DOI http://dx.doi.org/10.1590/1981-5344/62888

ANONIMIZAÇÃO DE DADOS PESSOAIS PARA FINS DE PESQUISA CIENTÍFICA E SUA ADEQUAÇÃO À LGPD

Fabrício Fracaroli Pereira
Universidade Estadual Paulista (UNESP)
São Paulo, São Paulo, Brasil

José Eduardo Santarem Segundo
Universidade de São Paulo (USP)
São Paulo, São Paulo, Brasil

Deise Maria Antonio Sabbag
Universidade Estadual Paulista (UNESP)
São Paulo, São Paulo, Brasil

RESUMO
A Lei Geral de Proteção de Dados Pessoais permite o tratamento de dados pessoais para estudos por órgãos de pesquisa, com a exigência de sua anonimização "sempre que possível". Todavia, a indeterminação desse dispositivo gera incertezas sobre o limite entre a proteção da privacidade e a preservação da utilidade dos dados, o que pode inviabilizar investigações científicas. Este artigo objetiva analisar e propor uma delimitação à necessidade de anonimização dos dados pessoais utilizados para fins de estudos por órgãos de pesquisa, representada na legislação pela expressão "sempre que possível". Trata-se de pesquisa qualitativa, exploratória e descritiva, desenvolvida mediante revisão bibliográfica e documental em doutrina especializada, legislação e Estudos Técnicos da Autoridade Nacional de Proteção de Dados. Como resultados, identificaram-se três possíveis interpretações do dispositivo legal, baseado na categorização de Greco (2008): "sempre que possível sem força"; "sempre que possível fraco"; "sempre que possível forte". Conclui-se pela adequação do sentido forte, que estabelece a anonimização como dever geral, mas viabiliza as pesquisas ao autorizar o uso de dados não anonimizados quando a anonimização for comprovadamente inviável ou destruir a utilidade científica do dado, desde que apresentada justificativa expressa. Essa abordagem equilibra a proteção à privacidade e a promoção do desenvolvimento científico.

Palavras-chave: Dados pessoais. Pesquisa científica. Anonimização.

ANONYMIZATION OF PERSONAL DATA FOR SCIENTIFIC RESEARCH PURPOSES AND ITS COMPLIANCE WITH THE LGPD

ABSTRACT
The General Data Protection Law (LGPD) allows personal data processing for studies by research institutions, requiring anonymization "*whenever possible*". However, the vagueness of this provision creates uncertainty regarding the boundary between privacy protection and the preservation of data utility, which can hinder scientific research. Given this problem, this article aims to analyze and propose a framework for applying the expression "*whenever possible*". This qualitative study, with exploratory and descriptive objectives, was conducted through a literature and documentary review of specialized doctrine, legislation, and Technical Studies from the National Data Protection Authority. As a result, three possible interpretations of the legal provision were identified based on Greco's (2008) categorization: "*whenever possible without force*"; "*weak whenever possible*"; and "*strong whenever possible*". The study concludes in favor of the strong interpretation, which establishes anonymization as a general duty while enabling research by allowing the use of non-anonymized data when anonymization is proven to be unfeasible or destroys the scientific utility of the data, provided that explicit justification is presented. This approach ensures a balance between privacy protection and the promotion of scientific development.

Keywords: Personal data. Scientific research. Anonymization.

DOI http://dx.doi.org/10.1590/1981-5344/62888

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS
data-in-article – Os dados de pesquisa estão disponíveis no corpo do documento.

Recebido em: 17/11/2025
Aceito em: 15/10/2026
Editora Chefe: Lorena Tavares de Paula

1 INTRODUÇÃO

A Lei Geral de Proteção de Dados Pessoais (LGPD) – Lei n. 13.709, de 14 de agosto de 2018 (Brasil, 2018) – prevê dentre as hipóteses autorizadoras do tratamento de dados pessoais, constantes de seu artigo 7º, a possibilidade de sua utilização para a realização de estudos por órgão de pesquisa, e exige, "*sempre que possível*", a anonimização dos dados pessoais. Trata-se de tema que envolve direitos fundamentais dos titulares dos dados pessoais, além de outros direitos constitucionalmente previstos, relacionados à liberdade de pesquisa e à promoção e incentivo do desenvolvimento científico, da pesquisa, da capacitação científica e tecnológica e da inovação.

Durante o desenvolvimento, primeiro foi trabalhado o conceito de *órgão de pesquisa*, descrito no artigo 5º, XVIII, da Lei 13.709/2018 (Brasil, 2018). Em seguida, abordou-se o tema da anonimização, que representa um processo que visa desvincular dados de uma pessoa específica, com o uso de diversas técnicas, de modo que, ao cabo, referidos dados perderão a qualidade de pessoais.

Por fim, debruçou-se sobre a expressão "*sempre que possível*", com apoio à classificação cunhada no âmbito do Direito Tributário pelo professor Marco Aurélio Greco, quando estudou o princípio da capacidade contributiva. Com isso, identificaram-se ao menos três possíveis interpretações acerca da expressão "*sempre que possível*" no contexto ora estudado: um "*sempre que possível sem força*", um "*sempre que possível fraco*" e um "*sempre que possível forte*", concluindo-se pela melhor adequação do último sentido.

