

UNIVERSIDADE FEDERAL DE SÃO CARLOS - UFSCar  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS - CECH  
CURSO DE BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO

ALEXANDRA ALMEIDA

***BIG DATA E PRIVACIDADE NA WEB:***  
**UMA ANÁLISE ACERCA DOS METADADOS COLETADOS**

SÃO CARLOS/SP  
2022

ALEXANDRA ALMEIDA

***BIG DATA E PRIVACIDADE NA WEB:***  
**UMA ANÁLISE ACERCA DOS METADADOS COLETADOS**

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de Bacharela em Biblioteconomia e Ciência da Informação pela Universidade Federal de São Carlos.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Ana Carolina Simionato Arakaki.

SÃO CARLOS/SP  
2022

*BIG DATA E PRIVACIDADE NA WEB: UMA ANÁLISE ACERCA DOS METADADOS  
COLETADOS.*

ALEXANDRA CRISTINA DOS SANTOS ALMEIDA

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de Bacharela em Biblioteconomia e Ciência da Informação pela Universidade Federal de São Carlos.

Aprovado em: \_\_\_\_/\_\_\_\_/\_\_\_\_.

**BANCA EXAMINADORA**

---

Orientadora

Professora Doutora Ana Carolina Simionato Arakaki  
Universidade Federal de São Carlos (UFSCar)

---

Membro da banca (1)

Professora Doutora Camila Carneiro Dias Rigolin  
Universidade Federal de São Carlos (UFSCar)

---

Membro da banca (2)

Letícia Guarany Bonetti  
Universidade Federal de São Carlos (UFSCar)

Dedico à minha mãe, Cristina e ao meu pai  
Ivanildo (em memória).

Amo vocês, infinitamente.

## AGRADECIMENTOS

Ao meu pai, Ivanildo,

por ser o melhor pai do mundo e em toda sua trajetória, ter me mostrado o valor da família e do amor. Você sempre estará presente em minha memória e meu coração.

À minha mãe, Cristina,

por sua bondade, cuidado e atenção. É a mulher mais forte que eu conheço, a base da minha existência e a melhor mãe do mundo.

Ao meu irmão, Alexandre,

pelo cuidado, apoio e por ter sido desde sempre o meu maior exemplo e inspiração, tenho muito orgulho de ser sua irmã.

À minha sobrinha, Sophie,

por sempre me mostrar que o universo é um lugar lindo, colorido, com unicórnios e que eu estou aqui para sonhar e realizar - tenho muita sorte de ser sua tia.

Aos meus grandes amigos, Jonatha, Mayara e Bianca (Jubs),

por crescerem comigo, por sempre me apoiarem em tudo e por estarem presentes em cada passo da minha jornada - *oito anos não são oito meses* - e que venham muito mais.

Às minhas irmãs, Milena e Nathália,

pelo amor, suporte, parceria, por partilharem comigo momentos inesquecíveis e terem se tornado parte de mim - minha família, *The Space Trio*.

Aos meus amigos, Bruno e Erick,

pela parceria na graduação e na vida, com risos fora de hora - e trabalhos de última hora - vocês tornaram essa *missão de espãs* muito mais leve e possível.

À minha amiga, Brenda,

pela parceria, ensinamentos, conversas aleatórias, conselhos e por todo o suporte prestado em minha vida profissional e acadêmica.

Às minhas amigas, Marília e Larissa,

Marília (*Jojinha*), por sempre me apoiar, me incentivar e aconselhar. Lari, por sempre estar disposta a me ouvir e a partilhar suas histórias comigo.

À minha professora e orientadora, Ana Carolina,

por todos os ensinamentos, pela paciência e todo o suporte prestado ao longo da graduação e principalmente, no desenvolvimento da minha pesquisa, muito obrigada Ana.

À professora Camila e à Letícia,

pela gentileza de aceitarem participar da minha banca, sou muito grata pelas sugestões e conhecimentos compartilhados, muito obrigada meninas.

À minha professora do ensino fundamental, Mônica,

por ter confiado em meu potencial e por ter me ajudado - direta e indiretamente - a chegar aonde eu estou.

Aos professores e técnicos do DCI,

por todo o conhecimento e suporte prestado ao longo da minha graduação, lembrarei de todos vocês com muito carinho.

À Universidade Federal de São Carlos (UFSCar),

por ser a melhor do Brasil, a Universidade que tenho muito orgulho de fazer parte e que me proporcionou inúmeros aprendizados.

À minha cidade do coração, São Carlos,

cidade onde vivi os melhores anos da minha vida e que estou pronta para viver muito mais.

