UNIVERSIDADE DE BRASÍLIA - UnB
FACULDADE DE DIREITO
CURSO DE GRADUAÇÃO EM DIREITO

EDUARDO VASCONCELOS GOYANNA

DADO PESSOAL E DADO ANONIMIZADO E A REGULAMENTAÇÃO NO BRASIL E NA FRANÇA.

BRASÍLIA - DF
2021

EDUARDO VASCONCELOS
GOYANNA

DADO PESSOAL E DADO ANONIMIZADO E A REGULAMENTAÇÃO NO BRASIL E NA FRANÇA.

Monografia apresentada como requisito parcial para a obtenção do grau de Bacharel em Direito pela Faculdade de Direito da Universidade de Brasília.

Orientadora: Professora Doutora Laura Schertel Mendes

BRASÍLIA - DF
2021

EDUARDO VASCONCELOS
GOYANA

DADO PESSOAL E DADO ANONIMIZADO E A REGULAMENTAÇÃO NO BRASIL E NA FRANÇA.

Monografia aprovada como requisito parcial para a obtenção do grau de Bacharel em Direito pela Faculdade de Direito da Universidade de Brasília, pela banca examinadora composta por:

_____

Professora Doutora Laura Schertel Ferreira Mendes
(Orientadora – Presidente)

_____

Professor Mestre Frank Ned Santa Cruz de Oliveira
(Membro da Banca Examinadora)

_____

Professor Mestrando Gabriel Campos Soares da Fonseca
(Membro da Banca Examinadora)

AGRADECIMENTOS

> “Longe do seu ninho, eu andar caminho
> Deixo onde passo meus pés no chão.
> Sou mais um na multidão”[1]

Este trabalho me chamou a atenção à importância em olhar cada pessoa como especial, um ser único portador de riquezas e dificuldades inerentes à vida.

Sendo assim, me sinto feliz em agradecer às pessoas que caminham comigo lado a lado, ombro a ombro, e, dessa maneira, me fazem sentir especial, mesmo quando não mereço tanto afeto e carinho.

Ao meu filho Eduardo, que tenho o orgulho de ter o mesmo nome dele, por me tornar especial sendo pai de pessoa tão única, de carisma e disciplina invejadas no Brasil e na Universidade de Harvard, onde cursa sua graduação.

À minha filha Ana Beatriz, de surpreendentes disciplina, propósito e inteligência. Ana me ensina diariamente que as batalhas da vida valem a pena serem encampadas com a cabeça erguida, mas sem medo de ser forte ao demonstrar suas próprias fraquezas.

Sem dúvidas, tenho orgulho em não ser anônimo para as pessoas que me amam e que luto para retribuir esse amor.

Nesse caso, não tenho nenhum medo de que esse sentimento seja distribuído a todos os cantos do grande meio digital da era da informação.

Obrigado a todos que tiveram a paciência de me acompanhar nessa “louca” jornada que é realizar uma segunda graduação.

Por fim, aos mestres com quem tive o orgulho de conviver nesses 4 anos, pessoas ilustres que não medem esforços em favor do conhecimento genuíno.

Que este trabalho possa ajudar, mesmo sendo uma gota no oceano do mundo jurídico, os operadores do Direito no Brasil que são muito bem representados pelo ilustre advogado e amigo Paulo Henrique Paiva Santos, cuja história de vida o torna único.

Eduardo Vasconcelos Goyanna

Brasília, 28 de outubro de 2021.

[1] Mais um na multidão, música de Erasmo Esteves / Antonio Carlos Santos De Freitas / Marisa De Azevedo Monte

RESUMO

O presente estudo tem como objetivo apresentar o conceito de anonimização de dados com o intento de esclarecer os processos que levam a um dado a ser considerado anônimo e encorajar a sua utilização em maior escala nas mais diversas estratégias de guarda e utilização de dados, facilitando, assim, o diálogo técnico entre os administradores de dados e os operadores do Direito. A anonimização de dados pessoais será desenvolvida a partir de um ponto de vista jurídico, ao apresentar a definição do tema e os riscos envolvidos nos processos de anonimização.
Por conseguinte, o estudo culminará na análise de como o debate sobre a anonimização influenciou a regulamentação na França (sob a luz do marco legal europeu sobre proteção de dados – GDPR) e, no Brasil, com a Lei Geral de Proteção de Dados (LGPD).

Palavras-chave: Proteção de Dados, Anonimização de dados, pseudoanonimização de dados, reidentificação.

ABSTRACT

The present study aims to present the concept of anonymization of data with the intention of clarifying the processes that lead to a data considered anonymous and encouraging its use on a larger scale in the most diverse strategies of data storage and use, thus facilitating the technical dialogue between data administrators and law operators. The anonymization of personal data will be developed from a legal point of view, by presenting the definition of the theme and the risks involved in anonymization processes. Thus, the study will culminate in the analysis of how the debate on anonymization influenced regulation in France (in the light of the European legal framework on data protection (GDPR) and also in Brazil with the General Data Protection Law (LGPD).

Keywords: Data Protection, Data Anonymization, Data Pseudoanonymization, Reidentification.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

1. INTRODUÇÃO

> Invenções recentes e métodos de negócios chamam a atenção para o próximo passo que deve ser dado para a proteção da pessoa, e para garantir ao indivíduo o que o juiz Cooley chama de direito "de estar sozinho". Fotografias instantâneas e empresas jornalísticas invadiram os distritos sagrados da vida privada e doméstica; e inúmeros dispositivos mecânicos ameaçam fazer bem a previsão de que "o que é sussurrado no armário acabará por ser proclamado a partir dos telhados da casa. [2]

> A sociedade da informação imprime uma nova dinâmica e novos desafios para a proteção da pessoa humana, a começar pela monetização dos seus dados pessoais. Tais dados, além de consolidar uma nova forma de prolongamento da pessoa, passam a interferir em sua própria esfera relacional, reclamando, por isso, uma normatização específica que justifica dogmaticamente a autonomia do direito à proteção dos dados pessoais e os desdobramentos da sua tutela jurídica (e.g., direito de acesso e retificação dos dados e oposição a decisões automatizadas, em especial de práticas discriminatórias).[3]

As análises jurídicas mencionadas sugerem, aparentemente, um debate contemporâneo no que tange a preocupação sobre a utilização dos dados pessoais, a proteção da pessoa e sua privacidade. Contudo, os dois textos possuem um lapso temporal de 131 anos.

A primeira discussão jurídica sobre o tema é o ensaio de Samuel Warren e Louis Brandeis – *The right to privacy* (1890) no decorrer da qual passou-se a compreender a privacidade sob ótica de um direito do indivíduo.

No mencionado artigo, os autores apontavam como novas tecnologias máquinas fotográficas que poderiam extrapolar limites e adentrar domínios invioláveis da vida privada e doméstica.

Desde então, o conceito evoluiu, a ponto de ser tutelado como um direito que importa à coletividade também, na medida em que a evolução tecnológica passou a reinventar modelos de negócios cada vez mais baseados em dados dos indivíduos.

Mais de um século depois, a preocupação de Warren e Brandeis é retomada por diversos países como expressa Bioni, em um contexto em que a privacidade e a proteção à pessoa não estão mais em torno de fotos divulgadas em jornais locais, mas

---

[2] WARREN, Samuel; BRENDEIS, Louis. Harvard Law Review, v. 4, n. 5., p.193-220, 1890

[3] BIONI, Bruno Ricardo. Proteção de dados pessoais: A função e os limites do consentimento. Rio de Janeiro:Editora Forense, 2019.

no conceito de uma “nova era”[4], momento em que os dados são o novo petróleo[5] ou a nova fonte de riqueza do século XXI.

São os dados que dizem respeito a todos e a tudo, desde níveis de estoque do grafeno e combustíveis fósseis até as opções alimentares, culturais, ideológicas e partidárias de pequenas cidades, bairros, famílias ou pessoas individualmente selecionadas. A maior parte dos dados encontra-se dentro de ambientes digitais[6,] mas alguns deles servem para dominar esse ambiente digital.

Efetivamente grandes quantidades de dados tornaram-se relevantes, em todos os sentidos, para as tradicionais indústrias e serviços, desde que esses passaram a depender da comunicação digital para seus negócios [7], bem como para o Estado, para a sociedade organizada e para grupos políticos. Pode-se mesmo falar em dependência das empresas e dos prestadores de serviços, em relação aos dados, seja para sobreviverem, para prosperarem ou para minimamente concorrerem de maneira eficaz e propagarem seus produtos, serviços e marcas.

De igual modo, é aplicável a todos demais envolvidos e interessados. Há nítida posição de vantagem que a obtenção dos dados acertados pode significar em qualquer ambiente, desde o econômico até o político, ideológico ou religioso. A nova luta no séc. XXI situa-se, assim, no campo dos dados.

Exatamente por isso é que se compreende a razão pela qual, neste século, a primazia pelo domínio sobre alguns dos chamados bens de produção tradicionais (terras, recursos naturais, máquinas e plantas industriais) está sendo pressionada e se desloca para o domínio sobre dados. Não se fala do fim das velhas indústrias ou serviços. Como bem coloca Zuboff, “[s]e o poder já foi uma vez identificado com a propriedade dos meios de produção, agora ele é identificado com a propriedade dos meios de modificação comportamental”.[8.]

Nesse sentido, percebemos nitidamente que essa revolução nos legou uma espécie de digitalização da vida em geral e, concomitantemente, a monetização.

Desses dados, mas sobretudo a fragilização do sujeito, viabilizada pelas

4 Klaus Schwab, Shaping the future of the fourth industrial revolution. New York: Currency, 2018, p. 69
5 Ibidem, p. 222
6 É a tese de Nick Srnicek (Plataform capitalism. Cambridge: Polity, 2017, p. 39)
7 Nick Srnicek; Plataform capitalism. Cambridge: Polity, 2017, p. 39
8 Como S. Zuboff, Big *other*: capitalismo de vigilância e perspectivas para uma civilização de informação. In: Fernanda Bruno e outros, Tecnopolíticas da vigilância. São Paulo: Boitempo, 2018, p. 40.

tecnologias que permitem colher as informações de todos nós, armazenar, refinar, minerar (como dizem alguns[9]) e utilizar com finalidades múltiplas, como o uso ideológico, econômico etc., inclusive com a possibilidade de reutilização infinita e redirecionamento aleatório desses usos. A modificação comportamental da sociedade é, certamente, o objetivo último. Suscetibilidade e vulnerabilidade passariam a ser princípios permanentes e supremos, para usar uma linguagem jurídica corrente, de fácil compreensão do que pretendo descrever.

Para bem gerir todos esses dados, é importante que se possa conectar os inúmeros dados colhidos, sobretudo na internet, mas não apenas por ela, ao R.G. de pessoas reais e bem determinadas. Aos dados basilares assim amealhados e devidamente refinados se poderá somar toda uma série de novos dados do cidadão-identificado, como sua rotina diária familiar, seus costumes, suas opções gerais de compra, de frequência, suas viagens, movimentações financeiras etc.

Nesse contexto, surge o conceito de anonimização de dados.

A anonimização de dados consiste na remoção ou no embaralhamento de dados pessoais de um repositório contenha informações sobre pessoas, de maneira que a pessoa natural não possa mais ser identificada ou identificável. É um processo geralmente utilizado quando o mantenedor de dados pessoais necessita compartilhar um com um terceiro que não deveria ter acesso aos dados pessoais que estão contidos no banco.

Ela é apresentada como uma solução pouco complexa que une o melhor de dois mundos: permitir que analistas possam encontrar informação útil ao mesmo tempo que usuários tenham sua privacidade preservada, atingindo um equilíbrio ideal entre os interesses empresarial e o privado.

Contudo, conforme bem marca o professor Paul Ohm, existe uma distorção gerada pela crença na eficácia completa da anonimização que produziu um desequilíbrio entre a preservação da privacidade e o interesse público. O professor, em seu ensaio, afirma: "dados podem ser úteis ou perfeitamente anônimos, mas nunca podem ser ambas as coisas".[10] Essa controvérsia acadêmica sobre a eficácia

---

[9] Ibidem.

[10] OHM, P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, Vol. 57, U of Colorado Law Legal Studies Research Paper No. 9-12. Boulder, EUA, 2010.

da anonimização de dados influenciou o debate do tema na regulamentação de diversos diplomas legais no mundo, notadamente na Europa (mais especificamente a França), nos Estados Unidos e no Brasil.

A causa principal da discussão está na tese de que, como o processo de anonimização pode ser revertido – a reidentificação -, dados anonimizados devem ser considerados dados pessoais e, portanto, devem estar sujeitos aos diversos regulamentos sobre proteção de dados pessoais. O professor Antonio Menezes, ao defender o critério da razoabilidade para utilização de técnicas de anonimização, afirma existir uma "impossibilidade fáctica de garantir a anonimidade absoluta dos dados recolhidos".[11]

O objetivo deste trabalho é apresentar as técnicas de anonimização de dados pessoais mais utilizadas internacionalmente, com o intento de esclarecer os processos que levam a um dado ser considerado anônimo e encorajar a reflexão sobre sua utilização em maior escala nas mais diversas estratégias de guarda e utilização de dados, facilitando, assim, o diálogo técnico entre os administradores de dados e os operadores do Direito.

Para tanto, as principais técnicas de anonimização de dados pessoais tais como a supressão, substituição, generalização, perturbação e agregação serão explanadas, assim como os dados que diferenciam os processos da pseudonimização, com foco na possível reidentificação dos dados.

A anonimização de dados pessoais será desenvolvida a partir de um ponto de vista jurídico, ao apresentar a definição do tema e os riscos envolvidos em seus processos.

Assim, o estudo culminará na análise de como o debate sobre a anonimização influenciou a regulamentação na França pela CNIL[12] (sob a luz do marco legal europeu sobre proteção de dados (GPDR)[13] e no Brasil com a Lei Geral de Proteção de Dados

[11] MENEZES, Barreto. Dados pessoais: conceito, extensão e limites, LLM, book Revista de Direito Civil 2, 2018

[12] CNIL (*Commission Nationale de l'Informatique et des Libertés* / Comissão Nacional de Informática e de liberdades) é considerada a autoridade oficial em proteção de dados na França.

[13] GDPR é uma sigla em inglês que significa *General Data Protection Regulation*. Ou, em português, Regulamentação Geral de Proteção de Dados. Trata-se de uma nova lei de proteção de dados que entrou em vigor na Europa. Proposta em 2012, a nova regulamentação foi aprovada pela União Europeia em 2016 e passou a valer oficialmente apenas em 2018.

Pessoais(LGPD) [14].

2. DADOS PESSOAIS E DADOS ANONIMZADOS – CONCEITUAÇÃO

O conceito de dado pessoal é essencial para compreensão acerca do alcance de leis e regulamentações sobre proteção de dados. Muito embora o conceito possa ser criticado por teóricos como Paul Ohm, que sustenta o abandono do conceito de "informação pessoal identificável" como pilar da proteção jurídica da privacidade, ele ainda é amplamente destacado no escopo do direito à proteção de dados pessoais.[15]

A conceituação de dado pessoal é elevada ao nível de proteção da própria pessoa natural, como descreve Laura Schertel Mendes:

> Tendo em vista que as informações pessoais constituem-se em intermediários entre a pessoa e a sociedade, a personalidade de um indivíduo pode ser gravemente violada com a inadequada divulgação e utilização de informações armazenadas a seu respeito. Por se constituírem em uma parcela da personalidade da pessoa, os dados merecem tutela jurídica, de modo a assegurar a sua liberdade e igualdade.[16]

Nesse sentido, tem-se que:

> O tratamento de dados pessoais, em particular por processos automatizados, é, no entanto, uma atividade de risco. Risco que se concretiza na possibilidade de exposição e utilização indevida ou abusiva de dados pessoais, na eventualidade desses dados não serem corretos e representarem erroneamente seu titular, em sua utilização por terceiros sem o conhecimento deste, somente para citar algumas hipóteses reais.
> Daí resulta ser necessária a instituição de mecanismos que possibilitem à pessoa deter conhecimento e controle sobre seus próprios dados – que, no fundo, são expressão direta de sua própria personalidade. Por este motivo, a proteção de dados pessoais é considerada em diversos ordenamentos jurídicos como um instrumento essencial para a proteção da pessoa humana e como um direito fundamental. [17]

2.1 BREVE ANÁLISE SOBRE OS DADOS RELATIVOS À PESSOA

O direito da proteção de dados considera que toda informação é considerada

---

[14] BRASIL, Lei 13.709 de 14 de agosto de 2018. Lei Geral de Proteção de Dados.
[15] OHM, Paul. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Review, n. 57, p. 1742, 2010
[16] Mendes, Laura – Revista dos Tribunais, Caderno Especial – novembro de 2019, p. 36
[17] DONEDA, 2011, p. 92.

relevante. Não deve haver informação pessoal não merecedora de respaldo jurídico, por mais que pareça insignificante ou mesmo fútil.[18]. Nesse ponto, vale ressaltar o sentido amplo de dados pessoais que vai além do conceito de direito da personalidade, muito menos se confunde privacidade com dados pessoais. Nesse sentido, o Tribunal de Justiça da União Europeia (TJUE) estende o conceito para além da vida pessoal, incluindo, assim, a vida social e a profissional:

> O fato de essa informação se inscrever no contexto de uma atividade profissional não lhe pode retirar a qualificação de conjunto de dados pessoais.[19]

Assim, a informação pode abranger elementos identificativos da pessoa: nome, data de nascimento, cadastros de pessoa física; características físicas: gênero, altura, peso, cor dos olhos ou do cabelo; considerações íntimas: crenças, opiniões, desejos, posições políticas ou religiosas; características profissionais e acadêmicas: títulos e graus ou estatutos profissionais e laborais; dados patrimoniais: conta bancária, imóveis, automóveis, investimentos e direitos de propriedade. Assim também, em sentido amplo, dados genéticos; dados biométricos; dados relativos à saúde; dados que revelem a origem racial ou étnica, as opiniões políticas, as convicções religiosas ou filosóficas, a filiação sindical, relativos à vida sexual ou orientação sexual; ou dados relacionados com condenações penais e infrações.[20] Assim, são infindáveis as modalidades passíveis de serem concebidas.

O TJUE amplifica ainda mais o entendimento do conceito de dados pessoais para imagens gravadas em vídeo, atas de reuniões contendo os nomes dos seus participantes, registros dos tempos de trabalhos de trabalhadores de uma empresa, observações de peritos sobre determinado projeto divulgadas em sítio da internet.[21]

As abordagens para conceituação do tema observam os conceitos restrito e amplo, ou reducionista e expansionista, conforme Solove e Schwartz[22].

---

[18] MENEZES, Barreto. Dados pessoais: conceito, extensão e limites, LLM, book Revista de Direito Civil 2, 2018

19 TJUE 16-nov.-2015, proc. C-615/13 P (ClienthEart v PAN Europe), 30.

[20] Artigo 16º, GDPR (General Data Protection Regulation)

[21] Menezes, Barreto. Dados pessoais: conceito, extensão e limites, LLM, book Revista de Direito Civil 2, 2018, p. 303

[22] SCHWARTZ, Paul M.; SOLOVE, Daniel J. The PII problem: privacy and a new con- cept of personally identifiable information. New York University Law Review, v. 86, p.
1871-1877, dec. 2011. A respeito do tema, vide também BIONI, Bruno. Xeque-mate: o tripé de proteção de dados pessoais no xadrez das iniciativas legislativas no Brasil. São Paulo: GPOPAI, 2015.