O objeto do estudo é a delimitação da aplicação da expressão "*sempre que possível*" (art. 7º, IV, da LGPD) no tratamento de dados pessoais por órgãos de pesquisa (Brasil, 2018). Para responder a essa problemática, adotou-se uma abordagem qualitativa, focada na análise conceitual e dogmática de textos normativos e teóricos, sem emprego de métodos quantitativos. Quanto aos objetivos, a pesquisa enquadra-se como exploratória, por discutir o alcance de um conceito jurídico indeterminado na interface entre Direito e Ciência da Informação, e descritiva, ao expor e categorizar as definições de órgão de pesquisa, as técnicas de anonimização e as interpretações da norma.

DOI http://dx.doi.org/10.1590/1981-5344/62888

O procedimento metodológico estruturou-se em duas etapas integradas. Primeiro, uma Revisão Bibliográfica, realizada mediante levantamento em bases de dados acadêmicas (em especial, Google Acadêmico e Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)) com os descritores *"dados pessoais", "pesquisa científica", "anonimização", "órgão de pesquisa" e "LGPD"*. Foram selecionadas obras doutrinárias e artigos científicos de referência sobre proteção de dados, privacidade e utilidade da informação. Segundo, uma Revisão Documental, consistente no levantamento e análise da legislação nacional (Constituição Federal de 1988 (CF) (Brasil, 1988) e Lei nº 13.709/2018 – LGPD (Brasil, 2018)), de normas internacionais (Regulamento Geral sobre a Proteção de Dados (GDPR) da União Europeia) e de documentos regulatórios oficiais, com destaque para os Estudos Técnicos publicados pela Autoridade Nacional de Proteção de Dados (ANPD) em 2022 e 2023 sobre o tratamento de dados para fins de pesquisa.

2 DESENVOLVIMENTO

A LGPD passou a vigorar completamente no Brasil na segunda metade do ano de 2020[2]. Como resultado de uma vasta discussão tanto nos Poderes Executivo e Legislativo, bem como objeto de debates na sociedade civil e nas universidades, a LGPD inaugurou em nosso país uma nova fase no tema da proteção dos dados pessoais, cuja salvaguarda em nível adequado já se mostrava presente na maioria dos países desenvolvidos (Doneda; Mendes; Bôas Cueva, 2020).

Como destacam Doneda, Mendes e Bôas Cueva (2020, p. 06):

> É bem verdade que já contávamos com dispositivos esparsos de proteção de dados pessoais. A LGPD, contudo, disciplina de maneira sistemática e coerente a coleta, o armazenamento, o tratamento e a disseminação de dados pessoais, que constituem o principal insumo da economia digital para a formação de perfis informacionais, mediante a precisa identificação de hábitos, preferências e decisões existenciais e políticas dos titulares dos dados.

[2] "No entanto, devido a pandemia do Covid-19, a Lei n. 13.709/2018 (LGPD) acabou entrando em vigor em setembro de 2020, após toda uma movimentação do Legislativo e do Executivo, com a aprovação da MP n. 959/2020 em 29-4-2020, mas que deixou de produzir efeitos com a publicação no Diário Oficial da União (DOU) da Lei Federal n. 14.058/2020. E posteriormente, com a aprovação do Projeto de Lei n. 1.179/2020 que foi convertido na Lei n. 14.010 que então prorrogou a aplicação das multas para 1º de agosto de 2021 [...]" (Pinheiro, 2021, p. 19).

CC BY

Diante da sua especial importância para o resguardo de uma série de direitos fundamentais, como a privacidade dos indivíduos, bem como para proporcionar um nível mais adequado de segurança para as relações jurídicas e, consequentemente, contribuir para o avanço tecnológico e o desenvolvimento econômico nacionais, a proteção de dados pessoais foi incluída na Constituição da República Federativa do Brasil de 1988 (Brasil, 1988), pela Emenda Constitucional nº 115, de 10 de fevereiro de 2022 (Brasil, 2022), no rol dos direitos e garantias fundamentais[3].

O tema central da legislação em discussão, o dado pessoal, é definido pela LGPD em seu artigo 5º, I, como a "informação relacionada a pessoa natural identificada ou identificável" (Brasil, 2018). Em outras palavras, a "antítese do conceito de dado pessoal seria um dado anônimo, ou seja, aquele que é incapaz de revelar a identidade de uma pessoa" (Bioni, 2021, p. 61).

Bem por isso, o artigo 7º da LGPD, que prevê as hipóteses autorizadoras do tratamento de dados pessoais, elenca no inciso IV a possibilidade de sua utilização "para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais" (Brasil, 2018)[4]. Na sequência, dispõe no artigo 12 que:

> 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 (Brasil, 2018).

Por se tratar de tema que envolve direitos fundamentais dos titulares dos dados pessoais, e tendo em vista a previsão legal acima descrita, questiona-se: qual é a dimensão da necessidade de anonimização dos dados pessoais utilizados para fins de estudos por órgãos de pesquisa, representada na legislação pela expressão "*sempre que possível*"?

A busca por esclarecimentos acerca da abrangência do texto legal ora debatido importa para assegurar um legítimo e adequado tratamento de dados pessoais quando realizado para fins de pesquisa científica, de modo a garantir tanto a proteção dos dados pessoais nela envolvidos, como também para resguardar direitos constitucionalmente previstos, relacionados à

[3] Foi incluído o inciso LXXIX no artigo 5º da CF, com o seguinte texto: "é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais" (Brasil, 2022).
[4] Tal determinação aparece, também, nos artigos 11, II, "c" e 13, da LGPD (Brasil, 2018).

DOI http://dx.doi.org/10.1590/1981-5344/62888

liberdade de pesquisa (artigo 206, CF) e à promoção e incentivo do desenvolvimento científico, da pesquisa, da capacitação científica e tecnológica e da inovação (artigo 218, CF). Da mesma forma, auxilia a amparar legalmente quem realiza tais pesquisas (Brasil, 1988).

