

**UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

GISLAINE CRISTINA DA CONCEIÇÃO

***BIG DATA* E DADOS PESSOAIS: QUESTÕES LEGAIS EMERGENTES NA
PERSPECTIVA DA CIÊNCIA DA INFORMAÇÃO**

**SÃO CARLOS/SP
2023**

GISLAINE CRISTINA DA CONCEIÇÃO

***BIG DATA E DADOS PESSOAIS: QUESTÕES LEGAIS EMERGENTES NA
PERSPECTIVA DA CIÊNCIA DA INFORMAÇÃO***

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Informação a Universidade Federal de São Carlos (UFSCar), como parte das exigências para a obtenção do título de Mestre em Ciência da Informação.

Área: Conhecimento, Tecnologia e Inovação

Linha: Tecnologia, Informação e Representação.

Orientador: Prof. Dr. Januário Albino Nhacuongue

**SÃO CARLOS/SP
2023**




UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Gislaíne Cristina da Conceição, realizada em 14/06/2023.

Comissão Julgadora:

Documento assinado digitalmente
 JANUARIO ALBINO NHACUONGUE
Data: 29/07/2023 13:47:39-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Januario Albino Nhacuongue (UFSCar)
Profa. Dra. Ana Carolina Simionato Arakaki (UFSCar)
Prof. Dr. Ricardo César Gonçalves Sant'Ana (UNESP)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Informação.

AGRADECIMENTOS

A Deus por ter me dado saúde e força e por me ajudar a ultrapassar todos os obstáculos encontrados durante a vida e ao longo do curso.

Aos meus pais Amarildo e Ivanir, pelo amor, incentivo e apoio incondicional. Obrigada por nunca me deixarem desistir e por estarem presentes em todos os momentos importantes da minha vida.

Ao meu marido Flávio que me incentivou desde a inscrição até o momento da defesa, compreendeu a minha ausência como esposa e me amou incondicionalmente durante os dias mais cinzentos destes dois anos e meio.

A minha irmã Gisele e meu sobrinho Heitor por me incentivarem a dar um passo por vez e compreenderem a minha ausência enquanto eu me dedicava à realização deste trabalho.

A minha amiga Janaína Goulart, com quem convivi intensamente durante os últimos anos, pelo companheirismo e pela troca de experiências que me permitiram crescer não só como pessoa, mas também como profissional.

A minha amiga Vanessa Custodio, obrigada por todos os conselhos, bem como palavras motivacionais, puxões de orelha e apoio incondicional. Gratidão por todo seu carinho e pelas risadas que compartilhamos durante esse momento da minha vida.

Ao meu orientador Prof. Dr. Januário Albino Nhacuongue pela orientação, apoio e confiança.

A Prof.^a. Dra. Ana Carolina Simionato Arakaki que sempre me apoiou, me incentivou e nunca desistiu de mim. Minha eterna gratidão!

Aos membros da banca, pela dedicação de seu tempo, leitura da dissertação e troca de experiências.

A todos que direta ou indiretamente fizeram parte de minha formação, o meu muito obrigado.

“Não fique preso a dogmas - que é viver com o resultado do pensamento de outras pessoas. Não deixe o barulho da opinião dos outros abafar a sua voz interior. E mais importante, tenha a coragem para seguir seu coração e sua intuição. Eles de alguma forma já sabem o que você realmente quer se tornar.”

Steve Jobs

RESUMO

Esta pesquisa tem por objetivo analisar as discussões acerca de aspectos legais no uso de dados na Ciência da Informação, no Brasil, nos últimos 10 anos. Desse modo, procura responder à seguinte questão norteadora: Qual vem sendo a atenção dedicada aos aspectos legais e de privacidade no uso de dados pessoais nas pesquisas científicas dos últimos dez anos da Ciência da Informação, no Brasil? O quadro teórico de referência foi estabelecido a partir dos conceitos: *Big Data*; Ciclo de Vida de Dados; Dados; Privacidade e Marco Regulatório Brasileiro. Quanto à metodologia, foi feita uma análise de domínio dos materiais bibliográficos referentes a dados e privacidade na Ciência da Informação, no Brasil, contemplando o período de janeiro de 2012 até dezembro de 2022. Neste estudo, a análise do domínio se concentrou na abordagem 5) Estudos históricos, conforme a classificação de Smiraglia (2015), e na abordagem 6) Estudos históricos de Hjørland (2002). A partir da análise dos trabalhos apresentados nas Bases de Dados BRAPCI, *Scielo*, BDTD, ISKO-Brasil e ENANCIB na forma de Comunicações Orais e de Pôsteres, ao longo da trajetória dos últimos dez anos. Do total de publicações recuperadas buscou-se identificar aqueles que tiveram os termos privacidade; dados pessoais; lei de acesso à informação e lei geral de proteção de dados pessoais, no título, resumo e/ou palavras-chave. Dos resultados obtidos das Bases de Dados 173 artigos estavam de acordo com os critérios estabelecidos para fazer parte da análise da pesquisa. Os trabalhos foram categorizados e a geração de gráficos ocorreu com informações de distribuição temporal de autores e suas filiações institucionais; área de aplicação do estudo de acordo com o Ciclo de Vida dos Dados em CI; Em seguida, foi feita uma análise histórica de tendências das discussões tratadas pelos dez autores mais colaborativos com o CVD-CI nos últimos dez anos. A análise realizada revelou que aproximadamente 75% dos trabalhos das Bases de Dados estão concentrados na etapa Coleta de dados do Ciclo de Vida (CVD-CI). Nota-se a carência de estudos nas temáticas das fases do ciclo de vida dos dados como armazenamento, recuperação e descarte de dados pessoais. Esses temas são fundamentais para compreender os aspectos legais envolvidos no uso de dados pessoais e para orientar possíveis ações na área da Ciência da Informação. Acredita-se que essa pesquisa é um ponto de partida importante para pesquisas futuras, pois destaca a necessidade de uma análise crítica e reflexiva sobre os aspectos que envolvem o uso de dados e informações. Essa abordagem é fundamental para subsidiar ações estratégicas em diversos setores da sociedade na atualidade.

Palavras-chave: *Big Data* e Dados Pessoais. Ciclo de Vida de Dados. Lei Geral de Proteção de Dados (LGPD). Privacidade. Ciência da Informação.

ABSTRACT

This research aims to analyze the discussions about legal aspects in the use of data in Information Science, in Brazil, in the last 10 years. In this way, it seeks to answer the following guiding question: How much attention has been paid to legal and privacy aspects in the use of personal data in scientific research over the last ten years of Information Science in Brazil? The theoretical frame of reference was established from the concepts: Big Data; Data Lifecycle; Data; Privacy and Brazilian Regulatory Framework. As for methodology, a domain analysis of bibliographic materials related to data and privacy in Information Science in Brazil was carried out, covering the period from January 2012 to December 2022. In this study, the domain analysis focused on approach 5) Historical studies, according to the classification of Smiraglia (2015), and in the approach 6) Historical studies of Hjørland (2002). From the analysis of the works presented in the BRAPCI, Scielo, BDTD, ISKO-Brasil and ENANCIB Databases in the form of Oral Communications and Posters, over the course of the last ten years. Of the total number of publications retrieved, we sought to identify those that had the terms privacy; personal data; law of access to information and general law of protection of personal data, in the title, abstract and/or keywords. Of the results obtained from the Databases, 173 articles were in accordance with the criteria established to be part of the research analysis. The works were categorized and the generation of graphs occurred with information on the temporal distribution of authors and their institutional affiliations; study application area according to the Data Lifecycle in CI; Then, a historical trend analysis of the discussions handled by the ten most collaborative authors with the CVD-CI in the last ten years was performed. The analysis carried out revealed that approximately 75% of the work on the Databases is concentrated in the Life Cycle Data Collection (CVD-CI) stage. There is a lack of studies on the themes of the stages of the data life cycle such as storage, retrieval and disposal of personal data. These themes are fundamental to understand the legal aspects involved in the use of personal data and to guide possible actions in the field of Information Science. It is believed that this research is an important starting point for future research, as it highlights the need for a critical and reflective analysis of aspects involving the use of data and information. This approach is essential to support strategic actions in various sectors of society today.

Keywords: *Big Data* and Personal Data. Data Lifecycle. General Data Protection Law (LGPD). Privacy. Information Science.

LISTA DE FIGURAS

Figura 1 – Infográfico sobre o volume de informação gerada a cada minuto na Internet em 2021	25
Figura 3 - Ciclo de Vida dos Dados para Ciência da Informação – (CVD–CI).....	33
Figura 4 – Quantidade mundial de dados criados por ano [zettabytes]	42
Figura 5 – Ranking dos autores mais colaborativos com o CVD - CI	71

LISTA DE GRÁFICOS

Gráfico 1 - Resultados das Buscas nas Bases de Dados	58
Gráfico 2 - Quantidade de Artigos por Fase do CVD - Cl.....	58
Gráfico 3 - Afiliações dos Autores.....	61
Gráfico 4 - Publicações por Instituições.....	62
Gráfico 5 - Frequência anual de publicações	62
Gráfico 6 - Publicações por Autores Categoria Coleta	64
Gráfico 7 - Assuntos Tratados por Autores na Categoria Coleta	65

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Problemática	12
1.2 Justificativa.....	16
1.3 Objetivos	20
2 CIÊNCIA DA INFORMAÇÃO, BIG DATA E DADOS PESSOAIS	22
2.1 O Fenômeno <i>Big Data</i>	24
2.2 Ciclo de Vida de Dados	31
2.2.1 Fases do Ciclo de Vida dos Dados e fatores paralelos – CVD-CI	33
2.3 Dados e Privacidade	35
2.4 Marco regulatório brasileiro	45
3 PROCEDIMENTOS METODOLÓGICOS	53
4 RESULTADOS E DISCUSSÃO	56
5. CONCLUSÃO.....	75
REFERÊNCIAS.....	78

1 INTRODUÇÃO

Na sociedade da informação, as tecnologias adotadas nos últimos dez anos permitem a coleta e o tratamento rápido de um conjunto cada vez mais vasto de dados pessoais. Embora esse processo traga vantagens em diferentes contextos de utilização de dados, como tomada de decisões estratégicas em empresas, em algumas vezes, ocorre em concomitância com violações de privacidade. Pode-se dar como exemplo a forma como os dados são reunidos, pois eles são coletados de perfis em redes sociais, transações comerciais e monitoramento de navegação na *Internet* – muitas vezes sem o conhecimento do usuário – e utilizados para as mais diversas finalidades (LIMA, 2019).

À medida que as redes de computadores e o uso da *Internet* cresceram e se intensificaram, as novas formas de comunicação e os registros de informações tornaram-se possíveis, facilitando o acesso e a recuperação de dados e informações. Neste novo cenário, onde computadores e dispositivos tecnológicos são cada vez mais utilizados para ampliar as habilidades de comunicação humana, surge a preocupação com a privacidade, uma vez que há o uso indevido de informações e dados pessoais, sejam elas de exposição da vida pessoal na rede mundial de computadores ou do uso de instrumentos de segurança (GRISOTO; SANT'ANA; SANTARÉM SEGUNDO, 2015).

O desenvolvimento tecnológico ocorrido nas últimas décadas contribuiu para o aumento do fluxo de informações e melhorias na tecnologia da informação. Esse cenário informacional passou a ser marcado pelo risco crescente de exposição e uso indevido ou abusivo dos dados (LIMA, 2019).

O acesso a dados vem transformando todas as esferas da atividade humana, principalmente nos últimos anos, pois as alternativas para coletar, armazenar e recuperar dados cresceram exponencialmente, superando até mesmo nossa capacidade de lidar com esses novos volumes de dados, perspectivas, diversidade e velocidade de acesso. Este novo cenário está dando espaço para o *Big Data* (SANT'ANA, 2016).

Dumbill (2012, p.3) explica que *Big Data* são dados que excedem a capacidade de processamento de sistemas de bancos de dados convencionais. Os dados são muito volumosos, se movem muito rápido, ou não se encaixam nas estruturas das arquiteturas de banco de dados.

Outra possibilidade para conceituar *Big Data* é apresentada por Davis (2012, p. 04):

Big Data são dados muito volumosos para serem tratados e analisados por protocolos de banco de dados tradicionais como SQL (o que faz *Big Data* um termo que pode evoluir ao longo do tempo, o que é considerado agora *Big Data* pode muito rapidamente tornar-se pequeno). Neste sentido, o tamanho é apenas um aspecto dessas novas tecnologias.

Todos os dias, segundo a *International Business Machines* (IBM):

São criados 2,5 quintilhões de bytes de dados - tanto que 90% dos dados no mundo de hoje foram criados nos últimos dois anos. Estes dados vêm de todos os lugares: sensores usados para coletar informações sobre o clima, mensagens para sites de mídia social, fotos e vídeos digitais, registros de transação de compras e sinais de telefone celular GPS, para citar alguns. Estes dados são *Big Data*. (IBM, 2014).

O advento da informática possibilitou o crescimento do fluxo de informações, bem como o aprimoramento das técnicas de coleta, de processamento e de utilização da informação (DIAS; VIEIRA, 2015).

A informatização, conforme entende Doneda (2006), operou mudanças de ordem quantitativa e qualitativa no tratamento de dados: maior quantidade de dados passou a poder ser processada em menos tempo e resultados mais valiosos passaram a ser obtidos a partir do uso de novos métodos, algoritmos e técnicas de processamento.

Com o crescimento, às vezes, desordenado e em larga escala da produção de dados e informações, surge a preocupação com a privacidade em diferentes contextos da coleta e do tratamento desses dados. Essas preocupações se estendem ao seu uso indevido por agentes de tratamento ou operadores (públicos e privados).

Nas palavras de Rodotà (2008, p.18), o direito à privacidade é entendido como “[...] o direito de manter o controle sobre as próprias informações [...]”, que é construído no eixo “[...] pessoas-informação-circulação-controle [...]” ao invés de “[...] pessoa-informação-sigilo [...]”, em torno do qual foi construído o conceito clássico de privacidade.

A informatização, conforme Doneda (2006), operou mudanças de ordem quantitativa e qualitativa no tratamento de dados: maior quantidade de dados passou

a poder ser processada em menos tempo e resultados mais valiosos passaram a ser obtidos a partir do uso de novos métodos, algoritmos e técnicas de processamento.

Nessa perspectiva, o conceito de privacidade se transforma para dar origem à disciplina da proteção de dados pessoais, tendo em vista o crescimento da importância da informação, que com o avanço crescente adquire uma importância cada vez maior em diversos aspectos da sociedade. Ela se torna um recurso valioso para tomada de decisões estratégicas, desenvolvimento de negócios, inovação tecnológica, pesquisa científica, criação de políticas públicas eficazes e até mesmo para o bem-estar individual dos cidadãos. A informação confiável e precisa é fundamental para o progresso e o sucesso em muitos setores, impulsionando o desenvolvimento econômico, social e cultural de uma sociedade. Além disso, a capacidade de acessar, analisar e utilizar a transferência de informação pode ser um diferencial competitivo tanto para empresas quanto para indivíduos.

Esta pesquisa insere-se nas discussões acerca do fenômeno *big data* e dados pessoais, com o objetivo de analisar as discussões acerca de aspectos legais no uso de dados na Ciência da Informação, no Brasil, nos últimos dez anos, utilizando o modelo de ciclo de vida de dados. O problema norteador da nossa discussão será apresentado na próxima seção.

1.1 Problemática

A Ciência da Informação (CI) estuda a prática de criação, gerenciamento e disseminação da informação para recuperação e uso, e está intimamente relacionada as novas mudanças que surgem com o aumento do fluxo de informações, principalmente por meio da criação de redes de computadores e da disseminação da *Internet*. A questão da privacidade surge como um novo desafio, principalmente no ambiente informacional digital, onde informações e dados são gerados em escala exponencial por meio do uso de dispositivos tecnológicos, sendo essencialmente importante para a CI discutir questões sobre esse tema (GRISOTO; SANT'ANA; SANTARÉM SEGUNDO, 2015).

Com o advento da infraestrutura digital, muitas transações econômicas e sociais estão sendo cada vez mais digitalizadas. Bancos *on-line*, redes sociais, redes inteligentes e redes industriais são exemplos de atividades que agora dependem fortemente de dados digitais (HUAWEY, 2021).

A quantidade e a velocidade com que os dados digitais são gerados e armazenados continuam aumentando exponencialmente. Embora a maioria dos dados seja trivial, uma quantidade significativa conterá informações confidenciais ou informações de identificação pessoal e precisa ser protegida. Isso requer defesas de segurança cibernética físicas e digitais claras para garantir que os dados permaneçam privados[...] à medida que as organizações adotam transações digitais, elas embarcam em uma jornada digital para transformar seu modelo de negócios. Em muitos casos, eles não têm especialistas internos para proteger dados de clientes e transações. Infraestruturas críticas, como abastecimento de água e eletricidade, também enfrentam riscos crescentes (HUAWEY, 2021, p.05).

Do ponto de vista da privacidade, a crescente digitalização de transações comerciais e sociais está criando oportunidades para criminosos cibernéticos violarem a integridade de dados das organizações, por meio da exploração de informações pessoais ou críticas para fraude, espionagem e sabotagem. Embora os regulamentos sobre proteção de dados devam ser modernizados para se adaptarem às novas tecnologias, eles não devem impedir a inovação digital ou bloquear a implementação dessas novas tecnologias. Essa dicotomia entre inovação digital e proteção de dados alimenta o debate em torno da quantidade necessária de regulamentação no espaço cibernético (HUAWEY, 2021).

De acordo com Huawei (2021, p.7), o custo total mundial da criminalidade cibernética é atualmente estimado em 525 bilhões de dólares, representando cerca de 0,7% do PIB mundial, e espera-se que cresça se as medidas apropriadas não forem tomadas.

O conceito de privacidade, enquanto direito ao autocontrole, à autonomia e à integridade da pessoa, também envolve a posse e o uso indevido de dados por entidades legitimamente qualificadas. Segundo Garfinkel (2000), as ameaças à privacidade estão intimamente relacionadas a dois fatores principais. Por um lado, aos avanços da tecnologia¹, como câmeras de monitoramento e computação ubíqua. Por outro lado, às estratégias dos governos, por exemplo, de restringir o uso de tecnologia de criptografia, para manter o poder de monitoramento e controle dos cidadãos.

Uma das principais preocupações é a violação da privacidade do indivíduo. Quando os detentores de dados coletam informações pessoais, como nome, endereço, número de telefone, hábitos de consumo, políticas comportamentais ou até

¹ Ao considerar a tecnologia uma das causas da ameaça à privacidade, Garfinkel (2000) alerta que ela não pode ser analisada de modo isolado, na medida em que existe na junção entre ciência, mercado e sociedade, isto é, ela é socialmente construída para preencher necessidades específicas.

os mesmos dados biométricos, eles têm a responsabilidade de proteger essas informações e usá-las apenas para os propósitos especificados e consentidos pelo titular dos dados. No entanto, em alguns casos, os dados são utilizados de maneira caseira, extrapolando os propósitos originais ou sendo compartilhados sem o devido consentimento (MAYER-SCHONENBERGER; CUKIER, 2013, p. 124-128).

Órgãos governamentais podem estar envolvidos em programas de vigilância em massa, nos quais coletam e monitoram informações de cidadãos em larga escala, muitas vezes sem o conhecimento ou consentimento desses indivíduos. Embora a justificativa para tais programas seja a segurança nacional, eles podem representar uma invasão significativa da privacidade e dos direitos individuais (FORNASIER; KNEBEL, 2021).

No caso das empresas, a utilização ilimitada de dados pode ocorrer de várias maneiras. Alguns exemplos incluem a venda de informações pessoais para terceiros sem o consentimento dos usuários, o rastreamento excessivo das atividades online das pessoas, o direcionamento de publicidade de maneira intrusiva e a análise de dados para consumo de consumidores ou práticas injustas de precificação (FORNASIER; KNEBEL, 2021).

Essas práticas irrevogáveis têm consequências provocadas para a sociedade. Elas minam a confiança dos indivíduos nos detentores de dados e podem levar a abusos, discriminação e injustiça. Além disso, a exposição ilimitada de informações pessoais pode levar ao roubo de identidade, fraude financeira e outras formas de crimes cibernéticos (MAYER-SCHONENBERGER; CUKIER, 2013, p. 124-128).

A natureza controversa do consentimento como afirmação dos direitos relacionados aos dados digitais surge devido à tentativa de estabelecer a liberdade e a autonomia pessoal em um contexto marcado pela desigualdade na gestão de dados. Essa desigualdade é evidente devido à disparidade na infraestrutura e no conhecimento relacionado à ciência e interpretação de dados em larga escala na era do *Big Data*, especialmente em relação aos avanços recentes em aprendizado de máquina e inteligência artificial. Diante disso, questiona-se a viabilidade do consentimento em relação a formas de análise e uso de dados que estão além da capacidade humana de processamento (MAYER-SCHONENBERGER; CUKIER, 2013, p. 124-128).

A ideia de cidadania digital também implica a existência de um sujeito digitalizado com direitos relacionados aos dados digitais. A condição de sujeito de

direito é um paradigma que surge e está intrinsecamente ligado ao capitalista, permaneceu como uma forma jurídica que media as trocas de mercadoria com base em relações completamente voluntárias e na expressão livre da vontade. Essa forma jurídica é essencial para a existência do capitalismo. No entanto, essa característica puramente formal, que considera as relações sociais de maneira abstrata, negligência a diversidade real entre os indivíduos e as complexidades das relações sociais (FORNASIER; KNEBEL, 2021).

Na evolução em direção a um capitalismo de vigilância, ocorre uma mudança na natureza da mercadoria oferecida pelos indivíduos, que não se limita mais à força de trabalho. Agora, há uma abstração do comportamento humano, envolvendo-o em uma mercadoria composta por dados comportamentais. Assim como o sujeito de direito fornece as condições jurídicas necessárias para a exploração da mais-valia do trabalho, de forma análoga, essa mesma forma jurídica é empregada para mediar a troca de dados, permitindo a segurança de uma mais-valia comportamental. Essa proteção é um processo essencial na nova economia do capitalismo de vigilância. Tais transformam-se são intrínsecas ao processo de reprodução social do capitalismo, pois cada grande fase desse sistema se baseia em um método específico de buscar de mais-valia e obtenção de lucro (MASCARO, 2013). Dessa forma, a transmissão de um novo tipo de excedente produzido pelas pessoas é uma característica fundamental da dinâmica do capital.

Da mesma forma que o sujeito de direito exerce o papel de mediador e carrega a ideologia da igualdade jurídica na relação capitalista de trabalho, através da livre vontade expressa em contratos, o processo de garantia de mais-valia comportamental também utiliza essa mediação no contexto da Lei Geral de Proteção de Dados (LGPD)². A LGPD consagra o instituto do consentimento como mecanismo de proteção e como condição essencial para a reprodução social do novo mercado de dados, que é justamente ganho por meio do processo de garantia de mais-valia descrito por Zuboff (2019, p. 97): “[...] os usuários fornecem seus dados por meio do consentimento e, como resultado, são transformados e analisados pelas empresas (que os comercializam no mercado de previsão de comportamentos)”.

Considerando o atual dilema entre a necessidade de segurança e a consequente perda de privacidade com o compartilhamento de dados pessoais, no

² Lei Geral de Proteção de Dados (LGPD) é a Lei nº 13.709/2018, que regula as atividades de tratamento de dados pessoais e que também altera os artigos 7º e 16 do Marco Civil da *Internet*.

momento em que grande parte da população mundial se torna cada vez mais dependente das novas tecnologias de informação e comunicação (LOTT; CIANCONI, 2018), será realizada uma pesquisa com o intuito de responder à seguinte questão: Qual vem sendo a atenção dedicada aos aspectos legais e de privacidade no uso de dados pessoais nas pesquisas científicas dos últimos dez anos da Ciência da Informação, no Brasil?

1.2 Justificativa

O aumento do volume de dados mudou a forma como a pesquisa, a governança, a socialização e o comércio são conduzidos. Podem ser produzidos por diversos setores da sociedade, governo, indústria, institutos de pesquisa e universidades. Eles são coletados, armazenados e compartilhados de diferentes maneiras por diferentes organizações. (EUROPEAN DATA SCIENCE ACADEMY, 2019).

