


Pentágono da privacidade no *Big Data Analytics*: proposta de modelo teórico

Brenner Lopes

Mestre em Administração

Centro Universitário Unihorizontes, Belo Horizonte, MG, Brasil

 <https://orcid.org/0000-0002-5807-0437> E-mail: brenner.lopes@gmail.com

Ricardo Rodrigues Barbosa

Doutor em Administração de Empresas

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

 <https://orcid.org/0000-0003-3366-7525> E-mail: rrbarb@gmail.com

Luander Cipriano de Jesus Falcão

Mestre em Administração

Universidade Fundação Mineira de Educação e Cultura, Belo Horizonte, MG, Brasil

 <https://orcid.org/0000-0003-2417-6345> E-mail: luanderfalcao@yahoo.com.br

Renato Rocha Souza

Doutor em Ciência da Informação

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

 <https://orcid.org/0000-0002-1895-3905> E-mail: rsouzaufmg@gmail.com

Submetido em: 13-09-2023

Reapresentado em: 21-12-2023

Aceito em: 30-12-2023

RESUMO

Vivemos num ambiente caracterizado como um oceano de dados, que cresce não só quanto ao seu volume e quantidade, mas também em termos de variedade, sendo criado e transitando em alta velocidade. Atualmente os dados estruturados estão em quantidade e importância bem menor, e os ajustes e aprimoramentos nas tecnologias e modelos analíticos

foram em parte realizados para se adaptarem a essa nova realidade, que convencionou-se chamar de *Big Data Analytics*. Entre as questões de grande preocupação, nessa nova realidade, estão as ameaças à privacidade. A questão posta como resultado de diversas pesquisas é que os procedimentos, técnicas, tecnologias e legislações, atualmente disponíveis, não conseguem dar garantia plena à privacidade. Diante desse complexo cenário, o objetivo dessa pesquisa foi propor um modelo teórico multifacetado no âmbito do *Big Data Analytics*, que garanta a privacidade, ao mesmo tempo em que não inviabilize sua extração de valor. A metodologia proposta para esse trabalho foi a revisão sistemática da literatura, com vistas à análise crítica dos apontamentos e conclusões de estudos anteriores, a identificação e proposição lógica de novas hipóteses e construtos, de maneira a formatar o desenho final de um modelo teórico. Como resultado é proposto o Pentágono da Privacidade no *Big Data Analytics*, que contempla um caleidoscópio de soluções capazes de garantir a privacidade ao mesmo tempo que dá garantias à extração de valor no *Big Data Analytics*. O construto obtido como resultado desse trabalho, traz uma resposta concisa e consistente à questão de partida desse trabalho.

Palavras-chave: privacidade; big data analytics; big data privacy; valor do big data; pentágono da privacidade.

Privacy pentagon in Big Data Analytics: theoretical model proposal

ABSTRACT

We live in an environment characterized as an ocean of data, which grows not only in terms of volume and quantity, but also in terms of variety, being created and moving at high speed. Currently, structured data is much smaller in quantity and importance, and adjustments and improvements in technologies and analytical models were partly carried out to adapt to this new reality, which is known as Big Data Analytics. One of the issues of great concern in this new reality are threats to privacy. The question raised as a result of several research is that the procedures, techniques, technologies, and legislation currently available cannot fully guarantee privacy. Given this complex scenario, the objective of this research was to propose a multifaceted theoretical model within the scope of Big Data Analytics, which guarantees privacy, while not making its extraction of value unfeasible. The methodology proposed for this work was supported by the systematic literature review approach, with a view to critically analyzing the notes and conclusions of previous studies, identifying, and logically proposing new hypotheses and constructs, in order to format the final design of a theoric model. As a result, the Pentagon of Privacy in Big Data Analytics is proposed, which includes a kaleidoscope of solutions capable of guaranteeing privacy while guaranteeing the extraction of value in Big Data Analytics. The construct obtained as a result of this work provides a concise and consistent answer to the starting question of this work.

Keywords: privacy; big data analytics; big data privacy; value of big data; privacy pentagon.

1 INTRODUÇÃO

Uma das características mais visíveis do nosso tempo é que literalmente vivemos cercados por um oceano de dados. E seu crescimento se dá não só quanto ao seu volume e quantidade, mas também em termos de variedade, sendo criado e transitando em alta velocidade. Se antes o foco e a disponibilidade de dados tinham como interface prioritária os chamados dados estruturados, hoje eles estão em quantidade e importância bem menor, e os ajustes e aprimoramentos nas tecnologias e modelos analíticos foram em parte realizados para se adaptarem a essa nova realidade, que convencionou-se chamar de *Big Data*.

Com toda essa disponibilidade de dados e a possibilidade de coleta, armazenamento, análise e criação de valor atrelado a múltiplos objetivos e interesses, disponíveis para praticamente qualquer pessoa e empresa, por um lado, surgiram novos problemas e, por outro, problemas já existentes foram agravados.

Como destacado por Oostveen (2016, p. 299, tradução nossa), a “enorme coleção e o uso de dados pessoais também levanta uma série de questões, entre elas e de maneira particular a questão da privacidade [...]”.

O foco dessa pesquisa estará assentado no contexto do *Big Data Analytics*, onde os times internos e/ou externos de analítica têm franco acesso aos dados, portanto, não há uma questão tão premente de segurança, mas sim de entender se conseguem garantir em seus processos, processamentos e análises a privacidade no âmbito dos dados acessados e utilizados.

No Brasil, no ano de 2018, foi aprovada a Lei Geral de Proteção de Dados (LGPD), que visa regulamentar o acesso e utilização aos dados pessoais, assim como salvaguardar o direito de todo cidadão à privacidade. Essas iniciativas reforçaram outras questões fundamentais na esfera da garantia da privacidade, como as abordagens tecnológicas e técnicas de preservação da privacidade, que já existiam e eram praticadas por alguns agentes, públicos e privados, mas que tomaram novas dimensões e perspectivas a partir da Lei.

Como colocado por Acquisti, Taylor e Wagman (2016, p. 483, tradução nossa), “na verdade, exploração do valor comercial dos dados pode muitas vezes implicar numa redução de sua utilidade privada, e às vezes, mesmo no bem-estar social em geral [...]”. Então aqui temos um grande dilema que integrará a estrutura principal desse trabalho, que é a importância de se

poder extrair o valor do *big data* ao mesmo tempo em que se garante a privacidade. Esse é um verdadeiro dilema que merece atenção: se é possível garantir privacidade total ao custo da nulidade das possibilidades de extração de valor do *Big Data Analytics*.

Tem-se, portanto, como objetivo dessa pesquisa propor um modelo multifacetado no âmbito do *Big Data Analytics*, que garanta a privacidade, ao mesmo tempo em que não inviabilize sua extração de valor.

2 REFERENCIAL TEÓRICO

2.1 SEGURANÇA

A análise de grandes volumes de dados é um problema até então sem solução definitiva, não só para a privacidade, mas também para a segurança. É praticamente impossível hoje em dia afirmar que um determinado sistema é totalmente seguro, principalmente se esse tiver interação com o ambiente da internet.

