Dados sensíveis de pesquisa relacionados à saúde numa perspectiva da Lei Geral de Proteção de Dados Pessoais

Sensitive research data related to health from a perspective of the General Law for the Protection of Personal Data

Paula Cotrim de Abrantes
ORCID: : https://orcid.org/0000-0003-0271-2186

Doutoranda em Ciência da Informação no Programa de Pós-graduação em Ciência da Informação pela Universidade Federal do Rio de Janeiro (IBICT/ UFRJ), Brasil. Arquivista na Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Brasil.

E-mail: pcotrimdeabrantes@gmail.com

Elisângela Marinho da Silva
ORCID: https://orcid.org/0009-0008-3971-7275

Mestranda em Ciência da Informação no Programa de Pós-graduação em Ciência da Informação pela Universidade Federal do Rio de Janeiro (IBICT/ UFRJ), Brasil.

E-mail: : elisangelamarinhodasilva4@gmail.com

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

RESUMO: Dados sensíveis de pesquisa relacionados à saúde é um tema de interesse social. É importante analisar e pesquisar como este assunto está relacionado à Lei Geral de Proteção de Dados Pessoais (LGPD) visto que se este tipo de dado não for bem tratado, cuidado, armazenado e anonimizado, pessoas podem sofrer estigmatização e bullying. Nesse sentido, observou-se uma lacuna na pesquisa sobre anonimização desses dados. Dessa forma, o estudo teve como objetivo refletir, analisar e descrever questões relacionadas aos dados sensíveis de saúde usados na pesquisa no que concerne à LGPD. Para isso, foi usada uma metodologia exploratória e bibliográfica, obtendo-se como resultado a descrição de algumas formas de anonimização de dados sensíveis de pesquisa relacionados à saúde.

PALAVRAS-CHAVE: Dados sensíveis de pesquisa; Saúde; LGPD; Anonimização.

ABSTRACT: Sensitive health-related research data is a topic of social interest. It is important to analyze and research how this subject is related to the General Law for the Protection of Personal Data (LGPD) since if this type of data is not well treated, cared for, stored and anonymized, people may suffer stigmatization and bullying. In this sense, a gap was observed in research on the anonymization of these data. Thus, the study aimed to reflect, analyze and describe issues related to sensitive health data used in research regarding the LGPD. For this, an exploratory and bibliographic methodology was used, resulting in the description of some forms of anonymization of sensitive research data related to health.

KEYWORDS: Sensitive research data; Health; LGPD; Anonymization.

1 Introdução

Cunha e Cavalcanti (2008, p. 113), no *Dicionário de Biblioteconomia e Arquivologia*, definem "dado", na sua forma mais simples, como a "menor representação da informação", podendo estar disponível de forma analógica ou digital e ser tratado de forma manual ou automática. Semeler e Pinto (2018) informam que desde a metade dos anos 2000 a biblioteconomia dos Estados Unidos, Reino Unido e Canadá vem se empenhando no estudo dos dados, criando formas de gerenciá-los e preservá-los. Para a Agência de Bibliotecas e Coleções Digitais da Universidade de São Paulo (2023), dados "são componentes centrais do processo de pesquisa. São

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

registros científicos que embasam os resultados de pesquisa publicados na forma de dissertações, teses, artigos, patentes e trabalhos científicos". No que se refere aos dados de pesquisa, para Sayão e Sales (2020, p. 32) o termo "dado de pesquisa"

> tem uma amplitude de significados que vão se transformando de acordo com domínios científicos específicos, objetos de pesquisas, metodologias de geração e coleta de dados e muitas outras variáveis. Pode ser o resultado de um experimento realizado num ambiente controlado de laboratório, um estudo empírico na área de ciências sociais ou a observação de um fenômeno cultural ou da erupção de um vulcão num determinado momento e lugar (SAYÃO; SALES, 2020, p. 32).

Os dados de pesquisa necessitam passar por alguns procedimentos para que não fiquem perdidos numa base de dados. Eles precisam ser "identificáveis, citáveis, visíveis, recuperáveis, interpretáveis, contextualizáveis, interoperáveis e reutilizáveis onde quesitos de consistência e procedência são considerados" (SEMELER; PINTO, 2018, p. 116).

Nesse sentido, com a criação da internet, com o avanço do aumento da capacidade de processamento dos computadores e com o surgimento de novas inovações tecnológicas, chegou-se ao Big Data na criação e manutenção de repositórios digitais, bem como nas melhores práticas para a curadoria digital. Tudo isso propiciou um melhor tratamento dos dados de pesquisa. Nesse contexto, Sayão e Sales (2020) afirmam que para se ter uma boa gestão de dados é preciso entender qual a função do dado na pesquisa e seu fluxo, pois dessa forma construir-se-ão critérios de pesquisa que tendem a facilitar que o pesquisador encontre o dado tendo condições de usá-lo (ou reusá-lo).

No entanto, os dados de pesquisa podem conter dados pessoais[1]. Esses dados precisam estar protegidos e dentro das normas da Lei Geral de Proteção de Dados Pessoais - LGPD. Na Europa, o Regulamento Geral sobre a Proteção de Dados

[1] Nesse contexto, é importante ressaltar que pesquisas científicas que contam com a participação de humanos precisam ser avaliadas "pelo Sistema CEP/CONEP, visando proteger o participante da pesquisa e assegurar que o estudo será realizado de acordo com princípios éticos" (SILVA, 2023, p. 281). Comitê de Ética em Pesquisa (CEP) - instância institucional, local -, e Comissão Nacional de Ética em Pesquisa (CONEP).

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

(RGPD) entrou em vigor em maio de 2016, e com aplicação a partir de maio de 2018 (Comissão Europeia, [2022?]). Mais detalhes podem ser vistos em EUR-Lex (2018)[2]. No que se refere ao Brasil, foi aprovada a Lei Geral de Proteção de Dados - LGPD, Lei nº 13.709, em agosto de 2018[3], sendo alterada pela Lei nº 13.853, de 8 de julho de 2019[4], iniciando sua aplicação em 18 de setembro de 2020.

2 https://eur-lex.europa.eu/legal-content/PT/TXT/?qid=1552642506978&uri=CELEX:32018R1725

3 http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm

4 http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Lei/L13853.htm#art1

A LGPD mudou normativas anteriores no que se refere à coleta e ao tratamento de dados, seja no setor público ou privado. Sua meta é proteger os dados pessoais de pessoas naturais, ou seja, vivas. Ela se baseia em regras relacionadas à coleta e ao tratamento desses dados, princípios constitucionais que se referem à inviolabilidade e à privacidade, previstos no artigo 5º da Constituição da República Federativa do Brasil (CRFB), que precisam ser cumpridos.

A transparência quanto ao manejo dos dados entre pessoas físicas e jurídicas também é basilar na LGPD. A Lei Geral de Proteção de Dados Pessoais informa, em seu artigo 5º, inciso II, que podem existir dados pessoais que precisam de ainda mais atenção e cuidado, que são os dados sensíveis:

> [...] 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 sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural (BRASIL, 2018, Cap. I, art. 5, inc. II).

No setor da saúde, informações contendo dados pessoais são usadas para estudos epidemiológicos. Essa prática pode trazer riscos para o indivíduo que fez parte da pesquisa, pois se acontecer um vazamento de dados, a pessoa pode sofrer *bullying* e estigmatização (RIBEIRO-ALVES; FRANCO, 2022). Concomitantemente a essas questões, existe a importância da produção dos dados e da disseminação e compartilhamento do conhecimento, o que pode favorecer novas descobertas e

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

estudos.