Sob a perspectiva da Ciência da Informação, o tratamento de dados pessoais para fins de pesquisa transcende a mera adequação jurídica, situando-se no âmbito da gestão, da curadoria e dos fluxos da informação científica[5]. Além disso, os processos de anonimização atuam como uma mediação informacional destinada a harmonizar o incentivo ao compartilhamento e reuso de dados[6] com o resguardo da privacidade dos sujeitos. Assim, a indeterminação da expressão, "*sempre que possível*" reflete um desafio central da Ciência da Informação: equacionar a preservação da utilidade e do valor científico da informação com a garantia da segurança informacional dos titulares.

Para responder ao problema suscitado, primeiro é importante uma definição do que se entende por *órgãos de pesquisa*, seguida de uma análise sobre o mecanismo da anonimização de dados (incluindo os pessoais) para, por fim, adentrar no tema do alcance da expressão acima mencionada.

2.1 Órgãos de pesquisa

A LGPD elenca, no seu artigo 5º, uma série de definições de termos e expressões nela previstos. Dentre eles, há o conceito de *órgão de pesquisa* (inciso XVIII), sendo o:

> […] órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado sem fins lucrativos legalmente constituída sob as leis brasileiras, com sede e foro no País, que inclua em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico (Brasil, 2018).

[5] "A Ciência da Informação tem um importante papel na construção de um arcabouço teórico que atenda a demanda pela busca de informações, uma vez que é papel da Ciência da Informação investigar o comportamento da informação, bem como o fluxo e o acesso a esta" (Affonso; Sant'Ana, 2017, p. 21).

[6] Para Silveira, Araujo e Dias (2023, p. 208), "[…] urge a necessidade de estimular o compartilhamento e o reuso de dados científicos de pesquisa, cujos benefícios perpassam por maior transparência e clareza da pesquisa, reconhecimento da propriedade dos dados, aumento nas citações, economia de tempo na coleta de dados e nos custos da pesquisa".

A partir da concepção legal, infere-se que um *órgão de pesquisa* pode fazer parte da Administração Pública[7], seja ela Direta ou Indireta. Por Administração Pública entende-se o universo de órgãos e pessoas que, num regime federativo, desempenham a função administrativa (Carvalho Filho, 2020). A Administração Pública Direta compreende o

> […] conjunto de órgãos que integram as pessoas federativas, aos quais foi atribuída a competência para o exercício, de forma centralizada, das atividades administrativas do Estado (Carvalho Filho, 2020, p. 487).

A Administração Pública Indireta o

> […] conjunto de pessoas administrativas que, vinculadas à respectiva Administração Direta, têm o objetivo de desempenhar as atividades administrativas de forma descentralizada (Carvalho Filho, 2020, p. 494).

Como exemplos de *órgãos de pesquisa* na Administração Pública, temos o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), o Instituto Nacional de Tecnologia (INT) e o Instituto Nacional de Pesquisas Espaciais (INPE), que consistem em unidades de pesquisa integrantes da estrutura do Ministério da Ciência, Tecnologia e Inovação, sendo parte da Administração Pública Direta; e as Instituições de Ensino Superior Públicas, em sua maioria constituídas sob a forma de autarquia de regime especial[8], componentes da Administração Pública Indireta.

Da mesma forma, se enquadram no conceito de *órgão de pesquisa* as pessoas jurídicas de direito privado sem fins lucrativos, legalmente constituídas sob as leis brasileiras, com sede e foro no País. Excluem-se desse conceito, portanto, as pessoas naturais (ou pessoas físicas), as pessoas jurídicas de direito privado *com* fins lucrativos e as não constituídas sob a lei brasileira ou com sede e foro no exterior.

[7] A grafia da expressão "Administração Pública" com letras iniciais maiúsculas tem a finalidade de diferenciar os diferentes sentidos a ela atribuídos pela doutrina jurídica. Quando escrita com as iniciais minúsculas, representa o sentido objetivo, ou seja, refere-se à própria atividade administrativa exercida pelo Estado, seja por seus agentes ou por seus órgãos, enquanto as iniciais maiúsculas exprimem o sentido subjetivo, que corresponde ao conjunto de agentes, órgãos e pessoas jurídicas que exercem a atividade administrativa, sentido ora empregado (Carvalho Filho, 2020).

[8] "Para as IES públicas, não se determinou qualquer forma amparada na Administração Pública indireta (art. 37, XIX, da CR/1988), o que Ranieri (2000) alega ser uma medida inócua, porque inexiste no sistema jurídico pátrio outra possibilidade de constituição. Dessa forma, em todos os sistemas de ensino, somente a lei, de iniciativa do chefe do Executivo, aprovada pelo Poder Legislativo, pode instituir IES públicas, universitárias ou não (arts. 48, XI, 37, XIX, da CR/1988), mediante a forma e regime jurídico previstos para a Administração indireta, isto é, pelo regime autárquico ou fundacional comum" (Gomes, 2009, p. 119).

DOI http://dx.doi.org/10.1590/1981-5344/62888

A exclusão das pessoas jurídicas de direito privado com fins lucrativos da incidência desta hipótese de tratamento de dados pessoais foi expressamente analisada pelo Poder Legislativo, sob o fundamento de que a exceção à necessidade do consentimento do titular dos dados concedida aos *órgãos de pesquisa* é necessária para facilitar a execução de políticas públicas e a prestação de serviços, ao passo que entidades de pesquisa privadas que possuem fins lucrativos, por não exercerem mandatos legais e terem o lucro por objetivo, não se enquadram no permissivo legal, devendo, portanto, obterem o consentimento (ANPD, 2022).