E à medida que novas tecnologias e aplicativos vão surgindo, a régua dos desafios da segurança vai também se alargando em proporção muito maior. “[...] para salvaguardar as informações pessoais, é crucial que os processos de armazenamento e transporte sejam integrados com medidas de segurança [...]” (Adams, 2017, p. 17, tradução nossa).

Segurança e privacidade são questões siamesas, ou seja, estão em geral sempre juntas, apesar de que ao mesmo tempo em que não existe privacidade sem uma ou mais camadas de segurança, a existência dessa última não consegue garantir em sua plenitude a privacidade. Constata-se ainda uma grande confusão entre os termos, que são utilizados em muitos casos como se fossem sinônimos.

Na visão de Sun, Pambel e Strang (2018, p. 4, tradução nossa), segurança refere-se às “políticas, procedimentos e medidas técnicas usadas para evitar acesso não autorizado, alternância, roubo de dados ou danos físicos a dispositivos e sistemas [...]”. Segurança é a prática de “defesa de informações e ativos de informação através do uso de tecnologia, processos e treinamento contra: acesso não autorizado, divulgação, interrupção, modificação, inspeção, gravação, destruição [...]” (Jain; Gyanchandani; Khare, 2016, p. 3, tradução nossa).

Segurança é confidencialidade, integridade e disponibilidade de dados. É a proteção contra o acesso não autorizado, permanecendo, portanto, confiável, íntegra e precisa, estando acessível sempre que solicitada, mas protegida contra ataques e roubos com finalidade escusas (Abouelmehdi; Beni-Hessae; Khaloufi, 2018, El Ouazzani; El Bakkali, 2020; Jain; Gyanchandani; Khare, 2019).

2.2 PRIVACIDADE

Abrimos essa subseção ponderando sobre uma importante e ampla reflexão trazida por Westin (1970, p. 34, tradução nossa), “o desenvolvimento da individualidade é particularmente importante nas sociedades democráticas, uma vez que qualidades de pensamento independente, diversidade de pontos de vista e não conformidade são consideradas características desejáveis para os indivíduos [...]”.

Talvez a característica mais relevante do conceito de privacidade e que lhe confere contornos múltiplos, sob muitos olhares, seja possuir um conceito dinâmico, que pode sofrer variações / alterações de acordo com as tecnologias e normas sociais vigentes (Hadar *et al.*, 2018).

A privacidade foi inserida como um direito em 1890, por Warren e Brandeis (1890). E conforme colocado por Politou, Alepis e Paysakis (2018, p. 2, tradução nossa) somente nas “últimas três décadas que ela foi amplamente discutida em suas várias formas e contextos, principalmente devido à computação e ciência da informação [...]”. No final da década de 1960, Politou, Alepis e Paysakis (2018, p. 2, tradução nossa) destacam que o conceito de privacidade surgiu dentro de um contexto mais filosófico, e desde então, “é discutida em grande controvérsia entre os círculos filosóficos, jurídicos, sociais e científicos [...]”.

Ainda segundo Politou, Alepis e Paysakis (2018, p. 2, tradução nossa), “não existe uma definição universalmente aceita de privacidade [...]”. Ela pode ser vista sob um caleidoscópio, ou seja, sob múltiplas visões e pontos de vista, como o controle sobre os dados de uma pessoa por ela própria, o direito de não ser monitorado e nem mesmo identificado, de ser deixado em “paz” e até mesmo ser “esquecido”.

Conforme colocado por Sun, Pambel e Strang (2018, p. 4, tradução nossa), a privacidade “consiste em dois pontos principais: confidencialidade e uso justo [...]”. A garantia

de privacidade, ou a tentativa dessa garantia, se refere ainda à possibilidade de ocultar a verdadeira identidade de uma pessoa. Por sua vez a “segurança, lida com confidencialidade, integridade e disponibilidade [...]” (El Ouazzani; El Bakkali, 2020, p. 143, tradução nossa).

Frequentemente a privacidade é definida como a capacidade de proteger informações confidenciais sobre as informações pessoalmente identificáveis, ou seja, dados que direta ou indiretamente possam identificar um indivíduo (Abouelmehdi; Beni-Hessane; Khaloufi, 2018).

Aprofundando o entendimento do que venha a ser privacidade, pode se dizer que privacidade é o querer do indivíduo de ser deixado em paz, livre de qualquer interferência, vigilância de outros indivíduos, organizações ou sistemas.

Por fim, também pode ser compreendida como a determinação de um indivíduo de deliberar sobre quais dados seus poderiam ser compartilhados e em que extensão, ao mesmo tempo em que possui algum controle sobre esses. Seria ainda a não divulgação de informações pessoais de forma pública (Jain; Gyanchandani; Khare, 2016; Rao; Krishna; Kumar, 2018; Sun; Pambel; Strang, 2018; Westin, 1970).

Solove (2002, p. 1129, tradução nossa), apresenta uma alternativa para essa situação, ao afirmar que a “maioria dos teóricos tenta conceituar a privacidade isolando um ou mais aspectos essenciais e comuns”. Nesse sentido propõe que a discussão e definição desse termo deveria ser discutida sob a ótica de seis temas conjuntamente: “1) o direito de ser deixado em paz; 2) acesso limitado a si mesmo; 3) sigilo; 4) controle de informações pessoais; 5) personalidade; 6) intimidade [...]” (Solove, 2002, p. 1094, tradução nossa).

Não seria razoável finalizar essa subseção, sem apontar um outro termo também muito relacionado e mais ainda confundido e utilizado de maneira indistinta e intercambiável quando da referência à privacidade que é a questão do conceito de “proteção” e “proteção de dados”, ambos com forte vínculo à questão da privacidade, mas que “na verdade, constituem duas noções distintas” (Politou; Alepis; Paysakis, 2018, p. 2, tradução nossa).

Com foco num maior entendimento dessa questão, Politou, Alepis e Paysakis (2018, p. 1, tradução nossa), definem a questão da privacidade como geralmente se referindo “à proteção do espaço pessoal de um indivíduo, enquanto a proteção de dados se refere a limitações ou condições no processamento de dados relativos a um indivíduo identificável [...]”.

2.3 BIG DATA E BIG DATA ANALYTICS

Como conceito, *big data* não é um conceito tão novo. Segundo Adams (2017, p. 13, tradução nossa), ele existe há pelos menos duas décadas, “desde que foi usado por Cox e Ellsworth (em 1997)”, quando esses autores se referiram ao termo no contexto de um conjunto imenso de dados científicos.

Jain, Gyanchandani e Khare (2019, p. 2, tradução nossa), dão uma ideia da dimensão desse fluxo de dados, ao apontarem que “todos os dias quintilhões de bytes de dados são criados, ou seja, 90% dos dados do mundo hoje foram criados apenas nos últimos dois anos [...]”.

