

Pesquisas em Análise de Redes Sociais e LGPD, análises e recomendações

Luiz Paulo Carvalho ¹, Jonice Oliveira ¹, Claudia Cappelli ¹

¹ Programa de Pós-Graduação em Informática – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

luiz.paulo.carvalho@ppgi.ufrj.br, jonice@dcc.ufrj.br,
claudia.cappelli@gmail.com

Abstract. *Data protection and data-driven solutions are two progressing areas permeating Brazilian society. This work presents an interdisciplinary theoretical approach related to the LGPD, from the Law studies perspective and the Social Network Analysis in Brazil, from the Informatics perspective. This research area utilizes personal data extensively for knowledge construction, with semantic contributions, analyzing the reality; or pragmatic, building artifacts. Challenges and inseparable issues are observed, exposed and debated in this work. We suggest good practices combining the two topics, personal data in the research field of social networks in Brazil respecting the LGPD.*

Resumo. *A proteção de dados e as soluções orientadas a dados são duas áreas em progresso que permeiam a sociedade brasileira. Este trabalho apresenta uma abordagem teórica interdisciplinar relacionada à LGPD, na perspectiva dos estudos de Direito e na Análise de Redes Sociais no Brasil, na perspectiva da Informática. Esta área de pesquisa utiliza dados pessoais extensivamente para construção do conhecimento, com contribuições semânticas, analisando a realidade; pragmáticos, na construção de artefatos. Desafios e questões inseparáveis são observados, expostos e debatidos neste trabalho. Sugerimos também boas práticas combinando os dois tópicos, dados pessoais no campo da pesquisa em redes sociais no Brasil, respeitando a LGPD.*

1. Introdução

“O recurso mais valioso do mundo não é mais o petróleo, são dados” ¹ [THE ECONOMIST, 2017], acompanha célebre expressão associada à Clive Humby, de 2006, “Dados são o novo petróleo” ². Prenuncia as influências do tema em diversas esferas, setores e âmbitos, como acadêmico e legal. Acompanha o aumento da capacidade computacional, como armazenamento e processamento de dados, permitindo procedimentos antes impossíveis ou altamente custosos, como cálculos multidimensionais a partir de atributos extraídos de Redes Sociais Online (RSO) [SUMPTER, 2018]. O caso *Cambridge Analytica* [ISAAK E HANNA, 2018], que tratou ilegalmente dados de usuários da RSO *Facebook*, expõe um uso antiético de dados.

A importância percebida dos dados influenciou diversos setores da sociedade, na área acadêmica podemos citar a Interação Humano-Dados [MORTIER *et al.*, 2014], Ecossistemas de Dados [OLIVEIRA E LÓCIO, 2018], Sociedade Orientada a Dados [PENTLAND, 2013], Economia de Dados [BÖRNER *et al.*, 2018], Ciência de Dados

¹ “The world's most valuable resource is no longer oil, but data”.

² “Data is the new oil”.

(*Data Science*) e Inteligência do Negócio (*Business Intelligence*) [LARSON E CHANG, 2016]], dentre outros. Com esta ascensão debruçada em alcançáveis tecnologias computacionais de alto desempenho e distribuição, questões éticas emergiram [MOOR, 2005] por má fé na manipulação de dados e seus impactos na sociedade leiga em relação a como seus próprios dados são utilizados. Governos reforçaram suas legislações sobre dados, trazendo os mesmos à luz e um novo patamar de importância.

A Regulamentação Geral de Proteção de Dados (*General Data Protection Regulation* – GDPR) [UE, 2018] está em vigor em todo alcance do bloco da União Europeia (UE) desde 2018. A GDPR possui caráter transnacional, influenciando países da UE e parceiros, como o Brasil. Determina que apenas países com uma legislação tão, ou mais, rígida quanto podem tratar dados pessoais de cidadãos da UE, mesmo aqueles com dupla cidadania. Isto é, RSO brasileiras, ou demais sistemas computacionais, só poderiam tratar dados de usuários europeus caso suas especificações e requisitos fossem compatíveis com a GDPR. Inclusas pesquisas na área de Análise de Redes Sociais (ARS) ocorridas no Brasil, mas com cidadãos da UE.

Em vista de adequar-se às diretivas da GDPR e construir uma soberania legal sobre os dados de seus cidadãos [BIONI, 2019], foi aprovada a Lei Geral de Proteção de Dados [BRASIL, 2018] no Brasil. Sua previsão de entrada em vigor é em 14/08/2020 e influenciará todo o tratamento de dados de pessoas naturais brasileiras, em território brasileiro ou exterior, por todos seus países parceiros, similar à diretiva europeia. Observa-se a relevância do tema a partir de uma Proposta de Emenda Constitucional (PEC) com a intenção de acrescentar a proteção de dados à Constituição Federal, fixando a competência privativa da União para legislar sobre a matéria [BRASIL, 2019].