Em todos os casos, para que possam proceder ao tratamento de dados pessoais com base no inciso IV do artigo 7º da LGPD, tais órgãos ou entidades (públicos ou privados) devem incluir em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico (Brasil, 2018).

Importa mencionar que as pessoas jurídicas de direito privado com fins lucrativos que exercem atividades de pesquisa não estão impedidas de proceder ao tratamento de dados pessoais, apenas por não se enquadrarem no conceito legal de *órgão de pesquisa*. Elas somente não poderão fazê-lo com base na hipótese do inciso IV, de modo que deverão enquadrar o tratamento dos dados em outro inciso da LGPD (Brasil, 2018), ou seja, exige-se "[...] para esses casos o amparo da situação concreta em outra hipótese legal, como as bases do consentimento do titular ou do legítimo interesse, observados os requisitos legais aplicáveis" (ANPD, 2022, p. 15).

O tratamento de dados pessoais por *órgãos de pesquisa*, com fundamento no inciso IV do artigo 7º da LGPD (Brasil, 2018), também exige, além do aspecto subjetivo acima tratado, um atributo objetivo, que é a anonimização dos dados pessoais. Portanto, é relevante a análise do método da anonimização de dados para, em seguida, verificar a extensão de sua indispensabilidade, legalmente estabelecida.

2.2 Anonimização de dados

Como mencionado, a LGPD exige aos *órgãos de pesquisa* um requisito objetivo para o tratamento de dados pessoais com base no seu artigo 7º, IV. Trata-se da anonimização dos dados, cujo conceito legal previsto no artigo 5º,

XI, consiste na "utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo" (Brasil, 2018).

A anonimização consiste em uma maneira de desvincular dados de uma pessoa específica, mediante diversas técnicas (aplicadas isolada ou conjuntamente), de modo que, realizado esse processo, referidos dados perderão a qualidade de pessoais. Com um dado anônimo (ou anonimizado), alguma informação relacionada a um indivíduo pode até ser conhecida, mas não poderá ser vinculada à sua identidade (Affonso, Oliveira, Sant'Ana, 2017).

Portanto, a anonimização pode ser entendida como o processo de remoção de identificadores, de modo que o vínculo entre o titular e o dado é desfeito. É tratada como um processo, pois se desenvolve em uma sequência "[...] de etapas que se inicia com o processamento de dado de caráter pessoal e tem o objetivo de, com a aplicação de técnicas variadas, desassociar identificadores do dado em seu estado originário ou bruto" (ANPD, 2022, p. 6). Como já foi mencionado, esse processo pode ser realizado mediante a utilização de diferentes técnicas, cujo objetivo é a eliminação desses elementos identificadores de um banco de dados (Bioni, 2020, p. 39).

> Tais técnicas podem ser entendidas como a modificação ou destruição de informações em bases de dados de forma que não seja mais possível identificar os indivíduos. Isso é feito, em geral, de três maneiras: supressão, quando alguns tipos de dados sensíveis são eliminados da base de dados; substituição, quando dados sensíveis são substituídos por outros dados não sensíveis ou falsos; e generalização, quando dados específicos são substituídos por categorias mais genéricas, por exemplo, a idade de indivíduos podem ser substituídas por intervalos como "entre 20 e 30 anos (Sousa *et al.*, 2020, p. 58).

Outra técnica trazida pela legislação é a chamada pseudonimização. No artigo 13 da LGPD, ao abordar a realização de estudos em saúde pública, a lei admite tanto a anonimização dos dados como também sua pseudonimização, cuja definição legal constante do parágrafo 4º do mencionado artigo a traduz como o

> [...] tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro (Brasil, 2018).

Embora muitos não considerem a pseudonimização como uma técnica de anonimização, já que são substituídos apenas

DOI http://dx.doi.org/10.1590/1981-5344/62888

> […] os identificadores diretos – *e.g.*, nome, CPF *etc*. – por pseudônimos – *e.g.*, números aleatórios, de modo que a pessoa permanece sendo identificável em razão de tais pseudônimos serem um retrato detalhado indireto delas (Bioni, 2020, p. 40).
>
> [Ela] se difere da anonimização no sentido de que a pseudonimização permite reidentificar e recuperar os dados originais enquanto a anonimização não (ao menos teoricamente) (Sousa *et al.*, 2020, p. 58).
>
> Por exemplo, o processo de pseudonimização conhecido como *Tokenização*, que consiste na substituição de dados por informações aleatórias conhecidas como *tokens*, é utilizado com sucesso em algumas aplicações financeiras para proteger dados sensíveis como contas bancárias e números de cartões de crédito (Sousa *et al.*, 2020, p. 58).

Diante dessa possibilidade de reidentificação a partir de informações que o controlador possui, isto é, de se rastrear seus titulares por meio do código-chave, a lei considera os dados pseudonimizados como dados pessoais (Doneda; Barreto; Almeida, 2019).

Pela leitura da definição legal de anonimização, acima exposta, nota-se que foi utilizado na elaboração do conceito o critério da razoabilidade, assim como também o adotou o Direito europeu[9]:

> Não basta a mera possibilidade de que um dado seja atrelado a uma pessoa para atrair o termo identificável. Essa vinculação deve ser objeto de um "esforço razoável", sendo esse o perímetro de elasticidade do conceito de dado pessoal como aquele relacionado a uma pessoa identificável (Bioni, 2020, p. 41).