Uma definição básica e mais comum de *big data* é a que o define como um conjunto de dados que são tão grandes e complexos que as aplicações tradicionais de processamento de dados não são suficientes, para armazená-los, processá-lo e analisá-los. Na atualidade, o gerenciamento dos grandes fluxos de dados, por meio de ambientes físicos e virtuais, e o processamento de *big data* se tornou um fator de criticidade para praticamente todas as empresas. E o mesmo tem apontado para uma nova realidade que é destacada não só pelo tamanho dos dados, mas também pela sua complexidade de organização e análise (Adams, 2017; Jain; Gyanchandani; Khare, 2016; El Ouazzani e El Bakkali, 2020; Sun; Pambel; Strang, 2018).

Mas não se pode obter uma compreensão completa sobre a dimensão e complexidade do *big data*, sem se entender minimamente algumas de suas características mais relevantes, que lhe dão os contornos e a singularidade que possui. Nesse âmbito, o *big data* poder ser descrito pelas características ou dimensões dadas por seus Vs. Originalmente estavam caracterizadas por seu volume, variedade e velocidade. Mais tarde, novos estudos e reflexões apontaram para a insuficiência dessas três características como definidoras do *big data* e, portanto, passaram a considerar nesse conjunto novas características, como valor, veracidade e variabilidade (El Ouazzani; El Bakkali, 2020; Jain; Gyanchandani; Khare, 2016; Sarkar, 2017; Sun; Pambel; Strang, 2018; Wilson; Belliveau; Gray, 2017).

Jain, Gyanchandani e Khare (2016, p. 1, tradução nossa), complementam esse conceito geral e aceito de *big data* ao pontuarem que o *big data* poder ser definido “como uma geração de tecnologias e arquiteturas, projetadas para volumes muito grandes [...]” de um amplo e variado espectro de dados (quantitativos, qualitativos, estruturados, semiestruturados e

desestruturados), possibilitando elevada velocidade de coleta, processamento, análise e geração de *outputs*.

Oostveen (2016, p. 302, tradução nossa) coloca que num nível mais básico, o *big data*, “entra em conflito com a privacidade e proteção de dados [...]”. Isso porque segundo os autores, “a coleta de dados na fase de aquisição pode revelar detalhes íntimos sobre a vida de uma pessoa [...]”, ou seja, infringe sua privacidade.

Mills (2018, p. 598, tradução nossa) complementa e aprofunda as colocações feitas por Oostveen (2016) ao destacar que o *big data* “é usado para objetivos específicos e usa algoritmos e análises específicas que podem causar desconforto [...]”, assim como “servir a interesses particulares que podem ou não estar alinhados [...]” com os interesses dos detentores dos dados analisados. É justamente nesse ponto em que muitas questões no âmbito da privacidade podem não ser levadas em conta.

Na esfera da privacidade, o *big data* “foi definido como dados sobre um indivíduo ou um grupo de indivíduos, que podem ser analisados para fazer inferências sobre esses indivíduos [...]” (Wilson; Belliveau; Gray, 2017, p. 3, tradução nossa).

Para o fechamento dessa seção é fundamental construir um vínculo mais consistente entre *big data* e *Big Data Analytics*. Apesar de serem termos difíceis de se “separar”, pois, se não há a análise dos grandes dados, na verdade não se têm nada, apenas grandes dados. Por isso mesmo, não é possível, sob uma condição lógica, ter-se “*big data*” se não se tem junto o “*analytics*”.

Sun, Sun e Strang (2016, p. 2, tradução nossa), apresentam uma consistente definição para a análise de *big data* (*Big Data Analytics*), onde sustentam que essa poderia ser definida como “o processo de coleta, organização e análise do *big data* para descobrir, visualizar e exibir padrões, conhecimento e inteligência [...]”.

2.4 LEGISLAÇÃO DE PROTEÇÃO DE DADOS

Em vários países ao redor do mundo, o debate sobre a questão da privacidade e da proteção de dados está muito ativo, isto porque apesar das legislações que passaram a vigorar nos últimos anos e mesmo algumas leis que já vigoram há mais tempo, o que os países e suas

lideranças têm visto é que a legislação não tem sido suficiente para fazer frente aos avanços do *Big Data Analytics*.

Esses instrumentos legislativos são imensamente relevantes, mas com certeza deverão ser continuamente revistos e melhorados. E mesmo essa revisão contínua não será suficiente para equacionar e/ou antecipar e prever todas as questões, dada a velocidade da análise de dados.

No Brasil a Lei Geral de Proteção de Dados (LGPD), Lei 13.709/18 (Brasil, 2018), foi sancionada em 14 de agosto de 2018, após vários anos de discussão. A lei de proteção de dados brasileira se inspirou de forma bem ampla no Regulamento Geral de Proteção de Dados (GDPR) da União Europeia. Ishii (2019, p. 517, tradução nossa) aponta que, como no caso da LGPD, embora a GDPR tenha o potencial de melhorar a proteção de dados, ainda precisa de “mais trabalho para formular diretrizes normativas e mecanismos práticos para colocar em prática os novos direitos e responsabilidades [...]”.

A LGPD se aplica a qualquer pessoa física ou jurídica (pública ou privada), que desenvolva projetos, iniciativas e ações que envolvam o tratamento de dados pessoais. Basicamente, ela estabelece regras no que diz respeito à coleta, armazenamento e compartilhamento de dados pessoais.

É relevante no âmbito dos dados pessoais entender-se os conceitos e diferenças existentes entre dados sensíveis, dados diretamente identificáveis e dados indiretamente identificáveis. Segundo a LGPD (Brasil, 2018, artigo 4º), denomina-se dado pessoal sensível como sendo aqueles vinculados a uma pessoa natural e aos aspectos raciais, étnicos, “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”.

Como já apontado, a questão da privacidade não é um tema de debate recente, mas sua concepção e garantia em tempos de *Big Data Analytics* é com certeza um tema recente de pesquisas, assim como de preocupações da sociedade. Por conseguinte, tornou-se uma questão de preocupação mundial.

3 METODOLOGIA

O trabalho de revisão da literatura cumpre diversos objetivos de grande relevância para todas as pesquisas, e de maneira específica para o caso da pesquisa em questão foram

fundamentais para dar o direcionamento e consistência necessários, assim como possibilitar o compartilhamento dos achados de outros estudos já realizados sobre a temática diretamente e/ou sobre temáticas adjacentes que compõe uma visão ampliada do campo de pesquisa proposto (Creswell, 2010).

Como definido por Noronha e Ferreira (2000, p. 182) a revisão da literatura pode ser entendida como “estudos que analisam a produção bibliográfica em determinada área temática, dentro de um recorte de tempo, fornecendo uma visão geral ou um relatório do estado-da-arte sobre um tópico específico [...]”. Tais análises cumprem um conjunto de objetivos, como a evidenciação de novas ideias, técnicas, métodos, metodologias, apontamento de opiniões e visões em comum a diversos pesquisadores, dentre outros.