A demanda por acesso e análise de dados privados e públicos tem aumentado, ganhando atenção de diversos segmentos da sociedade. A situação atual mostra que as empresas lucram, os governos influenciam as decisões relacionadas aos seus cidadãos e as pessoas são induzidas a aceitar conceitos, agravados pelo fato de que os dados que financiam essas ações nem sempre estão envolvidos de princípios de ética e justiça (KEMPER; KOLKMAN, 2018).

Dessa forma, a Sociedade da Informação traz uma verdadeira mudança no cenário econômico, político e social, em consonância com a disseminação acelerada das TIC, que facilitam a transformação das cadeias produtivas, com o uso massivo da tecnologia proporcionando melhorias e eficiências no processo de produção, melhores condições de uso de capital, matérias-primas e capital humano. Assim, na sociedade da informação, a tecnologia é um meio de impulsionar a mudança, por meio dos dispositivos tecnológicos e do acesso à *Internet*, alterando a forma de pensar a informação (GRISOTO; SANT'ANA; SANTARÉM SEGUNDO, 2015). A computação levou a uma mudança de paradigma onde a comunicação, ou seja, a transmissão e recepção de dados e conteúdo (imagens, músicas, livros etc.) passam a ter um leque de possibilidades e aplicações para os indivíduos.

A utilização das TIC e a combinação de várias ferramentas tecnológicas permitem a coleta, registro, processamento, cruzamento, organização e transmissão de dados, permitindo assim obter informação sobre os indivíduos, cujo valor é

agregado a possibilidade de tomar decisões econômicas, políticas e sociais (MENDES, 2008).

Mendes (2008, p. 17-18) discorre que a:

A evolução do direito à privacidade prosseguiu para se adaptar às novas transformações sociais ocasionadas pela revolução da tecnologia da informação que possibilitou a coleta e o processamento dos dados pessoais dos cidadãos de forma pioneira. Além de adquirir um caráter positivo e de ser reconhecido no âmbito internacional, o direito à privacidade transformou-se para ensejar o nascimento da disciplina de proteção de dados pessoais, à medida que surgiram novos desafios ao ordenamento jurídico a partir do tratamento informatizado dos dados.

A Lei Geral de Proteção de Dados (LGPD) trata de forma especial categorias de dados pessoais:

Art. 5º Para os fins desta Lei, considera-se:

I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

A confidencialidade não é tratada na LGPD, mas fornece as condições necessárias para que o tratamento dos dados seja realizado de forma segura e não exponha os dados do titular ao risco de exposição ou uso indevido. Outro aspecto muito importante é que a LGPD também garante o acesso à informação, porém na perspectiva do interesse individual, e não coletivo, diferenciando aqui o conceito de acesso, do de publicidade, pois nem todo acesso se dá através da disponibilidade pública da informação (OLIVEIRA; MOTTA; MELO; ESTEVES, p.16).

Com relação à privacidade, Davis (2012) explica que o *Big Data* está alterando o modo pelo qual entendemos tal palavra. Com o *Big Data* é possível fazer correlações sobre indivíduos gerando novos dados.

A reputação de uma pessoa está relacionada ao que as pessoas sabem e pensam sobre ela. O fenômeno do *Big Data* mudou algumas das características da reputação, principalmente no sentido exponencial, o número de pessoas que podem opinar sobre determinado sujeito, significando a perda do controle sobre a reputação virtual. Vale ressaltar que existem muitas empresas cujos modelos de negócios estão centrados na gestão da reputação (DAVIS, 2012).

A propriedade nos faz refletir sobre uma série de questões relacionadas à posse dos dados. Davis (2012, p.3) faz o seguinte questionamento: “Quem é o dono dos dados, os direitos a eles associados podem ser transferidos, e quais são as obrigações de pessoas que criam e utilizam esses dados?”. Davis (2012) finaliza questionando se o indivíduo deveria ter a legitimidade de controlar os próprios dados e em qual grau?

Levando em conta o período da pandemia, em que as pessoas estão mais em casa e aumentaram o uso de celular e computador nas compras *on-line*, o uso de serviços digitais e transações bancárias, o número de fraudes e golpes no ambiente digital disparou no Brasil. Segundo dados da FEBRABAN, em janeiro/fevereiro 2021 os ataques de *phishing*, a chamada pescaria digital, cresceu 100% em relação ao ano de 2020, enquanto os golpes da falsa central telefônica e falso funcionário de banco tiveram crescimento ainda maior, de 340%.

Diante desse contexto, o Observatório Febraban de junho de 2021 buscou investigar a visão da sociedade sobre a segurança de dados no Brasil e os crimes envolvendo violação de informações pessoais. O levantamento mostrou que os brasileiros estão atentos a essa questão, sobretudo quanto o uso que as empresas privadas fazem dos seus dados pessoais. Em alguns casos consideram a utilização dos dados aceitável, como por exemplo para a prevenção de crimes, de golpes em compras e fraudes com dados bancários. Outra constatação do estudo é que as pessoas passaram a adotar ou intensificar hábitos de proteção em relação a senhas, fornecimento de dados, aceitação de *cookies*, uso de cartão de crédito, entre outros. Mas ainda assim muitos se tornaram vítimas de crimes dessa natureza. Entretanto, apesar das notícias e experiências negativas a respeito, predomina uma expectativa de avanço na segurança dos dados nos próximos cinco anos (FEBRABAN, 2021).

A maioria dos brasileiros acredita que a privacidade nos meios eletrônicos se tornou um mito, com todas ou quase todas as informações acessíveis. É por isso que há uma grande demanda por eficiência e endurecimento da legislação de proteção de dados. Ao mesmo tempo, o cenário é otimista quanto à contribuição da LGPD e da Lei 14.155³ para a redução da criminalidade nos meios eletrônicos que segundo os

³ A Lei nº 14.155, de 27 de maio de 2021, altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tornar mais graves os crimes de violação de dispositivo informático, furto e estelionato cometidos de forma eletrônica ou pela *Internet*; e o Decreto-Lei nº 3.689, de 3 de outubro de 1941 (Código de Processo Penal), para definir a competência em modalidades de estelionato.

entrevistados, está relacionada principalmente à violação de dados de compras *on-line* e visitas a sites em geral (CAMPOS, 2021).

A transparência dos termos de uso e políticas de privacidade dos serviços digitais envolvendo material de natureza pessoal e sensível tem sido questionada, e há vários debates públicos em andamento para exigir que a lei garanta efetivamente a privacidade como um dos maiores direitos dos indivíduos. Nesse caso, grupos opostos se formaram - exigindo segurança e privacidade de dados ou apoiando políticas de compartilhamento de informações pessoais no interesse da segurança pública (LOTT; CIANCONI, 2018).

Diante deste cenário apresentado, o desenvolvimento da temática abordada nesta pesquisa se faz importante, pois o aproveitamento indevido de informações pessoais vem se tornando cada vez mais frequente, o que traz à tona discussões sobre os critérios de segurança e privacidade dos bancos de dados e a transparência das políticas de privacidade que devem assegurar aos usuários a propriedade sobre seus dados.

Sendo a CI, uma ciência interdisciplinar e social da Sociedade da Informação, preocupada em esclarecer o problema social concreto da informação no ciclo de construção, comunicação e uso (LE COADIC, 1996), os aspectos levantados justificam a relevância desta pesquisa, na medida em que oferecem novas direções de abordagem dentro do paradigma social do campo. Segundo Capurro (2003), ao contrário dos paradigmas físico e cognitivo, o paradigma social da CI não separa a informação dos condicionamentos sociais e materiais dos sujeitos.

Embora não esteja no escopo desta pesquisa, à medida que o volume e a variedade de dados aumentam, torna-se crucial refletir sobre questões éticas inerentes à sua coleta, ao seu armazenamento, ao seu processamento e ao seu compartilhamento.

O *Big Data* permite às empresas e organizações coletar, armazenar e analisar grandes dados de diversas fontes e em diversos formatos, incluindo dados não coletados. Isso ajuda a tomar decisões mais marcantes e identificar padrões e tendências que não eram visíveis em uma análise de dados tradicionais. Além disso, o *Big Data* permite que as empresas personalizem seus produtos e serviços de acordo com as necessidades e motivadas dos clientes, conseguindo assim a satisfação do cliente e aumentando a fidelidade. Também é útil em áreas como saúde, finanças e marketing, onde uma análise de grandes volumes de dados pode levar a descobertas

importantes e valiosas. No geral, o *Big Data* é uma ferramenta valiosa para a tomada de decisões em diversas áreas e setores. Nesse contexto, a privacidade se torna uma preocupação crucial.

Os avanços tecnológicos facilitam a coleta massiva de dados pessoais, levantando questões sobre como essas informações são utilizadas, compartilhadas e protegidas. É essencial garantir que as políticas de privacidade sejam claras, que os indivíduos tenham controle sobre seus dados e que sejam toleradas medidas de segurança adequadas para proteger a informação sensível.

Em suma, o direito à privacidade ganhou novos contornos de essencialidade na era da informação, à medida que aproveitamos oportunidades e enfrentamos desafios trazidos pelo crescente uso e disponibilidade de dados. A busca por um equilíbrio entre o uso eficiente da informação e a proteção dos direitos individuais é fundamental para o avanço tecnológico e o bem-estar da sociedade como um todo. A presente pesquisa irá abordar questões sobre dados, privacidade, ciclo de vida de dados, marco regulatório brasileiro e *Big Data*, mas sem adentrar em aspectos práticos de *profiling* ou *data mining*.

1.3 Objetivos

O objetivo geral desta pesquisa é analisar os estudos sobre dados e privacidade na Ciência da Informação, no Brasil, nos últimos dez anos, utilizando a análise de domínio e o modelo de ciclo de vida de dados. Para tanto, foram delineados os seguintes objetivos específicos:

- Apresentar o quadro teórico sobre dados e privacidade, destacando aspectos do marco regulatório brasileiro;
- Teoria de sustentação
- Caracterizar os estudos sobre dados e privacidade na Ciência da Informação quanto aos autores, suas filiações institucionais assuntos tratados e respectivas teorias de sustentação; na Ciência da Informação no país.

1.4 Estrutura do trabalho

O trabalho será dividido em cinco capítulos, sendo o primeiro uma introdução, contendo uma explanação inicial dos conceitos apresentados no texto, além da inserção da problemática, dos objetivos e da justificativa.

O segundo capítulo apresenta o referencial teórico do trabalho, delimitando os conceitos e as variáveis que embasam esta proposta. Nele serão tratados os temas: fenômeno *Big Data*, ciclo de vida dos dados, dado e privacidade, marco regulatório brasileiro.

O terceiro capítulo aborda os procedimentos metodológicos da pesquisa.

Os resultados e discussões da pesquisa são apresentados no quarto capítulo.

Por fim, o último capítulo apresenta as considerações finais, seguidas das referências bibliográficas.

2 CIÊNCIA DA INFORMAÇÃO, BIG DATA E DADOS PESSOAIS

A Ciência da Informação é um campo interdisciplinar que estuda as teorias, os métodos e os processos de coleta, organização, armazenamento, recuperação, uso e disseminação da informação em diferentes formatos e contextos. Ela busca entender como as informações são produzidas, processadas e compartilhadas por meio de tecnologias, sistemas e infraestruturas de informação, bem como investigar o impacto desse processamento em indivíduos, organizações e sociedade em geral.

De acordo com Santos e Sant'Ana (2015, p.199)

A Ciência da Informação refere-se à atividade direcionada à pesquisa de princípios e métodos que são partes da análise, do projeto e da evolução dos sistemas de informação. Nesses sistemas, os elementos constituintes são o ambiente, as pessoas, os recursos informacionais, as tecnologias e os procedimentos. Eles sustentam a capacidade para a busca de soluções e tomada de decisões como parte da vida diária, envolvendo a manipulação de dados, o acesso à informação e a apropriação do conhecimento.

A Ciência da Informação (CI), reconhecida por Saracevic (1996) como uma ciência interdisciplinar, dedica-se à resolução de problemas relacionados a dados, informações e conhecimento, estabelecendo-se como uma ciência que gerencia e organiza o conhecimento.

A CI sempre enfrentou desafios devido à revolução tecnológica e ao consequente acúmulo de conhecimento. Um dos desafios recentes para essa área é o fenômeno conhecido como *Big Data*, caracterizado pelo imenso volume de dados que não pode ser processado por meio de sistemas ou controle convencional.

Burke (1992) argumenta que o registro de eventos, tanto para informação imediata quanto para recuperação futura, tem sido uma prática presente ao longo da evolução da humanidade. Esses eventos são observados desde os desenhos nas paredes das cavernas os hieróglifos até egípcios esculpidos em pedra, os textos em pergaminhos, a invenção do papel, a criação de livros, a organização de bibliotecas, a imprensa, a máquina de perfurar cartões e, finalmente, a era atual da informática e da Web. Com o advento da era digital, testemunhamos um aumento exponencial no ritmo de geração, armazenamento, processamento e transmissão de registros de dados pela humanidade.

Na era da sociedade conectada, a velocidade e a volatilidade das ideias, notícias e decisões assumem um papel fundamental. O pensamento, o trabalho, o

entretenimento e os relacionamentos passaram a ser predominantemente remotos e digitais. O mundo parece ter “encolhido”, conforme previsto por McLuhan (1969), gerado na criação de uma verdadeira Aldeia Global.

Passarelli, Junqueira e Angeluci (2014) observam que estamos constantemente gerando e consumindo grandes volumes de dados por meio de dispositivos tecnológicos, que atualmente incluem dispositivos móveis, aplicativos, redes sociais, computação em nuvem e inteligência artificial.

O problema era (e, basicamente, ainda é) “a tarefa massiva de tornar mais acessível, um acervo crescente de conhecimento”; Bush (1945) identificou o problema da explosão informacional – o irreprimível crescimento exponencial da informação e de seus registros, particularmente em ciência e tecnologia. A solução por ele proposta era a de usar as incipientes tecnologias de informação para combater o problema. E foi mais longe, propôs uma máquina chamada MEMEX, incorporando (em suas palavras) a capacidade de associar ideias, que duplicaria “os processos mentais artificialmente”. É bastante evidente a antecipação do nascimento da CI e, até mesmo, da inteligência artificial.

Cavalcanti (2014, p. 34) observa que o *Big Data* não é simplesmente uma tecnologia. Reforçando essa perspectiva, Saracevic (1996, 1999) e Taurion (2012b) ressaltam que podemos estabelecer um paralelo entre a explosão de informações que originou a Ciência da Informação (CI) e a que ocorre atualmente. Mais uma vez, os recursos existentes revelam-se insuficientes para lidar com o volume de dados gerados, o que demanda a busca por novas ferramentas e metodologias.

A Ciência da Informação abarca avanços tecnológicos e se baseia na interdisciplinaridade com outras disciplinas relacionadas à informação. Ela se destaca por suas características singulares, conforme mencionado por Galvão e Borges (2000, p. 48):

[...] a Ciência da Informação não é uma ciência clássica, nem uma ciência básica. É uma ciência recursiva que se vale dos conhecimentos já existentes nas ciências (normais), dos avanços tecnológicos e suas possibilidades, bem como se define segundo os nichos de oportunidade (demandas sociais).

Como resultado do avanço contínuo dos espaços de estudo da informação fortalecidos pela interdisciplinaridade, destaca-se cada vez mais a integração entre pessoas, informação e tecnologia.

Diante do fenômeno *Big Data*, é factível admitir que a CI esteja diante da mesma necessidade que determinou o seu surgimento: expandir a capacidade de

guardar e recuperar informações que extrapolam as capacidades humanas de processamento, conforme Saracevic (1996, p. 42).

Nesta seção serão tratados os fundamentos teóricos do trabalho, visando delimitar os conceitos que embasam esta proposta. Aqui, serão tratados os temas sobre dado, privacidade, *Big data*, vigilância e o marco regulatório brasileiro sobre as principais leis de proteção de dados.

2.1 O Fenômeno *Big Data*

Macedo (2013) descreve que, para compreender o fenômeno do *Big Data*, podemos observar inúmeras pessoas e empresas com seus dispositivos móveis, computadores e suas atividades, como compras, transações bancárias e fiscais, além de milhões de *e-mails*, opiniões em redes sociais e sites especializados, telefonemas e registros históricos de vida.

De acordo com Chang, Mishra e Lin (2015), é possível observar sensores e dispositivos de controle, câmeras, veículos e diversos outros dispositivos conectados, todos gerando dados que são transmitidos pela *Internet*. Consequentemente, muitas de nossas ações são convertidas em dados.

Com o avanço tecnológico e sofisticação de dispositivos eletrônicos (por exemplo, celulares, automóveis, câmeras de segurança etc.), as organizações passaram a capturar trilhões de bytes de informações geradas em transações com os seus consumidores e fornecedores.

A figura 1 ilustra o volume de informações provenientes de redes sociais, mecanismos de busca, sites de entretenimento, e-mails, sites de compras e prestação de serviços. É importante ressaltar que o autor não considerou a participação e contribuição do volume de informações geradas no ambiente acadêmico e científico. Lewis (2021) destaca o imenso volume de informações geradas a cada minuto na *Internet*, fornecendo uma contribuição para a compreensão do fenômeno conhecido como *Big Data*.

Figura 1 – Infográfico sobre o volume de informação gerada a cada minuto na *Internet* em 2021



Fonte: Lewis (2021)

O *Big Data* está presente em vários segmentos sociais, empresariais e/ou governamentais, bem como em aplicações cotidianas. A partir dos dados coletados pelos usuários, aplicações Web como Google Now, Netflix e Amazon podem fazer recomendações personalizadas e melhorá-las através do aprendizado do comportamento de cada usuário (LOPES; SILVA, 2021).

Lewis (2021) nos traz uma perspectiva de que em apenas sessenta segundos são produzidos:

- Amazon Echo e Google Home - 932 alertas emitidos pelos dispositivos.
- E-mails - 197.6 milhões de E-mails são enviados.
- Facebook - 1,4 milhões de postagens são visualizadas.
- Facebook Messenger e Whatsapp - 69 milhões de mensagens são enviadas.

- Google Play e App Store - 414.764 apps são baixados.
- Imgur - 3 milhões de imagens são visualizadas
- Instagram - 695.000 stories são compartilhados.
- LinkedIn - 9.132 conexões são estabelecidas.
- Netflix - 28.000 assinantes estão assistindo algum conteúdo.
- SMS - 21,1 milhões de mensagens enviadas.
- Snapchat - 3.4 milhões de stories são publicados.
- Tik Tok - 5 milhões de visualizações são recebidas.
- Tinder - 2 milhões de perfis são "avaliados".
- Twitch - 2 milhões de visualizações são recebidas.
- Twitter - 200.000 tweets são publicados.
- YouTube - o equivalente à 500 horas de conteúdo é publicado.

A pesquisa também indica que leva apenas um minuto para 1.6 milhões de dólares sejam gastos online (LEWIS,2021).

De acordo, com a Pesquisa Brasileira de Mídia 2015, a *Internet* no Brasil é usada regularmente por 48% dos brasileiros, e o uso de smartphones já ultrapassou o de microcomputadores como forma de acesso à *Internet*.

Segundo o 26º Relatório Anual de Tecnologia da Informação da Fundação Getúlio Vargas (Brasil, 2015), há mais de 152 milhões de computadores em uso e somados a eles há também 154 milhões de smartphones em uso para uma população de 206 milhões de habitantes, ou seja, um alcance de 3 dispositivos para cada 2 habitantes (REIS; SÁ, 2020, p.2).

Com o surgimento dos smartphones, o aparelho celular deixou de ser apenas um aparelho para ligação e tornou-se um computador portátil, onde seu usuário pode realizar ligações, tirar fotografias, navegar na *Internet*, usar GPS de localização. Este fato quebrou o paradigma de produção e disseminação de informação que antes pertencia apenas aos veículos de mídias. Atualmente, todos os indivíduos que possuem um smartphone deixam de ser apenas consumidores, para se tornarem potenciais produtores massivos de dados.

Dumbill (2012) enfatiza que a velocidade e a volatilidade dos dados e dos formatos impulsionam o surgimento de novas tecnologias de suporte, demonstrando que as mudanças ocorrem de forma extremamente rápida.

Segundo Ferrer-Sapena e Sánchez-Pérez (2013, p.150) o fenômeno do *Big Data* está diretamente relacionado à necessidade de acesso não apenas à informação, mas também aos dados em formato aberto e público. Não se trata somente de compartilhar informação e permitir livre acesso às publicações, mas, agora deseja-se que os dados estejam acessíveis de maneira livre e gratuita. Os dados chegam de todos os contextos da sociedade, gerados pelos aplicativos da web 3.0, pelos governos, investigações científicas, publicações e organizações.

Os dados gerados formam um rastro digital, sendo de fácil recuperação. No entanto, existe uma dificuldade de processamento em massa e, pelo fato de serem essencialmente dados não-estruturados, o potencial da exploração do seu volume ainda não é plenamente aproveitado. Embora sua análise possa gerar informações significativas para empresas e governos, estima-se que apenas 1% desses dados seja efetivamente analisado (BREITMAN⁴, 2013 *apud* REIS; SÁ, 2020).

Big Data é um grande volume de dados, sendo eles de vários tipos, coletados massivamente de fontes heterogêneas (sensores, telefones celulares, redes sociais etc.), além de novos algoritmos e novas tecnologias de *hardware*, *software* que são usados para coletar, armazenar e processar esses dados com alta velocidade (FERREIRA; ROCKEMBACH; KREBS, 2017)

Desta forma, *Big Data* é um termo que descreve grandes conjuntos de dados complexos e em constante crescimento, que são difíceis de serem gerenciados e analisados com as ferramentas e técnicas tradicionais. Esses dados são gerados a partir de diversas fontes, incluindo dispositivos móveis, sensores, transações financeiras, redes sociais e outros. A análise desses dados pode fornecer informações valiosas para empresas e organizações em setores diversos, incluindo finanças, saúde, varejo, marketing e muitos outros.

A *Oracle Corporation*, empresa especializada em desenvolvimento e distribuição de banco de dados define o *Big Data* como:

[...] um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados. Esses conjuntos de dados são tão volumosos que o software tradicional de processamento de dados simplesmente não consegue gerenciá-los. No entanto, esses grandes volumes de dados podem ser usados para resolver problemas de negócios que você não conseguiria resolver antes (ORACLE, 2020).

⁴ BRIETMAN, K. (2013). *Big Data* Overview. Palestra apresentada no 1º. EMC Summer School on *Big Data*. EMC/NCE/UFRJ. Rio de Janeiro.

Outra possibilidade para conceituar *Big Data* é apresentada por Davis (2012, p.4):

Big Data são dados muito volumosos para serem tratados e analisados por protocolos de banco de dados tradicionais como SQL (o que faz *Big Data* um termo que pode evoluir ao longo do tempo, o que é considerado agora *Big Data* pode muito rapidamente tornar-se pequeno). Neste sentido, o tamanho é apenas um aspecto dessas novas tecnologias.

Big Data é um termo utilizado para descrever conjuntos de dados de grande volume, velocidade e variedade, que envolvem tecnologias e técnicas específicas para armazenamento, processamento e análise. O objetivo do *Big Data* é permitir a identificação de padrões, tendências e *insights* que possam ser úteis para empresas e organizações em diversos setores, incluindo finanças, saúde, varejo, marketing, entre outros. Para lidar com *Big Data*, são necessárias ferramentas e tecnologias específicas, como *Hadoop*, *Spark*, *NoSQL*, entre outras (LOPES; SILVA, 2021).

Na Ciência da Informação, *Big Data* é um conceito que se refere à análise e interpretação de grandes conjuntos de dados, com a finalidade de extrair informações úteis e conhecimento. A partir do uso de técnicas e ferramentas de análise de dados, a CI busca identificar padrões, tendências e *insights* em conjuntos de dados de grande volume e complexidade, gerados a partir de diversas fontes. A análise de *Big Data* é especialmente útil para empresas e organizações, pois pode fornecer informações valiosas para a tomada de decisões e o planejamento estratégico. Além disso, a Ciência da Informação também estuda questões éticas e legais relacionadas ao uso de *Big Data*, como privacidade, segurança e propriedade intelectual dos dados (LOPES; SILVA, 2021).