Além disso, formulou-se uma norma neutra tecnológicamente[10], ou seja, a lei não elege uma tecnologia específica por meio da qual deva ser efetuado o processo de anonimização, sob o risco de, com o passar do tempo, ela se tornar obsoleta. Ao contrário, estabelece que devem ser utilizados os meios técnicos disponíveis no momento do tratamento dos dados, permitindo os avanços tecnológicos sem, contudo, que a norma seja defasada.

De acordo com o artigo 12 da LGPD, os dados anonimizados não serão considerados dados pessoais, salvo quando seu processo de anonimização for

[9] O considerando n. 26 do GDPR assim estabelece: "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 evolução tecnológica" (União Europeia, 2016).

[10] Como ensina Bioni (2020, p. 42), "O conceito de 'technology-neutral regulation' tem sido evocado para se discutir o desenho de modelos regulatórios capazes de estimular e acompanhar o desenvolvimento tecnológico, sem engessá-lo nem ser permissivo a riscos."

CC BY

revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido (Brasil, 2018). Para se verificar se a viabilidade de reidentificação é razoável, a LGPD, seguindo o modelo europeu, estabelece que se devem considerar fatores objetivos, como o custo e o tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e um fator subjetivo, qual seja, a utilização exclusiva de meios próprios (artigo 12, §1º, LGPD) (Brasil, 2018). Diz-se subjetivo, pois leva "[...] em consideração quem é o agente de tratamento de dados e se ele dispõe de 'meios próprios' para reverter o processo de anonimização" (Bioni, 2020, p. 44).

Adota-se a razoabilidade como baliza, pois embora as técnicas de anonimização possam ser muito eficientes, não há como garantir o absoluto e irreversível anonimato. Com a utilização de instrumentos tecnológicos, há o risco de se reverter praticamente todo e qualquer processo de anonimização, permitindo a reidentificação dos titulares dos dados[11].

> Torna-se cada vez mais recorrente a publicação de estudos que demonstram ser o processo de anonimização algo falível. A representação simbólica de que os vínculos 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 (Bioni, 2021, p. 63).

Por isso, deve-se trabalhar com o escopo de mitigar os riscos inerentes aos processos de anonimização – e não os eliminar, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião do tratamento, conforme estabelece a legislação. Além disso, há de se procurar um equilíbrio – razoabilidade – entre a anonimização e a utilidade dos dados resultantes deste processo, pois caso contrário, pode-se prejudicar os fins a que se destinam, ou seja, podem se tornar inúteis às pesquisas científicas pretendidas.

Um exemplo diz respeito à utilização de bancos de dados compartilhados. Affonso, Oliveira e Sant'Ana (2017) constataram que as organizações se interessam por colaborar com pesquisas que envolvam o compartilhamento de dados, embora possam ocorrer riscos à confidencialidade de tais informações. Por esse motivo, tais compartilhamentos usualmente são realizados com identificadores pessoais suprimidos, a fim de

[11] "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" (Bioni, 2020, p. 40).

DOI http://dx.doi.org/10.1590/1981-5344/62888

garantir a privacidade. Todavia, ressaltam que "[...] a utilidade do conjunto de dados privados continua sendo um desafio, pois ao utilizar métodos de anonimização pode-se também diminuir a utilidade dos dados" (Affonso; Oliveira; Sant'Ana, 2017, p. 82).

> Assim, no processo de proteção da privacidade, observa-se de um lado o detentor de dados e de outro o conjunto de dados anonimizados, que é permeado pela díade utilidade dos dados e privacidade dos sujeitos referenciados por estes dados, visto que o processo busca o equilíbrio entre estas duas forças antagônicas. Logo, quanto maior a proteção dos dados, maior tendência à privacidade e à probabilidade de redução da utilidade dos dados disponibilizados. Uma proteção dos dados mínima pode garantir e manter a utilidade dos dados disponibilizados, no entanto, pode significar riscos ou violação da privacidade dos indivíduos envolvidos (Affonso; Oliveira; Sant'Ana, 2017, p. 82).

Diante das diferentes tecnologias disponíveis para a realização de um processo de anonimização, com demandas de tempo e custo as mais diversas possíveis, ou mesmo pela possibilidade de perda da utilidade dos dados depois de submetidos a um processo de anonimização, notadamente no que se refere a dados destinados a pesquisas cientificas, a legislação exige, "*sempre que possível*", que se realize a anonimização. Por isso, importa saber a amplitude dessa ressalva, a fim de que os tratamentos de dados realizados por órgãos de pesquisa possam garantir uma efetiva proteção dos dados pessoais dos titulares, bem como cumprir a determinação legal sem que sejam prejudicadas a eficiência e a liberdade de pesquisa, assim como o desenvolvimento científico e tecnológico do país.

2.3 O alcance da expressão "*sempre que possível*"

Como foi apresentado no decorrer do presente texto, dentre as hipóteses de tratamento de dados pessoais previstas na LGPD, há a possibilidade de sua utilização para a realização de estudos por órgão de pesquisa, devendo ser garantida, "*sempre que possível*", a anonimização dos dados pessoais (Brasil, 2018). Nos tópicos anteriores, foi esclarecido o que se entende por *órgão de pesquisa*, bem como discorrido sobre alguns contornos dos processos de anonimização. Resta, portanto, debruçar-se sobre o alcance da expressão "*sempre que possível*", para que o objetivo do presente artigo seja atingido.