As revisões de literatura, conforme proposto por Noronha e Ferreira (2000, p. 184), podem ser classificadas segundo “seu propósito, abrangência, função e tipo de análise desenvolvida [...]”. Com base nessa classificação proposta por Noronha e Ferreira (2000), essa revisão de literatura pode ser classificada quanto ao seu propósito como de base, que se advoga como aquela que visa suportar a comprovação ou não de hipóteses e questões de pesquisa. Quanto à sua abrangência ela mescla a abordagem temática com a temporal, por estar focada em um tema específico e dentro de um recorte temporal estabelecido (últimos cinco anos). Quanto à sua função, ela se classifica como de avaliação, pelo seu foco no levantamento e análise de estudos recentes, identificando informações, visões e resultados de outros trabalhos para o desenvolvimento de toda a lógica do modelo dessa pesquisa; por fim, quanto ao tratamento e abordagem dada aos trabalhos analisados, ela se enquadra como bibliográfica, ou seja, por ter servido para comparação das abordagens realizadas no âmbito dos múltiplos trabalhos analisados, possibilitando a identificação daqueles de maior relevância e aderência ao tema proposto para essa pesquisa.

A criação de um modelo de privacidade no *Big Data Analytics*, necessitou da identificação e análise das hipóteses e construtos já evidenciados em trabalhos anteriores, de forma a possibilitar uma base inicial e com respaldado em múltiplos trabalhos e autores de referência, formatando assim o corpo inicial da estrutura lógica do modelo.

Por outro lado, também possibilitou através da análise crítica e dos apontamentos relatados em estudos anteriores, a identificação e proposição “lógica” de novas hipóteses e construtos, formatando assim, o desenho final do modelo a ser proposto.

Bryman (1989, p. 17, tradução nossa) traz algumas reflexões e colocações sobre esse processo, apontando que “muitas vezes as hipóteses e seus conceitos associados são o produto de deliberações em conexão com a literatura relacionada a um campo substantivo [...]”. Complementando o raciocínio e fazendo o vínculo direto à abordagem aqui realizada, afirma que “questões teóricas anteriores podem surgir como justificativas para a inclusão de variáveis específicas ou para os padrões descritos nas hipóteses [...]”.

3.1 ESTRATÉGIA DE BUSCA E SELEÇÃO DOS CONTEÚDOS PARA REVISÃO DA LITERATURA

Partindo desse ponto, na revisão de literatura executada nessa pesquisa, optou-se por utilizar o Portal CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) (Brasil, 2023). Não como uma primeira camada de busca para posterior refinamento, mas como a principal ferramenta de busca e posterior seleção e refinamento dos resultados dessas buscas. Como exposto acima, isso se deu principalmente com base na característica mais visível e impactante da temática, que é seu contorno multifacetado, sendo discutido de um lado (*Big Data Analytics*) ou de outro (*big data privacy*) em múltiplos repositórios, não possuindo uma predominância, a princípio, nem mesmo naqueles com forte posicionamento multidisciplinar.

Foram realizadas um total de seis tentativas de buscas e respectivas extrações e análises de todo material, no período de 31/08/2020 a 11/09/2020. Na sexta tentativa, compreendida entre os dias 09 e 11/09/2020 foi possível se chegar a um algoritmo que retornou o melhor resultado. A estratégia de busca utilizada foi “Big Data Analytics” AND “Big Data Privacy”. No quadro 1 se detalham as características da estratégia.

Quadro 1 – Características da estratégia de busca

Termos	Qualquer campo	Data publicação	Tipo de material	Idioma	Filtro	Filtro
É exato para os dois termos	Título, autor, assunto	Últimos 5 anos	Artigos	Inglês	Ordenação por relevância	Mostrar somente periódicos revisados por pares

Fonte: Elaboração própria (2023).

A busca retornou 125 documentos classificados como artigos. Mais da metade (55%) dos materiais recuperados foram originários de duas bases principais: Scopus (31%) e

Aerospace Database Technology Research (25%). Numa proporção próxima, foram recuperados também documentos nas fontes Database, 16%; Web of Science (14%) e Onelife (14%).

A partir desse ponto, os artigos foram analisados com base nos seguintes critérios: a) condições de acesso aos artigos; b) confirmação de que todos os artigos estivessem em inglês; c) confirmação de que todos os documentos fossem efetivamente artigos; d) confirmação de não repetição dos documentos, visto que os mesmos poderiam estar publicados em mais de um repositório; e) análise da presença dos termos-chave de busca, principalmente do termo “privacy”; e) análise dos resumos.

Ao final de todo o processo criterioso de análise dos documentos chegou-se a um conjunto de trinta e três artigos, revisados pelos pares. A partir desse ponto procedemos à leitura completa e análise de todos os artigos.

4 ANÁLISE DE RESULTADOS

A literatura tem apontado consistentemente para o fato de que ainda não se possui um modelo suficientemente robusto que consiga garantir a privacidade dos dados em análises de *big data*, ao mesmo tempo que também garanta a extração do seu valor. A constatação é que o problema é complexo e envolve múltiplas dimensões como legislação, política e códigos de conduta, sensibilização, comportamento e tecnologia.

Uma primeira confirmação dessa constatação, com um foco na segurança e tecnologia, pode ser vista nas afirmações de Pragash e Jayabharathy (2017, p. 95, tradução nossa), quando apontam que as medidas de segurança consideradas tradicionais, como “Firewall, antivírus e sistemas de detecção e prevenção de intrusão, não estão mais fornecendo os níveis exigidos de granularidade, proteção e fiscalização [...]”, requisitados para o atendimento das regulamentações no campo da segurança e privacidade.

Como afirma Rao, Krishna e Kumar (2018, p. 9, tradução nossa) na discussão dos resultados da revisão sistemática da literatura que empreenderam, “observou-se que todos os mecanismos existentes de preservação da privacidade dizem respeito a dados estruturados [...]”. A questão que se coloca aqui é que mais de 80% dos dados gerados hoje não são estruturados.

Ohm (2010, p. 1704, tradução nossa) aprofunda a complexidade da discussão ao colocar que “os dados podem ser úteis ou perfeitamente anônimos, mas nunca ambos [...]”. O que já se sabe é que as legislações, técnicas e tecnologias disponíveis ainda não conseguiram desatar esse nó (Adams, 2017; Ishii, 2017; Oostveen, 2016; Rao; Krishna; Kumar, 2018; Sun; Pambel; Strang, 2018; Wilson; Belliveau; Gray, 2017).

Some-se a isso uma última questão fundamental: a de que a proteção / garantia da privacidade a nível das empresas (públicas e/ou privadas), está atrelada, de alguma forma à determinação dessas em seguir não só a legislação e efetivamente utilizar técnicas e tecnologias, mas também de instituírem códigos internos de ética, comportamento e procedimentos visando esse objetivo, já que a detecção de possíveis descumprimentos, por exemplo, a nível da legislação instituída é algo bem complicado, que só se torna mais visível em casos de grandes proporções ou com base em denúncias (que assim também precisam ser confirmadas).