Nesse sentido, a principal questão que essa pesquisa busca responder é: como usar/manejar dados sensíveis de pesquisa relativos à saúde sem que o cidadão possa ser identificado? Tendo como base esse ponto, o artigo tem como objetivo geral analisar a dinâmica entre dados sensíveis de pesquisa relativos à saúde numa perspectiva da LGPD, e como objetivos específicos:

1. Descrever conceitos da LGPD para entendimento do tema do artigo;
2. Refletir sobre a privacidade dos dados de saúde no que concerne às novas tecnologias;
3. Analisar o tratamento dos dados pessoais sensíveis no setor da saúde;
4. Descrever as práticas relacionadas à anonimização dos dados sensíveis de pesquisa relacionados à saúde.

Com os objetivos específicos elencados acima foi possível entender melhor como está acontecendo o manejo dos dados sensíveis de saúde numa perspectiva da LGPD. Sendo assim, o trabalho se justifica por vários motivos, dentre eles podemos citar: maior compreensão da Lei Geral de Proteção de Dados – LGPD no que se refere aos dados sensíveis relacionados à saúde; melhor entendimento das novas tecnologias e adequação às normativas da LGPD quanto à privacidade dos dados; melhor entendimento dos dados de pesquisa no contexto dos dados sensíveis no setor da saúde; e compartilhamento de informações sobre anonimização de dados sensíveis de saúde (questão não muito divulgada ao grande público de uma forma mais palatável).

Para fornecer os subsídios necessários para elaborar este trabalho, as autoras definiram uma metodologia: foi realizada uma pesquisa exploratória[5] e bibliográfica. A coleta do material bibliográfico ocorreu por meio de realização de busca de artigos no Google, no Google Acadêmico, no Portal Capes e em livros.

[5] "Visa desenvolver, esclarecer e modificar conceitos e ideias para estudos posteriores. (...) constituem a primeira etapa de uma investigação mais ampla. (...) envolvem levantamento bibliográfico e documental" (SILVA, 2018, p. 17).

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

Nesse sentido, o artigo está estruturado da seguinte forma: na primeira seção, explicita-se um pouco mais a metodologia da pesquisa; na segunda seção, são apresentados conceitos da LGPD a fim de introduzir o leitor ao tema; na terceira seção, busca-se refletir sobre a privacidade dos dados de saúde no contexto das novas tecnologias; na quarta, é realizada uma análise sobre os dados pessoais sensíveis no setor de saúde; na quinta, discute-se a questão da anonimização dos dados sensíveis de pesquisa relacionados à saúde; na sexta, aborda-se a discussão dos resultados do estudo; e na sétima seção, tem-se a conclusão da pesquisa.

2 Metodologia

A metodologia exploratória foi aplicada com o intuito de obter uma compreensão aprofundada do tema, gerar um debate e sugerir possíveis respostas para a questão da pesquisa. O objetivo do artigo foi responder à questão: "como usar/manejar dados sensíveis de pesquisa relativos à saúde sem que o cidadão possa ser identificado"? Por conta disso, a LGPD foi examinada com rigor. Em alguns momentos, de forma a oferecer uma visão mais geral da Lei, enquanto em outros buscou-se discutir o tema de forma mais específica.

Em relação à metodologia bibliográfica, envolveu-se pesquisas nas mais diversas fontes, como livros, artigos e publicações não seriadas. Para isso, recorreu-se ao Google, ao Google Acadêmico e ao Portal Capes. Essas plataformas foram utilizadas com a finalidade de compreender a questão da pesquisa de forma mais ampla, possibilitando assim identificar lacunas existentes nessa área e propor soluções.

3 Lei geral de proteção de dados – LGPD

A LGPD alterou normas para a coleta e o tratamento de dados pelos setores público e privado. Ela teve como objetivo garantir a privacidade e a proteção de dados pessoais. Como função, buscou criar normas para a coleta e o tratamento de

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

dados. No que se refere a sua meta, visou garantir a privacidade do cidadão, bem como proteger seus dados pessoais e estimular a transparência na relação entre pessoas físicas e jurídicas. Conforme indicação abaixo, o artigo 5º da LGPD nos traz os principais conceitos para compreender quem são os personagens que essa normativa se refere.

> [...] V - titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;
>
> VI - controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;
>
> VII - operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador; [...]
>
> VIII - encarregado: pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD);
>
> IX - agentes de tratamento: o controlador e o operador [...] (BRASIL, 2018, Cap. I, art. 5, inc. V, VI, VII, VIII, IX).

Nesse contexto, as instituições precisam se adequar, criar papéis institucionais, buscar os regulamentos e melhores práticas para oferecer ao cidadão todas as prerrogativas que a Lei lhe oferece. A maioria das organizações pedem os dados pessoais de seus clientes, no entanto, é preciso redobrar os cuidados com essas informações, pois para a LGPD dados pessoais são informações que podem identificar uma pessoa, seja de forma direta ou indireta.

Nome, sobrenome, data de nascimento, CPF, carteira de identificação civil, número do passaporte e número do título de eleitor são elementos que, correlacionados, fazem essa conexão entre os dados e o indivíduo. Apenas com o nome e o CPF uma pessoa já se torna identificada, e com dados como profissão, idade e empresa que trabalha, ela se torna muito facilmente identificável (LIMA et al., 2023). Sobre o tratamento dos dados sensíveis, o art. 11 destaca o seguinte:

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

> O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses: [...]
>
> II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para: [...]
>
> c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis; [...]
>
> e) proteção da vida ou da incolumidade física do titular ou de terceiro;
>
> § 1º Aplica-se o disposto neste artigo a qualquer tratamento de dados pessoais que revele dados pessoais sensíveis e que possa causar dano ao titular, ressalvado o disposto em legislação específica. [...]
>
> § 5º É vedado às operadoras de planos privados de assistência à saúde o tratamento de dados de saúde para a prática de seleção de riscos na contratação de qualquer modalidade, assim como na contratação e exclusão de beneficiários (BRASIL, 2018, Cap. II, art. 11, inc. II).

Conclui-se que o tratamento de dados pessoais sensíveis pode acontecer sem autorização do titular quando for utilizado para pesquisa ou quando houver interesse público. No entanto, sempre visando garantir as liberdades fundamentais. Da mesma forma, proíbe-se às operadoras de planos de saúde privados fazerem uma seleção dos titulares dos dados por conta do risco de algum prejuízo que eles possam oferecer (MACHADO, C.; DOURADO; MACHADO, F., 2021).

A preocupação maior com os dados sensíveis se deve ao fato de pessoas sofrerem penalizações sociais se seus dados mais íntimos de saúde vierem a público. Pessoas podem perder o emprego ou serem estigmatizadas no meio social que convivem. Nesse sentido, é primordial o zelo com o dado sensível para que ele não seja vazado, alterado, destruído ou compartilhado de forma ilegal (LIMA *et al.*, 2023). Sendo assim, houve preocupação da LGPD também com dados de pesquisa, o que pode ser observado no artigo 13, citado abaixo:

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

> [...] § 1º A divulgação dos resultados ou de qualquer excerto do estudo ou da pesquisa de que trata o caput deste artigo em nenhuma hipótese poderá revelar dados pessoais.
>
> § 2º O órgão de pesquisa será o responsável pela segurança da informação prevista no caput deste artigo, não permitida, em circunstância alguma, a transferência dos dados a terceiro.
>
> § 3º O acesso aos dados de que trata este artigo será objeto de regulamentação por parte da autoridade nacional e das autoridades da área de saúde e sanitárias, no âmbito de suas competências.
>
> § 4º Para os efeitos deste artigo, a pseudonimização é 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, Cap. II, art. 13).

Portanto, é aconselhável que os resultados das pesquisas que contenham dados pessoais sejam armazenados em local seguro e com a devida curadoria, para não ocorrerem riscos de vazamento. O artigo 13 também indica que deverá ser realizado um tratamento no dado para que não seja possível associar um dado a uma pessoa. Essa questão será tratada na quarta seção deste trabalho.

4 Privacidade dos dados de saúde num mundo tecnológico

O mundo está novamente entrando numa nova revolução. Alguns autores já denominam essa nova transformação tecnológica como Revolução Digital, e comparam as mudanças atuais com as que ocorreram a partir da Revolução Industrial com o surgimento da máquina a vapor (Vainzof, 2021). Essas transformações radicais mudam a forma com a qual a sociedade se relaciona, produz, protege e armazena a informação.