A abrangência da determinação legal acerca da anonimização de dados pode gerar interpretações diversas, ante o uso pelo legislador do que se

CC BY

pode considerar como um *conceito jurídico indeterminado*[12] – a expressão "*sempre que possível*". Por isso, para que seja conferida uma eficácia normativa compatível com os direitos e garantias fundamentais envolvidos na norma em estudo, é importante buscar uma interpretação que atenda a tais anseios, isto é, que não seja vaga a ponto de frustrar a proteção pretendida aos direitos dos titulares dos dados, nem muito rígida a ponto de inviabilizar o desenvolvimento das atividades inerentes aos órgãos de pesquisa.

Para analisar a expressão em debate, optou-se por buscar amparo em uma discussão advinda do Direito Tributário. Ao estudar o princípio da capacidade contributiva[13], o professor Marco Aurélio Greco[14], dentre outros autores, se debruçou sobre as interpretações da expressão "*sempre que possível*", constante do artigo 145, §1º, da CF (Brasil, 1988). A partir de suas conclusões, identificou três possíveis interpretações do dispositivo legal: uma primeira, no sentido de que a expressão seria apenas uma recomendação (algo como "se puder, faça"), à qual denominou de "*sempre que possível*" *sem força*[15]; uma segunda interpretação, no sentido de se tratar de um limite negativo ao legislador, ou seja, uma lei que não observe o princípio seria inconstitucional, vertente que chamou de "*sempre que possível fraco*"; e, por fim, uma terceira perspectiva, que considerou como sendo a mais adequada[16], pela qual haveria um enfoque positivo da expressão analisada, como se ela determinasse que "[...] só quando não for possível é que pode deixar de ser

[12] "Conceitos jurídicos indeterminados são termos ou expressões contidos em normas jurídicas, que, por não terem exatidão em seu sentido, permitem que o intérprete ou o aplicador possam atribuir certo significado, mutável em função da valoração que se proceda diante dos pressupostos da norma" (Carvalho Filho, 2020, p. 60).

[13] Meirelles (1997, p. 336) apresenta uma definição do princípio da capacidade contributiva: "Assim, dentro da árdua tarefa de conceituar capacidade contributiva, temos que pode ser ela considerada como a força econômica complexa e materializada do indivíduo, idônea a concorrer com as despesas públicas, à luz de exigências econômicas e sociais fundamentais, respeitando-se o mínimo vital para uma existência pessoal e familiar digna e livre".

[14] "A sua análise sobre o princípio da capacidade contributiva parte do disposto no §1º do artigo 145 da Constituição Federal, segundo o qual "sempre que possível, os impostos terão caráter pessoal e serão graduados segundo a capacidade econômica do contribuinte, facultado à administração tributária, especialmente para conferir efetividade a esses objetivos, identificar, respeitados os direitos individuais e nos termos da lei, o patrimônio, os rendimentos e as atividades econômicas do contribuinte" (Rocha, 2019, p. 23).

[15] E conclui o autor: "Desta perspectiva, estaríamos diante de recomendação do tipo "se der pra fazer, faça, mas se não der não faça que nada acontece". A meu ver, isto seria negar toda e qualquer eficácia ao dispositivo" (Greco, 2008, p. 326).

[16] "Este é um terceiro ângulo de interpretação do "sempre que possível", um "sempre que possível" forte que comanda, ou seja, só não se aplica o princípio quando isto não for possível; fora desta hipótese, a lei, sua interpretação e sua aplicação sempre terão de atender ao princípio" (Greco, 2008, p. 327).

DOI http://dx.doi.org/10.1590/1981-5344/62888

atendido o princípio da capacidade contributiva" (Greco, 2008, p. 327), tratando-se do chamado "*sempre que possível forte*".

Em analogia às conclusões alcançadas pelo mencionado tributarista, e usando as nomenclaturas cunhadas, pode-se identificar ao menos três possíveis interpretações acerca da expressão "*sempre que possível*" no contexto da necessidade de anonimização de dados pessoais:

a) um "*sempre que possível*" *sem força*, que serviria apenas como uma recomendação aos órgãos de pesquisa que tratam dados pessoais, para que efetuem o processo de anonimização quando embasados no artigo 7º, IV, da LGPD (Brasil, 2018);
b) um "*sempre que possível fraco*", atuando também como um limite negativo, no sentido de que, se não for possível realizar a anonimização dos dados, os órgãos de pesquisa devem se abster de realizar seu tratamento com base na hipótese autorizadora, sob o risco de praticar uma ilegalidade;
c) por fim, um "*sempre que possível forte*", cuja determinação seria no sentido de que apenas quando não for possível é que se poderia deixar de efetivar a anonimização, justificada tal impossibilidade.

Principalmente em razão da natureza dos direitos envolvidos, isto é, por se tratar de direitos fundamentais, bem como pela natureza e objetivos da norma jurídica em discussão – LGPD[17], não há como negar uma imperiosidade à determinação de anonimização. Logo, uma interpretação que considere apenas como se fosse uma recomendação legal não se mostra razoável. A razão de ser do processo de anonimização é justamente a desvinculação do dado do seu titular, com o fim de proteger sua privacidade, o que, para além de recomendado, é na realidade almejado.

Por sua vez, um "*sempre que possível*" compreendido sob um ponto de vista *fraco* seria insuficiente, pois restringiria a atuação dos órgãos de pesquisa que, sob o risco de infringirem a norma, poderiam deixar de realizar suas

[17] "Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural" (Brasil, 2018).

CC BY

atividades. Partindo dessa interpretação, se não for possível a anonimização dos dados, não haveria tratamento sob pena de recair em ilegalidade. Da mesma forma, não parece razoável tamanha restrição.