Como reforçam Sun, Pambel e Strang (2018, p.6, tradução nossa), “ainda existem lacunas na proteção da privacidade do indivíduo [...]”. El Ouazzani e El Bakkali (2020, p. 3, tradução nossa) apontam para uma das questões fundamentais quanto a motivação e justificativa da importância da temática dessa pesquisa, ao afirmarem que “a maioria das técnicas convencionais de privacidade de dados não suporta a escala completa do *big data* [...]”. Portanto, ainda na visão dos autores, “é obrigatório garantir a privacidade, garantindo que todas as tentativas de identificar o indivíduo falhem [...]” (El Ouazzani; El Bakkali, 2020, p. 2, tradução nossa).

Pode-se verificar de forma clara que a questão da privacidade está além, apesar de precisar e se estruturar a partir das questões legais e de segurança. Politou, Alepis e Paysakis (2018, p. 15, tradução nossa) reforçam essa constatação ao afirmarem em sua análise sobre a questão da vigência da GDPR na União Europeia que, “poucas organizações ainda são capazes de provar a conformidade real [...]” com essa norma. E segundo os mesmos autores, num aprofundamento dessa constatação, um dos principais fatores causadores desse panorama negativo é que a GDPR é um documento legal.

Tendo como pano de fundo esse desafio de segurança no âmbito do *big data*, Thomson e Thibadeau (2016) propõe três princípios a serem seguidos: a) a segurança deve ser incorporada à arquitetura e ao design (design por privacidade e design por padrão) dos

sistemas de informação da organização e aos ativos de tecnologia da informação; b) uma organização deve empregar uma estratégia de defesa em profundidade para lidar com todas as vulnerabilidades das soluções no campo do *big data*; c) não se devem implementar soluções com vulnerabilidades já previamente conhecidas.

Na visão de Sun, Pambel e Strang (2018, p. 5, tradução nossa) “as soluções de segurança tradicionais não são projetadas para proteger a privacidade individual na era do *big data* [...]”, o que reforça a importância de se pensar e propor modelos que consigam abarcar o maior número de questões determinantes da garantia da privacidade.

Jain, Gyanchandani e Khare (2016, p. 3, tradução nossa) reforçam essa visão ao afirmar que “embora a segurança seja fundamental para proteger os dados, não é suficiente para abordar a privacidade [...]”. E complementam ao raciocinar, refletir e afirmar que “a análise avançada de dados pode extrair informações valiosas do *big data*, mas ao mesmo tempo, representa um grande risco para a privacidade dos usuários [...]” (Jain; Gyanchandani; Khare, 2016, p. 10, tradução nossa).

A constatação anterior pode ser corroborada pelas colocações de Chanson *et al.* (2019, p. 5, tradução nossa) quando os autores são categóricos em afirmar que “apesar do corpo de conhecimento existente, faltam soluções viáveis [...]” à problemática da privacidade. Sendo complementadas pelas afirmações de Jain, Gyanchandani e Khare (2016, p. 23, tradução nossa) ao afirmarem nas conclusões do seu estudo que “como tal, existe uma enorme margem para mais pesquisas sobre métodos de preservação da privacidade em *big data* [...]”.

Abouelmehdi, Beni-Hessane e Khaloufi (2018, p. 15, tradução) são taxativos ao declararem que “métodos de privacidade precisam ser aprimorados[...]”. Esse mesmo autor ainda destaca questões sobre um uso secreto do *big data*, apontando, portanto, para um possível lado negro do *big data*, concluindo que as preocupações e garantias no âmbito da privacidade requerem muito mais do que ajustes nos atuais “protocolos” (Mills, 2018, p. 597).

Nesse sentido, parece muito claro que a proteção / garantia da privacidade deverá congrega soluções tecnológicas, legais, sociais, culturais e políticas, tanto a nível de nações quanto de empresas.

Partindo do ponto de entendimento de que os mecanismos atualmente disponíveis para prover a garantia da privacidade em análises no *Big Data Analytics*, não são suficientes para prover essa pretensa garantia, a questão de partida que motivou essa proposta de

pesquisa está amparada pela seguinte pergunta, para a qual propõe-se uma alternativa, “como garantir a privacidade e o valor em análises de *big data*?”, que enseja, portanto, a “proposição de um modelo multifacetado de garantia da privacidade no *Big Data Analytics*”.

Conforme apresentado por Hair *et al.* (2009, p. 545), um “modelo é uma representação de uma teoria [...]” (portanto, não pode ser pensado sem uma teoria ou teorias consistentes suportando-a), que poderia ser visualizada como um conjunto lógico de relações em torno de uma determinada questão, um determinado tema, capaz de fornecer uma explicação consistente e abrangente sobre a questão inicialmente levantada.

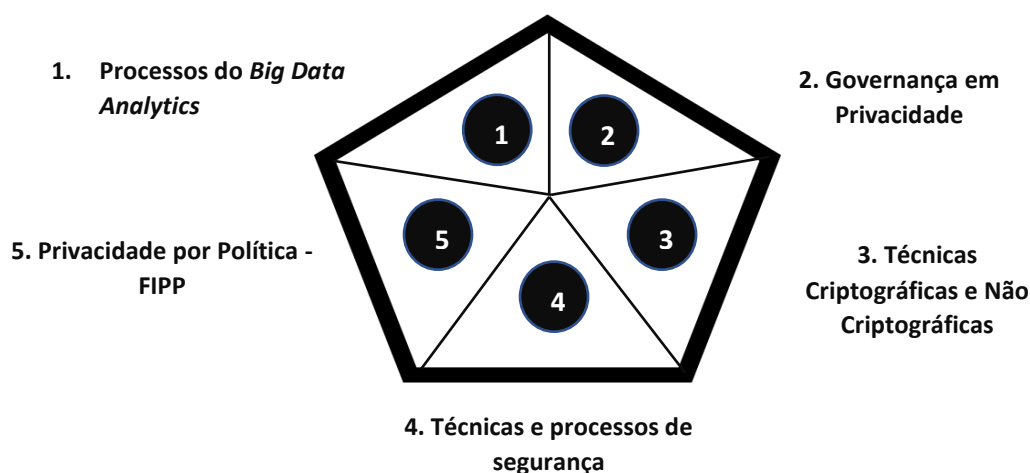
Esta pesquisa esteve assentada no esforço de identificação, análise lógica e validação dos construtos estruturantes de um modelo multifacetado que consiga prover efetiva garantia à preservação da privacidade, no que se refere aos dados pessoais, ao mesmo tempo em que garante a geração de valor, objetivo fundamental e sustentador da proposta do *Big Data Analytics*.

Os construtos identificados e o conjunto analítico aplicado a todo o conteúdo analisado apontou para a necessidade de se contemplar, como solução para o problema proposto nesse trabalho, cinco dimensões estratégicas e suas respectivas práticas componentes, que nominaremos nesse trabalho como modelo do “Pentágono da Privacidade no *Big Data Analytics*”.