Na abordagem restrita ou reducionista, dado pessoal é aquele que possui fatos sobre uma pessoa que possa ser identificada e além, possa ser individualizada em meio a um certo grupo, comunidade ou coletividade. Dessa maneira, existem identificadores diretos ou indiretos. A identificação direta tem como exemplo clássico o nome e sobrenome da pessoa[23].

Contudo nem sempre identificadores diretos são suficientes para diferenciação da pessoa do seu grupo ou comunidade. Nesse ponto, surge a necessidade dos indiretos: CPF, telefone, nacionalidade, Código de endereçamento postal (CEP) da residência ou até características fototípicas. A combinação dos identificadores segue o fenômeno das combinações únicas (conjunto de itens que identificam unicamente um objeto).[24]

Em contraponto, o conceito amplo vai além do conceito da pessoa identificada, alçando o significado para identificável: dados que tenham o potencial de conduzir à individualização da pessoa também se enquadram como dados pessoais. Ainda que o agente responsável pelo tratamento dos dados não possa identificar com precisão a pessoa natural a quem se referem as informações processadas, com algum esforço ele, ou terceiro, pode se valer de meios disponíveis para a obtenção dos dados adicionais aptos a fazê-lo. São informações que, sozinhas, não identificam ninguém, porém, no momento que são agrupadas com outros dados, permitem identificar uma pessoa. O dado identificável é como um quebra-cabeça, pois somente fará sentido juntando algumas peças. Ou seja, quanto mais informações combinadas de uma pessoa, mais fácil de ser identificada e maior risco associado à proteção de dados: número de cartão de crédito, IP do computador em que a pessoa acessou um serviço digital, nome da empresa, placa de Carro em nome da pessoa e, até mesmo, *cookies*[25] para rastreamento de ações de pessoas na internet.

A análise do potencial de identificação (ou identificabilidade) deve seguir alguns critérios como: o critério dos meios suscetíveis ser razoavelmente utilizados (limitados

---

[23] ARTICLE 29 DATA PROTECTION WORKING PARTY. Op. cit., p. 13.

24 ibidem

[25] Cookies são programas que armazenam informações sobre o que é feito, acessado e, sobretudo, o comportamento do usuário na internet. Nesse sentido, a CNIL (autoridade de dados francesa) publicou redomendação, em 23 de julho de 2019, para utilização do cookies para que haja conformidade com a regulação eruropeia de proteção de dados (GDPR) que podem ser acessadas em https://www.cnil.fr/en/cookies-and-other-tracking-devices-cnil-publishes-new-guidelines. Acesso em 10 de setembro em 2021)

por fatores como custo e tempo para identificação), a tecnologia disponível e ainda os riscos de falhas técnicas e de descumprimento dos deveres de confidencialidade do próprio dado. [26]

Assim, a definição de dados identificáveis, a depender do contexto em que estão inseridos, das tecnologias utilizadas no momento de sua existência e, principalmente da inteligência artificial e aprendizados de máquina faz da caracterização de um dado pessoal algo bem dinâmico, o que bem aponta Diego Machado:

> Esse caráter maleável do conceito, se de um lado contribui para que uma lei de proteção de dados acompanhe as transformações tecnológicas e socioeconômicas, por outro pode dar ensejo a insegurança em alguma medida, uma vez que os avanços das *data-driven technologies* se desenvolvem em passo acelerado e poderiam conferir caráter pessoal a dado antes considerado anônimo, devido à disponibilização de novas técnicas capazes de realizar essa reidentificação[27]

A lógica da identificação por dados pessoais segue a mesma lógica do fato jurídico. Fato jurídico é aquele que tem repercussão para o Direito, produzindo efeitos jurídicos, criando, modificando, extinguindo e defendendo relações jurídicas. Um fato sem tal adjetivação não tem repercussão e, portanto, não produz efeitos jurídicos, como é o exemplo da queda de uma folha seca da árvore que em nada tem relevância para a ciência jurídica. Assim também, não é qualquer dado que carregaria a repercussão jurídica, mas exclusivamente aquele que possa atrair o qualificador pessoal, na perspectiva expansionista ou reducionista[28].

A tabela 1 faz um resumo da diferenciação entre os conceitos explanados.

---

[26] Os critérios mais razoável de serem utilizados foram abordados pela diretiva 95/46/EC do Parlamento Europeu

[27] Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados, Diego Machado, – Revista dos Tribunais, Caderno Especial – novembro de 2018, p. - 108

[28] BIONI, Bruno Ricardo. Proteção de dados pessoais: A função e os limites do consentimento. Gen, Editora Forense, 2019.

Tabela 1 – Definição de dados pessoais

Amplo ou expansionista	Restrita ou reducionista
A pessoa é identificável	A pessoa é identificada
A pessoa é indeterminada ou inespecífica	A pessoa é específica ou determinada
O Vínculo é mediato, indireto, impreciso ou inexato	Vínculo é imediato, direto, preciso ou exato
Há a ideia de alargamento da qualificação como dado pessoal	Há a ideia de retração da qualificação do dado como pessoal

Fonte: Bioni, 2019

A tabela 2 exemplifica a conexão de vários dados relativos a uma pessoa no intuito de identificar um indivíduo e diferenciá-lo de seu grupo e comunidade.

Tabela 2 – Tabela de atributos de alunos de uma faculdade

CPF	Nome	Idade	Sexo	CEP	Renda	Data de	Curso
185.302.491-00	Ana Silva	18	M	13090-718	850,00	11/10/16	Direito
71.132.131-20	Ana Silva	20	M	13090-590	1200,00	12/10/16	Medicina
186.426.081-53	Ana Silva	23	F	13080-379	3200,00	15/10/11	Odontologia
098.705.201-20	Jose Aldo	25	F	13080-190	1900,00	18/10/16	Psicologia
297.099.701-00	Jose Aldo	67	F	13080-340	900,00	21/10/15	Direito
041.565.283-91	Bruno Carvalho	71	F	13080-350	1000,00	27/10/14	Medicina
258.842.631-04	Bruno Carvalho	45	M	13080-714	2500,00	02/11/17	Direito
275.503.971-04	Bruno Carvalho	52	M	13080-360	1250,00	07/11/14	Economia
098.705.201-20	Ataulfo Silva	21	F	13080-190	1900,00	07/11/19	Engenharia
610.384.661-72	Mateus Ferreira	33	M	13080-691	1500,00	07/11/11	Direito
296.416.591-20	Virgílio Alves	70	M	13080-369	900,00	09/11/14	Medicina
271.132.131-20	Rafael Silva	27	M	13090-590	1200,00	17/11/16	Odontologia

Fonte: elaborada pelo autor

O professor Bioni apresenta, de maneira didática, um critério de análise das perspectivas que pode ser utilizado para a tabela exposta.

A tabela é composta por uma lista de atributos das pessoas (CPF, Nome, Idade, Sexo, CEP, Renda Mensal, Data de Ingresso e Curso). O nome é um atributo direto de uma pessoa, contudo percebe-se que há homônimos na tabela, o que impede a individualização de uma parcela das pessoas inseridas. É preciso, então, na

perspectiva reducionista, a utilização do atributo CPF.

Com tal associação as pessoas tornam-se identificadas de maneira precisa, exata e inequívoca. Sendo assim, somente com o atributo CPF, os reducionistas considerariam a tabela acima como de atributos possíveis de identificação de uma pessoa natural.

A hipótese reducionista para o exemplo citado fica clara.

Contudo, caso houvesse a eliminação do atributo CPF, não haveria mais uma identificação única das pessoas da tabela, gerando, por exemplo, dúvidas sobre qual dos “Brunos” cursaria Direito, ou qual Ana cursaria Medicina. As informações não estão relacionadas com pessoas identificadas. Nesse sentido, a perspectiva expansionista vai agregar outros atributos como o CEP, por exemplo, para tentar eliminar a incerteza gerada por um grupo de pessoas que possuam o mesmo nome.

Nesse caso, em última análise, há somente um potencial de individualizar as pessoas, tendo em vista a ausência de um atributo que indique exclusivamente, de maneira direta e específica uma pessoa natural.

Portanto, verificar se um dado pode ser determinado como pessoal depende de uma análise contextual que depende de qual tipo de informação pode ser extraída de um conjunto ou uma tabela de dados. Essa análise circunstanciada pode ser mais dura ou mais flexível. Para os reducionistas, somente o primeiro contexto (com a presença do identificador CPF) seria dado pessoal. Para os expansionistas, os dois contextos seriam abarcados pelo conceito de dado pessoal.

2.2 BREVE ANÁLISE SOBRE O CONCEITO DE ANONIMIZAÇÃO E PSEUDOANONIMIZAÇÃO

Uma vez detalhado o conceito de dados pessoais e as abordagens da doutrina, surge o campo para análise do que seja o dado anonimizado.

O dado anônimo é a antítese do conceito de dado pessoal.[29]. No cerne do conceito está a impossibilidade de revelar, identificar, individualizar uma pessoa em relação a sua comunidade ou grupo de convivência. A própria terminologia de anônimo

[29] BIONI, Bruno Ricardo. Proteção de dados pessoais: A função e os limites do consentimento. Gen, Editora Forense, 2019.

traz esse contraponto ao dado pessoal: aquele que não tem nome e nem rosto[30]; algo ou alguém desconhecido[31]; "*Anonyme: Se dit de quelqu'un dont on ignore"* (Anônimo: diz-se daquele que se ignora)[32]; "*Anonymous: made or done by someone whose name is not known or not made public*"[33] (Anônimo: feito ou realizado por alguém cujo nome não é conhecido ou não foi tornado público).

Para Doneda[34], no contexto histórico, o marco da sociedade industrial altera o equilíbrio entre a vida rural e urbana e, entre todas suas consequências, na cidade, as relações tendem a ser impessoais e a sobrevivência, despida de várias formas de associativismo tipicamente rurais, tende a ter caráter mais individualista.

No ambiente rural, o contorno das relações tendia a ser diverso. O conceito de dados relativos a um só indivíduo era inexistente em sociedades desse tipo, como é exemplo a américa colonial da seguinte descrição:

> Em casas nas quais famílias e hóspedes costumavam dividir leitos por causa do frio, e nas quais membros da família se moviam livremente pelos seus cômodos para acender velas e lanternas, quando as pessoas conheciam os assuntos das demais, as oportunidades de se obter isolamento e anonimato, no sentido moderno, eram muito poucas. [35]

De fato, foram as grandes mudanças culturais trazidas pela revolução industrial, a diminuição dos membros das famílias e comunidades e, principalmente, os meios materiais à disposição que suscitaram pensamentos e desejos por privacidade e, mais tarde, a consideração sobre direitos da própria pessoa.

> Ao individualismo somam-se, portanto, os meios materiais à disposição, em um primeiro momento da burguesia, e que foram posteriormente massificados, meios esses que providenciavam de diversas maneiras a delimitação de espaços entre os ocupantes de uma mesma casa. Tais meios eram de regra fornecidos pela tecnologia, como a construção de habitações coletivas, a difusão da eletrificação e de toda a infraestrutura doméstica (hidráulica, aquecimento etc.), acompanhados de uma diminuição do número de membros para uma família média – o que implica diretamente que menos pessoas teriam necessidade de dividir seu quarto com outras.[36]

A nova revolução, a da era digital, trouxe um novo olhar sobre o dado pessoal,

---

[30] HOUAISS, Antônio; VILLAR, Mauro de Salles. Op. cit., p. 140-141.

[31] DICIO ONLINE, disponível em: www.dicio.com.br. Acesso em: 13 de setembro de 2021.

[32] DICIONÁRIO LARROUSSE, disponível em : www.larousse.fr, acesso em: 13 de setembro de 2021

[33] DICIONÁRIO DE CAMBRIDGE. Disponível em www.dictionary.cambridge.org , acesso em 13 de setembro de 2021

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

[35] Robert Ellis Smith. Ben Franklin's web site. Providence: Privacy Journal, 2000, p. 19.

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

privacidade e a consequente necessidade de haver dado anonimizado. Assim como nas antigas comunidades rurais, muito antes da Revolução Industrial, momento em que a noção de indivíduo era relativizada pelo estilo de vida, a grande rede de conexão que a internet proporciona quebra os vínculos de individualização da sociedade pós-industrial. Ironicamente, as informações pessoais estão, agora, espalhadas em uma incrível rede de informação não somente acessadas pelos membros de uma mesma comunidade, mas pelo mundo inteiro.

Essa quebra, graças ao poder de processamento da tecnologia atual, o volume de dados pode ser utilizado para um número variado de aplicações: pesquisa de mercado, levantamentos sociológicos, desenvolvimento de tratamentos médicos, melhoria de políticas públicas, entre outras.

2.2.1 DESIDENTIFICAÇÃO POR ANONIMIZAÇÃO

Ao se referir a uma pessoa indeterminada, o dado pode ser útil para várias finalidades às quais a informação tem valor a uma determinada coletividade ou um corte específico de indivíduos, com a vantagem de as pessoas envolvidas não serem nominadas. Um exemplo é o fluxo telefônico de uma determinada empresa de telecomunicações, sem que haja possibilidade de identificar as pessoas ou clientes envolvidos na análise do fluxo.[37]

Entretanto, o potencial dano proporcionado pelo acesso a tais informações também é grande. Essas informações, quando relacionadas às pessoas a quem elas se referem, podem ser acessadas por partes não autorizadas, fato que por si só já constitui uma violação da expectativa de privacidade do cidadão, podem ser utilizadas para fins de repressão, discriminação, retaliação ou chantagem.

O que se busca é o equilíbrio entre o benefício acadêmico, econômico, social com a análise dados pessoais de toda uma comunidade e a proteção dos valores íntimos da pessoa e do seu direito de não ter suas informações pessoais tratadas de forma indiscriminada e não consentida.

Esse cenário somente é possível com um processo de desidentificação de dados.

---

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

Esse processo é um gênero que cujas espécies são a pseudoanonimização e a anonimização[38].

A anonimização de dados é o processo que permite tornar o dado pessoal em anônimo. O procedimento tem por objeto não ser possível a reversão da identificação do titular dos dados pessoais tendo como base as informações anonimizadas. O esforço é realizado para que haja a quebra do vínculo entre o dado e seu titular[39]. Em última análise, permitir o tratamento dos dados sem as restrições impostas pelas diversas regulações sobre dados de pessoa (especialmente as sanções em caso de vazamento dos próprios dados).

Assim, resta ser uma técnica que exclui ou altera todas as informações que possam identificar direta ou indiretamente uma pessoa, resultando em dados anonimizados impossíveis de serem associados a nenhum indivíduo.

É importante destacar que o resultado do processo de anonimização é sempre dependente do contexto em que se insere, uma vez que uma técnica utilizada em determinado momento pode, em poucos anos, tornar-se ineficiente, permitindo que os dados sejam reidentificados e, caso isso ocorra, os dados anonimizados passarão a ser considerados dados pessoais novamente. Assim, a qualidade de uma técnica de anonimização leva em consideração o custo e o tempo necessários para que o processo seja revertido de acordo com as técnicas disponíveis em determinado momento do desenvolvimento tecnológico.

Portanto, sabiamente traduz o professor Bioni ao abordar o conceito dinâmico da anonimização de dados:

> Com maior ou menor grau de intensidade nota-se um método cujo mote é gerenciar circunstancialmente a identificabilidade de uma base de dados. As características de cada dado e a percepção de eles estarem inseridos em uma gama de informações devem orientar tal análise.
> Por isso, não há um único método ou uma combinação perfeita ex ante para parametrizar o processo de anonimização, devendo-se analisar contextualmente como este deve ser empreendido para que os titulares dos dados anonimizados não sejam reidentificados, nem mesmo por quem procedeu à sua anonimização.[40]

---

[38] “[Privacy is the] protection from being brought to the attention of others”. In: GAVISON, Ruth. Privacy and the Limits of Law. The Yale Law Journal, Vol. 89, No. 3 (Jan, 1980), pp. 421-471. The Yale Law Journal Company,1980. Disponível em: <http://www.jstor.org/stable/795891>. Acesso em: 15 jan. 2021

[39] DONEDA,2006, p. 44

[40] BIONI, Bruno Ricardo. Proteção de dados pessoais: A função e os limites do consentimento. Gen, Editora Forense, 2019

2.2.2 DESIDENTIFICAÇÃO POR PSEUDOANONIMIZAÇÃO

As regulamentações sobre dados pessoais tratadas neste documento têm como um dos eixos principais a identificação dos dados pessoais de alguma forma. Foi explicitado o conceito de anonimização que não deve ser confundido com pseudoanonimização.

Enquanto a anonimização é definida pela quebra total do vínculo original com o dado pessoal, uma técnica aplicada de tal forma aos dados pessoais que sua desidentificação se torna irreversível[41], a pseudoanonimização permite a recuperação dos dados originais por meio da reversão do processo utilizado para geração.

A pseudoanonimização é um meio caminho para anonimização no sentido em que as informações adicionais que permitiriam a identificação do titular são mantidas em separado pelos agentes de tratamento, que podem, assim, reidentificar os dados se fizerem uso dessa informação. Contudo, caso excluam essas informações adicionais, os agentes não mais poderão efetuar a reidentificação “por meios próprios”, caracterizando, assim, uma técnica de anonimização. [42]

Na pseudoanonimização, o dado não pode mais ser atribuído a um específico indivíduo sem a utilização de uma informação adicional que seja chave para a reidentificação. Surge o conceito de entropia da informação que é o uso de uma informação auxiliar para a reversão do processo de anonimização, normalmente em posse do agente de tratamento.[43]

Bioni exemplifica o conceito ao citar uma rede de lojas varejistas:

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

[41] ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 4/2007 on the concept of personal data. Bruxelas, 20077
[42] BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioteca digital Jurídica do Superior Tribunal de Justiça, 2019
[43] ibidem

> mantidas separadamente, para reverter o processo de anonimização. Ou seja, ele possui meios próprios para transmudar um dado aparentemente anonimizado em um dado pessoal, o que é revelado com base em uma análise subjetiva focada na sua própria capacidade de entropia de informação” [44]

Em um exemplo prático utilizando os dados descritos na tabela 2, a técnica consistiria em substituir o CPF por identificadores únicos e aleatórios, de maneira que, somente o agente de tratamento dos dados possuiria a relação entre os CPF e o novo identificador.

Tabela 3 - Tabela de atributos de alunos de uma faculdade pseudoanonimizada

CPF	Nome	Idade	Sexo	CEP	Renda Mensal	Data de	Curso
L8128210	Ana Silva	18	M	13090-718	850,00	11/10/16	Direito
L8128232	Ana Silva	20	M	13090-590	1200,00	12/10/16	Medicina
L8128265	Ana Silva	23	F	13080-379	3200,00	15/10/11	Odontologia
L8128278	Jose Aldo	25	F	13080-190	1900,00	18/10/16	Psicologia
L8128243	Jose Aldo	67	F	13080-340	900,00	21/10/16	Direito
L8128267	Bruno Carvalho	71	F	13080-350	1000,00	27/10/14	Medicina
L8128209	Bruno Carvalho	45	M	13080-714	2500,00	02/11/14	Direito
L8128214	Bruno Carvalho	52	M	13080-360	1250,00	07/11/14	Economia
L8128267	Ataulfo Silva	21	F	13080-190	1900,00	07/11/14	Engenharia
L8128243	Mateus Ferreira	33	M	13080-691	1500,00	07/11/11	Direito
L8128278	Virgílio Alves	70	M	13080-369	900,00	09/11/14	Medicina
L8128221	Rafael Silva	27	M	13090-590	1200,00	17/11/16	Odontologia

Fonte: elaborada pelo autor

No exemplo em questão, a coluna CPF foi totalmente substituída por identificadores aleatórios, o que não permite mais a identificação direta das pessoas e nem sua individualização.