Resumido o cenário de dados e da LGPD, a área de ARS é/será afetada. Pesquisas envolvendo pessoas naturais no domínio de ARS, contidas em RSO ou não, estarão sob os auspícios da lei, e sobre estas que trataremos. Redes Sociais compostas, por exemplo, de animais, empresas ou elementos abstratos não estão suscetíveis à LGPD.

De acordo com o relatório *Digital 2020 Reports*³, brasileiros: (i) passam em média 3h34m online em redes sociais por dia; (ii) são o segundo povo que mais utiliza Internet diariamente no mundo, 9h17m; (iii) terceiro em uso de Internet móvel, 4h41m; (iv) 79% estão preocupados com a forma com a qual empresas usam seus **dados pessoais**⁴; (v) 85% preocupam-se sobre o que é falso ou real na Internet; (vi) as redes sociais registraram um aumento de 8.2% no número de usuários, alcançando mais de cem milhões de usuários; (vii) cento e vinte milhões são alcançáveis por publicidade no *Facebook*. Sendo assim, observamos a influência da Internet e das RSO na sociedade brasileira, refletidas na importância e intenção da realização de pesquisas acadêmicas utilizando estes meios. O texto da LGPD não é simples ou de fácil assimilação, não foi redigido direcionado a pesquisadores. Aqui contextualizaremos a lei com práticas comuns em ARS.

A abordagem metodológica é teórico-analítica, detalhando ponto a ponto a LGPD, relacionando com o Estado da Arte que são realizadas as práticas em ARS [CAN *et al.*, 2014] [YANG *et al.*, 2017]. Pesquisas fora da conformidade com os ditames legais correm o risco de desindexação, exclusão ou, em casos extremos, processos administrativos envolvendo o pesquisador ou a instituição de pesquisa. Em primeiro momento já é esperado que haja determinado despreparo por parte das organizações [BIONI, 2019],

³ <https://datareportal.com/reports/digital-2020-brazil>. Disponível em 31/03/2020

⁴ “informação relacionada a pessoa natural identificada ou identificável;” [BRASIL, 2018]

inclusive de pesquisa, e um período de amadurecimento e transição rumo à conformidade legal, nossa proposta é preparar e antecipar.

O trabalho é estruturado da seguinte forma: Seção 2 resume fundamentos teóricos essenciais; Seção 3 expõe as contribuições; Seção 4 discute o tema e conclui o trabalho.

2. Breve fundamentação teórica

Não analisaremos profundamente a LGPD, apenas os aspectos gerais e, na Seção 3, sua associação com a pesquisa em ARS. A lei tem objetivo proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural [BRASIL, 2018], que tenha seus dados em RSO, sendo usuária dela ou não, *e.g.*, fotos publicadas expondo usuários não cadastrados na RSO e, ainda assim, tratadas.

No art. 2º da LGPD são listados os sete fundamentos da disciplina de proteção de dados, como respeito à privacidade e autodeterminação informacional. No art. 6º são listados os princípios que norteiam o tratamento de dados, precedidos, acima de tudo como determina no *caput*, pela boa-fé. A Tabela 1 apresenta os princípios, amplamente abordados neste trabalho, tangentes à ARS.

Tabela 1: Princípios da LGPD para tratamento de dados [BRASIL, 2018]

#	Princípio e definição
I	Finalidade: Realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;
II	Adequação: Compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;
III	Necessidade: Limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;
IV	Livre acesso: Garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;
V	Qualidade dos dados: Garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;
VI	Transparência: Garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;
VII	Segurança: Utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;
VIII	Prevenção: Adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;
IX	Não discriminação: Impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;
X	Responsabilização e prestação de contas: Demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

O termo **tratamento de dados**⁵ é definido em lista exemplificativa, não-exaustiva ou limitante de ações, diretas ou indiretas, envolvendo dados. A organização onde é realizada a pesquisa exerce o papel de **controlador**⁶, enquanto o pesquisador tratando os dados pode exercer ou não o papel de **operador**⁷ de dados. São **agentes de tratamento**⁸ da organização, seja pública, privada, terceiro setor, dentre outras. **Controlador** e o **operador** podem ser atores externos, terceirizados, à organização.

Estudos expõe a relação da GDPR com a epistemologia científica europeia [CHASSANG, 2017]. Apesar das similaridades entre as legislações, elas não são idênticas e possuem construtos textuais diferentes para conceitos semelhantes. Por exemplo, a GDPR não aponta explicitamente o princípio de não-discriminação e sua definição para **tratamento de dados e titular**⁹ dos dados é diferente da LGPD.

Sobre as diferenças entre a LGPD e a GDPR apontamos também a maturidade do tempo em vigor, a LGPD não está em vigor e a GDPR vigora há, mais ou menos, dois anos. A GDPR possui elementos dedicados para tratar, com detalhes, de pesquisa acadêmica e tratamento de dados envolvendo esta prática, art. 9, art. 89, encaminhando 159 [UE, 2018]. Conceitos e construtos da LGPD foram importados da GDPR, iguais ou parecidos, e apesar das diferenças a tendência é que as resoluções para casos brasileiros envolvendo este tópico acompanhem a interpretação europeia [MORIBE *et al.*, 2019].