4.1 MODELO DO PENTÁGONO DA PRIVACIDADE NO BIG DATA ANALYTICS

O modelo do “Pentágono da Privacidade no *Big Data Analytics*”, está estruturado sobre as cinco dimensões fundamentais e suas respectivas práticas visando garantir a privacidade no *Big Data Analytics*, ao mesmo tempo em que também consegue garantir a extração de valor a partir dessa abordagem. É, portanto, um modelo complexo, amplo e em constante evolução, pois sempre deve ser revisitado e atualizado, principalmente no que tange a suas práticas, *vis a vis* principalmente, os avanços das tecnologias e dos modelos analíticos no *Big Data Analytics*.

Na Figura 1 se apresenta uma visão geral do modelo estruturado com base nos levantamentos e achados dessa pesquisa; modelo esse que contempla um caleidoscópio de soluções capazes, como pontuado acima, de garantir a privacidade ao mesmo tempo que também dá garantias para a extração de valor no *Big Data Analytics*.

Figura 1 – Pentágono da Privacidade no *Big Data Analytics*

Fonte: Elaboração própria (2023).

Em linhas gerais as dimensões apresentadas no modelo podem ser entendidas da seguinte forma:

- 1) Processos do *Big Data Analytics*: essa dimensão contempla processos e práticas visando a garantia da privacidade em todo o ciclo de vida do *Big Data Analytics*: coleta, armazenamento, processamento e análise;
- 2) Governança em Privacidade: essa dimensão contempla questões como a existência de comitês de ética e revisão no âmbito do ciclo de vida do *Big Data Analytics*, privacidade por design, treinamento, inventário de ativos digitais, dentre outros;
- 3) Técnicas Criptográficas e Não Criptográficas: essa dimensão contempla o nível de utilização e maturidade da organização quanto a utilização de técnicas criptográficas e não criptográficas;
- 4) Técnicas e processos de segurança: essa dimensão contempla o nível de utilização e maturidade da organização quanto a utilização de técnicas e processos em segurança e proteção da informação;
- 5) Privacidade por Política-FIPP: essa dimensão contempla tanto o enfoque da privacidade por política, quanto a privacidade por arquitetura e a garantia dos princípios de prática de informação e privacidade justas. Em linhas gerais contempla a proteção da privacidade por meio de políticas, semelhante à lógica das legislações; assim como visam garantir a incorporação de funcionalidades de preservação da privacidade nos estágios iniciais do desenvolvimento de sistemas e processos.

No Quadro 2 são apresentados os principais elementos constituintes (práticas) de cada uma das dimensões do Pentágono da Privacidade no *Big Data Analytics*:

Quadro 2 – Elementos constituintes das dimensões do Pentágono da Privacidade no *Big Data Analytics*

PROCESSOS DO <i>BIG DATA ANALYTICS</i>	GOVERNANÇA EM PRIVACIDADE	TÉCNICAS CRIPTOGRÁFICAS E NÃO CRIPTOGRÁFICAS ADOTADAS PELA ORGANIZAÇÃO	TÉCNICAS E PROCESSOS DE SEGURANÇA ADOTADOS PELA ORGANIZAÇÃO	PRIVACIDADE POR POLÍTICA - FIPP (FAIR INFORMATION PRACTICES PRINCIPLES)
A organização trabalha sob a ótica do <i>smart data</i> (proteção de dados pessoais; incorporação de regras de acesso aos dados; respostas às solicitações de informações dependentes de suas regras de acesso)	A organização protege a privacidade de dados individuais por meio de políticas internas	Desidentificação/Pseudonimização (ex.: criptografia com chave secreta, função hash, função hash com chave armazenada, criptografia determinística, tokenização, dentre outros)	<i>Structured Map Reduced Layer</i>	Confidencialidade – as informações de identificação pessoal são protegidas com salvaguardas administrativas, técnicas e físicas
São identificados antecipadamente projetos e conjuntos de dados que contém dados pessoais, criando “zonas proibidas” para os mesmos	O controle do proprietário dos dados é um valor central nas soluções analíticas de nossa organização (acesso ao titular dos dados)	Anonimização/Generalização (ex.: diversidade L, proximidade T, k-anonimato baseado em generalização, dentre outros)	Segurança de dados – os dados pessoais são protegidos por salvaguardas de segurança	Especificação de propósito – as informações coletadas são usadas para um específico para o qual foram coletadas e não são utilizadas para outros fins sem a autorização devida
Utilizam-se combinações de controles (computação x inferência x uso) para gerenciar riscos em <i>big data</i>	A organização possui um profissional responsável pelo tratamento dos dados pessoais, cujos contatos estão publicamente divulgados em seu site também em outros canais de comunicação	Anonimização/Randomização (ex.: K-Anonimato baseado em supressão, permutação, substituição, privacidade diferencial, <i>nulling out</i> , mascaramento de dados, embaralhamento,	Controle de acesso (ex.: Autenticação, autorização, auditoria, MAC – Integração e controle de acesso obrigatório)	A organização possui uma política clara e de conhecimento de todos quanto aos procedimentos de comunicação ao mercado e stakeholders, sobre possíveis incidentes de vazamento ou roubo de dados
Há a garantia quanto ao armazenamento e compartilhamento seguro de dados	A organização incorpora funcionalidades de preservação da privacidade nos estágios iniciais do desenvolvimento de sistemas e soluções (<i>Privacy by Design</i>)	Outras técnicas criptográficas: triple DES, anonimização baseada em identidade, homomórfica	Segurança da informação	Consentimento – busca o consentimento do indivíduo para coleta, processamento, uso e transferência de seus dados e manutenção dos consentimentos individuais
É garantida a transparência algorítmica em nossos processos de análise	A organização possui um comitê de ética e revisão no âmbito da coleta, processamento e análise de dados em <i>Big Data</i>	Modelo de execução híbrida – HybrEx (confidencialidade e privacidade em computação em nuvem)	Processo foram de capacitação e certificação de todos o time da organização, que deve conhecer e trabalhar para conter as principais técnicas de ataque (link lateral,	Aviso – Informa ao titular sobre a coleta de seus dados

			palpite óbvio, reidentificação por meio de extremidades, dentre outros)	
A discriminação algorítmica não é um resultado analítico aceito pela organização	A organização tem estabelecido um processo e constante revisão e atualização de sua política de privacidade, refletindo e o acompanhamento e as melhorias nesse âmbito	Modelos de classificação de dados sensíveis (ex.: SVM Multi-Kernel)		Minimização de dados – impõe limitações sobre os tipos de informações e organização para coletar dados sobre um indivíduo
As atividades de tratamento de dados na organização observam a boa-fé quanto aos princípios da finalidade, adequação e necessidade	A organização realiza anualmente um inventário de ativos digitais	Técnicas de limitação de divulgação estatística (agregação, supressão, perturbação)		Existem políticas e procedimentos diferenciados quanto aos dados e análises de populações vulneráveis (ex.: crianças, mulheres grávidas, prisioneiros, etc.)
São definidos parâmetros para as permissões de transmissão, armazenamento, uso e descarte dos dados (“dados portadores de políticas”)	A organização possui políticas e práticas que consideram os cuidados e limites necessários quando da utilização de dados fornecidos por “corretores de dados”			Existência de políticas que disciplinam não só quem pode analisar, mas que tipos de dados podem ser analisados
Efetiva existência e consistência quanto aos procedimentos, políticas e práticas visando a proteção e preservação da privacidade em todo o ciclo de vida do <i>Big Data</i> (coleta, armazenamento, processamento e análise)	A organização requer e garante que seus fornecedores possuam protocolos de segurança capazes de garantir a privacidade e a confidencialidade dos dados			Retificação – permite que o titular dos dados exija retificação dos mesmos caso estejam imprecisos

Fonte: Elaboração própria (2023).