Contudo, o agente de tratamento possuirá uma tabela guardada em ambiente separado consistindo na relação dos identificadores com o próprio CPF (dado que identifica diretamente a pessoa).

[44] Idem, página 194

Tabela 4 – Relação entre os dados que torna possível a reidentificação das pessoas

Identificador	CPF
L8128210	185.302.491-00
L8128232	271.132.131-20
L8128265	186.426.081-53
L8128278	098.705.201-20
L8128243	297.099.701-00
L8128267	041.565.283-91
L8128209	258.842.631-04
L8128214	275.503.971-04
L8128267	098.705.201-20
L8128243	610.384.661-72
L8128278	296.416.591-20
L8128221	271.132.131-20

Fonte: elaborada pelo autor

Ora, fica claro que qualquer agente ou entidade que tiver posse dos dados da tabela 4 terá o poder de reidentificar o dado tido como anônimo, com o auxílio da entropia dos dados. Em resumo, a posse das relações de CPF com os identificadores aleatórios permitirá o dado ser totalmente pessoal e identificável.

2.2.3 ABRANGÊNCIA DA DESIDENTIFICAÇÃO

A correta utilização de técnicas de anonimização e mesmo de pseudoanonimização traz inúmeros benefícios para várias áreas sociais, econômicas e de saúde:

- Relatórios estatísticos ou de inteligência para a melhoria, medida da audiência ou o desempenho de produtos;
- Identificação do comportamento dos indivíduos (sem sua própria identificação) para auxílio ao desenvolvimento de novos produtos;
- Dados sintéticos (artificialmente criados por anonimização), tendo como fonte dados reais, podem ajudar no treinamento de algoritmos. Tais dados são anonimizados, pois os fatores identificativos são retirados e

substituídos por dados sintéticos;

- Compartilhamento entre áreas de negócios de uma mesma empresa para utilização em procedimentos internos para atendimento ao cliente;
- No sistema judiciário, a divulgação de teses debatidas, de métricas e estatísticas voltadas para melhoria de processos e mesmo utilização de Inteligência Artificial;
- Utilização por pesquisadores de saúde para validação de novas técnicas e tratamentos com a utilização de dados altamente sigilosos.

Caso sejam adequadamente implementadas, os resultados positivos trazidos pelas técnicas de anonimização são inúmeros. A sensibilidade dos dados trabalhados é tão importante que se deve sempre buscar o ponto de equilíbrio ideal entre a utilidade e segurança esperada pelos agentes de tratamento e a privacidade dos titulares dos dados.

O equilíbrio é sempre tênue pois um perfeito esquema que anonimize o dado pode simplesmente tornar o dado sem significado e, portanto, sem utilidade alguma para a entidade que busca trabalhar economicamente, socialmente ou academicamente a informação.

Para que seja eficiente, o processo de anonimização precisa preservar os aspectos semânticos e lógicos dos dados originais, a fim de que garanta a regularidade da posterior análise exploratória.

É fato que todo esforço deve ser eivado para que a intimidade e as informações ligadas a identificação sejam preservadas conforme a vontade do seu titular. Esse direito não deve ser relativizado e sempre buscado com todas as alternativas possíveis.

Assim, a busca pela proteção do dado que identifique a pessoa deve ser encampada por todos os agentes de tratamento em todo o ciclo de vida do dado que possa identificar um indivíduo.

Esse zelo deve existir muito antes do funcionamento dos sistemas, aplicações ou funcionalidades diversas, deve estar já na concepção da arquitetura de desenvolvimento de software ou projetos, abordagem chamada de *privacy by design* pelo qual o cuidado com o dado pessoal deve ser pensado passo a passo para que a

privacidade seja zelada durante todo o ciclo do tratamento[45].

Em acréscimo, a perspectiva *privacy by default,* é um conceito bem esclarecido por Frazão:

> [...] a diz respeito ao fato de que, ao lançar qualquer produto ou serviço ao público, as regras mais protetivas de tutela da privacidade devem ser aplicadas, sem que se exija do usuário qualquer iniciativa para tal propósito. Daí a importância das tecnologias reforçadoras da privacidade ou *PETs – Privacy Enhancing Technologies*, que são exemplos de como a tecnologia pode ser utilizada em prol da privacidade, o que se projeta desde a configuração de equipamentos eletrônicos, como *tablets* e *smart phones*, até mesmo à estruturação de produtos e serviços. [46]

A preocupação de um perigoso *trade-off* entre os direitos dos titulares de dados e as eficiências econômicas trazidas pelos mesmos passa a ser o cerne para a utilização adequada das técnicas de anonimização de dados dependentes do contexto do tratamento, da utilidade esperada e da sensibilidade dos dados tratados. Por tal razão, é preciso atentar para os riscos de implementação intrínsecos a tais técnicas.

3. ANONIMIZAÇÃO: APLICAÇÕES E MÉTODOS

Para uma segura guarda de dados sem identificação pessoal é necessário um diálogo permanente entre os agentes de tratamentos de dados, profissionais das ciências de privacidade (nesses inclusos os especialistas em tecnologia da informação) e, mais atualmente e não menos importante, com o advento de diversas leis para regulação do tratamento de dados pessoais, dos juristas.

O que se percebe é uma assimetria nesse diálogo que pode gerar desde falta de informação ao definir a abrangência da regulação até o aumento desnecessário no custo de empresas e órgãos governamentais, visto que as companhias concentram seus esforços na adequação abstrata de contratos e redação de políticas de privacidade, mas esquecem de atentar para a proteção técnica de suas bases de dados pessoais – diligência que, comparativamente, mostra-se mais importante.

---

[45] FRAZÃO, Ana. *Fundamentos da proteção dos dados pessoais – Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399. P. 26.

[46] ibidem

O mais grave na assimetria técnico-jurídica que surgiu com o contexto de dados pessoais e aqueles que não identificam unicamente o indivíduo é o risco gerar perigosos episódios de violação e vazamento de dados, como em julho de 2021, quando dados e documentos de mais de 220 milhões de brasileiros foram expostos por agentes criminosos e vendidos como mercadoria na internet.[47]

Este capítulo abordará as principais técnicas de anonimização utilizadas no mundo com o intuito de diminuir a distância entre os técnicos e agentes de tratamentos de dados e os profissionais do Direito.

Não há um único método ou combinação para realizar o processo de anonimização, deve-se passar a uma análise do contexto em que o dado esteja inserido sempre com o objetivo de que os dados anonimizados não sejam nunca reidentificados, sendo protegidos até mesmo de quem realizou a anonimização.

As principais técnicas abordadas serão: supressão, substituição, generalização, perturbação e agregação.[48]

3.1 ANONIMIZAÇÃO POR SUPRESSÃO

A supressão é a técnica mais radical em anonimização e, geralmente, é utilizada quando não é possível lançar mão de outras técnicas. A supressão de dados garantirá uma proteção ao dado pessoal e uma irreversibilidade do dado tratado, contudo a utilidade do dado estará comprometida.

Normalmente utiliza-se a supressão para eliminar identificadores diretos do titular do dado em um conjunto em que é possível excluir o identificador e não inutilizar o significado de toda a amostra trabalhada.[49]

A supressão é a exclusão de atributos de uma seção inteira de dados (ou a coluna de uma tabela de base de dados). Note-se que a supressão do dado deve eliminar de fato o dado identificador e não apenas escondê-lo, omiti-lo ou copiá-lo para

---

[47] Megavazamento expõe documentos e dados de mais de 220 milhões de brasileiros, . Olhar Digital, 2021 . Disponível em https://olhardigital.com.br/2021/07/29/seguranca/vazamento-dados-internet/. Acesso em 14/09/2021

[48] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 17. Item 9. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 set. 2021.

[49] ibidem

outra porção da base.

Em certos contextos, a informação é criptografada ou mascarada e não excluída, possibilitando a reversão pelo possuidor da chave de criptografia capaz de reverter o processo. Desta forma, o processo de anonimização pode ser desfeito e os dados recuperados. Contudo, essa forma de supressão enquadra-se mais propriamente como pseudoanonimização.

No exemplo abordado neste trabalho, mais especificamente na tabela 5, a anonimização por supressão consistiria em suprimir toda coluna CPF, pois ó único atributo que identifica diretamente as pessoas tratadas no conjunto de dados (note-se que pelo fato de haver homônimos, a coluna nome não mais se enquadra como identificador único no conjunto de dados trabalhados).

Tabela 5 – Tabela de atributos de alunos de uma faculdade anonimizados por supressão

Nome	Idade	Sexo	CEP	Renda	Data de	Curso
Ana Silva	18	M	13090-718	850,00	11/10/16	Direito
Ana Silva	20	M	13090-590	1200,00	12/10/16	Medicina
Ana Silva	23	F	13080-379	3200,00	15/10/11	Odontologia
Jose Aldo	25	F	13080-190	1900,00	18/10/16	Psicologia
Jose Aldo	67	F	13080-340	900,00	21/10/15	Direito
Bruno Carvalho	71	F	13080-350	1000,00	27/10/14	Medicina
Bruno Carvalho	45	M	13080-714	2500,00	02/11/17	Direito
Bruno Carvalho	52	M	13080-360	1250,00	07/11/14	Economia
Ataulfo Silva	21	F	13080-190	1900,00	07/11/19	Engenharia
Mateus Ferreira	33	M	13080-691	1500,00	07/11/11	Direito
Virgílio Alves	70	M	13080-369	900,00	09/11/14	Medicina
Rafael Silva	27	M	13090-590	1200,00	17/11/16	Odontologia

Fonte: elaborada pelo autor

Percebe-se, com o apoio da tabela 5, que nenhum atributo identifica uma pessoa natural diretamente, o que permite considerar esse conjunto uma série de dados anonimizados.

3.2 ANONIMIZAÇÃO TROCA, SUBSTITUIÇÃO OU PERMUTA

O processo consiste em reorganizar os dados no conjunto de dados de tal forma que os valores dos atributos individuais ainda estejam apresentados no conjunto, mas geralmente não correspondam ao registo original. [50]

Nesse caso, é também possível substituir os dados que possam levar à identificação de um indivíduo por outro dado, que não está diretamente identificado com o indivíduo a quem se referem os dados. No exemplo da tabela 2, seria possível substituir o dado do CPF por identificador aleatório e único, de maneira a eliminar a inferência direta de um CPF com uma pessoa natural. A tabela 3 mostra um conjunto de dados anonimizados por permuta. Note-se que, caso se guarde a relação entre o número aleatório e o CPF, passamos a ter uma pseudoanonimização e não mais uma anonimização pura.

3.3 ANONIMIZAÇÃO POR GENERALIZAÇÃO

Como discutido no capítulo 1, existem informações que embora não identifiquem indivíduos de forma explicita, podem ser combinadas de forma a identificar indivíduos em um banco de dados anonimizados. A combinação de vários atributos pode ser única para uma determinada pessoa. Atributos como esse são chamados de semi-identificadores e quando conjugados podem passar a identificar um indivíduo.

A questão é que, com a eliminação dos identificadores e semi-identificadores,

[50] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 17. Item 9. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 set. 2021.

os dados restantes possuirão pouca ou quase nenhuma utilidade.

Um banco de dados devidamente anonimizados não deve conter dados pessoais, mas se todos os identificadores e semi-identificadores fossem removidos do banco de dados, geralmente os dados restantes não possuiriam muita utilidade. Este certamente seria o caso se isto ocorresse em nosso exemplo:

Tabela 6 – Tabela de atributos de alunos de uma faculdade com todos os semi-identificadores excluídos

Curso
Direito
Medicina
Odontologia
Psicologia
Direito
Medicina
Direito
Economia
Engenharia
Direito
Medicina
Odontologia

Fonte: elaborada pelo autor

A tabela 6 mostra um conjunto de dados bem restrito que pouco pode ser utilizado com algum significado. A supressão de todos os identificadores e semi-identificadores traz extensos prejuízos ao próprio significado do dado.

A generalização surge como técnica para casos como esse. Ela consiste em uma redução deliberada na precisão dos dados, como por exemplo a conversão da idade de uma pessoa numa faixa etária, ou de um ponto preciso para um ponto menos preciso. Esta técnica também é denominada de recodificação.[51]

Essa técnica é muito útil quando os valores podem ser generalizados e ainda úteis para o objetivo pretendido.

Normalmente, a generalização cria faixas de dados cuidadosamente

[51] idem

trabalhadas conforme as necessidades da entidade que necessite trabalhar o dado.

Por exemplo, com os dados da tabela 1, os dados poderiam ser generalizados por ao eliminar o dia e o mês da data de ingresso (restando apenas o ano), assim como eliminar os últimos dígitos do CEP.

Tabela 7 – Tabela de atributos de alunos de uma faculdade a data de ingresso e CEP generalizados.

Nome	Idade	Sexo	CEP	Renda Mensal	Data de ingresso	Curso
Ana Silva	18	M	13090	850,00	2016	Direito
Ana Silva	20	M	13090	1200,00	2016	Medicina
Ana Silva	23	F	13080	3200,00	2011	Odontologia
Jose Aldo	25	F	13080	1900,00	2016	Psicologia
Jose Aldo	67	F	13080	900,00	2016	Direito
Bruno Carvalho	71	F	13080	1000,00	2014	Medicina
Bruno	45	M	13080	2500,00	2014	Direito
Bruno Carvalho	52	M	13080	1250,00	2014	economia
Ataulfo Silva	21	F	13080	1900,00	2019	engenharia
Mateus Ferreira	33	M	13080	1500,00	2011	Direito
Virgílio Alves	70	M	13080	900,00	2014	Medicina
Rafael Silva	27	M	13090	1200,00	2016	Odontologia

Fonte: elaborada pelo autor

Na tabela 7, não há mais o identificador direto CPF (que indicaria exclusivamente uma pessoa natural). Contudo se somente o CPF fosse excluído por supressão, ainda restariam semi-identificadores que, em conjunto, poderiam individualizar uma pessoa. Assim, a generalização da data de ingresso e do CEP da pessoa contribuem para o prejuízo da junção dos semi-identificadores. Outra opção seria também generalizar o curso pretendido, o que traria ainda mais segurança para a base anonimizada.

Tabela 8 – Tabela de atributos de alunos de uma faculdade a data de ingresso, CEP e curso generalizados.

Nome	Idade	Sexo	CEP	Renda Mensal	Data de ingresso	Curso
Ana Silva	18	M	13090	850,00	2016	Humanas
Ana Silva	20	M	13090	1200,00	2016	Saúde
Ana Silva	23	F	13080	3200,00	2011	Saúde
Jose Aldo	25	F	13080	1900,00	2016	Saúde
Jose Aldo	67	F	13080	900,00	2016	Humanas
Bruno Carvalho	71	F	13080	1000,00	2014	Saúde
Bruno Carvalho	45	M	13080	2500,00	2014	Humanas
Bruno Carvalho	52	M	13080	1250,00	2014	Humanas
Ataulfo Silva	21	F	13080	1900,00	2019	Exatas
Mateus Ferreira	33	M	13080	1500,00	2011	Humanas
Virgílio Alves	70	M	13080	900,00	2014	Saúde
Rafael Silva	27	M	13090	1200,00	2016	Saúde

Fonte: elaborada pelo autor

A generalização é muito utilizada para análises estatísticas que necessitem apenas de elementos comuns a várias pessoas, sem a necessidade de identificá-las pessoalmente. Por exemplo, na generalização do CEP, as análises sobre o perfil demográfico dos alunos ingressantes não estariam prejudicadas e, como o número de indivíduos é extenso, a identificação única torna-se difícil ou mesmo impossível.

O desafio da generalização é conceber faixas em dimensões apropriadas, assim como descreve o trabalho conjunto de Singapura:

> [...]Conceber as faixas etárias em dimensões apropriadas. Faixas etárias que sejam demasiado grandes podem implicar que os dados podem ser muito "modificados", enquanto que as faixas muito estreitas podem significar que os dados estão pouco alterados e por isso ainda fáceis de reidentificar. Se o k-anonimato for utilizado, o valor k escolhido irá afectar os alcances de dados também. Nota-se que a primeira e a última categoria podem ter um alcance maior para acomodar o número tipicamente mais baixo de registos nestes pontos; isto é frequentemente referido de codificação do topo/base.[52]

[52] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 20 Item 10.4. Disponível em: https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques. Acesso em: 14 set. 2021.

K-anonimidade é conceito desenvolvido por Latanya Sweeney com o objetivo de não possibilidade de identificar uma pessoa garantindo a utilidade dos dados e, ao mesmo tempo, classificando um nível de segurança em relação a não identificação. A ideia é que, ao atender ao critério de k-anonimidade, a combinação dos semi-identificadores que identifique uma pessoa ocorra, no mínimo, 2 vezes. Sendo assim, o valor mínimo de "k" para que se mantenha a privacidade do banco de dados é igual a 2. [53]

Em resumo, os semi-identificadores, quando agrupados em um único registro no qual, pelos menos k -1 registros compartilham os mesmos valores para seus atributos. Assim, é garantida a privacidade desde registro em 1/k[54]. Quanto maior o valor de k, menor a probabilidade de se reidentificar uma pessoa, muito embora a utilidade é prejudicada na mesma proporção.

Para que mais de uma pessoa compartilhe o mesmo valor do conjunto de registros, é preciso que o valor seja maior que 2, assim o conjunto de valores não identificariam uma pessoa natural somente. Contudo, vale ressaltar que um valor muito pequeno de k poderá gerar uma reidentificação por inferência. Caso o valor seja 2, uma pessoa poderá ser identificada com 50% de certeza.

Normalmente, o atendimento ao critério da k-anonimidade passa pela supressão de identificadores diretos e a generalização de semi-identificadores.

No exemplo da tabela 1, para que os dados atendam ao critério da k-anonimidade para garantia da privacidade dos dados, é preciso configurá-los como descrito na tabela 9.

[53] SWEENEY, L. Simple demographics often identify people uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh, EUA, 2000

[54] SHMATIKOV, Vitaly. k-Anonymity and Other Cluster-Based Methods. University of Texas, CS 380S: Theory and Practice of Secure Systems. Class presentation slides.

Tabela 9 – Dados anonimizados que atendem ao critério k-nanonimização

Nome	Sexo	CEP	Data de ingresso	Curso
Ana Silva	M	13090	2016	Humanas
Ana Silva	M	13090	2016	Saúde
Ana Silva	F	13080	2011	Saúde
Jose Aldo	F	13080	2016	Saúde
Jose Aldo	F	13080	2016	Humanas
Bruno Carvalho	F	13080	2014	Saúde
Bruno Carvalho	M	13080	2014	Humanas
Bruno Carvalho	M	13080	2014	Humanas

Fonte: elaborada pelo autor

Note-se que o conjunto de dados {nome, sexo, cep, data de ingresso} identificam, no mínimo duas pessoas diferentes, atendendo ao critério com k=2.

Esse modelo e algumas derivações como l-diversidade[55], t-proximidade e m-invariância[56] foram adotados por boa parte dos meios digitais de geração de conteúdo assim como se tornaram um padrão da indústria.