Sistemas de inteligência artificial, por exemplo, possuem condições de arma-

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

zenar enormes quantidades de dados, e questões relativas à segurança desses dados ainda estão num campo nebuloso (Magrini; Guedes, 2021). Mesmo os *smartphones* e *wearables* armazenam e compartilham dados, e não se sabe bem como eles são tratados. Durante a pandemia, algoritmos de *machine learning*[6] e *deep learning*[7] foram usados para acelerar a análise dos dados de saúde e trazer mais precisão aos diagnósticos (ARBIX; BRANDÃO; CAMARGO, 2021).

Corroborando com Arbix, Brandão e Camargo (2021), Fornazin e colaboradores (2021) apresentaram um estudo bibliométrico que reflete um aumento de termos relacionados à saúde digital, tais como: *machine learning* e *mobile health*. Ademais, percebeu-se que a partir de 2017 mais artigos científicos relacionam o tema saúde às tecnologias digitais relacionadas à inteligência artificial.

Nesse sentido, o setor de saúde possui muitos desafios no que se refere à gestão dos dados sensíveis, sendo preciso ressaltar que isso acontece tanto na saúde pública como na privada. O uso das novas tecnologias pode trazer mais acurácia e rapidez nos diagnósticos, mas como está a questão da proteção dos dados? Como são tratados os dados da telemedicina, por exemplo? As consultas podem ser gravadas sem que o paciente saiba, infelizmente. Informações sensíveis são proferidas numa consulta usando esse tipo de tecnologia. Quanto aos prontuários eletrônicos, as pessoas que têm seu acesso passaram por algum tipo de treinamento relativo à LGPD?

De acordo com Arbix, Brandão e Camargo (2021, p. 494), "os dados pessoais de saúde podem incluir dados demográficos, imagens, resultados de laboratório, dados de testes genéticos e gravações de dispositivos médicos ou sensores". Tudo

6 "O machine learning é um tipo de IA que permite que os computadores aprendam por si, automaticamente, a partir de dados a ele alimentados, contendo problemas e soluções, ou seja, 'aprendem' sem serem especificamente programados, como seria necessário nos antigos sistemas expertos, que aprenderiam a partir das experiências e conhecimento de humanos" (Machado, C.; Dourado; Machado, F., 2021, p. 505).

7 "O deep learning é uma subcategoria de machine learning que usa redes neurais artificiais que podem 'aprender' relações extremamente complexas entre dados de diversas características, com e sem a utilização de rótulos, capazes de manipular tais dados altamente complexos e heterogêneos, como os que são gerados no atendimento clínico moderno, tais como registros médicos, imagens clínicas, dados de monitoramento contínuo de sensores e dados genômicos, e produzir resultados capazes de auxiliar em previsões e diagnósticos clinicamente relevantes. O ponto chave é que são formulações dinâmicas que orientam decisões, e é essa variável conjugada com ações que atrai o debate jurídico" (Machado, C.; Dourado; Machado, F., 2021, p. 505).

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

isso faz parte de um vasto conjunto de tecnologias de geração e coleta de dados. Estão envolvidos servidores, aplicativos móveis, sensores, *wearables*, computadores, etc. Atualmente, a integração de diversos sistemas gerindo todos esses equipamentos é relativamente fácil, mas será que os dados sensíveis estão sendo protegidos? As autoridades de saúde e os órgãos de regulação estão cumprindo seu papel na fiscalização?

Arbix, Brandão e Camargo (2021) informam que os dados de saúde têm sua proteção garantida pelos *softwares* que fazem a gestão desses dados. Ainda de acordo com os autores, a Agência Nacional de Proteção de Dados Pessoais (ANPD) deve em breve trazer anúncios no que se referem às boas práticas de algoritmos de inteligência artificial no uso dos dados pessoais da saúde. Dessa forma, nesta sociedade altamente tecnológica e conectada do início do século XXI terão também regulações e o desenvolvimento de um ecossistema que protejam os dados sensíveis de saúde.

5 Dados pessoais sensíveis no setor da saúde

Os leitores mais atentos provavelmente já leram ou ouviram falar diversas vezes que "os dados são o novo petróleo", frase de Ajay Banga, CEO Global da Mastercard (Linkedin, 2019). Com a advento das grandes plataformas tecnológicas e *Big Techs*, como Google, Amazon, Facebook, Tik Tok, entre outras, essa questão do valor dos dados pessoais ficou mais evidente. Temos acesso gratuito, mas em compensação nossos dados são fornecidos para essas empresas ("se você não está pagando por algum serviço, então o pagamento é você", ou seja, seus dados)[8].

8 Citação do filme O Dilema das Redes (2020).

Dessa forma, é possível que os algoritmos direcionem o usuário para algum produto que ele nem pense em comprar naquele momento, mas em alguma outra circunstância já fez pesquisas sobre o produto na *internet*. As empresas usam essas pesquisas para oferecer mercadorias e/ou serviços aos usuários, e quando eles finalmente clicam nos anúncios, as *Big Techs* ganham dinheiro com dados dos

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

internautas e o círculo se fecha.

Similarmente acontece com as pesquisas sobre saúde que os usuários fazem na *internet*. A navegação do usuário fica registrada nos *cookies*[9] do computador, e em algum momento isso será usado pelas *Big Techs*, seja para oferecer um oxímetro, um medidor de pressão, um medidor do índice de glicose ou teste de gravidez. Tudo gira em torno dos dados. Pensando nesse espectro, os dados pessoais sensíveis são os que precisam de maior proteção, conforme representa a Figura 1.

9 De acordo com Cahn e colaboradores (2016), cookies são constituídos de arquivos de texto que são deixados no navegador do internauta quando ele acessa um servidor. "Embora os cookies sejam um elemento intrínseco dos aplicativos da web, seu uso tem implicações importantes na privacidade do usuário. De fato, um dos principais objetivos das corretoras de dados e dos anunciantes on-line é reunir o máximo possível de informações sobre os usuários com o objetivo de fornecer anúncios direcionados" (Cahn et al., 2016, p. 891).

Figura 01 – Dados

Fonte: Elaborada pelas autoras

Quando se pensa sobre ética, a confidencialidade é um dos pontos mais importantes. Ela impõe sigilo nas informações, sejam técnicas ou pessoais. Além disso, na saúde existe a questão do segredo profissional, que é um dispositivo moral nesse setor (VILLAS-BÔAS, 2015). Nesse cenário surgem preocupações sobre os dados sensíveis que são usados na saúde, pois empresas de saúde também usam a *internet*, trafegam, armazenam e disponibilizam dados. Sendo assim, a preocupação quanto à proteção desses dados é elevada a um outro patamar, pois como foi dito, pessoas podem ser estigmatizadas, sofrer *bullying*, se dados sensíveis de saúde não forem tratados da forma correta.