*It's something unpredictable  
But in the end it's right*

Muito obrigada.

## RESUMO

Perante o avanço tecnológico e a expansão da *Internet*, intensificou-se a produção e troca de informações entre indivíduos e organizações. Nessa conjuntura, surge o *Big Data*, tecnologia que oferece um potencial para o desenvolvimento de empresas de diversos setores, além de ser um novo campo de atuação para os bibliotecários e cientistas da informação. Apesar dos benefícios, ainda há desafios para a sua implementação, como questões de Privacidade Informacional. Com esse parâmetro, a presente pesquisa se configura como abordagem exploratória, a fim de compreender o *Big Data*, analisar e registrar como é realizada a coleta de dados e Metadados e identificar questões relacionadas à Privacidade na *Web*. Conclui-se que o *Big Data* e o seu uso é uma realidade que tende a aumentar. Contudo, ainda há muitos desafios a serem enfrentados. Necessita-se de alto investimento em sua Infraestrutura e Políticas de Segurança de Informação que visem a proteção e o cuidado em suas práticas de coleta, armazenamento e tratamento dos dados, garantindo além da segurança, mas em seu caráter ético e social, a transparência com os indivíduos.

**Palavras-chave:** *Big Data*. Metadados. Privacidade. Ciência da Informação.

## ABSTRACT

In view of technological advances and the expansion of the *Internet*, the production and exchange of information between individuals and organizations has intensified. In this context, *Big Data* emerges, a technology that offers potential for the development of companies in different sectors, in addition to being a new field of activity for librarians and information scientists. Despite the benefits, there are still challenges for its implementation, such as Informational Privacy issues. With this parameter, the present research is configured as an exploratory approach, in order to understand *Big Data*, analyze and record how data and Metadata are collected and identify issues related to Privacy on the *Web*. It is concluded that *Big Data* and its use is a reality that tends to increase. However, there are still many challenges to be faced. There is a need for high investment in its Infrastructure and Information Security Policies aimed at protecting and taking care in its data collection, storage and processing practices, ensuring in addition to security, but in its ethical and social character, transparency with the individuals.

**Keywords:** *Big Data*. Metadata. Privacy. Information Science.



## LISTA DE FIGURAS

<b>Figura 1</b> - Os sete V's do <i>Big Data</i> .....	
<b>Erro! Indicador não definido.</b> <b>Figura 2</b> - Formulário de Cadastro do <i>Facebook</i> .....	<b>Erro! Indicador não definido.</b>
<b>Figura 3</b> - Página inicial do Site da <i>Amazon</i> .....	40
<b>Figura 4</b> - Formulário de Cadastro.....	40
<b>Figura 5</b> - Notificação de Privacidade.....	<b>Erro!</b>
<b>Indicador não definido.</b> <b>Figura 6</b> - Notificação de <i>Cookies</i> .....	<b>Erro! Indicador não definido.</b>

## LISTA DE QUADROS

<b>Quadro 1</b> - Diferentes fontes de dados.....	
<b>Erro! Indicador não definido.</b> Quadro 2 - Tipos de Metadados e suas funções.....	<b>Erro! Indicador não definido.</b>
Quadro 3 - Classificações das informações coletadas <i>on-line</i> .....	27
Quadro 4 - Mecanismo da recolha de informações dos consumidores <i>on-line</i> .....	<b>Erro!</b>
<b>Indicador não definido.</b> Quadro 5 - Tipos de <i>Cookies</i> .....	<b>Erro! Indicador não definido.</b>
Quadro 6 - Conceitos de <i>Big Data</i> .....	45
<b>Quadro 7</b> - Etapas do desenvolvimento da Pesquisa.....	<b>Erro! Indicador não definido.</b>

## **LISTAS DE ABREVIATURAS E SIGLAS**

**ANPD** - Autoridade Nacional de Proteção de Dados

**BCI** - Biblioteconomia e Ciência da Informação

**BRAPCI** - Base de Dados em Ciência da Informação

**LGPD** - Lei Geral de Proteção dos Dados

**NoSQL** - Not Only SQL

**PII** - Personally identifiable information

**TI** - Tecnologia da Informação

**TIC** - Tecnologias da Informação e Comunicação

**UFSCar** - Universidade Federal de São Carlos

**UI** - Unidades de Informação

**XML** - Extensible Markup Language

## SUMÁRIO

1. INTRODUÇÃO	12
1.1 Definição do problema	14
1.2 Objetivos	14
1.3 Justificativa	15
1.4 Estrutura do trabalho	15
2. FUNDAMENTAÇÃO TEÓRICA	17
2.1 Embasamentos sobre o <i>Big Data</i>	17
2.2 Os conceitos e funções dos metadados	23
2.3 Aporte teórico sobre a privacidade na <i>Web</i>	26
3. PROCEDIMENTOS METODOLÓGICOS	32
4. RESULTADOS	34
4.1 Os temas <i>Big Data</i> , privacidade e metadados de uso na Ciência da Informação	34
4.2 Um exemplo da coleta de metadados pela <i>Amazon</i>	39
5. CONSIDERAÇÕES FINAIS	45
REFERÊNCIAS	49

