilidade de ligação (linkabilidade); 3) a inferência.

A distinção é a "possibilidade de se isolar alguns ou todos os registros que destacam uma pessoa em uma base de dados" (DONEDA & MACHADO, 2018). A possibilidade de ligação é a viabilidade de se conectar dois ou mais registros referentes a um mesmo indivíduo ou um mesmo grupo de pessoas. Por fim, a inferência é a possibilidade de dedução relativamente confiável de um atributo a partir de outros atributos disponíveis (DONEDA & MACHADO, 2018).

Dessa forma, por mais que a técnica de anonimização garanta um menor risco de distinção, ou seja, de se identificar na base os dados relacionados a um determinado indivíduo, a possibilidade de ligação e de inferência normalmente recebem pouca atenção na consolidação da técnica.

Isso porque a possibilidade de ligação e a inferência não necessariamente identificam um indivíduo específico, mas podem agregar informação a um dado, reduzindo a distância entre o anonimato e o titular.

Por causa desses dois fatores (a possibilidade de ligação e de inferência), mesmo dados que não poderiam ser considerados pessoais, ou dados indiretamente relacionados aos indivíduos (*quasi-identifiers*), podem fragilizar as técnicas de anonimização. Nesse sentido, os dados auxiliares ("AD") e mesmo fragmentos de dados, como é o caso de metadados, podem servir de apoio para a reidentificação.

Por causa dessa dificuldade de classificação, autores como Doneda e Machado afirmam que se continuarmos considerando apenas os "*quasi-identifiers*", ou os PII, como dignos de atenção para proteção da privacidade, muito provavelmente a anonimização terá o

mesmo destino de ser uma técnica autofágica, como são as técnicas criptográficas (DONEDA & MACHADO, 2018). Os autores usam a expressão para se referir à característica da criptografia, que tem sua evolução técnica marcada pela superação dos mecanismos existentes. Ou seja, uma técnica criptográfica tida por segura é colocada à prova constantemente até que é superada a partir do reconhecimento de um acesso ao conteúdo cifrado mesmo sem a chave criptográfica.

O resultado é que as técnicas criptográficas, uma vez quebradas, tornam frágeis toda a criptografia realizada naqueles moldes. A anonimização pode se deparar com situação semelhante, se tomada como uma técnica em si, dissociada do contexto em que é aplicada e dos dados disponíveis no momento da anonimização, para além da base a ser anonimizada. Nesse sentido, é extremamente arriscado confiar o aprimoramento das técnicas de anonimização à quebra de confiabilidade, principalmente em se tratando de dados pessoais alocados em contexto de *Big Data*. O resultado pode ser não apenas o não aprimoramento da técnica, mas também a exposição dos cidadãos a grandes riscos.

Nesse mesmo sentido, Ohm defende que nós devemos abandonar a ideia difundida de que podemos proteger a privacidade simplesmente removendo informações de identificação pessoal (PII). Isso porque PII é na verdade uma categoria em constante expansão, já que a reidentificação vem se mostrando possível a partir de dados não evidentemente ligados a indivíduo identificado ou identificável.

O autor chama esse tipo de abordagem como o "jogo da toupeira": assim que você acertar uma, outra aparecerá. Não importa a eficácia com que os reguladores incorporem os PII mais recentemente descobertos à previsão de seus regulamentos, os pesquisadores sempre encontrarão mais tipos de campos de dados que ainda não foram abordados, numa lista que não tem fim (OHM, 2010, p. 1742).

Ohm alerta então que os legisladores e reguladores devem se atentar não apenas aos dados que podem ser vinculados à identidade, evitando leis ou regras fundamentadas nessa distinção. Assim, ainda que sigamos com uma maior proteção aos PII já usualmente reconhecidos, devemos, na verdade, traçar um novo rumo em geral para a proteção de dados (OHM, 2010, p. 1742).

Paul Ohm exemplifica com dois estudos que permitiram a reidentificação de usuários a partir de informações que não são classificadas como PII. O primeiro estudo foi realizado pelo professor de ciência da computação Latanya Sweeney, que, observando o censo de 1990 dos Estados Unidos, percebeu que em posse de 3 dados: *ZIP code*, data de nascimento

e sexo, a probabilidade de eficácia na identificação de um titular seria de 87,1%. Nesse caso, esses três *quasi-identifiers* foram cruciais na identificação dos indivíduos.

O segundo exemplo é o conhecido desafio da Netflix.

A Netflix promoveu em 2006 um concurso público para melhorar seu algoritmo de sugestão de filmes, na esteira da personalização dos serviços que temos descrito. Para isso, ela liberou parte de seu banco de dados dos comentários sobre filmes contendo 100.408.507 avaliações criadas por 490.189 usuários do Netflix (NETO; MORAIS, 2017). O banco de dados disponibilizado continha apenas a nota atribuída pelo usuário e a data em que a avaliação foi feita, extraídos todos os identificadores dos clientes.

Além disso, os "*quasi-identifiers*" foram mascarados, acrescentando-se ruídos (por exemplo, modificando o dia e a hora da avaliação a dois valores, acima ou abaixo do real). Dessa forma, os organizadores do concurso tinham em mente que os dados disponibilizados estariam perfeitamente anonimizados, de forma que não poderiam ser considerados PII ou mesmo "*quasi-identifiers*". Estavam tão confiantes quanto à anonimização realizada que garantiram o completo cumprimento da política de privacidade da empresa (NARAYANAN & SHMATIKOV, 2008).

Os pesquisadores da Universidade do Texas, então, utilizaram a base de dados disponibilizada pela Netflix e a compararam a uma outra base de dados de acesso público: o *IMDB – Internet Movie Database*. Trata-se de site que também reúne reviews cinematográficos postados voluntariamente por internautas. Como resultado, eles conseguiram identificar quais usuários eram responsáveis pelos reviews da base de dados do Netflix, ou seja, re-identificaram os titulares dos comentários.

O trabalho foi capaz de identificar mais de 80% dos usuários a partir dos comentários e notas dadas aos filmes. A constatação dos professores foi de que, a partir de 6 avaliações de filmes mais conhecidas, 84% dos usuários poderiam ser reidentificados (OHM, 2010). E se o "adversário" soubesse também a data aproximada (duas semanas a mais ou a menos) que a avaliação teria sido feita, então a reidentificação teria a probabilidade de 99%.

Além disso, os pesquisadores conseguiram obter outros dados dos usuários, como número do cidadão, equivalente ao nosso "CPF", além de traços de personalidade, orientação religiosa, política e sexual, etc (NETO; MORAIS, 2017).

Entretanto, há que se apontar, no caso, que a reidentificação só foi possível por um terceiro sem qualquer outra informação sobre os usuários da Netflix, a partir da comparação dos dados disponibilizados pela Netflix com uma base de dados pública, ou seja, a partir de uma base de dados externa.

No caso, os professores usaram a base *IMDB* que deixava pública a pontuação que os usuários davam aos filmes assistidos, com a identificação do usuário. Dessa forma, a anonimização funciona em rede, já que cada nova informação disponibilizada publicamente oferece um grande risco para todas as bases anonimizadas.

Ohm descreve esse fenômeno como "problema do acréscimo" e, segundo o autor, por causa deste problema todo evento de disponibilização de dados e de reidentificação, por mais que pareça benigno, aproxima as pessoas dos danos. (OHM, 2010, p. 1747). Quanto mais dados se tem, mais fácil é reidentificar outras bases, pois aumenta a linkabilidade dos dados, ou seja, aumenta-se a possibilidade de ligação entre os dados disponíveis e as inferências possíveis. O problema não está somente na reidentificação, mas na menor entropia causada pelo acréscimo de informações, ou seja, a cada nova informação, mais próximo um adversário estará da reidentificação (OHM, 2010, p. 1746).

Por esse motivo, Ohm já apontava, em 2010, a necessidade de uma maior proteção de dados em *Big Data*. Segundo a abordagem do autor, para uma mais efetiva privacidade dos dados pessoais, além das preocupações com PII e com dados sensíveis ou que revelem fatos especialmente embaraçosos dos indivíduos, os reguladores deveriam se preocupar com proprietários de grandes bancos de dados, pois nesses ambientes há uma sensível diminuição da entropia.

Segundo o autor, a entropia está relacionada à aleatoriedade, que diminui à medida que se obtém informações sobre determinado objeto, fato ou evento. O autor ilustra com a brincadeira das 20 perguntas, onde um jogador pensa num objeto, pessoa, evento, etc. e dá o direito de o adversário fazer 20 perguntas às quais ele responderá sim ou não. A cada pergunta respondida a aleatoriedade da resposta diminui e também restam menos possibilidades plausíveis. Isso se caracteriza como a diminuição da entropia. (OHM, 2010, p. 1746-1748). No caso da anonimização, quanto menor entropia causada pelo acréscimo de informações, ou seja, a cada nova informação, mais próximo um adversário estará da reidentificação.

Para o autor, bancos de dados massivos são grandes diminuidores de entropia, por conter diversas informações interrelacionadas. O autor então sugere que esses tipos de bases sejam regulamentadas de forma mais rígida, seja com novas regras especialmente criadas para esta realidade, seja com a aplicação de regras preexistentes.

De toda forma, em se tratando de *Big Data* ou não, percebe-se que a possibilidade de inferência e de ligação entre os dados revela como a anonimização não é uma técnica estática, e deve respeitar não apenas o estado da tecnologia, mas também a flexibilidade de inferência que os dados externos repercutem na própria técnica.

Essas características da possibilidade de ligação e de inferência são particularmente importantes para se analisar as limitações da anonimização, pois, normalmente, os dados são anonimizados tomando-se em consideração apenas a própria base anonimizada.

Isso causa algumas distorções principalmente quando nos deparamos com uma enorme quantidade de dados externos facilmente acessíveis por outros meios que não a base anonimizada, como ocorre na atual sociedade informacional.

Nesse sentido, há uma grande probabilidade de que alguma informação, interna ou externa à base de dados, possa ser combinada com dados anonimizados para revelar informações privadas sobre um indivíduo.

A extração de informações da base de dados anonimizada a partir do tratamento com dados de bases externas pode ser ainda mais nociva quando o alvo é um indivíduo específico. Esses casos são chamados testes de perfil "*Prosecutor*", em que a coleção de informações acerca da mesma pessoa se torna mais fácil se agregada a informações já coletadas, tornando a reidentificação mais simples (PINHO, 2017, p. 38).

Percebemos, portanto, como de fato os dados não podem ser considerados nuclearmente em suas próprias bases, da mesma forma que devem ser protegidos de forma global. Nesse sentido, não apenas os dados pessoais merecem proteção, mas os dados de forma geral formam um conglomerado de informações que devem ser utilizados com cautela e boa-fé, respeitando padrões de boas práticas e de governança.

Defendemos isso, em especial para os dados anonimizados, que, por conterem informações referentes a pessoa, têm a flexibilização da proteção indicada pela lei, ao mesmo tempo em que contêm intrinsecamente a possibilidade sempre presente da violação da privacidade do indivíduo.

3.2.2. Accountability da técnica- fuga do modelo liberação-esquecimento de anonimização

Justamente por causa do problema do acréscimo e dos riscos que cada novo dado lançado nas redes traz para bases de dados anonimizadas, Paul Ohm é bastante crítico ao modelo "liberação-esquecimento", que é predominante na utilização dos dados anonimizados.

Nesse modelo, o gestor dos dados repassa (publicamente, para terceiros, ou dentro da própria corporação) os dados anonimizados, sem se preocupar com o que acontecerá com eles após a transferência (OHM, 2010, p. 1711-1712), ou seja, sem nenhum compromisso sobre o que será feito da informação.

O modelo “liberação-esquecimento” é, sem dúvidas, reflexo da crença de que uma anonimização robusta garantiria a privacidade, sendo desnecessários cuidados extraordinários para uma base de dados anonimizada de forma robusta.

A questão é, como ressaltado no trabalho, que a anonimização deve ser considerada não de forma isolada dentro do próprio banco de dados. Por causa da possibilidade de ligação e do poder de inferência, novas informações e novas técnicas podem tornar viáveis os processos de reidentificação. À medida que novos dados (PII, quasi-*identifiers* ou dados auxiliares (AD), como dados públicos em geral, dados não pessoais e metadados) são dispostos nas redes, maior a dificuldade de se anonimizar ou de se manter anonimizado um determinado banco de dados.

A livre disponibilização de dados fragiliza a técnica como um todo. Além disso, o modelo “liberação-esquecimento” acarreta dificuldades na responsabilização de agentes que, sob o pretexto da livre disposição de dados anônimos, acabam incorrendo em vazamento de dados pessoais (PII ou *quase-identifiers*).

Paul Ohm alerta sobre a inviabilidade de se esperar um aprimoramento das técnicas do modelo “liberação-esquecimento”. A livre disposição, como no modelo “liberação-esquecimento,” seria confiável apenas em situações de anonimização calcadas na segurança perfeita, o que é inviável já que há o *trade-off* entre utilidade e privacidade (OHM, 2010, p.1751).

Dessa forma, a manutenção desse modelo implica em pelo menos dois grandes problemas para a garantia da privacidade de dados. O primeiro trata da manutenção da anonimização dos bancos de dados anonimizados, já que eles precisam ser reajustados de acordo com a evolução da técnica e disponibilização de novos dados. No modelo “liberação-esquecimento” o gestor não realiza mais esse *accountability*, o que faz com que a técnica de anonimização se fragilize e reverta dados anonimizados em PII ou *quasi-identifiers*.

O segundo problema é quanto a responsabilização de agentes que disponibilizam publicamente dados cuja anonimização esteja ultrapassada ao longo do tempo, constituindo-se de fato em dados pessoais. Sem o controle sobre a cadeia de disponibilização de dados, surgem dificuldades em se responsabilizar agentes sobre os danos decorrentes de vazamentos.

3.2.3. Parâmetros de Governança da base e anonimização

O problema do acréscimo apresentado por Ohm, aliado com metodologias como a “liberação-esquecimento”, proporcionam um ambiente digital extremamente desafiador para

manutenção da anonimização. O problema ainda é acentuado quando consideramos plataformas de *Big Data*, onde ocorre significativa redução de entropia.

Por esse motivo, Domingo-Ferrer aponta que o anonimato em *Big Data* ainda é limitado (DOMINGO-FERRER, 2019). O autor ressalta que um dos principais motivos para essa limitação é o fato de que a confiança nos controladores de dados, concedida pelas legislações de proteção, é prejudicada pela falta de critérios de gerenciamento acionáveis para o tratamento da confidencialidade (DOMINGO-FERRER, 2019).

Além da ausência de controle quanto à disponibilidade de dados no modelo "liberação-esquecimento", também não é exigido dos controladores que sejam satisfeitos parâmetros de uso e gestão de seus bancos de dados de forma a minimizar riscos, o que seria particularmente importante em plataformas de *Big Data*.

De fato, de forma geral, mecanismos como o *Big Data* precisam passar por adaptações específicas. É por isso que, especialmente nesses ambientes, o anonimato deve ser combinado com outros mecanismos de proteção de dados (CARVALHO et al., 2020).

A Governança de Dados, por sua vez, apresenta-se como possível aliada em favor da privacidade (FOTHERGILL et al., 2019) (PIRAS et al. 2019) (CARVALHO et al., 2020) nesses contextos.

A Governança de Dados é, segundo Barbieri, um subdomínio da Governança Corporativa, que atua com independência frente à Governança da TI. Assim, ambos, a Governança de Dados e a Governança de TI seriam subdomínios da Governança Corporativa. A independência da Governança de Dados, frente à Governança de TI se justificaria pelo fato de que os dados cada vez mais são reconhecidos como um ativo da organização e não como uma propriedade ou um recurso da tecnologia (BARBIERI, 2019, p. 37-38).

Dessa forma, segundo o autor, a Governança de Dados poderia ser definida como:

> "(...) um conjunto de práticas, dispostas em um *framework*, com o objetivo de organizar o uso de dados e o controle adequado dos dados como um ativo organizacional. Seria, por assim dizer uma forma de pôr ordem na casa com relação aos aspectos de dados, visando disponibilidade, integridade, consistência, usabilidade, segurança, controle, etc. Metaforicamente, a Governança de Dados seria uma espécie de Legislativo e Judiciário dos dados, enquanto que a Gerência seria o Executivo dos dados. Juntas, formam a Gestão dos Dados" (BARBIERI, 2019, p. 62).

Adotando a mesma base conceitual, a legislação nacional, por meio do Decreto nº 10.046/2019, define governança de dados em seu art. 2º, XV, como sendo o "exercício de autoridade e controle que permite o gerenciamento de dados sob as perspectivas do

compartilhamento, da arquitetura, da segurança, da qualidade, da operação e de outros aspectos tecnológicos".

A governança é usada para promover a padronização e controle de qualidade na gestão de dados internos, garantindo maior documentação e organização. Ela também é importante para racionalizar o caro e expansivo processamento desses ativos. Justamente por essas características, a governança é uma boa aliada no desafio de promover maior proteção de dados, mitigando os riscos envolvidos na anonimização quando os dados são processados no contexto do *Big Data*.

E não apenas isso. A governança de dados tem demonstrado que é falacioso o seu estigma inicial, de que essa ferramenta seria um entrave ao livre desenvolvimento tecnológico. Atualmente a governança é vista como uma aliada na estratégia de negócios, justamente por aprimorar a gestão dos dados. Barbieri aponta que, por meio da Governança de dados,

> "as empresas hoje também definem mecanismos para analisar os processos que se abastecem de ou produzem os dados, criando um sentido maior de qualidade conjunta entre esses dois elementos seminais, dados e processos, contribuindo para o conhecimento da cadeia produtiva de informação e conhecimentos" (BARBIERI, 2019, p. 36).

Ohm também defende que a governança seja utilizada especialmente em plataformas de *Big Data*. (OHM, 2010. p. 1760). Ele sugere que analisar a forma de gestão dos dados anonimizados, os possíveis motivos que levariam a reidentificação e analisar se a anonimização poderia ser bem sucedida tomada em seu contexto seria mais útil do que se debruçar na robustez da anonimização em um dado isolado. Isso porque esse tipo de pergunta envolveria aspectos sociológicos, psicológicos e institucionais do funcionamento da base de dados e inclusive do caminho a ser percorrido por um possível "adversário" (OHM, 2010, p. 1761).

Justamente por definir parâmetros (requisitos) sobre como os dados serão utilizados, em qual contexto, além de estabelecer o *framework* dessa utilização, é que a governança é uma ferramenta que proporciona os meios para auditabilidade das decisões, promovendo por fim o aumento da confiabilidade nos controladores. Através desses modelos de atuação e seus requisitos de segurança da informação, estabelece-se maior transparência sobre como esses dados serão manipulados e quais as regras estabelecidas para essa manipulação.

O Decreto nº 10.046/2019 também conceitua, no inciso XXIII, requisitos de segurança da informação e comunicações, como "ações que objetivam viabilizar e assegurar a

disponibilidade, a integridade, a confidencialidade e a autenticidade das informações". Portanto, é dentro de estratégias de governança que se estabelecem métodos para cumprimento dos princípios da segurança da informação. Abordamos brevemente sobre esses princípios no subitem "o que é anonimização?" do capítulo 2.

Segundo Machado, os princípios da segurança da informação são três e podem ser assim definidos: 1) Confidencialidade: garante o nível de sigilo dos dados e previne contra a sua divulgação não autorizada, impedindo o acesso indevido; 2) Integridade: garante o rigor e a confiabilidade das informações, de forma que a informação original não seja alterada indevidamente por um agente sem autorização, ou danificada de qualquer forma; 3) Disponibilidade: garante que a informação estará disponível para aquele agente autorizado a utilizá-la toda a vez que for demandada a sua utilização. (MACHADO, 2014). Nesse sentido, a governança é uma boa aliada ao promover melhorias na segurança da informação desses bancos de dados, sendo um instrumento de gestão e acessibilidade.

Por exemplo, no caso da confidencialidade. Sabemos que, por causa do problema do acréscimo, quanto mais dados são disponibilizados para tratamento em conjunto com a base anonimizada, maiores os riscos de reidentificação ou ao menos diminuição de entropia. Em bases de dados desordenadas[67], que é o que normalmente ocorre quando não se estabelecem estratégias de governança, há grande possibilidade de se encontrar dados redundantes.

Dados redundantes são, como o nome indica, dados que se encontram duplicados na base, ou mesmo duplicados (triplicados, quadruplicados ou n-plicados) em mais de uma base de um mesmo sistema de banco de dados. Esses dados podem ser um problema, tanto na aplicação de técnicas de anonimização, quanto na própria gestão dos dados. Para as técnicas de anonimização, esses dados podem permitir a fácil reidentificação se um dado for submetido à anonimização em uma base enquanto se mantiver íntegro em uma outra base do mesmo banco de dados. Nesse sentido, a governança seria um estandarte da organização institucional, diminuindo essas falhas e aumentando a confidencialidade da informação pessoal.

Ao mesmo tempo, a partir de processos decisórios em governança, uma organização pode decidir pela manutenção de dados redundantes. Por exemplo, a redundância poderia ser útil em sistemas de *back up,* ou mesmo sistemas de gerenciamento de banco de dados poderiam garantir uma maior disponibilidade dos dados de forma distribuída através da redundância (CARVALHO, 2016).

67 Essas bases são chamadas pela literatura de "data swamp" ou "pântano de dados".

Um outro exemplo é que a governança seria importante na confidencialidade e integridade relacionada ao acesso de determinado agente de tratamento e na proteção da disponibilidade dos dados. A ISO/IEC 27001 define a disponibilidade como a propriedade de uma informação ser acessível e utilizada sob demanda por uma entidade autorizada (ISO/IEC 27001, 2005, p. 2). Em todos os elementos principiológicos da segurança da informação (confiabilidade, integridade e disponibilidade) se ressalta a questão do acesso restrito, pela chamada "entidade autorizada" e, por isso, são propriedades muito próximas (CARVALHO, 2016). A restrição de acesso à "entidade autorizada" é realizada por vários bancos de dados a partir de estratégias de governança, que definem o acesso às bases a partir de perfis setoriais.

Podem ser definidos três tipos de perfis de controle acesso: discricionário, mandatório e baseado em papéis.

O controle de acesso discricionário define privilégios de acesso em nível de conta e em nível de relação, e estabelece um controle binário de acesso a tudo ou o não acesso ao banco de dados (CARVALHO, 2016). Já no controle de acesso mandatório é possível a classificação de dados e usuários em classes de segurança (ELMASRI & NAVATHE, 2011, p. 570). O dado é classificado de acordo com a sua importância para o sistema, enquanto o usuário recebe uma classificação, chamada *clearence*, de acordo com a confiança que o sistema possui neste usuário (CARVALHO, 2016).

Por fim, o controle de acesso baseado em papéis permite a atribuição de permissões a papéis genéricos dentro do banco de dados, possibilitando posteriormente classificar o usuário em algum ou alguns dos papéis pré-estabelecidos. Os papéis podem ser, por exemplo, ligados a funções ou responsabilidades dos funcionários dentro de uma organização (CARVALHO, 2016).

O controle de acesso, de forma geral, permite o fomento a confiabilidade e a disponibilidade dos dados em um banco de dados, já que restringe o acesso à "entidade autorizada".

Esse tipo de controle facilita a preservação da anonimização, ou permite que menos usuários tenham acesso à base como um todo e assim aos dados que poderiam re-identificar a base. A decisão sobre que modelo de restrição é mais viável para um determinado banco de dados é definida pela melhor estratégia de governança institucional, de acordo com as especificidades de cada organização.

Além do controle de acesso, a disponibilidade pode também ser alcançada através de outras ferramentas de gestão. Um exemplo é a *Oracle Real Application Clusters (RAC)*. Trata-se de uma tecnologia de clusterização de bancos de dados, ou seja, que viabiliza o acesso

compartilhado de diversas máquinas a um mesmo banco de dados (CARVALHO, 2016). A sensação do usuário é de acessar a um banco de dados unificado, sendo que, na verdade, trata-se de bancos de dados descentralizados, permitindo o compartilhamento simultâneo de acesso. Dessa forma, há a possibilidade de se ter, ao menos sensorialmente, um banco de dados unificado, sem ter as desvantagens que uma unificação completa poderia trazer.

De fato, além de melhorias com relação à disponibilidade, esse tipo de tecnologia pode ser estratégica na defesa em ataques a bancos de dados, já que, por causa da clusterização, o atacante poderia ter real acesso apenas à parcela dos dados. A tolerância a falhas também está diretamente ligada ao conceito de clusterização, já que os computadores funcionam em nós, atuam de maneira simultânea e independente, fazendo com que, se um desses computadores falhe, não afete diretamente os demais (CARVALHO, 2016). A capacidade de processamento também pode ser aumentada à medida que novas máquinas são dispostas em nós na rede (CARVALHO, 2016).

Todas essas possibilidades tecnológicas apontam para um fato: a capacidade de melhor gestão dos dados é uma grande aliada no ajuste que a privacidade deve ter com os interesses do negócio. A partir de boas práticas de governança é possível se eleger as tecnologias que melhor se adequem ao perfil e aos objetivos almejados para um determinado banco de dados. Esse tipo de gestão apara as arestas das bases quanto a falhas de privacidade e, particularmente, para processos de anonimização, permitem a escolha das melhores técnicas e a auditabilidade sobre as técnicas aplicadas.

Por outro lado, a ausência de uma boa gestão aumenta a quantidade de falhas e a desorganização do fluxo de dados, que pode prejudicar não apenas a qualidade da técnica de anonimização, mas a própria eficiência do negócio e a utilidade dos dados a longo prazo.

Apesar dessas características da Governança, um estudo da Harvard Business Review, publicado em 2017, aponta que apenas 3% dos dados das empresas são submetidos à governança, atendendo a padrões básicos de qualidade (THOMAS et al., 2017).

A busca por metodologias mais ágeis de gestão faz com que surjam críticas aos métodos mais burocráticos e analíticos que são suscitados pela Governança de Dados, afirmando que eles seriam morosos e causariam o engessamento da gestão (BARBIERI, 2019, p. 250). Por vezes a documentação é deixada de lado para dar espaço a projetos céleres, o que pode acarretar em embates com sistemas tradicionais de Governança.

A Governança de Dados, por sua vez, tem se adaptado a essas novas demandas levantando possibilidades de gestão ágil por meio de novos *frameworks*, mais simplificados e mais voltados ao resultado. Barbieri aponta que diversos autores dos métodos *agile* “seguem

uma abordagem padrão de implementação de GD (Governança de Dados), porém com uma visão de otimização de tempo e recursos na definição de seus elementos mais básicos, com o objetivo de neutralizar percepções criadas de movimentos que se tornaram lentos e burocráticos" (BARBIERI, 2019, p. 243). Outros autores, no entanto, criaram gerências específicas de dados, voltados para o método ágil.

Sem desmerecer a busca por maior celeridade e objetividade, por vezes essenciais para viabilizar a configuração de projetos urgentes ou emergenciais, uma simplificação metodológica descuidada pode diminuir ou mesmo esvaziar características muito importantes da Governança de Dados para fins de proteção desses ativos. Isso porque elementos como a auditabilidade, transparência e *compliance* dependem por vezes de uma boa documentação ou mesmo do cumprimento de etapas de verificação. Além disso, muitas vezes a participação de *stakeholders* diversos nos processos de gestão, em especial o caso dos titulares na gestão de dados pessoais, normalmente demanda tempo de informação e decisão (CANEDO et al. 2019)[68].

Além disso, essas formas de acomodação de interesses suscitam questões importantes dentro da macrovisão de regulação que se utiliza da Governança como um ferramental. A dosagem da complexidade desses métodos e sua adequação em um determinado contexto se traduz como um verdadeiro desafio regulatório. Isso porque a aplicação de métodos sofisticados de Governança para todo e qualquer agente pode implicar em sérias dificuldades para empresas emergentes, por exemplo. A padronização de mecanismos pode resultar, no fim das contas, na inadequação do método para determinados agentes.