Há ampla literatura relacionando Ética, Pesquisa e Ciência [REIJERS *et al.*, 2017], com pontos incorporados neste trabalho. Entretanto, não há ainda um trabalho que associe a LGPD com ARS, considerando o contexto e a realidade brasileira. Privacidade, por sua vez, não é um assunto inédito em ARS [ZHELEVA E GETOOR, 2011], apesar disso não havia até a LGPD uma lei geral tão específica e abrangente direcionada especificamente a dados pessoais ou **dados pessoais sensíveis**¹⁰ no Brasil. Em relação à Pesquisa, a Associação Brasileira de Empresas de Pesquisa (ABEP) possui um guia simples e objetivo que facilita a associação da LGPD à Pesquisa em amplo espectro [ABEP, 2017], onde carece um detalhamento criterioso sobre ARS.

Uma das práticas essenciais em ARS envolve a coleta de dados em RSO. Neste âmbito, a coleta pode ocorrer envolvendo poucos titulares, por questionários, ou através de extração automatizada envolvendo muitos titulares, por *Application Programming Interfaces* (API) dedicadas. A RSO *Twitter*¹¹ é amplamente utilizada para pesquisas em ARS através de sua API, como podemos observar nos anais da *International Conference on Advances in Social Networks Analysis and Mining* (ASONAM)¹².

Resumidamente apontadas as bases teóricas que norteiam este trabalho, prosseguimos à análise teórica interdisciplinar.

⁵ “toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;” [BRASIL, 2018]

⁶ “pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;” [BRASIL, 2018]

⁷ “pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;” [BRASIL, 2018]

⁸ “o controlador e o operador;” [BRASIL, 2018]

⁹ “pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;” [BRASIL, 2018]

¹⁰ “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]

¹¹ <https://twitter.com/>. Acessado em 30/03/2020.

¹² <https://dblp.org/db/conf/asunam/index>. Acessado em 30/03/2020

3. LGPD e ARS, coexistindo em parcimônia

Nesta Seção detalhamos e analisamos os pontos mais relevantes da LGPD associados com ARS e, caso caibam, propostas recomendações às pesquisas no novo cenário com a lei que está prevista para vigorar a partir de 14/08/2020.

3.1. Coleta de dados

Informações inseridas por usuários em RSO não são domínio público ou dados públicos ou abertos, cada plataforma define a apropriação e direitos nos seus termos de uso, respeitando (como esperado) as legislações específicas onde operam [DONAHUE, 2016].

Bond *et al.* (2013) realizam uma pesquisa com usuários de fóruns relacionada com o que produzem nos ambientes virtuais, sobre a percepção da sua produção ser domínio público ou não. Não há um consenso, os usuários estão divididos entre aqueles que se contentam que sua produção pode ser instrumentalizada por qualquer um como bem entender, enquanto outros preferem deter a autoria e controle do que produzem.

No caso da LGPD, acompanhando o desenrolar da efetividade da GDPR, o titular dos dados, caso perceba que um objeto comunicacional fere sua privacidade, ou qualquer outro item do art. 2, pode requerer a tomada de medidas de acordo. Dois mecanismos da LGPD protegem a pesquisa e o pesquisador, o **consentimento**¹³ e o legítimo interesse.

Ao se tratar de pesquisas envolvendo uma quantidade pequena ou mediana de participantes, o pesquisador instrumentaliza um termo de consentimento livre e esclarecido. Esta condução é mais recorrente em pesquisas qualitativas [RECKER, 2013], que podem envolver dados pessoais sensíveis. Dados pessoais sensíveis tem uma seção dedicada a eles, Seção II, e o art. 13 aprofunda o uso destes dados em estudos de saúde pública, que podem envolver ou não ARS. O termo deve seguir todos os princípios da Tabela 1, expondo a finalidade, tempo de tratamento, local de armazenamento, dentre outras informações devem ser transparentes. Em caso de alteração relevante de algum dos elementos do termo um novo consentimento deve ser obtido, como a transmissão dos dados pessoais para outro grupo de pesquisa.

O legítimo interesse [COSTA *et al.*, 2019], art. 10, é uma saída para o tratamento de uma enorme quantidade de dados, como ocorre frequentemente em ARS. “O legítimo interesse do controlador somente poderá fundamentar tratamento de dados pessoais para finalidades legítimas, consideradas a partir de situações concretas, que incluem, mas não se limitam a: I - apoio e promoção de atividades do controlador;” [BRASIL, 2018]. Kotsios *et al.* (2019) realizaram uma pesquisa extraindo dados no Twitter, sem sucesso na obtenção de consentimento de cada um dos usuários titulares dos dados que coletaram. Mesmo que utilizando uma abordagem automatizada, o Twitter freou a emissão de mensagens diretas requerendo consentimento, alegando a prática de *spam*. Esse caso demonstra a impossibilidade atual de obter o consentimento generalizado em pesquisas com volume descomunal de dados provenientes de RSO.