Resta, portanto, como o mais adequado, o sentido *forte* da expressão. Por esta perspectiva, um órgão de pesquisa deve sempre efetuar o processo de anonimização dos dados pessoais quando tratados sob a hipótese do inciso IV, do artigo 7º, da LGPD (Brasil, 2018). Isto é, sempre que for possível a anonimização, deve-se realizá-la. Caso não seja possível a anonimização, ainda assim poder-se-á proceder ao tratamento dos dados, desde que justificada a impossibilidade, como, por exemplo, nos casos em que a utilidade dos dados será gravemente comprometida se submetidos ao processo, inviabilizando a pesquisa. Com a finalidade de expressamente demonstrar a boa-fé, bem como para asseverar a legalidade do tratamento, considera-se importante uma justificativa acerca dessa impossibilidade. Até porque, dentre os princípios expressos na LGPD, além da própria boa-fé, há a previsão de responsabilização e prestação de contas (*accountability*), descrito no artigo 6º, X, como a "demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas" (Brasil, 2018).

O processo de anonimização, portanto, é um requisito objetivo indispensável ao tratamento de dados pessoais com fundamento no artigo 7º, inciso IV, da LGPD, salvo expressa e justificada impossibilidade (Brasil, 2018).

Além disso, é possível fundamentar a indispensabilidade do procedimento de anonimização, salvo expressa e justificada impossibilidade, ante a relevância dos princípios *privacy by design* e *by default*, que norteiam o tema da proteção de dados pessoais, e são adotados tanto pela LGPD quanto pelo Regulamento europeu (GDPR), no sentido de que a proteção da privacidade deve ser resguardada desde a criação e desenvolvimento de um produto ou serviço e mantida por todo o ciclo de vida dos dados (Arbix, 2020).

> A proposta de *by Design and Default* é que o planejamento atento à privacidade preceda o processamento de dados pessoais, ou seja, que potenciais problemas com a privacidade dos titulares sejam identificados e solucionados no início do ciclo de desenvolvimento de um produto ou serviço ou de sua nova versão, antes de seu lançamento no mercado (Arbix, 2020, p. 57).

DOI http://dx.doi.org/10.1590/1981-5344/62888

Assim, seja para a elaboração de um produto ou serviço, ou mesmo para as atividades de pesquisa realizadas por órgãos públicos ou privados sem fins lucrativos, o respeito à privacidade deve estar profundamente imbricado nos procedimentos envolvendo dados pessoais, desde o início e durante todo o seu tratamento.

3 CONCLUSÕES

A LGPD representa um avanço tanto na proteção dos dados pessoais, como também no resguardo de direitos fundamentais correlatos, como é o caso da privacidade. A preocupação em respeitar os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural é expressamente mencionada logo no primeiro artigo da Lei. No mesmo sentido, também se alicerça na liberdade de expressão, de informação, de comunicação e de opinião. Portanto, não há como negar a necessidade de compatibilidade entre os diversos direitos por ela estimados.

A proteção à privacidade, que aparece como objetivo e fundamento da Lei, encontra uma de suas formas de concreção quando da anonimização dos dados pessoais destinados à realização de estudos por órgãos de pesquisa. Todavia, nesse processo de anonimização de dados, deve-se ter em mente um critério de razoabilidade em relação à proteção da privacidade e a necessidade de anonimização. Por isso, a lei exige que referido processo seja realizado *sempre que possível*.

Como resultado da análise da expressão *sempre que possível*, a categorização em três interpretações demonstrou que o sentido sem força esvazia a eficácia normativa ao reduzi-la a mera recomendação, enquanto o sentido fraco inviabiliza as pesquisas ao impor a abstenção do tratamento diante de qualquer obstáculo à anonimização. Em contraposição, como proposição central deste estudo para a delimitação do dispositivo legal, conclui-se pela adequação do sentido forte. Essa perspectiva estabelece a anonimização como dever geral, mas autoriza o uso de dados não anonimizados quando o processo for comprovadamente inviável ou destruir a utilidade científica da informação, desde que acompanhado de justificativa expressa e do cumprimento do dever de prestação de contas (*accountability*).

CC BY

Por estar mais de acordo com os princípios e fundamentos do tema da proteção de dados pessoais, insculpidos tanto em nível constitucional como na legislação, como os mencionados princípios do *privacy by design* e *by default*, a adoção do sentido forte consolida-se como a proposição que equaciona a tensão entre privacidade e utilidade dos dados, assegurando a proteção dos titulares sem comprometer os fluxos informacionais e o desenvolvimento científico nacional.

REFERÊNCIAS

AFFONSO, Elaine Parra; OLIVEIRA, Sandra Cristina de; SANT'ANA, Ricardo César Gonçalves. Análise do equilíbrio entre privacidade e utilidade no acesso a dados. **Informação & Sociedade**, v. 27, n. 1, 2017. Disponível em: https://periodicos.ufpb.br/index.php/ies/article/view/29422. Acesso em: 03 jun. 2024.

AFFONSO, Elaine Parra; SANT'ANA, Ricardo César Gonçalves. Preservação da privacidade no acesso a dados por meio do modelo k-anonimato. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 12, n. 2, 2017. DOI https://doi.org/10.22478/ufpb.1981-0695.2017v12n2.36774.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS [ANPD]. **Estudo técnico**: a LGPD e o tratamento de dados pessoais para fins acadêmicos e para a realização de estudos por órgão de pesquisa. Brasília, DF: ANPD, 2022.