Portanto, enfatizamos a grande importância da governança de dados como forma de conferir transparência, auditabilidade, maior proteção nos tratamentos de dados e melhor gestão desses ativos, mas não podemos deixar de reconhecer os desafios regulatórios que envolvem a questão.

3.3. Limites Externos: Desafios Jurídicos e Limites éticos na utilização dos dados anonimizados

Conforme ressaltamos, baseados no trabalho pioneiro de Paul Ohm, a anonimização apresenta limites técnicos tanto intrínsecos como extrínsecos, que fazem com que a completa dissociação entre o titular do dado e o dado anônimo encontre severas dificuldades.

[68] Há inclusive estudos sobre como associar os métodos ágeis a uma maior participação dos agentes envolvidos, através de oficinas de workshop, por exemplo (CANEDO et al., 2019).

Entretanto, para além das limitações técnicas, a anonimização também suscita outras discussões que repercutem e até mesmo aprofundam questões caras ao Direito no que tange aos dados pessoais.

A própria definição da natureza jurídica dos dados pessoais, por exemplo, conforme mencionamos, envolve complicados debates, ainda não solucionados, que podem afetar toda a estrutura legal de proteção. O resultado é a indefinição prática sobre como o Direito lida de forma geral com os dados pessoais e, de forma ainda mais delicada, de que forma ele lidará com os dados anônimos.

Somando-se a esse, outro grave problema de definição é enfrentado pelo conceito de privacidade, cuja tutela é o objetivo final da anonimização. A privacidade apresenta suas nuances de mutação e adaptação dentro de uma estrutura jurídica que se esforça para compreender e assimilar a nova era informacional. De um conceito iminentemente voltado à intimidade da esfera privada, a privacidade vem ganhando uma forma mais ampla e flexível, que desafia os contornos jurídicos de proteção. Todos esses fatores seguem suscitando dúvidas sobre os efeitos que a anonimização trará, na prática, para os titulares.

Enquanto essas questões seguem sem respostas, a liberdade no uso e compartilhamento de dados anônimos, assegurado legalmente, pode resultar por fim em um aprofundamento de graves problemas sociais já presentes. Retratamos neste trabalho dois aspectos dessas discussões, quais sejam: a acentuação da assimetria informacional e a formação de perfis comportamentais. A pretensão não é, de forma alguma, exaurir todas as discussões envolvendo esses dados, mas apenas contextualizar, para além da discussão técnica, as dificuldades práticas e éticas que a Lei de Proteção de Dados e sua interpretação deverão enfrentar na garantia dessa nova forma de privacidade que se delineia.

3.3.1. A ressignificação da Privacidade e a Anonimização

O paradigma informacional que vivenciamos tem transformado aspectos da existência humana e os direitos relacionados à personalidade. A privacidade tem sido particularmente afetada com a economia dos dados, fazendo com que até mesmo conceitualmente haja um ajuste deste direito. A privacidade começou como um direito de ser deixado sozinho ("*righ to be left alone*"), profundamente relacionado à intimidade, ou seja, afastando o indivíduo de quaisquer formas de intromissão não desejada por parte de outros sujeitos, que afetassem ou tivessem o potencial de afetar sua independência, individualidade, dignidade e honra (WARREN & BRANDEIS, 1890).

Com as novas tecnologias cada vez mais presentes no cotidiano das pessoas, “ser deixado sozinho” parece algo cada vez mais difícil de acontecer. A tendência é a hiperconexão constante.

Bauman aborda sobre o assunto ao afirmar que estamos em uma sociedade confessional, “em que microfones são instalados dentro de confessionários”, ou seja, em que a esfera pública é invadida por conteúdos eminentemente privados, trazendo a intimidade para fora de seus limites (BAUMAN, 2012, p. 164). Nessa sociedade, portanto, perdemos a firme barreira entre o que é da esfera pública e o que é da esfera privada, já que naturalmente o privado se expõe.

Segundo Bauman, “o advento da sociedade confessional assinalou o triunfo final da privacidade” uma vez que ao mesmo tempo em que a privacidade atingiu seu auge, invadindo, colonizando e conquistando o domínio público, isso lhe custou a “perda de seu direito ao sigilo – seu traço definidor e seu privilégio mais valorizado e defendido com tenacidade” (BAUMAN, 2012, p. 164). Bauman faz uma descrição de como as redes sociais têm ganhado espaço nas sociedades e como, de modo geral, os cidadãos têm se exposto nas redes, com seus “confessionários eletrônicos portáteis”. Ele aponta que não se trata mais de um fenômeno adolescente de inserção social, mas que a visibilidade do privado já ganhou tanto espaço na sociedade confessional que, aquele que preserva sua privacidade mantendo a vida íntima longe das redes passa a ser objeto de desconfiança, quando não completamente excluído do convívio social (BAUMAN, 2012, p. 165-166). A exposição pública do privado se transforma então numa virtude e numa obrigação pública (BAUMAN, 2012, p.166).

O autor ressalta que, na sociedade confessional, o âmbito privado deve ser exposto ao público já que as pessoas passam a ser produto de si mesmas. Devem se esforçar para fazerem a melhor propaganda de si nos mercados da atenção. Segundo o autor:

> (...)“forçadas a se vender no mercado e desejando vender-se pela maior oferta possível, são instigadas, induzidas ou obrigadas a promover uma atraente e desejável mercadoria; assim, fazem o possível, recorrendo aos melhores meios à sua disposição, para aumentar o valor de mercado dos produtos que vendem. A mercadoria que são estimulados a colocar no mercado, promover e vender são elas mesmas. Elas são, a um só tempo, promotoras de mercadorias e as mercadorias que promovem. São o produto e seus agentes de marketing, os bens e seus vendedores itinerantes (e permitam-me acrescentar que qualquer estudioso que já tenha se candidatado a um emprego na área de ensino ou a uma verba de pesquisa reconhecerá com facilidade sua condição nessa experiência). Seja qual for a categoria em que possam ser enquadrados pelos organizadores das tabelas estatísticas, todos habitarão o mesmo espaço social conhecido pelo nome de mercado” (BAUMAN, 2012, p. 167).

Dessa forma, Bauman afirma que a condição de se tornar um produto atrativo em meio à disputa por atenção dos mercados é que torna o indivíduo um membro legítimo dessa nova sociedade (BAUMAN, 2012, p. 168). Também por esse motivo, é gerada desconfiança com relação a quem não se esforça em apresentar publicamente a melhor versão de si. O autor completa:

> "tornar-se e continuar a ser uma mercadoria vendável é o mais forte motivo das preocupações do consumidor, mesmo que ele em geral seja latente e poucas vezes consciente, muito menos declarado" (BAUMAN, 2012, p. 168).

Nesse contexto, a privacidade é completamente ressignificada. O "direito de ser deixado sozinho" se torna cada vez mais dissonante no paradigma do indivíduo como mais um produto nesse mercado da atenção.

E isso é intensificado quando ressaltamos o regime 24/7 (CRARY, 2016, p. 19), consequente da globalização, apontado por Jonathan Crary, que estende o tempo de disponibilidade dos indivíduos a 24 horas, 7 dias por semana, já que há a necessidade contínua de se informar, de conectar a esfera privada com os acontecimentos ao redor do mundo.

O conceito de privacidade é então ajustado e, com a terceira geração legislativa, ganha um aspecto de domínio sobre o que se torna público, através da autodeterminação informativa. Dessa forma, a privacidade é considerada como respeitada, ainda que o indivíduo decida expor completamente sua vida privada, desde que essa exposição tenha sido minimamente espontânea.

Minimamente porque é possível se erigir críticas sobre essa livre autodeterminação. Até que ponto, considerada a assimetria informacional, a fragilidade dos indivíduos nessas relações de disputa por atenção, e a maior pressão pela objetificação dos titulares como mercadoria, podemos afirmar que há uma autodeterminação consciente e livre da exposição de seus dados privados?

De fato, com a quarta geração legislativa que se delineia, não sabemos como exatamente será definida a privacidade, já que os debates colocam esse direito para além do indivíduo isoladamente considerado, afirmando que nem sempre a autodeterminação informativa será viável. Na *Data-Driven Society*, é cada vez mais difícil de imaginar um domínio real da privacidade, através do consentimento livre e motivado. E, para além disso, entendemos que a privacidade deve sofrer ainda maiores alterações conceituais.

Se somos seres sociais e vivemos conectados uns aos outros, tanto na vida "real" quanto na "vida virtual" (ainda que essa divisão seja cada vez menos significativa), considerar

a privacidade como um critério eminentemente individual pode ser insuficiente na era dos dados.

A exposição de dados genéticos é um exemplo simples. Com os dados coletados de um único indivíduo, é possível obter informações de familiares próximos e inferir muitas outras de toda uma árvore genealógica. As informações sensíveis, portanto, extravasam o indivíduo, a exemplo de dados de patologias genéticas.

A mesma lógica pode ser utilizada para diversos outros tipos de dados. Por exemplo, as preferências, os hábitos ou a rotina de um indivíduo podem estar diretamente relacionados aos dados da família, ou do grupo social de que faz parte. Nesse sentido, a internet das coisas marca forte presença, já que, captando dados do ambiente, acabam por extrapolar a esfera dos indivíduos. Todos esses aspectos se refletem em uma "privacidade" coletiva, que pode ser violada pela exposição pontual de seus membros.

Nesse sentido, a busca pela proteção da privacidade mostra-se completamente falida no que tange à autodeterminação informativa. Isso porque, ainda que seja assegurado o livre consentimento motivado, o que já não seria uma tarefa fácil, a disposição consensual dos dados tem efeitos na privacidade de terceiros, o que deveria ser considerado nessa equação, principalmente com o evoluir das técnicas de inferência e mineração de dados, comuns em ambientes de dados massivos.

Toda essa discussão se aprofunda ainda mais no que tange aos dados anônimos. Isso porque, se por um lado as características da titularidade são minimizadas na anonimização, por outro, o uso desses dados anônimos permite tratamentos mais flexíveis, que viabilizam a obtenção de informações sobre grupos ou perfis de indivíduos de forma mais simples.

Nesse sentido, a anonimização ressalta o caráter difuso ou coletivo dessas informações, já que revelam importantes informações sobre indivíduos, ainda que não identificados. Essas informações podem ser usadas para fins de inferência e cálculo de probabilidades, a partir da análise de pontos de convergência ou divergência entre indivíduos com características comuns. No entanto, como ressaltado no item que discutimos a natureza jurídica dos dados anônimos, ainda não existe consenso sobre a atribuição de natureza de direito difuso ou coletivo a esses dados.

Portanto, a discussão sobre privacidade mostra grande potencial de transcendência e deve ser considerada nos aspectos orientadores da proteção de dados principalmente quando se pretende modelos efetivos a médio e longo prazo. Entretanto, ainda que esse tipo de demanda seja possível se vislumbrar mesmo com o atual estado da técnica, a legislação internacional e

nacional ainda está consideravelmente presa aos aspectos vinculados à autodeterminação informacional.

3.3.2. A questão comportamental de grupos e a anonimização

A formação de perfis de indivíduos ou a inferência de características de grupos é uma das possibilidades da anonimização. Insere-se nessas possibilidades a utilização de *Big Data Analytics* para estudo dos dados relacionados ao comportamento dos indivíduos, até mesmo para gerar predições confiáveis. Esse tipo de análise se torna comum à medida que as pessoas tornam suas vidas mais públicas, na sociedade-palco que vivemos. Dados sobre comportamento tornam-se presentes e facilmente obtidos, principalmente nas redes sociais e em plataformas ligadas à internet das coisas, já que elas podem coletar dados do cotidiano íntimo dos indivíduos, de hábitos e outros padrões.

As consequências da análise de dados comportamentais talvez não sejam autoevidentes. Por exemplo, Schneier afirma que "é contra-intuitivo, mas é preciso menos dados para identificar exclusivamente o que pensamos" (SCHNEIER, 2015, p. 36). De fato, em um estudo realizado pelo professor Christophe Rosenberg, pesquisador da Escola Nacional Superior de Engenheiros de Caen (França), várias informações diferentes foram possíveis de se obter a partir da análise da simples forma de digitação do usuário das redes. O professor afirma que, com apenas 5 frases escritas, é possível se identificar com 80% de acurácia, se o usuário é mulher ou homem. Além disso, o modo de digitação do usuário pode revelar até mesmo como anda o seu humor no dia (BBC, 2015).

Os perfis comportamentais têm sido bastante utilizados principalmente na disputa de atenção dos mercados, já que são úteis na criação de publicidade personalizada ou na sugestão de mercadorias que fazem parte da gama de interesses prováveis do consumidor.

É famoso o caso do Supermercado Target, que "descobriu" a gravidez de uma adolescente da cidade de Minneapolis, antes mesmo de seu pai. Isso através da análise dos produtos normalmente comprados pela consumidora, os quais estariam estatisticamente associados à gravidez, como sabonetes neutros, ou cosméticos sem perfume (HILL, 2012). Dessa forma, a partir do histórico de produtos comprados pela adolescente, foi possível prever a gravidez, e enviar cupons relacionados a produtos para gestantes que poderiam interessá-la. Esse é um exemplo prático de como os perfis comportamentais podem ser utilizados.

Mas a adequação de um indivíduo a um perfil preestabelecido de forma equivocada pode gerar efeitos nefastos e incompreensíveis ao consumidor. Isso porque a classificação em

um perfil atrela ao usuário uma série de informações que não são necessariamente verdadeiras e das quais dificilmente o consumidor conseguirá a retificação. Isso afeta oportunidade, disponibilidade e acesso a serviços, para citar alguns exemplos.

Além disso, as consequências da formação de perfis comportamentais são dispostas em mão dupla. Ou seja, não apenas o acesso aos bens de mercado é afetado, mas o próprio indivíduo que deseja a inserção é também atingido.

Nesse sentido, Bauman, analisando a formação de perfis e as técnicas de sugestão de consumo, faz emergir uma curiosa dúvida. Ele se pergunta se, afinal, as ferramentas de sugestionamento não estariam, na verdade, criando as demandas de seus próprios produtos, ao invés de estarem oferecendo produtos necessários para o suprimento de demandas preexistentes (BAUMAN, 2012, p. 168). Isso, na aquisição de produtos tanto para consumo quanto para o aperfeiçoamento pessoal relacionado à abordagem do homem como uma mercadoria.

Danilo Doneda também aborda a questão, apontando-a como um dos efeitos colaterais da formação de perfis. Segundo o autor "a partir do momento em que o perfil eletrônico é a única parte da personalidade de uma pessoa visível a outrem, as técnicas de previsão de padrões de comportamento podem levar a uma diminuição de sua esfera de liberdade, visto que vários entes com os quais ela se relaciona partem de pressupostos de que ela adotará um comportamento predefinido, acarretando uma efetiva diminuição de sua liberdade de escolha" (DONEDA, 2006, p. 174).

Em verdade, o poder de sugestionamento dos algoritmos baseados em dados comportamentais já é bastante conhecido nas plataformas digitais como o *spotify*, o *facebook*, as pesquisas do *google*, *Netflix*, etc. A questão que se aborda é, como apontado por Doneda e Bauman, em que proporção a própria liberdade estaria comprometida à medida que assimilamos os sugestionamentos. Schwab expõe essa crítica, quando aborda sobre os limites éticos das ferramentas de *Big Data Analytics*:

> Outra questão importante refere-se ao poder de previsão da inteligência artificial e da aprendizagem automática. Se nosso próprio comportamento torna-se previsível em todas as situações, qual seria o tamanho da liberdade pessoal que temos ou imaginamos ter para nos desviarmos da previsão? Será que isso poderia levar a uma situação em que os seres humanos começarão a agir como robôs? Isso também leva a uma questão mais filosófica: como manter nossa individualidade, a fonte da nossa diversidade e democracia, na era digital? (SCHWAB, 2016, p. 102).

Cath O'Neil alerta sobre essa ameaça aos princípios democráticos e de liberdade individual. A autora relata experimentos de manipulação cognitiva de usuários realizados pelo *Facebook*, apontando que o direcionamento de notícias para grupos de indivíduos é capaz de

modificar sua opinião política, como nas eleições dos EUA e da Índia, ou, em outros experimentos, capaz de modificar o comportamento ou o estado emocional dos usuários (O'Neil 2016, p. 79). Isso é possível uma vez que os sugestionamentos criam verdadeiras bolhas sociais, que tornam mais fáceis a manipulação a partir da análise do padrão comportamental humano.

Para além da ameaça à ampla liberdade de escolha e seus reflexos na democracia, a formação de perfis e os poderes de sugestionamento deles derivados podem causar efeitos ainda mais imediatos e fisiológicos nos indivíduos. A hiperconectividade é por vezes tratada como uma patologia psicológica moderna, onde o indivíduo sente constante medo de não estar disponível ou de estar incomunicável. A pressão pela inserção como produto do mercado da atenção, aponta Bauman, faz com que os indivíduos busquem sua adequação aos perfis mais desejados, de ampla disponibilidade, conectividade, caracterizado por estar sempre alerta e informado:

> "O medo de não conseguir se conformar é superado pelo temor da inadequação, mas nem por isso é menos apavorante. Os mercados de consumo são ávidos por lucrar com esse medo, e as empresas que produzem bens de consumo competem pelo *status* de guias e auxiliares mais confiáveis no interminável esforço de seus clientes para enfrentar o desafio. Elas fornecem "as ferramentas", os instrumentos necessários ao trabalho individualmente realizado de "autofabricação". Os produtos que elas representam como "ferramentas" de uso individual no processo de tomada de decisão são na verdade decisões tomadas por antecedência. Foram feitas sob encomenda muito antes de o indivíduo confrontar-se com o dever (representado como oportunidade) de decidir. É absurdo pensar nessas ferramentas como se fossem capazes de possibilitar a escolha individual do propósito. Esses instrumentos são cristalizações da irresistível "necessidade" – que, agora como antes, os seres humanos devem aprender a usar e a obedecer, e aprender a usar para obedecer a fim de serem livres. Será que o desconcertante sucesso do *Facebook* não é consequência de ele fornecer uma feira em que a necessidade pode encontrar-se todo dia com a liberdade de escolha? (BAUMAN, 2012, p. 168).

Assim, os mercados criam a necessidade e o produto. Mas os efeitos da disputa pela atenção são extremamente interruptivos para os indivíduos. É possível perceber isso nas relações humanas, com o maior distanciamento e dificuldades de empatia. Schwab aborda o tema, citando o escritor Nicholas Carr. Segundo o autor:

> O escritor de tecnologia e cultura, Nicholas Carr, afirma que quanto mais tempo passamos imersos nas águas digitais, mais superficiais se tornam nossas capacidades cognitivas devido ao fato de deixarmos de controlar nossa atenção: "A rede foi projetada para ser um sistema de interrupção, uma máquina voltada para dividir a atenção. As interrupções frequentes dispersam nossos pensamentos, enfraquecem nossa memória e nos deixam tensos e

> ansiosos. Quanto mais complexos forem os encadeamentos dos pensamentos em que estamos envolvidos, maior será o comprometimento causado pela distração" (SCHWAB, 2016, p. 103).

Os indivíduos encontram-se então imersos nesse mercado, um sistema que não dorme, e que cada vez mais se torna presente, opondo-se diretamente à fragilidade humana e sua necessidade de pausa, da reflexão, do "sono"[69] e desconexão. A consequência é a dispersão dos indivíduos na multidão de informações, resultando na própria fragmentação[70].

Dessa forma, a formação dos perfis comportamentais interfere de modo bidirecional tanto no acesso dos indivíduos a produtos e serviços, quanto impactam os próprios indivíduos na necessidade de adequação em perfis reconhecidos no mercado da atenção.

Outra consequência da formação de perfis comportamentais é o aprofundamento de assimetrias informacionais. A dedução ou previsão de informações sobre os indivíduos pelos detentores dos dados, a partir desses perfis, distancia o seu poder decisório em face dos usuários, que por vezes nem cogitam as informações que estão previamente inferidas nessas relações. É o que veremos a seguir.

3.3.2. O paradoxo da Anonimização e o Aprofundamento de Assimetrias Informacionais e Desigualdades Sociais

A anonimização vive o paradoxo entre a disponibilidade e transparência por um lado, e por outro, a privacidade e segurança. Isso porque a anonimização como técnica ganhou repercussão justamente com a promessa de se permitir o livre acesso ao dado, garantindo-se segurança e privacidade. Mas o que percebemos é que essa afirmação não é tão exata. Numa sociedade em que dado é valor pelo seu potencial econômico, social e político, quanto mais dados se mantêm acessíveis, maior o poder dos agentes. Isso porque, como ressaltado em todo o trabalho, os dados podem ser utilizados para gerar informações úteis, guiar os processos decisórios e até mesmo influenciar mudanças de comportamento e opinião dos usuários. Quanto maior a disposição pública de dados e quanto maior a liberdade de seu uso e compartilhamento,

69 O sono é visto como desnecessário e evitável. O sono e a espera são obstáculos às demandas, uma "interrupção sem concessões no roubo do nosso tempo pelo capitalismo" (CRARY, 2016, p.12-38). O autor faz a crítica sobre como o sono, como última barreira ao regime 24/7, passa a ser reordenado, despojado, como objeto de comércio, através dos medicamentos como estimulantes.

70 Conforme aponta Hartmut Rosa: "En resumen, la reacción del individuo a la aceleración social en la modernidad tardía parece resultar en una nueva forma de identidad situacional, en la cual el dinamismo de la modernidad 'clásica', caracterizado por un fuerte sentido de dirección (percibido como progreso), es reemplazado por una sensación de movimento frenético y sin rumbo que es, de hecho, una forma de inercia" (ROSA, 2011.p. 34).

maiores os riscos que são envolvidos na técnica de anonimização, de reidentificação e de obtenção de informações que aprofundam assimetrias.

A assimetria informacional é uma expressão cunhada por Stiglitz para denunciar as dificuldades envolvidas em garantir a livre concorrência no contexto de mercados financeiros (STIGLITZ, 2016). Nesses ambientes, o desequilíbrio de poder dos agentes é acentuado pelo desequilíbrio do acesso a informações, já que a volatilidade desses mercados faz com que a informação seja um ativo imprescindível para o poder de decisão em curto prazo. O termo, no entanto, pode ser estendido para outros contextos em que persista justamente esse desequilíbrio de poder informacional entre agentes nas relações sociais. É possível se identificar ao menos quatro situações diferentes onde essas assimetrias estão presentes: 1) na relação entre *Big Techs* e usuários; 2) na relação entre países emergentes e países detentores do *know how* tecnológico; 3) na relação entre empresas emergentes e plataformas consolidadas, para fins de concorrência; e 4) na relação entre usuários inseridos e indivíduos não inseridos.

Quando pensamos então em plataformas *Big Data* é possível se vislumbrar a acentuação da discrepância de poder e informação a que agentes dessas *Big Techs* têm acesso quando comparados com usuários comuns das redes.

Frank Pasquale denuncia essa discrepância de poder causada pela assimetria informacional cunhando a expressão "*one way mirror*". Segundo o autor, as *Big Techs* têm conhecimento sobre detalhes do cotidiano de seus usuários e podem utilizar esses dados para influenciar comportamentos e opiniões. Por outro lado, os usuários não têm a exata compreensão sobre que informações essas plataformas detêm e como as utilizam, como num espelho unidirecional (PASQUALE, 2015, p. 9).

Essa unidirecionalidade dificulta a compreensão sobe os mecanismos utilizados pelas ferramentas de Big Data Analytics para tomadas de decisão. Nesse sentido, os professores Fabiano Hartmann e Roberta Zumblick afirmam a importância de uma maior reflexão sobre esses mecanismos, de forma a gerar maior transparência, e estimulando "mecanismos de detecção de erros e aplicações inadequadas, numa especialização de sistemas de governança" (PEIXOTO & SILVA, 2019, p. 73-74).

Schwab também aponta para essa questão, quando aborda a possibilidade de que a quarta revolução industrial implique em verdadeiro aprofundamento de desigualdades, já que assimetrias informacionais geram assimetrias de poder (SCHWAB, 2016, p. 71).

E de fato, a assimetria informacional, associada a pouca transparência dos processos decisórios pode resultar em decisões discriminatórias, que aprofundam ainda mais desigualdades sociais. Isso sobre a pretensa neutralidade das decisões, tornando-se em "Armas

de Destruição Matemática" –WMDs. (O'NEIL, 2016), uma vez que minam o acesso da população já fragilizada a serviços, cargos e oportunidades. Assim são criados óbices ao acesso que não podem ser sequer questionados pelos usuários, já que os padrões decisórios e as informações que deram ensejo às decisões não estão claras, além de serem muitas vezes protegidas como estratégias de negócio.

Schwab também aponta que essa assimetria não ocorre apenas entre as *Big Techs* e os usuários, mas também entre países emergentes e os detentores de tecnologia. Nesse sentido, quanto mais alguns países despontam como detentores do acesso total à tecnologia, dominando aspectos técnicos, outros países se distanciam cada vez mais da compreensão e controle dessas tecnologias, passando a ser meros "usuários passivos de uma tecnologia que não entendem" (SCHWAB, 2016, p. 77).

O autor afirma que além de contar com uma infraestrutura digital precária, os países em desenvolvimento também encontram outros desafios para se inserir competitivamente nos mercados digitais:

> O trabalho do fórum sobre *Data-Driven Development* (Desenvolvimento por meio dos dados) destaca que o acesso à infraestrutura digital não é tudo o que importa para poder aproveitar essas oportunidades. A abordagem do "déficit de dados" também é crucial para muitos países, particularmente no hemisfério sul, em virtude das limitações sobre como os dados podem ser criados, coletados, transmitidos e utilizados. Fechar as quatro "lacunas" que contribuem para este déficit — sua existência, o acesso, a governança e a usabilidade — oferece aos países, regiões e cidades competências adicionais que podem melhorar seu desenvolvimento; por exemplo, acompanhamento do surto de doenças infecciosas, melhores respostas às catástrofes naturais, aumento do acesso aos serviços públicos e financeiros para os pobres e compreensão dos padrões de migração das populações vulneráveis (SCHWAB, 2016, pg. 81).

Para poderem usufruir das vantagens existentes na análise massiva de dados, esses países ainda precisam suprir deficiências na coleta e armazenamento de dados. Em vários desses países, incluindo o Brasil, não há o histórico de integração de bases de dados, ou de gestão padronizada. Uma legislação que restrinja de forma aprofundada essa coleta ou armazenamento pelos poderes públicos acaba por inviabilizar o uso de tecnologias de *Big Data* pelo Estado.

Além disso, não é apenas a posição frente ao mercado internacional que está em jogo quando se fala no domínio dessas tecnologias. Na verdade, o que se percebe é que a sociedade informacional envolve até mesmo os fatores de soberania. Schwab aponta que

dificilmente uma guerra na atualidade não envolveria uma faceta cibernética[71]. As operações militares internacionais podem se iniciar e tomar força a partir das redes, inclusive no exercício de guerras não necessariamente declaradas:

> A guerra cibernética apresenta uma das mais graves ameaças de nosso tempo. O ciberespaço tem se tornado um teatro de operações semelhante ao que o solo, o mar e o ar foram no passado. Posso afirmar com segurança que, enquanto qualquer conflito futuro entre agentes razoavelmente avançados poderá ou não ocorrer no mundo físico, ele provavelmente incluirá uma ciberdimensão, simplesmente porque nenhum adversário moderno resistirá à tentação de perturbar, confundir ou destruir os sensores, as comunicações e a capacidade de decisão de seu inimigo. Isso não só irá diminuir o limiar da guerra, mas também irá embaçar a distinção entre guerra e paz, porque quaisquer redes ou dispositivos conectados, tanto os sistemas militares de infraestrutura quanto os civis — tais como fontes de energia, redes de eletricidade, saúde ou controles de tráfego ou abastecimento de água — podem ser hackeados e atacados. O conceito de adversário também é afetado. Ao contrário do passado, não há como ter certeza de quem está atacando você — e até mesmo se foi realmente atacado.
> (...)
> Desde 2008, vem ocorrendo vários casos de ataques cibernéticos, dirigidos a países e empresas específicos, mas as discussões sobre essa nova era da guerra estão ainda no início, e a lacuna entre aqueles que entendem as questões altamente técnicas da guerra eletrônica e aqueles que estão desenvolvendo as políticas cibernéticas amplia-se a cada dia (SCHWAB, 2016, p. 87-88).