Importante reiterar, caso não haja consentimento livre e esclarecido e caso o objeto informacional da pesquisa atente contra os objetivos e princípios legais, as partes envolvidas estão passíveis às sanções administrativas previstas em lei, controlador e operador, e haverá intercessão sobre a comunicação científica, caso a caso. Este mesmo

¹³ “manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma finalidade determinada;”. [BRASIL, 2018]

encaminhamento se aplica a pesquisas realizadas pelas pessoas naturais autônomas com fins exclusivamente particulares e não econômicos, detalhada na Seção 3.3, onde há infração à LGPD e ampla divulgação pública contendo dados pessoais, sensíveis ou não.

3.2. A lei protege dados e, conseqüentemente, o titular, não o inverso

Mesmo que haja consentimento livre e esclarecido e a comunicação científica faça uso de dados pessoais ou dados pessoais sensíveis potencialmente danosos ao titular dos dados ocorre uma potencial infração. Uma das essências epistemológicas essenciais da LGPD está na reinterpretação do paradigma de “proteção”, onde a legislação protege os dados, individuais ou coletivos, visando alcançar seus objetivos [BIONI, 2019]. Essa interpretação é premente quando se tratam dados pessoais sensíveis, que podem causar uma sorte de danos sociais ao titular e atenta contra o princípio da não discriminação.

Um cidadão leigo ou ingênuo em relação à possível magnitude que o tratamento dos seus próprios dados pode trazer a si por soluções computacionais complexas ou obscuras não tem a habilidade de discernir o dano que lhe será causado, *i.e.*, os impactos ou influências efetivos daquele tratamento de dados a partir do momento que consentiu com o tratamento dos mesmos. Este raciocínio é validado pelo art. 42, II, § 2º, possibilitando inversão do ônus da prova a favor do titular dos dados quando, a seu juízo, for verossímil a alegação, houver hipossuficiência para fins de produção de prova [BRASIL, 2018], e pelos próprios princípios do art. 6, na Tabela 1.

Desinformação, como *fake news*, e informações potencialmente polêmicas utilizando dados pessoais, sensíveis ou não, são disseminadas e espalhadas de forma rápida e descontrolada em RSO [AVELAR, 2019], comportamento intensificado em usuários da Internet não nativos digitais ou menos letrados digitalmente [GUESS *et al.*, 2019]. Um exemplo são as desinformações veiculadas e compartilhadas através da RSO *YouTube* em relação ao COVID-19 e sua pandemia [DIAS E RIBEIRO, 2020]. Dados de pesquisas publicadas podem ser legalmente reinterpretadas e apropriadas por veículos de comunicação e divulgadas alcançando proporções inesperadas, colocando em risco os envolvidos, consentindo ou não.

3.3. Uso para fins acadêmicos, científicos ou estudos

Em três momentos a lei toca nestes assuntos: (i) **Fins acadêmicos (art. 4, II, b)**. Esta Lei não se aplica ao tratamento de dados pessoais para fins exclusivamente acadêmicos; [BRASIL, 2018]; (ii) **Realização de estudos (art. 7, IV e art. 11, II, c)**. O tratamento de dados pessoais somente poderá ser realizado para a realização de estudos por **órgão de pesquisa**¹⁴, garantida, sempre que possível, a anonimização dos dados pessoais. O art. 7 trata de dados pessoais e o art. 11 trata de dados pessoais sensíveis; (iii) **Órgão de pesquisa (art. 5, XVIII, art. 7, IV e art. 11, II, c)**. Pesquisa, seja fim acadêmico ou fim de estudo, só está abarcada nestes domínios na LGPD. Somado a estes três pontos, o art. 4, I: “Esta Lei não se aplica ao tratamento de dados pessoais realizado por pessoa natural para fins exclusivamente particulares e não econômicos;” [BRASIL, 2018].

Belmudes (2020) apresenta uma análise da relação entre LGPD e redes sociais no

¹⁴ “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]

geral, inclusive no âmbito da comunicação científica. A Corte de Justiça Europeia reforçou e ratificou sua posição em relação ao uso de dados pessoais, sensíveis ou não, provenientes de redes sociais, “mesmo que um dado pessoal seja utilizado de forma particular e não econômica, se seu processamento der visibilidade a um número indefinido de pessoas, a exceção não se aplica” [BELMUDES, 2020]. Considerando que um órgão de pesquisa como o conceituado no art. 5, XVIII, tem fins não-lucrativos, ao que concerne aspecto econômico, uma comunicação científica disponibilizada e divulgada à sociedade em geral está sob determinação da LGPD.

Sendo assim, caso um pesquisador construa e disponibilize amplamente uma comunicação científica tácita que fira os objetivos da LGPD, art. 2, será observada a discordância com os princípios, na Tabela 1. O titular dos dados, neste caso, poderá requerer a eliminação dos dados pessoais, art. 18, que o identifiquem ou levem a identifica-lo, podendo chegar à exclusão ou desindexação do objeto como um todo.