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

No Brasil, esse tipo de dado tem sido armazenado e tratado nos mais diferentes sistemas de informação, seja para estudos relativos às epidemias, à demografia ou para elaborar novos produtos e serviços. A partir desse contexto, diversas variáveis surgem no que se refere a "necessidades coletivas, privacidade, inviolabilidade da intimidade, dignidade da pessoa humana, autodeterminação informativa, livre desenvolvimento de personalidade, desenvolvimento tecnológico, direitos humanos, sigilo de dados" (Aragão; Schiocchet, 2020, p. 699). No Ministério da Saúde existem vários sistemas com os mais diversos dados de saúde, dentro eles, de acordo com o Quadro 1, podemos citar:

Quadro 1 – DataSUS – Sistemas de Saúde

Nome do Sistema	Sigla dos Sistema	Função do Sistema
Sistema de Controle de Diagnóstico Laboratorial	DSTAIDS-DIAG	"Avaliar o desempenho individual dos laboratórios mediante a precisão e a acuracidade dos resultados, quando comparados à média geral obtida por todos os laboratórios incluídos na Rede Nacional de Diagnóstico do PN-DST/AIDS."
Sistema de Cadastro das Pessoas para o teste de HIV através do método de fluido oral	DSTAIDS-FLUIDO ORAL	"O sistema tem o objetivo de manter o cadastro das pessoas que fazem o teste de HIV usando o método de fluido oral."
Sistema de Informação de Agravos de Notificação	NOVO SINAN	"O Sistema de Informação de Agravos de Notificação – Sinan é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de doenças de notificação compulsória (Portaria de Consolidação nº 4, de 28 de setembro de 2017, anexo V – Capítulo I)."
Serviço de Atendimento Móvel de Urgência do SUS	E-SUS SAMU	"Sistema utilizado para registro de ocorrências médicas que, baseado na avaliação do médico regulador, podendo enviar ou não uma ambulância."

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

Sistema de Copagamento para Expansão da Farmácia Popular do Brasil	FARMACIA POPULAR	"O Programa Farmácia Popular do Brasil vem a ser uma iniciativa do Governo Federal que cumpre uma das principais diretrizes da Política Nacional de Assistência Farmacêutica. O Sistema permite que as Farmácias/Drogarias cadastradas no programa realizem a dispensação de medicamentos para as patologias: hipertensão, asma e diabetes (Programa saúde não tem preço)."

Fonte: Elaborado pelas autoras a partir de DataSUS (BRASIL, [201?]).

Como pode se observar, os sistemas armazenam os mais diversos dados sensíveis de saúde. Sejam de HIV, de doenças cujos medicamentos são atendidos pelo Programa Farmácia Popular ou informações de alguém que precisou ser atendido pelo SAMU. Nesse sentido, há preocupações relevantes quanto aos riscos envolvidos no vazamento desses dados. É preciso que eles estejam protegidos pela LGPD. Assim sendo, seja empresa pública ou privada, é preciso se adequar à legislação, ter responsabilidade e cuidado com os dados para que os cidadãos tenham seus direitos garantidos.

Portanto, seja funcionário, terceirizado, contratado ou estagiário, todos necessitam passar por treinamentos para saber lidar com esse tipo de dado. Caso os profissionais envolvidos não passem por um treinamento, podem ocorrer vazamento de dados sensíveis de saúde, e entre as punições, de acordo com o artigo 52 da LGPD, temos:

> [...] I - advertência, com indicação de prazo para adoção de medidas corretivas;
>
> II - multa simples, de até 2% (dois por cento) do faturamento da pessoa jurídica de direito privado, grupo ou conglomerado no Brasil no seu último exercício, excluídos os tributos, limitada, no total, a R$ 50.000.000,00 (cinquenta milhões de reais) por infração;
>
> III - multa diária, observado o limite total a que se refere o inciso II;

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

> IV - publicização da infração após devidamente apurada e confirmada a sua ocorrência;
>
> V - bloqueio dos dados pessoais a que se refere a infração até a sua regularização;
>
> VI - eliminação dos dados pessoais a que se refere a infração (BRASIl, 2018, Cap. VIII, art. 52, inc. I, II, III, IV, V e VI).

Nota-se que o não cumprimento da lei pode trazer diversas consequências. Dessa maneira, o DataSUS implementou algumas políticas institucionais para assegurar a privacidade dos dados dos cidadãos e trazer mais segurança a esses dados de saúde conforme citado abaixo:

Regras de firewall de redes para filtro de pacotes e bloqueio de portas de acesso;

- Firewall de aplicação web para proteção contra ataques como falsificação de solicitação entre sites, cross-site-scripting (XSS), inclusão de arquivos e SQL Injection. Sendo uma defesa de protocolo da camada 7 (no modelo OSI);
- Software de proteção antivírus e antimalware para servidores;
- Rotinas de proteção de dados;
- Segurança das comunicações com a utilização de protocolos de comunicação seguros – como TLS/HTTPS – e aplicativos com criptografia fim a fim;
- Processo de Gestão de Acessos que busca garantir o uso de serviços a usuários autorizados e, ao mesmo tempo, prevenir que usuários não autorizados tenham acesso a esses serviços. Além disso, as concessões deverão ser aplicadas respeitando o princípio de privilégios mínimos e apenas pela duração de tempo necessário. O processo é implementado para decidir quem deverá ter direitos de acesso sobre os ativos de TI (BRASIL, 2022, p. 16-17).

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

Assim sendo, fica evidente a preocupação relacionada às questões que envolvam vazamento de dados. Há utilização de *softwares* adequados, processos de gestão de dados e segurança nas comunicações internas e externas. Todas essas ferramentas propiciam uma melhor segurança e privacidade dos dados.

No entanto, segundo Aragão e Schiocchet (2020), há ainda um caminho longo a trilhar sobre a segurança dos dados pessoais sensíveis. As autoras destacam preocupações relativas à transparência, à regulamentação e à adaptação do Sistema Único de Saúde (SUS) à LGPD. Elas ressaltam que não há clareza suficiente sobre como as entidades de saúde gerenciam e protegem os dados dos usuários. Além disso, enfatizam que existe falta de normatização relativa à troca de informações dentro do SUS com instituições que ele se relaciona. As autoras também se preocupam sobre a falta de discussão no âmbito acadêmico sobre a aplicação da LGPD às informações que o SUS produz, trata e compartilha.

A partir desse apontamento preocupante, na próxima seção veremos questões relacionadas à anonimização dos dados para que não ocorra acesso e identificação indevida.

6 Dados sensíveis de pesquisa relacionados à anonimização na saúde

O artigo 7º da LGPD informa no seu inciso IV que o tratamento de dados pessoais somente poderá ser realizado "[...] para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização" (BRASIL, 2018, Cap. II, art. 7, inc. IV). C. Machado, Dourado e F. Machado (2021) entendem que quando a LGPD menciona dados pessoais para uso em órgãos de pesquisa, seriam dados usados numa universidade, fundação ou instituto. Nesse caso, seriam seguidas normas da Sociedade Brasileira para Progresso da Ciência (SBPC), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), dentre outros.

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

No entanto, primeiramente é preciso entender o que são dados de pesquisa para posteriormente contextualizar e compreender seu uso no setor da saúde. Esses tipos de dados, entendidos como científicos, podem ter as seguintes categorias quanto a sua origem: dados originados a partir da observação, dados elaborados por modelos computacionais e dados criados a partir da experimentação científica. Dessa forma, abaixo é possível destacar um breve resumo para o melhor entendimento de cada categoria:

Dados observacionais: são dados em grandes volumes onde a observação não poderá ser repetida. São necessários cuidados redobrados no que se refere ao armazenamento destes dados/metadados, recomendando-se guarda permanente (*National Science Board*, 2005). Estes dados podem ser monitorados por algum tipo de equipamento ou sensor, como, por exemplo, observar e registrar o ruído de um aeroporto (Mac Dewitt Wallace Library, 2023). No que se refere aos estudos epidemiológicos, Costa e Barreto (2003) informam que os estudos a partir dos dados observacionais podem ser categorizados como descritivos e analíticos. Os descritivos buscam determinar quando, onde e quem foi acometido pela doença. No que se refere aos estudos observacionais analíticos, eles buscam pesquisar se existe uma correlação entre uma "exposição e uma doença ou condição relacionada à saúde" (COSTA; BARRETO, 2003, p. 194).

Dados computacionais/ dados de simulação: imitam o mundo real por meio de algum modelo de computador. Temos como exemplo a previsão do tempo, modelos econômicos, atividade sísmica, etc. O modelo de computador faz uma predição a partir dos dados. O modelo pode ser mais importante que os dados (MAC DEWITT WALLACE LIBRARY, 2023). Horvitz e Kristan (2009, p. 84) apontam que "análises computacionais intensivas servirão como base para modelagem e visualização dos dados populacionais intrinsecamente de alta dimensão". Esses modelos computacionais foram bastante usados na pandemia da COVID-19. Os dados semanais de

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

óbitos, hospitalizações e contaminados eram inseridos nos modelos e a partir disso se concluía se determinado local estaria entrando numa onda de contaminação ou não.