Um dos maiores repositórios de dados, que trabalha diretamente com estatísticas e estudos sobre dados de pessoas com o objetivo de definir perfis de comportamento e consumo de produtos e informação, é a base de dados do *google*. A empresa afirma utilizar a anonimização de dados no intuito de resguardar a privacidade e segurança de seus usuários

> [...] Com a análise de dados anonimizados, podemos criar produtos e recursos seguros e valiosos, como o preenchimento automático de uma consulta de pesquisa inserida, e detectar com mais precisão as ameaças à segurança, como sites de phishing e malware, além de proteger a identidade dos usuários. Podemos também compartilhar externamente e com segurança os dados anonimizados, tornando-os úteis para outras pessoas sem colocar a privacidade dos nossos usuários em risco. [57]

Em sua declaração de política de privacidade, a google afirma utilizar a

[55] GOOGLE. Como o Google Anonimiza os Dados. In: Privacidade & Termos do Google. Disponível em: https://policies.google.com/technologies/anonymization?hl=pt-BR. Acesso em: 15 set. 2021.
[56] XIAO, Xiaokui; TAO, Yufei. m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets. Hong Kong: Chinese University of Hong Kong, Department of Computer Science and Engineering, 2007. P. 2.
[57] 4 GOOGLE. Como o Google Anonimiza os Dados. In: Privacidade & Termos do Google. Disponível em: . Acesso em: 15 set. 2021.

generalização de dados, com a preocupação de respeitar o conceito de Swenney sobre o k-anonimato:

> [...]generalização nos permite alcançar o k-anonimato, um termo padrão do setor usado para descrever uma técnica que esconde a identidade dos indivíduos em um grupo de pessoas semelhantes. No k-anonimato, k é um número que representa o tamanho de um grupo. Se, para qualquer indivíduo do conjunto de dados, houver pelo menos k-1 indivíduos que tenham as mesmas propriedades, teremos alcançado o k-anonimato para esse conjunto de dados. Por exemplo, imagine um conjunto de dados específico em que k seja igual a 50 e a propriedade seja o CEP. Se observarmos qualquer pessoa desse conjunto de dados, sempre encontraremos 49 outras pessoas com o mesmo CEP. Portanto, não conseguiremos identificar nenhuma pessoa a partir do CEP dela. [58]

3.4 ANONIMIZAÇÃO POR PERTURBAÇÃO

A perturbação difere da generalização no sentido em que os dados são alterados e não sofrem somente uma redução na precisão. A ideia é selecionar semi-identificadores que combinados com outras fontes de dados, possam gerar novos valores aceitáveis.[59] É uma técnica bastante utilizada, contudo não deve ser implementada no caso de uma necessidade de precisão do dado. A precisão vai depender do tamanho da base que possui os dados reais, caso a base possua poucos valores a anonimização será precária, se a base for muito grande, os valores finais serão bem divergentes dos originais comprometendo a utilidade do dado.

A perturbação pode envolver várias técnicas como arredondamento, e agregação, contudo a mais utilizada é a adição de ruído, com seu detalhamento de privacidade diferencial[60]

Na adição de ruído, o dado é alterado de maneira aleatória decrescendo ou acrescendo valores. Quando o conjunto de dados é expressivo, essas pequenas alterações de faixa de dados não altera a análise estatística pretendida pois a variância é garantida com um desvio padrão reduzido.

No exemplo dos alunos da faculdade (tabela 1), a adição de ruído consistiria na alteração da coluna “data de ingresso” com, por exemplo, a adição ou subtração de 4

---

[58] ibidem

[59] SINGAPORE (Personal Data Protection Commission). Guide to Basic Data Anonymization Techniques. Publicado em 25 jan. 2018. P. 20 Item 10.4. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 set. 2021.

[60] 2 PILLOW, Timothy. A Review of Synthetic Tabular Data Tools and Models: Anonymization methods that are revolutionizing how we share data. Medium: 02 jul. 2020. Disponível em: . Acesso em: 15 nov. 2020.

anos para todos os indivíduos. Nesse caso, perguntas estatísticas como “Quantos alunos ingressaram na faculdade no mesmo ano?” poderiam ser respondidas, contudo, a pergunta como “Quantos alunos entraram na faculdade em 2016?” estariam prejudicadas.

Um dos pontos positivos desta técnica é a possibilidade de se mensurar adequadamente o grau de privacidade em detrimento da utilidade dos dados pessoais com base na escala e força de interferência do desvio escolhido.

Tabela 10 – Dados anonimizados por adição de ruído à data de ingresso

Nome	Idade	Sexo	CEP	Renda Mensal	Data de ingresso	Curso
Ana Silva	18	M	13090	850,00	2020	Humanas
Ana Silva	20	M	13090	1200,00	2020	Saúde
Ana Silva	23	F	13080	3200,00	2015	Saúde
Jose Aldo	25	F	13080	1900,00	2020	Saúde
Jose Aldo	67	F	13080	900,00	2020	Humanas
Bruno Carvalho	71	F	13080	1000,00	2018	Saúde
Bruno Carvalho	45	M	13080	2500,00	2018	Humanas
Bruno Carvalho	52	M	13080	1250,00	2018	Humanas
Ataulfo Silva	21	F	13080	1900,00	2023	Exatas
Mateus Ferreira	33	M	13080	1500,00	2015	Humanas
Virgílio Alves	70	M	13080	900,00	2018	Saúde
Rafael Silva	27	M	13090	1200,00	2020	Saúde

Fonte: elaborada pelo autor

Essa técnica é amplamente utilizada por empresas da área digital, indústrias e atividades de pesquisa acadêmica. Assim como na generalização, a google cita a adição de ruído como uma técnica utilizada por suas empresas empreas para garantir a anonimização e privacidade de seus dados:

> A privacidade diferencial (outro termo padrão do setor) descreve uma técnica para adição de ruído matemático aos dados. Com a privacidade diferencial, é difícil determinar se um indivíduo faz parte de um conjunto de dados, porque o resultado de um algoritmo específico parecerá essencialmente o mesmo, independentemente de as informações dos indivíduos estarem incluídas ou omitidas. Por exemplo, imagine que estejamos medindo a tendência geral nas pesquisas sobre a gripe em uma região geográfica. Para alcançarmos a privacidade diferencial, adicionamos ruído ao conjunto de dados. Isso significa que podemos adicionar ou subtrair o número de pessoas que estão pesquisando sobre a gripe em uma área específica. Isso, no entanto, não afetaria nossa medição da tendência em uma região geográfica mais ampla.

> É importante notar também que a adição de ruído a um conjunto de dados pode torná-lo menos útil. [61]

4. IMPLICAÇÕES JURÍDICAS DA ANONIMIZAÇÃO E DA REINDENTIFICAÇÃO

Os conceitos sobre dados pessoais, anonimização e reidentificação, assim como a exemplificação das técnicas mais utilizadas hoje são essenciais para o entendimento sobre como as legislações acerca dos dados pessoais foram pensadas, discutidas, ordenadas e sancionadas.

O cerne da questão é: há a possibilidade de anonimizar um dado de maneira que ele nunca possa ser reidentificado em nenhum momento do tempo?

A temática sobre a validade de tais técnicas começou com o emblemático caso da reidentificação dos dados do governador Weld[62], nos anos 90, que culminou no modelo de k-anonimidade e na reidentificação dos dados do prêmio Netflix em 2006.[63]

4.1 O CASO DO GOVERNADOR WELD

O estado de Massachusetts, nos Estados Unidos da América, nos anos 90, por meio de sua agência administradora dos planos de saúde dos funcionários públicos do estado, compartilhou dados médicos para todos os pesquisadores que necessitassem.

Em meio às preocupações sobre a privacidade dos registros médicos das pessoas envolvidas, o governador do estado á época, William Weld, veio a público para garantir que todas as informações pessoais dos dados compartilhados seriam removidas.

Latana Sweeney, estudante no *Massassuchets Institute of Technology* (MIT) conseguiu, ao intentar a fragilidade dos métodos de anonimização, utilizando técnicas estatísticas identificar os registros completos do próprio governador do estado.

---

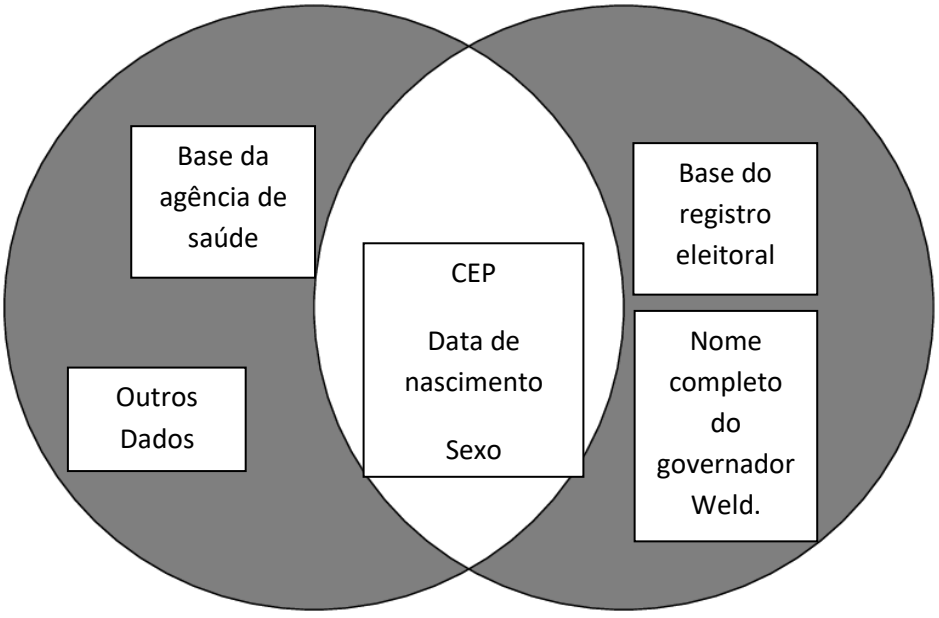
[61] GOOGLE. Como o Google Anonimiza os Dados. In: Privacidade & Termos do Google. Disponível em: https://policies.google.com/technologies/anonymization?hl=pt-BR. Acesso em: 15 set. 2021.
[62] SWEENEY, L. Simple demographics often identify people uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh, EUA, 2000.
[63] NETFLIX INC. The Netflix Prize Rules. 2006. Disponível em: <http://www.netflixprize.com/rules.html> Acesso em: 06 nov. 2016.

A técnica utilizada por Sweeney foi a combinação das informações presentes no banco de dados anonimizado com outras retiradas do registro público dos eleitores da cidade – o acesso ao registro foi feito legalmente pelo preço de 20 (vinte) dólares.

A base de dados dos eleitores possuía o nome completo de cada eleitor, o endereço, a filiação partidária, a data da última votação, o CEP, o sexo e a data de nascimento. Sweeney identificou o governador com a ajuda dos dados de CEP, sexo e data de nascimento. Ao fazer a combinação desses atributos do banco de dados eleitoral e aquele disponibilizado pela agência de saúde, foi possível identificar unicamente o governador e expor todos os seus dados.

Na verdade, o agente de tratamento de dados não considerou que, uma vez os dados ditos anonimizados fossem publicados, eles poderiam ser combinados com outras fontes públicas de dados e a interseção entre eles identificar unicamente uma pessoa.

Figura 1 – Intersecção entre os dados pessoais do governador Weld



Fonte: elaborada pelo autor

Na base de registro eleitoral não anonimizada, mas de domínio público, as informações constantes de CEP, data de nascimento e Sexo (que constavam também

da base da agência de saúde) indicavam diretamente o nome do governador. Assim, Sweeney identificou todo o prontuário e dados do próprio governante.

A resposta parece tender ao descrédito em técnicas infalíveis de anonimização e, de fato, muitos estudos têm corroborado com a fabilidade dos algoritmos de anonimização.[64]Conforme Arvind Narayan defende, a teoria de que os atributos de identificação de uma base de dados poderiam ser completamente eliminados, garantindo-se, com 100% (cem por cento) de eficiência, o anonimato das pessoas, é um mito.[65]

O fato é que, por mais complexas que sejam as técnicas de anonimização, o risco de reidentificação persiste em todas elas. Como lembra o professor Bioni:

> Por essa lógica, qualquer dado pessoal anonimizado detém o risco inerente de se transmudar em um dado pessoal. A agregação de diversos "pedaços" de informação (dados) pode revelar (identificar) a imagem (sujeito) do quebra-cabeça, a qual era até então desfigurada (anônimo) – o chamado efeito mosaico.
> Por isso, leis que adotam o conceito expansionista de dados pessoais e, ao mesmo tempo, estabelecem uma dicotomia deste com dados anônimos correriam o risco de serem tautológicas. Isso porque haveria uma redundância normativa, já que dados anônimos seriam, em última análise, potencial e provavelmente, dados relacionados a uma pessoa identificável. [66]

Nesse momento, vale uma importante alusão que servirá de base para análise nas próximas seções deste trabalho, assim como o estudo sobre como as regulamentações debruçaram-se sobre o tema e o que ainda falta para a devida confiança nos processos de anonimização.

Após o fatídico caso do governador Wels, as informações geográficas e de data passaram a fazer parte da lista obrigatória de dados que necessitam sofrer anonimização, sobretudo com o processo de generalização. Essa determinação foi realizada pela HIPPAA[67] e obriga que dados médicos não podem conter subdivisões geográficas menores que um estado e que nenhuma informação de data deve ser

---

[64] BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioetca digital Jurídica do Superior Tribunal de Justiça, 2019

[65] NARAYANAN, Arvind; SHMATIKOV, Vitaly. Myths and fallacies of "personally identifiable information". Communications of the ACM, Nova York, v. 53, n. 6, p. 24-26, 2010. Disponível em: http://bit.ly/30G9CVq. Acesso em: 15 set. 2021.

[66] BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioetca digital Jurídica do Superior Tribunal de Justiça, 2019, p. 191-192

[67] Health Insurance Portability and Accountability Act of 1996 (HIPPAA) é uma lei federal que definiu os padrões nacionais para proteger dados sensíceis de pacientes de serem utilizados sem ao consentimento ou conhecimento.

mais específica que os 4 dias de ano.[68]

4.2 O CASO DO PRÊMIO NETFLIX

Fica claro que o processo de anonimização do governador Weld não utilizou os requisitos mínimos da k-anonimização e o evento foi inspiração para Latanya Sweeney desenvolver o conceito.

Contudo, dois cientistas de dados, Armind Narayanan e Vitaly Shmatikov, conseguiram provar que mesmo dados devidamente anonimizados e com k-anonimização presente ainda é possível reidentificar usuários. A dupla publicou em 2008 um artigo comprovando esta hipótese, após um árduo tempo de trabalho.[69]

Trata-se de um trabalho baseado no prêmio de uma grande empresa de transmissão direta de filmes, a Netflix, em 2006. [70]

No prêmio, a empresa desafiava a comunidade a desenvolver um algoritmo melhor que o seu (o Cinematch) que obtivesse sucesso em prever como os usuários iriam avaliar seus filmes mais recentes. Qualquer algoritmo que conseguisse acertos 10% maiores que o Cinematch ganharia a recompensa.

A empresa procedeu um processo de anonimização e publicou as seguintes informações: ID de usuário (número aleatório); nome do filme, data da avaliação anterior e a própria avaliação, com valores de 1 a 5. Para garantir a anonimização a Netflix adicionou ruído aos dados com a supressão, substituição e inserção de algumas avaliações. Toda essa modificação, garantiu o veículo, conferia a devida anonimização, mas ainda assim não descontruía o valor estatístico dos dados.

O conjunto de dados publicado após este processo de anonimização continha as seguintes informações: ID de usuário; nome do filme; data da avaliação e quantidade de estrelas dadas pelo cliente ao filme, um número inteiro de um a cinco. Além disso, a empresa perturbou parte dos dados publicados. Isto foi feito através da

---

[68] ESTADOS UNIDOS DA AMÉRICA. Other requirements relating to uses and disclosures of protected health information. Code of Federal Regulations. Título 45, Seção 164.514. Disponível em: <https://www.law.cornell.edu/cfr/text/45/164.514> Acesso em: 30 out. 2016.

[69] NARAYANAN, A.; SHMATIKOV, V. Robust De-anonymization of Large Datasets. SP ’08 Proceedings of the 2008 IEEE Symposium on Security and Privacy. p. 111- 125. Washington, EUA, 2008. Disponível em:<https://www.cs.cornell.edu/~shmat/shmat_oak08netflix.pdf> Acesso em: 09 set. 2021.

[70] NETFLIX INC. The Netflix Prize Rules. 2006. Disponível em: http://www.netflixprize.com/rules.html

remoção, modificação e inserção de algumas das avaliações dos usuários. Apesar disto, segundo os responsáveis pela competição, o conjunto de dados publicado não era estatisticamente diferente dos dados originais.

Os pesquisadores analisaram os dados não no intuito de produzir um algoritmo mais assertivo que o *Cinematch,* mas para provar que os dados poderiam ser reidentificados.

Eles obtiveram sucesso na reidentificação, mesmo com a utilização de dados meramente comportamentais, simplesmente utilizando outra base de dados: a do IMDB[71]. O cruzamento das informações obtidas pelo prêmio com as mesmas do IMDB possibilitou aos pesquisadores reidentificar usuários cadastrados na base de dados da Netflix.[72]

A argumentação da dupla é que toda informação que distinga as pessoas de outras pode ser usada para fins de reidentificação e que informações como históricos de pesquisa, relações de contatos, ou avaliações de filmes. Um tratador de dados que busque reidentificar um banco de dados anonimizados pode utilizar para este fim qualquer característica sobre uma pessoa que seja razoavelmente estável, e suficientemente variável de modo que seja improvável que duas pessoas possuam a mesma combinação de características.

> Cada vez menos sentido à medida que a quantidade de informação disponível publicamente sobre indivíduos aumenta exponencialmente"[73]

Paul Ohm vai além ao afirmar que muitas pessoas utilizam o mesmo nome de usuário em vários veículos da internet e que os pesquisadores poderiam identificar os usuários do Netflix em outros veículos, como Instagram ou Facebook.[74]

---

[71]IMDB (internet Movie Database" , pode ser acessado em https://www.imdb.com/, É um banco de dados online de informações relacionadas a filmes, programas de televisão, vídeos caseiros, videogames e conteúdo de streaming on-line – incluindo elenco, equipe de produção e biografias pessoais, resumos de enredos, curiosidades, classificações e críticas de fãs e críticas

[72] NARAYANAN, A.; SHMATIKOV, V. Robust De-anonymization of Large Datasets. SP'08 Proceedings of the 2008 IEEE Symposium on Security and Privacy. p. 111-125. Washington, EUA, 2008

[73] ibidem

[74] OHM, P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, Vol. 57, p. 1701-1770, U of Colorado Law Legal Studies Research Paper No. 9-12. Boulder, EUA, 2010

4.3 CONSIDERAÇÕES SOBRE OS PROCESSOS DE ANONIMIZAÇÃO.

O constatado, tanto por Paul Ohm quanto por Armind Narayanan e Vitaly Shmatikov é que mesmo dados que pareçam inofensivos como os de ingênuas avaliações de filme podem terminar por reidentificar uma pessoa natural. Tal reidentificação pode revelar traços dos perfis dos usuários como: preferências sexuais, políticas ou religiosas, e essas informações podem ser utilizadas para crimes ou delitos.

Atualmente, os estudos tendem a afirmar que a anonimização é um processo que pode falhar. Alguns autores afirmam que a anonimização com eficácia plena é um mito[75].

O professor Bioni corrobora com a preocupação:

> Por essa lógica, qualquer dado pessoal anonimizado detém o risco inerente de se transmudar em um dado pessoal. A agregação de diversos "pedaços" de informação (dados) pode revelar (identificar) a imagem (sujeito) do quebra-cabeça, a qual era até então desfigurada (anônimo) – o chamado efeito mosaico.
> Por isso, leis que adotam o conceito expansionista de dados pessoais e, ao mesmo tempo, estabelecem uma dicotomia deste com dados anônimos correriam o risco de serem tautológicas. Isso porque haveria uma redundância normativa, já que dados anônimos seriam, em última análise, potencial e provavelmente, dados relacionados a uma pessoa identificável. [76]

De fato, por mais que técnicas atuais de anonimização de dados demonstrem-se consistentes e eficazes, nada impede que novas tecnologias, novos recursos ou algoritmos, ou mesmo novos cruzamentos de dados, possam reverter o processo antes tido como infalível.