Por exemplo, o art. 4º pode levar à interpretação de que, como a lei não incide sobre pesquisas acadêmicas, os dados pessoais coletados, sensíveis ou não, podem ser transferidos entre partes livremente, e estão isentos de requisitos não-funcionais de segurança ou libera a exposição dos titulares. Esse entendimento é absurdo, expondo que mesmo fins exclusivamente acadêmicos devem guiar-se pela LGPD.

3.4. Dados pessoais sensíveis, referentes à saúde ou não

A LGPD é rígida e objetiva em relação à segurança das informações envolvendo dados pessoais sensíveis e, enfaticamente, dados pessoais sensíveis referentes à saúde, art. 11, II, g, § 4º e § 5º; art. 13. Um exemplo no caso de dados relacionados à saúde, são estudos de Epidemiologia e ARS [STATTNER E VIDOT, 2011]. O consentimento para o uso desta categoria de dados deve ser especificado de forma específica e detalhada, não obscurecido de forma generalizada.

Em trabalhos envolvendo esta categoria de dados o cuidado e precaução devem ser requisitos primários da pesquisa, de forma a não possibilitar o rastreamento reverso dos dados aos seus titulares, identificando-os, oferecendo risco a eles e à pesquisa.

Na Rússia, um estudante foi expulso da instituição de ensino por ter seus dados pessoais das suas RSO analisados, categorizado como “gay” pelos dirigentes [THE MOSCOW TIMES, 2019]. Configura perseguição moral e uso inescrupuloso dos dados.

3.5. Anonimização, pseudonimização e identificação direta ou indireta

A LGPD lida com **dado anonimizado**¹⁵, **anonimização**¹⁶ e **pseudonimização**¹⁷. Dados pessoais que não tenham sido processados a partir de uma abordagem de anonimização ou pseudonimização associam o titular aos dados, identificando-o, direta ou indiretamente. Associação direta quando o dado identifica o titular apenas com o que é apresentado, como Registro Geral (RG); indiretas quando há necessidade de esforço para a associação. A identificação indireta não é trivial e exige procedimentos específicos, por

¹⁵ “dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;” [BRASIL, 2018]

¹⁶ “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]

¹⁷ “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]. A definição de pseudonimização não está no art. 5, mas destacado no art. 13, § 4º.

exemplo, inferir a orientação sexual de usuários em uma RSO baseado em suas interações e produção [JERNIGAN E MISTREE, 2009].

Pseudonimização é citada unicamente no art. 13, especificamente para estudos em saúde pública, e respectiva comunicação. Em seguida há a recomendação de que estes estudos sigam os padrões éticos relacionados com estudos e pesquisas, sem especificar quais. Conceitos relacionados à Ética só tem uma ocorrência no texto da lei, neste ponto. Ademais, cabe a interpretação de “boa-fé” no *caput* do art. 6.

O art. 12 determina que dados anonimizados não são dados pessoais para os fins da LGPD, salvo em circunstância onde a anonimização for revertida ou seja deliberadamente identificado pelas partes, tornando-o novamente um dado pessoal para o fim da lei. Sendo assim, a recomendação é que a anonimização seja regra e a associação, direta ou indireta, a exceção. Anonimização razoável garante a não incidência da LGPD.

A LGPD incorpora o conceito de “meios técnicos razoáveis e disponíveis”, tendo em vista a impossibilidade de construir uma solução infalível ou inquebrantável e que a anonimização é uma atividade complexa. Uma Faculdade de pequeno porte que realize pesquisas em ARS não terá os mesmos recursos disponíveis para anonimização de suas bases de dados comparada com uma Universidade de grande porte.

A anonimização também é uma saída recomendada para quando o prazo de tratamento dos dados consentido pelo titular for alcançado. Caso o próprio titular não possa ser identificado ou identificável, não há infração à lei. A LGPD determina que após a realização de todas as finalidades envolvendo o tratamento, neste caso a pesquisa ou estudo em questão, os dados sejam excluídos. A anonimização é uma saída para que os dados não sejam excluídos, já que não são mais considerados “dados pessoais”.

Pesquisas qualitativas e que dependem da identificação de pessoas naturais devem assumir um cuidado redobrado neste aspecto, já que o (ou um dos) objeto da pesquisa é um titular de dados e seus dados pessoais são protegidos pela LGPD. Uma possível alternativa é divulgar o trabalho aos titulares de dados antes de submetê-la à externalização de fato, demonstrando que ele detém a posse de seus dados pessoais.