Dados experimentais: são dados que provêm da intervenção do pesquisador que produz mudanças num determinado objeto de pesquisa (MAC DEWITT WALLACE LIBRARY, 2023). Não é necessária a guarda permanente do dado num ambiente controlado, mas é importante destacar que por condições que independem da vontade do pesquisador, às vezes o ambiente da pesquisa experimental muda de alguma forma, e o resultado não poderá ser replicado. Portanto, isso precisa ser avaliado pela instituição, visto que algumas pesquisas têm custos muito altos e seus resultados não podem ser perdidos (NATIONAL SCIENCE BOARD, 2005). Refletindo sobre estudos epidemiológicos, a partir dos dados experimentais, esse dado irá provir de um estudo onde houve uma tentativa de mudar o curso da doença, ou por mudança de hábitos do paciente ou por algum tipo de tratamento (BONITA; BEAGLEHOLE; KJELLSTRÖM, 2010).

Entretanto, Borgman (2015) caracteriza dados de saúde quanto a sua origem numa **quarta categoria**, nomeando-os como "**registros**". Nesse sentido, fazendo uma reflexão sobre o que foi exposto acima, observa-se que os registros se relacionam aos outros três tipos de dados de pesquisa; eles apresentam grandes riscos de vazamento dos dados sensíveis de saúde. Inclusive, nos dados computacionais que geram modelos há também grande preocupação, visto que para gerar modelos no computador é preciso que haja dados, e esses dados podem ser de uma pessoa, e precisam estar enquadrados na LGPD.

No que tange aos riscos de divulgação, Templ *et al.* (2014) expõem três tipos de riscos: risco de identidade, risco de atributo e risco de inferência. O risco de identidade acontece quando uma pessoa é associada a um determinado registro por alguém não autorizado a ter acesso a essa informação no sistema. O risco de atributo ocorre quando informações específicas de um determinado grupo, que estão num local específico, são divulgadas. Um exemplo hipotético: "Hospital Nossa

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

Senhora de Lourdes divulga que na ala feminina, todas as mulheres de 25 a 35 anos estão com COVID". A partir dessa informação, todos que conhecem as mulheres nessa faixa etária internadas nesse hospital saberão dessa informação, ou seja, a informação se tornou pública.

No que se refere ao risco de inferência, este acontece quando uma pessoa não autorizada tem acesso ao sistema e consegue extrair valor de alguma característica de um indivíduo, que não seria possível se o dado não fosse divulgado, nem que para isso essa pessoa estranha ao sistema usasse um modelo computacional de regressão.

Para Ribeiro-Alves e Franco (2022), os riscos de divulgação podem ser enquadrados em alguns grupos; abaixo citamos dois:

1. Identificadores diretos: CPF, identidade civil, telefones, nomes e endereços;
2. Identificadores variáveis (quase identificadores): estado se saúde, sexo, renda, idade, nacionalidade, local que o indivíduo mora (região) – muito usados para estatísticas.

É possível observar alguns dados sensíveis acima, mas, para o que interessa para esta pesquisa, iremos focar no dado "saúde". Como foi exemplificado acima, uma das formas possíveis de ocorrer um vazamento acontece quando um indivíduo não autorizado acessa a base de dados e consegue ter acesso aos dados sensíveis de saúde das pessoas.

Entretanto, no que se refere à proteção desses dados, Almeida *et al.* (2020, p. 2489), descrevem que existem formas de protegê-los. Por exemplo, a anonimização que consiste na "aplicação de medidas técnicas para impossibilitar a associação direta ou indireta dos dados ao indivíduo, e a pseudoanonimização que geralmente remove identificadores e os substitui por um código-chave único". Assim sendo , os dados estariam protegidos. Nesse contexto, entende-se que existem técnicas que podem ser úteis para realizar essa desidentificação do indivíduo.

Ribeiro-Alves e Franco (2022) esclarecem que existem métodos para garantir a segurança desses dados. Estes métodos medem o risco de divulgação dos dados,

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

tornam os microdados[10] anônimos ou não identificados e comparam dados originais e modificados. Todos esses procedimentos são realizados com cautela e cuidado, pois, caso contrário pode haver vazamento de dados sensíveis de saúde. Para realizar esse trabalho existe uma ferramenta computacional, o RStudio, que pode auxiliar na anonimização dos dados.

10 Dados estatísticos de uma pessoa, empresa ou família que fizeram parte de uma pesquisa (Silva, 2015).

"O R é um ambiente computacional e uma linguagem de programação que vem progressivamente se especializando em manipulação, análise e visualização gráfica de dados" (Carvalho, 2020, p. 3). Esse *software* faz a "desidentificação de informação pessoal em microdados de pesquisa usando ferramentas conjuntamente conhecidas como de controle de divulgação estatística, SDC, (Statistical Disclosure Control)" (Ribeiro-Alves; Franco, 2022, p. 5). Para isso, segundo os autores, o R usa o pacote sdcMicro'[11]. Esse pacote[12] faz a anonimização dos microdados.

11 http://cran.nexr.com/web/packages/sdcMicro/index.htm (Templ; Kowarik; Meindl, 2018).

12 "O pacote R sdcMicro serve como uma implementação de classe S4 orientada a objeto e fácil de manusear métodos SDC para avaliar e anonimizar conjuntos de microdados confidenciais. Inclui todos os métodos populares de risco e perturbação de divulgação. O pacote realiza o recálculo automatizado de contagens de frequência, medidas de risco individuais e globais, perda de informações e estatísticas de utilidade de dados após cada etapa de anonimização. Todos os métodos são altamente otimizados em termos de custos computacionais para poder trabalhar com grandes conjuntos de dados. Os recursos de relatórios que resumem o processo de anonimização também podem ser facilmente usados pelos profissionais" (Templ; Kowarik; Meindl, 2015, p. 1).

Métodos para medir o risco de divulgação

Deste método fazem parte o risco individual, o risco global e o risco domiciliar. Eles fazem uma avaliação dos dados disponíveis e o ambiente onde ele é divulgado. Alguns dados podem ter identificadores diretos, o que pode levar à identificação de uma pessoa, por isso a importância da avaliação de risco. Neste artigo, o foco é o risco individual.

Fazem parte desse método, por exemplo, o **k-anonimato; *l*-diversidade** e o **t-proximidade**. O k-anonimato é elaborado em cenários onde as bases de dados são pequenas e a pessoa não autorizada no sistema sabe de quem são os dados da pesquisa. Visando manter o titular do dado anônimo, constrói-se um padrão de

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

variáveis-chave onde o k representa os registros da amostra. Normalmente estabelece-se o valor 3 (três) para k. Dessa forma, três registros da pesquisa irão possuir um padrão, no entanto, sempre será $k \geq 3$ (Silva, 2015). Entretanto, para Affonso e Sant'ana (2017), o valor seria $k > 1$.

O *l*-diversidade vem como alternativa ao método anterior, pois tem havido críticas quanto à suficiência do k-anonimato para anonimizar dados sensíveis (Silva, 2015; Ribeiro-Alves; Franco, 2022). Silva (2015, p. 18) explica que "tem *l*-diversidade se contém pelo menos l valores para as variáveis sensíveis". No entanto, esse método também sofre crítica, uma vez que o dado pode ser interpretado por questões de assimetria/simetria. Essas situações podem ocorrer na hora da distribuição dos valores de uma variável sensível em um banco de dados.

Quanto ao t-proximidade, este método consegue atingir os objetivos relacionados à anonimização dos dados. Diferentemente dos métodos anteriores, este método faz uma distribuição dos dados sensíveis de forma que não seja possível perceber simetrias nem assimetrias na base. No entanto, poderá trazer consequências para a utilidade dos dados no que se refere a fazer uma correlação das variáveis (SILVA, 2015).