# 1. INTRODUÇÃO

A Lei Geral de Proteção de Dados (LGPD), nº 13.709 de 14 de agosto de 2018, é a lei brasileira que dispõe sobre o tratamento dos dados pessoais, inclusive nos meios digitais, por pessoa natural ou jurídica de direito público ou privado. De acordo com o Art. 2º da Lei nº 13.709, de 14 de agosto de 2018, a LGPD tem como objetivo garantir o direito à segurança e privacidade dos dados pessoais, através dos fundamentos ligados ao respeito à privacidade, à autodeterminação informativa, à liberdade de expressão, à inviolabilidade da intimidade, à defesa do consumidor, aos direitos humanos e ao desenvolvimento econômico e tecnológico (BRASIL, 2018).

A LGPD entrou efetivamente em vigor em 18 de setembro de 2020, o que trouxe mudanças significativas para as organizações, que tiveram que se adequar às normas estabelecidas na lei, revisar políticas e processos internos que implicam na coleta, tratamento e disseminação de dados pessoais.

Para assegurar o cumprimento da LGPD, o governo criou a Autoridade Nacional de Proteção de Dados (ANPD), o órgão da administração pública responsável pela regulamentação, implementação e fiscalização da LGPD. Dentre as competências da ANPD, pode-se destacar a elaboração de diretrizes para a política de proteção de dados, fiscalizações e aplicação de sanções para tratamento de dados em descumprimento à legislação, promoção e disseminação de conhecimento das normas e políticas e implementação de mecanismos para registro de reclamações sobre o tratamento inadequado de dados (GOV BR, 2022).

Em relação aos profissionais da Ciência da Informação e sua atuação na implementação da LGPD, Lima e Presser (2022) destacam:

Nesse contexto, o cientista da informação configura-se como um profissional promotor de mudanças positivas para a gestão organizacional. A partir dele, é possível criar um ambiente regido por políticas voltadas para a organização, o armazenamento, a recuperação e a proteção de dados. Todavia, a inserção desse profissional nas organizações ainda é escassa por diversos motivos e, dentre eles, estão o desinteresse dos gestores em investir em mão de obra qualificada, a não valorização das informações e dados e o descompromisso com as normas legais de proteção de dados (LIMA; PRESSER, 2022 p. 118).

É válido destacar que o desenvolvimento e a implementação da Lei Geral de Proteção dos Dados representaram um marco histórico em questões relacionadas à segurança e privacidade informacional no Brasil. Contudo, para que se tenha resultados efetivos, cabe às empresas se adequarem às normas pré-estabelecidas e aos órgãos desempenharem suas funções de fiscalização e aplicação de sanções.

Desde a publicação e o exercício da LGPD, tornou-se possível perceber inúmeras mudanças no que tange o funcionamento da sociedade, implicando em diversos aspectos, sendo eles econômicos, políticos, sociais e culturais. “A popularização crescente da *Internet* (rede mundial de computadores) modificou não só as práticas sociais, mas também a forma de produzir e consumir informação” (MAGALHÃES et al., 2014).

Nesse contexto, relaciona-se a grande quantidade de informações que estão sendo produzidas, coletadas, processadas, armazenadas e disseminadas, dão luz ao fenômeno conhecido como *Big Data*. O *Big Data* possui um conceito abstrato e é caracterizado pelo grande volume de dados e metadados que são gerados e processados a partir de fontes heterogêneas. Além destes atributos, também podemos destacar a velocidade, variedade, veracidade, variabilidade, volatilidade e principalmente, o valor desse volume de dados.

A análise dessa massa de dados pode gerar informações valiosas em todos os setores da sociedade humana, desde a compreensão do governo sobre as demandas da população, até mesmo para as estratégias competitivas das empresas privadas que almejam por um melhor posicionamento no mercado (TAURION, 2013).

Uma pauta que tem gerado discussões quando o assunto é *Big Data*, é em relação à privacidade dessas informações. Haja vista que, com todo esse excesso de informações, dados pessoais como a identidade, geolocalização, dados bancários e até mesmo os hábitos dos indivíduos podem, de certa forma, correr o risco de serem acessados por terceiros.