Nesse sentido, vemos grandes falhas no caso brasileiro na gestão de dados que envolvem estratégias de segurança nacional. Um exemplo é o vazamento de dados denunciado pelo TCU, pelo programa clube de descontos oferecido aos servidores federais. O vazamento expôs agentes de inteligência da ABIN, tornando públicos seus dados e locais onde estariam infiltrados (VELEDA, WALTENBERB, 2020). Falhas como essa podem colocar estratégias nacionais inteiras em perigo, comprometendo a segurança de todo o país.

A anonimização, nesse contexto, pode trazer soluções para maior proteção de informações sigilosas, assim como manter-se como ferramenta estratégica em casos de guerra cibernética. Por exemplo, a rastreabilidade de um ataque cibernético poderia ser confundida através de técnicas de anonimização, fazendo com que tanto a expertise do mascaramento quanto a da reidentificação se tornem particularmente importantes em contextos de conflito (VAZ, 2018). Nesses casos, a anonimização pode ser uma ferramenta útil para o cumprimento

71 Guerras cibernéticas são definidas pelo manual de Guerra Cibernética do Exército Brasileiro, como o "uso ofensivo e defensivo de informação e sistemas de informação para negar capacidades de C2 ao adversário, explorá-las, corrompê-las, degradá-las ou destruí-las, no contexto de um planejamento militar de nível operacional ou tático ou de uma operação militar. Compreende ações que envolvem as ferramentas de TIC para desestabilizar ou tirar proveito dos sistemas de informação do oponente e defender os próprios Sist Info. Abrange, essencialmente, as ações cibernéticas. A oportunidade para o emprego dessas ações ou a sua efetiva utilização será proporcional à dependência do oponente em relação às TIC" (MINISTÉRIO DA DEFESA, 2017, p. 2-2).

do princípio da dissimulação, evitando, por exemplo, o rastreamento de ataques cibernéticos, mascarando endereços de IP e camuflando o rastro do tráfego de dados na rede.

Entretanto, todas essas técnicas, como mencionamos, têm sua eficácia medida a partir do contexto da técnica e das informações prévias já obtidas. Ohm alerta para isso, afirmando que os agentes se consolidaram sem os óbices agora existentes para o uso e armazenamento de dados terão vantagem evidente na assimetria informacional. Segundo o autor:

> "Mesmo que os cientistas da computação desenvolvam amanhã uma técnica inovadora que protege os dados com muito mais robustez do que qualquer coisa feita hoje - e esse é um improvável "se" -, a nova técnica funcionará apenas nos dados protegidos no futuro; não fará nada para proteger os dados que foram armazenados ou divulgados no passado. Um banco de dados, uma vez lançado, pode se tornar mais fácil de se identificar, mas nunca mais difícil. Longas cadeias de inferências de reidentificação passada não podem ser quebradas com os avanços de amanhã." (OHM, 2010, p.1757).

Ou seja, a pré-existência de grandes bancos de dados já compartilhados dificulta ainda mais a manutenção da anonimização, principalmente quando levamos em consideração agentes com grande assimetria informacional pré-existente. Isso se reflete não apenas na questão da fragilidade dos países emergentes frente aos países desenvolvidos, mas também em questões locais de assimetria de poder.

Já no caso da relação entre empresas emergentes e plataformas consolidadas, a assimetria informacional implica em questões desafiadoras para a livre concorrência.

A professora Ana Frazão alerta que, nesses casos, a assimetria não se concretiza apenas na desigualdade de acesso aos dados, mas também na obtenção do ferramental tecnológico adequado para o tratamento (FRAZÃO, 2021). Ou seja, ainda que, num cenário ideal, fosse possível uma melhor disponibilização dos dados para empresas emergentes, se elas não tivessem acesso às tecnologias de tratamento, denominadas *Big Data Analytics*, persistiriam os abismos concorrenciais.

Além disso, o próprio acesso aos dados deveria ser qualificado. Sabemos que a análise em *Big Data Analytics* é aprimorada à medida que a base é alimentada de dados confiáveis. Se alimentada por dados não confiáveis, as informações obtidas ou os serviços prestados podem não atender às expectativas. Portanto, a qualidade dos dados de tratamento, e não só o acesso genérico a "dados", também interfere em questões concorrenciais (FRAZÃO, 2021).

A crítica da autora é válida para todas as formas de assimetria aqui expostas, mas ganha especial importância na relação entre empresas emergentes e plataformas consolidadas,

uma vez que o modelo de negócio desses mercados é justamente vinculado ao poder de extração de informações a partir do melhor tratamento de dados nos parâmetros da *Data-Driven Economy*. Dessa forma, é muito difícil para os novos atores competirem nos mercados com plataformas já maduras.

Além disso, a professora questiona até mesmo a legitimidade dessas plataformas maduras, que teriam se consolidado na lacuna regulatória, violando princípios como os da boa-fé:

> "Vale ressaltar que, não obstante as inúmeras inovações e benefícios que justificaram a expansão das plataformas, pelo menos parte do poder de muitas delas foi conquistada (i) de forma ilícita, como resultado da exploração indevida dos dados dos usuários, muitas vezes sem o consentimento ou até mesmo sem a ciência destes; ou (ii) aproveitando-se da ausência de regulação de dados, cujo vácuo foi preenchido por uma autorregulação abusiva e sem limites, estabelecida apenas em favor dos interesses das próprias plataformas."

Nesse contexto turbulento, a anonimização surge como mais um fator de complexidade. Com a anonimização, mais dados são disponibilizados para as plataformas maduras, gerando mais poder decisório para os seus negócios além de poder de inferência para quebra da anonimização. Ao mesmo tempo, a anonimização também é uma forma de disponibilização de dados para empresas emergentes, que não têm como competir sem acesso a dados.

Além disso, empresas com maior capacidade de processamento de dados podem obter mais informações dos dados anonimizados. O tratamento de dados mais desenvolvido tem um maior poder de diminuição de entropia dos dados através de maior linkabilidade e inferência entre os dados disponíveis, podendo até mesmo levar à reidentificação.

Importante ressaltar que, numa economia movida a dados, a utilização massiva de informações gera impactos globais. A opção por métodos regulatórios que sejam estritamente rígidos pode aprofundar ainda mais essas assimetrias informacionais, a que estamos submetidos. Países chamados subdesenvolvidos, ou em desenvolvimento, assim como empresas emergentes, para mencionar os exemplos citados, já iniciam essa corrida tecnológica em grande desvantagem competitiva.

Dessa forma, percebemos que, em todas as situações de assimetria exemplificadas, a anonimização funciona como uma via de mão dupla. Se por um lado, a anonimização pode significar em alguma medida a maior disponibilidade de informações para os agentes fragilizados, por outro, o acesso a esses dados pode aumentar ainda mais o poder de inferência das plataformas já maduras, além de dificultar a própria técnica de anonimização, facilitando a

reidentificação. Nessas situações, a assimetria pode ser aprofundada ou reduzida, a depender da disposição, do uso, e da finalidade desses dados.

Lessig já apontava essa característica de ambiguidade latente dessas tecnologias. Para o autor, as tecnologias têm esse caráter de ser ferramentas com potencial para uso benéfico ou maléfico em termos de democracia e liberdades fundamentais.

No caso da ambiguidade latente entre anonimização e identificação, a ausência de identificação é por vezes tratada como um problema com reflexos no mundo real. Isso pode ser percebido principalmente quando lidamos com técnicas de anonimização voltadas para usuário final, ou seja, ferramentas que promovem o anonimato do usuário, de forma diversa das técnicas de anonimização de dados em bancos de dados que temos tratado até aqui.

Lessig narra um experimento que descreve como o anonimato pode fomentar discursos odiosos. Ele descreve que em uma das aulas dele em ambiente digital, em Yale, Lessig permitiu que as pessoas se apresentassem da forma que quisessem, com o nome próprio ou pseudônimo. O autor aponta que apesar de o grupo começar como um ambiente tranquilo, com o passar do tempo, sobressaíam-se os perfis diferentes dos alunos. Em alguns casos, destacavam-se alunos que ele denomina *helpers*, que seriam ajudadores de outros alunos. Em outros casos, surgiam situações emblemáticas como a do aluno de pseudônimo Aidex. Lessig descreve que o aluno era extremamente violento em seu discurso, e as agressões desferidas geravam ainda mais agressão por parte dos demais alunos (LESSIG, 2006). O autor conclui que as regras do espaço, que foram feitas como a maior possibilidade de liberdade de expressão, também permitem esse discurso agressivo. Dessa forma, o anonimato teria o risco de permitir discursos violentos ou incentivar comportamentos espúrios.

Para além do abuso no discurso, as inúmeras possibilidades de violações cometidas por perfis não identificados fazem com que os governos se preocupem com a formação de *backdoors* em tecnologias de criptografia, por exemplo, ou de meios para se rastrear informações (REIDENBERG, 1997) e identificar usuários. Esse paradoxo também foi identificado por Danilo Doneda que a descreve como:

> "como identificou Herbert Burkert, um verdadeiro desafio: hoje, o mesmo legislador que permite (e eventualmente promove) a pseudonímia e a criptografia na Internet, também busca formas de identificar quem é a pessoa atrás de cada e-mail, em um árduo debate entre segurança e liberdades individuais cujas proporções merecem ser examinadas em outra sede" (DONEDA, 2006, p. 180).

Lessig também aborda essa necessidade de saídas para identificação por meio da arquitetura, exemplificando isso com alguns casos, dentre eles o caso *Digital Telephony Act*,

em que a Suprema Corte determina que as companhias telefônicas deveriam utilizar uma arquitetura que viabilizasse a interceptação telefônica pelo poder público (LESSIG, 1996).

Nesse sentido, as discussões sobre regulação do anonimato no ciberespaço envolvem a defesa de mecanismos de escape, para que a anonimização realizada seja viável de reversão, quando necessário para identificar o autor ou autores de um ilícito.

Segundo Schneier, por esse motivo, surgem várias propostas de se eliminar o anonimato na Internet. "A ideia é que, se tudo o que alguém fez fosse atribuível - se todas as ações pudessem ser rastreadas até a origem -, seria fácil identificar criminosos, *spammers*, *stalkers* e *trolls* da Internet" (SCHNEIER, 2015, p. 97).

Interessante destacar que mesmo sobre essa regra de reversão do anonimato, ou seja, de captação da identidade por meio da arquitetura dos códigos, há ambiguidade. Isso porque, enquanto no anonimato do usuário final (*ex ante*) a preocupação é a reversão da técnica para sua reidentificação, na anonimização em bancos de dados (*ex post*) a preocupação é em se promover um anonimato irreversível de forma a impossibilitar a reidentificação.

De fato, a regulação de bancos de dados prioriza que haja a completa dissociação entre o dado e o seu titular, ou seja, fomenta que a anonimização seja completa e irreversível, já que o dado só será considerado anônimo se não puder mais ser, direta ou indiretamente, vinculado ao seu titular. O processo não pode ser revertido, ou seja, não é considerado anônimo, para efeito das leis, o dado que puder ser novamente vinculado ao seu titular por reversão, caso essa reversão possa ser feita por meios razoáveis.

Conclui-se então que a lei não permite que o controlador desses dados anonimizados em bancos de dados deixe em aberto "*backdoors"*, ou seja, não permite saídas na arquitetura do código para que a anonimização seja desfeita. Além disso, a lei encoraja o completo anonimato desses dados para viabilizar que os mercados se utilizem dessas informações de forma livre, sem prejudicar a privacidade dos usuários. Por outro lado, como visto, para a anonimização promovida pelos usuários finais ou derivada da própria arquitetura das redes, a preocupação é justamente a inversa, de se permitir que o usuário seja de alguma forma reidentificado.

É claro que não desconsideramos aqui o fato de que os processos de anonimização em bancos de dados *(ex post)* e o anonimato do usuário final *(ex ante)* estão inseridos em contextos de abstração distintos, conforme ressaltamos no capítulo 2, no item "O que é anonimização", fazendo com que, quanto mais próximo da realidade, maiores os riscos de que a anonimização seja utilizada de forma maliciosa.

Como ressaltamos, a abstração cria um *gap* semântico (DALL'OGLIO, 2007) (FALBO & SOUZA, 2006) entre o mundo real e o objeto de estudo, que se apresentará de forma simplificada para permitir a formação de modelos. Assim há um maior *gap* na anonimização *ex post* quando comparada com a anonimização *ex ante*.

De forma geral, a anonimização *ex ante* tem seu controle de riscos e gerenciamento difuso nas redes. Nesse sentido, as discussões sobre a anonimização *ex ante* se relacionam mais aos temas de governança e regulação da internet. Já na anonimização de bancos de dados (*ex post*), o controlador de bancos de dados normalmente pode ser responsabilizado legalmente pelo mau uso de dados, o que conferiria alguma saída estatal para responsabilização nesse contexto.

Mas não deixa de ser curioso que a lei tenha se preocupado em definir um conceito e uma tipologia tão restrita e voltada à impossibilidade de reversão quando se trata de proteção de dados, na anonimização de bancos de dados. Se isso parece, em um primeiro momento, atender aos anseios populares pela privacidade, talvez, com um olhar mais acurado, se revele como uma forma de atender de fato aos anseios dos mercados.

Isso porque, com a anonimização, é possível que as plataformas mantenham grandes bancos de dados, aptos à utilização de ferramentas como o *Big Data Analytics*, sem que sejam barrados pelas premissas de privacidade de dados pessoais. Por outro lado, caso as legislações encampassem a mesma defesa de reversibilidade dos dados anonimizados pela via da anonimização de banco de dados, a justificativa para diferenciar dados anônimos e dados pessoais sofreria um enfraquecimento[72]. Com isso, o livre uso dos dados anônimos estaria ameaçado e, com ele, as diversas estratégias de mercado para captação de informação dos dados em contextos massivos.

De todo modo, as nuances da ambiguidade da técnica de anonimização suscitam discussões sobre a possibilidade de regulação e controle sem que fosse necessária a completa identificação do usuário, mantendo algum nível de anonimato.

Seguindo na linha das ambiguidades que circundam a anonimização, temos uma última forma de relação assimétrica identificada, que se consubstancia do paradoxo da conectividade. Trata-se do crescente abismo informacional que existe entre os conectados e os desconectados. Por mais curioso que possa parecer, a impossibilidade de se posicionar enquanto indivíduo como mercadoria nesses mercados é fator de exclusão de visibilidade e de acesso. As informações que chegam pelas redes e as oportunidades que dela derivam são fatores

72 Isso ocorre, por exemplo, com as técnicas de pseudonimização, onde a capacidade de reversão do tratamento é presumida, fazendo com que a lei interprete os dados pseudonimizados como dados pessoais.

dificilmente alcançados por quem não está inserido nessa conectividade. A visibilidade nas redes repercute então em visibilidade social como um todo, gera poder, dá acesso à informação e inserção nos mercados.

De fato, a visibilidade é também um bem, na era dos dados, que faz com que produtos e políticas públicas sejam direcionados para aqueles que estão incluídos na dinâmica dos fluxos de dados. Nesse mesmo sentido, discutimos sobre a ressignificação da privacidade e a reconfiguração da exposição como algo não necessariamente intrinsecamente ruim numa economia dos dados.

A anonimização e a divulgação de dados, desse ponto de vista, contribuem para a acessibilidade, sendo que grupos marginalizados, que não têm seus dados envolvidos nesses fluxos, se encontram prejudicados pela ausência de visibilidade. Dessa forma, a desconexão é sinônimo de marginalização, e sofre as consequências da ausência de visibilidade e, por fim, da falta de voz nos processos decisórios que são perpetrados nas redes.

Esse é um tipo de assimetria que é aprofundada com a disponibilização de dados públicos nas redes. É difícil imaginar que qualquer cidadão, ainda que não tenha acesso a internet em si, não tenha nenhum dado seu disposto nas redes, justamente devido aos programas de transparência e acessibilidade governamental.

A inclusão digital se torna então um dos aspectos da democracia e deve ser levada em consideração. As decisões sobre os dados, sua anonimização ou sua gestão devem ganhar o mais amplo espectro de discussão pública quanto possível, já que interfere diretamente na vida dos cidadãos.

4. Desafios do Framework de Requisitos de Anonimização da LGPD: entre o legislado e os limites da Anonimização.

Os limites expostos acerca da anonimização trazem um melhor panorama sobre a técnica e sobre os desafios de sua manutenção e da utilização dos dados anonimizados.

Partimos das principais críticas de Ohm e seus desdobramentos para refletir sobre pontos sensíveis da anonimização que devem ser levados em consideração pelos mais diversos agentes, em especial, pelos reguladores, em prol da privacidade de dados.

Como mencionado no início desse trabalho, a anonimização faz parte das técnicas de regulação "*privacy by design"*, ou seja, aquelas técnicas que buscam a privacidade através da própria arquitetura da ferramenta. O objetivo desse tipo de técnica é evoluir a arquitetura de forma que a privacidade seja um dos aspectos relevantes na própria configuração dos dados.

Paul Own exemplifica esse tipo de técnica afirmando que, dentre a infinidade de dados gerados pelos usuários ao ingressar um servidor web, grande parte dos dados não é coletada, porque há um limite para essa coleta estipulado pelo próprio padrão do software[73] (OHM, 2010, p. 1710). É isso inclusive que permite um maior grau de abstração entre a anonimização *ex post*, quando comparada à anonimização *ex ante.*

Por sua vez, a anonimização *ex post* garante que os dados, uma vez coletados, serão tratados para que as informações que deles se possa extrair sejam limitadas de forma suficiente a proteger o titular.

A ideia de uma regulação para além dos instrumentos legais de proibição e repressão tem estado muito presente quando se trata da regulação das novas tecnologias. Lessig traz esse debate já nas primeiras linhas do seu livro CODE 2.0, relacionando conceitos como regulação, liberdades, direitos e arquitetura. Para o autor, haveria uma regulação intrínseca, realizada pelos mercados e pelos governos, concretizada através do código. É o que Lessig chama de regulação pela arquitetura (LESSIG, 1997).

O autor entende que a regulação é inevitável, posto que o código é inevitável. Então caberia aos agentes desenhar e codificar as tecnologias de modo a proteger princípios fundamentais, ou desenhar e codificar o ciberespaço de modo que esses princípios desapareçam. (LESSIG, 2006).

Sabendo que a própria técnica é uma forma de regulação e que a lei pode apontar diretrizes para o desenvolvimento desta técnica, nos propusemos neste trabalho a elencar quais são as diretrizes que a Lei de Proteção de Dados Pessoais Brasileira aponta para o tratamento da anonimização enquanto técnica.

E, como toda forma de regulação pela arquitetura, o conhecimento dos limites que a técnica de anonimização apresenta pode orientar os reguladores a ajustar os pontos frágeis inclusive por alterações na arquitetura da própria técnica.

A fim de mapear o que a lei dispõe acerca da anonimização, elaboramos um *framework* que expõe quais são os requisitos trazidos pela LGPD para se considerar um dado como anonimizado.

Faremos ainda a análise crítica dos requisitos legais do *framework* e a avaliação se eles são suficientes para orientar os desenvolvedores sobre os critérios da anonimização, com

73 Nas palavras do autor: “The vast majority of web servers collect much less than the maximum amount of information available about your visit, not due to the principled privacy convictions of their owners, but because the software saves only a limited amount of information by default” (OHM, 2010, p. 1710).

base nos limites já expostos da técnica. A partir dessa análise, caso seja identificada alguma lacuna nos requisitos legais, pretendemos descrevê-la e analisá-la.

Desta forma, pretendemos, nesses próximos itens, responder duas principais perguntas que se consubstanciam no problema de pesquisa deste trabalho: 1) Qual *framework* pode ser traçado sobre os requisitos legais para o anonimização na LGPD?; e 2) Os requisitos definidos pela Lei Geral de Proteção de Dados Brasileira para anonimização, elencados no *framework*, levam em consideração os principais limites da técnica?

É o que discorreremos a seguir.

4.1. Elicitação de requisitos e formulação do framework legal da anonimização

A engenharia de requisitos tem um importante papel na busca pela maior confiabilidade dos softwares pelos usuários. Trata-se de um ramo da engenharia de software que se preocupa com o atendimento das expectativas dos mais diversos *stakeholders* (ALTARTURI et al., 2017) (RUEDA et al., 2020). Dessa forma, é através dos requisitos e para atender as demandas ali elencadas que são pensadas as soluções tecnológicas.

A atividade de extrair dos clientes suas necessidades e expectativas sobre a ferramenta e transformar essas demandas em especificações textuais passíveis de codificação, implementação e/ou proposta de soluções (requisitos) é denominada "elicitação" (CHEMUTURI, 2013).

Portanto a elicitação é a etapa onde são levantados, em diálogo entre os desenvolvedores, engenheiros de software e usuários finais, os requisitos que a ferramenta deve cumprir na conclusão do projeto. Através da elicitação de requisitos definem-se os objetivos da ferramenta a ser construída, que diretrizes ela deve respeitar durante o processamento e qual o passo a passo a ser seguido pelo desenvolvedor, sempre tendo em mente as demandas levantadas pelo usuário final.

Os requisitos podem ser classificados em funcionais e não funcionais e são estabelecidos de diferentes maneiras[74], como entrevistas com as partes interessadas, análise documental das características essenciais da ferramenta, *brainstorming, design thinking*, além de outras técnicas ou mesmo do uso combinado de técnicas (ALTARTURI et al., 2017) (CHEMUTURI, 2013).

74 Algumas dessas técnicas são explicadas didaticamente no site: http://retraining.inf.ufsc.br/guia/app/classificacoes/tecnicas-de-elicitacao-de-requisitos.

Os requisitos funcionais são aqueles que se relacionam à funcionalidade em si do *software* em desenvolvimento, voltados aos objetivos finais da ferramenta, ou seja, aos resultados que dela se espera. Dessa forma, normalmente são requisitos voltados ao resultado, à qualidade, ao tempo ou eficácia da ferramenta em atender a demanda a que ela se propõe, específica de cada projeto de *software*.

Já os requisitos não-funcionais são aqueles que estão relacionados com restrições ou medições de qualidade da ferramenta, voltados não necessariamente à funcionalidade do software, mas aos aspectos éticos, de segurança da informação e demais diretrizes voltadas ao controle operacional e finalístico. São requisitos que buscam balizar a ferramenta e seus objetivos em princípios que vão além do simples cumprimento dos objetivos funcionais do *software*.

A lei, por exemplo, pode dar ensejo a um requisito funcional ou não funcional, a depender da sua repercussão na funcionalidade do *software.* Mas de qualquer forma, podemos elencar a lei e o *compliance* à legislação como requisitos que o desenvolvedor deve perseguir.

Questões como transparência, privacidade e ética são cada vez mais suscitadas como requisitos, não apenas não funcionais, mas também propriamente funcionais, já que os usuários demandam, como mencionamos, o controle sobre esses parâmetros no produto final desenvolvido. Há, inclusive, uma tendência crescente nos próprios modelos de desenvolvimento de *software,* de uma maior participação do usuário final em cada etapa do processo de construção das ferramentas. Nesse sentido, a evolução das técnicas de *predictive process* em meados dos anos 70, para técnicas mais interativas, como *Spiral, RAD, RUP*, até os processos ágeis que ganham adeptos hoje em dia, como o *SCRUM, XP, Lean, Opens UP, FDD, Crystal,* etc.

Essas técnicas basicamente significaram um maior peso decisório dentro dos processos de desenvolvimento da aplicação, reformulação ou reiteração dos requisitos às partes interessadas, promovendo uma maior sinergia e engajamento entre os desenvolvedores e os usuários finais.

Escolhemos para esta pesquisa os requisitos definidos pela lei sobre a anonimização, portanto temos como fonte para estabelecer os requisitos a análise do texto trazido pela LGPD.

Isso porque a lei de proteção de dados traça diretrizes sobre que tipo de tratamento pode ser considerado válido para anonimização, já que aponta que dado é considerado anonimizado. Essas disposições legais interferem diretamente nos requisitos que um *software* ou um tratamento de anonimização deve possuir para estar em *compliance* legal. Além disso, a

lei estabelece diretrizes gerais para o tratamento de qualquer dado pessoal, que devem ser levadas em consideração durante o processo de anonimização.

Assim, para listar corretamente os requisitos legais para a anonimização na LGPD, é necessário diferenciar os **requisitos preliminares** do tratamento de dados, dos **requisitos internos** ao tratamento de anonimização, que se dividem em princípios gerais para o tratamento de dados pessoais e requisitos específicos da anonimização, sendo que todos estes requisitos devem ser devidamente cumpridos pelo desenvolvedor.

Os "**requisitos preliminares para o tratamento de todo e qualquer dado pessoal"** são requisitos que precedem o processamento. Eles não estão atrelados à finalidade do tratamento, portanto são requisitos não funcionais e devem estar presentes, em alguma de suas hipóteses, antes de qualquer tratamento de dados pessoais.

Por outro lado, os "**Requisitos internos ao tratamento de anonimização**" são aqueles definidos por lei para considerar dados como anonimizados, ou seja, estão relacionados ao objeto resultado do tratamento. Dentro do processo de anonimização, podemos listar os requisitos funcionais e não funcionais.

Os requisitos internos ao tratamento da anonimização dividem-se em duas modalidades: 1) **"princípios gerais para o tratamento de dados pessoais**"; e 2) "**requisitos específicos da anonimização como tratamento de dados**".

Os "princípios gerais para o tratamento de dados pessoais" são princípios orientadores para o processamento de dados. São, na sua maioria, requisitos não funcionais, uma vez que são diretrizes para o processamento de dados em geral e, portanto, não vinculados às funcionalidades da ferramenta. Com exceção da fase posterior da responsabilização (*accountability*), que tem um aspecto funcional na anonimização, as exigências desta etapa estão mais relacionadas aos aspectos éticos do tratamento.

Por sua vez, os "requisitos específicos da anonimização como tratamento de dados" são requisitos funcionais relacionados à qualidade do processo de anonimização, que precisa gerar dados que não estão mais associados a uma pessoa específica.

Para explorar ainda mais os tipos de requisitos legais, discutiremos cada um nos tópicos a seguir.

4.1.1. Anonimização como uma forma de tratamento de dados pessoais

Antes de avançar para os requisitos legais da anonimização de dados é muito importante destacar que a anonimização se classifica como uma forma de processamento de

dados pessoais. Isso porque a anonimização lida, como seu objeto de entrada, com dados pessoais *lato sensu*, o que inclui as categorias, dados pessoais *stricto sensu* e dados pessoais sensíveis.

Por esse motivo, conceituamos a anonimização, no segundo capítulo, como uma técnica de tratamento de dados pessoais, na modalidade "processamento de dados", cujo objetivo é a dissociação de dados, dispostos em banco de dados, públicos ou privados, organizados ou desorganizados, dos seus respectivos titulares, considerando meios legais de razoabilidade.

Portanto, para que o tratamento seja possível, a anonimização deve respeitar os critérios legais para o tratamento dos dados pessoais, de forma semelhante a qualquer outro tratamento de dados.

Entretanto, cumpre ressaltar que a anonimização é um tratamento de dados pessoais *sui generis*, com peculiaridades que devem ser observadas, mas que não a descaracteriza como tratamento de dados pessoais.

A peculiaridade se traduz no fato de que, diferente dos outros tratamentos em que o objeto de processamento consiste em dados pessoais tanto na entrada quanto na saída, na anonimização teremos como resultado do tratamento dados anônimos.