Partindo desse pressuposto, a doutrina e os pesquisadores resolveram por propor certos limites e condições para que o dado possa, de fato, ser considerado anonimizado e, portanto, fora do escopo das leis que restringem e criam regras para manejo de dados pessoais.

Bioni propõe que seja adotado um filtro de elasticidade ao conceito expansionista, com o intuito de garantir a segurança e confiabilidade na anonimização

---

[75]NARAYANAN, Arvind; SHMATIKOV, Vitaly. Myths and fallacies of "personally identifiable information". Communications of the ACM, Nova York, v. 53, n. 6, p. 24-26, 2010. Disponível em: http://bit.ly/30G9CVq. Acesso em: 28 set. 2021.

[76] BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioteca digital Jurídica do Superior Tribunal de Justiça, 2019, p. 191-192

do dado.

De fato, a elasticidade citada deve levar em conta um conceito chave em todo o processo que é a razoabilidade. Nesse conceito, não basta somente ser possível que um dado anonimizado seja reidentificado, mas é preciso que o esforço de reidentificação seja razoável. A razoabilidade seria então os limites da elasticidade considerada.

> A contrário sensu, se para a correlação entre um dado e uma pessoa demanda-se um esforço fora do razoável, não há que se falar em dados pessoais. Nessa situação, o dado é considerado como anônimo, uma vez que o "filtro da razoabilidade" barra o seu enquadramento como aquele relacionado a uma pessoa identificável.
> Com isso, há coerência em se estabelecer conceitos diferentes para tais espécies de dados, sobretudo sob o ponto de vista de uma dicotomia mutualmente excludente entre eles, que é delimitada pelo fator da razoabilidade. Do contrário, repita-se, haveria uma redundância normativa, na medida em que dados anônimos – sem o critério da razoabilidade – seriam sempre enquadrados dentro do conceito de dado pessoal, como aquele relacionado a uma pessoa identificável.[77]

Bioni defende que as regulamentações devem alinhar-se com uma linha neutra tecnológica.

Tal linha defende que os modelos regulatórios devem ser capazes de estimular, monitorar e acompanhar o desenvolvimento tecnológico, sem engessá-lo e nem ser permissivo aos riscos inerentes à tecnologia. Assim, não há a adoção ou indicação de tecnologias específicas, sob o risco inerente de se tornarem obsoletas tornando a reidentificação do dado pessoal um processo possível, a discussão então sobre a razoabilidade deve ser ressignificada e atualizada pelo próprio desenvolvimento científico.[78].

Contudo, além dos estudos dos estudiosos e cientistas, balizas para a redução da discricionariedade do conceito de razoável podem ser seguidas:

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

[77] BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioteca digital Jurídica do Superior Tribunal de Justiça, 2019, p. 191-192
[78] ibidem

> é o **agente de tratamento de dados e se ele dispõe de "meios próprios"** para reverter o processo de anonimização. Ao invés de considerar quais são os padrões sociais acerca da reversibilidade de um dado anonimizado, foca-se em analisar qual é a capacidade individual de engenharia reversa de quem processa tais dados.[79]

Vale lembrar que o segundo eixo citado por Bioni nada mais é que o conceito de pseudoanonimização já descrito neste trabalho.

Rubstein acrescenta que deve ser analisada uma matriz de risco a ser considerada para aquisição da confiança nos processos de anonimização e reidentificação. O fato é que a análise e certificação das duas abordagens permitirá saber o risco da produção de uma engenharia reversa que possibilitará a reidentificação do dado tido como anonimizado.[80]

Para Laura Schertel Mendes[81,] o século XX presenciou uma inexorável reinvenção da privacidade pois de um direito com dimensão estritamente negativa, passou a ser considerado uma garantia de controle do indivíduo sobre as próprias informações e um pressuposto para qualquer regime democrático.

Nesse sentido, tem-se que:

> O tratamento de dados pessoais, em particular por processos automatizados, é, no entanto, uma atividade de risco. Risco que se concretiza na possibilidade de exposição e utilização indevida ou abusiva de dados pessoais, na eventualidade desses dados não serem corretos e representarem erroneamente seu titular, em sua utilização por terceiros sem o conhecimento deste, somente para citar algumas hipóteses reais. Daí resulta ser necessária a instituição de mecanismos que possibilitem à pessoa deter conhecimento e controle sobre seus próprios dados – que, no fundo, são expressão direta de sua própria personalidade. Por este motivo, a proteção de dados pessoais é considerada em diversos ordenamentos jurídicos como um instrumento essencial para a proteção da pessoa humana e como um direito fundamental.[82]

Em seu estudo, ele sugere 06 (seis) cenários que aumentam o risco de reidentificação:

- Grandes volumes de dados
- Natureza dos dados – tem relação com o interesse de terceiros e quão recompensador seria o esforço para reidentificar o dado;

---

[79] Idem

[80] RUBINSTEIN, Ira; HARTZOG, Woodrow. Anonymization and risk. *91 Washington Law Review*, Washington, DF, v. 703, 2016; *NYU School of Law*, *Public Law Research Paper*, Nova York, n. 15-36. Disponível em: http://bit.ly/2TKMmo8. Acesso em: 29 set 2021

[81] 2014. p. 29

[82] .DONEDA, 2011, p. 92

- Cadeia de atividade de tratamento de dados – quanto maior a quantidade de atores para a geração da base anonimizada ou mesmo o uso, maior o risco
- Gerenciamento de identidade – quando menor o controle de acesso aos dados anonimizados, maior o risco
- Ausência de atualização contínua sobre as técnicas e tecnologias de anonimização
- Cláusulas contratuais – a ausência de cláusulas que obriguem a não reidentificação, delimitem papéis e obriguem um ciclo de vida da informação anonimizada.

Por tais razões, é bastante válida a posição de Paul Ohm em que é necessário que o atual modelo de anonimização, baseado apenas na análise técnica dos dados, seja substituído por um modelo mais completo, que envolva uma análise dos riscos e benefícios envolvidos no compartilhamento de dados. Assim, pode-se afirmar que as recomendações feitas pelos participantes que defendiam que dados anonimizados deveriam estar dentro do escopo da lei eram as mais adequadas, tendo em vista a literatura científica atual.

Bioni relata que a análise que deve ser feita para que o dado seja, de fato, considerado anonimizado envolve, essencialmente, o termo circunstancial:

> Ao invés de considerar anonimização como algo cujo resultado (output) é infalível, foca-se em uma abordagem que considera a aplicação sistemática de técnicas de anonimização com o objetivo de agregar consistência ao processo como um todo. Por essa razão, a análise acerca de se um dado deve ser, de fato, considerado como anonimizado é eminentemente circunstancial. Os dois critérios de análise – objetivo e subjetivo – acima mencionados, ganharão vida somente a partir do contexto no qual está inserida uma atividade de tratamento de dados, sobre a qual se busca retirar, ao máximo, seus respectivos identificadores.[83]

O conceito de razoabilidade pode ser delimitado pelos fatores descritos acima e serve de balizamento para todos os agentes envolvidos no tratamento e anonimização de dados.

Tal balizamento poderá viabilizar a escolha da melhor técnica a ser utilizada e, principalmente, considerando os riscos associados à possível reversão e

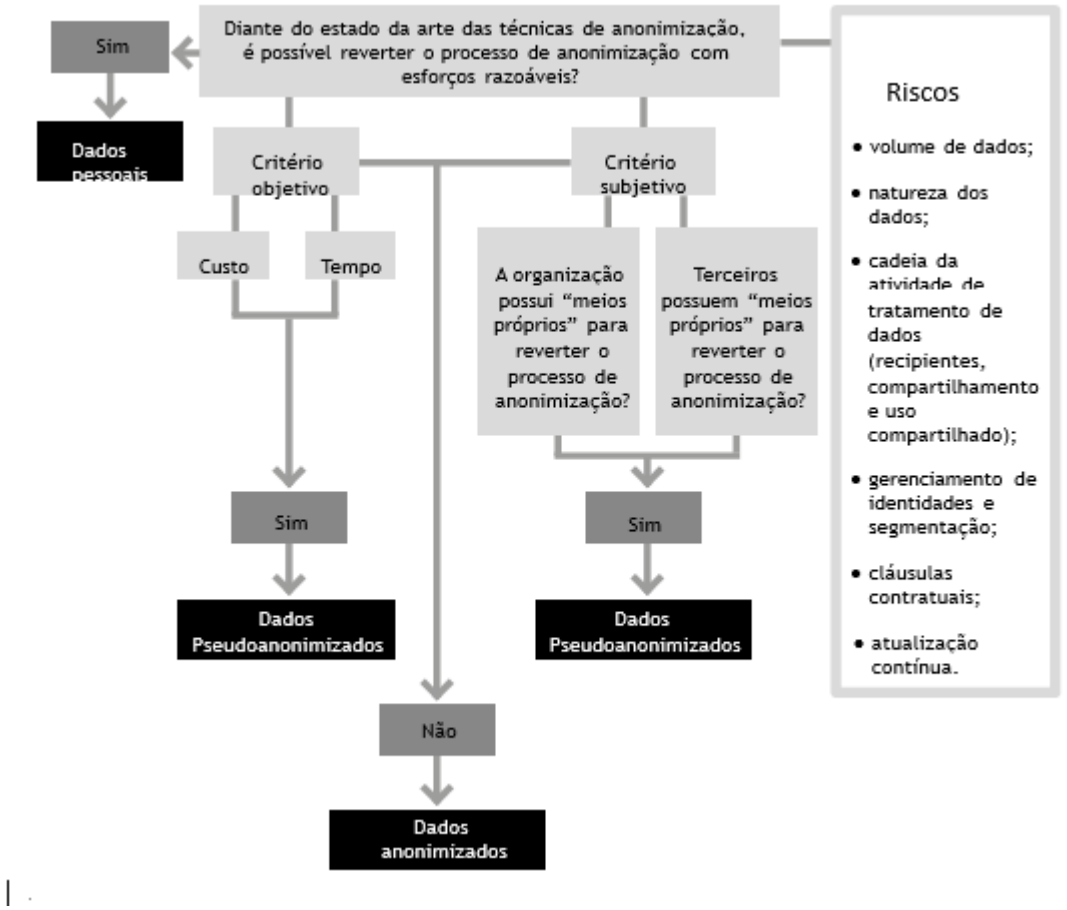
[83] BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioteca digital Jurídica do Superior Tribunal de Justiça, 2019, p. 191-192

reidentificação dos dados.

Por isso, não há um único método ou uma combinação perfeita para parametrizar o processo de anonimização. Deve-se proceder à análise do contexto, das circunstâncias, dos riscos, dos meios próprios e outros fatores para garantia de que os dados não sejam reidentificados, ou caso haja risco, este seja o mais mitigado possível.

Por fim, Bioni elucida graficamente os fatores a serem considerados pelos agentes de tratamento de dados, de forma a tornar minimamente seguro o processo de anonimização:

Figura 2 – Fluxograma de decisão sobre dado anonimizado[84]



Fonte: Bioni, 2019

Sendo assim, passa-se à análise de como as regulamentações brasileira e

[84] Ibidem, página 200

francesa abordaram o tema e intentaram trazer segurança aos titulares que porventura tenham seus dados pessoais tratados com técnicas de anonimização para os mais diversos fins.

Mais especificamente o contexto da anonimização de dados, no que tange o aspecto regulatório, é mister, pois as legislações costumam excluir do seu escopo de exigências, restrições e responsabilização os dados considerados anonimizados, diferindo apenas entre anonimização e pseudoanonimização.

5. REGULAMENTAÇÃO EUROPEIA

Não seria possível estudar as regulamentações francesa e brasileira sem antes analisar a regulamentação europeia que serve de base não somente para os países membro da União Europeia com para muitos outros países como o mais extenso da América Latina.

5.1 HISTÓRICO NORMATIVO SOBRE PROTEÇÃO DE DADOS

Na Europa, os sistemas normativos de proteção de dados são divididos em gerações de leis. A primeira geração surgiu na década de 1970, no contexto do estado de bem-estar social onde, para se proverem serviços sociais, era necessário realizar a coleta de dados dos cidadãos. A primeira geração de proteção dos dados pessoais teve foco na esfera governamental e na premissa de se estabelecer normas rígidas que domassem o uso da tecnologia[85].

Fala-se nesse sentido, de política regulatória de proteção de dados pessoais (regulação ex ante). À luz dessa posição, concorda-se que:

> Característica marcante do modelo europeu de proteção de dados é a exigência de que o controlador só possa tratar dados se tiver amparado em uma base legal, o que pode ser compreendido como uma racionalidade ex-ante de proteção de dados. (...) Há grande semelhança quando se compara os direitos europeu e brasileiro sob esse ponto de vista. Afinal, a grande motivação que a LGPD operou no ordenamento jurídico brasileiro pode ser compreendida exatamente na instituição de um modelo ex-ante de proteção de dados. (...) Esse modelo está amparado em três características centrais: i)

[85] MAYER-SCHÖNBERGER, 2013, p. 223, *apud* BIONI, 2019, p. 114

> um conceito amplo de dado pessoal; ii) necessidade de que qualquer tratamento de dados tenha uma base legal; e iii) legítimo interesse como hipótese autorizativa e a necessidade de realização de um teste de balanceamento de interesses[86]

Com o desenvolvimento tecnológico, novas leis surgiram e se adequaram a novos paradigmas do tratamento de dados realizado digitalmente, de modo que, nesse contexto, a proteção dos direitos fundamentais passou a exigir uma atuação positiva do Estado que procurou, a partir de novas normas, promover ao cidadão, o acesso, a correção e o cancelamento do tratamento de seus dados pessoais, reconhecendo-lhe o direito de delimitar a utilização desses dados pelos diversos bancos de dados existentes.

Exsurge daí, portanto, a segunda geração de leis de proteção de dados pessoais já no final da década de 70. Como exemplo, tem-se a própria Lei Francesa de Proteção de Dados Pessoais de 1978, a qual considerava a privacidade e a proteção de dados como uma liberdade negativa a ser exercitada pelo próprio cidadão contra as bases de dados estatais e da esfera privada[87].

A terceira geração de leis de proteção de dados pessoais surgiu em 1980 e passou a abranger, além da liberdade de fornecer ou não seus dados pessoais, a preocupação em garantir a efetividade desta liberdade[88]. "A ampliação do papel protagonista do indivíduo na proteção dos seus dados pessoais foi o divisor de águas para a terceira geração de leis"[89]. Com a participação ativa do cidadão e um envolvimento contínuo, desde a coleta até o compartilhamento dos seus dados pessoais, foi garantido ao indivíduo além do controle[90], sua "autodeterminação informacional"[91].

A década de 1980 fomentou o desenvolvimento econômico a partir do papel estratégico dos dados pessoais. Assim, a Organização para o Desenvolvimento e Cooperação Econômica (OCDE) passou a emitir as chamadas *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, diretrizes que contribuíram não só para a consolidação do conceito de privacidade ligado à proteção

---

[86] BIONI; MENDES, 2019, p. 810-811.
[87] DONEDA, 2006, p. 209
[88] DONEDA, 2006, p. 211
[89] BIONI, 2019, p. 115
[90] MENDES, 2014, p. 65-66
[91] PANEBIANCO, 2000, p. 187, *apud* DONEDA, 2006, p. 196

de dados pessoais, como também para o desenvolvimento do fluxo transfronteiriço de dados.

Outro passo importante para a consolidação de um sistema de proteção foi a Convenção nº 108/1981[92] da União Europeia, também conhecida como Convenção de Estrasburgo. A norma assegura a proteção das pessoas em relação ao processamento automático de seus dados e reconhece o caráter de direito fundamental destes, estabelecendo uma série de princípios para o tratamento e regulando também o fluxo transfronteiriço de dados pessoais. Outro ponto muito importante da norma foi a previsão de uma autoridade independente com o poder de zelar pela proteção dos dados pessoais. O órgão regulador, desde então, passou a ser fundamental para o sistema de proteção de dados pessoais.

Vinte e cinco anos depois do primeiro marco normativo, em 1995, a União Europeia promulgou a importante Diretiva 95/46/CE[93], um marco no campo da proteção de dados pessoais. Seu objetivo foi uniformizar as formas e as normas de tratamento de dados e fortalecer o mercado interno no âmbito da União Europeia. Todo esse arcabouço jurídico, formado pelas leis de quarta geração, se caracterizou pela utilização de instrumentos que elevaram o padrão coletivo de proteção, fortalecendo a posição do indivíduo frente às entidades coletoras e processadoras de seus dados.

6.2 A REGULAMENTAÇÃO GERAL DE PROTEÇÃO DE DADOS NA UNIÃO EUROPEIA

Devido ao rápido avanço tecnológico das últimas duas décadas, preocupada com os desafios da era digital, sobretudo para garantir uma maior eficácia na proteção dos dados pessoais, a Comissão Europeia propôs, em janeiro de 2012, um novo regulamento que garantiu uma reforma geral do regramento Europeu[94]. Assim, em maio de 2018, a Diretiva 95/46/CE foi substituída pelo Regulamento n° 2016/679[95], a nova Lei Geral de Proteção de Dados da União Europeia, mais conhecida como

[92] Em 2012, a Convenção 108 foi modernizada, reforçando a proteção à privacidade no campo digital, através de um protocolo que altera a Convenção 108 (Protocolo CETS 223)
[93]Disponível em :https://eur-lex.europa.eu/legalcontent/PT/TXT/PDF/?uri=CELEX:31995L0046&from=PT Acesso em 27/09/2021
[94]COPETTI, 2018, p. 172
[95]Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679. Acesso em 27/09/2021

General Data Protection Regulation (GDPR) e pela Diretiva 2002/28/EC, conhecida como ePrivacy, por regular aspectos de privacidade em serviços de comunicação eletrônica.

Esses instrumentos jurídicos são aplicados por *Data Protection Authorities*, ou seja, autoridades nacionais independentes de proteção de dados, que no topo hierárquico europeu são orientadas pela *European Data Protection Supervisor* que aconselha as instituições e os organismos da UE sobre todos os aspectos do tratamento dos dados pessoais, das políticas e da legislação neste domínio.

A aplicabilidade do GDPR está restrita apenas aos dados de pessoas naturais localizadas no âmbito da União Europeia, mas a todo o fluxo de dados existente entre os países membros e os demais países ao redor do mundo, ou seja, todos aqueles que possuem pontos de contato com o mercado europeu.

Devido à sua abrangência e amplitude legislativa, o GDPR se tornou um modelo mundial, e diversos países (entre eles o Brasil), nele se espelham em busca de padrões normativos uniformes para a proteção de dados pessoais. Como se trata de um regulamento, e não de uma diretiva, o GDPR é diretamente aplicável aos 28 Estados Membros, já que não há a necessidade de qualquer transposição para cada jurisdição nacional, sendo este considerado como uma norma interna, diferente da Diretiva 95/46/CE[96].