3.6. Minimização de dados, perfilamento e especificidade de termos

O princípio da necessidade, na Tabela 1, operacionaliza o conceito de minimização de dados, determinando claramente que apenas os dados imprescindíveis e absolutamente necessários deverão ser tratados para dado fim. Esse ponto é complexo em relação à extração de conteúdo de redes sociais, onde centenas de milhares de dados podem ser tratados, não necessariamente relacionados com o objetivo final ou protocolo da pesquisa. Sendo assim, uma opção é adequar a base de dados, excluindo os registros e dados não associados com a pesquisa e, então, anonimizá-los, caso a essência da pesquisa permita.

Sobre pesquisas qualitativas ou com coleta de dados direta com participantes, a recomendação é que seja requerido estritamente o mínimo necessário de dados à sua realização. Este princípio cerceia o famoso costume de “coletar o máximo, para depois aproveitar o que for possível”, este raciocínio também cria uma fragilidade de dados em uma possível brecha de segurança, passível de penalização na ocorrência de uma fiscalização relacionada ao comprometimento dos dados. Uma recomendação é que uma tabela de correspondência seja elaborada, associando dados com sua finalidade concreta.

Uma das preocupações da LGPD, como da GDPR [KOTSIOS *et al.*, 2019], é a

construção de perfis determinísticos relacionados com dados pessoais, principalmente sensíveis. Esta prática é conhecida como perfilamento (*profiling*). A pesquisa constrói “perfis” ou “modelos”, onde categoriza envolvidos. Esta ainda é uma questão em aberto, não abordada na lei, porque a pesquisa pode construir perfis que venham a trazer consequências negativas aos próprios titulares. Quais são os limites destas pesquisas? Padrões éticos? Possivelmente prejudicar uma ou algumas pessoas é motivo necessário e suficiente para frear o processo da pesquisa? Quais são os limites deste perfilamento?

Especificidade de termos será dividida em dois pontos, vício de consentimento e generalização de termos. Vício de consentimento ocorre “se a pessoa, ou seja, o declarante tivesse real conhecimento da situação, não teria manifestado sua vontade da forma a qual foi declarada.” [LOBO, 2015], se o titular dos dados soubesse que a orientação sexual dele seria utilizada em uma pesquisa de ARS, ele publicaria este dado no *Facebook*? Se um funcionário de uma empresa soubesse que o teste de personalidade que ele realizou na empresa, através de uma organização privada de pesquisa organizacional, poderia prejudicar sua integridade empregatícia, ele poderia negar sua participação? Sem represália? Se um usuário do *Twitter* soubesse que suas publicações são utilizadas para pesquisas de ARS de análise de comportamental, ele ainda publicaria as mesmas coisas? Ou, simplesmente, ainda publicaria? Os termos de consentimento e as justificativas de legítimo interesse devem ser compreensíveis, claros e simplificados.

A generalização de termos se refere ao art. 8, § 4º, “O consentimento deverá referir-se a finalidades determinadas, e as autorizações genéricas para o tratamento de dados pessoais serão nulas.”. O tratamento de dados não pode ser baseado em finalidade genéricas, eternas ou que levem o titular à enganação. Neste sentido é um desafio expor os termos e requisitos da pesquisa de forma transparente, informativa e que respeite os requisitos não-funcionais de entendimento [CARVALHO *et al.*, 2019]. O pesquisador não pode condensar todo o consentimento à “pesquisa” ou “estudo”. Inclusive, a condução de uma pesquisa científica deve atender a rigor metodológico e protocolos específicos [RECKER, 2013], de forma que o pesquisador deve estar consciente dos dados envolvidos, das finalidades, prazo de tratamento e ambiente de armazenamento.

4. Discussão e conclusão

Neste trabalho relacionamos a epistemologia científica de ARS com a LGPD, no contexto brasileiro onde incide a legislação. A relevância desta contribuição se dá na conscientização do uso de dados pessoais, sensíveis ou não, pelos pesquisadores de ARS interessados nesta categoria de insumos no desenvolvimento de suas pesquisas com legislações regulamentando as atividades de tratamento de dados, que influenciam as respectivas pesquisas. A não adequação à lei acarreta a possibilidade potencial de prejudicá-la, sob responsabilidade do pesquisador e da instituição em questão, que é a controladora de dados e a parte anuente do tratamento, legal ou ilegal, dos dados pessoais.

Apesar do aparente rigor que parece incidir sobre o tratamento de dados exposto por este trabalho, este tópico encontra-se imaturo e incipiente no Brasil, denotando também o caráter inovativo desta contribuição. Muitas questões ainda não foram abordadas na lei, estão em aberto, ou abrem margem para interpretações oportunistas ou obscuras, sobretudo no âmbito acadêmico, dos estudos e das pesquisas. Nesse sentido, a recomendação não é negligenciar o cenário e posicionar-se fingindo não entender o que está acontecendo para tirar vantagem ou para garantir o seu bem próprio; sim de assumir as rédeas desta agenda acadêmica e construir soluções que, de fato, norteiem o uso de

dados pessoais em ARS no Brasil, com ética e maturidade.