ARBIX, Daniel. A importância da Privacidade por Design e por Default (Privacy by Design and by Default). *In*: DONEDA, Danilo; MENDES, Laura Schertel; BÔAS CUEVA, Ricardo Villas. **Lei Geral de Proteção de Dados (Lei n. 13.709/2018)** – A caminho da efetividade: contribuições para a implementação da LGPD. São Paulo: Thomson Reuters, 2020.

BIONI, Bruno Ricardo. Compreendendo o conceito de anonimização e dado anonimizado. *In*: DONEDA, Danilo; MENDES, Laura Schertel; BÔAS CUEVA, Ricardo Villas. **Lei Geral de Proteção de Dados (Lei n. 13.709/2018)** – A caminho da efetividade: contribuições para a implementação da LGPD. São Paulo: Thomson Reuters, 2020.

BIONI, Bruno Ricardo. **Proteção de dados pessoais**: a função e os limites do consentimento. 3. ed. Rio de Janeiro: Forense, 2021.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/ConstituicaoCompilado.htm. Acesso em: 01 maio 2024.

DOI http://dx.doi.org/10.1590/1981-5344/62888

BRASIL. **Emenda constitucional nº 115, de 10 de fevereiro de 2022**. Altera a Constituição Federal para incluir a proteção de dados pessoais entre os direitos e garantias fundamentais e para fixar a competência privativa da União para legislar sobre proteção e tratamento de dados pessoais. Brasília, DF: Presidência da República, 2022. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc115.htm. Acesso em: 22 out. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm. Acesso em: 01 maio 2024.

CARVALHO FILHO, José dos Santos. **Manual de direito administrativo**. 34. ed. São Paulo: Atlas, 2020.

DONEDA, Danilo; BARRETO, Maurício Lima; ALMEIDA, Bethânia de Araújo. de. Uso e proteção de dados pessoais na pesquisa científica. **Direito Público**, Porto Alegre, v. 16, n. 90, 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3895. Acesso em: 03 jun. 2024.

DONEDA, Danilo; MENDES, Laura Schertel; BÔAS CUEVA, Ricardo Villas. **Lei Geral de Proteção de Dados (Lei n. 13.709/2018)** – A caminho da efetividade: contribuições para a implementação da LGPD. São Paulo: Thomson Reuters, 2020.

GOMES, Magno Federici. Natureza, regime jurídico e níveis de especialização das instituições de ensino superior. **Direito Público**. Porto Alegre, v. 6, n. 29, p. 113-131, set./out. 2009. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/1626. Acesso em: 22 out. 2026.

GRECO, Marco Aurélio. **Planejamento tributário**. 2. ed. São Paulo: Dialética, 2008.

MEIRELLES, José Ricardo. O princípio da capacidade contributiva. **Revista de informação legislativa**, Brasília (DF), v. 34, n. 136, p. 333-339, out./dez. 1997. Disponível em: https://www2.senado.leg.br/bdsf/bitstream/handle/id/315/oprincipiodacapacidade.pdf?sequence=6&isAllowed=y. Acesso em: 01/05/2024.

PINHEIRO, Patrícia Peck. **Proteção de dados pessoais: comentários à Lei n. 13.709/2018 (LGPD)**. 3. ed. São Paulo: Saraiva Educação, 2021.

ROCHA, Sergio André. **Planejamento Tributário na Obra de Marco Aurélio Greco**. Rio de Janeiro: Editora Lumen Juris, 2019.

SILVEIRA, Crislaine Zurilda; ARAÚJO, Paula Carina de; DIAS, Thiago Magela Rodrigues. Aspectos éticos no reuso de dados científicos de pesquisa: uma revisão sistematizada da literatura. **Informação & Informação**, v. 28, n. 2, p. 207–224, 2024. DOI https://doi.org/10.5433/1981-8920.2023v28n2p207.

SOUSA, Thiago R.; COUTINHO, Murilo; COUTINHO, Lilian; ALBUQUERQUE, Robson. LGPD: Levantamento de Técnicas Criptográficas e de Anonimização para Proteção de Bases de Dados. *In*: SIMPÓSIO BRASILEIRO DE SEGURANÇA DA INFORMAÇÃO E DE SISTEMAS COMPUTACIONAIS (SBSEG), 20., 2020, Petrópolis. **Anais** […]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 55-68. DOI https://doi.org/10.5753/sbseg.2020.19227.

UNIÃO EUROPEIA. **Regulamento (UE) 2016/679 Do Parlamento Europeu E Do Conselho, de 27 de abril de 2016**. Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Bruxelas: Parlamento Europeu, 2016. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679. Acesso em: 03/06/24.

CONFLITOS DE INTERESSE

Os autores declaram não haver conflitos de interesse relacionados à realização e à publicação deste artigo.

DOI http://dx.doi.org/10.1590/1981-5344/62888

CONTRIBUIÇÕES DAS AUTORIAS

Informa-se nesta seção as funções de cada autoria, de acordo com a taxonomia CRediT, conforme orientado na página da revista PCI:

Função	Definição
Conceituação	Fabrício Fracaroli Pereira.
Curadoria de dados	—
Análise Formal	—
Obtenção de financiamento	—
Investigação	—
Metodologia	—
Administração do projeto	—
Recursos	—
Software	—
Supervisão	José Eduardo Santarem Segundo; Deise Maria Antonio Sabbag.
Validação	—
Visualização [de dados (infográfico, fluxograma, tabela, gráfico)]	—
Escrita – primeira redação	Fabrício Fracaroli Pereira.
Escrita – revisão e edição	Fabrício Fracaroli Pereira; José Eduardo Santarem Segundo; Deise Maria Antonio Sabbag.