Métodos de desidentificação de microdados

Ribeiro-Alves e Franco (2022) descrevem alguns métodos que fazem parte da desidentificação de microdados, como, por exemplo: recodificação; supressão local; pós-aleatorização (pram); microagregação; adição de ruído; embaralhamento. De forma geral, fazem transformações nos dados originais por meio de agregação de informação ou reorganização dos dados. Dessa forma, podem levar um certo ruído aos dados, o que torna difícil sua identificação para quem não conhece os dados estruturados da pesquisa.

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

Métodos de medida da perda de informação

De acordo com o *Manual prático de anonimização de dados de pesquisa com o R*[13], elaborado na Fiocruz, os dados anonimizados não devem ter estruturas muito diferentes dos dados originais e precisam conter uma precisão elevada caso precisem ser analisados. Para isso, podem ser usadas algumas ferramentas gerais e específicas do RStudio. Após configurados os dados, eles são comparados para verificar se atendem ao que o pesquisador precisa, e se realmente houve minimização dos riscos de vazamento dos dados; caso seja observado que a amostra não atende aos padrões necessários, aconselha-se refazer o trabalho.

13 https://www.arca.fiocruz.br/bitstream/handle/icict/56398/Manual%20Pr%c3%a1tico%20de%20Anonimiza%c3%a7%c3%a3o%20de%20Dados%20de%20Pesquisa%20com%20o%20R.pdf?sequence=2&isAllowed=y

Dados de pesquisa na saúde precisam de maior zelo no seu tratamento para invasores não conseguirem acesso ou para não sofrerem vazamento. No entanto, a meta é sempre que eles continuem com alta serventia dentro da instituição para manter seu valor de pesquisa.

7 Discussão dos resultados

Como resultados obtidos a partir da metodologia descrita, o estudo disponibilizou algumas formas possíveis de anonimização de dados pessoais sensíveis de pesquisa relacionados à saúde. De forma geral, é possível especificar nos Quadros 2 e 3 os métodos sugeridos por este estudo para anonimização dos dados sensíveis de pesquisa em saúde.

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

Quadro 2 – Métodos para medir o risco de divulgação

k-anonimato	Método que garante que cada pessoa esteja indistinguível de pelo menos de k ≥ 3 indivíduos nos dados (Silva, 2015), ou de k > 1 (Affonso; Sant'ana, 2017).
l-diversidade	Método que garante que cada grupo de dados com variáveis-chave compartilhadas tenha no mínimo *l* valores distintos para variáveis sensíveis (Silva, 2015; Ribeiro-Alves; Franco, 2022).
t-proximidade	Método que assegura a distribuição dos dados sensíveis de forma que não seja possível perceber simetrias nem assimetrias na base (Silva, 2015).

Fonte: Elaborada pelas autoras

O objetivo do k-anonimato é garantir que a anonimização de um registro (pessoa) seja resguardada ao assegurar que existam outros k−1 registros com valores iguais no que se refere a cada um de seus atributos (Barreto; Henrique, 2021). Dessa forma, por exemplo, um conjunto de pessoas que possuam diabetes seriam agrupadas numa determinada faixa etária e num determinado intervalo de níveis de glicose com seus dados identificáveis suprimidos. Os atributos sensíveis seriam os mesmos de outras pessoas, sendo definida a quantidade de pessoas com os mesmos atributos pelo valor de k. Como resultado, o conjunto de dados após esse tratamento teria informações menos específicas e menos identificáveis.

O *l*-diversidade pode ser usado junto com o k-anonimato, ou separadamente (Alves, 2021). Usando os dois métodos, além de "k", com o mesmo atributo, dentro de um conjunto de dados, ele teria "l" diversidade. Isso significa que além de cada pessoa ter seu atributo igual a ao menos duas pessoas (k > 1), onde a identificação não seria possível, esse conjunto de dados teria uma diversidade de condições médicas dentro de cada grupo de registros com atributos compartilhados. Por exemplo: teria um conjunto de dados de pessoas numa faixa etária de 40-50 anos, com faixa de glicose em jejum acima de 126 mg/dl, onde k > 1. Desse grupo fariam parte pessoas na mesma faixa etária com hipotireoidismo. Cada atributo k-anonimato

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

precisaria possuir ao menos *l* valores diferentes representados em cada respectiva classe de equivalência (Prata et al., 2020). Entrando outros atributos, dificultar-se--ia ainda mais a identificação dos dados pessoais de saúde. É importante ressaltar, entretanto, que o uso do *l*-diversidade em um conjunto de dados pequenos pode levar à identificação de uma pessoa justamente por esse fator da diversidade, o que introduz, neste ponto, as questões de assimetria/simetria dos dados.

No que se refere ao t-proximidade, ele é usado para proteger a privacidade em conjuntos de dados sensíveis, limitando a variação de diagnósticos entre indivíduos com atributos parecidos. Se, por exemplo, 30% dos pacientes no conjunto total tem diabetes, cada subgrupo de pacientes deve ter uma porcentagem semelhante de pacientes diabéticos, tornando difícil a identificação individual baseada nessas informações. Isso seria possível porque o t-proximidade garante que a frequência com que cada característica sensível dentro de um conjunto de indivíduos com características semelhantes seja parecida com a sua distribuição geral no conjunto de dados (Mendonça, 2018).

Quadro 3 - Métodos de desidentificação de microdados

Recodificação; supressão local; pós-aleatorização (PRAM); microagregação; adição de ruído; embaralhamento	Fazem transformações nos dados originais por meio de agregação de informação ou reorganização dos dados.

Fonte: Elaborado pelas autoras a partir de Ribeiro-Alves e Franco (2022).

De forma mais detalhada, nos métodos do Quadro 3, de acordo com apontamentos de Ribeiro-Alves e Franco (2022), a recodificação combina categorias para formar uma nova, diminuindo assim a identificabilidade dos dados. A supressão local, remove valores em ao menos uma variável categórica, o que possibilita o anonimato dos indivíduos. A pós-aleatorização (PRAM) transforma variáveis categóricas em outras; para isso, usa um modelo probabilístico de transição predefinida, dificultando a identificação dos registros. A microagregação divide os registros e

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

grupos, onde cada grupo terá sua variável agregada. Quanto à adição de ruído, esta técnica introduz variações aleatórias nos microdados, salvaguardando-os contra um matching com dados externos.

No que concerne ao embaralhamento, este método sintetiza valores sensíveis baseados em variáveis não confidenciais e classifica os valores simulados de acordo com os valores originais por conta da aplicação do mapeamento reverso. Dessa forma, "valores classificados dos valores simulados são substituídos pelos valores classificados dos dados originais" (Ribeiro-Alves; Franco, 2022, p. 23). Isso torna difícil rastrear os dados originais, protegendo assim a privacidade das informações.

De maneira específica, observou-se particularmente que o software RStudio possui ferramentas como Controle de Divulgação Estatística (SDC) com potencial de auxiliar na anonimização dos dados. De acordo com Ribeiro-Alves e Franco (2022), essa evolução no software destaca a rápida mudança e a necessidade de adaptação constante no campo da proteção de dados pessoais sensíveis, especialmente dados de saúde.

Todos esses métodos podem proteger a privacidade dos dados sensíveis de saúde em pesquisas científicas. A partir de suas particularidades, cada instituição precisa avaliar e concluir qual se adequa mais a sua realidade.

8 Conclusão

O artigo buscou trazer ao leitor reflexões e respostas sobre os dados sensíveis de pesquisa relacionados à saúde numa perspectiva da Lei Geral de Proteção de Dados. Expôs que as novas tecnologias vieram para somar, mas também precisam ser reguladas.