Desta forma, a anonimização é o único processamento de dados que implica na transubstanciação de seu objeto, já que, os dados entram como dados pessoais e saem como dados não mais pessoais, mas anonimizados. Esses dados anonimizados, como vimos, são submetidos a regras mais flexíveis pela legislação, uma vez que já não são considerados pela lei como pessoais. É o que descrevemos nas Figuras 4 e 5:

Figura 4: Processamento de dados pessoais

Figura 5: Anonimização como tratamento de dados pessoais

Ressaltamos, no entanto, que essa peculiaridade da anonimização não a descaracteriza como um tratamento de dados pessoais. Isso porque a lei não excepciona os dados pessoais das regras de proteção e os dados só poderão ser considerados como desvinculados de seu titular em algum momento no decurso do tratamento da anonimização. Nesse sentido, o teor do art. 5º, X, da LGPD, que define tratamento de dado pessoal como “toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção,

classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração".

Dessa forma, apesar de o tratamento da anonimização manifestar particularidades que destacaremos ao longo da construção do *framework*, entendemos que todos os requisitos inerentes ao tratamento de dados pessoais devem também ser observados pela anonimização à medida de sua factibilidade.

4.1.2. Requisitos preliminares para processamento de todo e qualquer dado pessoal

A LGPD dispõe que o tratamento de dados pessoais só pode ser realizado se estiver dentro de alguns parâmetros previamente estabelecidos. O artigo 7º enumera 10 casos em que o tratamento de dados pessoais é autorizado. São eles, o tratamento:

1) mediante consentimento do titular;
2) para cumprimento de obrigação legal ou regulatória;
3) para a execução de políticas públicas pelo governo;
4) para a realização de estudos por órgão de pesquisa, garantindo, sempre que possível, a anonimização dos dados pessoais;
5) para execução de contrato ou procedimentos preliminares relativos a contrato, a pedido do titular;
6) para o exercício regular de direitos em processos judiciais, administrativos ou arbitrais;
7) para a proteção da vida ou incolumidade física do titular ou de terceiros;
8) para a tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária;
9) para atender aos legítimos interesses do controlador ou de terceiro;
10) para proteção de crédito.

Portanto, o tratamento de dados pessoais está condicionado às hipóteses previstas na lei. O consentimento do titular é um dos mais conhecidos, e é definido como a "manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados

pessoais para uma finalidade determinada"[75]. Ele se dará, normalmente, por escrito[76], exceto quando o próprio titular tenha tornado os dados públicos[77].

Mas existem ainda outras formas legais de tratamento dos dados pessoais, conforme listado. A maioria das hipóteses prevê situações bem específicas, como o tratamento de dados pessoais para fins de políticas públicas pela administração pública, ou para cumprimento de obrigação regulatória, pelo controlador.

Vale ressaltar que, para as hipóteses 3, 4 e 8 elencadas, a permissão de tratamento é específica para determinados agentes, a saber, governo, órgão de pesquisa e profissionais de saúde, serviços de saúde ou autoridade sanitária, respectivamente. Destacamos também a preferência legal pela anonimização na hipótese 4.

Há, no entanto, previsão mais abstrata de tratamento legal, como quando o controlador ou terceiro tenha interesse legítimo. Essa última hipótese é bastante controvertida, uma vez que o próprio conceito e os limiares do que é um interesse legítimo ainda não estão claros. Ressalta-se que o GDPR tem previsão muito semelhante[78] e igualmente vaga acerca do tratamento mediante legítimo interesse, excepcionando dessa possibilidade de tratamento apenas os casos em que os dados pessoais digam respeito à criança.[79]

Já a LGPD aponta duas hipóteses, em rol não taxativo, sobre o que seria considerado legítimo interesse, quais sejam, o apoio e promoção de atividades do controlador; e a "proteção, em relação ao titular, do exercício regular de seus direitos ou prestação de serviços que o beneficiem", desde que se respeitem as legítimas expectativas do titular, seus direitos e liberdades fundamentais.[80]

Ressalte-se que, conforme veremos adiante, o legítimo interesse não pode ser utilizado para tratamento de dados sensíveis, já que o tratamento desses dados está disposto em rol taxativo no art. 11 da LGPD.

De forma geral, pode-se inferir que o legítimo interesse não pode desrespeitar os limites principiológicos estabelecidos pela lei, principalmente a finalidade a que se destina o tratamento, a adequação e também ao princípio da transparência, princípios que veremos com mais profundidade nos próximos subitens.[81] Portanto, mesmo em hipóteses mais abstratas,

75 Conforme art. 5º, XII.
76 Conforme art. 8º caput da LGPD.
77 Confirme art. 7º, §4º da LGPD.
78Conforme destacamos na Tabela Comparativa (Figura 1).
79 Art. 6 (1)(f); "Considerandos" precedentes ao texto do GDPR, ponto 47.
80 Conforme art. 10 da LGPD.
81 Conforme art. 10, §2º.

como no legítimo interesse, há a garantia aos titulares que eles sejam informados sobre o tratamento[82], ainda que não se exija o consentimento.

Por fim, pontue-se que, quanto ao legítimo interesse, há previsão de que a autoridade nacional pode solicitar relatório de impacto para tratamentos baseados nessa modalidade de pré-requisito,[83] e que o controlador e o operador deverão manter registro dessas operações de tratamento[84].

Além desses casos elencados no artigo 7º e seus incisos, temos três outras hipóteses de pré-requisitos de tratamento de dados pessoais que podem ser extraídas do texto legal. São elas:

11) O tratamento de dados pessoais públicos, ou seja, de dados pessoais cujo acesso é público, nos termos do art. 7º, §3º.
12) O tratamento de dados tornados manifestamente públicos pelo titular, nos termos do art. 7º, § 4º, que dispensam a necessidade de consentimento;
13) O tratamento ao término do tratamento original, o que permite a manutenção dos dados para uso exclusivo do responsável após finalizado o tratamento, vedado o acesso de terceiros, e desde que os dados sejam anonimizados, de acordo com o art. 16, IV.

Quanto aos dados pessoais públicos (hipótese 11) a lei permite, no art. 7º, §3º, que o tratamento seja realizado mesmo que não preenchidos os requisitos ordinários de tratamento (como é o caso do consentimento), mas desde que considere o princípio da finalidade, a boa-fé e o interesse público que justificaram a disponibilização do dado. Já os dados tornados públicos pelo titular (hipótese 12) podem ser tratados com dispensa de consentimento, desde que respeitados os princípios gerais dispostos na lei, e os direitos do titular [85].

Quanto à última hipótese (hipótese 13), há tripla restrição: o agente de processamento é específico (o controlador responsável originalmente pelo tratamento dos dados), o tipo de tratamento é específico (deve se ater à modalidade de processamento de dados da anonimização), e o uso de dados anonimizados também é limitado, por se restringir ao uso exclusivo do controlador. O acesso de terceiros é proibido. Essa restrição de uso deve ser

82 Conforme art. 6º, VI.
83 Conforme art. 10, §3º.
84 Conforme art. 37.
85 Conforme art. 7º, §4º.

enfatizada, pois é exclusiva dessa forma de tratamento. Assim, enquanto se atendidos outros requisitos preliminares o controlador pode utilizar livremente os dados anonimizados, neste caso de anonimização ao término do tratamento, os dados ficam restritos à sua própria base e ao acesso exclusivo do controlador.

Os casos listados estão dispostos no *framework* descrito na Figura 6:

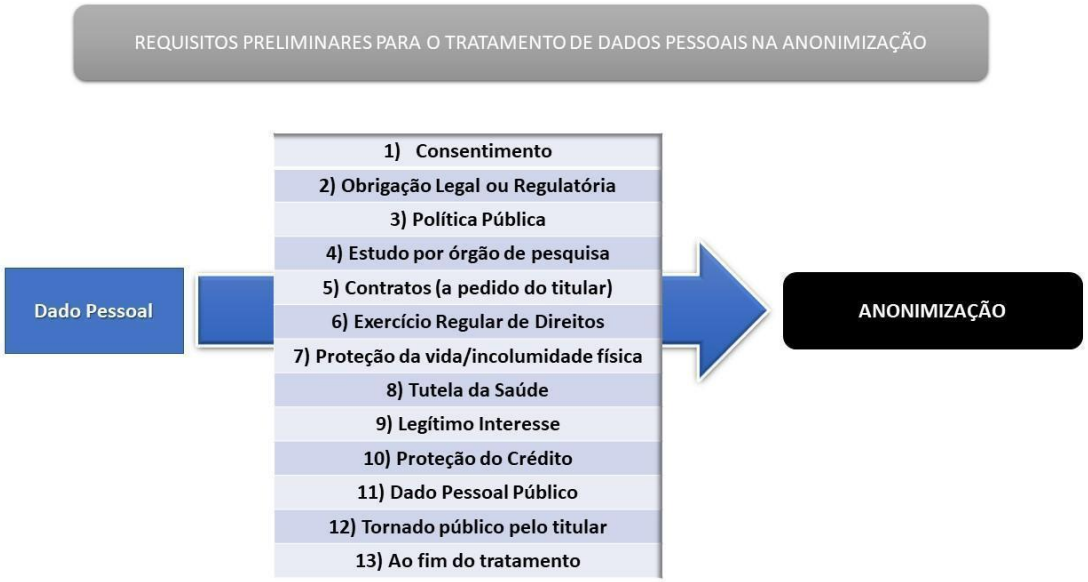

Figura 6: Requisitos preliminares para processamento de dados pessoais na anonimização

Dessa forma, o dado só poderá ser submetido ao processo de anonimização se estiver relacionado a alguma ou algumas dessas hipóteses estabelecidas como requisitos preliminares.

Por outro lado, a lei especifica que o tratamento de dados pessoais sensíveis ocorre em casos mais restritos de permissão. Dados pessoais sensíveis são aqueles relacionados a raça ou etnia, religião, posição política, saúde, orientação sexual, dados biométricos e genéticos, para citar exemplos (artigo 5º, II). São dados que recebem maior proteção por possuírem um caráter existencial proeminente e, portanto, potencial de discriminação contra os indivíduos. O artigo 11 versa sobre o tratamento de dados sensíveis, que se restringe aos casos de tratamento:

1) mediante consentimento do titular;
2) para cumprimento de obrigação legal ou regulatória;
3) para a execução de políticas públicas pelo governo;
4) para a realização de estudos por órgão de pesquisa, garantindo, sempre que possível, a anonimização dos dados pessoais sensíveis;

...

6) para o exercício regular de direitos em processos judiciais, administrativos ou arbitrais;

7) para a proteção da vida ou incolumidade física do titular ou de terceiros;

8) para a tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária;

Portanto, os casos 5, 9, 10, 11, 12 e 13, apresentados como pré-requisitos para tratamento de dados pessoais, são excluídos como hipóteses de tratamento para os dados pessoais sensíveis. Os casos 5, 9 e 10 [“5) para execução de contrato ou procedimentos preliminares relativos a contrato, a pedido do titular”; “9) para atender aos legítimos interesses do controlador ou de terceiro”; “10) para proteção de crédito”] foram excluídos por falta de disposição expressa. Por sua vez, os casos 11, 12 e 13 [“11) De dados pessoais públicos”; “12) Que dispensa consentimento, para dados tornados manifestamente públicos pelo titular”; “13) Que permite a manutenção dos dados após o término do tratamento para uso exclusivo do responsável pelo tratamento, vedado o acesso de terceiros, e desde que os dados sejam anonimizados”] foram excluídos por dedução, uma vez que a sua previsão se refere aos dados pessoais *stricto sensu*, enquanto os dados sensíveis têm uma maior restrição de tratamento por intenção legislativa.

A lei traz ainda um caso específico de tratamento de dados sensíveis, que chamaremos de 14º caso, ou seja, o tratamento:

14) para garantir a prevenção de fraudes e segurança do titular, nos processos de identificação e autenticação de cadastro em sistemas eletrônicos (art. 11, II, g).

Delineamos essas hipóteses de requisitos preliminares para tratamento de dados sensíveis por meio da Figura 7:

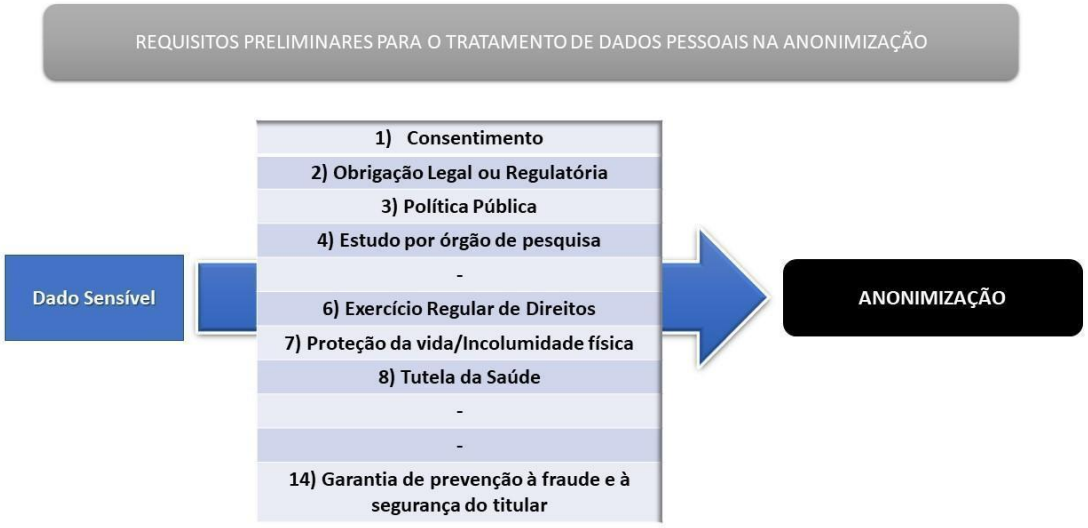

Figura 7: Requisitos preliminares para o tratamento de dados pessoais sensíveis na anonimização

Portanto, para que seja possível dar início ao processo de anonimização, algumas dessas hipóteses jurídicas descritas na Figura 6 e Figura 7 devem preceder o tratamento, a depender do tipo de dado a ser tratado. Atendido esse requisito, pode-se dar início ao tratamento de anonimização que, por sua vez, possui seus requisitos próprios. Isso é o que analisaremos a seguir.

4.1.3. Requisitos internos ao tratamento de anonimização

Dentre esses requisitos, que fazem parte já do processamento da anonimização, teremos aqueles que são requisitos não-funcionais, caracterizados pelos princípios que devem reger internamente o tratamento de dados pessoais, e aqueles que são funcionais, ou seja, que balizam o resultado do tratamento, qual seja, a produção de dados anonimizados.

4.1.3.1. Princípios Gerais para o Tratamento de Dados Pessoais

Como em qualquer outro tratamento de dados pessoais, a anonimização deve respeitar os princípios previstos na lei, levando-se em consideração, no entanto, a característica *sui generis* desse tipo de processamento de dados. O Artigo 6 prevê um total de dez princípios, cuja definição transcrevemos literalmente:

1) Finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;

2) Adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;

3) Necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;

4) Livre Acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;

5) Qualidade dos Dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

6) Transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;

7) Segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;

8) Prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;

9) Não Discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;

10) Responsabilização e Prestação de Contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

Dessa forma, esses princípios estabelecem limites éticos e práticos para o controlador em qualquer tratamento de dados e garantem que o titular tenha um maior controle de seus dados, através do acesso à informação.

Alguns princípios, como finalidade, adequação e livre acesso, esclarecem que, mesmo nos casos em que não é necessário o consentimento, o titular deve ser informado sobre o tratamento e ter acesso a esses dados utilizados, enquanto o controlador deve fazer exatamente o que informou como finalidade do tratamento ao titular. Assim, todo o tratamento é pautado

pelos fins legítimos definidos pelo controlador, que serve de parâmetro para limitar os dados utilizados ao estritamente necessário para esses fins, conforme dispõe o princípio da necessidade.

Portanto, ressalta-se que é dever do controlador informar ao titular qualquer alteração sobre a finalidade do tratamento de seus dados, permitindo que o titular revogue as permissões que concedeu ao uso de seus dados. [86]

Por isso, além de obedecer aos princípios, o tratamento deve respeitar a boa-fé, ou seja, as expectativas geradas pelo titular dos dados. É isso que aponta o caput do art. 6º ao expressamente prever a boa-fé como regra obrigatória de qualquer tratamento aliada aos princípios elencados.

Dessa forma, podemos descrever que todo o tratamento de dados pessoais deve obedecer ao *framework* não-funcional relacionado aos princípios gerais para o tratamento disposto a seguir, na figura 8:

Figura 8: Diretrizes Principiológicas para o tratamento de dados

É preciso, no entanto, realizar a observação de que, em algumas das hipóteses de requisitos preliminares para o processamento de dados pessoais, a lei flexibiliza a incidência desses princípios. É o caso do tratamento de dados pessoais públicos, ou tornados públicos pelo titular.

86 Conforme art. 8º, § 6º e art. 9º, § 2º.

Para ambos os casos, a lei flexibilizou o princípio da finalidade, que é o princípio que vincula o tratamento do dado à finalidade previamente estabelecida pelos agentes de tratamento. Essa flexibilização se dá uma vez que a lei definiu como possível o tratamento posterior desses dados para novos objetivos (que não aqueles previamente estabelecidos), sendo que nesses casos a lei exige apenas que sejam "observados os propósitos legítimos e específicos para o novo tratamento", além dos direitos do titular, os fundamentos e princípios da lei[87].

De toda a forma, conforme o texto legal, os demais princípios permanecem intactos, sendo necessária sua observância para os tratamentos posteriores realizados com novos objetivos.

Por fim, cabe ressaltar que, para o tratamento de dados públicos pessoais a lei estabelece que a finalidade, a boa-fé e o interesse público que justificaram a disponibilização do dado devem ser levados em consideração na interpretação das diretrizes principiológicas.

Podemos exemplificar essa situação com a Figura 9, que dispõe sobre os princípios para o tratamento de dados públicos pessoais:

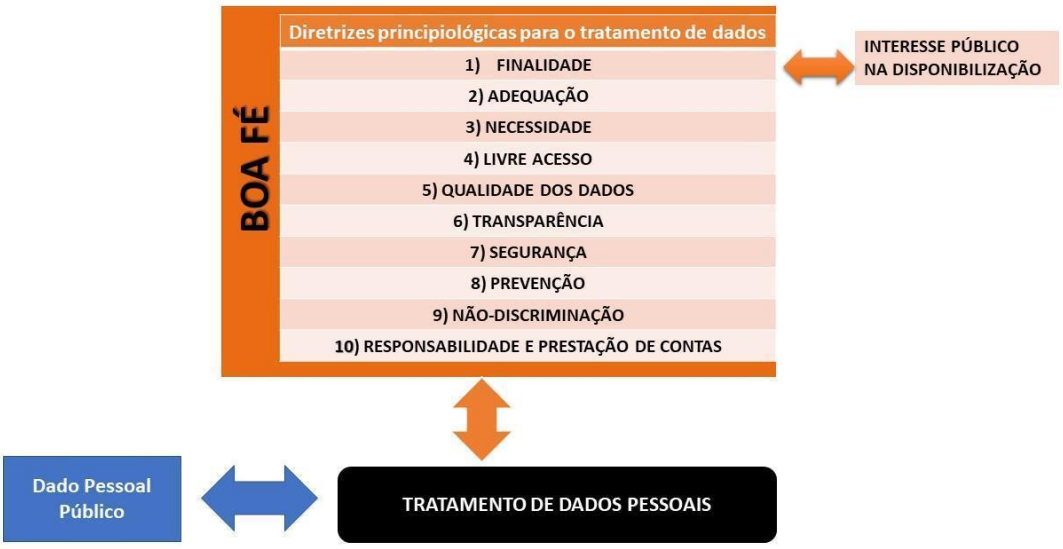

Figura 9: Diretrizes Principiológicas para o tratamento de dados públicos pessoais

Por sua vez, no caso da anonimização, o cumprimento de alguns desses princípios é desafiador. O princípio da finalidade, por exemplo, prevê que o tratamento deve ter "finalidades legítimas, específicas, explícitas e informadas para o titular, sem possibilidade de processamento posterior de forma incompatível com essas finalidades", e encontra equivalente

87 Conforme art. 7º, §7º.

no GDPR no art. 5º (1)(b) (princípio da limitação das finalidades). De forma semelhante, o princípio da adequação determina a "compatibilidade do tratamento com as finalidades informadas ao titular". No entanto, uma vez anonimizados, os dados podem ser utilizados de forma mais flexível, mesmo para finalidades que não foram originalmente definidas pelo responsável pelo tratamento. Isso ocorre porque os dados anônimos não são mais dados pessoais e as restrições LGPD não se aplicam mais a eles.

Neste caso, o objetivo imediato do tratamento seria o próprio processo de anonimização. Por exemplo, a parte deve ser informada de que seus dados estão sendo anonimizados e que a finalidade do tratamento é a própria anonimização.

Desta forma, percebemos que os princípios foram estipulados para incidir no tratamento de dados pessoais *lato sensu.* No caso da anonimização, no entanto, em certo momento do tratamento, os dados perdem a característica de serem dados pessoais, desobrigando-se, portanto, da incidência dessas diretrizes principiológicas.

Nesses casos, os princípios encontram aplicação direta, ao menos de acordo com a interpretação literal da lei, apenas enquanto os dados ainda estiverem no processo de anonimização. Após anonimizados, os dados se desincumbem do cumprimento estrito dos princípios elencados. É o que se descreve na Figura 10:

Figura 10: Visão Geral das Diretrizes Principiológicas para a anonimização

Da mesma forma, o princípio da necessidade também é aplicado de forma peculiar na anonimização. Este princípio, também tratado pelo GDPR como princípio da minimização

dos dados[88], refere-se à "limitação do tratamento ao mínimo necessário para a realização de suas finalidades"[89]. O princípio da necessidade limita a utilização dos dados, fazendo com que sejam tratados apenas os dados "pertinentes, proporcionais e não excessivos" para se atingir a finalidade almejada e informada.

No caso de anonimização, é desejável que o tratamento seja realizado com a maior quantidade de dados possível, uma vez que os dados estariam mais protegidos na forma anonimizada. A fronteira sobre quais dados seriam viáveis de anonimização seria traçada pela perda de utilidade dos dados, conforme observado por alguns autores (ALTARTURI et al., 2017) (DOMINGO-FERRER, 2019) (BRASHER, 2018) (OHM, 2010).

Apesar dessas incompatibilidades na aplicação integral dos princípios à anonimização, entendemos que eles devem ser aplicados tanto quanto possível, pelo menos durante o processo de anonimização. Isso porque, como mencionado, durante o período de processamento ainda temos dados pessoais que devem ser protegidos pela LGPD.

Dessa forma, princípios como o Livre Acesso, a Qualidade dos Dados e a Transparência devem ser assegurados ao titular durante o processamento da anonimização. Isso garante aos indivíduos o acesso a quais dados estão sendo anonimizados, e até mesmo a retificá-los caso exista incompatibilidade.

Os princípios da Segurança e da Prevenção, nos moldes da LGPD, também se mantêm aplicáveis na medida em que o processamento lida com dados pessoais. É factível, no entanto, que elementos destes princípios (ainda que não nos moldes da LGPD) sejam mantidos mesmo com os dados anonimizados, já que a segurança da informação é um dos elementos da segurança dos dados observados de maneira mais disseminada pela Governança Corporativa.

Por sua vez, o princípio da Não Discriminação se traduz num princípio de conduta ética que não encontra limitações à sua aplicação em todas as etapas do tratamento da anonimização. Apesar de não existir a exigibilidade expressa da aplicação deste princípio após o tratamento da anonimização, é certo que este é um princípio que atende à demanda por algumas diretrizes éticas no tratamento de dados de forma geral, e não apenas dos dados pessoais. De qualquer forma, uma vez ausente a exigibilidade legal, designamos o princípio para as etapas que ocorrem durante o processamento da anonimização.

Por fim, destacamos a especificidade do princípio da responsabilização e prestação de contas quando aplicada à anonimização. Esse princípio se refere à "demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento

88 Conforme art. 5(1)(c) do GDPR.
89 Conforme art. 6º, III da LGPD.

das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas". Trata-se, portanto, de um princípio de *accountability*, que atribui o ônus aos agentes de tratamento para indicarem e comprovarem a adequação dos tratamentos aos princípios estabelecidos. Este princípio tem equivalência no GDPR com o princípio da responsabilidade, previsto no art. 5º (2).

No caso da anonimização, o *accountability* se destaca tanto como um requisito não-funcional, posto que é diretriz principiológica para o tratamento, tanto como um requisito funcional. Isso porque a anonimização requer ação do controlador além do momento do tratamento, ou seja, depois de os dados estarem anonimizados e com a finalidade de mantê-los nesse "*status* legal" da anonimização. O cumprimento deste requisito funcional deve centrar-se tanto na demonstração da eficácia da anonimização, como na sua manutenção ao longo do tempo, à medida que a técnica evolui. Trataremos melhor desse aspecto temporal da anonimização na subseção seguinte.

Por este princípio se desdobrar como requisito funcional e não funcional, ele foi bipartido tanto como diretriz principiológica, quanto como uma fase específica do *framework* da anonimização de dados, que se estende para além do processamento, conforme apresentado na Figura 11:

Figura 11: Diretrizes Principiológicas para a anonimização

Por fim, é importante ressaltar a exceção sobre incidência principiológica que é feita nos casos de anonimização ao final do tratamento (hipótese 13 dos requisitos preliminares). Nesses casos, a própria lei estabelece restrições de uso ao dado mesmo após anonimizados, já

que os dados só podem ser utilizados e acessados pelo próprio controlador. Portanto, para essa hipótese de requisito preliminar de tratamento, a lei especifica a permanência do princípio da segurança, mesmo que o dado esteja anonimizado, já que só quem pode ter acesso ao dado é o controlador que o anonimizou, conforme Figura 11 a seguir:

Figura 12: Diretrizes Principiológicas para a anonimização na hipótese de anonimização ao fim do tratamento

4.1.3.2. Requisitos específicos da anonimização como tratamento de dados

Os requisitos específicos da anonimização são aqueles requisitos funcionais que, de acordo com a lei, são essenciais para que um dado seja considerado anonimizado.

A LGPD define dado anônimo como aquele que não pode ser relacionado ao seu titular (art. 5º, III). No conceito de anonimização (art. 5º, XI) a lei deixa claro que se houver associação direta ou mesmo indireta entre os dados e seu titular não há anonimização. Por este motivo, os dados utilizados para formar o perfil comportamental de um determinado indivíduo, caso seja identificado, são dados pessoais (e não dados anônimos) (artigo 12, § 2º).

No entanto, seguindo o conceito europeu de anonimização, a LGPD reconhece que a total ausência de relação entre o titular e os seus dados é difícil ou impossível de se obter através da anonimização. Vários autores já abordaram as dificuldades de obtenção de anonimato perfeito, ou mesmo os riscos de reidentificar dados anonimizados (OHM, 2010) (BRASHER, 2018) (DOMINGO-FERRER, 2019) (PIRAS et al., 2019) (RYAN; BRINKLEY, 2017) (GJERMUNDRØD et al.,2016) (CARVALHO et al., 2020).

Portanto, a lei estipula que a relação entre os dados e o titular seja quebrada, levando-se em consideração alguns parâmetros estabelecidos.

Nesse sentido, a lei considera a tecnologia disponível no momento do processamento e os meios de que dispõe o controlador (art. 12, § 1º). Portanto, podemos observar que a lei inseriu parâmetros objetivos e um parâmetro subjetivo para analisar o anonimato, qual seja, os meios disponíveis ao próprio controlador. Isso se justifica porque diferentes atores podem ter tecnologias muito diferentes para realizar o processo de anonimização. Quanto mais sofisticados os meios disponíveis ao controlador para o processamento, a anonimização exigida também será mais sofisticada.

A lei equilibra esse aspecto subjetivo com os aspectos objetivos sobre o que é uma dissociação razoável (art. 12, §1º). Se o processo de conexão entre os dados e o titular for muito caro, demorado e/ou trabalhoso, por exemplo, os dados são considerados anônimos. Neste caso, o parâmetro não é o meio próprio do controlador, mas a tecnologia disponível no momento do processamento. Portanto, o critério vai além da análise da realidade do controlador. A lei enumera dois critérios objetivos: tempo e custo de processamento, mas explica que não são critérios exaustivos.