O *Big Data* é um termo associado ao conjunto de dados do qual o tamanho ou tipo está acima da capacidade de bancos de dados tradicionais de capturar, gerenciar e processar os dados.

Os especialistas em *Big Data* trabalham com três tipos básicos de dados, sendo eles:

- *Social Data*: são dados utilizados na análise de campanhas de marketing. Eles são coletados em ambientes de interação entre usuários e nas redes sociais, normalmente comportamentais e demográficos, ditando um padrão de determinado grupo com as mesmas características (LOPES; SILVA, 2021).

- *Enterprise Data*: os dados empresariais são atributos sobre funcionários e setores diferentes dentro de um ambiente empresarial. Eles são coletados pelos diversos setores da empresa, incluindo o RH. Esse tipo de dado é um marco de investimento estratégico de grandes empresas, que visam minimizar gastos e otimizar lucros, pois podem ser utilizados para otimizar processos e identificar falhas ou fraudes dentro de uma determinada seção (LOPES; SILVA, 2021).
- *Personal Data*: os dados pessoais, são dados obtidos através de smartphones, geladeiras, televisões, carros etc. Esses dados são utilizados para determinar preferências pessoais de um indivíduo (LOPES; SILVA, 2021).

De acordo com Lobo (2013), a sociedade se beneficia imensamente do grande volume de dados disponíveis, tanto pela disponibilidade da informação quanto pela qualidade dos serviços orientados por essas informações. Além do interesse acadêmico, o autor enfatiza que profissionais de negócios, indústria e governo descobriram que o *Big Data* oferece vastas oportunidades para o comércio, inovação e engenharia social.

Todas as ações realizadas por pessoa deixam rastros sobre quem ela é e o que faz, de modo que as características de um determinado indivíduo podem ser armazenadas em algum banco de dados espalhado pelo mundo, e que grandes organizações e governos possam acessar tais dados para as mais diversas finalidades. O principal uso de todos esses dados são análises na busca a fim de definir padrões e comportamentos de usuários, pacientes, empregados, doenças, produtos, entre outros (CONEGLIAN; SANTARÉM SEGUNDO; SANT'ANA, 2017).

Ainda, pela evidência da assimetria informacional, a mesma lei também reconhece a vulnerabilidade dos titulares de dados ao se apoiar também no direito do consumidor (MIRAGEM, 2019, p. 27-28). Todavia, por mais que isso possibilite práticas protetivas, reconhece-se pela LGPD o papel do usuário nessa economia de dados, de consumidor, protegido quanto ao que compra, não participe das complexas relações que movimentam a gestão de dados no contexto do *Big Data*.

Essa assimetria entre usuários e corporações é a fundação da nova economia orientada por dados, em que os dados são o motivo da expansão de grandes empresas de tecnologia (*Big Techs*) que têm um baixo custo em expandir seus

serviços enquanto lucram em massa com os ativos de dados (CIURIK, 2018, p. 14-15). A mediação digital de tudo, conforme afirma Morozov (2018, p. 160-166), pressupõe o extrativismo de dados promovido pelas grandes empresas de tecnologia, tendo a noção de que os usuários são estoques de informações valiosas, concebendo então complexas formas de fazê-los abdicar o controle exclusivo sobre seus dados pessoais comportamentais e de compartilhá-los voluntariamente — principalmente mediante o uso de inteligências artificiais com aprendizado profundo, capazes de administrar os milhões de usuários produtores de dados no contexto do *Big Data*. A gestão de dados massivos pressupõe a administração de toda a informação, em tempo real, sendo feita pelas *Big Techs*, assumindo o titular de dados pessoais o papel não um administrador, mas de um conjunto produtivo de dados administrado — e a autodeterminação informativa, nesse cenário, é um princípio jurídico aplicável a conflitos específicos, não sustentáveis do ponto de vista da economia política.

As *Big Techs* são as principais empresas de tecnologia do mundo, que têm uma influência significativa na indústria e na economia global. Elas são empresas dominantes no mercado hoje em dia. Guitarrara (2021) afirma que algumas das maiores empresas desse setor econômico, com ampla presença no mercado e na nossa vida cotidiana, possuem sede nos Estados Unidos e formam o grupo conhecido como *Big Five*, representado pelo acrônimo GAMAM: *Google, Apple, Microsoft, Amazon* e *Meta*:

1. *Apple*: A *Apple* é conhecida por seus produtos de hardware, como *iPhones, iPads* e *Macs*, bem como por seus serviços, incluindo o *iCloud, Apple Music* e *App Store*.
2. *Amazon*: A *Amazon* é uma das maiores empresas de comércio eletrônico do mundo. Além disso, a empresa oferece serviços em nuvem por meio da *Amazon Web Services (AWS)* e possui produtos como o *Kindle* e o *Amazon Echo*.
3. *Google (Alphabet Inc.)*: A *Google* é uma empresa de tecnologia diversificada oferecem uma variedade de serviços e produtos, incluindo seu mecanismo de busca, *Android* (sistema operacional móvel), *Google Cloud, YouTube* e muitos outros.
4. *Facebook*: O *Facebook* é a maior rede social do mundo e é proprietário de outras plataformas populares, como *Instagram* e *WhatsApp*.

5. *Microsoft*: A *Microsoft* é conhecida por seus sistemas operacionais, como o *Windows*, e sua suíte de produtividade, o *Microsoft Office*. A empresa também oferece serviços em nuvem com o *Microsoft Azure* e possui consoles de videogame, como o *Xbox*.

Essas são algumas das principais empresas consideradas *Big Techs*. Vale ressaltar que a classificação pode mudar ao longo do tempo, à medida que a indústria evolui e novas empresas emergem como líderes.

2.2 Ciclo de Vida de Dados

A Ciência da Informação tem como dever contribuir para o novo cenário de acesso e uso intensivo de dados, explorando fatores e características que permitam a ampliação do equilíbrio entre os usuários do processo e as entidades envolvidas no gerenciamento. Esses fatores e características podem ser compreendidos a partir do ciclo de vida de dados.

Rahul e Banyal (2020) consideram que o ciclo de vida dos dados apresenta todo o processo de dados no sistema, através de uma estrutura de alto nível para planejar, organizar e gerenciar todos os aspectos durante suas fases (coleta, armazenamento, usabilidade, compartilhamento e descarte dos dados). Logo, para uma implementação bem-sucedida do *Big Data*, é necessário manter o ciclo de vida dos dados sob um sistema ou modelo seguro, completo e eficaz de gerenciamento de dados.

O problema é que existem vários modelos de ciclo de vida de dados, com diferentes características e aplicações. Por exemplo, Sant'Ana (2013) identificou 18 iniciativas de modelos de ciclo de vida de dados entre 1997 e 2013 (SAYÃO; SALES, 2015).

Shah, Peristeras e Magnisalis (2020), por meio de uma Revisão Sistemática da Literatura, identificaram e analisaram 76 modelos de ciclos de vida de dados para propor uma estrutura de ciclo de vida de dados para governos. Os modelos analisados eram referentes a diversos domínios de aplicação, como pesquisa científica, Web Semântica, bancos de dados, qualidade de dados, processamento de dados, etc. Além disso, envolviam diferentes etapas do ciclo de vida, algumas confusas e difíceis de associá-las às práticas de gestão em organizações.

O modelo de ciclo de vida de dados para governos orientados a dados, proposto por Shah, Peristeras e Magnisalis (2020), envolve 14 fases: planejamento,

coleta, preparação, análise, visualização, acesso, compartilhamento, uso e reuso, arquivamento, descarte, governança, qualidade, armazenamento e segurança ou proteção de dados. Para cada fase existem funções principais a serem consideradas. Por exemplo, na fase de coleta deve-se apresentar um programa de conscientização sobre proteção de dados durante a coleta de dados; coletar metadados com base em padrões de metadados, para garantir a interoperabilidade etc.

Embora o modelo dos autores possa ser utilizado em várias áreas do setor público, como saúde, agricultura, educação; seja iterativo e envolva a completude do em relação aos desafios do *Big Data* (volume, variedade, velocidade e veracidade), na nossa análise, seria de difícil compreensão e aplicação no contexto desta pesquisa.

Em vista disso, nesta pesquisa, escolhemos o modelo de Ciclo de Vida dos Dados proposto por Sant'Ana (2016), que foi desenvolvido levando em consideração as características e especificidades da Ciência da Informação.

O panorama atual do *Big Data* indica um novo conjunto de possibilidades de uso e consequências referentes a alta disponibilidade de dados, apresentando novos requisitos dos usuários a partir de suas necessidades informacionais (SANT'ANA, 2016, p.2). Não se pode esquecer os eventuais danos causados pelo uso massivo dos dados como a assimetria informacional (AKERLOF, 1970), e esses danos podem ser perigosos para os indivíduos e para a sociedade como um todo.

De acordo com Sant'Ana (2016, p.2)

Esta potencial assimetria se sustenta na necessidade de camadas de recursos tecnológicos para uso destes dados que tendem a ser mais profundas e complexas com o aumento das características de *Big Data*, o que justifica um olhar para as entranhas desta tecnologia. Neste cenário emerge a necessidade de uma estrutura básica em que se possa, ao menos, contextualizar momentos, características e requisitos bem como fatores que permeiam estes momentos (SANT'ANA, 2016, p.2).

O *Big Data* tem uma grande importância para a Ciência da Informação, pois oferece uma grande oportunidade para a coleta, armazenamento e análise de grandes volumes de dados. Isso permite que os profissionais da área possam extrair insights úteis e valiosos, a partir dos dados, que podem ser utilizados para processos aprimorados, tomadas de decisões e soluções informacionais.

Dessa forma, o *Big Data* se torna uma ferramenta importante para a Ciência da Informação, ajudando na exploração e interpretação de grandes volumes de dados e permitindo que os profissionais dessa área possam utilizar essas informações para

aprimorar seus projetos e soluções informacionais, além de contribuir para a construção de uma sociedade mais resistente e conectada.

Sant'Ana (2016) afirma ser necessário pensar em uma estrutura que permita a percepção destes momentos e características, que no caso dos dados podem ser considerados como cíclicos, proporcionando a reflexão sobre um ciclo de vida dos dados (SANT'ANA, 2016, p.3)

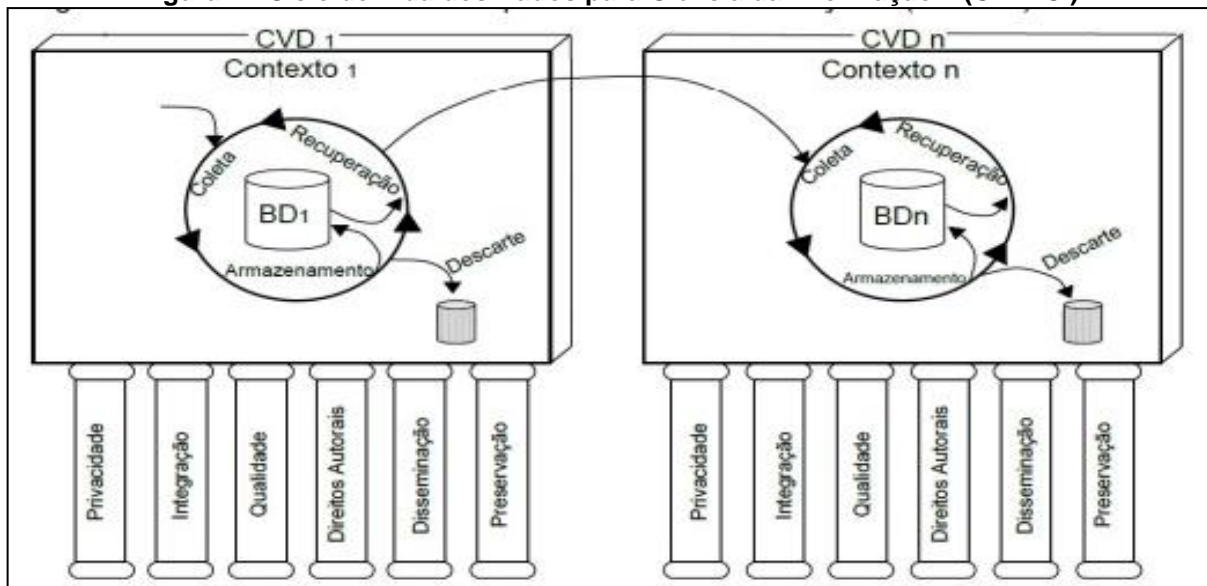
2.2.1 Fases do Ciclo de Vida dos Dados e fatores paralelos – CVD-CI

A Ciência da Informação tem como dever contribuir para o novo cenário de acesso ao uso intensivo de dados, explorando fatores e características que permitam a ampliação do equilíbrio entre os usuários do processo e a otimização do uso dos dados.

Conforme afirma Sant'Ana (2016, p.8):

Nesta tarefa se faz necessário estruturar esta análise, e para tanto, propõe-se a utilização de uma delimitação de fases (momentos em que distintas necessidades e competências são necessárias) envolvidas no acesso e uso dos dados, mantendo-se como ponto central os próprios dados e para tanto se propõe o uso do Ciclo de Vida dos Dados - CVD como forma de evidenciar os diferentes momentos e fatores envolvidos neste processo.

Figura 2 - Ciclo de Vida dos Dados para Ciência da Informação – (CVD-CI)



Fonte: Sant'Ana (2016)

A Figura 3 ilustra as fases do ciclo de vida dos dados que devem ser consideradas na perspectiva epistemológica e pragmática da Ciência da Informação: coleta, armazenamento, recuperação e descarte. Para cada fase do ciclo de vida

devemos considerar os seguintes fatores norteadores: privacidade, integração, qualidade, direitos autorais, disseminação e, preservação.

Coleta: a princípio, é necessário obter dados que possam ser utilizados para atender a uma necessidade específica ou demanda esperada relacionada a um contexto específico. Neste primeiro momento são necessárias medidas para implementar um plano de ação, análise de viabilidade e realização da coleta de dados. Outras questões importantes nesta fase incluem: como os dados podem ser coletados? Em que formato estão? Quais são os dados necessários? Nesta etapa, o fator privacidade recai principalmente sobre as características e fontes de dados e o processo envolvido na sua coleta. Quanto à integração, torna-se necessário considerar os principais atributos para identificação dos dados, incluindo aspectos da sua granularidade. Nesta fase também se torna imprescindível considerar os aspectos inerentes à propriedade intelectual e à ética no tratamento de dados, por exemplo, respeitando os limites estabelecidos na LGPD. A disseminação nesta fase diz respeito a identificação de elementos contextuais dos dados que possibilitem a localização e interpretação na fase de recuperação. A preservação, por sua vez, diz respeito a estruturas de armazenamento e a linguagens de representação (SANT'ANA, 2016).

Armazenamento: Posteriormente, depois da Coleta dos Dados, pode ser necessário e útil manter os mesmos disponíveis de alguma forma para referência futura. Neste momento passa-se, então, a uma segunda fase que é aquela em que os esforços são no sentido de manter estes dados em uma determinada base de dados. Esta fase recebe o nome de armazenamento. Outras questões importantes nesta fase incluem: Qual estrutura (física e lógica) será utilizada para seu armazenamento? Estes dados podem representar um risco a privacidade dos indivíduos ou instituições neles referenciados de alguma forma? (SANT'ANA, 2016).

Recuperação: Esta fase tem como objetivo a viabilização da recuperação dos dados. Dentre outras questões fundamentais desta fase pode-se destacar: quais dos dados armazenados serão disponibilizados? Existe algum público-alvo específico? Durante o processo de recuperação quais são os riscos à privacidade dos indivíduos ou entidades referenciados pelos conteúdos recuperados? O acesso será feito diretamente a base em que se encontra armazenado ou será necessário retornar a fase de armazenamento para definição de novas estruturas de armazenamento específicas para recuperação? (SANT'ANA, 2016).

Descarte: Nesta fase são analisados quais dados são importantes para continuar sendo mantidos em uma base e dados e dos demais são descartados, ou seja, excluídos. Quais dados já não são mais necessários? Estes dados estão replicados em outras bases? Como garantir e explicitar que estes dados foram realmente excluídos e não simplesmente ocultos? Para o descarte foi considerada a necessidade de preservação em seus diversos aspectos? (SANT'ANA, 2016).

De acordo com Sant'Ana (2016, p. 134-137) se estamos considerando que um dado não está disponível na fase de recuperação, o principal fator envolvido é o "disseminação", ou seja, este dado não está disponível para que outros colem, no entanto, o motivo que leva a esta falha na disseminação pode estar relacionado a diversos motivos, exemplificando:

[1] o dado não foi disseminado por uma questão de erro na gestão dos dados, portanto seria também uma forma de considerar o fator "qualidade";

[2] o dado não foi disseminado por ser considerado um dado sensível, portanto seria também uma forma de considerar o fator "privacidade";

[3] o dado não foi disseminado por não ter dados complementares que permitissem sua vinculação a pesquisa que foi realizada pelo usuário em outro conjunto de dados, portanto seria também uma forma de considerar o fator "integração";

[4] o dado não foi disseminado por não ter formalmente autorização de detentores para sua publicação, portanto seria também uma forma de considerar o fator "legislação", e;

[5] o dado não foi disseminado por não ter sido mantido íntegro, seja em sua forma física ou lógica, pelo tempo necessário, portanto seria também uma forma de considerar o fator "preservação"

2.3 Dados e Privacidade

O conceito de dado precisa ser redimensionado e se faz necessário o seu entendimento e a sua percepção como elemento básico, reconhecendo sua importância como um elemento fundamental nos fluxos de informações. Especialmente em um momento em que a questão do acesso aos dados é amplamente estendida e legislada na administração de conteúdo, no favorecimento de sua visibilidade e na sua utilização e reutilização (SANTOS; SANT'ANA, 2015).

Davenport (1998) retoma a ideia a partir da definição de que

[...] dados são simples observações sobre o estado do mundo, podendo ser facilmente estruturados, obtidos, tratados e transferidos por máquinas. Ainda mais utilitarista, e baseado na visão da área de tecnologia, o autor retoma o conceito, descrevendo dados como “conjunto de fatos distintos e objetivos, relativos a eventos, sendo descritos utilitariamente, no contexto organizacional, como registros estruturados de transações” (DAVENPORT; PRUSAK, 1999, p.2).

De acordo com Santos e Sant’Ana (2002), o termo "dado" é definido como um elemento fundamental que consiste em um signo ou conjunto limitado de signos, que não possui um significado intrínseco, mas apenas elementos sintáticos. Isso incorporava a perspectiva do dado como um símbolo ou elemento que não possuía significado suficiente para ser interpretado por si só, o que minimizava, na época em que o conceito foi proposto, o desafio da delimitação dos dados.

Ao considerar a obtenção de dados, é fundamental começar a criar um método de coleta de alta qualidade, aos seus objetivos pretendidos, a fim de assegurar que obtenha as informações desejadas de forma precisa.

Garfinkel (2000, p.92) relata que estamos à beira de uma crise de informação. Nunca houve tanta informação sobre tantas pessoas coletadas em tantos lugares diferentes. Nunca tanta informação foi disponibilizada tão facilmente para tantas instituições em tantas maneiras diferentes e para tantos propósitos diferentes.

Alan Westin cunhou o termo sombra de dados ou *data shadow* na década de 1960. Westin, professor da *Columbia University em Nova York*, alertou que registros de crédito, registros bancários, seguros registros e outras informações que compunham a infraestrutura digital emergente da América poderiam ser combinados para criar um dossiê digital detalhado. A metáfora, com seu caráter ligeiramente sensação sinistra, foi estranhamente preciso: assim como poucas pessoas estão cientes de onde seus as sombras caírem, poucos titulares de dados no futuro, conjecturou Westin, seriam capazes de manter acompanhamento de seus dossiês digitais.

Mas esse arquivo de fatos e sentimentos é uma lâmina que pode cortar de duas maneiras diferentes. Mais do que meu próprio diário digital, também tenho lançado uma vasta "sombra de dados" que revela os segredos da minha vida diária para qualquer um que possa lê-lo (GARFINKEL, 2000, p.92).

De acordo com Garfinkel (2000, p.92) *data shadow* ou sombra de dados são rastros que os titulares de dados vão deixando, principalmente, em transações digitais diárias, como caixa eletrônico, pagamento com cartão de crédito, reserva de horário etc. Esses rastros de dados são mais difíceis de controlar e garantir a sua proteção contra usos indevidos.

Desta forma, sombra de dados ou *data shadow* é um termo informal usado para descrever a transmissão de pequenos fragmentos de informações deixados por um indivíduo através de suas atividades aprendidas. Esses dados são gerados quando alguém envia um e-mail, atualiza um perfil em uma rede social, usa um cartão de crédito, realiza uma transação em uma caixa eletrônica, entre outras ações. A noção de sombra de dados tornou-se uma preocupação significativa, uma vez que é difícil controlar quem tem acesso aos dados de uma pessoa, quais foram extraídas dessas informações e quais medidas são tomadas com base nestas elaboradas (ROUSE, 2014, p. 5).

Existem leis de privacidade de dados que estão sendo criadas para prevenir o abuso das sombras de dados de um indivíduo. Por exemplo, isso pode ocorrer quando um empregador demite um funcionário com base em informações sobre seus amigos ou fotos no Facebook. No entanto, é lamentável que a legislação de privacidade geralmente esteja aquém da capacidade de uma organização de coletar, compilar e analisar dados (ROUSE, 2014, p. 6).

Os dados são frequentemente usados como entrada para algoritmos de computador e análises estatísticas para extrair informações úteis e conhecimento a partir deles. A Ciência da Informação se preocupa com a coleta, organização, armazenamento, recuperação e análise de dados para facilitar a sua utilização eficaz e eficiente.

Desta forma, a Ciência da Informação, um dado é um elemento básico de informação que pode ser armazenado, processado e transmitido por meio de tecnologias da informação. Um dado pode ser um número, um texto, uma imagem, um som, um vídeo, entre outros tipos de dados digitais.

Dados e privacidade são dois conceitos intimamente relacionados à era digital. Dados referem-se às informações coletadas e recebidas sobre indivíduos, organizações ou qualquer outra entidade. Isso pode incluir informações pessoais,

como nomes, endereços, números de telefone, dados financeiros, histórico de navegação na web e muito mais.

De acordo com Doneda (2019, p.17), o contexto atual, em que a informação pessoal e a privacidade se relacionam, é diretamente moldado pela tecnologia, juntamente com algumas mudanças sociais. Doneda argumenta que isso ocorre principalmente devido ao reconhecimento do direito à privacidade como um direito fundamental, o que resulta em sua funcionalidade e ampliação em diversos direitos subjetivos. Esses direitos incluem não apenas o direito à vida privada e à vida familiar, mas também o direito à proteção dos dados pessoais (DONEDA, 2019, p.43).

Doneda empresta para si a definição que Rodotà dava para privacidade, entendida como: o direito de manter o controle sobre as próprias informações e de determinar as modalidades de construção da própria esfera privada (RODOTÀ apud DONEDA, 2019, p. 132).

Privacidade é uma questão de acesso a coisas consideradas privadas, desta forma, ter privacidade é ter controle sobre o que consideramos privado (MACNISH, 2019, p.7).

A privacidade é violada quando as informações são compartilhadas em contextos inapropriados: quando nosso médico vende nossas informações médicas para empresas de publicidade ou nossos amigos compartilham nossa conversa privadas no Twitter. Macnish (2019) continua argumentando que o valor da privacidade reside em sua capacidade de proteger os indivíduos e contribuir para as liberdades democráticas.

A privacidade está sujeita a mudanças à medida que ocorrem as transformações sociais, políticas e culturais. No entanto, é importante ressaltar que o conceito fundamental de privacidade é considerado inviolável, pois busca garantir ao indivíduo o direito à sua vida privada, bem como o controle sobre a disponibilização de seus dados e a utilização de suas informações por terceiros (GRISOTO; SANT'ANA; SANTARÉM SEGUNDO, 2015, p.169)

O direito à privacidade é um direito humano de primeira dimensão que se configura no direito do indivíduo de estar só, considerando-se a possibilidade que deve ter toda pessoa de excluir do conhecimento de terceiros aquilo que a ela só se refere, e que diz respeito ao seu modo de ser no âmbito da vida privada. O direito à privacidade é universal,

inalienável e indivisível e está intimamente ligado à dignidade humana. (GALVÃO; RICARTE, 2012, p.98)

Desta forma, a privacidade é o direito que os indivíduos têm de controlar o acesso, o uso e a divulgação de suas informações pessoais. Essas informações podem ser qualquer dado que possa ser usado para identificar uma pessoa, como nome, endereço, número de telefone, dados financeiros, histórico médico, entre outros. A privacidade é fundamental porque permite que as pessoas mantenham o controle sobre sua identidade e vida privada, evitando o uso indevido de suas informações pessoais por terceiros, como empresas, governos ou outras pessoas. A privacidade é protegida por leis e regulamentações específicas em muitos países, que visam garantir que os indivíduos tenham o direito de escolher como suas informações pessoais são coletadas, armazenadas e compartilhadas.