Em relação à definição de dados pessoais, o GDPR adotou um conceito expansionista[97] no qual dado pessoal é qualquer tipo de informação que permita sua identificação, ainda que o vínculo não seja estabelecido de imediato, mas de maneira indireta, ou seja, esses dados se referem a circunstâncias pessoais ou materiais de um indivíduo identificado ou identificável[98]. Segundo o GDPR, consideram-se dados pessoais quaisquer informações utilizadas para identificar uma pessoa, tais como, dados de localização de usuário, IDs de dispositivos móveis e até endereços IP.

O conceito expansionista tem, de fato, guiado interpretações mais amplas do conceito de dados pessoais como o acórdão do Tribunal de Justiça da União Europeia em 2016 no emblemático caso de Patrick Breyer contra Bundesrepublik Deutschland[99].

---

[96] GUIDI, 2017, p. 3
[97] BIONI, 2019, p. 68
[98] MENDES, 2014, p. 56
[99] TJUE 19-out.-2016, proc. C-582/14 (*Breyer v Bundesrepublik Deutschland*).

O tribunal europeu considerou que um endereço de protocolo Internet dinâmico (endereço único de comunicação entre todas as máquinas em rede para encaminhamento dos dados[100]) registrado por um prestador de serviços de meios de comunicação em linha aquando da consulta por uma pessoa de um sítio Internet que esse prestador disponibiliza ao público constitui, relativamente a esse prestador, um dado pessoal, quando este disponha de meios legais que lhe permitam identificar a pessoa em causa graças às informações suplementares que o fornecedor de acesso à Internet dessa pessoa dispõe. Muito embora o argumento de que esse endereço não revela diretamente a identidade da pessoa singular proprietária do computador a partir do qual se efetua a consulta de um sítio na Internet, nem a de outra pessoa que possa utilizar esse computador, o TJUE acredita haver conexão entre esse dado com um dado pessoal identificável, excluindo somente o cenário em que seja necessário um esforço desproporcional em termos de tempo, custos e mão de obra ou mesmo razoavelmente impraticável (os conceitos de razoável, esforço e custo serão abordados com mais detalhes nos próximos títulos deste trabalho)

6.3 DADO ANONIMIZADO E PSEUDOANONIMIZADO NO ESCOPO DA GDPR

No que tange ao escopo de dados anonimizados, a GDPR, em seu considerando 26, dispõe expressamente que os dados anônimos não estão compreendidos em sua regulamentação, contudo ela não exclui, de maneira expressa, os dados pseudoanonimizados:

> 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 pseudoanonimizados, que possam ser atribuídos a uma pessoa singular mediante a utilização de informações suplementares, deverão ser considerados informações sobre uma pessoa singular identificável.
> Para determinar se uma pessoa singular é identificável, importa considerar todos os meios suscetíveis de ser razoavelmente utilizados, tais como a seleção, quer pelo responsável pelo tratamento quer por outra pessoa, para identificar direta ou indiretamente a pessoa singular.
> Para determinar se há uma probabilidade razoável de os meios serem utilizados para identificar a pessoa singular, importa considerar todos os fatores objetivos, como os custos e o tempo necessário para a identificação, tendo em conta a tecnologia disponível à data do tratamento dos dados e a

[100] Kurose, James F.; Ross, Keith W. (2013). *Redes de computadores e a internet: uma abordagem top-down* 6ª ed. São Paulo: Pearson Education do Brasil.

> evolução tecnológica.
> 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 anónimos que o seu titular não seja ou já não possa ser identificado.
> O presente regulamento não diz, por isso, respeito ao tratamento dessas informações anónimas, inclusive para fins estatísticos ou de investigação.[101]

Vale ressaltar que a GDPR, no considerando, elenca os princípios da razoabilidade e os limites balizadores citados neste trabalho.

Assim, a lei de ouro do regulamento europeu não se aplica a dados anonimizados, notadamente dados que não se relacionam com uma pessoa natural identificada ou identificável ou a dados pessoais tornados anônimos de tal forma que o sujeito de dados não é ou não é mais identificável.

O considerando 26 prevê que para determinar se uma pessoa natural é identificável, deve-se levar em conta todos os meios razoavelmente propensos a serem usados, seja pelo controlador ou por outra pessoa para identificar a pessoa física direta ou indiretamente. Para verificar se os meios são razoavelmente propensos a serem utilizados para identificar a pessoa física, deve-se levar em conta todos os fatores objetivos, como os custos e o tempo necessário para identificação, levando em consideração a tecnologia disponível no momento do processamento e desenvolvimento tecnológico.

Portanto, todos os fatores objetivos devem ser considerados, como o custo e o tempo necessário para identificação, levando em consideração a tecnologia disponível no momento do processamento e desenvolvimentos tecnológicos. A GDPR define categorias especiais de dados pessoais (ou "dados confidenciais") como "dados pessoais revelando origem racial ou étnica, opiniões políticas, crenças religiosas ou filosóficas, ou associação sindical, e o processamento de dados genéticos, dados biométricos com o propósito de identificar exclusivamente uma pessoa natural, dados relativos à saúde ou dados relativos à vida sexual ou orientação sexual de uma pessoa natural". O processamento desta categoria de dados pessoais é proibido a menos que exceções se apliquem, como é o caso dos dados anonimizados.[102]

---

[101] Disponível em https://gdpr-text.com/pt/read/recital-26/#para_gdpr-r-026. Acesso em 28/09/2021
[102] Artigos 4 e 9 da GDPR (General Dataprotection Regualtion)

A pseudonimização é definida na GDPR como:

> O processamento de dados pessoais de tal forma que os dados pessoais não podem mais ser atribuídos a um determinado sujeito de dados sem o uso de informações adicionais, desde que tais informações adicionais sejam mantidas separadamente e estejam sujeitas a medidas técnicas e organizacionais para garantir que os dados pessoais não sejam atribuídos a uma pessoa natural identificada ou identificável."[103]

A regulação europeia afirma claramente que os dados pessoais que sofreram pseudoanonimização, e que poderiam ser atribuídos a uma pessoa física pelo uso de informações adicionais devem ser considerados como informações sobre uma pessoa natural identificável, sendo, portanto, objeto de das restrições de segurança e tratamento afetos aos próprios dados pessoais identificados ou identificáveis.

Ressalta-se ainda que o Regulamento europeu determina, em seu artigo 32, o dever de os agentes de tratamento de dados implementarem adequadas medidas técnicas e organizacionais de segurança informacional, tendo em vista a gradação ou nível de risco causado pela atividade de tratamento de informações pessoais a direitos e liberdades fundamentais dos titulares dos dados. A encriptação é expressamente mencionada pelo legislador europeu como modo de cumprimento de tal dever.

Quanto às técnicas de anonimização, foi criado um o grupo de trabalho focado no artigo 29[104] que criou um relatório de indicações sobre as melhores práticas para utilização das técnicas de anonimização como a supressão, a permuta, a generalização e a perturbação.[105]

Ainda, o relatório considera as restrições de tempo, custo, atualização tecnológica assim como a matriz de risco estudada neste trabalho.

> Neste parecer, o GT analisa a eficácia e os limites das técnicas de anonimização existentes no contexto jurídico comunitário de proteção de dados e apresenta recomendações para lidar com essas técnicas, tendo em conta o risco residual de identificação inerente a cada uma delas. O GT reconhece o valor potencial da anonimização de dados pessoais, em particular enquanto estratégia para colher os benefícios dos «dados abertos» para as pessoas e a sociedade em geral, reduzindo, simultaneamente, os riscos para as pessoas em causa. No entanto, alguns estudos de caso e publicações científicas demonstraram a dificuldade de criar um conjunto de

[103] Artigos 4 e 5 da GDPR (General Dataprotection Regualtion)
[104] ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 5/2014 on Anonymisation Techniques. Bruxelas.
[105] https://ec.europa.eu/justice/article-29/documentation/opinion-ecommendation/files/2014/wp216_en.pdf

> dados verdadeiramente anónimo que mantenha simultaneamente as informações subjacentes suficientes exigidas para a tarefa em questão.[106]

Em complemento à GDPR, o Regulamento (UE) 2018/1807[107], de 14 de novembro de 2018, faz a distinção de dados pessoais e dados não pessoais e faz alusão à atenção pela análise circunstancial e sobre a atualização tecnológica:

> A internet das coisas, a inteligência artificial e a aprendizagem automática, que estão em expansão, representam grandes fontes de dados não pessoais, por exemplo, em consequência da sua utilização em processos automatizados de produção industrial. Exemplos concretos de dados não pessoais incluem conjuntos de dados agregados e anonimizados utilizados para a análise de grandes volumes de dados, os dados relativos à agricultura de precisão que podem ajudar a controlar e a otimizar a utilização de pesticidas e de água ou ainda dados sobre as necessidades de manutenção de máquinas industriais. Se os progressos tecnológicos permitirem transformar dados anonimizados em dados pessoais, esses dados devem ser tratados como dados pessoais, e o Regulamento (UE) 2016/679 deve ser aplicado em conformidade.[108]

A União Europeia admite existir um limite móvel entre os “dados pessoais” e os “dados não pessoais”, onde os dados são entendidos com uma tendência expansiva à medida que a tecnologia o permita. O que exige uma atitude de prevenção e de precaução permanentes por parte de quem assume beneficiar do respetivo tratamento, com os inerentes riscos e sem exclusão das respectivas responsabilidades

6. REGULAMENTAÇÃO FRANCESA

A França faz parte dos países europeus onde surgiram as primeiras leis de proteção de dados. Sendo um país de vanguarda no que concerne às políticas públicas e regulamentações sobre a conceituação, o uso e o tratamento de dados pessoais.

De fato, a lei número 78-17, promulgada em 6 de janeiro de 1978[109], chamada *“loi Informatique et Libertés*” ou lei da informática e das liberdades, ou somente LIL, traz a conscientização das questões que envolvem o processamento de dados pessoais e ainda está em vigor, embora tenha sofrido várias modificações profundas.

---

[106] Ibidem, p. 3
[107] Regulamento (UE) 2018/1807 do Parlamento Europeu e do Conselho.
[108] Ibidem, item (9)
[109] A lei n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés - Légifrance pode ser acessada em https://www.legifrance.gouv.fr/

Tem como objetivo proteger os direitos dos indivíduos, regulando as possibilidades de processamento de dados pessoais, em especial por meio da criação de uma Comissão Nacional de Informática e Liberdades (CNIL), autoridade de dados francesa.

Em 2004, sob as regulações da diretiva europeia de 24 de outubro 1995[110] sobre a proteção de dados de caráter pessoal, a lei foi emendada. O texto aumentou os poderes do CNIL no que diz respeito a verificações e sanções in loco. Também criou o "Correspondent Informatique et Libertés" (CIL). São profissionais, que dentro de sua organização (empresa, administração ou autoridade local), garantem o cumprimento da Lei de Proteção de Dados.[111]

Na década seguinte, a GDPR harmoniza consideravelmente os textos nacionais ao restabelecer os principais princípios da proteção de dados pessoais, deixando espaço para manobras aos Estados-Membros sobre cerca de cinquenta pontos. Como a lei europeia é um padrão mais alto do que a lei francesa, torna a grande maioria da lei francesa vigente nula.

Assim, em 20 de junho de 2018, o Parlamento aprovou a Lei nº 2018-493[112] sobre a proteção de dados pessoais. O legislador optou por manter a Lei de Proteção de Dados, mas a modificou profundamente com base em todos os elementos da GDPR, podendo citar:

- A idade de consentimento ao uso dos dados pessoais é fixada em 15 anos;
- A possibilidade de os agentes do CNIL realizarem verificações on-line sobre a identificação de dados pessoais;
- A possibilidade do CNIL ser consultado pelo presidente da Assembleia Nacional, pelo presidente do Senado ou pelas comissões competentes da Assembleia Nacional e do Senado, bem como a pedido de um presidente do grupo parlamentar sobre qualquer projeto de lei relativo à proteção de dados pessoais;

---

[110] A diretiva europeia de 24 de outubro 1995 sobre a proteção de dados de caráter pessoal, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31995L0046

[111] Lei n° 2004-801 de 6 de outubro 2004 pode ser acessada em https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000441676/#:~:text=Dans%20les%20r%C3%A9sum%C3%A9sLoi%20n%C2%B0%202004%2D801%20du%206%20ao%C3%BBt%202004%20relative,et%20aux%20libert%C3%A9s%20(1).

[112] Lei nº 2018-493 https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037085952.

- A possibilidade de os indivíduos serem representados por associações e organizações já autorizadas a representar diversos indivíduos em ações em grupo, em suas reclamações ao CNIL;
- A obrigação do controlador, no caso de decisões administrativas individuais, incluindo processamento algorítmico, garantir controle do processamento algorítmico e suas evoluções, a fim de poder explicar, detalhadamente e de forma inteligível, aos dados o sujeito da forma como o processamento foi implementado em relação a ele.

Enfim, assim como a Portaria de 12 de dezembro de 2018, foi editada uma lei em 2019, a lei 2019-536, de 29 de maio, que também objetivou melhorar a legibilidade do marco legal nacional e alinhar as disposições regulatórias com o direito europeu e as medidas legislativas nacionais tomadas para implementá-lo. Além disso, esse texto adapta certas regras processuais perante a Comissão Nacional de Informática e Liberdades. Retorna em especial à autorização dos agentes dos serviços do CNIL no controle da implementação dos tratamentos, possibilitando o controle *in loco*.

6.1 A AUTONOMIA DA CNIL

Percebe-se que a autoridade de dados francesa (CNIL), criada ainda 1978, foi fortalecida em seu papel regulador e fiscalizatório em um esforço de adequação às normas da GDPR, a qual a França deve se vincular por ser um país membro da União Europeia.

No que tange à anonimização de dados, assim como pseudoanonimização, a legislação francesa também confere a sua autoridade de dados (CNIL) a análise e recomendações:

> A Comissão nacional de l'informatique et des libertés é uma autoridade administrativa independente. Será a autoridade nacional de supervisão dentro do significado e para a aplicação do Regulamento (UE) 2016/679 de 27 de Abril de 2016. Suas missões são as seguintes:
> 1 ° Informa todas as pessoas em causa e todos os controladores de seus direitos e obrigações e pode, para isso, fornecer informações às autoridades locais, seus grupos e pequenas e médias empresas;
> 2 ° Assegura que o processamento de dados pessoais seja implementado de acordo com as disposições desta lei e outras disposições relativas à proteção de dados pessoais previstas em leis e regulamentos, direito da União

Europeia e compromissos internacionais da França.[113]

A força, autonomia e independência da autoridade de dados francesa foi construída ao longo das últimas décadas a sua composição demonstra isso.[114]

Os dezessete membros que formam a comissão são, em sua maioria, eleitos pelas assembleias e jurisdições a que pertencem.

- O presidente do CNIL é nomeado por decreto do Presidente da República.
- 4 parlamentares (2 membros da Assembleia, 2 Senadores);
- 2 membros do Conselho Econômico, Social e Ambiental francês;
- 6 representantes de jurisdições elevadas (2 membros do Conselho de Estado, 2 membros do Tribunal de Cassação e 2 membros do Tribunal de Contas;
- 5 figuras públicas qualificadas indicadas por: o Presidente da Assembleia Nacional (1 figura pública), o presidente do Senado (1 figura pública), o Gabinete francês (3 figuras públicas). O mandato dos comissários é de 5 anos, ou, para os parlamentares, desde a duração de seu mandato;
- Presidente do CADA (Comissão de Acesso a Documentos Administrativos).

Os membros do CNIL realizam uma sessão plenária uma vez por semana, de acordo com uma agenda definida pela Presidência. Parte substancial dessas sessões é dedicada à revisão de projetos de lei e decretos apresentados pelo governo para um parecer oficial do CNIL. Analisa as consequências das novas tecnologias na vida privada dos cidadãos.

Além disso, a CNIL possui um comitê restrito com 5 membros e um presidente diverso do da comissão. O comitê pode impor várias sanções aos controladores de dados que não cumprem a lei. Após a adequação à GDPR, a quantidade de sanções financeiras pode chegar a até 20 milhões de euros, ou para empresas até 4% de seu faturamento anual global. [115]

---

[113] Lei de informática e liberdades, Chapitre II : La Commission nationale de l'informatique et des libertés, artigo 8.

[114] https://www.cnil.fr/fr/status-composition, acesso em 05/10/201

[115] ibidem

O poder fiscalizatório e autonomia da Comissão ficaram explícitos em 12 de janeiro de 2021, quando sancionou o próprio ministério do interior francês por ter utilizado ilegalmente drones equipados com câmeras, em especial para monitorar o cumprimento das medidas de contenção. Ele orienta o ministério a parar todos os voos de drones até que um quadro normativo o autorize.[116]

6.2 A DEFINIÇÃO DE DADOS PESSOAIS PELA CNIL

Muito embora e, conforme já explanada neste trabalho, a identificação de dados pessoais tenha um conceito expansionista e o cruzamento de dados e informações possa possibilitar a identificação de um pessoal natural, a CNIL, em seu guia, de maneira mais didática indica o que são dados pessoais.

Para a comissão, são dados pessoais: sobrenome, primeiro nome, pseudônimo, data de nascimento, fotos, gravações de som de vozes, telefone fixo ou celular, endereço postal, endereço de e-mail, endereço IP, ID de login do computador ou ID de cookies, impressão digital, rede venosa ou palmar da mão, impressão de retina, número da placa, número da previdência social, número de identificação, dados de uso de aplicativos, comentários, entre outros.

Ainda, ela explica a questão do dado identificável ao relatar que a **identificação de pessoas naturais pode ser** realizada a partir de um único pedaço de dados (exemplos: sobrenome e primeiro nome) ou a partir do cruzamento de um conjunto de dados (exemplo: uma mulher vivendo em um determinado endereço, nascida em um determinado dia e membro de uma determinada associação).[117]

A CNIL segue a ideia de que alguns dados são considerados particularmente sensíveis e lembra que a GDPR proíbe a coleta ou o uso desses dados, a menos que, em particular, o sujeito de dados tenha dado seu consentimento expresso (ativo, explícito e preferencialmente escrito, que deve ser livre, específico e informado).

6.3 O GUIA SOBRE ANONIMIZAÇÃO DA CNIL

[116] https://www.cnil.fr/fr/drones-la-cnil-sanctionne-le-ministere-de-linterieur, acesso em 07/05/2021
[117] https://www.cnil.fr/fr/identifier-les-donnees-personnelles, acesso em 07/05/2021

A GDPR aplica-se apenas a dados pessoais, ou seja, informações relacionadas a um indivíduo identificado ou identificável. Dados não pessoais, incluindo dados pessoais que foram anonimizados, estão fora do escopo do GDPR.

Assim, as autoridades de dados dos países europeus atualizaram suas orientações sobre a anonimização no âmbito do GDPR.

Nesse sentido, a CNIL, em maio de 2020, divulgou orientações[118] para esclarecer como a identificação deve ser compreendida e avaliada no âmbito da regulamentação europeia. A orientação da CNIL está alinhada com a regulação (EU) 2016/679 (GDPR) e fornece sugestões práticas para a aplicação de métodos de anonimização.

A Comissão, ao sugerir tais práticas atualizadas sob a GDPR, permitiu que as organizações não precisem mais buscar esclarecimentos fora de sua jurisdição para garantir sua conformidade com o GDPR e considerar as melhores práticas.

6.3.1 O CONCEITO E OS CRITÉRIOS DE ANONIMIZAÇÃO CONFORME A CNIL

No geral, a CNIL adere às normas e aos critérios anteriores de anonimização estabelecidos pelo Grupo de Trabalho do artigo 29º[119]. Define a anonimização como "processamento que consiste em utilizar um conjunto de técnicas para impossibilitar, na prática, identificar a pessoa por qualquer meio e irreversivelmente"[120]. Acrescenta que a anonimização visa à eliminação de qualquer possibilidade de reidentificação. A anonimização é, portanto, apresentada como um estado irreversível e absoluto.