Assim, para serem considerados anônimos, os dados não podem ser associados, direta ou indiretamente, ao seu titular, considerando os próprios meios do controlador (critério subjetivo) e esforços razoáveis (critérios objetivos), conforme mostrado na Figura 13:

Figura 13: Requisitos para considerar um dado como anonimizado

Lembrando que esses requisitos funcionais da técnica de anonimização devem ser constantemente reanalisados, de acordo com o requisito funcional de *accountability*, já que, com a evolução das tecnologias disponíveis, novos parâmetros objetivos são estabelecidos para a anonimização.

Além disso, caso exista uma melhoria das condições de processamento do próprio controlador, a técnica também precisará ser revista.

Destaque-se, portanto, que a anonimização, de acordo com os próprios critérios legais, é um tratamento que promove um aspecto situacional de anonimização ao dado, ou seja, o dado "está anonimizado", em um certo contexto tecnológico e operacional do controlador. Não se trata, portanto, de uma característica imutável do dado, e por isso, ressaltamos o caráter de "*status*" da anonimização.

4.1.4. *Framework* completo dos requisitos legais para anonimização

Neste item apresentamos o panorama completo das fases já relatadas sobre o processo de anonimização, conforme mostrado na Figura 14:

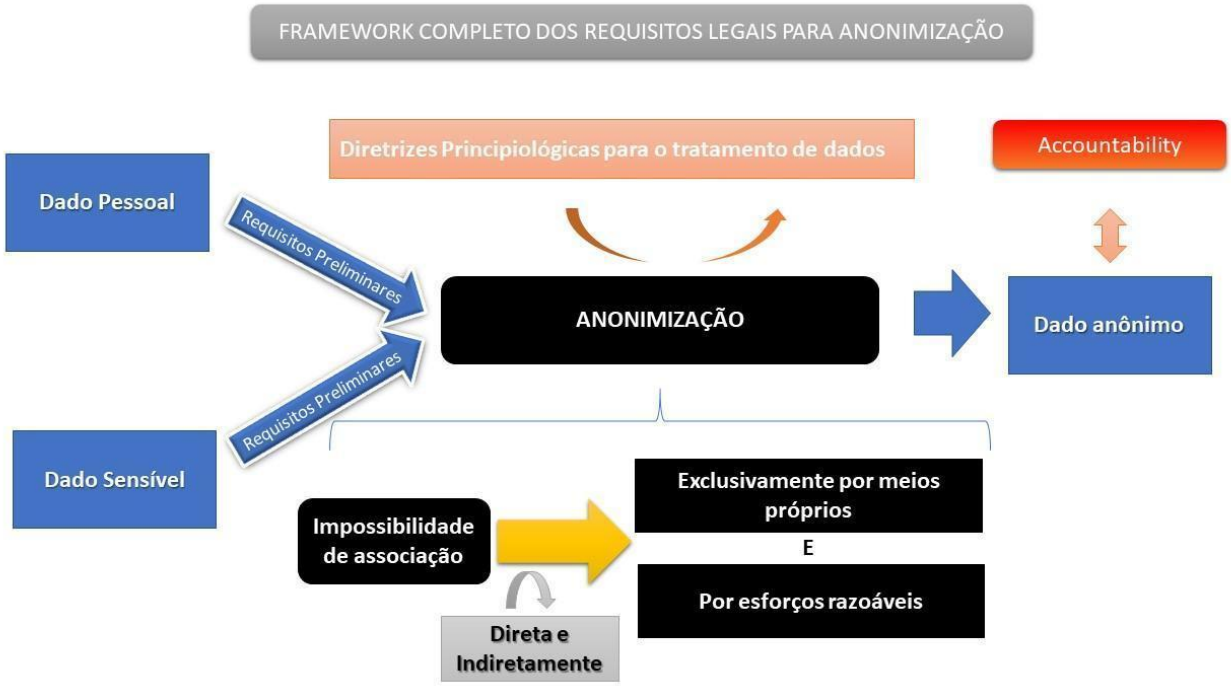

Figura 14: *Framework* completo dos requisitos legais para anonimização

Vemos, portanto, que a anonimização é precedida pelo cumprimento dos requisitos preliminares, estabelecidos legalmente de acordo com a característica do dado, ou seja, se ele é

um dado pessoal *stricto sensu*, ou se é um dado pessoal sensível (conforme figuras 5 e 6, respectivamente).

Cumprida uma ou mais hipóteses legais de tratamento, pode-se dar início à anonimização. Dentro do processamento da anonimização, devem ser respeitados os requisitos não funcionais e os requisitos funcionais. Os requisitos não-funcionais se consubstanciam no cumprimento da base principiológica necessária ao tratamento dos dados pessoais. Já os requisitos funcionais são especificados com as características que fazem com que o dado seja considerado anonimizado e permaneça nesse *status*, ao longo da evolução técnica e das alterações dos meios de tratamento do próprio controlador.

Desta forma, respondemos a primeira parte do problema de pesquisa sobre que tipo de *framework* pode ser apresentado acerca dos requisitos legais da anonimização na LGPD. É importante destacar que o *framework* foi elaborado com base em critérios exclusivamente legais, sem a intenção de esgotar todas as fases do processo de anonimização de dados pessoais.

Portanto, é possível que seja necessário elencar outros requisitos práticos numa elicitação que abarque todo o processamento da anonimização, ou seja, outros requisitos específicos de um projeto de solução de *software*. Por exemplo, um desenvolvedor deveria, em diálogo com outros *stakeholders*, determinar que tipo de técnicas de anonimização seriam mais interessantes para uma determinada base de dados, elaborando uma "instância" deste *framework* de desenvolvimento específica para o projeto em questão.

Entretanto, como dispomos anteriormente, a intenção deste *framework* foi levantar apenas os requisitos apontados pela Lei de Proteção de Dados Pessoais (LGPD). Portanto, trata-se de um *framework* limitado aos requisitos legais.

Um *framework* integral, na prática, deve incorporar os requisitos legais mencionados e complementá-los de acordo com as escolhas práticas de processamento dos diversos atores.

4.1.5. Papel da Autoridade Nacional

A Lei Geral de Proteção de Dados Pessoais deixou em aberto a possibilidade de regulação mais específica do *framework* apresentado. Isso porque algumas matérias da nossa LGPD são delegadas à Autoridade Nacional de Proteção de Dados (ANPD), que é o órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento da Lei em

todo o território nacional[90]. De forma geral, a ANPD poderá implementar novos parâmetros e mesmo dispor acerca dos casos omissos na LGPD, mas necessários à proteção de dados pessoais.

As atribuições da ANPD estão descritas no art. 55-J. Entre elas destacamos o seu papel regulatório, disposto por exemplo nos incisos III e XIII, que preveem, respectivamente, a elaboração de diretrizes para a Política Nacional de Proteção de Dados Pessoais e da Privacidade; e a edição de regulamentos e procedimentos sobre proteção de dados pessoais e privacidade, bem como sobre relatórios de impacto à proteção de dados pessoais para os casos em que o tratamento representar alto risco à garantia dos princípios gerais de proteção de dados pessoais.

Ressalte-se também que o inciso VIII aponta como papel da ANPD o estímulo à adoção de padrões para serviços e produtos que facilitem o exercício de controle dos titulares sobre seus dados pessoais, o que pode, por exemplo, gerar diretrizes de governança para as bases de dados de forma geral, inclusive para as bases anonimizadas.

O papel de fiscalização e auditorias da ANPD pode ser destacado nos incisos IV e XVI, que prescrevem a fiscalização e aplicação de sanções em caso de tratamento de dados realizado em descumprimento à legislação; e a realização de auditorias, ou determinação de sua realização, no âmbito da atividade de fiscalização.

Além disso, a Autoridade Nacional deve estar aberta a ouvir os agentes de tratamento e a sociedade em matérias de interesse relevante, conforme inciso XIV, e deve promover critérios diferenciados para que microempresas, empresas de pequeno porte e iniciativas empresariais de caráter incremental ou disruptivo adequem-se à lei, conforme inciso XVIII.

Portanto, vemos a preocupação da Lei em observar elementos subjetivos do regulado, para além da capacidade técnica do controlador, disposto como requisito funcional da anonimização. Nesse sentido, é desejável que a ANPD estabeleça regras diferenciadas na análise dos critérios de *compliance* à LGPD na anonimização para pequenas, médias e grandes empresas, a fim de evitar estrangulamentos no surgimento de novos agentes e o desincentivo à inovação.

A ANPD também deve traçar diretrizes acerca da interpretação da LGPD, em situações que ela seja ambígua ou na integração de casos omissos, conforme inciso XX.

90 Conforme art. 5º, XIX.

Portanto, é possível que a ANPD defina critérios mais específicos sobre requisitos preliminares, como por exemplo, em quais casos se aplicaria o legítimo interesse.

Por fim, destacamos que a ANPD pode solicitar aos responsáveis pelo tratamento relatórios de impacto à proteção de dados pessoais,[91] os quais devem conter, "no mínimo, a descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados".[92]

No caso específico da anonimização, o art. 12 § 3º prevê que 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 ANPD também contará com o auxílio do Conselho Nacional de Proteção de Dados Pessoais e da Privacidade, para a elaboração da Política Nacional de Proteção de Dados Pessoais e da Privacidade, e para sugestões sobre a atuação e novas ações da ANPD, conforme as atribuições dispostas no art. 58-B da LGPD.

Por fim, apontamos que a autoridade nacional deve ser informada sobre vazamento de dados juntamente com os titulares potencialmente atingidos, conforme descreve o art. 48 da LGPD.

Isso é particularmente importante quando lidamos com anonimização porque, como ressaltado neste trabalho, os dados anonimizados recebem esse *status* enquanto cumprirem os padrões legais de *compliance*. Caso a anonimização seja fragilizada, ou pela evolução da técnica ou por aprimoramento dos meios do próprio controlador, esses dados voltam a ser considerados como pessoais e podem estar sujeitos às previsões de vazamento de dados.

Nesses casos de difusão indevida de dados pessoais, a Autoridade Nacional é que verificará a "gravidade do incidente e poderá, caso necessário para a salvaguarda dos direitos dos titulares, determinar ao controlador a adoção de providências", além de avaliar se o controlador adotou medidas técnicas adequadas para tornar os dados pessoais vazados ininteligíveis.[93]

Portanto, conforme os dispositivos legais ressaltados, é possível que em breve a Autoridade Nacional complemente os requisitos elencados no *framework*.

---

91 Conforme art.4º, §3º.
92 Conforme art. 38 parágrafo único.
93 Conforme art. 48 §§2º, 3º.

Entretanto esclarecemos que, no item seguinte, a comparação entre os limites identificados na técnica de anonimização e os requisitos legais dispostos no *framework* só abarcam os requisitos já dispostos na legislação. Dessa forma, não se exclui a possibilidade de que a ANPD possa aparar as lacunas que porventura sejam identificadas. Na verdade, é até mesmo desejável que os limites que, por hipótese, não tenham sido considerados pela Legislação, sejam expostos até para abertura do debate sobre que tipo de complementação regulamentar é necessária e como a ANPD poderia auxiliar no reparo dos riscos ainda persistentes.

4.2. Análise crítica dos requisitos legais demonstrados no *framework* da LGPD à luz dos Limites da Anonimização: A LGPD leva em consideração esses limites?

Como uma segunda parte do problema de pesquisa, propomos responder se os requisitos definidos pela Lei Geral de Proteção de Dados Brasileira para anonimização, elencados no *framework*, contemplam os principais limites da técnica que exploramos no capítulo 3 deste trabalho.

Para isso, faremos uma breve retrospectiva dos pontos levantados, comparando-os aos elencados no *framework.*

Ressaltamos que os requisitos definidos na lei não são exaustivos e podem, como relatado nas linhas acima, ser complementados pela autoridade nacional e inclusive por outras fontes de regulamentação, como normas internacionais, etc.

Mas a identificação das lacunas que porventura existam é um passo importante no aprimoramento do tratamento enquanto técnica e no uso dos dados anonimizados de forma a se assegurar a permanência e a confiabilidade da anonimização.

4.2.1. Limites Intrínsecos:

4.2.1.1. Informação Teoricamente Segura X Segurança Perfeita

Dentro dos limites intrínsecos da anonimização, o primeiro ponto a se ressaltar é o de que a LGPD discerne a anonimização como uma técnica distante da chamada “segurança perfeita”. De fato, a LGPD define a anonimização como uma técnica situacional, ou seja, que promove um *status* de anonimizado ao dado, enquanto persistirem as características legais de dissociação. Não se trata, portanto, de uma técnica definitiva, no sentido da completa irreversibilidade.

Isso foi demonstrado na figura 13, quando afirmamos que a lei define como dado anonimizado aquele que, por esforços razoáveis e levando-se em consideração o aspecto subjetivo dos meios disponíveis ao controlador, não puder ser associado direta ou indiretamente ao seu titular.

Por esse motivo, em atenção às orientações legais, definimos neste trabalho dado anônimo como o dado pessoal que em um determinado momento e em um determinado contexto pode ser considerado dissociado de seu titular, deixando sua "pessoalidade" enquanto seguir as regras legais que lhe conferem o mencionado *status* de anonimizado.

Assim como a regulação europeia, a LGPD prevê que o dado só será anônimo se não puder ser reidentificado. Se uma base de dados supostamente anonimizada puder ser revertida, o tratamento deve se ajustar aos requisitos dispostos para dados pessoais. Além disso, pelo conceito de anonimização, a lei resguarda sua aplicação imediata quando a técnica de anonimização for superada pelo aprimoramento de *softwares* maliciosos, ou por novas possibilidades de inferência.

Por causa dessas características da Lei, Ohm tece a crítica de que o conceito de anonimização funciona sem que existam limites definidos, como um gás que se ajusta ao seu recipiente. (OHM, 2010, p. 1741). O autor afirma que isso seria prejudicial, uma vez que, em uma visão mais rigorosa, quase todo dado anonimizado poderia ser, dessa forma, submetido a lei de proteção de dados, inutilizando, por fim, a ferramenta. Além disso, o autor aponta que esse tipo de efeito expansivo da lei faz com que, na prática, exista pouca segurança jurídica na anonimização.

No caso da LGPD, a lei não define parâmetros fechados para a caracterização desse tratamento. No que se refere aos critérios objetivos, a lei chega a apontar que os dois elencados são meramente exemplificativos, a saber, o tempo e os custos do processamento (conforme Figura 13). E mesmo para esses dois critérios objetivos estabelecidos, nenhuma medida foi definida sobre o que é um custo razoável ou um tempo de processamento razoável.

Portanto, a crítica de Ohm é pertinente também ao nosso sistema legislativo, mas não incontornável. Nesse sentido, entendemos que o conceito amplo de anonimização pode ser melhor delimitado pela Autoridade Nacional, fixando padrões mais objetivos (e que levem em consideração a realidade das organizações) sobre os requisitos razoáveis para se considerar um dado como anonimizado.

O Artigo 12, §3, indica, como explicitamos alhures, que a Autoridade Nacional "poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar

verificações acerca de sua segurança". Isso mostra que a própria lei prevê um papel para a Autoridade Nacional na especificação desses parâmetros.

Dessa forma, concluímos que, apesar de a LGPD levar em consideração que a segurança perfeita é praticamente inatingível no que se refere ao tratamento da anonimização, ela ainda deixa lacunas sobre os parâmetros necessários para se alcançar a classificação do dado como "informação teoricamente segura". Apesar da lacuna, a própria lei prevê a complementação de requisitos pela Autoridade Nacional, que, quando cumprida, virá ao encontro da necessidade de maior segurança jurídica a estes parâmetros.

4.2.1.2. *Trade-off* Utilidade e anonimização.

Quanto ao limite intrínseco referente ao *trade-off* de utilidade na anonimização, apontamos que a lei, apesar de não se debruçar especificamente sobre essa característica, acolhe-a quando deixa de considerar os dados anonimizados como dados pessoais (conforme especificado na Figura 5). A lei ressalta que a anonimização implica em perdas, quando a define justamente como a perda da possibilidade de associação[94] direta ou indireta (conforme Figura 13). A consequência dessa perda é a descaracterização do dado como um dado pessoal.

De qualquer forma, entendemos que este limite intrínseco da técnica é uma preocupação que atinge muito mais os agentes de tratamento, já que a anonimização implica em perda de utilidade e consequentemente perda de valor dos seus ativos.

E, justamente por isso, a decisão sobre a anonimização é uma escolha gerencial, que deve ter por premissa o fato de que aqueles dados certamente perderão valor econômico. Nesse sentido, apesar de, como mencionamos anteriormente, o Princípio da Necessidade ser mitigado no processo da anonimização, ele é equilibrado pela existência desse *trade off*.

O Princípio da Necessidade limita o tratamento de dados pessoais ao "mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados"[95]. Na anonimização, o princípio da necessidade é mitigado já que é desejável que os dados estejam mais protegidos e a lei considera a anonimização um mecanismo de proteção de dados pessoais. Assim, em tese, a anonimização poderia ser utilizada de forma irrestrita e generalizada para fins de *compliance* à Lei e Proteção de Dados.

94 Conforme art. 5º, XI da LGPD.
95 Conforme art. 6º, III, da LGPD.

No entanto, justamente porque os dados anonimizados perdem parcialmente seu potencial de fornecer informação, diminuindo sua utilidade e seu valor para esses mercados, a anonimização é realizada de forma mais parcimoniosa.

A governança de dados nesse caso também é uma excelente aliada para direcionar a decisão sobre quais dados podem ser anonimizados numa estrutura institucional sem grandes perdas para a estratégia de negócio.

Portanto, concluímos que o *trade-off* da utilidade foi contemplado, ainda que de forma indireta pela LGPD. Ressaltamos, contudo, que este limite é, mais do que uma preocupação legislativa, um desafio para a gestão institucional, que deve ponderar a escolha sobre quais dados podem sofrer essa perda de utilidade e valor, sem inviabilizar estratégias de negócio.

4.2.1.3. A integridade dos dados e as técnicas de anonimização

Se por um lado, o elemento da perda é tomado como característica da técnica de anonimização pela LGPD, não há na lei uma maior preocupação com as consequências dessa supressão de informação íntegra e de seu distanciamento da realidade. Tratamos sobre essa característica quando falamos da perda da integridade dos dados na anonimização, como limite intrínseco à técnica.

Há uma relação direta entre a perda de utilidade do dado e a perda de sua integridade já que, como afirmamos anteriormente, a utilidade e, consequentemente, o valor de mercado dos dados encontram-se justamente na precisão das predições e inferências passíveis de se extrair deles. Quanto mais próximos da realidade maior a potencialidade de correção e eficácia das decisões tomadas a partir da análise dos dados.

Assim, a LGPD, conforme abordamos no tópico anterior, reconhece o fenômeno da perda que envolve a anonimização, mas não se debruça sobre os efeitos dessas perdas sobre os dados anonimizados e os tratamentos dele decorrentes. O que a lei não considera é que a dissociação entre a realidade e o dado anonimizado pode causar consequências nefastas e de difícil retificação. Principalmente no que tange a dados públicos, como apontamos no item de referência, a perda da integridade do dado pode levar a conclusões equivocadas e afetar toda uma cadeia decisória. É o que acontece, por exemplo, na adição de ruído aleatoriamente num banco de dados. O resultado é que o agente de tratamento não saberá quais informações são verídicas e, consequentemente, poderá tomar decisões precipitadas, baseadas em informações ruidosas.

Além disso, a transparência na publicização de dados quando anonimizados é sempre, de alguma forma, minimizada ou comprometida. Dessa forma, a publicização de dados de forma anonimizada acaba por implicar em perdas para a transparência pública a que se objetiva.

Se a finalidade de publicização do dado pessoal público é justamente a transparência das informações, a anonimização poderia ser vista como técnica incompatível com a manutenção do interesse público que justifica a própria disponibilização, conforme exposto na Figura 9 do *framework*. Dessa forma, a escolha dos diversos agentes em anonimizar dados deve balancear as vantagens e desvantagens em se ter uma base menos confiável para o acesso à informação e para nortear os processos decisórios.

No caso de bases de dados públicas, acreditamos ser necessárias medidas para além dessa prévia reflexão sobre conveniência da anonimização. Especialmente caso os dados sejam publicizados, é necessário o esclarecimento acerca da aplicação da anonimização, inclusive, com a especificação dos métodos utilizados, até para viabilizar a finalidade pública de acesso à informação e de transparência na gestão pública.

Além disso, a privação de transparência dos dados públicos também deve ser sopesada, antes da escolha pela anonimização. Nesses casos, se a exigência legal for de plena transparência de determinados dados, a anonimização não pode ser considerada um método recomendável.

Portanto, quanto ao dilema da integridade dos dados, concluímos que a legislação reconhece que os dados anonimizados perdem, seja a utilidade, seja a integridade, ao salientar que esses dados se distanciam dos dados pessoais. No entanto, a LGPD não dedica atenção às consequências dessas perdas, em especial no que concerne a veracidade das informações que podem ser extraídas dos dados anonimizados e as consequências disso, principalmente para a transparência e confiabilidade de dados públicos.

Podemos então concluir, quanto aos limites intrínsecos, que, apesar de a lei levar em consideração alguns desses limites, ela não se debruça nas questões decorrentes deles. Há lacunas importantes sobre critérios mais sólidos para a conceituação da anonimização e sobre orientações mais práticas quanto a como lidar com a perda de utilidade e integridade dos dados, ainda que reconhecidas pela lei.

Esses fatores deverão ser melhor especificados pela Autoridade Nacional e devem ser levados em consideração tanto pelo controlador, dentro da governança de seus dados, quanto pelos agentes públicos, no que tange às perdas de utilidade e integridade derivadas da técnica.

4.2.2. Limites Extrínsecos

4.2.2.1. Linkabilidade e poder de inferência de bases de dados externas

Por sua vez, quanto aos limites extrínsecos à técnica, destacamos, a exemplo dos aspectos conceituais da anonimização, que a LGPD deixa em aberto os critérios de aferição de linkabilidade e poder de inferência de bases externas e sua repercussão em considerar um dado como anonimizado.

Levantamos três questões lacunosas quanto a matéria: 1) a diminuição de entropia e os parâmetros auditáveis de anonimização; 2) a consideração de bases de dados externas no *accountability*; e 3) a publicização de dados anonimizados.

Quanto à primeira questão, acerca da **diminuição de entropia e os parâmetros auditáveis de anonimização**, destacamos que, aparentemente, a LGPD se preocupa apenas com o fator de distinção para apontar se um dado é anonimizado, utilizando a classificação de Doneda e Machado (DONEDA & MACHADO, 2018). Assim, um dado seria anônimo se não fosse possível isolar algum ou algumas informações que destacam uma pessoa numa base de dados, tomando-se em consideração essa mesma base.

Por sua vez, a possibilidade de ligação e a inferência, ainda que presentes, não necessariamente interferem na caracterização ou qualificação do dado anonimizado. Podemos citar um exemplo. Segundo Doneda e Machado, a capacidade de ligação é a possibilidade de se conectar dois ou mais registros referentes a um mesmo indivíduo ou um mesmo grupo de pessoas. Pela LGPD, se esse indivíduo não pudesse ser identificado, os dados, ainda que ligados, não perderiam seu caráter de anonimizados. Da mesma forma, a inferência seria viável, desde que não expusesse o titular dos dados. É o que se observa dos requisitos específicos da anonimização, que delimitamos na Figura 13 do *framework*.

No entanto, como vimos, tanto a possibilidade de ligação quanto a inferência são fatores relevantes para a diminuição de entropia, fragilizando a técnica de anonimização. Dessa forma, é razoável que a linkabilidade e a inferência possam se reverter em parâmetros auditáveis para classificação da anonimização e seu *compliance*, ainda que a reidentificação imediata não seja possível.

Interessante ressaltar que algumas técnicas da anonimização já utilizam essas características de linkabilidade e inferência como parâmetro para a própria metodologia, como acontece no chamado K-anonimato. Nessa técnica, a impossibilidade de identificação é definida quando se restringe os dados prováveis de um titular a uma quantidade delimitada de perfis.

Assim, há a padronização sobre a linkabilidade e a inferência consideradas aceitáveis para a manutenção da técnica de anonimização.

Além disso, podemos ressaltar iniciativas que tem justamente se utilizados da linkabilidade e do poder de inferência para mensurar a confiabilidade da anonimização, sendo inclusive uma das estratégias apresentadas pela ISO 20889 e pela ISO 19944. Os professores Domingo-Ferrer, Muralidhar e Bras-Amor´os também apontam a possibilidade de utilização desses fatores como métrica no artigo "*General Confidentiality and Utility Metrics for Privacy-Preserving Data Publishing Based on the Permutation Model*" (DOMINGO-FERRER, et al., 2020).

Nesse sentido, entendemos que esses parâmetros de diminuição de entropia podem ser utilizados para aferição da técnica. Não apenas na delimitação sobre a reidentificação, mas também para classificação de níveis de confiança a serem depositadas na anonimização e nos controladores.

A segunda lacuna que observamos na lei relaciona-se à **consideração de bases de dados externas para fins de *accountability*** da técnica de anonimização.

Por causa do fenômeno descrito por Ohm como "problema do acréscimo", bases de dados externas interferem na possibilidade de reidentificação de dados anonimizados, fazendo com que, quanto mais dados estejam disponíveis, maior seja a rastreabilidade e mapeamento entre pessoas e dados. Por esse motivo, a técnica de anonimização deve levar em consideração as bases públicas disponíveis no momento do tratamento, ainda que essas bases não se refiram a PII ou a *quasi-identifiers*.

A LGPD não considera de forma explícita essas bases nos parâmetros objetivos sobre o que é um dado anonimizado, conforme Figura 13 do *framework*. Também nos parâmetros subjetivos, o englobamento dessas bases não é claro, já que a lei prescreve a razoabilidade da anonimização segundo os meios próprios de que dispõe o controlador. Entendemos, no entanto, que seria possível definir o que seriam os "meios próprios" do controlador seguindo parâmetros semelhantes ao que o direito civil utiliza para o "homem médio", ou seja, considerando não apenas os meios de que dispõe efetivamente o controlador, mas também os meios a que tem acesso, mesmo que não se utilize de fato.

Esse tipo de ajuste interpretativo ou de métrica poderá ser feito pela Autoridade Nacional ao regular de forma mais minuciosa quais os parâmetros objetivos e subjetivos para se considerar o dado como anonimizado, como discorremos na Figura 13 do *framework*.

Desta forma, os parâmetros de razoabilidade da anonimização como "informação teoricamente segura" devem averiguar a possibilidade de reidentificação dos dados levando-se

em consideração não apenas a base a ser anonimizada, mas a interação entre essa base e outras facilmente disponíveis ao controlador, na medida da acessibilidade a esses dados.

Por fim, a terceira lacuna legal que identificamos acerca do poder de inferência é a ausência de parâmetros para a **publicização de dados anonimizados**.

A LGPD não estabeleceu parâmetros para a disponibilização de dados anonimizados, o que faz com que o modelo liberação-esquecimento pareça um método compatível com a legislação como veremos mais à frente, no próximo tópico. Esse fator de livre fluxo e disponibilização de dados anônimos interfere diretamente no aumento do poder de inferência, na linkabilidade e na diminuição de entropia das bases de dados de forma geral. Sabemos que quanto mais dados estão acessíveis livremente, maiores os riscos para as técnicas de anonimização, que precisam se ajustar ao novo contexto, alterando os dados que podem ser reidentificados a partir das novas ligações e inferências possíveis.

Nesse sentido, a ausência de diretrizes acerca da disponibilização pública de dados e, como vimos no tópico anterior, acerca da relação entre esses novos dados acessíveis e as bases anonimizadas, confere menor segurança à técnica de anonimização. O limite da anonimização consubstanciado no problema do acréscimo deve levar os gestores e legisladores a uma maior reflexão sobre a disponibilização dos dados e a forma de sua utilização.