A lógica geral do modelo se aplica quando todos esses elementos estão presentes, atuantes e integrando as práticas e a cultura das organizações. Em última instância, a depender do estágio de maturidade em que se encontra a organização, pode ser visto como um roteiro de melhores práticas, um caminho para se alcançar a maturidade em garantia da privacidade no âmbito do *Big Data Analytics*.

5 CONCLUSÕES

Apesar da preocupação com a privacidade não ser uma questão tão recente, a partir dos avanços no âmbito do *Big Data Analytics*, poderíamos afirmar que passa a ser uma das grandes preocupações não só de governos, mas também das organizações públicas, privadas e do terceiro setor.

Fazendo um paralelo com a história do mito de guerra mais famoso da Antiguidade, o Cavalo de Troia, poderíamos comparar o *big data* como sendo o cavalo de Troia da privacidade. Muitos elementos já foram considerados em diversas esferas: técnica, legal, comportamental, dentre outras; mas aquele com maior potencial de impacto e, por enquanto, menor controle ou até mesmo conhecimento, está situado no âmago das análises avançadas de grandes massas de dados.

O construto obtido como resultado dessa pesquisa, contém um conjunto ótimo de elementos, que levou em consideração as pesquisas e debates mais recentes na esfera acadêmica. É um campo de pesquisa em exponencial crescimento que deve ser atentamente acompanhado pois espera-se que traga soluções de alto impacto para a problemática desenvolvida nessa pesquisa.

Esse esforço esteve baseado num levantamento exaustivo com o objetivo de identificar os principais elementos necessários para a constituição teórica de um arcabouço, técnico, metodológico, comportamental para fazer frente às questões trazidas pelo *Big Data Analytics* no âmbito da privacidade. Portanto, com a apresentação de um modelo teórico multifacetado de garantia da privacidade em análises de big data, o “Pentágono da Privacidade no *Big Data Analytics*”, entendemos ter constituído uma resposta concisa à questão colocada como ponto de partida desse trabalho: como garantir a privacidade e o valor em análises de big data,

Um próximo passo natural seria a submissão do construto aqui apresentado para ser testado empiricamente por meio de uma pesquisa quantitativa junto às organizações públicas e privadas no Brasil, num primeiro ciclo.

REFERÊNCIAS

- ABOUELMEHDI, Karim; BENI-HESSANE, Abderrahim; KHALOUFI, Hayat. *Big healthcare data: preserving security and privacy*. **Journal of Big Data**, v. 5, n. 1, p. 1-18, Jan. 2018. DOI: <https://doi.org/10.1186/s40537-017-0110-7>. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0110-7>. Acesso em: 11 set. 2020.
- ACQUISTI, Alessandro; TAYLOR, Curtis; WAGMAN, Liad. The economics of privacy. **Journal of Economic Literature**, v. 54, n. 2, p. 442-492, 2016. DOI: <http://dx.doi.org/10.1257/jel.54.2.442>. Disponível em: <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jel.54.2.442>. Acesso em: 9 set. 2020.
- ADAMS, Mackenzie. Big Data and Individual Privacy in the Age of the Internet of Things. **Technology Innovation Management Review**, v. 7, n. 4, p. 1-24, Apr. 2017. Disponível em: <https://timreview.ca/article/1067>. Acesso em: 31 ago. 2020.
- BRASIL. Lei Nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm. Acesso em: 3 set. 2020.
- BRASIL. Ministério da Educação. **Portal de Periódicos, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes)**. 2023. Disponível em: www.periodicos.capes.gov.br. Acesso em: 31 ago. 2020.
- BRYMAN, Alan. **Research Methods and Organization Studies**. London: Routledge, 1989. (Contemporary Social Research, 20).
- CHANSON, Mathieu; BOGNER, Andreas; BILGERI, Dominik; FLEISCH, Elgar. Blockchain for the IoT: Privacy-Preserving Protection of Sensor Data. **Journal of the Association for Information Systems**, v. 20, n. 9, p. 1274-1309, Mar. 2019. DOI: <http://dx.doi.org/10.17705/1jais.00567>. Disponível em: <https://aisel.aisnet.org/jais/vol20/iss9/10/>. Acesso em: 8 set. 2020.
- COX, Michael; ELLSWORTH, David. **Application Controlled Demand Paging for Out-of-Core Visualization**. NASA, Ames Research Center, 1997. Disponível em: <https://www.nas.nasa.gov/assets/pdf/.../1997/nas-97-010.pdf>. Acesso em: 10 jul. 2019.
- CRESWELL, John W. **Projeto de Pesquisa: métodos qualitativo, quantitativo e misto**. Porto Alegre: Artmed, 2010.
- EL OUAZZANI, Zakariae; EL BAKKALI, Hanan. A classification of non-cryptographic anonymization techniques ensuring privacy in big data. **International Journal of Communication Networks and Information Security (IJCNIS)**, v. 12, n. 1, p. 142-152, April 2020. Disponível em: https://www.researchgate.net/publication/342009466_A_Classification_of_non-

Cryptographic_Anonymization_Techniques_Ensuring_Privacy_in_Big_Data. Acesso em: 2 set. 2020.

HADAR, Irit; HASSON, Tomer; AYALON, Oshrat; TOCH, Eran; BIRNHACK, Michael; SHERMAN, Sofia; BALISSA, Arod. Privacy by designers: software developer's privacy mindset. **Empirical Software Engineering**, v. 23, n. 1, p. 259-289, Feb. 2018. DOI: <https://doi.org/10.1007/s10664-017-9517-1>. Disponível em: <https://link.springer.com/article/10.1007/s10664-017-9517-1>. Acesso em: 31 ago. 2020.

HAIR, Joseph F. Jr.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E.; TATHAM, Ronald L. **Análise multivariada de dados**. Tradução: Adonai Schlup Sant'Anna. 6 ed. Porto Alegre: Bookman, 2009.