Com os avanços tecnológicos e a crescente digitalização da sociedade, a coleta e o uso de dados completaram-se onipresentes. Empresas, governos e outras organizações coletam dados em grande escala para diversos fins, como serviços melhorar, segmentar anúncios, tomar decisões de negócios e muito mais.

No entanto, a coleta e o uso indiscriminado de dados podem representar riscos para a privacidade. Quando informações pessoais são mal protegidas, podem ser acessadas por terceiros não autorizados, provocadas em roubo de identidade, fraude, abuso e outras violações. Além disso, uma combinação de diferentes conjuntos de dados pode revelar informações sensíveis sobre uma pessoa, mesmo que os dados individuais não sejam acessíveis.

Uma das lições mais importantes da última década é que a privacidade não só não é contrária à segurança, mas é parte integrante dela. Para manter os cidadãos seguros, temos de garantir a sua privacidade. E para proteger a privacidade, precisamos garantir a segurança de nossos sistemas digitais (MACNISH, 2019, p.26).

Para abordar essas questões, muitos países têm leis e regulamentos que visam proteger a privacidade dos indivíduos, como o Regulamento Geral de Proteção de Dados (RGPD) na União Europeia e a Lei Geral de Proteção de Dados (LGPD) no Brasil. Essas legislações estabelecem requisitos para a coleta, o armazenamento e o uso de dados pessoais, além de garantir os direitos dos indivíduos em relação aos seus dados.

Além das medidas legais, é importante que as empresas adotem práticas de privacidade por design, ou seja, considerem a privacidade desde o início ao desenvolver produtos e serviços. Isso inclui a implementação de medidas de segurança robustas, obtenção de consentimento adequado para a coleta de dados, transparência sobre as práticas de privacidade e concessão de opções para que os indivíduos possam controlar suas informações.

Em última análise, a proteção da privacidade no mundo digital requer uma abordagem multifacetada que envolve uma combinação de legislação adequada, práticas éticas por parte das organizações e conscientização dos indivíduos sobre seus direitos e opções em relação aos seus dados pessoais.

Garantir a privacidade dos dados requer a adoção de práticas e políticas adequadas, como a obtenção de consentimento explícito para a coleta e uso de informações pessoais, o armazenamento seguro dos dados, a implementação de medidas de segurança cibernética e a transparência nas práticas de coleta e uso de dados. Além disso, é essencial educar e conscientizar as pessoas sobre seus direitos de privacidade e as melhores práticas para proteger suas informações pessoais.

Ao equilibrar a utilização dos dados para fins legítimos, como melhorar serviços, personalizar experiências e impulsionar avanços tecnológicos, com a proteção da privacidade individual, é possível estabelecer uma abordagem responsável e ética em relação aos dados.

O aumento do volume de dados tem modificado a maneira como pesquisas, governança, socialização e negócios estão sendo realizados. Podem ser produzidos por diferentes classes sociais, governos, indústrias, institutos de pesquisa e universidades. Eles são coletados, armazenados e divulgados de diferentes maneiras e com diferentes organismos (ANDRADE; GONÇALEZ; BERTI JUNIOR; BAPTISTA; CONEGLIAN, 2020).

De acordo com Milagre, Santarém Segundo (2015, p.2) dado é qualquer elemento identificado em sua forma bruta que, por si só, não conduz a uma compreensão de determinado fato ou situação.

Para Souza e Almeida (2021, p. 4): “[...] os dados não apresentam significados próprios, porém, representam a matéria prima como uma série de observações, medidas ou fatos que podem ser representados como números, palavras, sons ou imagens ao qual é produzida a informação.”

Desta maneira, é possível dizer que o dado é o estado bruto de uma futura informação, pois o mesmo após ser coletado, processado e analisado será transformado em uma informação que será utilizado por uma organização.

De acordo com a Lei nº 12.527/2011 - Lei de Acesso à Informação (LAI) – e a Lei nº 13.709/2018 - Lei Geral de Proteção de Dados Pessoais (LGPD) - os dados pessoais são "[...] informações relacionadas a uma pessoa identificada ou identificável". (Brasil, 2011, 2018). De acordo com a legislação brasileira, os dados pessoais são originários de qualquer indivíduo que possa ser identificado de forma única por meio de dados e informações. Exemplos de dados pessoais incluem números de documentos como o Cadastro Geral (RG), o Cadastro de Pessoas Físicas (CPF), o próprio nome, nomes dos pais e outras informações que possam identificar uma pessoa física (VIGNOLI; VERONEZ JUNIOR; SILVA; MONTEIRO, 2020).

De acordo com Silveira, Avelino e Souza (2016, p.3):

Dado pessoal pode ser definido como qualquer informação relativa a uma pessoa singular identificada ou identificável. O Relatório do Fórum Econômico Mundial qualificou o dado pessoal como as informações e metainformações criadas por e sobre as pessoas, abrangendo: dados oferecidos voluntariamente (exemplo: perfil na rede social), dados observados (como: dados de localização ao usar os celulares) e dados inferidos (exemplo: análise de informações oferecidas ou observadas com a finalidade de construir uma pontuação de crédito) (SILVEIRA; AVELINO; SOUZA, 2016, p.3).

Para o Grupo das Nações Unidas para o Desenvolvimento (2018, p. 12):

Dados pessoais significam dados, em qualquer forma ou meio, relacionados a um indivíduo identificado ou identificável que possa ser identificado, direta ou indiretamente, por meios razoavelmente prováveis de serem usados, inclusive quando um indivíduo possa ser identificado por vincular os dados a outras informações razoavelmente disponíveis. Os dados pessoais são definidos por muitos instrumentos regionais e nacionais e podem ser referenciados como informações pessoais ou informações de identificação pessoal (UNITED NATIONS DEVELOPMENT GROUP, 2018, p.2).

Durante os cadastros solicitados pelas empresas, pagamentos de compras presenciais ou *on-line* acontece o fornecimento de dados pessoais. Desta maneira, o próprio indivíduo fornece voluntariamente seus dados pessoais. A disponibilização de dados pessoais a terceiros através do consentimento livre ocorre todos os dias sem que o indivíduo tenha conhecimento de que está a fornecer os seus dados (VIGNOLI; VERONEZ JUNIOR; SILVA; MONTEIRO, 2020).

Cada vez mais relevante na sociedade informacional o mercado de dados pessoais gera interações econômicas voltadas a compra e venda das informações de uma pessoa identificada ou identificável. Ele é baseado nas necessidades de informação das empresas, instituições públicas e usuários finais (SILVEIRA; AVELINO; SOUZA, 2016).

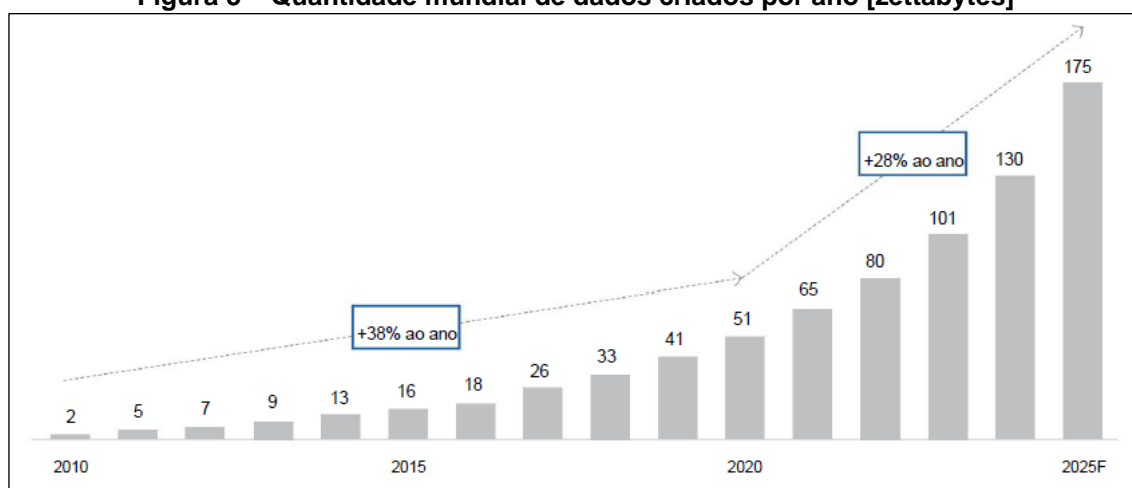
De acordo com o texto da *Organisation for Economic Co-operation and Development* (2013, p.4):

[...] os dados são normalmente utilizados para servir melhor os clientes, melhorar a eficiência das transações e a qualidade dos produtos, bem como para identificar as macrotendências em um número de diferentes setores, incluindo saúde, transporte e segurança (OECD, 2013, p. 4).

A população global adota cada vez mais o uso de suas atividades ON-LINE, buscando flexibilidade e conveniência comparado aos métodos tradicionais. Somente em 2019, o número de usuários globais de *Internet* cresceu 7%, atingindo 4,5 bilhões de usuários. Essa taxa de crescimento é seis vezes maior que o crescimento da população em geral. Principalmente no Canadá, a taxa de utilização da *Internet* é de 94%, representando mais de 35 milhões de usuários ativos (DATAREPORTAL, 2020).

As possibilidades revolucionárias oferecidas pelas novas tecnologias para a economia digital são enormes, no entanto, representam grandes desafios para as políticas e os legisladores que lutam para fazer cumprir a regulamentação em um ambiente de ritmo acelerado. Os riscos de violações de segurança e má gestão de privacidade aumentam diariamente com o surgimento de novas tecnologias (HUAWEY, 2021).

Figura 3 – Quantidade mundial de dados criados por ano [zettabytes]



Fonte: Huawei (2021)

Tecnologias específicas como a Análise de *Big Data*, *Internet das Coisas* (IoT) e a Computação em Nuvem levantam questões sobre a privacidade dos dados (HUAWEY, 2021).

- Computação em nuvem: a computação em nuvem levanta questões complexas relacionadas a transferências internacionais de dados. Ao conectar os usuários globalmente, os provedores de serviços em nuvem podem armazenar dados e se comunicar em locais externos. Para limitar a exposição potencial de dados domésticos, alguns países implementaram regras de localização de dados exigindo que provedores de serviços estrangeiros armazenem dados internamente. Não há um acordo unânime sobre formas de regular a computação em nuvem (HUAWEY, 2021).
- *Internet das Coisas* (IoT): o número de dispositivos conectados e a quantidade de dados que eles geram cresce rapidamente, acarretando uma maior troca de informações através das fronteiras internacionais. Desta forma, há a necessidade de cooperação e regulamentação internacional, visto que os ataques cibernéticos de IoT são cada vez mais prevalentes. Por exemplo, 80% dos tomadores de decisão de segurança da saúde na China, Alemanha, Japão, Reino Unido e EUA relataram ataques a seus dispositivos conectados entre 2018 e 2019 (IRDETO, 2019).
- Análise de *Big Data*: as organizações buscam diariamente coletar dados de usuários para entender melhor a dinâmica de mercado. Entretanto, o princípio implícito do *Big Data* em coletar todos os dados, não apenas os dados pertinentes estão em oposição aos princípios globais de limitação de coleta de dados (conforme a estrutura de privacidade do bloco econômico APEC). Esses princípios oferecem espaço de manobra para as organizações, afetando negativamente a privacidade dos indivíduos (HUAWEY, 2021).

Nos últimos anos houve um aumento da conscientização pública sobre privacidade de dados, após grandes violações de dados documentadas e casos de abuso de coleta de dados. A privacidade é uma questão particularmente difícil de regular em um ambiente político e empresarial em constante mudança, por um lado, pela busca por maior desenvolvimento e adoção tecnológica e, por outro, por exemplos de má gestão de dados (HUAWEY, 2021).

O Grupo das Nações Unidas para o Desenvolvimento (2018, p. 12) define privacidade como:

[...] a presunção de que os indivíduos devem ter uma área de desenvolvimento autônomo, interação e liberdade, uma 'esfera privada' com ou sem interação com outros, livres de intervenção do Estado e de intervenção excessiva não solicitada de outros indivíduos (UNITED NATIONS DEVELOPMENT GROUP, 2018, p.12).

Para Mendes (2008, p.23) privacidade é a reivindicação de indivíduos, grupos ou instituições para determinar, quando, como e em que extensão, informações sobre si próprios devem ser comunicadas a outros.

De acordo com o Art. 5º, inciso X da Constituição brasileira de 1988: “[...] são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação”. (BRASIL, 1988).

Portanto, a privacidade sofre alterações à medida que as transformações sociais, políticas, culturais e econômicas vão ocorrendo, partindo do pressuposto que o núcleo duro de seu conceito é inviolável, pois tenta garantir ao indivíduo o direito sobre a sua vida privada e sobre a disponibilização de seus dados e a utilização de suas informações por terceiros (GRISOTO; SANT’ANA; SEGUNDO,2015).

A privacidade, é alterada à medida que ocorrem as transições sociais, políticas, culturais e econômicas, partindo do pressuposto de que o seu conceito é a inviolável, pois busca garantir o direito do indivíduo à vida privada e disponibilização de seus dados e a utilização de suas informações por terceiros (GRISOTO; SANT’ANA; SEGUNDO,2015).

O direito à privacidade é um direito humano de primeira dimensão que se configura no direito do indivíduo de estar só, considerando-se a possibilidade que deve ter toda pessoa de excluir do conhecimento de terceiros aquilo que a ela só se refere, e que diz respeito ao seu modo de ser no âmbito da vida privada. O direito à privacidade é universal, inalienável e indivisível e está intimamente ligado à dignidade humana. (GALVÃO; RICARTE, 2012, p.98)

A criação do Marco Civil da *Internet* dispõe:

[...] sobre princípios, garantias, direitos e deveres para o uso da *Internet* no Brasil e determina as diretrizes para atuação da União, dos Estados, do Distrito Federal e dos Municípios, busca garantir dentre outras coisas, a segurança no acesso e na disponibilização dos dados. (BRASIL, 2014).

Medidas estão sendo tomadas nos termos da Lei de Acesso à Informação e da Lei Marco de Cidadania visando garantir a integridade e a dignidade humana na proteção de suas informações e no direito de acesso à informação pública (GRISOTO; SANT'ANA; SEGUNDO,2015).

Mendes (2008, p. 17-18) discorre que a:

A evolução do direito à privacidade prosseguiu para se adaptar às novas transformações sociais ocasionadas pela revolução da tecnologia da informação que possibilitou a coleta e o processamento dos dados pessoais dos cidadãos de forma pioneira. Além de adquirir um caráter positivo e de ser reconhecido no âmbito internacional, o direito à privacidade transformou-se para ensejar o nascimento da disciplina de proteção de dados pessoais, à medida que surgiram novos desafios ao ordenamento jurídico a partir do tratamento informatizado dos dados.

A privacidade é um dos principais desafios éticos no contexto do *Big Data*. Isso porque o processo de coleta, armazenamento e análise de grandes volumes de dados pode afetar a privacidade das pessoas, expondo informações pessoais e sensíveis sem o seu consentimento ou conhecimento.

Para lidar com esse problema, é necessário estabelecer políticas e práticas que garantam a privacidade dos dados. Isso inclui o uso de técnicas de anonimização e criptografia para proteger as informações coletadas e garantir que elas não possam ser identificadas ou associadas a indivíduos específicos. Além disso, é importante informar claramente as pessoas sobre quais dados estão sendo coletados e para que finalidades, bem como garantir que elas tenham controle sobre o uso dessas informações.

2.4 Marco regulatório brasileiro

Em 2018, segundo um estudo da *McAfee* publicado na revista *Veja*, o Brasil registrou perdas progressivas com crimes virtuais, chegando a R\$ 10 bilhões por ano. Nosso país é uma das “potências” mundiais nesse quesito, ao lado de Índia, Vietnã, Rússia e Coreia do Norte.

Empresas e usuários desde o início da década vêm buscando respostas para as questões de segurança virtual, uma vez que os cibercrimes não param de crescer. A proteção dos dados pessoais e o direito à privacidade foi tratado de uma forma mais abrangente na Constituição Federal de 1988.

Na Constituição Federal de 1988, a palavra privacidade é tratada de forma implícita pelo inciso X do art. 5º da Carta Magna, que visa garantir o que pode ser

considerado como preceitos fundamentais da privacidade, escrito da seguinte forma: são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação”. (BRASIL, 1988)

No inciso XII do art. 5º da Constituição Federal é citado sobre a proteção dos dados, porém de uma perspectiva diferente tratado atualmente, contudo com a possibilidade de interpretação e adequação da norma, garante a proteção dos dados pessoais de qualquer pessoa, independente do cenário em que está aplicado. De acordo com a Constituição Federal (1988):

[...] é inviolável o sigilo da correspondência e das comunicações telegráficas, de dados e das comunicações telefônicas, salvo, no último caso, por ordem judicial, nas hipóteses e na forma que a lei estabelecer para fins de investigação criminal ou instrução processual penal. (BRASIL, 1988).

Evidentemente, a sociedade brasileira passou por diversas mudanças desde a entrada em vigor da Constituição Federal em 1988, que ainda vigora nos dias atuais, por isso não compreende todos os fatos novos que surgem com o advento das novas tecnologias, principalmente da *Internet*, para o Brasil Juridicamente falando, é um desafio garantir a preservação dos dados pessoais dos cidadãos disponibilizados na *Internet* (BALDISSERA, 2019).

Em relação a esses novos fatos que começam a surgir devido à informatização de nossa sociedade, segundo Baldissera (2019):

Por meio da proteção de dados pessoais, garantias a princípio relacionadas à privacidade passam a ser vistas em uma ótica mais abrangente, pela qual outros interesses devem ser considerados, abrangendo as diversas formas de controle tornadas possíveis com a manipulação de dados pessoais. Para uma completa apreciação do problema, estes interesses devem ser considerados pelo operador do direito pelo que representam, e não somente pelo seu traço visível – a violação da privacidade. (DONEDA, 2011, p. 95).

As medidas tomadas nesse sentido é a promulgação da Lei de Acesso Informação (Lei nº 12.527/2011), Lei de Crimes Informáticos (Lei nº 12.737/2012) e a instituição do Marco Civil da *Internet* (Lei nº 12.965/2014), segundo Vinicius Borges Fortes (2016). Recentemente foi criada a Lei nº 13.709 de 2018, que trata exclusivamente da proteção de dados pessoais, a qual foi sancionada em 14 de agosto pelo Presidente da República Michel Temer (BALDISSERA, 2019).

A Lei de Acesso à Informação, segundo Baldissera (2019) pode ser explicada resumidamente como:

A Lei de Acesso à Informação determina que o tratamento das informações pessoais detidas por entidades e instituições nela abrangidas seja realizado de modo transparente, respeitando o direito fundamental à proteção da intimidade, da vida privada, da honra e da imagem das pessoas, o que, nos fundamentos defendidos nesta obra, corresponde à proteção do direito fundamental à privacidade. A lei impõe restrições substanciais de acesso a informações pessoais, como o acesso restrito às informações, pelo prazo máximo de cem anos, a agentes públicos autorizados, bem como a possibilidade de acesso ou divulgação a terceiros, mediante prévio consentimento do titular das informações, exceto nos casos previstos no regulamento. (FORTES, 2016, p. 118).

A Lei de Crimes Informáticos, que foi popularmente conhecida como Lei Carolina Dieckmann⁵, pois a atriz teve suas imagens obtidas e divulgadas na *Internet*, pela invasão de aparelhos eletrônicos pessoais da atriz (BALDISSERA, 2019). A situação evidenciou a necessidade de penalidade específica para esse tipo de crime, o que foi feito com a criação da Lei 12.737/2012, que inseriu o artigo 154-A no Código Penal, tendo como teor:

Invadir dispositivo informático alheio, conectado ou não à rede de computadores, mediante violação indevida de mecanismo de segurança e com o fim de obter, adulterar ou destruir dados ou informações sem autorização expressa ou tácita do titular do dispositivo ou instalar vulnerabilidades para obter vantagem ilícita.
Pena – detenção, de 3 (três) meses a 1 (um) ano, e multa. (BRASIL, 2012).

O Marco Civil da *Internet* criou uma lei para contemplar a maior quantidade possível de situações que poderiam acontecer envolvendo a *Internet*.

De acordo com Teixeira (2016, p.18):

O Marco Civil da *Internet* no Brasil, como se sabe, é a Lei que regula o uso da *Internet* no Brasil, por meio da previsão de princípios e estabelecimento de garantias aos usuários. O texto trata de temas como neutralidade da rede, privacidade e retenção de dados, impondo obrigações para os provedores de serviços de *Internet*.

⁵ Diante do primeiro escândalo do gênero no país sobre várias fotos íntimas divulgadas na *Internet*, não tardou para que o caso ganhasse os olhos do público e da justiça. Em menos de um ano após o caso, a lei Nº 12.737/2012, apelidada de Lei Carolina Dieckmann, foi sancionada no dia 30 de novembro de 2012. A criação da lei se deu em virtude do caso da atriz que, na época do crime, não recebeu amparo de uma legislação específica para a devida penalização dos criminosos.

A Lei nº 12.965, de 23 de abril de 2014, estabelece normas reguladoras sobre o uso da *Internet* no Brasil. No art. 3º é abordado a proteção dos dados pessoais e no inciso III o direito à privacidade (BRASIL, 2014).

[...] o Marco Civil determina que a guarda e a disponibilização dos registros de conexão e de acesso a aplicações de *Internet*, bem como de dados pessoais e do conteúdo de comunicações privadas devem atender à preservação da intimidade, da vida privada, da honra e da imagem das partes direta ou indiretamente envolvidas.” (FORTES, 2016, p. 129).

Fortes (2016) cita outros registros que existem nesta lei quanto à proteção de dados pessoais, como restrições quanto às finalidades para as quais os dados pessoais são coletados e não podem ser utilizados de outras formas, sem o consentimento do titular, devendo ser fornecidos em o contrato de serviço ou a *Internet*, conforme indicado nos termos de uso do aplicativo.

A Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709, de 14 de agosto de 2018, foi promulgada para proteger os direitos fundamentais de liberdade e privacidade e a livre formação da personalidade de cada pessoa. A lei trata do tratamento de dados pessoais disponibilizados em meio físico ou digital por pessoas físicas ou jurídicas de direito público ou privado, incluindo diversas operações que possam ocorrer em meio manual ou digital.

No Art. 5º da LGPD, considera-se:

I – dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

II – dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

III – dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;

IV – banco de dados: conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico;

V – titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;

VI – controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

VII – operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador; [...]

A proteção dos dados pessoais é tratada de forma mais complexa, estabelecendo novas ideias, conceitos e normas. A LGPD protege qualquer operação de tratamento e utilização de dados no Brasil. Ela regula a forma como os dados são coletados e processados, com foco maior nos meios digitais, como dados cadastrais ou até mesmo textos ou fotos publicadas em redes sociais (BALDISSERA, 2019).

Uma das inovações propostas é que tanto os dados pessoais que permitem a identificação do seu titular, a lei também regulamenta os dados isolados, dados esses que não permitem a identificação do seu titular, mas que em conjunto com outros dados pode estabelecer a identificação do seu dono (BALDISSERA, 2019).

No Brasil todas as normas estabelecidas pela LGPD devem ser aplicadas nas operações de tratamento de dados.

X – tratamento: 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;

XI – anonimização: 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;

XII – consentimento: manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma finalidade determinada;

As empresas estrangeiras do mesmo modo estão sujeitas a esta legislação independente do país em que será realizado o tratamento dos dados. Deverá o país de destino dos dados garantir a proteção compatível com esta Lei por meio de contratos ou normas corporativas (BRASIL, 2018).