Podemos então concluir norteando as três questões lacunosas que devem ser direcionadas tanto à autoridade nacional quanto ao debate público acerca do poder de inferência das bases. São elas: 1) a linkabilidade e a inferência como parâmetros auditáveis em bases anonimizadas, ainda que a reidentificação imediata não seja possível; 2) a necessidade de que os parâmetros do que é considerado um dado anonimizado leve em consideração não apenas a base em tratamento, mas também bases de dados públicas ou facilmente acessíveis; 3) a maior ponderação na disponibilização de dados anonimizados, já que essa disponibilização aumenta o poder de inferência, interferindo no que pode ser considerado um dado anonimizado.

4.2.2.2. Accountability da técnica- fuga do modelo liberação-esquecimento de anonimização

Com os conhecimentos que já possuímos acerca da anonimização, é necessário abrir mão da falácia de que a anonimização proporciona o melhor dos dois mundos: os benefícios do fluxo de informações e fortes garantias de privacidade.

Dessa forma, é preciso pensar cuidadosamente na transferência de informações tendo em vista a impossibilidade de se ter o ideal de utilidade e privacidade. Essa

impossibilidade deve nos fazer refletir acerca de princípios e estatutos antes tidos como perfeitamente equilibrados pela anonimização. É o que afirma Ohm quando discorre que a anonimização permitia aos legisladores ignorar uma necessária mediação entre valores como segurança, inovação e livre fluxo de informações. Isso porque a ideia prevalecente era a de que o risco à privacidade era mínimo ou nulo quando se trata de compartilhamento de dados anônimos (OHM, 2010, p. 1736).

Mas, uma vez ciente dos riscos, esse balanceamento passa a ser imprescindível. Ainda mais porque a confiança persistente na anonimização robusta pode pender a balança para a manutenção de livres fluxos de informação, em detrimento da privacidade. É necessário um reequilíbrio entre esses fatores e ainda outros, como o acesso à informação e a inovação (OHM, 2010, p. 1740).

Justamente devido aos riscos da disponibilização irresponsável dos dados, **o modelo liberação-esquecimento** se torna incompatível com a segurança e a privacidade na anonimização. Os problemas que envolvem o modelo liberação-esquecimento podem ser analisados no contexto da nossa legislação sob dois enfoques: **1) quanto ao livre fluxo e publicização de dados; 2) quanto à responsabilização por dados anônimos publicizados.**

Quanto ao primeiro aspecto, ou seja, relativo ao **livre fluxo e publicização de dados**, ressaltamos que o modelo liberação-esquecimento tem encontrado respaldo no ordenamento jurídico nacional antes mesmo da questão da anonimização de dados.

Como relatamos no segundo capítulo deste trabalho, a segunda onda legislativa brasileira se caracterizou pelo fortalecimento do princípio da transparência e da auditabilidade do poder público, através da disponibilização de grande quantia de dados. Ohm ressalta que, se o poder público decide publicizar seus dados, a preocupação com segurança deve ser redobrada (OHM, 2010, p. 1729). Entretanto, no contexto brasileiro, de forma geral, o fomento ao livre fluxo de dados é o que tem prevalecido, sem que exista enfoque nos parâmetros de segurança ou de como serão utilizados os dados após disponibilizados.

Por exemplo, além do Cadastro Base, o Decreto nº 10.046/2019 viabiliza a extensão das bases temáticas das entidades do governo federal através do compartilhamento de dados entre elas. O Decreto prevê três formas de categorização dos níveis de compartilhamento: 1) Amplo, quando o dado pode ser partilhado livremente; 2) Restrito, referente a dados sigilosos, que podem ser compartilhados com todos os órgãos para fins de políticas públicas; e 3) Específico, referente a dados sigilosos que podem ser compartilhados apenas com entidades

específicas[96]. Nessa linha, o decreto aponta as diretrizes para o compartilhamento de dados entre os órgãos e as entidades da administração pública federal direta, autárquica e fundacional e os demais Poderes da União, que estipula, em seu art. 3º, I:

> I - a informação do Estado será compartilhada da forma mais ampla possível, observadas as restrições legais, os requisitos de segurança da informação e comunicações e o disposto na Lei nº 13.709, de 14 de agosto de 2018 - Lei Geral de Proteção de Dados Pessoais;

O art. 11 traça ainda as diretrizes para o chamado compartilhamento amplo[97], que dispensa autorização prévia pelo gestor de dados e será realizado pelos canais existentes para dados abertos e para transparência ativa, na forma da legislação. Há, portanto, previsão de que esses dados de compartilhamento amplo sejam catalogados no Portal Brasileiro de Dados Abertos em formato aberto. (art. 11 § 5º).

Por um lado, o Decreto permite o compartilhamento de dados ainda que não haja a celebração de convênio, acordo de cooperação técnica ou instrumentos congêneres entre os órgãos e entidades públicas. Por outro, o Decreto prevê a conformidade com as restrições de uso e compartilhamento com os limites traçados pela Lei de Proteção de Dados, que, de forma geral, apontam padrões mais flexíveis para os dados de acesso público.

De fato, na LGPD os dispositivos se alinham à maior disponibilidade dos dados pessoais quando públicos. A lei traz um capítulo específico sobre esse tema, em observância à já mencionada Lei de Acesso à Informação (LAI). Nesses casos, a LGPD prevê que os dados devem ser mantidos em formato interoperável e estruturado para viabilizar o uso compartilhado para fins de políticas públicas, serviços públicos, descentralização da atividade pública e disseminação e acesso das informações.[98] O poder público pode compartilhar esses dados já estruturados com entidades privadas quando esses dados forem acessíveis publicamente[99]. Os dados também poderão ser compartilhados mediante contratos, convênios ou instrumentos congêneres, desde que comunicados à autoridade nacional. [100] Portanto, principalmente no que tange ao compartilhamento de dados públicos, a regra é a disponibilização dos dados, assegurada a veracidade e a interoperabilidade.

---

96 Conforme art. 4º do Decreto nº 10.046/2019.

97 Pontue-se, no entanto, que o Decreto prevê a incidência de mecanismos de governança, como a gestão, auditabilidade, confidencialidade desses dados, nas plataformas de interoperabilidade, definidos pelo Comitê Central de Governança de Dados.

98 Conforme Art. 25.

99 Conforme art. 26, §1º, III.

100 Conforme art. 26, §1º, V, e § 2º.

A LGPD, no entanto, estabelece como restrição ao tratamento dos dados pessoais públicos, a existência de compatibilidade entre o tratamento a ser realizado e o interesse público na disponibilização dos dados, como demonstramos na Figura 9 do *framework*. Neste ponto, a LGPD atua como balizadora dos efeitos da disponibilização, estabelecendo um limite finalístico ao modelo liberação-esquecimento: a exposição pública de dados não deve transgredir os objetivos iniciais de transparência e auditabilidade da disponibilização.

Por sua vez, no que se refere aos dados anonimizados, especificamente, ressaltamos que a LGPD não estabelece diretrizes para publicização desses dados nem pelos particulares nem pelo poder público. Ao que parece a lei permite que, caso os dados estejam devidamente anonimizados, eles possam ser livremente usados, compartilhados e disponibilizados. Inclusive, não existem óbices legais expressos à comercialização desses ativos. E, na visão dos mercados, essa é justamente uma das funções da anonimização, ou seja, permitir maior flexibilidade aos gestores, que podem extrair valor desses dados de formas diversas e em *compliance* com a lei.

Ressalvamos a exceção disposta no art. 16, IV, que se refere à anonimização ao fim do tratamento de dados. Nesse caso, a lei estabelece limites à disposição dos dados, já que prevê que o uso desses ativos será exclusivo do controlador. Além disso, a lei dispõe expressamente que é vedado o acesso destes dados por terceiros. Nós apontamos as nuances desse requisito preliminar ao tratamento de dados na Figura 12 do *framework.*

Quanto à anonimização realizada sob alguma outra das modalidades de requisitos preliminares, não existem óbices à livre disposição dos dados após a anonimização. Uma vez anonimizados, os dados não têm limitações legais ao seu livre fluxo. Dessa forma, a anonimização segue como "solução" das demandas por publicidade e privacidade. O modelo liberação-esquecimento se torna ainda mais forte nesse contexto, já que os dados pessoais públicos têm garantias principiológicas envolvidas em seu tratamento enquanto os dados anonimizados deixam de ser pessoais e, portanto, perdem esse arcabouço protetivo.

Essas ferramentas são utilizadas por vezes até mesmo pelo setor público, para proporcionar maior flexibilidade de gestão. Temos, por exemplo, a recentíssima Lei 14.129/2021, que dispõe sobre alguns princípios do governo digital, que busca a implantação do governo como plataforma, promovendo o uso de dados, preferencialmente anonimizados, para formulação de políticas públicas, pesquisas científicas, negócios e controle social.[101]

101 Conforme teor: "Art. 3º São princípios e diretrizes do Governo Digital e da eficiência pública:" (...) "XXIII - a implantação do governo como plataforma e a promoção do uso de dados, preferencialmente anonimizados, por pessoas físicas e jurídicas de diferentes setores da sociedade, resguardado o disposto nos arts. 7º e 11 da Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais), com vistas, especialmente, à formulação de políticas públicas, de pesquisas científicas, de geração de negócios e de controle social";

Essas questões apontam sobre a complexidade do tema da disponibilização de dados, principalmente quando se busca o equilíbrio de princípios tão importantes e sensíveis como a transparência na gestão pública, a privacidade dos cidadãos e a disponibilidade dos dados.

A solução, no entanto, não se distancia do necessário debate público e do estabelecimento de parâmetros para disponibilização de dados e para aferição de seu uso. A disponibilização pública e o livre fluxo de dados devem ser balizados pelos limites que são inerentes à técnica, como a perda de utilidade e integridade dos dados anonimizados e a fragilização derivada do poder de inferência, sob pena de violação de direitos.

Por sua vez, o segundo problema derivado do modelo liberação-esquecimento é a **responsabilização por dados publicizados**.

A LGPD não discorre explicitamente sobre a responsabilidade dos controladores acerca dos dados anonimizados, após liberados, compartilhados ou mesmo comercializados. A lei deixa em aberto se a responsabilidade na manutenção dos dados anonimizados se limita a quem os possui em suas bases no momento da anonimização, da auditoria, ou se haverá impactos quanto à responsabilidade na cadeia de compartilhamento desses dados. Isso porque, apesar de prever, como vimos na Figura 10, a necessidade de *Accountability* acerca da permanência da anonimização, a lei não deixa claro se esse tipo de responsabilidade de prestação de contas se refere apenas aos dados em posse do controlador no momento do tratamento. Essas lacunas quanto à responsabilidade contribuem para o aumento da insegurança que envolve a técnica da anonimização.

O art. 44 da LGPD prevê que será considerado irregular o tratamento que não fornecer a segurança esperada pelo titular, ou seja, que os riscos não estejam entre aqueles razoavelmente esperados tendo em consideração as técnicas disponíveis na época do tratamento[102]. Os riscos que circundam os dados anonimizados nem sempre são compreendidos pelo titular, principalmente se considerarmos o livre uso, compartilhamento e disponibilização dos dados anonimizados. Podemos imaginar um exemplo: caso um titular consinta na anonimização de seus dados por determinada prestadora de serviços, o uso destes dados não encontrará o óbice do art. 16, IV; não se restringindo ao uso pelo controlador.

Art. 4º Para os fins desta Lei, considera-se: (...) "VII - governo como plataforma: infraestrutura tecnológica que facilite o uso de dados de acesso público e promova a interação entre diversos agentes, de forma segura, eficiente e responsável, para estímulo à inovação, à exploração de atividade econômica e à prestação de serviços à população".

102 Conforme art. 44, caput, II, III.

Nesse caso, a livre disposição desses dados pode fazer com que um terceiro, não envolvido na relação entre o titular e o prestador de serviços, tenha acesso aos dados. Caso em posse desse terceiro os dados sejam reidentificados a violação por vezes não será nem mesmo conhecida pelo titular e, da mesma forma, dificilmente este risco teria sido considerado pelo titular no momento do consentimento.

Dessa forma, a responsabilização em cadeia diminuiria a lacuna entre os riscos razoavelmente esperados e a atribuição de responsabilidade aos agentes. Todas essas nuances afetam a forma como deverá ser realizada a responsabilização se porventura os dados deixarem seu *status* de anônimos. Também causa implicações diretas sobre a regulação e a auditabilidade dos controladores. Todavia, a lei foi omissa quanto à forma de prestação de contas, incorrendo em falta de métricas acerca da responsabilidade e dos agentes a serem responsabilizados.

Concluímos, portanto, que a LGPD não aboliu o modelo liberação-esquecimento. Isso porque a lei não estabelece limites ao livre fluxo e publicização de dados anônimos, com exceção do caso em que os dados são anonimizados ao término do tratamento; e também não estabelece diretrizes claras acerca da responsabilidade por dados publicizados. Apesar de prever como necessário o *accountability* na anonimização, a lei não deixa claro se o controlador se responsabiliza pela dissociação apenas dos dados em sua posse no momento do tratamento, ou se os dados compartilhados também seguem na cadeia de responsabilidade do controlador original.

Para fins de sugestão quanto ao tópico em destaque, ressaltamos a proposta que o autor Paul Ohm faz como alternativa ao modelo liberação-esquecimento.

O autor recomenda que este modelo seja abolido e sugere o monitoramento efetivo dos dados após a anonimização e compartilhamento. Isso faz com que a responsabilidade do gestor ultrapasse a própria base de dados e abarque a gestão dos dados compartilhados (OHM, 2010, p. 1755- 1756).

A proposta é a formação de trilhas de auditoria e controle de acesso. Assim, ainda que os controladores liberem seus dados, os mesmos deverão ser controlados por *softwares* que limitam o acesso e rastreiam o uso. O resultado é o registro de todas as movimentações dos dados, dispostos em trilha que será relatada ao administrador e poderá ser auditada, como um "cão de guarda de terceiros" (OHM, 2010, p. 1756).

Interessante ressaltar que os Sistemas de Gerenciamento de Banco de Dados (SGBD) — do inglês *Data Base Management System* (DBMS) podem ter extensões que atuam de forma semelhante à sugerida por Ohm. Esse tipo de sistema é conceituado como o conjunto de *softwares* responsáveis pelo gerenciamento de um banco de dados, e tem como

objetivo gerenciar o acesso, a persistência, a manipulação e a organização dos dados (CARVALHO, 2016).

Além disso, infere-se de Ohm a proposta de uma cadeia de responsabilidade que não se limita apenas ao detentor dos dados no momento da reidentificação, mas a todos os participantes da cadeia de transferência.

Isso é particularmente interessante quando consideramos que o titular, como mencionamos, em muitos desses casos, só teve acesso ao primeiro gestor que anonimizou os dados. Numa possível reidentificação na base de outro controlador, dentro da cadeia de compartilhamento, não haveria sequer ciência do titular quanto à disposição de seus dados, os quais, uma vez reidentificados, passam ao *status* de dado pessoal novamente.

Ressalte-se ainda que é necessário ajustes no controle de acordo com o tipo de agente, nos moldes do que a lei aparenta estabelecer. Isso para se evitar que técnicas tão robustas, como a do rastreio por meio de software exemplificado acima, seja exigível de todo e qualquer agente de tratamento, por vezes inviabilizando pequenos negócios que importam até mesmo em menores riscos à privacidade.

Nem todos os atores, seriam obrigados a utilizar práticas tão avançadas de auditoria e governança de dados. Mas de forma geral, para todos os controladores de dados, é primordial que exista transparência na utilização e na gestão desses ativos e responsabilidade na disponibilização dos dados, nos distanciando da irresponsabilidade do modelo liberação-esquecimento.

4.2.2.3. Parâmetros de Governança da base e anonimização

Ressaltamos neste trabalho que a governança é cada vez mais necessária para todos os agentes que lidam com dados, sejam eles pessoais ou não. A sua ausência, como um dos limites extrínsecos à ferramenta da anonimização, ocasiona a desconfiança frente aos gestores, a falta de transparência e maior dificuldade de auditoria acarretando, por fim, uma maior fragilidade da técnica. Por este motivo defendemos que, para que a lei de proteção de dados seja eficiente, é necessário que ela se alie a projetos de governança interna desses bancos de dados, proporcionando transparência na sua gestão.

A LGPD, reconhecendo a importância da governança na proteção de dados pessoais, dispõe expressamente, em seu art. 49, sobre a necessidade de que o tratamento seja estruturado e atenda aos requisitos de segurança, aos padrões de boas práticas de governança e aos princípios dispostos na lei. Portanto, para fomentar a segurança dos sistemas e melhorar a

gestão dos dados pessoais em prol da privacidade, a LGPD sugere que boas práticas e governança seriam instrumentos a serem adotados.

Nesse sentido, o art. 50 dispõe sobre a possibilidade de que controladores e operadores formulem suas próprias regras de boas práticas e de governança que estabeleçam condições de organização, regime de funcionamento, procedimentos, normas de segurança, padrões técnicos, mecanismos internos de supervisão e de mitigação de riscos, dentre outros aspectos.[103] A Autoridade Nacional terá ainda um papel importante no reconhecimento e divulgação de boas práticas e na adoção de padrões técnicos que "facilitem o controle pelos titulares dos seus dados pessoais", conforme art. 50 § 3º e art. 51.

Apesar de não ser imperativo pela lei um rol taxativo de regras de governança, ao final, como os operadores devem demonstrar as medidas realizadas para minimização de riscos, alguma medida dessas boas práticas se faz inevitavelmente necessária. Ressalte-se, portanto, que a lei permite que os padrões de governança sejam ajustáveis ao controlador e seu negócio, desde que atendidas as exigências de transparência e prestação de contas.

---

103 Conforme teor: "Art. 50. Os controladores e operadores, no âmbito de suas competências, pelo tratamento de dados pessoais, individualmente ou por meio de associações, poderão formular regras de boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, incluindo reclamações e petições de titulares, as normas de segurança, os padrões técnicos, as obrigações específicas para os diversos envolvidos no tratamento, as ações educativas, os mecanismos internos de supervisão e de mitigação de riscos e outros aspectos relacionados ao tratamento de dados pessoais.
§ 1º Ao estabelecer regras de boas práticas, o controlador e o operador levarão em consideração, em relação ao tratamento e aos dados, a natureza, o escopo, a finalidade e a probabilidade e a gravidade dos riscos e dos benefícios decorrentes de tratamento de dados do titular.
§ 2º Na aplicação dos princípios indicados nos incisos VII e VIII do caput do art. 6º desta Lei, o controlador, observados a estrutura, a escala e o volume de suas operações, bem como a sensibilidade dos dados tratados e a probabilidade e a gravidade dos danos para os titulares dos dados, poderá:
I - implementar programa de governança em privacidade que, no mínimo:
a) demonstre o comprometimento do controlador em adotar processos e políticas internas que assegurem o cumprimento, de forma abrangente, de normas e boas práticas relativas à proteção de dados pessoais;
b) seja aplicável a todo o conjunto de dados pessoais que estejam sob seu controle, independentemente do modo como se realizou sua coleta;
c) seja adaptado à estrutura, à escala e ao volume de suas operações, bem como à sensibilidade dos dados tratados;
d) estabeleça políticas e salvaguardas adequadas com base em processo de avaliação sistemática de impactos e riscos à privacidade;
e) tenha o objetivo de estabelecer relação de confiança com o titular, por meio de atuação transparente e que assegure mecanismos de participação do titular;
f) esteja integrado a sua estrutura geral de governança e estabeleça e aplique mecanismos de supervisão internos e externos;
g) conte com planos de resposta a incidentes e remediação; e
h) seja atualizado constantemente com base em informações obtidas a partir de monitoramento contínuo e avaliações periódicas;
II - demonstrar a efetividade de seu programa de governança em privacidade quando apropriado e, em especial, a pedido da autoridade nacional ou de outra entidade responsável por promover o cumprimento de boas práticas ou códigos de conduta, os quais, de forma independente, promovam o cumprimento desta Lei.
§ 3º As regras de boas práticas e de governança deverão ser publicadas e atualizadas periodicamente e poderão ser reconhecidas e divulgadas pela autoridade nacional."

No caso de bases de dados de entes públicos, são cada vez mais comuns regulamentos prevendo governança e boas práticas de gestão. É o caso do citado Decreto 10.046/2019, que prevê a incidência de mecanismos de governança, como a gestão, auditabilidade e confidencialidade desses dados nas plataformas de interoperabilidade, definidos pelo Comitê Central de Governança de Dados. Dessa forma, cada vez mais se reconhece a necessidade de gestão de dados, sejam eles pessoais ou não, submetendo-os a critérios básicos de organização gerencial, o que só é possível através da adoção de boas práticas e técnicas de governança.

Entretanto, para dados anonimizados, a LGPD não dispõe sobre esse tipo de exigência de padrões mínimos de governança. Dessa forma, temos que a lei não estabelece como obrigatória a adoção de boas práticas em bancos de dados que não contenham dados pessoais, dados sensíveis ou sigilosos. Nesse sentido, a Lei perde inclusive em potencial fiscalizatório, na medida em que bases desordenadas encontram maiores dificuldades de submissão a auditorias de forma geral. Essa omissão legal é, desta forma, inclusive contraditória, já que a anonimização é uma técnica que, pela sua definição legal, exige constante monitoramento.

Interessante ressaltar que a governança não é uma ferramenta impeditiva de inovação e de tratamento de dados. Ao contrário, as práticas de governança permitem uma gestão metodológica dos dados, da forma que melhor atenda aos critérios do negócio. Assim também tem sido as previsões legais quanto à governança, já que não há a exigência do cumprimento de um modelo específico, mas sim de transparência acerca dos modelos selecionados.

Além do mais, a governança é fundamental para se imaginar a aplicação da segurança da informação de forma multinível, unindo-se formas diversas de práticas de segurança. Essa heterogeneidade cria barreiras que impedem que falhas atinjam toda a cadeia de segurança da informação de uma instituição, gerando reflexos na confiabilidade, e diminuindo o risco de que ataques sejam efetivos (CARVALHO, 2016).

E, de fato, a privacidade deve ser pensada numa cadeia multinível de segurança, sem que se baseie toda a confiança em apenas uma técnica como a anonimização que já demonstrou não ser perfeita em diversos aspectos.

Portanto, concluímos que, apesar de extremamente importante, não só para subsistência da técnica, mas até mesmo para viabilizar o *accountability* previsto como necessário pela LGPD, não há previsão expressa nesta lei acerca da imprescindibilidade da governança para bases de dados anonimizadas. A omissão legal, novamente, poderá ser

contornada pela atuação da Autoridade Nacional, direta ou indiretamente, através do estabelecimento de diretrizes de governança ou de parâmetros de auditabilidade que só serão aferíveis a partir de uma boa gestão dos dados.

4.2.3. Limites Externos

4.2.3.1. A ressignificação da Privacidade e a Anonimização

Por fim, quanto ao que denominamos limites externos, que correspondem aos desafios e limites éticos na utilização dos dados anonimizados, o que percebemos é a continuidade de uma série de omissões legais e por vezes até mesmo o que parece uma lacuna de consciência e debate público acerca de algumas dessas questões.

Quanto à ressignificação da privacidade e a anonimização, ressaltamos que o paradigma da autodeterminação informativa, apesar de adotado pela LGPD, encontra exceções legais de aplicabilidade que desfavorecem o titular.

Podemos afirmar que a autodeterminação informativa, típica da terceira onda legislativa que compreende os dados pessoais como manifestações do direito da personalidade, foi recepcionada pela LGPD. É possível constatar essa afirmação a partir de diversos dispositivos, por exemplo, na previsão da necessidade de consentimento para o tratamento, como destacado nas Figuras 6 e 7 do *framework*. A necessidade de prestação de contas e esclarecimentos ao usuário, além da vinculação à finalidade previamente estabelecida e informada ao titular, descritas nas diretrizes principiológicas dispostas na Figura 8 do *framework*, também garantem um maior controle do usuário quanto aos seus dados. Além disso, a lei aponta, em alguns dispositivos, como nos arts. 1º e 2º, VII, a vinculação entre a proteção dos dados pessoais e a salvaguarda de direitos fundamentais, da privacidade e do livre desenvolvimento da personalidade do titular.

Entretanto, é importante ressaltar que a lei distancia os dados anonimizados dos parâmetros de autodeterminação informativa. Isso porque esses dados, posto que não mais considerados pessoais, podem ser livremente utilizados, fazendo com que os princípios e diretrizes necessários ao tratamento de dados pessoais sejam dispensados para dados anonimizados. Por exemplo, a lei permite a manutenção dos dados anonimizados de um usuário, ainda que ele tenha solicitado a portabilidade de seus dados para outro fornecedor, conforme art. 18, V e § 7º da LGPD, conforme teor:

> Art. 18. O titular dos dados pessoais tem direito a obter do controlador, em relação aos dados do titular por ele tratados, a qualquer momento e mediante requisição:
> V - portabilidade dos dados a outro fornecedor de serviço ou produto, mediante requisição expressa, de acordo com a regulamentação da autoridade nacional, observados os segredos comercial e industrial; (Redação dada pela Lei nº 13.853, de 2019) Vigência
> § 7º A portabilidade dos dados pessoais a que se refere o inciso V do caput deste artigo não inclui dados que já tenham sido anonimizados pelo controlador.

Dessa forma, percebe-se que uma vez anonimizado, pela lei, a portabilidade do dado fica obstada. Isso pode ser interpretado também como uma perda do direito sobre o dado anonimizado, posto que não mais vinculado ao titular. Essa questão ressalta que, também na nossa legislação, pairam dúvidas acerca da natureza jurídica dos dados anonimizados, que ainda seguem sem respostas.

Entretanto, a nosso ver, um dos dispositivos legais particularmente intrigantes acerca da autodeterminação informativa é aquele relacionado à possibilidade de anonimização ao final do tratamento, como uma das formas viáveis de cumprimento dos requisitos preliminares para o tratamento de dados pessoais, descrito na Figura 6 do *framework*. Essa hipótese de requisito preliminar se destaca porque incorre na flexibilização da autodeterminação de dados pessoais, ou seja, antes da anonimização. Entendemos que a lei foi incongruente ao permitir esse tipo de tratamento de forma direta, posto que, nesses casos, pela interpretação literal do dispositivo, o tratamento de um dado que é pessoal depende apenas da vontade do controlador, e nenhum outro elemento externo de legitimação.

A polêmica em torno do assunto se dá, por exemplo, se seria direito do titular impedir que, ao término de todo e qualquer tratamento, seus dados fossem anonimizados. Principalmente quando consideramos os riscos que envolvem a anonimização a pergunta se torna ainda mais pertinente. Nesse caso, em que medida os usuários são obrigados a arcar com os riscos da anonimização involuntária de seus dados ao final de todo e qualquer tratamento, com o respaldo legal que proporciona o art. 16, IV?

Dessa forma, o dispositivo legal corrobora com um risco não consentido e por vezes nem mesmo conhecido e mensurado pelos usuários, que dificilmente terão a possibilidade de fiscalizar o processo e a manutenção dessa anonimização.

Além disso, o uso restrito desses dados ao controlador, como menciona o art. 16, IV, exigiria, no mínimo, a existência de alguma governança desses dados, já que seria necessária a auditabilidade de seus fluxos, conforme exposto na Figura 12 do *framework*. No entanto, como abordamos, a lei não dispõe expressamente acerca da necessidade de governança

para esses dados. A ausência de governança inviabiliza a fiscalização do princípio da segurança, legalmente previsto para incidir nos dados anonimizados sob essa previsão legal do art. 16, IV.

Sugerimos então o estabelecimento de limites aos riscos que permeiam essa espécie de requisito preliminar de tratamento de dados. A **primeira limitação** que sugerimos é o aspecto já abordado de interpretação restritiva da lei acerca dos dados que podem se enquadrar nesse tipo de requisito preliminar de tratamento. Apontamos que os dados sensíveis não podem se enquadrar nessa previsão legal de tratamento de anonimização ao término do tratamento de dados pessoais, posto que as hipóteses de tratamento de dados sensíveis encontram-se restritas àquelas estabelecidas no artigo 11, conforme ilustrado na Figura 7.