A pesquisa descreveu e contextualizou tópicos da legislação, discutiu abordagens sobre o ambiente tecnológico e de anonimização dos dados. No que se refere a este ponto, este estudo buscou expor e detalhar métodos que atendessem ao objetivo geral da pesquisa. Foi explicado tanto o método para medir o risco de divulgação

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

de dados sensíveis de saúde, como o método de desidentificação de microdados. Dessa forma, respondeu à questão do artigo: "como usar/manejar dados sensíveis de pesquisa relativos à saúde sem que o cidadão possa ser identificado"?

Esse debate trouxe à tona diversas questões pertinentes à legislação vigente e à necessidade de adaptação às normas, sobretudo em ambientes de pesquisa que lidam frequentemente com esse tipo de informação. Explorou questões relacionadas à infraestrutura tecnológica e à seleção de ferramentas que possam garantir a privacidade dos dados. No entanto, este estudo não esgota o assunto, trouxe apenas uma contribuição visando encontrar formas de responder às lacunas encontradas com relação a esse tema.

Para pesquisas futuras, estudos mais profundos sobre a privacidade de dados sensíveis de saúde numa perspectiva do uso da inteligência artificial são muito relevantes e necessários. Normativas sobre isso ainda estão sendo regulamentadas no Congresso Nacional brasileiro, o que torna essa questão bem preocupante e carente de estudos.

Para o contexto deste artigo, pode-se concluir que ao seguir a lei e se adaptar às regulamentações, as instituições podem contribuir para um ambiente de pesquisa mais seguro e confiável.

Referências

AFFONSO, Elaine Parra; SANT'ANA, Ricardo César Gonçalves. Preservação da privacidade no acesso a dados por meio do modelo k-anonimato. PontodeAcesso, v. 11, n. 1, p. 20-41, 2017. Disponível em: https://periodicos.ufba.br/index.php/revistaici/article/view/13754/14661. Acesso em: 24 jun. 2023.

AGÊNCIA DE BIBLIOTECAS E COLEÇÕES DIGITAIS DA UNIVERSIDADE DE SÃO PAULO. Dados de Pesquisa. 2023. Disponível em: https://www.abcd.usp.br/apoio-pesquisador/dados-pesquisa/. Acesso em 20 mar. 2023.

ALMEIDA, Bethania de Araujo et al. Preservação da privacidade no enfrentamento da COVID-19: dados pessoais e a pandemia global. Ciênc. saúde coletiva, v. 25 n. 1, p. 2487-2492, jun., 2020. Disponível em: https://www.scielosp.org/article/csc/2020.v25suppl1/2487-2492/#. Acesso em 20 mar. 2023.

ALVES, Daniel Versoza. Técnicas de anonimização de dados pessoais e a lei n. 13.709/2018. 2021. 36 f. Trabalho de Conclusão de Curso. (Graduação em Direito). Faculdade de Direito da Universidade Federal do Paraná. Paraná, 2021. Disponível em: https://acervodigital.ufpr.br/handle/1884/71173. Acesso em: 30 jun. 2023.

ARAGÃO, Suéllyn Mattos; SCHIOCCHET, Taysa. Lei Geral de Proteção de Dados: desafio do Sistema Único

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

de Saúde. Reciis – Rev Eletron Comun Inf Inov Saúde. v. 14, n. 3. p. 692-708, jul./set., 2020. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2012/2391. Acesso em 20 mar. 2023.

ARBIX, Glauco; BRANDÃO, Rodrigo; CAMARGO, Laura Simões. Pandemia, Ciência Conectada e IA. In: VAINZOF, Rony et al. Inteligência Artificial: Sociedade Economia e Estado. São Paulo: Thomson Reuters Brasil, 2021, p. 357-374.

BARRETO, Fabíola Gonçalves; HENRIQUE, Fabricio Gustavo. LEI GERAL DE PROTEÇÃO DE DADOS E A APLICABILIDADE NA ANONIMIZAÇÃO. IV Workshop de Tecnologia da Fatec Ribeirão Preto – v. 1, n. 4, p. 1-14, Dez., 2021. Disponível em: http://www.fatecrp.edu.br/WorkTec/edicoes/2021-2/trabalhos/IV-Worktec-LEI_GERAL_DE_PROTEC%CC%A7A%CC%83O_DE_DADOS_E_A_APLICABILIDADE_NA_ANONIMIZAC%CC%A7A%CC%83O.pdf. Acesso em: 30 jun. 2023.

BONITA, R.; BEAGLEHOLE, R.; KJELLSTRÖM, T. Epidemiologia básica. 2 ed. São Paulo: Santos editora, 2010. 203 p.

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 18 mar. 2023.

BRASIL. Lei nº 13.853, de 8 de julho de 2019. Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Lei/L13853.htm#art1. Acesso em 18 mar. 2023.

BRASIL. Ministério da Saúde. Assessoria Especial de Proteção de Dados. Programa de Governança em Privacidade. Brasília: Ministério da Saúde, 2022. 24 p. https://bvsms.saude.gov.br/bvs/publicacoes/programa_governanca_privacidade.pdf
Brasil. Ministério da Saúde. Secretaria Executiva. Plano diretor de tecnologia da informação e comunicação 2019/2021. Disponível em: https://datasus.saude.gov.br/wp-content/uploads/2019/08/PDTIC-2019-A-2021-FINAL-14-DE-AGOSTO-2019.pdf. Acesso em: 20 mar. 2023.

BRASIL. Ministério da Saúde. Catálogo de Produtos – DATASUS. [201?]. Disponível em: https://datasus.saude.gov.br/wp-content/uploads/2019/08/Catalogo-de-Produtos-DATASUS.pdf. Acesso em: 20 mar. 2023.

BORGMAN, Christine L. Big data, little data, no data: scholarship in the networked world.London: The MIT Press, 2015. 416 pp.

CAHN, Aaron et al. An Empirical Study of Web Cookies. In: 25th International Conference on World Wide Web (WWW '16), International World Wide Web Conferences Steering Committee. Republic and Canton of Geneva, Switzerland, April 11–15, 2016, Montréal, Québec, Canada. pp. 891-901. 2016. Disponível em: https://pages.cs.wisc.edu/~pb/www16_final.pdf. Acesso em: 24 mar. 2023.

CARVALHO, Cristiano. Introdução ao R. Universidade Federal de Minas Gerais. Departamento de Estatística. 2020. 19 p. Disponível em: http://www.est.ufmg.br/~cristianocs/Pacotes2021/Intro.html#3. Acesso em: 20 mar. 2023.

COMISSÃO EUROPEIA. Proteção de dados na EU. O Regulamento Geral sobre a Proteção de Dados (RGPD), a Diretiva sobre a Proteção de Dados na Aplicação da Lei e outras regras relativas à proteção de dados pessoais. [2022?]. Disponível em: https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_pt#:~:text=Regulamento%20Geral%20sobre%20a%20Prote%C3%A7%C3%A3o%20de%20Dados%20(RGPD),-Regulamento%20(UE)%202016&text=Este%20regulamento%20constitui%20uma%20medida,p%C3%BAblicos%20no%20mercado%20%C3%BAnico%20digital. Acesso em: 25 mar. 2023.

COSTA, Maria Fernanda Lima; BARRETO, Sandhi Maria. Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento. Epidemiologia e Serviços de Saúde. v. 12, n. 4, p. 189 – 201. 2003. Disponível em: http://scielo.iec.gov.br/pdf/ess/v12n4/v12n4a03.pdf. Acesso em: 30 mar. 2023.

CUNHA, Murilo; CAVALCANTI, Cordélia Robalinho de Oliveira Bastos da. Dicionário de biblioteconomia e

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

arquivologia. Brasília, DF: Briquet de Lemos, 2008. 472 p. Disponível em: https://repositorio.unb.br/handle/10482/34113. Acesso em: 19 mar. 2023.