O problema de regular o uso de dados pessoais a partir do conceito de privacidade, segundo Garkinfel (2000), está em tratar a privacidade como propriedade. Para o autor, ao contrário de outras formas tradicionais de propriedade, a informação possui características próprias: não se desgasta e o seu valor aumenta com o uso. Os dados podem simultaneamente serem transferidos para terceiros e continuarem vinculados ao titular, produzindo vários efeitos jurídicos nas duas instâncias.

A situação colocada por Garfinkel (2000) torna-se complexa ao analisarmos a LGPD. Conforme o inciso XII, o consentimento é a manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para

uma finalidade determinada. Ora, ao autorizar o uso de seus dados pessoais, o titular o faz em condições específicas, que deveriam ser formuladas em linguagem clara e simples. Porém, na maioria das vezes, não é assim que acontece.

Numa citação à Pamela Samuelson, professora de direito autoral da Universidade da Califórnia, nos Estados Unidos, Garfinkel (2000) nos mostra o dilema do titular de dados. É contraditório ou paradoxal concordar com o tratamento de seus dados pessoais para uma finalidade determinada, para depois se preocupar com possíveis desvios e abusos por parte do agente de tratamento desses dados. Várias questões podem ser colocadas neste dilema: quais são os mecanismos que o titular de dados possui para aferir se o agente está tratando os dados nas condições pré-estabelecidas e acordadas entre as partes? Considerando a complexidade tecnológica atual, por exemplo, nos campos da Inteligência Artificial ou da Ciência de dados, os órgãos de fiscalização possuem competências e condições suficientes para acompanhar o ritmo das empresas de alta tecnologia, como Google ou Facebook?

A LGPD, no inciso III, estabelece o conceito de dado anonimizado – dado relativo ao titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento. Este conceito, conforme Garfinkel (2000), também pode ser problemático e controverso. Citando casos de tratamento de amostras de sangue por Centros de Controle de Doenças, o autor afirma que:

[...] de acordo com as diretrizes éticas atuais, os testes que utilizam partes do corpo humano não são considerados violações dos direitos do paciente se o nome do paciente for removido das amostras. Supostamente, remover o nome de uma pessoa torna a amostra anônima (GARFINKEL, 2000, P. 188).

O problema é que algumas análises cruzam dados e podem identificar com facilidade o seu titular. As questões que acabamos de levantar reforçam a necessidade de mais discussões sobre dados e privacidade, principalmente no que tange ao tratamento.

Houve diversos casos em que a autorização do uso de dados pessoais se tornou uma questão judicial. A seguir, citamos alguns exemplos notáveis.

Caso Facebook-Cambridge Analytica: em 2018, veio à tona o envolvimento do Facebook e da empresa de consultoria política Cambridge Analytica. A Cambridge Analytica coletou ilegalmente dados pessoais de milhões de usuários do Facebook por meio de um aplicativo de teste de personalidade. Esses dados foram utilizados

para eleições eleitorais e campanhas políticas, o que levou a consideração e processos judiciais contra as empresas envolvidas.

Caso Google Street View: em 2010, foi descoberto que os carros do *Google Street View* estavam coletando dados pessoais de redes Wi-Fi desprotegidas enquanto mapeavam as ruas. Esses dados incluíam informações como senhas, histórico de navegação e e-mails. Vários países abriram inquéritos e processos contra o Google devido à violação de privacidade.

Caso da Uber: vazamento de dados de 57 milhões de usuários e motoristas: Em 2016, a Uber sofreu um vazamento de dados em larga escala, no qual informações pessoais de 57 milhões de usuários e motoristas, incluindo nomes, endereços de e-mail e números de telefone, foram comprometidas. No entanto, a Uber não divulgou o incidente publicamente até 2017. Essa falta de divulgação oportuna levou a várias ações judiciais e investigações governamentais em vários países, questionando a proteção dos dados pessoais dos usuários e a resposta da empresa ao incidente.

Caso Serasa Experian: vazamento de dados do Serasa Experian em 2021, o Serasa *Experian*, uma das maiores empresas de análise de crédito do Brasil, sofreu um vazamento que expôs informações pessoais de milhões de brasileiros. Os dados vazados incluíam nome completo, Cadastro de Pessoa Física (CPF), data de nascimento, endereço e informações de contato. Esse incidente levou a ponderar e processar contra a empresa por supostas falhas na segurança e proteção dos dados pessoais.

Caso Tribunal Superior Eleitoral (TSE): vazamento de dados do Tribunal Superior Eleitoral (TSE) em 2020, o Tribunal Superior Eleitoral do Brasil gerou um vazamento de dados que expôs informações pessoais de milhões de eleitores. O vazamento incluiu dados como nome completo, CPF, data de nascimento e filiação. Esse incidente levou a consideração e ações de processamento para apurar a responsabilidade pelo líder e a implementação de medidas de segurança mais robustas.

O Regulamento Geral de Proteção de Dados (RGPD) entrou em vigor na União Europeia em 2018 e estabeleceu regras rigorosas para a proteção de dados pessoais. Desde então, várias empresas enfrentam multas por violação de privacidade. Por exemplo, o Google recebeu uma multa de 50 milhões de euros em 2019 por falta de transparência e informações recebidas sobre o uso de dados pessoais.

Esses são apenas alguns exemplos de casos em que a autorização do uso de dados pessoais se tornou um problema judicial. A crescente conscientização sobre a privacidade e a proteção de dados levou a uma maior regulamentação e ações legais para garantir que as empresas e organizações cumpram as leis de proteção de dados.

A regulação em torno da proteção de dados reconhece o problema do extrativismo de dados, mas busca fornecer segurança jurídica por meio da liberdade contratual e da disponibilidade desses dados. No contexto brasileiro, a Lei Geral de Proteção de Dados (LGPD) representa um marco ao estabelecer a figura do titular de dados pessoais, um sujeito de direito capaz de consentir na disponibilização de seus dados comportamentais. Embora a autodeterminação informacional seja um dos pilares dessa legislação, assim como ocorre com a autonomia privada baseada na igualdade jurídica, o titular carece de condições materiais para exercer plena liberdade sobre seus dados pessoais. Sua escolha está restrita ao consentimento sobre como os dados serão cedidos aos canais de serviços digitais.

Para enfrentar esses desafios, um dos caminhos pode ser uma melhor especificação das diretrizes para coleta, processamento e armazenamento de dados pessoais, para garantir a transparência e o acompanhamento de ações dos agentes de tratamento; ampliação e reforço das ações dos órgãos controladores; e imposição de taxas e sanções para desestimular o uso indevido de dados.

3 PROCEDIMENTOS METODOLÓGICOS

O campo da Ciência da Informação, segundo Bawden e Robinson (2012), pode ser entendido a partir da sua preocupação com o estudo da informação humana registrada, com foco nos componentes da cadeia de comunicação, investigados sob a perspectiva da análise de domínio. Nesse sentido, este trabalho utiliza a análise de domínio para investigar as discussões acerca de aspectos legais e de privacidade no uso de dados na Ciência da Informação, no Brasil.

Segundo Guimarães (2014), as abordagens teóricas e aplicadas da Ciência da Informação utilizam a análise de domínio para caracterizar e avaliar a ciência, identificando as condições da sua construção e socialização.

Smiraglia (2015) apresenta a análise de domínio como paradigma metodológico para a organização do conhecimento e como conjunto de técnicas e ferramentas para identificação de bases de conhecimento específicas para os diversos setores científicos.

De acordo com Beagrie (2006, p.4), um domínio refere-se a um conjunto de conhecimentos que são compartilhados dentro de um determinado discurso.

Na Ciência da Informação, a análise de domínios de conhecimento como comunidades discursivas ganhou espaço a partir dos anos 1995, com a proposta de Hjørland e Albrechtsen.

Mais tarde, o delineamento das onze abordagens para análise de domínio, de Hjørland (2002), clarificou o quadro para sua aplicação em estudos de organização do conhecimento do campo. Essas abordagens são: 1) Produção de guias de literatura e entradas de assunto; 2) Construção de classificações especiais e tesouro; 3) Indexação e recuperação de especialidades; 4) Estudos empíricos de usuários; 5) Estudos bibliométricos 6) Estudos históricos; 7) Estudos de Documentos e Gênero; 8) Estudos Epistemológicos e Críticos; 9) Estudos Terminológicos, linguagem para propósitos especiais (na língua inglesa, *Special Purpose Languages – SPL*), bases de dados semânticas e estudos de discurso; 10) Estruturas e Instituições na Comunicação Científica, e 11) Cognição Científica, Especialistas do Conhecimento e Inteligência Artificial: (TENNIS, 2003; GUIMARÃES, 2014; HJØRLAND, 2017).

[...] a melhor forma de entender a informação em SI é estudar os domínios do conhecimento como comunidades de pensamento ou discurso, que fazem parte da divisão do trabalho da sociedade. A organização do conhecimento, a estrutura, os padrões de cooperação,

a linguagem e as formas de comunicação, os sistemas de informação e os critérios de relevância são reflexos dos objetos de trabalho dessas comunidades e de seu papel na sociedade. A psicologia, o conhecimento, as necessidades de informação e os critérios de relevância subjetiva de cada pessoa devem ser vistos nesta perspectiva.' (HJØRLAND; ALBRECHTSEN, 1995, p. 400).

Smiraglia (2015) fez um levantamento de estudos analíticos conduzidos e relatados na literatura do domínio e observou que, apesar do crescimento do paradigma de análise de domínio na organização do conhecimento ao longo do tempo, com embasamento nas onze abordagens de Hjørland, alguns contextos continuavam fora da análise. Por exemplo, guias de literatura; especialidade de indexação; e estudo de gênero e comunicação científica. Conseqüentemente, apresentou a seguinte taxonomia revisada de abordagens: 1) guias de assuntos ou guias literatura; 2) classificações especiais e tesouros; 3) estudos empíricos de usuários; 4) estudos informativos; 5) estudos históricos; 6) documentação e estudos de gênero; 7) estudos epistemológicos e críticos; 8) estudos terminológicos; 9) semântica do banco de dados; 10) análises de discurso e 11) cognição, conhecimento especializado e inteligência artificial.

Segundo Guimarães (2014), as abordagens teóricas e aplicadas da Ciência da Informação utilizam a análise de domínio para caracterizar e avaliar a ciência, identificando as condições da sua construção e socialização.

Neste estudo, a análise do domínio se concentrou na abordagem 5) Estudos históricos, conforme a classificação de Smiraglia (2015), e na abordagem 6) Estudos históricos de Hjørland (2002).

Tennis (2003) nos mostra que o principal problema no uso da análise do domínio não está no processo investigativo do objeto, mas sim na especificação da clareza do objeto a ser investigado – o domínio. Como solução, o autor propõe dispositivos analíticos – eixos, de modo a delinear o que o analista de domínio está estudando. O eixo Áreas de Modulação fornece os parâmetros sobre o nome e a extensão do domínio, respondendo à pergunta: como se chama o domínio e o que abrange? O eixo Graus de Especialização qualifica e define a intenção do domínio, diminuindo a sua extensão.

Assim, quanto ao parâmetro nomeação, o domínio da pesquisa é produção científica da Ciência da Informação. A extensão do domínio é formada por estudos sobre aspectos legais no uso de dados pessoais. Considerando a diversidade de

abordagens possíveis neste domínio, a intenção ou o foco do domínio é: autores e suas filiações institucionais; assuntos tratados considerando as fases do CVD-CI; Em seguida, foi feita uma análise histórica de tendências das discussões nos últimos dez anos dos autores que mais contribuíram com o Ciclo de Vida dos Dados.

Para o levantamento bibliográfico da produção científica, a seguinte estratégia de busca foi used: “dados pessoais” AND “ciência da informação” (“personal date” AND “information science”), in the bases de dados: Biblioteca Digital Brasileira de Teses e Dissertações do Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT (BDTD), Base de Dados de Referência de Artigos de Periódicos em Ciência da Informação (BRAPCI), *Scientific Electronic Library Online* (SciELO) Brasil, Anais do Congresso Brasileiro de Organização e Representação do Conhecimento organizado pela ISKO-Brasil; Anais do Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB.

Dos resultados obtidos, apenas as publicações que apresentaram no título, resumo e/ou foram contabilizadas as palavras-chave, o termo “dados pessoais” e que estivessem em português e de origem brasileira.

Os critérios de inclusão dos estudos foram: publicação dos últimos dez anos; publicação em português; acesso ao texto completo; o estudo aborda o tema na área de modulação e no foco do domínio.

Assim, a análise limitou-se a artigos, trabalhos apresentados em eventos científicos e trabalhos acadêmicos: teses, dissertações, trabalhos de conclusão de curso e relatórios de iniciação científica. Publicações até dezembro de 2022.

As publicações selecionadas para compor o corpus de análise foram sistematizados em tabela, computando as informações: autor; título; ano de publicação; filiação institucional; fase CVD; fatores norteadores; assuntos tratados; base de dados.

As áreas de aplicação do estudo foram especificadas a partir das etapas do ciclo de vida de dados: coleta, armazenamento, recuperação e descarte (SANT’ANA, 2016).

4 RESULTADOS E DISCUSSÃO

O objetivo de analisar os estudos sobre dados e privacidade na Ciência da Informação, no Brasil, nos últimos dez anos, utilizando a análise de domínio e o modelo de ciclo de vida de dados foi alcançado utilizando como método o estudo qualitativo e quantitativo, tendo a literatura disponível como subsídio para a definição do processo de análise de domínio. Foi feita a coleta e categorização dos estudos, com a devida identificação de informações sobre os autores, suas respectivas instituições, assuntos tratados, respectivas teorias de sustentação e áreas de aplicação de estudos.

A partir da compilação de dados, os trabalhos foram categorizados e a geração de gráficos ocorreu com informações de distribuição temporal de autores e suas filiações institucionais; área de aplicação do estudo de acordo com o CVD - CI; Em seguida, foi feita uma análise histórica de tendências das discussões nos últimos dez anos dos dez autores mais colaborativos com o CVD - CI.

O dado pessoal é coletado para atender uma finalidade específica. A Lei Geral de Proteção de Dados Pessoais (LGPD) considera como tratamento:

Art. 5º Para os fins desta Lei, considera-se:

X - Tratamento: 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 (LGPD, art. 5, X);

A Lei Geral de Proteção de Dados Pessoais (LGPD) e a Lei de Acesso à informação (LAI) igualmente apresentam regras específicas para o acesso a documentos de interesse público e que em sua grande maioria contém também dados pessoais.

Neste processo, aprofundou-se a compreensão acerca de cada categoria com a análise de domínio, realizada a partir da proposta de Sant'Ana (2016, p.6)

A Ciência da Informação pode e deve contribuir para que este cenário de acesso e uso intenso de dados se desenvolva da melhor maneira possível, buscando identificar e estudar fatores e características que propiciem ampliação do equilíbrio entre os atores envolvidos no processo e a máxima otimização do uso dos dados. Nesta tarefa se faz necessário estruturar esta análise, e para tanto, propõe-se a utilização de uma delimitação de fases (momentos em que distintas necessidades e competências são necessárias) envolvidas no acesso

e uso dos dados, mantendo-se como ponto central os próprios dados e para tanto se propõe o uso do Ciclo de Vida dos Dados - CVD como forma de evidenciar os diferentes momentos e fatores envolvidos neste processo (SANT'ANA, 2016, p.6)

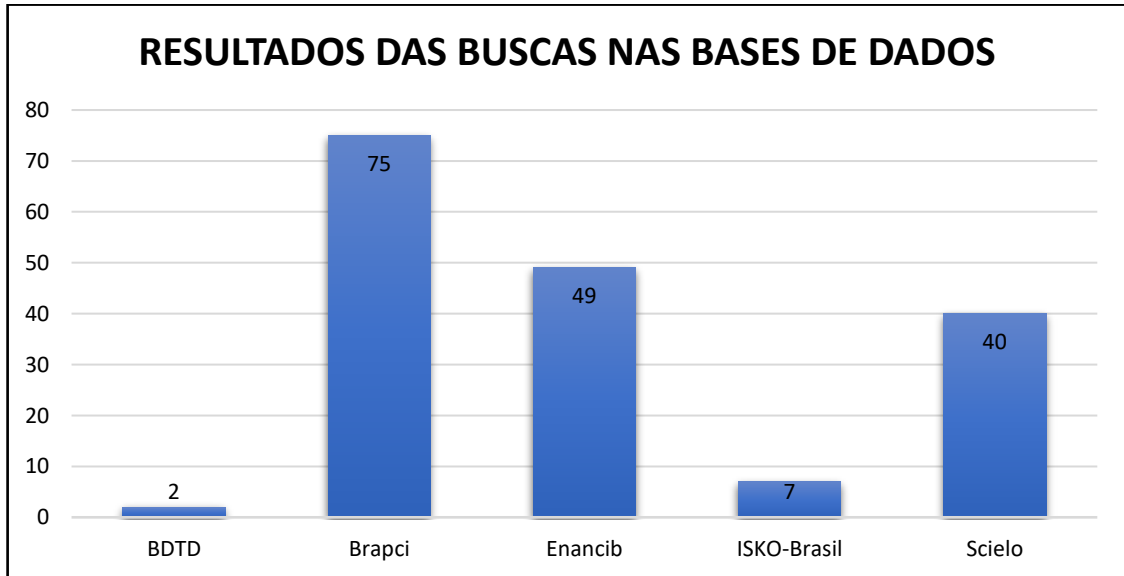
Sant'Ana (2016, p.8) categoriza o acesso e o uso de dados em quatro fases, sendo elas: Coleta, Armazenamento, Recuperação e Descarte. Além de explicitar os fatores que estão presentes em cada uma destas fases: Privacidade, Integração, Qualidade, Direitos Autorais, Disseminação e Preservação (SANT'ANA, 2013).

Para orientar a prática do tratamento e apresentar os ativos institucionais envolvidos, o presente trabalho abordará o ciclo de vida do tratamento dos dados pessoais em cinco fases: coleta, armazenamento, processamento, compartilhamento e descarte. Com relação aos fatores que estão presentes em cada uma destas fases será utilizado: Privacidade, Integração, Qualidade, Direitos Autorais, Disseminação e Preservação (SANT'ANA, 2013).

A partir da análise dos trabalhos apresentados nas Bases de Dados BRAPCI, *Scielo*, BDTD, ISKO-Brasil e ENANCIB na forma de Comunicações Orais e de Pôsteres, ao longo da trajetória dos últimos dez anos (2012-2022). Do total de publicações recuperadas (350), buscou-se identificar aqueles que tiveram os termos privacidade; dados pessoais; lei de acesso à informação e lei geral de proteção de dados pessoais, no título, resumo e/ou palavras-chave. Dos resultados obtidos das Bases de Dados apenas 173 artigos estavam de acordo com os critérios estabelecidos para fazer parte da análise da pesquisa.

Na base de dados BRAPCI foram recuperados 75 artigos, seguidos da ENANCIB com 49 artigos, *Scielo* com 40 artigos, ISKO-Brasil com 7 e a BDTD com 2 artigos recuperados.

Gráfico 1 - Resultados das Buscas nas Bases de Dados

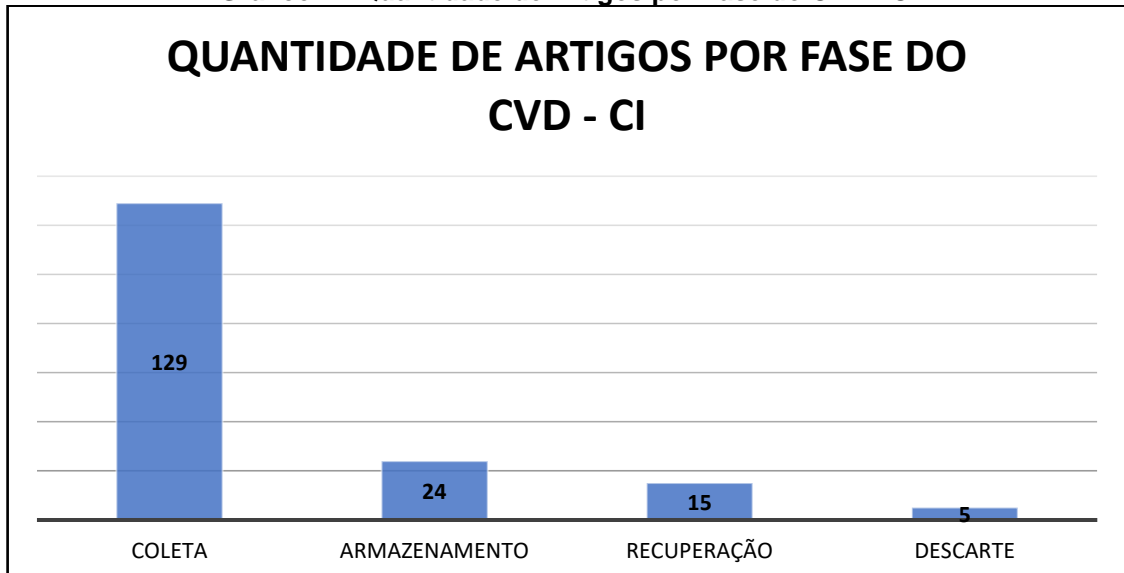


Fonte: elaboração própria

Assim, para este estudo, considerando a inclusão definida e critérios de exclusão na metodologia, foram analisadas 173 publicações. A análise limitou-se a artigos, trabalhos apresentados em eventos científicos e trabalhos acadêmicos.

O Gráfico 2 apresenta a distribuição dos artigos nas categorias em estudo.

Gráfico 2 - Quantidade de Artigos por Fase do CVD - CI



Fonte: elaboração própria

A partir da compilação dos dados coletados, foi possível identificar que 75% dos trabalhos dizem respeito a fase coleta de dados, seguidos das fases de armazenamento com 14%, recuperação com 8,7% e descarte com 2,8%.

A fase coleta de dados pode ser encontrada como tema de estudo na área de Ciência da Informação em uma publicação na décima terceira edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XIII ENANCIB 2012), baseada no artigo de Vidotti (2012) sobre como os órgãos governamentais através de tecnologias realizam a coleta dos dados para investigar de que modo o *Second Life* (SL) pode potencializar o acesso, o uso e a disseminação de dados públicos para a promoção da transparência pública. A fase coleta de dados reaparece na publicação de Ventura (2012), na décima terceira edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XIII ENANCIB 2012), intitulado Acesso à informação pública de qualidade: estudo da aplicação da Lei nº 12.527/2011 pelas Instituições Federais de Ensino Superior da região nordeste. A partir de 2013, o uso da fase coleta de dados do ciclo de vida de dados do CVD-CI torna-se mais recorrente, principalmente em sua dimensão digital, experimentando um aumento significativo a partir de 2015.

A fase armazenamento pode ser encontrada como tema de estudo no artigo Requisitos Funcionais para Dados Imagéticos Digitais (RFDID): um modelo conceitual de entidades e relacionamentos das autoras Simionato e Santos (2013), apresentado na décima quarta edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XIV ENANCIB 2013) que. O artigo trata sobre como o ambiente digital oferece a possibilidades de armazenamento e acesso instantâneos às imagens capturadas a todo o momento por diversos dispositivos, como celulares e câmeras; No entanto, enfrentar o desafio de transformar a enorme quantidade de dados disponíveis diariamente em informações consistentes, que asseguram uma economia no processo de acesso e recuperação da informação, é fundamental. A fase armazenamento aparece no artigo de Santarém Segundo (2015), intitulado Web semântica: introdução a recursos de visualização de dados em formato gráfico, publicado na décima sexta edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XVI ENANCIB 2015). O artigo trata sobre como a Web Semântica tem provocado mudanças na estrutura e ligação entre dados disponibilizados na Internet, a fim de permitir que estes dados atendam de forma mais adequada às necessidades dos usuários. Grande parte dos estudos está diretamente ligada à estrutura e armazenamento dos dados em formato semântico, entretanto, posteriormente a este processo, a apresentação destas informações para os usuários é deficitária.