ISHII, Kaori. Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects. **AI & Society**, v. 34, p. 509-533, Aug. 2019. DOI: <https://doi.org/10.1007/s00146-017-0758-8>. Disponível em: <https://link.springer.com/article/10.1007/s00146-017-0758-8>. Acesso em: 1 set. 2020.

JAIN, Priyank; GYANCHANDANI, Manasi; KHARE, Nilav. Big data privacy: a technological perspective and review. **Journal of Big Data**, v. 3, n. 1, p. 1-25, 2016. DOI: <https://doi.org/10.1186/s40537-016-0059-y>. Disponível em: <https://journalofbigdata.springeropen.com/counter/pdf/10.1186/s40537-016-0059-y.pdf> . Acesso em: 10 set. 2020.

JAIN, Priyank; GYANCHANDANI, Manasi; KHARE, Nilay. Enhanced Secured Map Reduce layer for Big Data privacy and security. **Journal of Big Data**, v. 6, n. 60, p. 1-17, Mar. 2019. DOI: <https://doi.org/10.1186/s40537-019-0193-4>. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0193-4>. Acesso em: 1 set. 2020.

MILLS, Kathy A. What are the threats and R potentials of big data for qualitative research? **Qualitative Research**, v. 18, n. 6, p. 591-603, 2018. DOI: <https://doi.org/10.1177/1468794117743465>. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/1468794117743465>. Acesso em: 4 set. 2020.

NORONHA, Daisy Pires; FERREIRA, Sueli Mara S. P. Revisões de literatura. In: CAMPELLO, Bernadete Santos; CONDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (org.) **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000.

OHM, Paul. Broken promises of privacy: responding to the surprising failure of anonymization. **UCLA Law Review**, v. 57, p. 1702-1777, Aug. 2010. Disponível em: <https://www.uclalawreview.org/pdf/57-6-3.pdf>. Acesso em: 4 set. 2020.

OOSTVEEN, Manon. Identifiability and the applicability of data protection to big data. **International Data Privacy Law**, v. 6, n. 4, p. 299-309, Nov. 2016. DOI: <https://doi.org/10.1093/idpl/ipw012>. Disponível em: <https://academic.oup.com/idpl/article-abstract/6/4/299/2525426?login=false>. Acesso em: 1 set. 2020.



POLITOU, Eugenia; ALEPIS, Efthimios; PAYSAKIS, Constantinos. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. **Journal of Cybersecurity**, v. 4, n. 1, p. 1–20, 2018. DOI: <https://doi.org/10.1093/cybsec/tyy001>. Disponível em: <https://academic.oup.com/cybersecurity/article/4/1/tyy001/4954056>. Acesso em: 10 set. 2020.

PRAGASH, K.; JAYABHARATHY, J. A Survey on Big Data Privacy and Security Issues in Healthcare Information System. **Advanced in Natural and Applied Sciences**, v. 11, n. 12, p. 95-99, Oct. 2017. Disponível em: <https://www.aensiweb.net/AENSIWEB/anas/anas/2017/October/95-99.pdf>. Acesso em: 3 set. 2020.

RAO, P. Ram Mohan; KRISHNA, S. Murali; KUMAR, A. P. Silva. Privacy preservation techniques in big data analytics: a survey. **Jornal of big data**, v. 5, n. 22, p. 1-12, Mar. 2018. DOI: <https://doi.org/10.1186/s40537-018-0141-8>. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0141-8>. Acesso em: 8 set. 2020.

SARKAR, Bikash Kanti. Big data for secure healthcare system: a conceptual design. **Complex Intell. Syst.**, v. 3, p. 133-151, June 2017. DOI: <https://doi.org/10.1007/s40747-017-0040-1>. Disponível em: <https://link.springer.com/article/10.1007/s40747-017-0040-1#citeas>. Acesso em: 2 set. 2020.

SOLOVE, Daniel J. Conceptualizing Privacy. **California Law Review**, v. 90, n. 4, p. 1087-1155, July 2002. Disponível em: https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2086&context=faculty_publications. Acesso em: 17 dez. 2023.

SUN, Zhaohao; PAMBEL, Francisca; STRANG, Kenneth David. Privacy and security in the big data paradigm. **Journal of Computer Information Systems**, v. 60, n. 3, p. 1-10, Feb. 2018. DOI: <https://doi.org/10.1080/08874417.2017.1418631>. Disponível em: https://www.researchgate.net/publication/323056850_Privacy_and_security_in_the_big_data_paradigm. Acesso em: 2 set. 2020.

SUN, Zhaohao; SUN, Lee Lizhe; STRANG, Kenneth. Big data analytics services for enhancing business intelligence. **Journal of Computer Information Systems**, v. 58, n. 2, p. 162-168, 2016. DOI: <http://dx.doi.org/10.1080/08874417.2016.1220239>. Disponível em: https://www.researchgate.net/publication/309389413_Big_Data_Analytics_Services_for_Enhancing_Business_Intelligence. Acesso em: 1 set. 2020.

THOMSON, Lucy L.; THIBADEAU, Robert. American Bar Association. Security challenges of the big data ecosystem require a laser- like focus on risk. **The SciTech Lawyer**, v. 12, n. 2, Jan. 2016.

WARREN, Samuel D.; BRANDEIS, Louis D. The Right to Privacy. **Harvard Law Review**, v. 4, n. 5, p. 193-220, 1890. Acesso em: 29 set. 2020.

WESTIN, Alan. **Privacy and Freedom**. Nova York: Atheneum, 1970.



WILSON, Rebecca J.; BELLIVEAU, Kiley M.; GRAY, Leigh Ellen. Busting the Black Box: Big Data, Employment and Privacy. **Defense Counsel Journal**, v. 84, n. 3, p. 1-34, July 2017. Disponível em: <https://www.iadclaw.org/defensecounseljournal/busting-the-black-box-big-data-employment-and-privacy/>. Acesso em: 11 set. 2020.

Declaração de Contribuição dos Autores

Brenner Lopes – Conceptualização – Curadoria dos Dados – Análise Formal – Investigação – Metodologia – Administração do Projeto – Recursos – Supervisão – Validação – Visualização – Escrita (rascunho original) – Escrita (análise e edição).

Ricardo Rodrigues Barbosa – Conceptualização – Análise Formal – Metodologia – Administração do Projeto – Supervisão – Validação – Escrita (análise e edição).

Luander Cipriano de Jesus Falcão – Metodologia – Software - Validação – Visualização – Escrita (análise e edição).

Renato Rocha Souza – Validação – Visualização – Escrita (análise e edição).

Como citar o artigo:

LOPES, Brenner; BARBOSA, Ricardo Rodrigues; FALCÃO, Luander Cipriano de Jesus; SOUZA, Renato Rocha. Pentágono da Privacidade no Big Data Analytics: proposta de modelo teórico. **Revista Informação na Sociedade Contemporânea**, Natal, v. 8, p. e33898, 2024. DOI: <https://doi.org/10.21680/2447-0198.2024v8n1ID33898>.