A GDPR apresenta artigos e encaminhamentos associados com fins acadêmicos, passíveis de importação para o nosso cenário legal. Por exemplo, no art. 17 [UE, 2018] há uma exceção relacionada ao Direito ao Esquecimento [BIONI, 2019] e exclusão de dados pessoais instrumentalizados legalmente em pesquisas, *i.e.*, no caso de tratamento legal de dados pessoais, o titular não pode requerer sua exclusão, já que isso pode descaracterizar ou arruinar a pesquisa em questão.

Alguns pontos são questões em aberto e serão desafios acadêmicos, em ARS podemos listar alguns: (a) a interpretação do art. 7º, § 4º: “É dispensada a exigência do consentimento previsto no caput deste artigo para os dados tornados manifestamente públicos pelo titular, resguardados os direitos do titular e os princípios previstos nesta Lei.” [BRASIL, 2018], sendo assim informações configuradas como “públicas” ou sem nenhuma limitação de alcance em RSO são dados tornados manifestadamente públicos pelo respectivo titular? (b) a LGPD retroagirá sobre dados pessoais coletados e armazenados antes de sua entrada em vigor? Na UE os termos para coleta de dados, como consentimento, anteriores à vigência da GDPR só poderiam ser considerados válidos se estivessem em conformidade com a GDPR, caso contrário era mandatório elaborar e disponibilizar novos termos, e obter novo consentimento de **todos** os titulares [UE, 2018]; (c) como será diferenciado o tratamento de dados para pesquisas ou estudos em, ou para, organizações públicas ou privadas? (d) como instruir os respectivos titulares de seus dados a agir caso identifiquem uso ilegal de seus dados pessoais em uma comunicação acadêmica ou estudo? (e) qual o limites do legítimo interesse quando se trata da extração de enormes volumes de dados, inclusive pessoais, de RSO? Por exemplo, o esforço para obter consentimento é desproporcional e impraticável no *Twitter* [KOTSIOS, *et al.*, 2019]; (f) quais os limites da boa-fé em pesquisas que utilizam dados pessoais em ARS?

Muitas questões em aberto só podem ser respondidas pela Autoridade Nacional de Proteção de Dados (ANPD). A ANPD ainda não foi formada, mesmo faltando menos de seis meses para lei vigorar. Casos omissos à redação da LGPD ou conflitos de interpretação serão decididos pela ANPD, como os que envolvem fins acadêmicos. A recomendação de tomar as rédeas do discurso, de forma ética e responsável, é, adicionalmente, para criar conteúdo basilar que a ANPD, e demais interessados, utilizem como guia para suas tomadas de decisão e formulação de conteúdo.

A construção de uma pesquisa ou estudo epistemologicamente baseado em requisitos não-funcionais éticos é uma área de pesquisa específica [MANZLOOR, 2016], não trivial. Considerando a complexidade, ou até impossibilidade, de alcançar um nível de segurança de dados onde, por exemplo, a anonimização seja irreversível, a LGPD utiliza a razoabilidade de esforços e tecnologias para aplicação dos seus termos. A condução legal, ética e de boa fé está na aplicação concreta dos melhores meios de anonimização disponíveis, ao contrário de negligenciá-los sob o pensamento de “isso vai ser muito fácil de reverter, deixa para lá”.

Como trabalhos futuros apontamos o encaminhamento das questões aqui presentes; elaboração de mais trabalhos interdisciplinares contendo percepções e teorizações entre as áreas de ARS e Direito, dialogicamente; construção de artefatos que auxiliem pesquisadores na conformidade com a LGPD em suas pesquisas, como guias ou sistemas de recomendação; análise aprofundada do tratamento de dados pessoais de crianças e adolescentes, sendo um tópico específico e destacado até na legislação.