A **segunda limitação** é aquela inerente ao tratamento de qualquer dado pessoal, que exige o respeito aos princípios do tratamento de dados mesmo na anonimização tanto quanto seja tecnicamente viável. Nesse caso, os princípios da finalidade e da adequação são viáveis e, portanto, asseguram ao usuário de que ele seja informado acerca da anonimização de seus dados pela modalidade estabelecida no art. 16, sendo que os dados terão por finalidade o uso exclusivo do controlador. Dessa forma, é assegurado ao titular se cientificar acerca do tratamento da anonimização e observar se o tratamento será realizado de maneira adequada, sem que haja a possibilidade de vinculação entre o dado e seu titular ao fim do tratamento.

Também sugerimos, como **terceira limitação**, o estabelecimento de um prazo máximo para armazenamento desses dados pelo controlador. Guedes e Meireles apontam que poderia ser utilizado nesses casos o prazo estabelecido pelo Marco Civil da Internet (Lei 12.965/2014), em seu art. 15, que prevê seis meses para armazenamento do registro de acesso aos provedores de aplicação na internet (GUEDES & MEIRELES, 2019).[104] Entretanto, entendemos que o prazo poderia ser melhor definido pela Autoridade Nacional, observando a congruência de interesses dos agentes envolvidos.

Como **quarta limitação**, entendemos como necessário o estabelecimento de parâmetros de governança desses dados, até para viabilizar a fiscalização do cumprimento do princípio da segurança, outrora mencionado.

Esses pontos de incongruência da lei ressaltam uma questão ética delicada acerca de formas de distribuição dos riscos e dos ganhos decorrentes do uso de dados anonimizados, principalmente pelas grandes plataformas. Isso porque a permissão de anonimização dos dados

104 Conforme teor: "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".

após qualquer tratamento é uma das previsões que têm maior potencial de fomentar o uso da ferramenta da anonimização para consolidação de plataformas de dados massivos, sem maiores restrições legais.

Por fim, destacamos que a LGPD não traz explicitamente o debate acerca das contingências da autodeterminação informativa, como a questão da assimetria informacional que trataremos adiante. Também não há indícios de uma preocupação da proteção dos dados para além do indivíduo, assunto que abordamos como emergente na quarta onda legislativa.

Concluímos, então, que a LGPD caminha ainda na consolidação da autodeterminação informativa, sem se atentar para as críticas contemporâneas que permeiam essa sistemática de proteção à privacidade. Nos esforços pela garantia da autodeterminação informativa, a lei apresenta lacunas, como é o caso da previsão do art. 16, IV, sobre a anonimização ao final do tratamento. Entendemos que a previsão necessita dos ajustes interpretativos e regulatórios de forma imediata, quais sejam: 1) a inviabilidade de aplicação desse requisito preliminar de tratamento para dados sensíveis; 2) a obediência aos princípios do tratamento de dados; 3) o estabelecimento de prazo máximo de armazenamento dos dados anonimizados nessa forma de tratamento; 4) a exigência de boas práticas de governança para viabilizar auditoria quanto ao requisito da segurança. De forma mediata, por sua vez, entendemos também como necessária uma discussão mais aprofundada sobre os riscos que devem ser assumidos de forma involuntária pelos cidadãos na anonimização de seus dados.

4.2.3.2. A questão comportamental de grupos e a anonimização

A coleta de informações comportamentais, ainda que anonimizadas, suscita questionamentos acerca dos limites éticos na utilização dos dados pessoais. Podemos questionar, por exemplo, se um usuário de rede social imagina que cada uma de suas ações e inações no uso desses aplicativos poderá ser utilizada para classificar perfis e personalidades e viabilizar ou não o acesso a serviços e produtos posteriormente.

Os dados, mesmo anonimizados, permitem a extração de informações de perfis de indivíduos e grupos que só são úteis em termos negociais porque são verossímeis. Os dados são úteis porque permitem o compartilhamento de parcela da personalidade, comportamento, ou visão de mundo dos usuários, de forma legal, segundo os parâmetros de privacidade atualmente estabelecidos. Portanto, esse tipo de uso dos dados não é necessariamente vedado pela lei, principalmente quando respaldados na anonimização.

A LGPD prevê no art. 12, §2º, que o perfil comportamental poderá ser considerado dado pessoal, caso o titular seja identificado. Desta forma, a lei permite que esses dados comportamentais sejam manipulados quando não se sabe ao certo de qual indivíduo, particularmente considerado dentro dos grupos do perfil, a informação foi extraída. A formação desses perfis é utilizada, como mencionamos anteriormente, para prestação de serviços personalizados, para persuasão dos consumidores, mas também para a criação de demandas antes inexistentes e mesmo para sugestionamentos e controle comportamental.

O indivíduo nesse cenário não é apenas o consumidor, mas o próprio produto, já que se torna fornecedor dos dados para fins de *Big Data Analytics*, e também define seu próprio "valor de mercado" através de sua exposição nas redes em busca de um espaço no mercado da atenção. No entanto, é no mínimo ardiloso que dados dispostos pelos usuários em momentos de lazer, situações em que a resistência e a diligência encontram-se sensivelmente diminuídas, sejam posteriormente utilizadas para finalidades totalmente diversas e não previsíveis, com o anteparo legal da anonimização.

Concluímos que a LGPD apesar de prever como dado pessoal a formação de perfis de pessoas identificadas, em seu art. 12, § 2º, passa ao largo da discussão ética acerca do uso de dados anonimizados para formação de perfis comportamentais genericamente considerados. As informações extraídas desses dados anonimizados podem ser igualmente prejudiciais e, portanto, é necessário a estipulação de parâmetros éticos para a sua utilização.

4.2.3.3. O paradoxo da Anonimização e o Aprofundamento de Assimetrias Informacionais e Desigualdades Sociais

Apesar dos dilemas e desafios que envolvem a anonimização, deve-se reconhecer sua importância como ferramenta relativamente simples e de fácil acesso, que fomenta a privacidade. A própria LGPD aponta a anonimização como um direito do titular quando o permite exigir que os dados "desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei" sejam anonimizados, conforme teor do art. 18, IV. Ressalte-se que Ohm, mesmo apesar de suas críticas, indica que a anonimização não deve ser desprezada como técnica, já que em muitos casos garante uma proteção difícil de ser revertida (OHM, 2010, p. 1716). Nesse sentido, a técnica tem sua importância inegável para o fomento da privacidade.

A crítica, no entanto, é dura, justamente porque os agentes têm deixado de levar em consideração as limitações da própria ferramenta. Essas limitações não podem ser ignoradas,

sob pena de acarretarem maiores danos do que benefícios, principalmente quando se considera a livre divulgação dos dados.

A confiança irrefletida na anonimização pode ser um fator de aprofundamento das diversas assimetrias informacionais que abordamos. O desequilíbrio de poder, anterior à sociedade informacional, se intensificou com a capacidade de processamento massiva de dados do *Big Data Analytics* e com um período considerável de vácuo regulatório que permitiu o armazenamento desenfreado e o tratamento de dados de formas escusas. A assimetria informacional decorrente desse cenário pode ser fomentada ou minorada de acordo com as diretrizes que traçaremos para o uso dos dados, uma vez cientes dos riscos que os envolvem.

Dessa forma, a escolha legislativa da LGPD em flexibilizar excessivamente os parâmetros regulatórios para dados anonimizados pode fomentar esse abismo informacional, acentuando o desequilíbrio de poder. Nem mesmo um direcionamento principiológico foi estabelecido para a utilização desses dados, que permanecem com potencial de dano em sua utilização ímproba. Ressalte-se que esse tipo de direcionamento não deveria implicar em excesso regulatório para a *Data-Driven Economy*, mas a livre utilização desses dados implica em riscos desarrazoados, não apenas para a privacidade, mas para o equilíbrio de poder em cenários democráticos. A própria LGPD traz princípios que poderiam ser aproveitados para o estabelecimento dessas diretrizes. É o caso, por exemplo, do princípio da Não Discriminação, que dispõe sobre "a impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos". Trata-se de um princípio de conduta ética que poderia ser aplicado aos danos anonimizados sem implicar em excessiva interferência nesses mercados. Esse tipo de direcionamento garantiria um arcabouço legislativo de proteção a esses dados, atualmente inexistente.

Talvez essa lacuna principiológica descrita venha ao encontro do interesse em se conferir a natureza jurídica puramente patrimonial aos dados anonimizados. Mas essa escolha legislativa acaba por ignorar que os dados anonimizados ainda tem o potencial de acarretar danos a personalidade dos indivíduos e, justamente para mitigar esses riscos, haveria a necessidade da estipulação de diretrizes éticas.

Nesse sentido, a legislação brasileira segue a predição que Ohm faz acerca da balança entre privacidade dos dados e manutenção dos fluxos da informação. Como mencionamos, Ohm aponta que normalmente essa balança não tem mantido um equilíbrio, já que tem pendido em desfavor da privacidade, por vezes, com a excessiva confiança na anonimização robusta, e por outras vezes, na restrição severa dos fluxos, prejudicando o acesso à informação e a inovação (OHM, 2010, p. 1740). Ohm sugere então uma regulação que

reequilibre essa balança. O autor propõe que os riscos de dano pela reidentificação ou má utilização dos dados sejam quantificados, de acordo com o contexto, na denominada integridade contextual. Além disso, uma vez identificado o *score* de risco, devem ser tomadas medidas para que esses riscos sejam amenizados. Ohm propõe 5 fatores para se mensurar o risco de dano à privacidade, que aplicamos aqui especificamente aos dados anonimizados (OHM, 2010, p. 1765-1768):

1) Estabelecer parâmetros mais objetivos para mensurar riscos para a técnica. Nesse sentido, a nossa crítica é sobre a insuficiência dos parâmetros de razoabilidade hoje existentes na LGPD para se considerar um dado como anonimizado. Esses parâmetros devem ser especificados pela Autoridade Nacional, de forma a facilitar a auditabilidade dessa técnica. Além disso, a mensuração da linkabilidade e do poder de inferência que abordamos também seriam ferramentas úteis de auditabilidade de riscos. É possível inclusive pensar em ferramentas automatizadas de auditoria com esses parâmetros e *softwares* que pudessem prever essa classificação de risco.

2) Preferência pelo compartilhamento privado de dados ao invés da divulgação de dados publicamente, ou seja, a restrição da divulgação pública em prol do compartilhamento de dados entre parceiros. O autor aborda que a divulgação pública de dados tem beneficiado grandes empresas com a sabedoria das multidões, mas que essa liberação causaria riscos que deveriam ser remediados. Concordamos com o autor que a divulgação pública faz com que esses dados sejam facilmente assimilados em grandes bancos, diminuindo a entropia. No entanto, entendemos que há também um ponto negativo na retenção de informações em setores privados. Principalmente no que tange ao fato de a economia de dados se alimentar desses ativos. A restrição à publicização de dados poderia aumentar ainda mais as lacunas informacionais que já compõem o cenário atual, tornando cada vez mais difícil o surgimento de novas iniciativas no setor. Afinal de contas, a retenção de acesso aos dados faz com que o principal ativo circule apenas entre aqueles que já têm sua posição firme nos mercados. Outra questão é que a disponibilização de dados públicos deveria levar em consideração o dever de transparência, dificultando ainda mais a escolha da decisão pelo compartilhamento apenas interno.

3) Regular não só a qualidade dos dados (sensíveis, pessoais ou pseudonimizados), mas a quantidade dos dados e o tempo que eles podem ser retidos. O autor sugere estabelecer limites quantitativos e tempo para descarte. A LGPD já prevê essa questão para os dados pessoais em sentido *lato*, com a incidência do princípio da necessidade, mas não há nenhuma limitação para os dados não considerados pessoais. Dessa forma, a

sugestão seria então uma regulação Setorial, não baseada necessariamente em PII, ou *quasi-identifiers*, mas na gestão dos dados como um todo. Esse tipo de regulação levaria em consideração a inferência que se pode fazer da combinação de determinados dados, balanceando os riscos e os ganhos decorrentes da possibilidade dessa acumulação (OHM, 2010, p. 1775).

4) Flexibilizar as restrições de acesso para grupos específicos, como por exemplo, os dados não publicizados para fins de pesquisa científica. Esse fator seria aliado principalmente da confiança tratada a seguir.

5) Atribuir confiança, não mais na tecnologia da anonimização, mas nos controladores, estabelecendo os parceiros confiáveis (governos, pesquisadores, etc.). Ohm sugere a institucionalização da confiança social depositada nos controladores, permitindo aos agentes confiáveis até mesmo mais informação do que seria disponível nos sistemas de anonimização. Para nós, a confiabilidade nesses agentes de tratamento deveria ser estruturada por uma regulação para além da legislação e balizada em 4 importantes parâmetros, quais sejam:

a) A utilização de boas práticas de governança,

b) A certificação de padrões de segurança,

c) Os resultados obtidos em auditorias acerca do *compliance* à lei, ou seja, o histórico de atuação do agente,

d) Os níveis de preservação da autodeterminação dos titulares, para além daqueles parâmetros de *compliance* já estipulados pela lei.

Essas cinco sugestões de Ohm facilitariam a adequação da regulamentação aos casos concretos, de acordo com o contexto.

Ohm também alerta acerca da inviabilidade de uma regulação legislativa exaustiva. Segundo o autor, a legislação não deve nem ser tão específica a determinados setores, nem tão genérica. Nesse sentido, Ohm critica a legislação protetiva dos EUA, ressaltando que ela é exageradamente específica e não aponta requisitos básicos para o tratamento de dados pelos agentes de forma geral. Por outro lado, também critica o regulamento europeu ao legislar de forma muito dura quanto ao tratamento de dados para qualquer que seja o agente e sua base de dados. Para ele, a anonimização, nos moldes do que tem ocorrido com o GDPR, acaba esvaziando seu propósito pelo excesso regulatório (OHM, 2010, p. 1762). A sugestão é, então, um *framework* geral, mas que em sua aplicação e verificação leve em conta o contexto específico e sua adaptabilidade (OHM, 2010, p.1762). É importante que exista uma lei balizadora de requisitos mínimos de tratamento de dados para todos os agentes e que as especificidades sejam reguladas na medida de sua potencialidade de dano.

A LGPD, nesse ponto, parece em consonância com a sugestão do autor, já que estabelece uma estrutura que pode ser complementada por normas administrativas e regulamentares, de acordo com as necessidades. Justamente por causa desse caráter mais geral da LGPD, a professora Ana Frazão, por exemplo, se questiona acerca da necessidade de um endereçamento especializado da legislação concorrencial para os problemas decorrentes do tratamento de dados pessoais nessa vertente do direito (FRAZÃO, 2021). [105] O mesmo se aplica para demais vertentes do Direito envolvidas na questão, como o Direito do Consumidor, Direito Civil, Direito Penal, etc.

Entretanto, de qualquer forma, muitos desses parâmetros desejáveis não poderão ser estabelecidos apenas pela lei, já que a padronização nesses contextos nem sempre é uma boa solução, tendo em consideração as assimetrias. A regulamentação deve se atentar às desigualdades preexistentes, a fim de não inviabilizar novos atores e sua estratégia de negócio. As regras de governança devem ser cobradas de forma mais ou menos rigorosa de acordo com as capacidades dos atores envolvidos na regulação.

Em consonância com o contexto delicado de recursos regulatórios escassos, os reguladores, em sua atuação infralegal, devem direcionar sua atenção aos bancos de dados massivos que contribuem para um risco ainda mais acentuado, sejam eles públicos ou privados. Ohm propõe a criação de uma categoria própria para os proprietários de grandes bancos de dados, chamado de "grandes redutores de entropia" (OHM, 2010, p. 1760). Esses "grandes redutores de entropia" seriam entidades que acumulam bancos de dados maciços, contendo maior possibilidade de ligação pela diversidade e extensão de seus dados e mantendo um grande potencial de linkabilidade e inferência mesmo que excluam de seus bancos de dados todas as informações particularmente sensíveis e diretamente vinculáveis (OHM, 2010, p.1760). Ohm sugere que essas bases sejam regulamentadas de forma mais rígida, com novas regras especialmente criadas para esta realidade ou com a aplicação de regras preexistentes (OHM, 2010, p. 1760-1761), de modo a propor diretrizes restritivas ao comportamento dos controladores dessas bases.

Com base nas observações de Ohm, o que propomos é, então, a avaliação e mensuração dos riscos e dos possíveis retornos sociais decorrentes dessas plataformas *Big Data*,

---

[105] Nas palavras da autora: "A grande questão é saber em que medida outras searas deverão assumir, de forma conjunta e harmônica com a LGPD, um papel mais incisivo no que diz respeito à regulação dos dados. Entretanto, no tocante ao Direito da Concorrência, nem mesmo deveria haver tal questionamento, considerando a constatação de que os dados são hoje fontes importantíssimas de poder econômico. Dessa maneira, a pergunta a ser feita não deveria ser se caberia ao Direito da Concorrência intervir ou não na questão dos dados, mas sim em que medida e para que propósitos a análise antitruste deve endereçar a questão dos dados" (FRAZÃO, 2021).

vedando condutas discriminatórias e antidemocráticas. Nesse sentido, é necessário pensar a questão em duas vertentes: 1) na estipulação de limites aos controladores e 2) na conscientização do cidadão quanto aos riscos, equilibrando-os com o retorno social do uso desses dados, e mesmo com formas de reparação de danos pelos riscos que os usuários estão, em alguma medida, sempre submetidos.

De fato, urge a necessidade de que a sociedade debata com mais clareza acerca da exposição a esses riscos. Stiglitz, por exemplo, sugere acabar com o foco excessivo no benefício de curto prazo do consumidor, que normalmente é hipervalorizado quando se trata dos bens e serviços de plataformas privadas (STIGLITZ, 2019). Dessa forma, a inovação e as facilidades delas decorrentes devem servir ao usuário e não o contrário, e devem levar em consideração esse titular não apenas como consumidor imediato, mas como indivíduo sujeito de direitos fundamentais, como ser social dotado de personalidade.

O usuário deve ter em mente que a anonimização é convencionada, e não absoluta. Ou seja, ao consentir com a anonimização de seus dados, o titular deve ter como expectativa não a impossibilidade total de reidentificação, mas o cumprimento de padrões de exigência que determinam a razoabilidade dos riscos.

Existem ainda propostas de distribuição dos riscos de exposição, a partir da exigibilidade de que as plataformas de *Big Data Analytics* obtenham seguros contra riscos ou sejam submetidas a premissas específicas de responsabilidade civil. Mas de toda a forma, a decisão entre os riscos que envolvem a anonimização, a disponibilização desses dados, o livre fluxo de informação e os benefícios gerados desses fluxos sempre seria por fim uma decisão difícil dos diversos agentes.

A questão não é, de forma alguma, simples, principalmente quando consideramos a assimetria informacional envolvida nessas relações e o contexto ambíguo que a tecnologia proporciona, envolvendo o dilema da privacidade, transparência, acesso a produtos e serviços personalizados e a inovação. Entretanto, ela deve ser debatida, estabelecendo-se limites éticos ao uso dos dados e privilegiando sempre a boa-fé dos usuários. A estipulação de parâmetros de confiança para os controladores e principalmente a auditabilidade de suas ações é essencial para que a sociedade tenha maior controle sobre os riscos e maior poder de decisão sobre as formas de regulação.

Concluímos, portanto, que a LGPD é uma legislação que aponta diretrizes gerais, mas não se aprofunda no debate sobre assimetrias informacionais, e destacamos que este se trata de um campo de disputas de poder que deve ser cuidadosamente avaliado pela sociedade. A questão pode se desdobrar em legislações específicas em diversos campos do Direito, mas

também é possível realizar alguns ajustes na legislação de proteção de dados para nortear soluções. Sugerimos o estabelecimento de diretrizes principiológicas para o uso de dados anonimizados e trouxemos as cinco sugestões práticas de Ohm para a questão, quais sejam: 1) parâmetros mais objetivos para mensurar riscos da técnica; 2) compartilhamento privado de dados ao invés da divulgação pública; 3) regulação da quantidade dos dados e o tempo que eles podem ser retidos; 4) flexibilização de restrições de acesso para grupos específicos; e 5) Atribuição de níveis de confiança para os controladores. Especificamente quanto à sugestão 5, ressaltamos que a confiança deveria ser orientada a partir de quatro fatores: a) boas práticas de governança, b) certificação de padrões de segurança, c) resultados obtidos em auditorias acerca do *compliance* à lei, e d) níveis de preservação da autodeterminação dos titulares. Todas essas sugestões vêm ao encontro de uma legislação diferenciada para os "grandes redutores de entropia", sejam eles atores públicos ou privados, balizando riscos e o retorno social no uso dos dados.

5. Conclusões

A anonimização desponta na legislação brasileira e, em especial, na LGPD como uma ferramenta importante para a proteção de dados, pois combina a preocupação com a privacidade com a manutenção de formas de negócios que se utilizam de bancos de dados massivos.

Comparada a outras técnicas de privacidade, a anonimização é uma das mais otimistas, já que é bem menos complexa, cara e com menores limites técnicos do que as demais (OHM, 2010, p. 1751).

No entanto, sem ignorar a importância da ferramenta, até pela maior proteção que ela confere à privacidade e a relativa simplicidade de sua aplicação, procuramos delinear alguns dos limites que apresenta. Tomamos por base o pioneiro texto de Paul Ohm para realizar uma classificação entre os limites intrínsecos, extrínsecos à técnica e os limites externos, que surgem na utilização prática desses dados.

Procuramos ainda estabelecer um *framework* para orientar os desenvolvedores sobre os requisitos listados pela LGPD para a anonimização, tanto os requisitos específicos quanto os requisitos gerais para qualquer processamento de dados pessoais.

Ao comparar criticamente os requisitos legais da anonimização expostos no *framework* que construímos e os limites que extraímos do texto de Ohm, foi possível encontrar lacunas jurídicas quanto a diversas questões abordadas. Em todos os níveis de classificação,

seja os limites intrínsecos, extrínsecos ou externos, a lei não é exaustiva em sua abordagem, delimitando apenas parâmetros gerais para anonimização.

Em relação aos **limites intrínsecos,** temos dificuldades em definir, apenas com base na LGPD, o que seria uma "informação teoricamente segura" no contexto da anonimização. Também observamos lacunas acerca do dilema da integridade dos dados, principalmente para a transparência e confiabilidade de dados públicos, ainda que a lei reconheça o *trade-off* de utilidade e confiabilidade desses ativos.

Quanto aos **limites extrínsecos**, a lei deixa brechas ao poder de inferência que não reidentifique imediatamente o dado e, por esse motivo, fizemos três sugestões de adequação: 1) a utilização da linkabilidade e da inferência como parâmetros auditáveis em bases anonimizadas; 2) a inclusão de bases de dados externas e acessíveis nos parâmetros de caracterização da anonimização; 3) a maior ponderação na disponibilização de dados anonimizados.

No que se refere à disponibilização de dados, constatamos, com pesar, que a LGPD não é incompatível com modelo liberação-esquecimento para dados anonimizados, já que não existem limites claros, nem mesmo principiológicos, ao livre fluxo e publicização desses dados, à exceção do art. 16, IV da LGPD. Também não existem diretrizes claras acerca da responsabilização por danos decorrentes de dados publicizados ou compartilhados. Sobre a exigibilidade de adoção de boas técnicas de governança para esses dados a lei é omissa, ainda que esse tipo de gestão seja necessário para viabilizar o *accountability* da anonimização.

Por fim, quanto aos **limites externos**, constatamos que a LGPD caminha ainda na consolidação da autodeterminação informativa, sendo incipientes as discussões sobre as falhas que esse modelo de privacidade apresenta, quando consideradas as assimetrias informacionais e o caráter coletivo e difuso dos dados anonimizados. Mesmo a autodeterminação informativa é mitigada na previsão do requisito preliminar de tratamento disposto no art. 16, IV. Sugerimos ajustes interpretativos para mitigar os riscos que envolvem essa previsão legal, quais sejam: 1) a inviabilidade desse requisito para dados sensíveis; 2) a obediência aos princípios do tratamento de dados; 3) o estabelecimento de prazo máximo de armazenamento dos dados anonimizados nessa forma de tratamento; e 4) a exigência de boas práticas de governança para auditoria da segurança.

A lei também viabiliza a formação de perfis comportamentais com dados anonimizados, em seu art. 12, § 2º, sem se atentar para os riscos na utilização maliciosa desses recursos e sem estabelecer parâmetros éticos principiológicos para nortear o uso de dados

anonimizados. Sugerimos como exemplo a necessidade da aplicação de princípios como o da Não Discriminação, também para o tratamento desses dados.

Para minimizar os efeitos das assimetrias informacionais e das assimetrias de poder que envolvem a matéria, trouxemos as cinco sugestões práticas de Ohm, das quais destacamos a atribuição de níveis de confiança para os controladores a partir da análise de quatro fatores de confiabilidade, quais sejam: 1) a utilização de boas práticas de governança; 2) a certificação de padrões de segurança; 3) a obtenção de bons resultados em auditorias acerca do *compliance* à lei; e 4) a preservação de altos níveis de autodeterminação dos titulares em seus tratamentos. Ressaltamos ainda que o enfoque regulatório deve estar nos "grandes redutores de entropia", sejam eles atores públicos ou privados,

Feitas essas considerações, é possível se observar que, por um lado, o caráter geral da LGPD permite uma maior liberdade regulatória para a Autoridade Nacional, que encontra respaldo legal para estabelecer parâmetros mais objetivos de fiscalização da técnica. Além disso, é viável o estabelecimento de regulação setorial sobre anonimização de dados a partir de relações de confiança com os agentes, por meio do cumprimento de padrões internacionais, certificações e auditabilidade.

Por outro lado, a ausência de previsões mais cautelosas para nortear o que a lei considera como dado anonimizado e para direcionar a gestão desses dados acarreta a fragilização da técnica de forma geral. Por ser um estado situacional e por estar vinculada em rede a outros dados através da linkabilidade e da inferência, a anonimização sofre abalos, de forma generalizada, quando dados são disponibilizados de forma irresponsável ou quando a reidentificação é facilmente alcançada.

É importante destacar que os critérios de anonimização estarão sempre ligados à evolução da técnica, o que exige que os desenvolvedores atualizem constantemente os parâmetros utilizados. Desta feita, a anonimização não pode ser considerada uma característica imutável do dado, mas apenas um *status*, que permanece enquanto cumpridas as exigências legais.

Concluímos que orientações mais objetivas são necessárias para que exista maior segurança jurídica em torno da técnica. Sabemos que muitos desses critérios serão definidos pela Autoridade Nacional de Proteção de Dados, e devem ser estabelecidos levando em consideração não apenas os limites intrínsecos e extrínsecos à técnica, mas principalmente as questões éticas que envolvem o uso dos dados anonimizados, as assimetrias informacionais e os direitos dos titulares.

A exposição destes critérios proporcionará maior segurança jurídica aos controladores, quanto à qualidade e cumprimento legal nos processos de anonimização. Uma vez que os requisitos não estão suficientemente especificados objetivamente, a lei não pode ser considerada exaustiva para os desenvolvedores. No entanto, essas lacunas não impedem a aplicação de técnicas de anonimização, desde que os profissionais respeitem os requisitos legais e os princípios erigidos pela LGPD, conforme enumerados no *framework.* Os desenvolvedores devem então nortear a anonimização e o preenchimento das lacunas legais priorizando as diretrizes gerais da lei de proteção de dados e os critérios de privacidade.

Além disso, julgamos extremamente necessário e urgente o fomento ao debate público acerca dos riscos que estão envolvidos na anonimização e principalmente no livre fluxo de dados anonimizados. Entendemos também como necessária a discussão sobre os riscos que serão assumidos de forma involuntária pelos cidadãos na anonimização de seus dados.

A decisão sobre que riscos estamos dispostos a correr não é simples e deve ser entendida de forma abrangente, tomando-se em consideração o maior número de fatores possível, inclusive as assimetrias de poder que envolvem a questão.