A fase recuperação é tratado no artigo publicado na Revista Digital de Biblioteconomia e Ciência da Informação (RDBCI) pelos autores Lima e Campos (2022). O artigo é intitulado *Acessibilidade web na perspectiva dos sistemas de recuperação da informação*, onde os autores abordam que para o efetivo desempenho de sistema de recuperação da informação aconteça é necessário a qualidade com a qual a organização da informação é realizada, o que implicará em uma recuperação da informação mais relevante e pertinente, visto que esses procedimentos são condicionados um ao outro fazendo uma ponte entre a entrada e a saída da informação. A fase recuperação reaparece no artigo de Modesto, Ferneda e Sant'Ana (2013) intitulado *Proposta de metadados para descrição de recursos informacionais digitais criados dinamicamente*. O artigo apresentado na décima quarta edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XIV ENANCIB 2013) destaca a necessidade de propor um modelo de representação descritiva de Recursos Informacionais Digitais criados dinamicamente disponibilizados pelos governos em seus sítios oficiais usando metadados que possam ser utilizados em processos de recuperação. Desta forma, o cidadão poderá localizar o ponto exato em que é gerada uma tabela de resultado de uma determinada consulta sem passar pela estrutura (caminho) que hoje é hierarquizado. Vale lembrar que o foco deste trabalho está na recuperação e não na preservação do recurso.

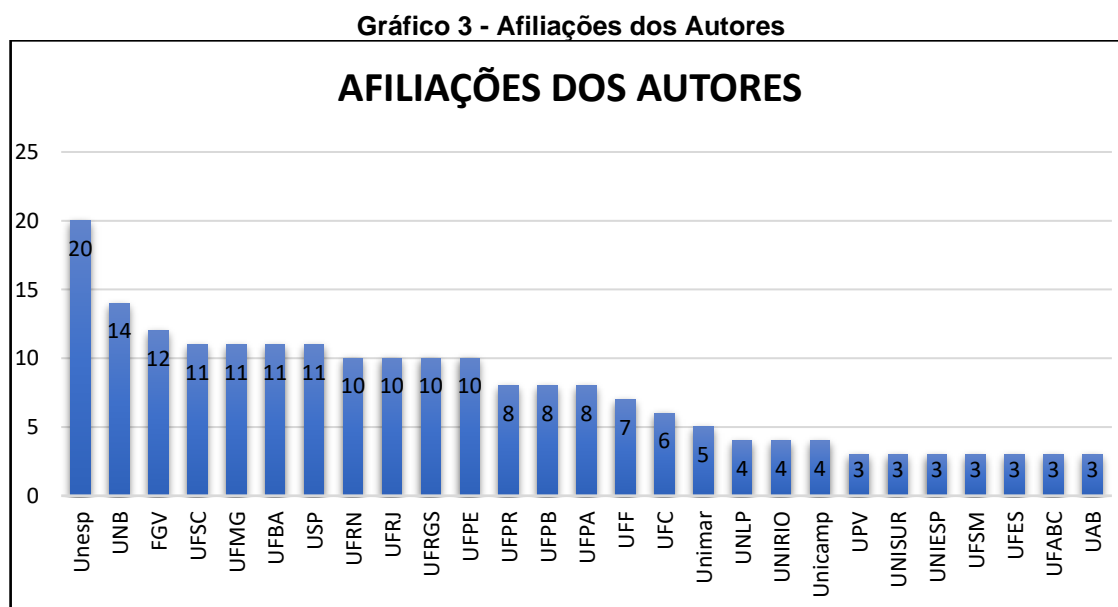
A fase descarte é tratado pelos autores Silva e Araújo (2014) no artigo da *Modelo para o descarte seguro da informação em suporte digital*. Os autores tratam como o descarte de uma informação (dado) é de suma importância e que ao descartar uma informação é necessário garantir que não haverá como esta informação ser recuperada novamente. Inclusive os autores ainda ressaltam a lacuna sobre este tema na Ciência da Informação.

Desta forma, a análise de domínio foi focada na Fase Coleta de dados, onde será realizada a geração de gráficos com informações de distribuição temporal, de autores e suas filiações institucionais; assuntos tratados considerando as fases do CVD-CI; Em seguida, foi feita uma análise histórica de tendências das discussões nos últimos dez anos dos autores que mais contribuíram com o Ciclo de Vida dos Dados.

A Categoria Coleta de Dados conta com duzentos e quarenta e três autores distribuídos por cinquenta e cinco instituições. O Gráfico 3 demonstra as Instituições que mais possuem autores afiliados que abordam sobre o tema são Universidade

Estadual Paulista (UNESP) com 20 autores afiliados, seguida da Universidade de Brasília (UNB) com 14 autores, a Fundação Getúlio Vargas (FGV) com 12 autores afiliados.

Já a Universidade Federal de Santa Catarina (UFSC), Universidade Federal de Minas Gerais (UFMG), Universidade Federal da Bahia (UFBA) possuem 11 autores afiliados por instituições, seguidas da Universidade de São Paulo (USP), Universidade Federal do Rio Grande do Norte (UFRN), Universidade Federal do Rio de Janeiro (UFRJ), Universidade Federal do Rio Grande do Sul (UFRGS), Universidade Federal de Pernambuco (UFPE) com 10 autores afiliados por instituições e as demais seguem com oito ou menos autores por instituições.



Fonte: elaboração própria

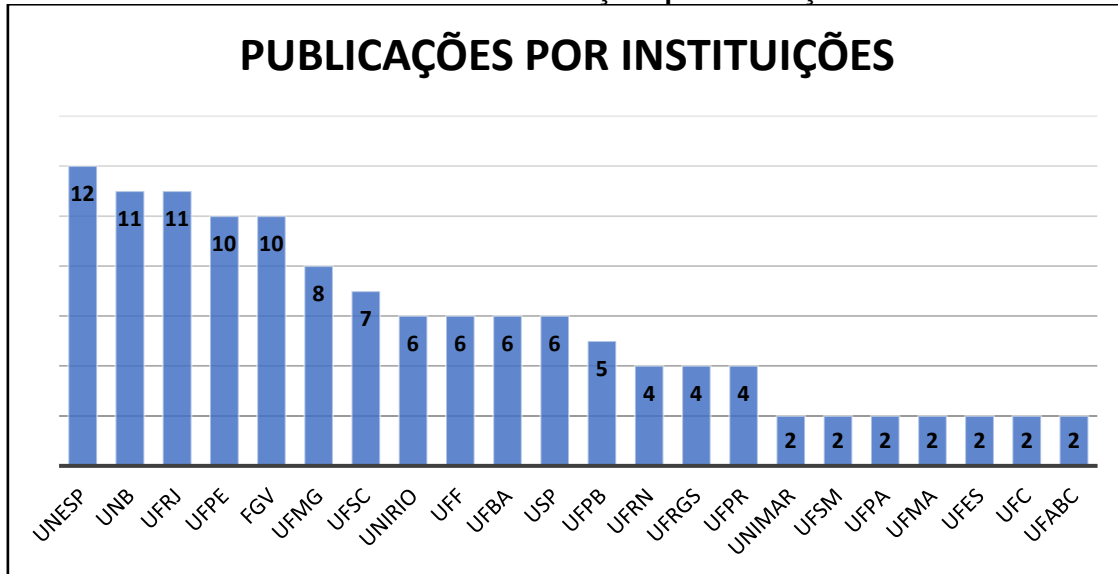
Com as informações coletadas nas bases de dados, cinquenta e cinco instituições tiveram publicações referentes a Coleta de Dados.

No Gráfico 4 está representado apenas as instituições que tiveram duas ou mais publicações sobre o tema.

Nota-se que a Unesp teve 12 publicações sobre a Fase da Coleta de Dados, seguidas das instituições Universidade de Brasília (UnB) e Universidade Federal do Rio de Janeiro (UFRJ) com 11 publicações cada uma, já a Universidade Federal de Pernambuco (UFPE) e a Fundação Getúlio Vargas (FGV) contribuíram

com 10 publicações cada uma. As demais instituições tiveram uma variação de oito a dois artigos.

Gráfico 4 - Publicações por Instituições

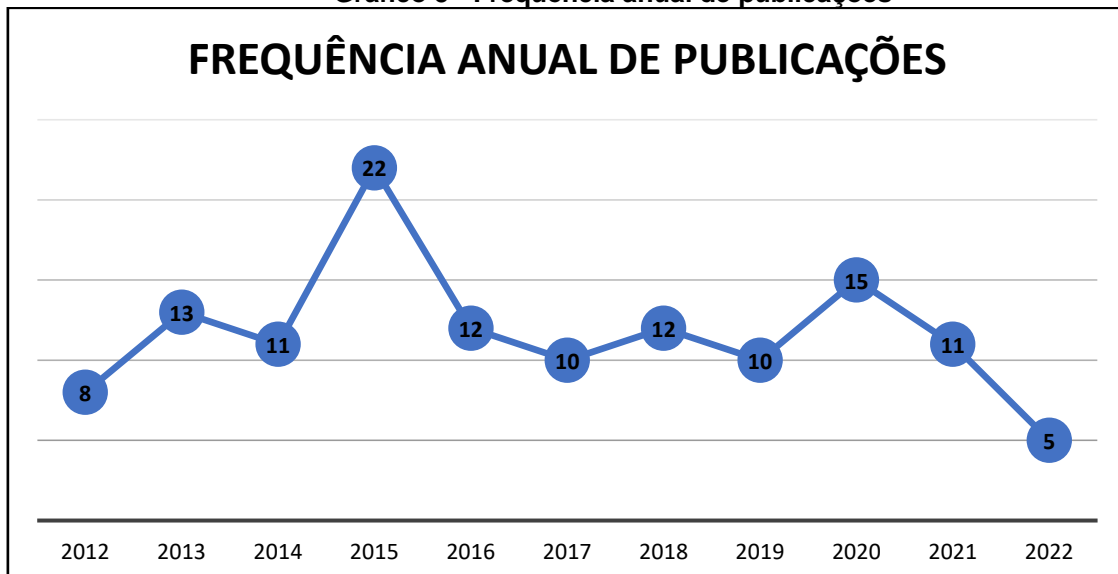


Fonte: elaboração própria

Em comunidades acadêmicas como as de Ciências Sociais e Humanas, a análise de citações fornece vestígios de evidências de relações temáticas e teóricas. Nesse contexto, os dados das publicações foram organizados de forma que fosse possível apresentar padrões de reconhecimento explícito de cada publicação.

Assim, quantitativamente, o Gráfico 5 mostra a frequência de publicações segundo o ano.

Gráfico 5 - Frequência anual de publicações



Fonte: elaboração própria

Os últimos anos testemunharam uma preocupação crescente com os dados e a privacidade. À medida que as más práticas de dados vêm à tona, os cidadãos estão começando a entender o custo real do uso de tecnologias digitais online. Acertar na privacidade dos dados é um dos maiores desafios desta nova década do século XXI.

Após a regulamentação da Lei de Acesso à Informação (LAI) nº 12.527 que entrou em vigor no dia 16 de maio de 2012 que tem como objetivo promover a transparência e a prestação de contas dos órgãos públicos, além de determinar que o tratamento das informações pessoais detidas por entidades e instituições nela abrangidas seja realizado de modo transparente, respeitando o direito fundamental à proteção da intimidade, da vida privada, da honra e da imagem das pessoas, o que, nos fundamentos defendidos nesta obra, corresponde à proteção do direito fundamental à privacidade.

Nota-se um aumento significativo das publicações com relação a fase Coleta de Dados referente aos anos de 2013 até 2017. Uma vez que na fase de Coleta de Dados, de acordo com Sant'Ana(2016, p.121) a fase de obtenção dos dados que pode ser identificada como aquela em que ocorrem: a definição das necessidades informacionais que irão nortear as escolhas e definições sobre quais dados são necessários; são estabelecidas estratégias sobre como localizar e avaliar estes dados; são escolhidos os mecanismos que serão utilizados para sua obtenção, e; são elaboradas as metodologias e ferramentas necessárias para consecução destes dados.

No ano de 2018 é regulamentada a Lei Geral de Proteção de Dados (LGPD), legislação brasileira que visa regulamentar o tratamento de dados pessoais no país. A LGPD tem como objetivo principal proteger os direitos fundamentais de privacidade e garantir o controle dos indivíduos sobre suas informações pessoais. Ela estabelece diretrizes claras para a coleta, o armazenamento, o uso, o compartilhamento e a transferência de dados pessoais por parte de organizações públicas e privadas.

Percebe-se que após a regulamentação da LGPD o número de publicações sobre a fase de Coleta da Dados do Ciclo de Vida de Dados passa a ter um declínio, com exceção do ano de 2020 que contou com 15 publicações. Acredita-se que o declínio de publicações se dá pelo motivo de a partir de 2018 haver uma lei para

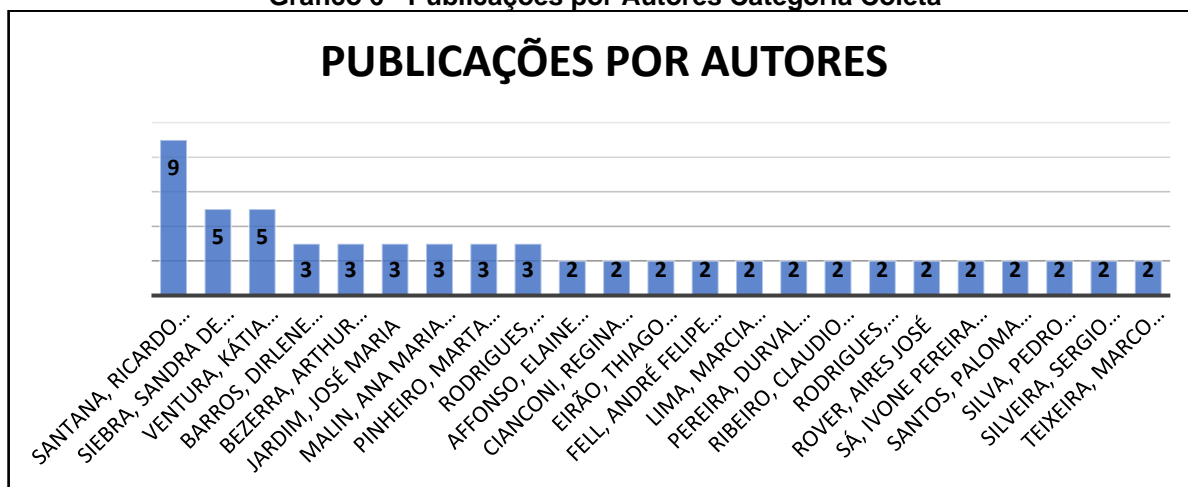
proteger os direitos fundamentais de privacidade e garantir o controle dos indivíduos sobre suas informações pessoais.

A categoria Coleta, conforme Gráfico 6, foi concebida abarcando estudos de análise, estudo comparado, relato de experiência, estudo de caso. Esta categoria conta com 129 artigos escritos por 238 autores.

A apresentação de citações de trabalhos dessa categoria demonstra como se deu a identificação na categoria Coleta;

[...] Em um primeiro momento, depara-se com a fase de obtenção dos dados que pode ser identificada como aquela em que ocorrem: a definição das necessidades informacionais que irão nortear as escolhas e definições sobre quais dados são necessários; são estabelecidas estratégias sobre como localizar e avaliar estes dados; são escolhidos os mecanismos que serão utilizados para sua obtenção, e; são elaboradas as metodologias e ferramentas necessárias para consecução destes dados (SANT'ANA , 2016, p.9).
 [...] A coleta é a obtenção, recepção ou produção de dados pessoais independente do meio utilizados (documento papel, documento eletrônico, sistema de informação etc.) (SEF, 2021, p.2).

Gráfico 6 - Publicações por Autores Categoria Coleta



Fonte: elaboração própria

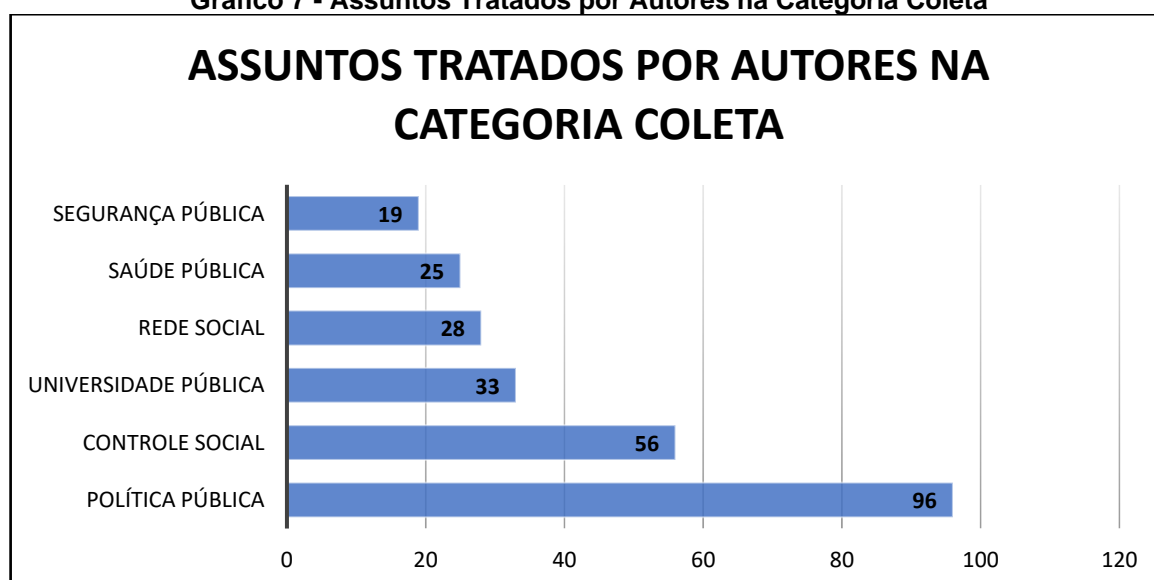
[...] Em um primeiro momento, depara-se com a fase de obtenção dos dados que pode ser identificada como aquela em que ocorrem: a definição das necessidades informacionais que irão nortear as escolhas e definições sobre quais dados são necessários; são estabelecidas estratégias sobre como localizar e avaliar estes dados; são escolhidos os mecanismos que serão utilizados para sua obtenção, e; são elaboradas as metodologias e ferramentas necessárias para consecução destes dados (SANT'ANA , 2016, p.9).
 [...] A coleta é a obtenção, recepção ou produção de dados pessoais independente do meio utilizados (documento papel, documento eletrônico, sistema de informação etc.) (SEF, 2021, p.2).

A apresentação de citações de trabalhos dessa categoria demonstra como se deu a identificação na categoria Coleta;

O Gráfico 6 traz o *ranking* dos 23 autores que mais publicaram sobre a Categoria Coleta. O primeiro lugar é liderado pelo autor SANT'ANA, R. com nove publicações, o segundo lugar é liderado pelas autoras SIEBRA, S. e VENTURA, K. com cinco publicações sobre o tema Coleta de Dados. Já o terceiro lugar é ocupado pelos autores BARROS, D., BEZERRA, A., JARDIM, J.M., MALIN, A.M., PINHEIRO, M. e RODRIGUES, F.A., com três publicações por autores.

Conforme o Gráfico 7, nos últimos dez anos, ocorreu um aumento significativo no número de autores dedicados à exploração de diversos recursos relevantes principalmente no campo da Política Pública, onde 96 autores se dedicaram a abordar o assunto, enquanto 56 autores se concentraram em pesquisas sobre Controle Social. Além disso, 28 autores se aprofundaram na temática da Rede Social, buscando compreender suas explicações e efeitos. Outros temas considerados incluem Segurança Pública, com 19 autores confiantes com seu conhecimento, Saúde Pública, com 25 autores e Universidade Pública, que atraiu a atenção de 33 investigadores. Essa diversidade de autores e áreas de estudo reflete a importância e a complexidade desses assuntos nos últimos dez anos.

Gráfico 7 - Assuntos Tratados por Autores na Categoria Coleta



Fonte: elaboração própria

O assunto Política Pública pode ser encontrado como tema de estudo na área de Ciência da Informação na publicação dos autores Santarém Segundo et al. (2020)

na Revista *Perspectivas em Ciência da Informação*, v.25, número 2, de 2020. O artigo *A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet* traz como o uso da Internet para acessar conteúdos gerados com a operacionalização de programas governamentais é crescente e gera possibilidades para o cidadão acompanhar a execução das políticas públicas. No contexto de políticas públicas para a agricultura, essas constituem um fator preponderante no desenvolvimento social e econômico para os pequenos produtores, que por sua vez necessitam de dados e informações para usufruírem destas ações. O Governo é responsável por garantir a ampla divulgação dos programas, contudo, observa-se que muitas vezes os dados não são disponibilizados como deveriam, tendo principalmente problemas relacionados à qualidade; os autores Bianchini et al. (2021) publicaram na Revista da Avaliação da Educação Superior (Campinas), v. 26, n. 3, de 2022, o artigo *Política pública de interiorização e de ação afirmativa: uma possibilidade de ampliação de liberdades instrumentais e substantivas da população residente no litoral do Paraná*, o mesmo tem como propósito refletir sobre o acesso da população local à educação superior por meio da existência de uma instituição no local e essa com Ações Afirmativas, a luz do desenvolvimento humano de Amartya Sen. A UFPR no Litoral do Paraná trouxe oportunidades de acesso à educação superior para a população local, possibilitando a ampliação das liberdades instrumentais desta população, e possivelmente por meio da educação superior, as liberdades substantivas.

O assunto Controle Social pode ser encontrado como tema de estudo na publicação dos autores Melo e Fuchigami (2019) no artigo *Proposta de Índice Bidimensional de Transparência da Informação Público-Eletrônica como Ferramenta para Participação e Controle Sociais*. A publicação foi realizada na Revista *Eletrônica de Administração* (Porto Alegre), v. 25 de 2019. O artigo aborda como a Lei de Acesso à Informação (LAI) reafirma a importância da transparência como elemento fundamental para a confiança e aprovação das ações concedidas. No entanto, quando as informações em sites públicos são indisponíveis, incompletas ou de difícil compreensão, isso compromete a implementação efetiva da LAI e a promoção da transmissão. Através dos índices obtidos é permitido verificar com alta precisão a qualidade das informações nos sites, revelando que elas são pouco transparentes e insuficientes para serem compreendidas pelos cidadãos. Isso sugere um baixo

engajamento tanto dos agentes públicos na aplicação da lei quanto dos stakeholders no exercício do controle social. Portanto, além de um índice confiável de transparência, a conscientização, a união e a atuação social são fundamentais para promover a participação e o controle social; os autores Gonçalves e Varella (2018) também abordam sobre o assunto Controle Social em uma publicação para a Revista Direito GV, v. 14, de 2018, intitulada Os desafios da Administração Pública na disponibilização de dados sensíveis. O artigo relata que no contexto da Lei de Acesso à Informação – em que a publicidade é a regra e o sigilo a exceção –, do Marco Civil da Internet e de decretos como o que trata da interoperabilidade de bases de dados, um dos desafios da Administração Pública brasileira é lidar com as bases de dados que contêm dados sensíveis. Nesse cenário, buscou analisar o modelo de disponibilização de bases de dados em ambientes denominados seguros. Foram examinados os modelos do Instituto Brasileiro de Geografia e Estatística (IBGE), do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e do Instituto de Pesquisa Econômica Aplicada (Ipea), apontando suas principais semelhanças e diferenças.

O assunto Universidade Pública é tratado no artigo Transparência e acesso à informação: um estudo da demanda por informações contábeis nas universidades federais brasileiras, dos autores Rezende e Rodrigues (2016). A publicação trata sobre como a Lei de Acesso à Informação e o Brasil vivenciam o início da implementação de uma política de acesso às informações públicas, instrumento de uma ampla política nacional com foco na transparência e no controle social dos atos públicos. As universidades federais brasileiras, como entidades públicas, estão obrigadas a cumprir as regras de transparência e prestação de contas. O trabalho apresenta resultados parciais de uma pesquisa de doutorado que, dentre outros objetivos, busca verificar o comportamento da demanda por informações contábeis nas universidades públicas federais após a edição da Lei de Acesso à Informação, compreendendo uma amostra de 59 universidades, nos dois primeiros anos de implementação da lei; Souza e Alvarenga (2014) também abordam sobre o assunto Universidade Pública no Encontros Bibli da revista eletrônica de biblioteconomia e ciência da informação, v.19 de 2014. O artigo intitulado A Universidade Federal de Minas Gerais no contexto do acesso aberto à informação científica: identificação de seus sistemas de informação aborda sobre os obstáculos enfrentados pela

comunidade científica na disseminação e na garantia de acesso às próprias produções, contextualizados na necessidade de acesso aberto à informação científica, impulsionaram a criação dos Repositórios Institucionais - RIs. Trata-se de tecnologias adotadas por instituições de ensino e de pesquisa que visam gerir e disponibilizar a produção científica local. A pesquisa teve como objetivo geral caracterizar sistemas de informação da Universidade Federal de Minas Gerais analisando a percepção dos atores responsáveis por sua existência/manutenção dentro do contexto do acesso aberto à informação científica.