A autoridade francesa de proteção de dados em sua orientação de maio de 2020, descreve diferentes técnicas de anonimização e distingue entre pseudoanonimização e anonimização. Além disso, a orientação destaca que, para a construção de um processo de anonimização, é aconselhável identificar as informações a serem mantidas de acordo com sua relevância, remover os elementos

[118] https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles, acesso em 07/05/2021

[119] ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 5/2014 on Anonymisation Techniques.

[120] https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles, acesso em 03/05/2021

de identificação direta, bem como os dados que poderiam permitir uma fácil reidentificação das pessoas, distinguir informações importantes de informações secundárias ou inúteis que podem ser excluídas, e definir a granularidade ideal e aceitável de cada pedaço de dados armazenados.

A CNIL especifica os critérios cumulativos sobre os quais deve ser feita uma avaliação sobre se os dados foram anonimizados com sucesso:

- Individualização; Não deve ser possível isolar um indivíduo em um conjunto de dados.
- Correlação; Não deve ser possível vincular conjuntos separados de dados relativos ao mesmo indivíduo.
- Inferência; Não deve ser possível inferir, quase com certeza, novas informações sobre um indivíduo.

Se esses três critérios não forem totalmente atendidos, os agentes de tratamentos de dados devem realizar uma avaliação aprofundada dos riscos de identificação para demonstrar que "o risco de reidentificação com meios razoáveis é zero"[121]. Deve-se, também, considerar a realização de verificações regulares para monitorar e verificar se as informações tratadas permanecem anônimas.

Por fim, a Orientação observa que se um conjunto de dados publicado online como 'anônimo' realmente contém dados pessoais, e nenhuma das exceções mencionadas no artigo L.312-1-2 do Decreto nº 2018-1117 de 10 de dezembro de 2018 relativo às categorias de documentos administrativos que podem ser tornados públicos sem ser objeto de um processo de anonimização não é aplicável, então tal publicação será considerada uma violação de dados.

6.3.2 OS MÉTODOS DE ANONIMIZAÇÃO SUGERIDOS

A CNIL classifica os métodos de anonimização de dados como enquadrados em duas categorias principais (nesse sentido a comissão alinha-se com o GDPR e o grupo de trabalho do artigo 29):

- Aleatorização: o conceito da comissão consiste em modificar os atributos

[121] ibidem

em um conjunto de dados para que sejam menos precisos, preservando a distribuição geral.

- Generalização: A CNIL define a generalização como modificar a escala dos atributos dos conjuntos de dados, ou sua ordem de magnitude, a fim de garantir que eles sejam comuns a um conjunto de pessoas. Este método pode abordar o risco de individualização e correlação.

Ainda, em seu guia, a autoridade francesa, para identificar o risco de identificação e o método de anonimização adequado, sugere que devem ser considerados os seguintes elementos:

- As categorias de dados pessoais a serem anonimizadas e armazenadas;
- A identificação e remoção de identificadores diretos, o que poderia permitir a fácil identificação dos sujeitos de dados envolvidos;
- A distinção entre informações cruciais e informações secundárias ou desnecessárias e a exclusão desta última;
- O objetivo do processamento de dados dos conjuntos de dados anonimizados.

A orientação parece negligenciar as habilidades de tecnologias interconectadas e orientadas por inteligência artificial que desafiam o conceito de anonimização irreversível, conceito que a GDPR enfatiza como a preocupação com as atualizações tecnológicas.

Quanto à pseudonimização, a CNIL esclarece que pseudônimo é a retenção dos dados originais e a produção de conjuntos de dados anonimizados. Refere-se ao processamento de dados pessoais realizado de tal forma que os dados relativos a uma pessoa física não podem mais ser atribuídos sem recurso a informações adicionais.

Em sua orientação, a Comissão afirma que a GDPR insiste que essas informações adicionais devem ser mantidas separadamente e estar sujeitas a medidas técnicas e organizacionais, a fim de evitar a reidentificação dos sujeitos dos dados. Ao contrário da anonimização, a pseudônimo é um processo reversível.

Ela retoma o conceito da própria regulação europeia ao dizer que, na prática, um processo de pseudoanonimização consiste em substituir os dados de identificação direta (sobrenome, primeiro nome,etc.) de um conjunto de dados com dados de

identificação indireta (nome de usuário, número em um ranking, entre outros) a fim de reduzir sua sensibilidade. Eles podem resultar de um *hash* criptográfico de dados de indivíduos, como seu endereço IP, ID do usuário, endereço de e-mail.

Por fim, a CNIL, mais uma vez em consonância com o comitê europeu, transcreve:

> Os dados resultantes da pseudônimo são considerados **dados pessoais e, portanto, permanecem sujeitos às obrigações do GDPR.** No entanto, a regulamentação europeia incentiva o uso da pseudônimo no contexto do processamento de dados pessoais. Além disso, o GDPR considera que a pseudônimo reduz os riscos para os sujeitos de dados e contribui para o cumprimento do Regulamento.[122]

7. REGULAMENTAÇÃO BRASILEIRA

7.1 CONTEXTO HISTÓRICO DA PROTEÇÃO DE DADOS PESSOAIS

No Brasil, há alguns anos discute-se a proteção de dados pessoais como direito fundamental. Para os estudiosos, um fator extremamente preocupante é a garantia de real controle dos dados pelo cidadão. Nesse sentido, Laura Schertel Mendes defende que a proteção de dados pessoais deve ter status de direito fundamental e afirma que:

> A tutela jurídica para a proteção de dados da personalidade em face do tratamento de dados pessoais envolve o estabelecimento de uma série de procedimentos, princípios e direitos, que limitam o processamento de dados pessoais ao mesmo tempo que empoderam o cidadão para controlar o fluxo de seus dados [123]

Para a autora, função precípua da proteção de dados não é proteger os dados *per se,* mas garantir proteção à pessoa que é o titular desses dados.

Inspirada na GDPR e criada para estabelecer regras mais claras e transparentes ao tratamento de dados pessoais em nosso país, a LGPD (lei geral de proteção de dados pessoais) traz a definição de dado pessoal em seu art. 5º

> Art. 5º Para os fins desta Lei, considera-se:
> I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;
> II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida

[122] https://www.cnil.fr/fr/identifier-les-donnees-personnelles, acesso em 07/05/201
[123] (MENDES, 2014, p. 189).

> sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;
> III - dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;
> IV - banco de dados: conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico;[124]

Para o marco regulatório da proteção de dados no Brasil, dado pessoal é a informação relacionada a pessoa natural identificada ou identificável e seu âmbito de proteção engloba o processamento e tratamento de dados realizado em meio físico ou digital, seja por pessoa natural ou jurídica, de direito público ou privado.

A evolução do conceito de proteção de dados pessoais, a partir do direito à privacidade, levou a doutrina à criação de um direito autônomo em vários países, alguns inclusive, dando-lhe o status de direito fundamental. Uma decisão importante tendo em vista o atual fenômeno da ubiquidade[125], por meio do qual, o processamento dos dados tornou-se onipresente, perpassando atualmente todas as áreas da vida humana[126].

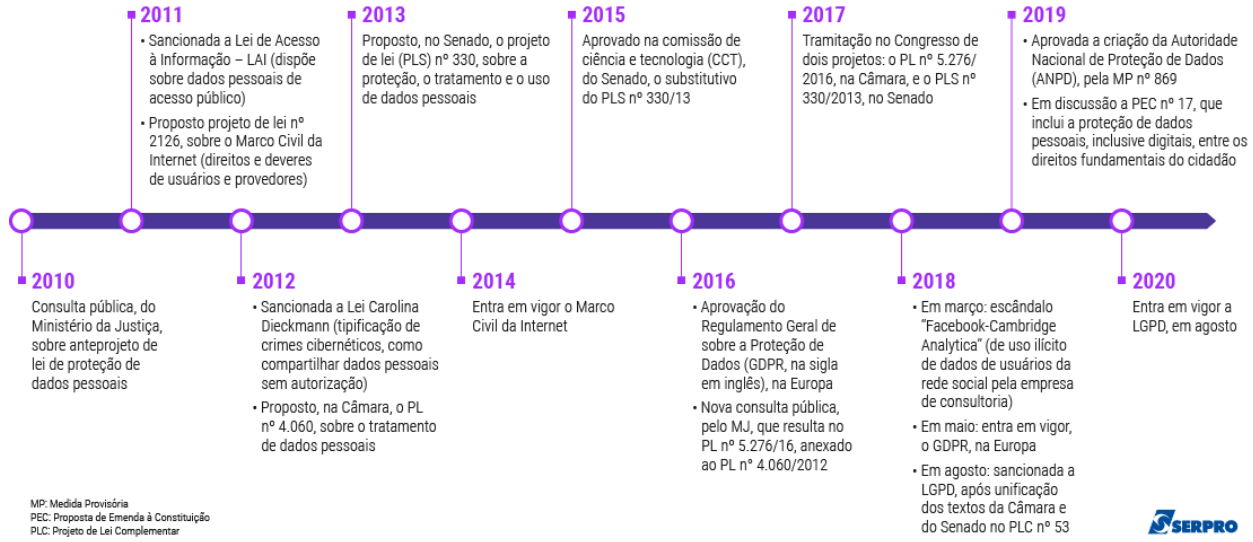
O panorama e a construção histórica das leis de proteção de dados pessoais pelo mundo, especificamente no continente europeu, demonstram quase 50 anos de atraso legislativo do Brasil em relação ao tema. No país, o direito à proteção de dados pessoais encontra-se tutelado de forma reflexa em nossa Constituição da República, a partir da interpretação conjunta dos artigos 1º, III; 3º, I e IV, 5º, X, XII e LXXII. Porém, diante na nova realidade trazida pela evolução tecnológica do século XXI, alinhada ao fluxo transfronteiriço de dados pessoais ao redor do globo, exige-se que haja uma evolução normativa, tornando-se essencial uma adequação do nosso ordenamento jurídico para que a proteção de dados pessoais seja resguardada constitucionalmente.

No contexto histórico, vale salientar as iniciativas regulatórias ocorridas no Brasil que culminaram na promulgação da lei nº 13.709, de 14 de agosto de 2018**, a lei geral de proteção de dados pessoais do Brasil:**

---

[124] Lei geral de proteção de dados, Lei nº 13.709, de 14 de agosto 2018
[125] MENDES, 2014, p. 78-79
[126] HOFFMANN-RIEM, 2008, p. 1010, *apud* MENDES, 2014, p. 79

Figura 3 -linha do tempo das iniciativas legislativas para a regulação da proteção de dados no Brasil[127]



Fonte: SERPRO, 2021

7.2 A LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS

Com a sanção da LGPD, defende-se que a proteção de dados pessoais deve ser tratada em nosso país como um direito autônomo que é, não podendo apenas ser tratada somente sob o aspecto da privacidade, mas como um direito fundamental independente, como já se faz em outros países[128]. Tal entendimento não obsta, por exemplo, o reconhecimento da fundamentalidade da proteção de dados através do princípio da dignidade da pessoa humana como fonte de direitos materialmente fundamentais não expressos na Constituição.

Conforme análise de Mendes e Doneda, é possível identificar 5 eixos principais para a articulação da lei em torno doa proteção dos dados da pessoa:

i. Unidade e generalidade da aplicação da Lei;
ii. legitimação para o tratamento de dados (hipóteses autorizativas);
iii. princípios e direitos do titular;
iv. obrigações dos agentes de tratamento de dados;
v. responsabilização dos agentes.[129]

---

[127] https://www.serpro.gov.br/lgpd/menu/arquivos/linha-do-tempo-1/view
[128] Outras iniciativas legislativas brasileiras são a (PEC) 17/2019 e Medida Provisória n° 954, de 2020
[129] REFLEXÕES INICIAIS SOBRE A NOVA LEI GERAL DE PROTEÇÃO DE DADOS, revista do Direito do Consumidor, vol. 120/2018 | p. 469 - 483 | Nov. - Dez / 2018

A LGPD só protege explicitamente os dados pessoais de pessoas físicas. Portanto, os dados das pessoas jurídicas também não são cobertos. Em seu artigo 5, ela esclarece que um sujeito de dados é uma pessoa natural a quem se referem os dados pessoais objeto de processamento[130].

A lei brasileira aplica-se a controladores de dados e processadores de dados, juntos chamados de agentes de tratamento, que podem ser empresas, órgãos públicos, instituições, bem como organizações sem fins lucrativos. A LGPD define um controlador como pessoa natural ou jurídica que é responsável por tomar decisões sobre o processamento de dados pessoais e define um operador como pessoa física ou jurídica, de direito público ou privado, que trata dados pessoais em nome do controlador.

A LGPD refere-se a qualquer operação de tratamento, definida como qualquer operação realizada com dados pessoais, como coleta, produção, recebimento, classificação, uso, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, exclusão, avaliação ou controle das informações, modificação, comunicação, transferência, divulgação ou extração.

Vale ressaltar a definição de "dados pessoais sensíveis" como dados pessoais relativos à origem racial ou étnica, crença religiosa, opinião política, associação sindical ou religiosa, filosófica ou política, dados relativos à saúde ou vida sexual, dados genéticos ou biométricos, quando relacionados a uma pessoa física.[131]

7.3 O CONCEITO E OS CRITITÉRIOS DE ANONIMIZAÇÃO CONFORME A LGPD

Em 28 de janeiro de 2015, o Ministério da Justiça lançou consulta pública para discutir a proteção de dados pessoais armazenados tendo como base o anteprojeto de lei que veria a se tornar o marco civil da proteção de dados no Brasil.

A preocupação do ministro da justiça à época ao lançar a consulta pública ressalta a questão do processamento de dados pessoais:

> [...} o que se protege é o dado da pessoa e suas referências pessoais. É claro que tem um ponto de ligação. Por exemplo, os dados pessoais que

[130] Lei geral de proteção de dados, LEI Nº 13.709, DE 14 DE AGOSTO DE 2018, artigo 13
[131] Ibidem, artigos 3-5, 11 -12

> você tem cadastrados numa loja podem ser passados para alguém ou negociados? Os dados do governo podem ser cedidos? Esse é o enfoque do projeto.[132]

Tal consulta teve um grande debate entre os interessados no próprio conceito de dado pessoal e dado anonimizado.[133]

Empresas de telecomunicações, associações de indústrias e tecnologia e associações de empresas de tecnologia da informação[134] defenderam que dados anonimizados não deveriam ser considerados pessoais.

Em contraponto, interessados na defesa do consumidor e na defesa da privacidade, centros de pesquisa de linhas acadêmicas e pesquisadores da área[135] defenderam que os dados anonimizados deveriam ser considerados dados pessoais pois sempre existiria a possibilidade de serem reidentificados.

A definição de dados pessoais (e por extensão, a de dados anônimos) é particularmente importante pois ela define o escopo da lei. Durante a consulta pública, houve intenso embate entre aqueles que buscavam expandir esta definição, de forma a incluir explicitamente dados anonimizados, e aqueles que buscavam restringi-la, de forma a deixar claro que dados anônimos não deveriam ser considerados dados pessoais.

De fato, o debate terminou por culminar nos entendimentos dos artigos 12 da lei:

> 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.
> § 1º A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios.
> § 2º Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.
> § 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua

---

132 https://www.camara.leg.br/noticias/449278-consulta-publica-sera-base-para-projeto-de-lei-sobre-protecao-de-dados-pessoais/, acesso em 05/10/2021

133 http://pensando.mj.gov.br/dadospessoais/texto-em-debate/anteprojeto-de-lei-para-a-protecao-de-dados-pessoais/, acesso em 05/10/2021

134 Empresas como a Cisco, Abranet(associação brasileira de internet, a BRasscom (Associação Brasileira das Empresas de Tecnologia e da Informação e Comunicação) e a ITI (Information Techonlogy Industry Council)

135 Organizações e pesquisadores como: Centro de Tecnologia e Sociedade da FGV - Direito/RIO, Bruno B. Bioni, GPoPAI-USP, Intervozes, PROTESTE e Coletivo Antivigilância

> segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.[136]

A lei parece ter assumido um meio termo para atender aos anseios dos atores da consulta pública, ela define dados pessoais como informações relacionadas a uma pessoa natural identificada ou identificável, contudo exclui os dados anonimizados, que são definidos como dados relativos a um sujeito que não pode ser identificado, direta ou indiretamente, considerando o uso de meios técnicos razoáveis disponíveis no momento do processamento[137].

De acordo com a LGPD, os dados não são considerados anonimizados quando o processo de anonimização ao qual foram submetidos é reversível, considerando esforços razoáveis. No entanto, além dos critérios objetivos (custo, tempo e tecnologia disponível), a LGPD também adota um conceito subjetivo, que é o "uso de recursos próprios" do controlador ou processador para determinar se os esforços feitos para o processo de anonimização foram razoáveis.

Vale ressaltar que a lei brasileira difere da lei europeia no § 2º ao abrir uma exceção podem ser considerados dados pessoais aqueles utilizados para formação de perfil comportamental de determinada pessoa natural, se identificada ainda que o tratamento tenha sido iniciado com origem em dados anonimizados.

Esse ponto não é explorado pela legislação europeia e mostra a preocupação da regulamentação brasileira com a formação e utilização de perfis comportamentais que possam identificar uma pessoa.

Mendes e Machado atentam para um exemplo em que a formação de perfis comportamentais com base em dados anônimos, caso identificados, passam a ser escopo da lei assim como os dados pessoais:

> [...] anonimizar as informações pessoais para então processá-las e minerá-las para formar modelos estatísticos ou perfis personalizados ou de grupo (group profiling) – v.g., modelos preditivos de mobilidade populacional baseado em dados agregados de localização de usuários de telefone celular. (ii), visto que cuida de uma particularidade do assunto versado no caput da disposição legal. Significa isso dizer que, dados a princípio anonimizados – portanto, qualificados preliminarmente como dados não pessoais – se utilizados para “formação de perfil comportamental de determinada pessoa, se identificada” hão de ser considerados dados pessoais para fins de aplicação da LGPD.[138]

---

[136] Lei geral de proteção de dados, LEI Nº 13.709, de 14 de agosto de 2018, artigo 12

[137] Ponto em que a LGPD se aproxima da GDPR e, também, das diretivas da CNIL quanto às limitações de um possível processo de reidentificação.

[138] TECNOLOGIAS DE PERFILAMENTO E DADOS AGREGADOS DE GEOLOCALIZAÇÃO NO

No que tange à pseudoanonimização, a lei brasileira entende, aos moldes da europeia, como o processo pelo qual os dados não podem mais ser associados direta ou indiretamente a um indivíduo, exceto usando informações adicionais mantidas separadamente pelo controlador em um ambiente controlado e seguro.[139]

Contudo, A LGPD não afirma explicitamente que os dados pseudônimos devem ser considerados como dados pessoais como a GDPR o faz expressamente.

No entanto, é possível interpretar como tal, uma vez que a definição dada indica a possibilidade de reidentificação, sendo escopo do artigo 12.

Em outra característica díspar em relação ao entendimento europeu, A LGPD prevê que, em estudos relacionados à saúde pública, entidades de pesquisa que tenham acesso a dados pessoais devem processar os dados exclusivamente dentro da entidade e estritamente com a finalidade de realizar estudos e levantamentos em ambiente controlado e seguro, incluindo, sempre que possível, a anonimização ou pseudônimo dos dados.