A eficácia das medidas possíveis para minimizar esses riscos encontra obstáculos muito menos nas limitações das técnicas do que nas escolhas tomadas pelos indivíduos envolvidos. Não é sem razão que Barbieri aponta que “a Governança de Dados, como desafio, é menos de Governança, menos de dados e muito mais de pessoas” (BARBIERI, 2019, p. 32).

Dessa forma, é necessário ter em mente que o problema não está nas ferramentas, mas nas decisões tomadas sobre elas em seu uso para finalidades escusas. A tecnologia em si não pode ser considerada a principal causa da desigualdade social e nem mesmo da desigualdade informacional. As assimetrias estariam mais voltadas às relações de poder que orientam o uso das tecnologias.

Nesse sentido, não é possível descartar a importância da anonimização como técnica em prol da privacidade, responsabilizando-a pelos excessos derivados de sua disposição ou mal uso dos dados anonimizados. É o que aponta Paul Krugman quando afirma que muitas vezes o discurso prevalente “culpa os robôs” por decisões na verdade tomadas para manutenção das relações de poder. A tecnologia se traduz diversas vezes em aumento da produtividade e em ganhos econômicos significativos que não são repassados às “vítimas do progresso tecnológico”. O discurso privilegia a inevitabilidade da existência de mártires e repassa a culpa inclusive para esses atores, indicando "lacuna de habilidades" para lidar com a nova era digital. Mas a verdade é que o discurso é uma forma de se desviar a atenção das más políticas praticadas e das decisões motivadas pela manutenção de privilégios (KRUGMAN, 2020, p. 122). Da

mesma forma, a flexibilização da privacidade ou a exposição de dados não pode ser vista como resultado da simples inaptidão digital dos usuários, descartando os interesses por trás da captura desses ativos.

Como aponta Schwab, o objetivo das tecnologias é o fim a que as sociedades as destinam. Precisamos ter em mente qual o nosso objetivo comunitário no uso dessas ferramentas:

> A tecnologia não é uma força externa sobre a qual não temos nenhum controle. Não estamos limitados por uma escolha binária entre "aceitar e viver com ela" ou "rejeitar e viver sem ela", Na verdade, tomamos a dramática mudança tecnológica como um convite para refletirmos sobre quem somos e como vemos o mundo, Quanto mais pensamos sobre como aproveitar a revolução tecnológica, mais analisamos a nós mesmos e os modelos sociais subjacentes que são incorporados e permitidos por essas tecnologias (SCHWAB, 2016, p. 14).

Dessa forma, Schwab afirma que não são os limites técnicos que restringirão o uso dessas tecnologias, já que o avanço científico ocorreria de forma a suprir as lacunas. A forma de limite do uso equivocado dessas tecnologias seria os ditames da ética por meio dos regulamentos e legislações (SCHWAB, 2016, p. 30). São esses os parâmetros que não devem, de forma alguma, ser desconsiderados quando pensamos na regulação dessas tecnologias. No entanto, como vimos, são justamente os parâmetros éticos que mais encontram omissão legal e que são muitas vezes suprimidos do debate público, favorecendo a manutenção de assimetrias de poder, incremento de riscos e violações.

Todavia, justamente essas escolhas humanas sobre limitar eticamente o uso dessas tecnologias é que serão decisivas para determinar como definimos e definiremos a privacidade e a sua importância para a sociedade informacional. São essas escolhas que definirão em que medida teremos de fato o controle sobre esses ativos, que ao fim e ao cabo, são nossa verdadeira expressão de ser na era digital.

6. Referências Bibliográficas:

ALTARTURI, Hamza Hussein; NG, Keng-Yap; NINGGAL, Mohd Izuan Hafez; NAZRI, Azree Shahrel Ahmad; GHANI, Abdul Azim Abd. A requirement engineering model for big data software. In 2017 **IEEE Conference on Big Data and Analytics (ICBDA)**. IEEE, 2017. p. 111–117.

ALTMAN, Steven, A.; BASTIAN, Phillip. **DHL Global Connectedness Index 2020: The State of Globalization in a Distancing World.** NYU Stern School of Business. Disponível

em: https://www.dhl.com/content/dam/dhl/global/dhl-spotlight/documents/pdf/spotlight-g04-global-connectedness-index-2020.pdf. Acesso em: 20/04/2021.

ARBUCKLE. Luk; EL EMAM, Khaled. **Anonymizing Health Data: Case Studies and Methods to Get You Started**. 1ª Edição. Estados Unidos da América: O'Reilly, 2013.

BAUMAN, Zygmunt**.** 25 DE FEVEREIRO DE 2011: Sobre Facebook, intimidade e extimidade. In: **Isto não é um diário.** Tradução: Carlos Alberto Medeiros. Rio de Janeiro: Editora Zahar, 2012.

BARBIERI, Carlos. **Governança de Dados: Práticas, Conceitos e Novos Caminhos**. 1ª Edição: Rio de Janeiro: Editora Alta Books, 2020.

BBC, **Como a forma de digitar revela a sua identidade**. BBC News Brasil. 12 de agosto de 2015. Disponível em: https://www.bbc.com/portuguese/noticias/2015/08/150812_estudo_teclado_seguranca_lgb. Acesso em: 02/03/2021.

BBC NEWS. **Snowden: leaks that exposed US spy programme**. BBC News. 17 de Janeiro de 2014. Disponível em: http://www.bbc.com/news/world-us-canada-23123964. Acesso em 02/08/2018.

BRASHER, E. A. **Addressing the Failure of Anonymization: Guidance from the European Union's General Data Protection Regulation.** Colum: Bus. L. Rev. v. 1, n. 1, p. 8–23. 2018.

BOYLE, James. **A theory of law and information: Copyright, spleens, blackmail, and insider trading**. California Law Review, v. 80, p. 1413, 1992. Disponível em: https://scholarship.law.duke.edu/faculty_scholarship/166/. Acesso em: 18/08/2020.

BRODSKY, Paul. **Internet Traffic and Capacity in Covid- Adjusted Terms**. Blog Telegeography. 27 de agosto de 2020. Disponível em: https://blog.telegeography.com/internet-traffic-and-capacity-in-covid-adjusted-terms. Acesso em: 24/04/2021.

BUGHIN, Jacques; LUND, Susan. **The ascendancy of international data flows.** McKinsey Global Institute. 9 de Janeiro de 2017. Disponível em: https://www.mckinsey.com/mgi/overview/in-the-news/the-ascendancy-of-international-data-flows. Acesso em: 01/03/2019.

BYRNE, Michael. **O que é Privacidade Diferencial?** Motherboard, tech by Vice. 8 de Janeiro de 2015. Disponível em: https://www.vice.com/pt_br/article/nz3avw/o-que-e-privacidade-diferencial. Acesso em: 05/09/2020.

CANEDO, Edna Dias; KOSLOSKI, Ricardo Ajax Dias; MARTINS, Hugo Ferreira; OLIVEIRA, Edgard Costa; OLIVEIRA JÚNIOR, Antônio Carvalho de; PALDÊS, Roberto Ávila. Design Thinking: Challenges for Software Requirements Elicitation. In: **Information 2019**, MDPI, v.10, n.371, 2019.

CARVALHO, Artur Potiguara; CANEDO, Edna Dias; CARVALHO, Fernanda Potiguara; CARVALHO, Pedro Henrique Potiguara. Big Data, Anonymisation and Governance to

Personal Data Protection. In: **The 21st Annual International Conference on Digital Government Research.** 2020. p. 185–195.

CARVALHO, Pedro Henrique Potiguara. **A segurança dos dados: uma perspectiva de banco de dados.** Trabalho Final de Conclusão de curso de Pós-Graduação. AVM- Faculdade integrada. Brasília. 2016.

CASTELLS, Manuel. **A sociedade em Rede**. Vol.I - Tradução: Roneide Venâncio Majer. Ed. Paz e Terra, 1999.

CELLAN-JONES, Rory. **Amazon completa 20 anos de muito sucesso e pouco lucro**. BBC News, 15 de junho de 2015. Disponível em: https://www.bbc.com/portuguese/noticias/2015/07/150715_amazon_20_anos_rb. Acesso em: 01/06/2019.

CHEMUTURI, Murali. **Requirements Engineering and Management for Software Development Projects**. New York: Springer, 2013.

CIRANI, Simone; FERRARI, Gianluigi; PICONE, Marco; VELTRI, Luca. **Internet of Things: Architectures, Protocols and Standards**. Ed. Wiley, 2019.

COMAS, Jordi Soria. **Improving data utility in differential privacy and K-anonymity**. Tese de Doutorado. Departamento de Engenharia da Computação e Matemática. Universidade de Tarragona. 2013

COMITÊ EUROPEU PARA PROTEÇÃO DE DADOS**. Parecer n. 5 de 2014**. Grupo de Trabalho de Proteção de dados do art. 29. disponível em://oemmndcbldboiebfnladdacbdfmadadm/https://www.gpdp.gov.mo/uploadfile/2016/0831/20160831042518381.pdf. Acesso em 25.08.2019.

COSTA, Frederico Lustosa da. Brasil: 200 anos de Estado; 200 anos de administração pública; 200 anos de reformas. In: **Revista de Administração Pública- RAP**. Rio de Janeiro: 42(5):829-74, Set/Out. 2008.

CRARY, Jonathan. **24/7: Capitalismo tardios e os fins do sono**. São Paulo: Ubu, 2016.

CRUZ, Bruna Souza**. Por que o governo compra dados dos próprios cidadãos?** 18 de Junho de 2018. Disponível em: https://tecnologia.uol.com.br/noticias/redacao/2018/06/18/entenda-por-que-o-governo-compra-dados-publicos-dos-proprios-cidadaos.htm. Acesso em 03/08/2018.

CUEVAS, Ricardo Villas Bôas. A proteção de dados pessoais na jurisprudência do Superior Tribunal de Justiça. In**: Lei geral de proteção de dados pessoais: suas repercussões no direito brasileiro.** São Paulo: Ed. RT, 2019.

DAMA-DMBOK, The Data Management Association. **O Guia da DAMA para o corpo de conhecimento em gestão de dados DAMA-DMBOK**. Editor Desenvolvimento: Mark Mosley; Editor Produção: Michael Brackett; Editora Assistente: Susan Earley; Patrocinadora do Projeto: Deborah Henderson. Tradução: Rossano Soares Tavares. Dama Brasil, Westfield: Ed. Technics Publications, LLC, 2012.

DATE, Christopher J. **Introdução a sistemas de bancos de dados**. Elsevier Brasil, 2004.

DOMINGO-FERRER, Josep; MURALIDHAR, Krishnamurty; BRAS-AMOR´OS, Maria. General Confidentiality and Utility Metrics for Privacy-Preserving Data Publishing Based on the Permutation Model. In**: IEEE Transactions on Dependable and Secure Computing**. Jan.2020

DOMINGO-FERRER, Josep. Personal Big Data, GDPR and Anonymization. In: **13th International Conference on Flexible Query Answering System**. Springer, 2019. p. 7–10.
DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**. 1ª Edição. Rio de Janeiro: Ed. Renovar, 2006.

DONEDA, Danilo. A proteção dos dados pessoais como direito fundamental. In: **Espaço Jurídico Joaçaba**, v. 12, n. 2, p. 91-108, jul./dez. 2011.

DONEDA, Danilo; MACHADO, Diego. Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados. In. **Revista dos Tribunais Caderno Especial Vol. 998**. São Paulo, Ed. RT, p. 99-128, dezembro. 2018.

DONEDA, Danilo. Panorama Histórico da Proteção de Dados Pessoais. In: DONEDA, Danilo; SARLET, Ingo Wolfgang; MENDES, Laura Schertel Mendes, RODRIGUES JÚNIOR, Otávio Luiz; BIONI, Bruno Ricardo. **Tratado de proteção de dados pessoais**. Rio de Janeiro: Ed. Forense [eBook], 2021.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. 6ª Edição. São Paulo: Editora Pearson, 2011.

FALBO, Ricardo de Almeida Falbo; SOUZA, Vítor. **Introdução à Orientação a Objeto.** Departamento de Informática Universidade Federal do Espírito Santo. 11 de abril de 2006. Disponível em: http://www.inf.ufes.br/~vitorsouza/archive/2020/wp-content/uploads/academia-br-cursooo-slides03.pdf. Acesso em: 01/12/2020.

FERNANDES, Aguinaldo A; ABREU, Vladmir F. **Implantando a Governança de TI: Da Estratégia à Gestão dos Processos e Serviços**. 4ª Edição. Brasport, 2014.

FOTHERGILL, Dr. B; KNIGHT, William; STAHL, Bernd Carsten; ULNICANE, Inga. **Responsible Data Governance of Neuroscience Big Data**. Frontiers in neuroinformatics. 24 de abril de 2019. Disponível em: https://www.frontiersin.org/articles/10.3389/fninf.2019.00028/full. Acesso em: 05/05/2020.

FRAGOSO, Nathalie; MASSARO, Heloísa. **Cadastro Base e amplo compartilhamento de dados pessoais: a que se destina?** Internet Lab: Pesquisa em Direito e Tecnologia. 20 de dezembro de 19. Disponível em: https://www.internetlab.org.br/pt/privacidade-e-vigilancia/cadastro-base-e-amplo-compartilhamento-de-dados-pessoais-a-que-se-destina/. Acesso em 02/02/2020

FRAZÃO, Ana. **Geopricing e geoblocking: as novas formas de discriminação de consumidores. Os desafios para o seu enfrentamento**. JOTA. 15 de Agosto de 2018. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-

mercado/geopricing-e-geoblocking-as-novas-formas-de-discriminacao-de-consumidores-15082018. Acesso em: 10/01/2021.

FRAZÃO, Ana. *Big Data* e Aspectos Concorrenciais do Tratamento de Dados Pessoais. In: DONEDA, Danilo; SARLET, Ingo Wolfgang; MENDES, Laura Schertel Mendes, RODRIGUES JÚNIOR, Otávio Luiz; BIONI, Bruno Ricardo. **Tratado de proteção de dados pessoais**. Rio de Janeiro: Ed. Forense [eBook], 2021.

GARFINKEL, Simson**. Database nation**. Sebastopol: O'reilly, 2000.

GDPR. **General Data Protection Regulation**. EU data protection rules.2018. Disponível em: https://ec.europa.eu/commission/priorities/justice-and-fundamentalrights/data protection/2018-reform-eu-data-protection-rules_en. Acesso em: 06/07/2019.

GJERMUNDRØD, Harald; DIONYSIOU, Ioanna Dionysiou; COSTA, Kyriakos. Privacy-Tracker: a privacy-by-design GDPR-compliant framework with verifiable data traceability controls. In: **International Conference on Web Engineering.** Springer, 2016. p. 3–15.

GOODRICH, Michael T; TAMASSIA, Roberto. **Introdução à Segurança de computadores.** Tradução: Maria Lúcia Blanck Lisbôa. Revisão Técnica: Raul Fernando Weber. Porto Alegre: Editora Bookman, 2012.

GOODWIN, Tom. **The Battle is for the Costumer Interface**. TechCrunch. 3 de março de 2015. Disponível em: https://techcrunch.com/2015/03/03/in-the-age-of-disintermediation-the-battle-is-all-for-the-customer-interface/. Acesso em: 29/04/2021.

GOVERNO FEDERAL. Ministério da Justiça e Segurança Pública. **O que é RIC?.** Disponível em:https://www.justica.gov.br/Acesso/governanca/ric. acesso em: 28/05/2020.

GUEDES, Gisela Sampaio da Cruz; MEIRELES, Rose Melo Vencelau. Término do Tratamento de dados. In: FRAZÃO, Ana; OLIVA, Milena Donato; TEPEDINO, Gustavo (coord.). **Lei Geral de Proteção de Dados Pessoais: e suas repercussões no Direito brasileiro**. São Paulo: Revista dos Tribunais, 2019.

HADAIR, Rodrigo. **TSE mantém nulo convênio e dá prazo para Serasa**. 13 de agosto de 2013. Disponível em: https://www.conjur.com.br/2013-ago-13/tse-mantem-nulo-convenio-serasa-prazo-defesa-empresa. Acesso em: 05/08/2018.

HILL, Kashmir. **How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.** Forbes. 16 de fevereiro de 2012. Disponível em: https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=615fcb5e6668. Acesso em 09/03/2021.

KAHNEMAN, Daniel. **Rápido e devagar: duas formas de pensar**. Tradução Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.

KHAN, Lina M. **Amazon's Antitrust Paradox.** The Yale Law Journal. 126:710. 2017

KRUGMAN, Paul. **Arguing with Zombies: economics, politics, and the fight for a better future**. W.W. Norton & Company [eBook], 2020.

LESSIG, L**. Code: Version 2.0.** New York: Basic Book, 2006.

LESSIG, L. The Architecture of Privacy. In: **Vanderbilt Journal of Entertainment Law & Practice**, v. 1, p. 56–65, 1999.

MACHADO, Felipe Nery Rodrigues. **Segurança da Informação: princípios e controle de ameaças**. 1ª Edição. São Paulo: Érica [eBook], 2014.

MAIA, Roberta Mauro Medina. A titularidade de dados pessoais prevista no Art. 17 da LGPD: direito real ou pessoal? In: FRAZÃO, Ana; OLIVA, Milena Donato; TEPEDINO, Gustavo (coord.). **Lei Geral de Proteção de Dados Pessoais: e suas repercussões no Direito brasileiro.** São Paulo: Revista dos Tribunais, 2019. p. 131-156.

MANCINI, Mônica**. Internet das Coisas: História, Conceitos, Aplicações e Desafios**. São Paulo: PMI, 2018. Disponível em https://pmisp.org.br/documents/acervo-arquivos/241-internet-das-coisas-historia-conceitos-aplicacoes-e-desafios/file. Acesso em 12/08/2018.

MAURER, U. Information-theoretic cryptography. In: **Advances in Cryptology—CRYPTO'99**. Springer, 1999. p. 785–785.

MAURER, U. M. The Role of Information Theory in Cryptography. In: **Fourth IMA Conference on Cryptography and Coding**, 1993. p. 49–71.

MAYER-SCHÖNBERGER, V., & CUKIER, K. **Big Data: A revolution that will transform how we live, work, and think**. New York: Houghton Mifflin Harcourt, 2013.

MAZZUCATO, Mariana. **The value of everything: making and taking in the global economy.** Ed. Allen Lane [eBook], 2018.

MEHMOOD, A. ABID MEHMOOD, Abid; NATGUNANATHAN, Iynkaran; XIANG, Yong; HUA, GUANG; GUO, Song. Protection of big data privacy. In: **IEEE Access**, v. 4, p. 1821–1834, 2016.

MENEZES NETO, Elias Jacob. **Surveillance, democracia e direitos humanos: os limites do Estado na era do *Big Data***. Tese de Doutorado. Rio Grande do Sul: Universidade do Vale do Rio dos Sinos - Unisinos. 2016.

MINISTÉRIO DA DEFESA, Exército Brasileiro, Comando de Operações Terrestres. **Manual de campanha. Guerra Cibernética**. 1ª Edição. 2017. Disponível em: https://bdex.eb.mil.br/jspui/bitstream/1/631/3/EB70MC10232.pdf. Acesso em: 11/02/2020.

MOORE, Martin. **Tech Giants and Civic Power**. King's College London. Abril 2016. Disponível em: https://www.kcl.ac.uk/sspp/policy-institute/cmcp/tech-giants-and-civic-power.pdf. Acesso em 14.06.2018.

NARAYANAN, Arvind; SHMATIKOV, Vitaly. **How to break anonymity of the Netflix Prize dataset**. Cornell University. 2007. p. 3. Disponível em: https://goo.gl/RxggOU. Acesso em: 30.08.2018.

NAGLE, Thomas C. Redman Tadhg; SAMMON, David Sammon. **Only 3% of Com-panies' Data Meets Basic Quality Standards**. 2020. Disponível em https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards. Acesso em:20/04/2020.

NARAYANAN, Arvind; SHMATIKOV, Vitaly. Robust De-anonymization of Large Sparse Datasets. In: **Proceedings of the 2008 IEEE Symposium on Security and Privacy**. Washington: IEEE Computer Society, 2008.

NETO, Elias Jacob de Menezes; MORAIS, Jose Luis Bolzan de; BEZERRA, Tiago José de Souza Lima. O projeto de lei de proteção de dados pessoais (PL 5276/2016) no mundo do *big data*: o fenômeno da dataveillance em relação à utilização de metadados e seu impacto nos direitos humanos. In: **Revista Brasileira de Políticas Públicas**. Brasília: v. 7, nº 3, pp. 184-198, 2017.

OHM, Paul. Broken promises of privacy: responding to the surprising failure of anonymization. In: **UCLA Law Review**, v. 57, n. 6, p. 1701–1777, 2009.

O'NEIL, Cath. **Weapons of math destruction. How big data increases inequality and threatens democracy**. New York: Broadway books, 2016.

PAAR, C.; PELZL, J**. Understanding cryptography: a textbook for students and practitioners**. London: Springer International Publishing, 2010.

PASQUALE, Frank. **The black box society. The secret algorithms that control money and information**. Cambridge: Harvard University Press. 2015.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Inteligência Artificial e Direito. Coleção Direito, Racionalidade e Inteligência Artificial. Curitiba: ed. Alteridade, 2019.

PINHO, Frederico António Sá Oliveira. **Anonimização de bases de dados empresariais de acordo com a nova Regulamentação Européia de Proteção de Dados**. Dissertação. Universidade do Porto, Porto, 2017.

PIRAS, Luca; AL-OBEIDALLAH, Mohammed Ghazi; PRAITANO, Andrea; MOURATIDIS, Aggeliki Tsohou, Haralambos; CRESPO, Beatriz Gallego-Nicasio; BERNARD, Jean Baptiste; FIORANI, Marco; MAGKOS, Emmanouil; SANZ, Andres Castillo. DEFeND Architecture: A Privacy by Design Platform for GDPR Compliance. In International Conference on Trust and Privacy in Digital Business. Springer, 2019, 78–93.

RODOTÀ, Stefano. **A vida na sociedade de vigilância: a privacidade hoje**. Rio de Janeiro: Renovar, 2008.

ROSA, Hartmut. Aceleración social: consecuencias éticas y políticas de una sociedad de alta velocidad desincronizada. In: **Persona y Sociedad**. v. 25, n. 1, pp. 9-49, 2011.

RUARO, Regina; REIS, Fernando. Anonimização dos dados como forma de relativização da proteção de informações sigilosas e a atuação fiscalizatória dos Tribunais de Contas. In: **Revista de Estudos e Pesquisas Avançadas do Terceiro Setor**. v.5, n.2, p. 157-187, 2019

RUEDA, Silvia; PANACH, Jose Ignacio; DISTANTE, Damiano. Requirements Elicitation Methods based on Interviews in Comparison: A Family of Experiments. Information and Software Technology. In: **Information and Software Technology**. v.126, October, 2020.

RYAN, Melissa; BRINKLEY, Mark. Navigating privacy in a sea of change: new data protection regulations require thoughtful analysis and incorporation into the organization's governance model. In: **Internal Auditor** v.74, n.,3 p. 61–63. 2017.

SARLET, Ingo Wolfgang. Fundamentos Constitucionais: O Direito Fundamental à proteção de dados. In: DONEDA, Danilo; SARLET, Ingo Wolfgang; MENDES, Laura Schertel Mendes, RODRIGUES JÚNIOR, Otávio Luiz; BIONI, Bruno Ricardo. **Tratado de proteção de dados pessoais**. Rio de Janeiro: Ed. Forense [eBook], 2021.

SCHNEIER, Bruce. **Data and Goliath: The hidden battles to collect you Data and Control Your World**. New York- London: W.W. Norton & Company, 2015.

SCHWAB, Klaus **A quarta revolução industrial. Tradução Daniel Moreira Miranda**. São Paulo: Edipro, 2016.

STF. **MS 27.091/DF**. Min. Luís Roberto Barroso. Data de Julgamento: 03/04/2017. Data de Publicação: DJ 04/04/2017

SIEWERT, Sam B. **Big data in the cloud: data velocity, volume, variety, veracity**. IBM developers Works, 2013.

SOLOVE, Daniel J. **Understanding privacy**. Cambridge, Massachusetts; London, England: Harvard University Press, 2008.

SPINDLER, Gerald; SCHMECHEL, Philipp. Personal data and encryption in the European General Data Protection Regulation. In: **Journal of Intellectual Property, Information Technology and Electronic Commerce Law**, v. 7, 2016.

STIGLITZ, Joseph. **Power, and Profits. Progressive Capitalism for an Age of Discontent**. New York: W.W. Norton Company. 2019

STIGLITZ, Joseph E. **Rewriting the Rules of the American Economy: an agenda for growth and shared prosperity.** The Roosevelt Institute. New York: W.W. Norton Company, 2016.

STUCHE, Maurice; GRUNES, Allen. **Big data and competition policy.** New York: Oxford University Press, 2016

SWEENEY, Latanya. Simple demographics often identify people uniquely. In: **Data Privacy Working Paper**. Carnegie Mellon University, Pittsburgh, v. 3. 2000.

TAURION, Cezar. **Big Data.** Rio de Janeiro: Ed. Brasport [eBook], 2013.

TCU. **Acórdão 1835/2007**. Plenário. TC 025.686/2006-7. Rel.: Min. Marcos Vinicios Vilaça. Data de Julgamento: 5/9/2007.

TCU. **Acórdão 1958/2015.** Plenário. TC 017.090/2015-6. Rel.: Min. Raimundo Carreiro. Data de Julgamento: 5/8/2015.

TCU. **Acórdão 785/2016**. Plenário. TC 005.619/2015-7. Rel.: Min. Raimundo Carreiro. Data de Julgamento: 6/4/2016.

TCU. **Acórdão 1391/2016**. Plenário. TC 017.090/2015-6. Rel.: Min. Walton Alencar Rodrigues. Data de Julgamento: 1/6/2016.

THE ANNOTATED 8 PRINCIPLES OF OPEN GOVERNMENT DATA. Open Gov Data. Disponível em: https://public.resource.org/8_principles.html. Acesso em: 05/08/2018.

THE NEW YORK TIMES. **Cambridge Analytica and Facebook: The Scandal and the Fallout So Far**. 04 de abril de 2018. Disponível em: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html. Acesso em: 02/08/2018.

VAZ. CapCom Asael da Silva. **Proposta de Procedimentos para Anonimização da Origem de Ataques Cibernéticos. Escola de Aperfeiçoamento de Oficiais**. Especialização em Ciências Militares com ênfase em Gestão Organizacional. Rio de Janeiro. 2018.

VELEDA, Raphael ; WALTENBERG, Guilherme. Dados sigilosos: programa do governo federal expõe até agentes secretos. Metrópoles. 8 de fevereiro de 2020. Disponível em: https://www.metropoles.com/brasil/servidor-brasil/dados-sigilosos-programa-do-governo-federal-expoe-ate-agentes-secretos. Acesso em: 11/02/2020.

VIANNA, Marcelo. Um novo "1984"? O projeto renape e as discussões tecnopolíticas no campo da informática brasileira durante os governos militares na década de 1970. In: **Oficina do Historiador EPHIS/PUCRS**, EDIPUCRS, Porto Alegre. p.1448-1471.maio.2014.

W3C BRASIL. **Dados Abertos Governamentais.** Disponível em: http://www.w3c.br/divulgacao/pdf/dados-abertos-governamentais.pdf. Acesso em: 14/08/2018.

WARREN, Samuel D.; BRANDEIS, Louis D. The right to privacy. In: **Harvard Law Review**, v. 4, n. 5. p. 193-220. Dezembro, Disponível em: 1890. Disponível em http://www. cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf. Acesso em: 28 jul. 2018.

WU, Tim**. The attention merchants: The epic scramble to get inside our heads**. New York: Alfred Knopf, 2016.

ZUBOFF, Shoshana. **A Era do Capitalismo da Vigilância: a luta por um futuro humano na nova fronteira de poder**. Tradução: George Schlesinger. 1ª Edição. Ed. Intrínseca [eBook]. 2021.