O assunto Rede Social é tratado pelo autor Bezerra (2015) na décima sexta edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XVI ENANCIB 2015) em seu artigo intitulado Vigilância e filtragem de conteúdo nas redes digitais: desafios para a competência crítica em informação. No artigo serão abordados os fenômenos de vigilância e monitoramento de dados pessoais na internet e a filtragem de informação em plataformas digitais – como redes sociais e mecanismos de busca – que ocorre por meio de algoritmos que tendem a reificar a experiência dos usuários a partir de seu histórico de navegação e de seu comportamento prévio na rede; os autores Oliveira e Raminelli (2014) também abordam sobre o tema no artigo O direito ao acesso à informação na construção da democracia participativa: uma análise da página do conselho nacional de justiça no Facebook. A publicação fala sobre o direito à informação que ganhou importância dentro da sociedade informacional, sendo que o Brasil, apesar de expressamente consagrá-lo na Constituição Federal de 1988, apenas em 2011 promulgou sua Lei de Acesso à Informação, com o objetivo de aumentar a transparência da Administração Pública. Esse caminho para a transparência governamental é uma tendência ampliada diante do conceito de governo eletrônico, que, além de modernizar os serviços prestados, proporciona maior interação e democratização na relação entre o cidadão e o Governo;

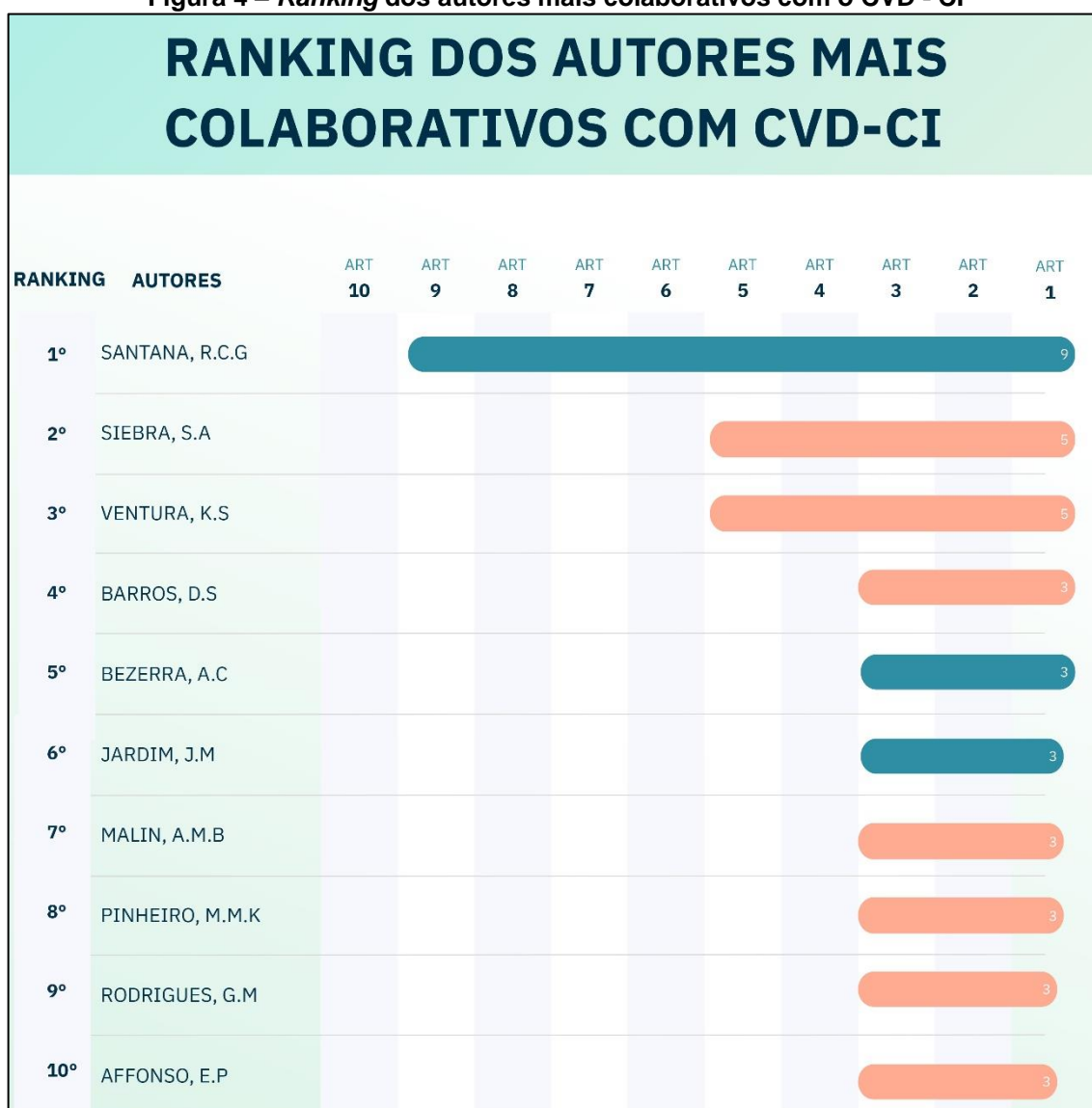
Os autores Assis e Sant’Ana (2012) abordam o assunto Saúde Pública na décima terceira edição da *Encontro Nacional de Pesquisa em Ciência da Informação* (XIII ENANCIB 2012) na publicação intitulada Restrições tecnológicas e de acesso a dados disponíveis sobre destinos de repasses financeiros federais para a saúde pública em ambientes informacionais digitais. O artigo ressalta que a criação e o uso de ambientes democráticos permitem a sociedade participar, ainda que indiretamente, nas discussões sobre serviços prestados pelo Estado. Nos últimos anos, a discussão

da qualidade dos serviços públicos em saúde ganhou ênfase pelos meios de comunicação. Em 2011, o acesso a dados governamentais tornou-se obrigatório com a sanção da “Lei de Acesso à Informação”. Entretanto, sítios oficiais do Governo Federal, que contém recursos informacionais com conjunto de dados sobre destinos de repasses de recursos financeiros para a saúde pública, possuem interfaces heterogêneas, restrições de acesso e restrições tecnológicas, formando uma barreira para o cidadão na recuperação dos dados; o assunto Saúde Pública também é tratado por Santos et al. (2015) no artigo Conselhos de saúde e a publicização dos instrumentos de gestão do SUS: uma análise dos portais das capitais brasileiras. A publicação pode ser lida na Revista Ciência & Saúde Coletiva, v. 25 de 2020. O artigo tem como objetivo demonstrar que o acesso à informação é uma condição imprescindível para que não só conselheiros, mas a sociedade civil como um todo, possa propor, monitorar e avaliar as ações empreendidas na área da saúde. A partir desse entendimento e da orientação advinda da Lei 141/2012 sobre dar visibilidade aos instrumentos de gestão do SUS. O presente estudo buscou verificar como os portais das prefeituras das capitais brasileiras têm divulgado seus conselhos de saúde e os instrumentos necessários para a análise, o monitoramento e o acompanhamento da política de saúde.

Rangel e Fernandes (2016) tratam sobre o assunto Segurança Pública no artigo intitulado Os impactos da publicação de dados de agentes públicos para a segurança nacional e segurança pública. O artigo analisa os fatores de riscos expostos na comunicação entre o governo e a sociedade a partir da publicação dos dados e informações de agentes da administração pública federal que atuam na segurança nacional e segurança pública. Participaram da investigação sete agentes públicos que atuam nas atividades relativas à segurança nacional e vinte e sete que atuam na segurança pública. O estudo revelou a facilidade de obtenção de dados pessoais por qualquer pessoa, utilizando o método proposto no estudo, que não requer o emprego de ferramentas automatizadas. Os resultados apresentados são preocupantes e demonstram a fragilidade do Estado brasileiro com a publicação de dados pessoais dos seus agentes públicos e a inexistência de orientações que estabeleça regras claras e objetivas sobre publicação de dados pessoais nos diversos portais de governo que promovem a transparência pública; o autor Macedo (2014) também trabalha o assunto Segurança Pública em seu artigo Apontamentos para a

política de informação na área de defesa cibernética no Brasil, onde ele traça um histórico das relações entre a noção de informação, na Ciência da Informação, e as estratégias para a defesa e a segurança militares, na perspectiva da Ciência Política. A metodologia deverá ser de natureza qualitativa, valorizando-se a descrição histórica do processo de sistematização da informação científica e tecnológica como parte integrante dos objetivos estratégicos dos Estados, a partir dos pós- 2ª Guerra Mundial, focando no desenvolvimento da infraestrutura militar, até o advento da concepção de “ciberguerra”. Enfocar, nesse contexto, a defesa cibernética como um dos três setores estratégicos sobre os quais se estabelece a Estratégia Nacional de Defesa do Brasil, a saber: o espacial, o cibernético, e o nuclear.

A figura 5 é constituída com base na análise das informações coletadas nas Bases de Dados, foi possível criar um *ranking* dos dez principais autores que mais publicaram artigos durante a fase de coleta, abordando temas como políticas públicas, saúde pública, universidade pública, sociedade e jornalismo.

Figura 4 – *Ranking* dos autores mais colaborativos com o CVD - CI

Fonte: elaboração própria

No *ranking* dos dez principais autores que mais publicaram artigos relativos à fase de coleta, o primeiro lugar é ocupado por Sant'Ana, R.C.G com nove publicações, sendo que destas, cinco tratam sobre políticas públicas. O autor traz sua preocupação com problemas que afetam a sociedade como um todo. O autor levanta ações do regime público frente a situações socialmente problemáticas e busca soluções para essas questões, ou pelo menos propõe ideias de tratá-las. Ainda encontramos em suas publicações problemáticas como Rede Social, Saúde Pública, Universidade Pública e Economia. A metodologia trabalhada na maioria de suas obras são pesquisas bibliográficas, investigações teóricas e análise de conteúdo.

O segundo e terceiro lugar são ocupados com cinco publicações pelas autoras SIEBRA e VENTURA. Ambas costumam tratar sobre o tema Políticas Públicas, visando

contribuir para a ampliação do acesso à informação pelos cidadãos brasileiros, independente de limitações, por meio da análise e discussão da relação entre a promoção do acesso à informação e a acessibilidade nos espaços físicos dos Serviços de Informação ao Cidadão. Em termos metodológicos, trata-se de uma pesquisa quali-quantitativa, do tipo bibliográfica, descritiva e aplicada.

A autora Barros, com três publicações, ocupa o quarto lugar no *ranking*. Seu foco principal está nas Políticas Públicas do Nordeste. Após a aprovação da Lei de Acesso à Informação (LAI), nº 12.527/2011 (BRASIL. Presidência da República, 2011a), a autora investiga, em suas pesquisas, a percepção de resistência à regulamentação do acesso à informação por parte dos estados da federação brasileira. As metodologias utilizadas são o levantamento bibliográfico e documental, com abordagem qualitativa e descritiva.

O autor Bezerra, com três publicações, ocupa a quinta posição, abordando em seus textos temas como a Sociedade, as Redes Sociais e a Segurança Pública. Ao incorporar a contribuição da teoria crítica em seus estudos, ele busca um pensamento embasado em fundamentos sociológicos que consideram os aspectos políticos e psicológicos do atual regime de informação. Além disso, ele examina as práticas difusas de vigilância informacional realizadas por instituições estatais e grandes corporações do século XXI, confrontando-se com questões e dilemas relacionados à informação em rede, incluindo várias formas de opacidade, a privacidade dos indivíduos e os sistemas de governança. A metodologia utilizada é uma revisão bibliográfica.

A sexta posição é do autor Jardim, com três publicações, onde no âmbito das Políticas Públicas, o autor realiza uma análise do contexto histórico-social em que surge a Lei de Acesso à Informação Pública (Lei nº 12.527, de 18 de novembro de 2011). Seus estudos abrangem investigações detalhadas das características desse marco legal, bem como sua regulamentação nos Poderes Executivos Federais e nos estados de Minas Gerais, Paraná, Rio de Janeiro, Rio Grande do Sul e São Paulo. Essas regiões são o universo empírico que embasam as pesquisas, permitindo uma compreensão mais aprofundada do impacto e da implementação da referida lei em cada contexto específico. A metodologia trabalhada é a Revisão Bibliográfica.

A autora Malin ocupa a sétima posição com três artigos que abordam o tema de Políticas Públicas. Em seus trabalhos, ela analisa a publicação da Lei de Acesso à Informação (Lei nº 12.527), destacando sua importância para a consolidação

democrática do Brasil e o avanço das medidas de combate à corrupção no país. Ao possibilitar uma maior participação popular e o controle social das ações governamentais, a autora ressalta que o acesso da sociedade às informações públicas contribui para uma gestão pública mais eficiente. Além de gerar resultados imediatos, a nova Lei de Acesso à Informação no Brasil representa um marco para um regime de transparência pública e democratização. A metodologia adotada para as pesquisas consiste em revisões bibliográficas e pesquisas sobre o tema.

Na oitava posição, a autora Pinheiro tem três publicações, abordando o tema de Políticas Públicas e Universidades Públicas. Em suas pesquisas, ela se concentra nos movimentos de governo aberto de dados, que flutuam discretamente no Brasil desde 2005. No entanto, foi somente após a promulgação da Lei de Acesso à Informação Pública, Lei nº 12.527, em 18 de novembro de 2011, que regulamenta o acesso a dados e informações detidas pelo governo, que se iniciou a disponibilização de dados governamentais abertos. As universidades públicas federais, como órgãos integrantes da administração direta do poder Executivo, autarquias ou fundações públicas, estão sujeitas a essa lei e representam espaços reservados para inovação, criação e disseminação do conhecimento. O trabalho empregou uma metodologia baseada em revisão bibliográfica,

O autor Rodrigues, com três publicações, ocupa a nona posição e concentra seus trabalhos nas Políticas Públicas e Universidades Públicas. Em suas publicações, ele destaca que o Brasil está passando pelo início da implementação de uma política de acesso à informação pública, que é um instrumento essencial de uma ampla política nacional adotada para a transparência e o controle social dos atos adotados. Ele também ressalta que as universidades federais brasileiras, como entidades públicas, estão obrigadas a cumprir as normas de transparência e prestação de contas, mas muitas vezes enfrentam dificuldades para se adaptar à nova legislação. Quanto à metodologia, o autor adota uma abordagem qualitativa, realizando pesquisa bibliográfica e documental para coleta de informações.

A autora Affonso ocupa a décima posição e seu foco é nas Políticas Públicas, com ênfase nas Universidades. Em seus estudos, ela destaca a importância dos repositórios de dados científicos como digitais implementados nas universidades, que auxiliam os pesquisadores no gerenciamento, disponibilização e acesso a dados científicos, promovendo sua reutilização. Além disso, ressalta-se a necessidade de considerar aspectos relacionados à privacidade dos sujeitos mencionados nas

pesquisas nos Planos de Gerenciamento de Dados (PGD), tanto dos investigadores quanto dos repositórios. A autora adotou uma abordagem metodológica quantitativa e qualitativa, utilizando o método exploratório para analisar os documentos das universidades.

Os demais autores também catalogados na fase Coleta de Dados do CVD-CI foram de suma importância para a Ciência da Informação.

5. CONSIDERAÇÕES FINAIS

Os estudos sobre Dados e Privacidade na Ciência da informação teve como objetivo examinar os aspectos legais envolvidos na coleta, armazenamento, recuperação e descarte de dados pessoais. Foi realizada uma análise dos impactos das tecnologias de informação e comunicação na privacidade e proteção de dados, com foco na investigação da evolução das práticas de coleta de dados, o uso dos dados por parte das empresas e as disposições relacionadas aos direitos à privacidade dos indivíduos.

O fenômeno *Big Data* é um termo que descreve grandes conjuntos de dados complexos e em constante crescimento, que são difíceis de serem gerenciados e analisados com as ferramentas e técnicas tradicionais. Esses dados são gerados a partir de diversas fontes, incluindo dispositivos móveis, sensores, transações financeiras, redes sociais e outros. A análise desses dados pode fornecer informações valiosas para empresas e organizações em setores diversos, incluindo finanças, saúde, varejo, marketing e muitos outros.

Além disso, os estudos sobre dados e privacidade na Ciência da Informação também exploram a legislação e as políticas governamentais relacionadas à proteção de dados. No Marco Regulatório Brasileiro é abordado as leis: Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709, de 14 de agosto de 2018, foi promulgada para proteger os direitos fundamentais de liberdade e privacidade e a livre formação da personalidade de cada pessoa. A Lei trata do tratamento de dados pessoais disponibilizados em meio físico ou digital por pessoas físicas ou jurídicas de direito público ou privado, incluindo diversas operações que possam ocorrer em meio manual ou digital; A Lei de Acesso Informação (Lei nº 12.527/2011) determina que o tratamento das informações pessoais detidas por entidades e instituições nela abrangidas, seja realizado de modo transparente, respeitando o direito fundamental à proteção da intimidade, da vida privada, da honra e da imagem das pessoas, o que, nos fundamentos defendidos nesta obra, corresponde à proteção do direito fundamental à privacidade; A Lei de Crimes Informáticos (Lei nº 12.737/2012) é uma legislação brasileira que busca combater os crimes cibernéticos e estabelecer condutas ilícitas relacionadas ao uso indevido de computadores e informações digitais. Essa lei, entrou em vigor no dia 2 de abril de 2013; A instituição do Marco Civil

da *Internet* oficialmente conhecido como Lei nº 12.965/2014, é uma legislação brasileira que estabelece princípios, direitos e responsabilidades para o uso da Internet no país. Ele foi sancionado em abril de 2014 e entrou em vigor em junho do mesmo ano. O objetivo principal do Marco Civil da Internet é garantir a liberdade de expressão, a privacidade dos usuários, a neutralidade da rede e a proteção dos dados pessoais, além de estabelecer diretrizes para a atuação dos provedores de serviços na internet.

O Ciclo de Vida de Dados (CVD-CI) é um conceito importante na Ciência da Informação que descreve as etapas pelas quais os dados passam, desde a sua coleta até o seu arquivamento ou descarte. Essas etapas ajudam a compreender como os dados são gerados, organizados, armazenados, preservados e utilizados ao longo do tempo. O CVD é importante para garantir a qualidade, a integridade e a usabilidade dos dados. Considera-se os aspectos éticos, legais, técnicos e organizacionais. Ao entender e gerenciar esse ciclo de forma adequada, é possível melhorar a eficiência, a transparência e a confiabilidade das práticas de gerenciamento de dados na área da Ciência da Informação.

A Análise de Domínio é uma atividade realizada no campo da Ciência da Informação que busca compreender e explorar um determinado domínio de conhecimento. Essa análise é essencial para identificar e estruturar as informações relevantes dentro desse domínio, bem como para facilitar o acesso, a recuperação e o uso efetivo dessas informações. A análise de domínio envolve o estudo dos conceitos, termos, explicações, relações e características específicas de um determinado campo de conhecimento. Ela busca compreender como o conhecimento é organizado, representado e utilizado nesse domínio, a fim de identificar os principais oferecidos, temas e áreas de interesse.

O presente trabalho consiste no mapeamento dos trabalhos das principais Bases de Dados Referenciais de Artigos de Periódicos em Ciência da Informação, abordando os aspectos legais relacionados ao uso de dados pessoais no período de janeiro de 2012 a dezembro de 2022. A categorização temática foi estruturada de acordo com as etapas do ciclo de vida de dados propostos por Sant'Ana (2016), buscando identificar os principais focos de interesse das Bases de Dados durante o período estudado, além de analisar tendências, distribuições de autoria e institucionais.

A partir da compilação de dados, os trabalhos foram categorizados e a geração de gráficos ocorreu com informações de distribuição temporal de autores e suas filiações institucionais; área de aplicação do estudo de acordo com o Ciclo de Vida dos Dados em CI; Em seguida, foi feita uma análise histórica de tendências das discussões nos últimos dez anos (2012-2022) dos dez autores mais colaborativos com o CVD - CI.

A análise realizada revelou que aproximadamente 75% dos trabalhos das Bases de Dados estão concentrados na etapa Coleta de dados do Ciclo de Vida (CVD-CI). Nota-se a carência de estudos nas temáticas das fases do ciclo de vida dos dados como armazenamento, recuperação e descarte de dados pessoais. Esses temas são fundamentais para compreender os aspectos legais envolvidos no uso de dados pessoais e para orientar possíveis ações na área da Ciência da Informação.

Essa pesquisa corrobora com as ideias propostas por Sant'Ana (2016), ao adotar o Ciclo de Vida de Dados (CVD-CI) para gerenciar os dados pessoais em um ambiente de *Big Data*, desde a fase de coleta até o descarte. Além disso, a aplicação da Lei Geral de Proteção de Dados (LGPD) é essencial para garantir a privacidade dos dados e garantir o uso adequado das informações.

Acredita-se que essa pesquisa é um ponto de partida importante para pesquisas futuras, pois destaca a necessidade de uma análise crítica e reflexiva sobre os aspectos que envolvem o uso de dados e informações. Essa abordagem é fundamental para subsidiar ações estratégicas em diversos setores da sociedade na atualidade.

REFERÊNCIAS

ADREJEVIC, M.; GATES, K. **Big Data surveillance**: introduction. *Surveillance & Society*, v. 12, n. 2, p. 185-196, 2014.

ANDRADE, M. C.; GONÇALEZ, P. R.; BERTI JUNIOR, D. W.; BAPTISTA, A. A.;

CONEGLIAN, C. S. Ciência responsável dos dados: imparcialidade, precisão, confidencialidade, e transparência dos dados. **Informação & Informação**, v. 25, n. 2, p. 26-48, 2020. DOI: 10.5433/1981-8920.2020v25n2p26 Acesso em: 08 set. 2022.

BALDISSERA, W. A. **Jurisdição e Transferência de Dados**: desafios para a proteção do direito à privacidade. 2019. Disponível em: <https://periodicos.unisantos.br/leopoldianum/article/view/912>. Acesso em: 01 set. 2022.

BAWDEN, D., ROBINSON, L. (2012). **Introdução à Ciência da Informação**. Faceta. Doi:10.29085/9781783300761

BEAGRIE, N. Digital curation for science, digital libraries, and individuals. **The International Journal of Digital Curation**, Edinburgh, v. 1, Iss. 1, 2006. Disponível em: <http://www.ijdc.net/article/view/6>. Acesso em: 23 mai. 2023.

BEZERRA, M. A. A.; ARAÚJO, E. A.; Uma ética da informação para pensar o orkut: reflexões sobre a informação e a liberdade no contexto da sociedade da informação. **Informação & Sociedade: Estudos**, v. 18, n. 2, 2008. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/91053>. Acesso em: 29 ago. 2022.

BEYER, M. A.; LANEY, D. **The importance of Big Data**: A definition. 2012. Disponível em: <https://www.gartner.com/doc/2057415/importance-big-datadefinition>>. Acesso em: 13 ago. 2022.

BIONI, B. R. **Proteção de dados pessoais**: a função e os limites do consentimento. 2019. Rio de Janeiro: Forense.

BRASIL. **Constituição Federal de 1988**. Brasília, DF, out. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 08 set. 2022.

BRASIL. Lei n.º 12.527, de 18 de novembro de 2011. **Lei de Acesso à Informação (LAI)**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília: Presidência da República, 2011.

BRASIL. Lei nº 12.737, de 30 de novembro de 2012. **Dispõe sobre a tipificação criminal de delitos informáticos**; altera o Decreto-Lei no 2.848, de 7 de dezembro

de 1940 - Código Penal; e dá outras Providências. Brasília, DF, nov. 2012. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm Acesso em: 08 set. 2022.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. **Marco Civil da Internet**. Brasília, DF, abr. 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 08 set. 2022

BRASIL. Lei nº 13.709 de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Brasília, DF, ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 08 set. 2022.