EUR-LEX. Regulamento (UE) 2018/1725 do Parlamento Europeu e do Conselho, de 23 de outubro de 2018, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais pelas instituições e pelos órgãos e organismos da União e à livre circulação desses dados, e que revoga o Regulamento (CE) n.° 45/2001 e a Decisão n.° 1247/2002/CE. 2018. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?qid=1552642506978&uri=CELEX:32018R1725. Acesso em: 23 mar. 2023.

HORVITZ, Eric; KRISTAN, William. Toward a Computational Microscope for Neurobiology. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.). The Fourth Paradigm: data-Intensive Scientific Discovery. Washington: Microsoft Research, 2009. pp. 83-90. Disponível em: https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf. Acesso em: 25 mar. 2023.

FORNAZIN, Marcelo et al. From medical informatics to digital health: a bibliometric analysis of the research field. In: Americas' Conference on Information Systems, 9-13 ago. 2021. Proceedings. 18. Disponível em: https://aisel.aisnet.org/amcis2021/healthcare_it/sig_health/18. Acesso em: 23 jun. 2023.

LIMA, Ana Paula Canto de et al. Manual do cidadão: privacidade, proteção de dados pessoais. Recife: Editora Império, 2023. 242 p.

LINKEDIN. "Os dados são o novo petróleo", diz o CEO global da Mastercard. 2019. Disponível em: https://www.linkedin.com/pulse/os-dados-s%C3%A3o-o-novo-petr%C3%B3leo-diz-ceo-global-da-mastercard-/?originalSubdomain=pt. Acesso em: 20 mar. 2023.

MAC DEWITT WALLACE LIBRARY. Research Guides. Data Module #1: What is Research Data? 2023. Disponível em:
https://libguides.macalester.edu/c.php?g=527786/&p=3608643#:~:text=Experimental%20data%20are%20collected%20through,projectable%20to%20a%20larger%20population. Acesso em: 02 abr. 2024.

MACHADO, Caio César Vieira; DOURADO, Daniel de Araújo; MACHADO, Flávio Roberto Naval. Bases legais para a pesquisa, desenvolvimento e uso de inteligência artificial na saúde e bem-estar. In: VAINZOF, Rony et al. Inteligência Artificial: Sociedade Economia e Estado. São Paulo: Thomson Reuters Brasil, 2021, p. 501-528.

MAGRINI, Eduardo; GUEDES, Paula. Sistemas de recomendação impulsionados por inteligência artificial: desafios éticos e jurídicos. In: VAINZOF, Rony et al. Inteligência Artificial: Sociedade Economia e Estado. São Paulo: Thomson Reuters Brasil, 2021, p. 103-136.

MENDONÇA, André Luís da Costa. Uma abordagem de privacidade diferencial para dados correlacionados utilizando técnicas de agrupamento. 2018. 94 f. Dissertação (Mestrado em Ciência da Computação). Universidade Federal do Ceará, Ceará, Fortaleza, 2018. Disponível em: https://repositorio.ufc.br/bitstream/riufc/38796/3/2018_dis_alcmendon%C3%A7a.pdf. Acesso em: 24 jun. 2023.

NATIONAL SCIENCE BOARD. Long-lived digital data collections: enabling research and education in the 21st Century. Arlington: National Science Foundation, 2005. 92 pp. Disponível em: http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf. Acesso em: 21 mar. 2023.

O DILEMA DAS REDES. Direção: Jeff Orlowski; Roteiro Jeff Orlowski, Davis Coombe; Elenco: Skyler Gisondo, Kara; Hayward, Vincent Kartheiser. Título original: The Social Dilemma. Estados Unidos. 2020, 1h 29 min.

PRATA, Paula et al. Garantia de Privacidade Versus Utilidade dos Dados em Anonimização: um estudo no ensino superior. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação, v. 1, n. 40, p. 112-127, 2020. Disponível em: https://ubibliorum.ubi.pt/bitstream/10400.6/11463/1/RISTI-ver-cameraReady.pdf. Acesso em: 01 jul. 2023.

RIBEIRO-ALVES, Marcelo; FRANCO, Carolina Mendes. Manual prático de anonimização de dados de pesquisa com o R. Fundação Oswaldo Cruz, Rio de Janeiro, 2022. 89 p. Disponível em: https://www.arca.fiocruz.br/

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023

bitstream/handle/icict/56398/Manual%20Pr%c3%a1tico%20de%20Anonimiza%c3%a7%c3%a3o%20de%20Dados%20de%20Pesquisa%20com%20o%20R.pdf?sequence=2&isAllowed=y. Acesso em: 25 mar. 2023.

SAYÃO, Luis Fernando; SALES, Luana. AFINAL, O QUE É DADO DE PESQUISA? Biblos: Revista do Instituto de Ciências Humanas e da Informação. Rio Grande v. 34, n. 02, p. 32-51, jul./dez. 2020. Disponível em: https://doi.org/10.14295/biblos.v34i2.11875. Acesso em: 19 mar. 2023.

SEMELER, Alexandre Ribas; PINTO, Adilson Luiz. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. Ci.Inf., Brasília, DF, v. 48 n.1, p. 130-129, jan./abr, 2019. Disponível em: https://webcache.googleusercontent.com/search?q=cache:1VZtYtIf2TYJ:https://revista.ibict.br/ciinf/article/download/4461/4102/14051&cd=2&hl=pt-BR&ct=clnk&gl=br. Acesso em: 22 mar. 2023.

SILVA, Bethânia de A. Almeida SILVA. ÉTICA EM PESQUISA E DADOS SENSÍVEIS: O QUE MUDA? In: SILVA, Angélica Baptista; CUNHA, Francisco José Aragão Pedroza. Lei Geral de Proteção de Dados e o controle social da saúde. Alegre, RS: Editora Rede Unida. 2023, p. 280-291. E-book. (Série Participação Social e Políticas Públicas, v. 13).

SILVA. Daniel Fernando Alves da. Geração Sintética de Microdados utilizando algoritmos de data mining. 2015. 125 f. Dissertação (Mestrado em Economia) Universidade do Porto, Porto, 2015. Disponível em: https://repositorio-aberto.up.pt/bitstream/10216/80015/2/36274.pdf. Acesso em: 25 mar. 2023.

SILVA, Sérgio Conde de Albite. Projeto de Trabalho de Conclusão de Curso – TCC: Diretrizes para a sua elaboração no escopo da disciplina Metodologia da Pesquisa. Rio de Janeiro: Programa de Pós-Graduação em Gestão de Documentos e Arquivos, UNIRIO, 2018, 40 p.

TEMPL, Matthias et al. Introduction to Statistical Disclosure Control (SDC). IHSN Working Paper n. 007. August, 2014. 25 pp. Disponível em: https://www.ihsn.org/sites/default/files/resources/ihsn-working-paper-007-Oct27.pdf. Acesso em: 25 mar. 2023.

TEMPL, Matthias; KOWARIK, Alexander; MEINDL, Bernhard. sdcMicro: Statistical Disclosure Control Methods for Anonymization of Microdata and Risk Estimation. 2018. Disponível em: http://cran.nexr.com/web/packages/sdcMicro/index.html. Acesso em: 25 set. 2023.

TEMPL, Matthias; KOWARIK, Alexander; MEINDL, Bernhard. Statistical Disclosure Control for Micro-Data Using the R Package sdcMicro. Journal of Statistical Software, v. 67, n 4, pp. 1–36. 2015. Disponível em: https://www.jstatsoft.org/article/view/v067i04. Acesso em: 26 mar. 2023.

VAINZOF, Rony et al. Inteligência Artificial: Sociedade Economia e Estado. São Paulo: Thomson Reuters Brasil, 2021. 696 p.

VILLAS-BÔAS, Maria Elisa. O direito-dever de sigilo na proteção ao paciente. Rev. Bioét. v. 23, n. 3, p. 513-523, 2015. Disponível em: https://bit.ly/2nLeWUP. Acesso em: 25 mar. 2023.

Submetido: 12/07/2023
Aceito: 25/09/2023
Publicado: 31/12/2023