Referências ¹⁸

- ABEP (2017) "GUIA DE PROTEÇÃO DE DADOS – LGPD para Profissionais de Pesquisa de Mercado, Opinião e Mídia". Disponível em: cutt.ly/FtmXfA6
- Avelar, D. (2019) "WhatsApp fake news during Brazil election ‘favoured Bolsonaro’". The Guardian. Disponível em: cutt.ly/DtmXD5D
- Belmudes, G. (2020) "Aplicação da LGPD nas redes sociais". JOTA Info. Disponível em: cutt.ly/EtmXJie
- Bioni, B. (2019) "Proteção de Dados Pessoais - A Função e os Limites do Consentimento". ed. 2, Forense. ISBN: 978-8530988623.
- Bond, C. S., Ahmed, O. H., Hind, M., Thomas, B., Hewitt-Taylor, J. (2013) "The Conceptual and Practical Ethical Dilemmas of Using Health Discussion Board Posts as Research Data". JMIR. DOI: 10.2196/jmir.2435
- Brasil (2018) “LEI Nº 13.709, DE 14 DE AGOSTO DE 2018”. Lei Geral de Proteção de Dados. Disponível em: <http://bit.ly/2YgUqMZ>
- Brasil (2019) “Proposta de Emenda à Constituição nº 17/2019”. Disponível em: cutt.ly/utmXL3q
- Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., Wu, L., A. Evans, J. A. (2018) "Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy". Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.1804247115
- Can, F., Özyer, T., Polat, F. (2014) "State of the Art Applications of Social Network Analysis". Springer International Publishing. DOI: 10.1007/978-3-319-05912-9.
- Carvalho, L. P., Oliveira, J., Cappelli, C., Majer, V. (2019) "Desafios de Transparência pela Lei Geral de Proteção de Dados Pessoais". VII WTrans. DOI: 10.5753/wtrans.2019
- Chassang, G. (2017) "The impact of the EU general data protection regulation on scientific research". eCancer. DOI: doi.org/10.3332/ecancer.2017.709
- Costa, D., Oliveira, D., Bousso, F., Caparelli, M., Campos, R., Lerner, V. (2019) "O interesse legítimo como justificativa para o tratamento de dados pessoais". Disponível em: cutt.ly/ZtmXCU0
- Dias, T., Ribeiro, P. V. (2020) “Youtube faz canais de direita ganharem dinheiro espalhando mentiras sobre coronavírus”. Disponível em: cutt.ly/WtmMG8a
- Donahue, L. (2016) "Who Owns the Content Posted on Social Media?". Disponível em: cutt.ly/CtmXBCR
- Guess, A., Nagler, J., Tucker, J. (2019) "Less than you think: Prevalence and predictors of fake news dissemination on Facebook". Science Advances. DOI: 10.1126/sciadv.aau4586
- Isaak, J., Hanna, M. J. (2018) "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection,". Computer, vol. 51, no. 8, pp. 56-59.
- Jernigan, C., Mistree, B. (2009) "Gaydar: Facebook friendships expose sexual orientation". First Monday. DOI: 10.5210/fm.v14i10.2611.
- Kotsios, A., Magnani, M., Rossi, L., Shklovski, I., Vega, D. (2019) "An Analysis of the

¹⁸ Todas as referências online foram acessadas e estavam disponíveis em 30/03/2020.

- Consequences of the General Data Protection Regulation on Social Network Research". *ACM Transactions on Social Computing*. DOI: 10.1145/3365524
- Larson, D., Chang, V. (2016) "A review and future direction of agile, business intelligence, analytics and data science". *IJIM*. DOI: 10.1016/j.ijinfomgt.2016.04.013
- Lobo, H. (2015) "O que é vício de consentimento?". *Jus Navigandi*. Disponível em: cutt.ly/1tmX1Qs
- Manzoor, A. (2017). "Ethics of Social Media Research". *Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelligence*. DOI: 10.4018/978-1-5225-2031-3.ch013
- Moor, J. H. (2005) "Why we need better ethics for emerging technologies". *Ethics and Information Technology*. DOI: 10.1007/s10676-006-0008-0.
- Moribe, G., Júnior, O., Monteiro, R. (2019) "Comparing privacy laws: GDPR v. LGPD". Disponível em: cutt.ly/8tmX0ST
- Mortier, R., Haddadi, H., Henderson, T., McAuley, D., Crowcroft, J. (2015) "Human-Data Interaction: The Human Face of the Data-Driven Society". *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2508051
- Oliveira, M. I., Lóscio, B. F. (2018) "What is a data ecosystem?". *ICPS Proceedings*. DOI: 10.1145/3209281.3209335
- Pentland, A. (2013). "THE DATA-DRIVEN SOCIETY". *Sci. American*, 309(4), pp. 78-83.
- Recker, J. (2013) "Scientific research in information systems: a beginner's guide". Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-3-642-30048-6
- Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., Gordijn, B. (2018). "Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations". *Sci. Eng. Ethics*. DOI: 10.1007/s11948-017-9961-8
- Stattner, E., Vidot, N. (2011) "Social network analysis in epidemiology: Current trends and perspectives". *V ICRCIS*. DOI: 10.1109/RCIS.2011.6006866
- Sumpter, D. (2018) "Outnumbered: From Facebook and Google to Fake News and Filter-bubbles – The Algorithms That Control Our Lives". Bloomsbury Sigma. ISBN: 9781472947413.
- The Economist (2017). "The world's most valuable resource is no longer oil, but data". Disponível em: cutt.ly/XtmX3K5
- The Moscow Times (2019) "'We Tracked Your Social Media, You're Gay': Russian Student Threatened With Expulsion". Disponível em: cutt.ly/jtmX4cW
- UE - União Europeia (2018) General Data Protection Regulation (GDPR): Regulation (EU) 2016/679. Disponível em: <https://gdpr-info.eu/>
- Yang, S., Keller, F. B., Zheng, L. (2017) "Social Network Analysis: Methods and Examples". SAGE Publication Inc. ISBN: 9781483325217.
- Zheleeva, E., Getoor, L. (2011) "PRIVACY IN SOCIAL NETWORKS: A SURVEY". Em *Social Network Data Analytics*, Org. Aggarwal, C. C. Springer US. DOI: 10.1007/978-1-4419-8462-3_10.