BRUNO, F. **Dispositivos de vigilância no ciberespaço: duplos digitais e identidades simuladas**. 2006. *Fronteiras: Estudos Midiáticos*, São Leopoldo, v. 7, n. 2, p. 152- 159. Disponível em: <http://revistas.unisinos.br/index.php/fronteiras/article/view/6129>. Acesso em: 29 ago. 2022.

BURKE, P. (Org.). *A escrita da história: novas perspectivas*. São Paulo: Unesp, 1992.

BUSH, V. et al. *As we may think*. *The atlantic monthly*, Boston, v. 176, n. 1, p. 101-108, 1945.

CAMPOS, A. **Maioria dos brasileiros teme fraudes, diz pesquisa da Febraban**. 2021. Disponível em: <https://valorinveste.globo.com/produtos/servicos-financeiros/noticia/2021/07/02/maioria-dos-brasileiros-teme-fraudes-diz-pesquisa-da-febraban.ghtml>. Acesso em: 03 jul. 2022.

CAPURRO, R. **Ética para provedores e usuários da informação**. In: Orgs. Anton Kalb, Reinhold Esterbauer Hans-Walber Ruckenbauer. *Cibernética – Responsabilidade em mundo interligado pela rede digital*. São Paulo: Loyola, 2001.

CAPURRO, R. **Epistemologia e Ciência da informação**. In: V ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., Belo Horizonte, 2003. *Anais...* Belo Horizonte: Escola de Ciência da Informação da UFMG, 2003. Disponível em: http://www.capurro.de/enancib_p.htm#:~:text=A%20primeira%20parte%20menciona%20alguns,s%C3%A3o%20aprofundados%20tr%C3%AAs%20paradigmas%20epistemol%C3%B3gicos. Acesso em: 12 jan. 2022.

CAVALCANTI, M. Big data não é tecnologia. **O Globo**, Rio de Janeiro, 23 mar. 2014. Disponível em: <https://g1.globo.com/especial-publicitario/cidade-em-movimento/renault/noticia/2021/08/05/navegacao-big-data-e-a-Internet-das-coisas.ghtml>. Acesso em: 01 jun. 2023.

CIURIAK, D. *Unpacking the Valuation of Data in the Data-Driven Economy*. 2019 Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3379133. Acesso em: 23 mai. 2023.

CONEGLIAN, C. S. *et al.* O papel da web semântica nos processos do *Big Data*. **Encontros Bibli**, v. 23, n. 53, p. 137-146, set./dez. 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p137>. Acesso em: 15 ago. 2022.

CONEGLIAN, C. S.; SEGUNDO, J. E. S.; SANT'ANA, R. C. G. *Big Data*: fatores potencialmente discriminatórios em análise de dados. **Em Questão**, v. 23, n. 1, p. 62-86, 2017. DOI: 10.19132/1808-5245231.62-86 Acesso em: 18 nov. 2022.

CHANG, H. T.; MISHRA, N.; LIN, C. C. (2015). **IoT Big-Data Centred Knowledge Granule Analytic and Cluster Framework for BI Applications: A Case Base Analysis**. *PloS one*, 10(11), e0141980. <https://doi.org/10.1371/journal.pone.0141980>

DATAONE. *What is DataONE? Online*. Disponível em: <https://www.dataone.org/what-dataone>. Acesso em: 20 mai. 2023.

DAVENPORT, T. H. **Big Data at work**: dispelling the myths, uncovering the opportunities. Harvard: Harvard Business School Publishing, 2014.

DONEDA, D. **Da privacidade à proteção de dados pessoais**. Rio de Janeiro: Renovar, 2006.

DONEDA, D. **Privacidade e transparência no aceso à informação pública**. Zaragoza: Prencsas Universitárias de Zaragoza, 2010. Disponível em: <http://www.egov.ufsc.br/portal/sites/default/files/lefis11-09.pdf>. Acesso em: 08 jun.2023

DATAREPORTAL. (2020). **Digital 2020**: Canada. Disponível em: <https://datareportal.com/reports/digital-2020-canada>. Acesso em: 13 ago. 2022.

DAVIS, K. **Ethics of Big Data**. Sebastopol: O`Reilly Media, 2012.

DIAS, G. A.; VIEIRA, A. A. N. *Big Data*: questões éticas e legais emergentes. **Ciência da Informação**, v. 42, n. 2, 2013. DOI: 10.18225/ci.inf..v42i2.1380 Acesso em: 13 ago. 2022.

DONEDA, D. **Da privacidade à proteção de dados pessoais**. Rio de Janeiro: Renovar, 2006.

DUMBILL, E. **What is Big Data?** In: O`Reilly Media Inc. *Big Data Now: current perspectives*. O`Reilly Media: California. 2012. Disponível em: <http://www.oreilly.com/data/free/files/big-datanow-2012.pdf>. Acesso em: 9 abr. 2022.

DUTRA, M. L.; MACEDO, D. D. J. Curadoria digital: proposta de um modelo para curadoria digital em ambientes *Big Data* baseado numa abordagem semi-automática para a seleção de objetos digitais. **Informação & Informação**, v. 21, n. 2, p. 143-169, 2016. DOI: 10.5433/1981-8920.2016v21n2p143 Acesso em: 13 ago. 2022.

EUROPEAN DATA SCIENCE ACADEMY. **About EDSA**. (2019). Disponível em: <http://edsa-project.eu/overview/about-edsa/>. Acesso em: 20 mar. 2022.

FEBRABAN. **Segurança de Dados no Brasil: a visão da sociedade**. Observatório Febraban, Julho de 2021. Disponível em: https://cmsarquivos.febraban.org.br/Arquivos/documentos/PDF/RELATO%CC%81RIO%20OBSERVATO%CC%81RIO%20FEBRABAN%20-%20JUNHO%202021_%20SEGURANC%CC%A7A%20DE%20DADOS%20NO%20BRASIL_VF.pdf. Acesso: 24 mar. 2022

FERRER-SAPENA, A.; SÁNCHEZ-PÉREZ, E. Open data, big data: hacia donde nos dirigimos? **Anuário ThinkEPI** 2013, v. 7, p. 150-156, 2013. Disponível em: <http://eprints.rclis.org/21006/1/150-156-Ferrer-Sanchez-Open-data-Big-data.pdf>. Acesso em: 27 maio. 2023.

FERREIRA, L. B.; ROCKEMBACH, M.; KREBS, L. M. Reflexões conceituais e éticas sobre *Big Data*: limites e oportunidades. **Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, n. XVIII ENANCIB, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/105088>. Acesso em: 18 nov. 2022.

FORNASIER, M. DE O.; KNEBEL, N. M. P. O titular de dados como sujeito de direito no capitalismo de vigilância e mercantilização dos dados na Lei Geral de Proteção de Dados. **Revista Direito e Práxis**, v. 12, n. 2, p. 1002–1033, abr. 2021.

FORTES, V. B. **Os direitos de privacidade e a proteção de dados pessoais na Internet**. Editora Lumen Juris, Rio de Janeiro, 2016.

FREIRE, G. H. A. **Ética da Informação**. 2010. Disponível em: https://lti.pro.br/uploads/posts_files/148/5174bc63b1722a7b0a923f3f8fe63f.pdf. Acesso em: 16 ago. 2022.

GALVÃO, M. C. B.; RICARTE, I. L. M. **Prontuário do Paciente**. 1. ed. Rio de Janeiro: Guanabara Koogan, 2012. v. 1, 344p.

GALVÃO, M. C. B.; BORGES, P. C. R. Ciência da informação: ciência recursiva no contexto da sociedade da informação. **Ciência da Informação**, [S. l.], v. 29, n. 3, 2000. DOI: 10.18225/ci.inf.v29i3.872. Disponível em: <https://revista.ibict.br/ciinf/article/view/872>. Acesso em: 5 jun. 2023.

GARFINKEL, S. **Database nation: the death of privacy in the 21st Century**. Sebastopol: O'Really, 2000.

GUITARRARA, Paloma. **O que são big techs?**; Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/o-que-e/geografia/o-que-sao-big-techs.htm>. Acesso em 03 de junho de 2023.

GRISOTO, A. P.; SANT'ANA, R. C. G.; SEGUNDO, J. E. S. A questão da privacidade no contexto da ciência da informação: uma análise das teses e dissertações do programa de pós-graduação em Ciência da Informação da Unesp

Campus de Marília. **Revista Ibero-Americana de Ciência da Informação**, v. 8 No 2, n. 2, p. 165-181, 2015. DOI: 10.26512/rici. v8.n2.2015.2066 Acesso em: 09 abr. 2022.

GUIMARÃES, J. A. C. “**Análise de domínio como perspectiva metodológica em organização da informação**”. *CI Inf.*, vol. 41 no. 1, jan./apr. 2014, p.13-21, <http://revista.ibict.br/ciinf/article/view/1415/1593>. Acesso em: 13 ago. 2022

HJØRLAND, BIGGER E ALBRECHTSEN, HANNE. “Rumo a um novo horizonte na Ciência da Informação: domínio Glushko, Robert J. (ed.). **A disciplina da organização. The MIT Press, 2013, 540 p. análise**”, *Journal of the American Society for Information Science*, vol. 46, nº. 6, 1995, pág. 400- 425, [https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/\(SICI\)1097-4571\(199507\)46:6%3C400::AID-ASI2%3E3.0.CO;2-Y](https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-4571(199507)46:6%3C400::AID-ASI2%3E3.0.CO;2-Y). Acesso em: 13 ago. 2022.

HJØRLAND, BIRGER. “Análise de Domínio em Ciência da Informação: onze abordagens – tradicionais e Harvey, Ross. **Curadoria digital: um manual de como fazer. Neal-Schuman Publishers, Inc., 2010. Inovativa**”. *Jornal de Documentação*, vol. 58, nº. 4, 2002, p. 422-462 <https://doi.org/10.1108/00220410210431136>. Acesso em: 13 ago. 2022

HUAWEY. **Cyber Security and Data Privacy**: key considerations for policymaker. 2021. Disponível em: <https://www.huawei.com/en/annual-report/2021>. Acesso em: 16 ago. 2022.

IBM. **What is Big Data?** 2014. Disponível em: <<http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>>. Acesso em: 19 mar. 2022.

IBM. **O que é Análise de Big Data**. 2015. Disponível em: <https://www.ibm.com/br-pt/analytics/hadoop/big-data-analytics#:~:text=Big%20data%20%C3%A9%20um%20termo,alta%20velocidade%20ou%20grande%20variedade>. Acesso em: 13 ago. 2022.

IRDETO. **Pesquisa de segurança cibernética da Irdeto Global Connected Industries**: ataques cibernéticos de IoT. 2019 Disponível em: <https://irdeto.com/privacy-portuguese>. Acesso em: 13 ago. 2022.

KEMPER, J. ; KOLKMAN, D. **Transparente para quem?** Nenhuma responsabilidade algorítmica sem um público crítico. 2018. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1477967>. Acesso em: 13 ago. 2022.

LANEY, D. **3D data management**: Controlling data volume, velocity and variety. META Group Research Note 6, fev. 2001.

LA TAILLE, Y. de. **Moral e Ética**: dimensões intelectuais e afetivas. São Paulo: Artmed, 2011. 192 p.

LE COADIC, Yves-François. A Ciência da Informação. Brasília, DF: Briquet de Lemos/Livros, 1996.

LEWIS, L. **Volume de informação gerada por minuto em 2021**. Disponível em: <https://lorilewismedia.com/>. Acesso 03 jun.2023

LIMA, C. F. de. **O Profiling e a Proteção de Dados Pessoais**. 2019. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/199951/001100797.pdf?sequence=1&isAllowed=y>. Acesso em: 15 jul. 2022.

LOBO, A. P. **Governo de Minas contrata big data da Microsoft**. Convergência Digital, 2013. Disponível em: <http://convergenciadigital.uol.com.br/cgi/cgilua.exe/sys/start.htm?infolid=SRVMstCUk>. Acesso em: 07 mai. 2023.

LOPES, L. B. P.; SILVA, M. B. Tecnologia *Big Data* em bibliotecas: estudo contributivo a partir de pontuações literárias. **Revista Brasileira de Biblioteconomia e Documentação**, v. 17, p. 1-23, 2021. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/168759>. Acesso em: 13 ago. 2022.

LOTT, Y. M.; CIANCONI, R. B. Vigilância e privacidade, no contexto do *Big Data* e dados pessoais: análise da produção da Ciência da Informação no Brasil. **Perspectivas em Ciência da Informação**, v. 23, n. 4, p. 117-132, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/108454>. Acesso em: 28 mar. 2022.

LYON, D. **Cultura da vigilância**: envolvimento, exposição e ética na modernidade digital. 2019. Em: BRUNO, Fernanda, et al. (Org.). *Tecnopolíticas da Vigilância*. São Paulo: Boitempo, p. 151-179.

MACEDO, J. Big data: transformando a indústria de telefonia móvel. **Revista CanalTech**. Disponível em: <https://arquivo.canaltech.com.br/big-data/Big-Data-transformando-a-industria-de-telefonia-movel/> Acesso em: 28 mai. 2023

MASCARO, Alysson Leandro. *Filosofia do Direito*. 4 ed. São Paulo: Atlas, 2014.

MACNISH, K. Data, Privacy, and the Individual. Madri: **Centro para a Governança da Mudança, IE University**, 2019. Disponível em: www.ie.edu/cgc/research/data-privacy-individual. Acesso em: 21 jun. 2023.

MAYER-SCHONBERGER, V.; CUKIER, K. **Big Data**: A revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt, 2013.

MCLUHAN, M. *Os meios de comunicação como extensões do homem*. São Paulo: Cultris, 1969.

MENDES, L. S. **Transparência e privacidade**: violação e proteção da informação pessoal na sociedade de consumo. 2008. 156 f. Dissertação (Mestrado em Direito) –

Faculdade de Direito, Universidade de Brasília, DF, 2008. Disponível em: <https://repositorio.unb.br/bitstream/10482/4782/1/DISSERTACAO%20LAURA.pdf>. Acesso em: 9 set. 2022.

MILAGRE, J.; SANTARÉM SEGUNDO, J. E. A propriedade dos dados e a privacidade na perspectiva da Ciência da Informação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S. l.], v. 20, n. 43, p. 47-76, 2015. DOI: 10.5007/1518-2924.2015v20n43p47. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20n43p47>. Acesso em: 9 set. 2022.

MIRAGEM, B. A Lei Geral de Proteção de Dados (Lei 13.709/2018) e o Direito do Consumidor. **Revista dos Tribunais**, v. 1009, nov/2019.

MORA, J. F. **Dicionário de Filosofia**: Tomo II (E-J). Tradução: Maria Stela Gonçalves et al. São Paulo: Edições Loyola, 2001, p. 931 – 935.

MOREIRA, F. M.; SANT'ANA, R. C. G.; JORENTE, M. J. V.; MOREIRA, F. M.; SANT'ANA, R. C. G.; JORENTE, M. J. V. A complexidade na disponibilização e acesso a dados governamentais na web. **Perspectivas em Ciência da Informação**, v. 21, n. 3, p. 70-88, 2016. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/36561>. Acesso em: 19 nov. 2022.

NEGRI, S. M. C. V.; BATISTA, N. P. R. Do rastreamento de contato à vigilância: um estudo sobre o tracetogether app. **Liinc em revista**, v. 17, 2021. DOI: 10.18617/liinc.v17i1.5710 Acesso em: 09 set. 2022.

NIST. *Big Data* Definitions. Disponível em: <http://dx.doi.org/10.6028/NIST.SP.1500-1>. Acesso em: 13 ago. 2022. INTEL. **Peer Research Report: Big Data Analytics**. Disponível em: <http://www.intel.co.za/content/www/za/en/big-data/data-insights-peer-researchreport.html>. Acesso em: 14 ago. 2022.

OECD. **Exploring the economics of personal data**: a survey of methodologies for measuring monetary value. OECD Digital Economy Papers, No. 220, 2013. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en. Acesso em: 14 ago. 2022

OLIVEIRA, A. C. S.; MOTTA, D. B.; MELO, J. H.; ESTEVES, R. C. S. P. A. Empoderamento digital, proteção de dados e IgpD. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 15, n. 3, p. 247-261, 2020. DOI: 10.22478/ufpb.1981-0695.2020v15n3.54698 Acesso em: 14 abr. 2022.

ORACLE. **O que é Big Data?** 2020. Disponível em: <https://www.oracle.com/br/big-data/what-is-big-data/#:~:text=A%20defini%C3%A7%C3%A3o%20de%20big%20data,de%20>. Acesso em: 14 ago. 2022.

PASSARELLI, B. e JUNQUEIRA, A. H. e ANGELUCI, A.C. B. **Os nativos digitais no Brasil e seus comportamentos diante das telas**. Matrizes, v. 8, n. ja/ju 2014, p.

159-178, 2014 Tradução. Disponível em: <http://dx.doi.org/10.11606/issn.1982-8160.v8i1p159-178>. Acesso em: 05 jun. 2023.

PINTO, V. A.; CARDOSO, A. M. P.; PINHEIRO, M. M. K.; PARREIRAS, F. S. Interdisciplinarity in data science over *Big Data*: findings for mining industry. **Informação & Sociedade: Estudos**, v. 29, n. 4, p. 61-74, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/147954>. Acesso em: 13 ago. 2022.

QUIVY, R.; CAMPENHOUDT, L. V. **Manual de Investigação em Ciências Sociais**. 4. ed. Lisboa: Gradiva, 2005. (Coleção Trajectos). 275 p.

RAHUL, K.; BANYAL, R. Data Life Cycle Management in Big Data Analytics. **Procedia Computer Science**. v. 173. 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050920315465#cebibsec1>. Acesso em 28 abr. 2023.

REIS, L. C. R.; Sá, M. I. F. E. *Big Data*: um novo campo de atuação para bibliotecários. **Prisma.com (Portugual)**, n. 41, p. 231-250, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/135715>. Acesso em: 14 ago. 2022.

RIBEIRO, C. J. S. *Big Data*: os novos desafios para o profissional da informação. **Informação & Tecnologia**, v. 1, n. 1, p. 96-105, 2014. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/40838>. Acesso em: 13 ago. 2022.

RODOTÀ, S. **A vida na sociedade da vigilância**: a privacidade hoje. Rio de Janeiro: Renovar, 2008.

ROUSE, Margaret. **Sombra de Dados**. 2014. Disponível em: <https://www.techopedia.com/definition/28091/data-shadow>. Acesso em: 19 jun. 2023.

RUMBAUGH, J. et al. **Modelagem e projetos baseados em objetos**. Rio de Janeiro: Campus, 1994.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, [S.l.], v. 21, n. 2, p. 116–142, dez. 2016. ISSN 1981-8920. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 25 mar. 2022. DOI: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>.

SANT'ANA, R. C. G. Ciclo de vida dos dados e o papel da ciência da informação. In: **Encontro Nacional De Pesquisa Em Ciência da Informação**, 14., 2013, Florianópolis. Anais [...]. Florianópolis, 2013. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000013772/fac7d2fec2c83f1e9208204b2fef10>. Acesso em: 14 mai. 2023.

SANTOS, P. L. V. A. da C.; SANT'ANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, [S. l.], v. 42, n. 2, 2015. DOI:

10.18225/ci.inf.v42i2.1382. Disponível em:
<https://revista.ibict.br/ciinf/article/view/1382>. Acesso em: 26 abr. 2023.

SARACEVIC, T. Ciencia da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.1, p.41-62, jan/jun. 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>. Acesso em: 26 mai. 2023.

SARACEVIC, T. *Information science*. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1051-1063, Oct. 1999.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://carpedien.ien.gov.br/bitstream/ien/1624/1/GUIADEDADOSDEPESQUISA.pdf>. Acesso em: 14 mai. 2023.

SHAH, S. I. H.; PERISTERAS, V.; MAGNISALIS, L. DaLiF: a data lifecycle framework for data-driven governments. **Journal of Big Data**. n. 8, v. 89, 2021. <https://doi.org/10.1186/s40537-021-00481-3>. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00481-3>. Acesso em 3 mai. 2023.

SILVEIRA, S. A.; AVELINO, R.; SOUZA, J. A privacidade e o mercado de dados pessoais. **Liinc em Revista**, v. 12, n. 2, p. 217-230, 30 nov. 2016

SEF - Secretaria de Estado da Fazenda. **O Ciclo de Vida do Tratamento dos Dados Pessoais**. 2021. Disponível em: <http://www.fazenda.mg.gov.br/transparencia/lgpd/LGPD-SEF-Ciclo-de-Vida-fases.pdf>. Acesso em: 19 nov. 2022.

SIEBRA, S. A.; BORBA, V. R. Análise, tendências e perspectivas da produção científica em curadoria digital. In: OLIVEIRA, H. P. C.; VIDOTTI, S. A. B. G. (org.). **Informação e tecnologias: desenhando fronteiras científicas**. João Pessoa: Editora Ufpb, 2018. p. 97-112.

SILVA, V. M. da. **Big Data**: definição e um breve histórico. Definição e Um Breve Histórico. 2019. Disponível em: <https://medium.com/@stratiobd/big-data-defini%C3%A7%C3%A3o-e-um-breve-hist%C3%B3rico-a389abcf6a3>. Acesso em: 01 set. 2022.

SMIRAGLIA, R. P. Domain Analysis for Knowledge Organization: **Tools for Ontology Extraction**. Oxford: Chandos Publishing, 2015.

SOUZA, M.; ALMEIDA, F. G. O comportamento do termo dado na ciência da informação. **Ciência da Informação em Revista**, v. 8, n. 2, p. 39-54, 2021. DOI: 10.28998/cirev.2021v8n2c. Acesso em: 08 set. 2022.

TAURION, C. O caos conceitual e os 5V's do big data. **Revista CIO**, 11 maio 2012a. Disponível em: <https://cio.uol.com.br/opiniaio/2012/05/11/o-caos-conceitual-e-os-5vs-do-bigdata/>. Acesso em: 13 jan. 2023.

TAURION, C. **Você realmente sabe o que é big data?** 30 abr. 2012b. Disponível: <https://www.oracle.com/br/big-data/what-is-big-data/#:~:text=A%20defini%C3%A7%C3%A3o%20de%20big%20data%20crescentes%20e%20com%20mais%20velocidade>. Acesso em: 13 jan. 2023.

TEIXEIRA, T. **Marco Civil da Internet**. São Paulo: Leya. 2016. Disponível em: https://books.google.com.br/books?hl=ptBR&lr=&id=JwADDAAAQBAJ&oi=fnd&pg=3&dq=marco+civil+da+Internet&ots=_xY0ZiKLa&sig=yf1qus8zuPSXloLX2A2LHi3wSQ#v=onepage&q&f=false. Acesso em: 08 set. 2022.

TURKLE, S. Os efeitos do mundo digital no nosso comportamento. Globosat Play. UNITED NATIONS DEVELOPMENT GROUP. **Data privacy, ethics and protection guidance note on Big Data for achievement of the 2030 agenda**. 2018. Estados Unidos. Disponível em: <https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda>. Acesso em: 01 jul. 2022.

VÁSQUEZ, S.; Ética. Tradução: João Dell'Anna. Rio de Janeiro: Civilização Brasileira. 1995.

VASALOU, A.; JOINSON, A.; HOUGHTON, D. Privacy as a Fuzzy Concept: A New Conceptualization of Privacy for Practitioners. **Journal of the Association for Information Science and Technology**, v. 66, n. 5, May 2014.

VIGNOLI, R. G.; VERONEZ JUNIOR, W. R.; SILVA, A. R.; MONTEIRO, S. D. Mediação da informação e a proteção da privacidade e de dados sensíveis por bibliotecários. **Informação & Informação**, v. 25, n. 4, p. 280-305, 2020. DOI: 10.5433/1981-8920.2020v25n4p280 Acesso em: 29 ago. 2022.

ZUBOFF, Shoshana. The age of surveillance capitalism: The fight for a human future at the new frontier of power. Londres: Profile Books, 2019.

WOOD, D. M. et al. **A report on the surveillance society: report for the UK information commissioner's office**. *Surveillance Studies Network*, 2 nov. 2006. Disponível em: http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/02_11_06_surveillance.pdf. Acesso em: 13 ago. 2022.