Uma entidade de pesquisa é definida na LGPD como um órgão ou pessoa jurídica do governo direto (federal, estadual e local e seus órgãos correlatos sem personalidade jurídica) ou indireta (agências, fundações, empresas estatais, etc. que tenham personalidade jurídica própria) órgãos governamentais ou uma pessoa jurídica sem fins lucrativos de direito privado, legalmente estabelecida pela lei brasileira, com sede e jurisdição no Brasil, que inclua em sua missão institucional ou em seus propósitos estatutários o objetivo de fazer pesquisas básicas ou aplicadas de natureza histórica, científica, tecnológica ou estatística.

O fato é que a regulamentação brasileira ainda não traçou os limites da anonimização assim como intentaram a GDPR, o grupo de estudos do artigo 29, o Parecer 05/2014 sobre Técnicas de Anonimização e própria CNIL francesa. Assim, os operadores do Direito buscam subsídios na experiência europeia derivada da aplicação de sua legislação sobre proteção de dados, que serviu de inspiração para a LGPD.

COMBATE À COVID-19 NO BRASIL: UMA ANÁLISE DOS RISCOS INDIVIDUAIS E COLETIVOS À LUZ DA LGPD, Diego Carvalho Machado, Laura Schertel Mendes, 2020, página 136

[139] Lei geral de proteção de dados, LEI Nº 13.709, de 14 de agosto de 2018, artigo 13

Espera-se que tal lacuna seja preenchida por estudos da Autoridade Nacional de Proteção de Dados brasileira (ANPD), cujos detalhes de sua instituição e competências são trabalhados na próxima seção.

7.2 A AUTORIDADE DE PROTEÇÃO DE DADOS BRASILEIRA

> 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. [...]
> § 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.[140]

Assim como a regulamentação francesa, a legislação brasileira preocupou-se em instituir uma autoridade nacional de proteção de dados atribuições similares à CNIL.

Assim, em seu capítulo X, a LGPD trata da Autoridade Nacional de Proteção de Dados (ANPD) e do Conselho Nacional de Proteção de Dados[141]

A composição da ANPD é definida pela lei:

> Art. 55-C. A ANPD é composta de:
> I - Conselho Diretor, órgão máximo de direção;
> II - Conselho Nacional de Proteção de Dados Pessoais e da Privacidade;
> III - Corregedoria;
> IV - Ouvidoria;
> V - órgão de assessoramento jurídico próprio;
> VI - Unidades administrativas e unidades especializadas necessárias à aplicação do disposto nesta Lei. [142]

A ANPD terá autoridade para emitir sanções por violações à LGPD. O Conselho da ANPD tem autoridade para, entre outras coisas:

- Supervisionar a proteção de dados pessoais;
- Emitir regulamentos e procedimentos relacionados à proteção de dados pessoais;

---

[140] Lei geral de proteção de dados, LEI Nº 13.709, DE 14 DE AGOSTO DE 2018, artigo 12
[141] Idem, artigos 55-58
[142] Idem art. 55

- Deliberar, em nível administrativo, sobre a interpretação da LGPD e assuntos omitidos em sua redação;
- Supervisionar e aplicar sanções em caso de processamento de dados realizado em violação à legislação;
- Implementar mecanismos simplificados para registrar reclamações sobre o processamento de dados pessoais em violação da LGPD.

Além disso, o Conselho da ANPD é responsável, entre outras funções:

- Propor diretrizes estratégicas e subsídio para a criação da Política Nacional de Proteção de Dados Pessoais e para o funcionamento da ANPD;
- Sugerir ações a serem realizadas pela ANPD;
- Elaborar estudos e realizar debates e audiências públicas sobre a proteção de dados pessoais.

O artigo 12 da LGPD diz que os dados anonimizados não serão considerados dados pessoais sob a LGPD, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.

Portanto, o processamento de dados anonimizados está fora do escopo da lei

O parágrafo terceiro da regulamentação brasileira delega à ANPD a autoridade para emitir normas e técnicas de anonimização e verificar sua segurança, após consulta ao Conselho Nacional de Proteção de Dados.

A regulamentação sobre os processos seguros de anonimização de dados deve ser incentivada e intensamente disseminada pela autoridade de dados brasileira, pois é fundamental para a inovação e para a economia digital.

Tal fato permitirá não apenas o processamento de dados em geral, mas também as organizações continuarem tratando e extraindo valor a partir de contextos fora dos propósitos originais para os quais os dados foram coletados. A ANPD deve fornecer padrões e técnicas de anonimização o mais rápido possível, bem como emitir orientação sobre o que constitui "meios razoáveis” para os processos de reidentificação de dados anônimos.

8. CONSIDERAÇÕES FINAIS

O presente trabalhou buscou contribuir com o debate sobre o conceito de dado pessoal e dado anonimizado e como essa conceituação influenciou as regulamentações de proteção aos dados pessoais, primeiro na Europa, berço da regulamentação mais madura sobre guarda, processamento e tratamento de dados, a GDPR. Posteriormente, foi realizada análise sobre as adequações do regulamento francês com base na lei europeia e nas estratégias da autoridade de dados do país gaulês para dar segurança aos interessados em tratar dados anonimizados.

Ao conceituar a anonimização e os processos atuais para sua produção, o texto analisou as possibilidades e riscos inerentes a uma possível reidentificação de dados considerados anônimos.

De fato, os dispositivos legais estudados neste documento são unânimes em excluir das restrições impostas ao tratamento de dados pessoais, os dados considerados anônimos, ou de não possibilidade de identificação da pessoa natural.

Contudo, conforme relatado, pesquisadores entendem que não há como haver um processo plenamente seguro que garanta que o dado permanecerá anônimo, principalmente com as constantes atualizações tecnológicas e a possibilidade de cruzamento de informações com vários provedores de conteúdo.

Se a anonimização não é feita corretamente, pode-se sim argumentar que é fácil a reidentificação desses dados.

Assim, é necessário que as estratégias de anonimização sejam substituídas, constantemente, por modelos mais completos, que envolvam uma análise dos riscos e benefícios envolvidos no compartilhamento de dados, principalmente focando atenção a variáveis como o volumes de dados trabalhado, a natureza dos dados, a cadeia de atividade de tratamento de dados, o gerenciamento de identidade, a análise da atualização contínua sobre as técnicas e tecnologias de anonimização e as devidas cláusulas contratuais que tratem o tratamento dos dados pessoais e as possíveis responsabilizações.

Em acréscimo, é imperativa a análise se é razoável o esforço para reidentificação do dado, fato que é abordado em todas as regulamentações.

A razoabilidade considera o estado da arte da tecnologia, tanto ao momento da anonimização, quanto nos tempos seguintes, o custo estimado para uma possível reidentificação e o tempo necessário para execução. Todos esses fatores devem ser razoáveis.

Por fim, deve-se verificar quem é o agente de tratamento de dados e se ele dispõe de meios próprios para reidentificar o dado.

A complexidade das variáveis impostas para identificação de um dado seguramente anônimo exige análise e acompanhamento específico das autoridades de tratamento de dados.

De fato, a França, com base nos dispositivos e estudos da própria União Europeia, lançou um guia sobre definições e técnicas sobre anonimização por meio de sua autoridade (a CNIL) e fortaleceu o órgão, concedendo poderes para instruir processos, propor multas e responsabilizações aos agentes de tratamento de dados, empresas e até mesmo entidades públicas.

O Brasil, no entanto, não possui ainda tais definições por meio de sua ANPD. Enquanto tal esclarecimento não vier, os agentes de tratamento de dados deverão desenvolver políticas robustas de controle de anonimização, com o apoio dos respectivos encarregados de proteção de dados e de profissionais com conhecimento técnico, jurídico e regulatório, de sorte a mitigar os riscos de eventual descumprimento da LGPD, causando prejuízos aos cidadãos e riscos de elevadas multas.

Não somente as autoridades de dados dos países europeus, como a França, mas de todos os países interessados no tratamento dos dados pessoais com respeito à privacidade, honra e própria dignidade da pessoa, devem realizar acordos de cooperação técnica e legislativa.

O intercâmbio de experiências técnicas e jurídicas permitirá o acompanhamento dos contextos das técnicas de anonimização e, também, das tecnologias de reidentificação possíveis, com o objetivo de tornar as regulações mais robustas e seguras para o titular do dado.

9. REFERÊNICIAS BIBLIOGRÁFICAS

WARREN, Samuel; BRENDEIS, Louis. Harvard Law Review, v. 4, n. 5, 1890

SCHWAB Klaus*, Shaping the future of the fourth industrial revolution*. New York: Currency, 2018

SRNICEK Nick, Plataform capitalism. Cambridge: Polity.

ZUBOFF, Shoshana, Big *other*: capitalismo de vigilância e perspectivas para uma civilização de informação. In: Fernanda Bruno e outros, Tecnopolíticas da vigilância. São Paulo: Boitempo, 2018

OHM, Paul, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, Vol. 57, U of Colorado Law Legal Studies Research Paper No. 9-12. Boulder, EUA, 2010

MENEZES, Barreto. Dados pessoais: conceito, extensão e limites, LLM, book Revista de Direito Civil 2, 2018

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Acesso em: 17 nov. 2020.

MENDES, Laura. Revista dos Tribunais, Caderno Especial, novembro de 2019

TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA, processo C-615/13 , ClienthEart v PAN Europe, 30.

UNIÃO EUROPEIA. **Directiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de Outubro de 1995**. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A31995L0046>. Acesso em: 22 set 2021.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília DF, 1988.

MENDES, Laura Schertel. Transparência e privacidade: violação e proteção da informação pessoal na sociedade de consumo. Dissertação (Mestrado em Direito), Faculdade de Direito da Universidade de Brasília, Brasília, 2008

Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho (General Data Protection Regulation). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679>. Acesso em: 15 set. 2021.

SCHWARTZ, Paul M.; SOLOVE, Daniel J. The PII problem: privacy and a new concept of personally identifiable information. New York University Law Review, v. 86, p. 1871-1877, dec. 2011

ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 4/2007 on the concept of personal data**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf >. Acesso em: 03 out. 2021.

MACHADO, Diego, Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados, Revista dos Tribunais, Caderno Especial – novembro de 2018.

BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019. XXVII, 314 p., il, 24 cm. Inclui notas explicativas, bibliográficas, de jurisprudência e bibliografia. ISBN 9788530981686.

HOUAISS, Antônio. Dicionário Houaiss da Língua Portuguesa. Rio de Janeiro, Ed. Objetiva, 2001.

DICIO ONLINE. Disponível em: www.dicio.com.br. Acesso em: 13 de setembro de 2021.

DICIONÁRIO LARROUSSE, Disponível em: www.larousse.fr. Acesso em: 13 de setembro de 2021.

DICIONÁRIO DE CAMBRIDGE. Disponível em www.dictionary.cambridge.org.Acesso em 13 de setembro de 2021.

GAVISON, Ruth. **Privacy and the Limits of Law**. The Yale Law Journal, Vol. 89, No. 3 (Jan, 1980), pp. 421-471. The Yale Law Journal Company,1980. Disponível em: <http://www.jstor.org/stable/795891>. Acesso em: 15 ago. 2021.

DONEDA, Danilo (Org.). **A regulação da criptografia no direito brasileiro**. São Paulo: Revista dos Tribunais, 2018. ISBN 978-85-203-6888-6.

_____. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Editora Revista dos Tribunais, 2019.

BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. Biblioteca digital Jurídica do Superior Tribunal de Justiça, 2019

FRAZÃO, Ana. *Fundamentos da proteção dos dados pessoais – Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399.

Megavazamento expõe documentos e dados de mais de 220 milhões de brasileiros, Olhar Digital, 2021. Disponível em https://olhardigital.com.br/2021/07/29/seguranca/vazamento-dados-internet/. Acesso em 14/09/2021

SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 17. Item 9. Disponível em: <https://www.pdpc.gov.sg/help- and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 set. 2021.

SWEENEY, Latanya. Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000. Disponível em: <https://dataprivacylab.org/projects/identifiability/paper1.pdf?fbclid=IwAR2qeD2uXtZmaUUPQzAStNoO1n2pAQpfzzhnDYg4WPN-yhjZ8oTBHmpmhow>. Acesso em: 10 ago. 2021

SHMATIKOV, Vitaly. k-Anonymity and Other Cluster-Based Methods. University of Texas, CS 380S: Theory and Practice of Secure Systems. Class presentation slides. Disponível em: < https://www.cs.utexas.edu/~shmat/courses/cs380s_fall09/>. Acesso em: 15 set. 2021.

GOOGLE. **Como o Google Anonimiza os Dados**. In: Privacidade & Termos do Google. Disponível em: https://policies.google.com/technologies/anonymization?hl=pt-BR. Acesso em: 10 set. 2021.

XIAO, Xiaokui; TAO, Yufei. m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets. Hong Kong: Chinese University of Hong Kong, Department of Computer Science and Engineering, 2007.

PILLOW, Timothy. **A Review of Synthetic Tabular Data Tools and Models: Anonymization methods that are revolutionizing how we share data**. Medium: 02 jul. 2020. Disponível em: <https://towardsdatascience.com/a-review-of-synthetic-tabular-data-tools-and-models-d83b232aae25>. Acesso em: 12 ago. 2021.

NETFLIX INC. The Netflix Prize Rules. 2006. Disponível em:

<http://www.netflixprize.com/rules.html> Acesso em: 15 out. 2021.

MINISTÉRIO DA JUSTIÇA. **Anteprojeto de Lei para a Proteção de Dados Pessoais.** Disponível em: <http://pensando.mj.gov.br/dadospessoais/texto-em-debate/anteprojeto-de-lei-para-a-protecao-de-dados-pessoais/> Acesso em: 09 nov. 2016.

NARAYANAN, Arvind; SHMATIKOV, Vitaly. **Privacy and Security Myths and Fallacies of "Personally Identifiable Information"**. Communications of the ACM. Jun 2020, Vol. 53, No. 08.

NARAYANAN, A.; SHMATIKOV, V. **Robust De-anonymization of Large Datasets. SP'08 Proceedings of the 2008 IEEE Symposium on Security and Privacy**. p. 111-125. Washington, EUA, 2008. Disponível em:<https://www.cs.cornell.edu/~shmat/shmat_oak08netflix.pdf>, acessado em 27/09/2021

IMDB.COM INC. **IMDb Database Statistics**. Disponível em: http://www.imdb.com/stats Acesso em: 06 set. 2021.

OHM, P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, Vol. 57, U of Colorado Law Legal Studies Research Paper No. 9-12. Boulder, EUA, 2010

RUBINSTEIN, Ira; HARTZOG, Woodrow. Anonymization and risk. *91 Washington Law Review*, Washington, DF, v. 703, 2016; *NYU School of Law*, *Public Law Research Paper*, Nova York, n. 15-36. Disponível em: http://bit.ly/2TKMmo8. Acesso em: 29 set 2021

COPETTI, Rafael; CELLA, José Renato Gaziero. **Autoridade Nacional de Proteção de Dados: natureza jurídica e a Lei nº 13.079/2018.** Caderno de Direito, governança e novas tecnologias II do CONPEDI, 2018. Disponível em: http://conpedi.danilolr.info/publicacoes/34q12098/15d3698u/Rc9sMnwxjkLPh2o4.pdf. Acesso em: 05 ago. 2021.

GUIDI, Guilherme. **Modelos regulatórios para proteção de dados pessoais.** Coletânea de artigos do ITS Rio, 2017. Disponível em: https://itsrio.org/wp-content/uploads/2017/03/Guilherme-Guidi-V-revisado.pdf. Acesso em: 13 out. 2021.

TRIBUNAL DE JUSTIÇA DA UNIÃO EUROPEIA, Processo C-582/12 (Breyer v Bundesrepublik Deutschland), 19 out. 2016

Kurose, James F.; Ross, Keith W. (2013). **Redes de computadores e a internet: uma abordagem top-down,** 6ª ed. São Paulo: Pearson Education do Brasil.

Regulamento (UE) 2018/1807 do Parlamento Europeu e do Conselho, de 14 de novembro de 2018, relativo a um regime para o livre fluxo de dados não pessoais na União Europeia (Texto relevante para efeitos do EEE.), disponível em: https://eur-lex.europa.eu/eli/reg/2018/1807/oj#:~:text=REGULATION%20%28EU%29%202018%2F1807%20OF%20THE%20EUROPEAN%20PARLIAMENT%20AND,in%20the%20European%20Union%20%28Text%20with%20EEA%20relevance%29. Acesso em 29 set 2021.

Lei n° 78-17 de 6 de janeiro de 1978 **relative à l'informatique, aux fichiers et aux libertés**. Légifrance. Disponível em https://www.legifrance.gouv.fr/, acesso em 21 set. 2021

Lei n° 2004-801 du 6 août 2004, disponível em https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000441676/#:~:text=Dans%20les%20r%C3%A9sum%C3%A9sLoi%20n%C2%B0%202004%2D801%20du%206%20ao%C3%BBt%202004%20relative,et%20aux%20libert%C3%A9s%20(1), acesso em 27 de setembro de 2021.

Lei nº 2018-493, disponível em https://www.legifrance.gouv.br/jorf/id/JORFTEXT000037085952. Acesso em 01 set. 2021.

Sítio oficial da CNIL. Disponível em https://www.cnil.fr. Acesso em 05/10/2021

REFLEXÕES INICIAIS SOBRE A NOVA LEI GERAL DE PROTEÇÃO DE DADOS, revista do Direito do Consumidor, vol. 120/2018, p. 469 – 483, Nov. - Dez / 2018.

MENDES Schertel Laura; Machado Diego**, TECNOLOGIAS DE PERFILAMENTO E DADOS AGREGADOS DE GEOLOCALIZAÇÃO NO COMBATE À COVID-19 NO BRASIL: UMA ANÁLISE DOS RISCOS INDIVIDUAIS E COLETIVOS À LUZ DA LGPD**, 2020.

MENDES, Laura Schertel. **Privacidade, proteção de dados e defesa do consumidor: linhas gerais de um novo direito fundamental**. São Paulo: Saraiva, 2014.

Klaus Schwab*, Shaping the future of the fourth industrial revolution*. New York: Currency, 2018, p. 69

MENDES, Gilmar Ferreira; SARLET, Ingo Wolfgang; COELHO, Alexandre Zavaglia. **Direito, inovação e tecnologia** São Paulo: Saraiva, 2015. v. 1. p. 205]

MALDONADO, Nobrega Viviane; BLUM, Opice Renato; LGPD – **Lei Geral de Proteção de Dados comentada**. São Paulo: Ed. RT, 2019.

WARREN, Samuel D.; BRANDEIS, Louis D. **The right to privacy**. Harvard Law Review, p. 193-220, 1890.

ZUBOFF, Shoshana, Big other: capitalismo de vigilância e perspectivas para uma civilização de informação.

BIONI, Bruno Ricardo; MENDES, Laura Schertel. **Regulamento Europeu de Proteção de Dados Pessoais e a Lei Geral brasileira de Proteção de Dados: mapeando convergências na direção de um nível de equivalência**. In.: Gustavo Tepedino; Ana Frazão; Milena Donato Oliva (Org.). Lei Geral de Proteção de Dados

Pessoais e suas repercussões no direito brasileiro. 1 ed. São Paulo: Thomson Reuters Brasil, 2019.

COPETTI, Rafael; CELLA, José Renato Gaziero. **Autoridade Nacional de Proteção de Dados: natureza jurídica e a Lei nº 13.079/2018.** Caderno de Direito, governança e novas tecnologias II do CONPEDI, 2018. Disponível em: <http://conpedi.danilolr.info/publicacoes/34q12098/15d3698u/Rc9sMnwxjkLPh2o4.pdf> . Acesso em: 22 ago. 2021