Taurion (2013) destacou em sua obra que:

É importante definir uma política de uso das informações, considerando os aspectos de privacidade e segurança. Privacidade implica em aderências às legislações e aspectos regulatórios que estejam em vigor. Também é importante manter um acompanhamento constante pelo fato de *Big Data* ainda ser uma novidade e muita legislação nova deve ser gerada nos próximos anos (TAURION, 2013).

O *Big Data* é uma tecnologia oriunda da ciência da computação, mas que devido a sua natureza, conquistou o seu espaço em áreas correlatas, como a Biblioteconomia e a Ciência da Informação. Reis e Sá (2020) salientam que o mercado do *Big Data* é um campo de atuação para os bibliotecários com perfis tecnológicos, haja vista que seus conhecimentos de classificação, análise, indexação e recuperação, coincidem com atividades realizadas por profissionais do *Big Data*.

É de conhecimento geral que os bibliotecários e cientistas da informação desempenham um papel fundamental no processo de gestão e organização da informação, portanto, lidar com diferentes tipos de dados é uma realidade para esses profissionais. A padronização dos recursos

por meio de metadados é um exemplo prático de como os bibliotecários utilizam os dados em seu cotidiano.

Portanto, esse estudo buscará compreender os conceitos sobre o *Big Data*; as definições e relações com os metadados, por fim, o trabalho descreve as convergências com a privacidade na *Web* e a sua relação com o fenômeno do *Big Data*.

## 1.1 Definição do problema

Perante os benefícios e desafios que contemplam o *Big Data*, o presente estudo tem por finalidade compreender os seus conceitos e delimitações, bem como suas tecnologias. Isto é, por meio de um levantamento bibliográfico que contempla autores de diversas áreas, busca-se reunir alguns dos conceitos, atributos e tecnologias que contemplam o *Big Data*. Do mesmo modo, busca-se destacar suas vantagens e os seus desafios para a sociedade, tal como o grande volume de dados que são gerados continuamente e sua relação com a privacidade da informação no ambiente da *Web*.

Além disso, propõe-se analisar e registrar como são coletados os dados e metadados na *Web* e identificar questões relacionadas à Política de Privacidade nesses meios. No contexto da Biblioteconomia e Ciência da Informação, busca-se através deste estudo compreender o papel dos profissionais diante desse novo campo de atuação.

Deste modo, o questionamento deste estudo é pontuado pela necessidade em compreender o *Big Data* e suas implicações, essencialmente àquelas relacionadas à privacidade e aos metadados, competências essas ligadas aos profissionais da informação.

## 1.2 Objetivos

O objetivo geral deste estudo é compreender o *Big Data*, analisando a coleta de dados e metadados e as questões relacionadas à privacidade informacional.

Os objetivos específicos que apoiam o objetivo geral, foram delineados em:

- Apresentar uma revisão de literatura acerca dos conceitos que definem o *Big Data*, suas tecnologias, benefícios e desafios;
- Descrever os metadados e seu uso em ambientes digitais;
- Relacionar a privacidade no contexto da *Web*, conjunto aos movimentos do *Big Data*;
- Caracterizar a Política Corporativa da empresa *Amazon* a partir de fontes documentais;
- Levantar o estado da arte sobre o fenômeno do *Big Data* e da privacidade informacional no cenário da Ciência da Informação.

### 1.3 Justificativa

O presente estudo tem como finalidade contribuir cientificamente no campo da Biblioteconomia e Ciência da Informação por meio de um objeto de estudo que contempla o desenvolvimento tecnológico, a informação e a sociedade. Não se limitando apenas ao campo da Ciência da Informação, este estudo visa contribuir com o desenvolvimento científico ao dispor de uma fundamentação teórica a ser utilizada como embasamento para pesquisadores de diversas áreas, tais como a ciência da computação, estatística e a matemática.

Como motivação pessoal, surge o interesse acerca das temáticas devido aos conhecimentos adquiridos no curso de Bacharelado em Biblioteconomia e Ciência da Informação (BCI) pela Universidade Federal de São Carlos (UFSCar). Outro aspecto a ser destacado como motivação pessoal, se dá por questões que contemplam o âmbito profissional, tal como a experiência com *Marketing* - um dos setores em que o *Big Data* possui forte participação. A ampliação deste conhecimento, somada à experiência acadêmica e profissional, contribuem para novas possibilidades de atuação.

Como motivação social, se evidencia a necessidade de levantar discussões acerca de pautas que vêm se destacando nos últimos anos: a privacidade das informações pessoais na *Web*. A conscientização da sociedade no que diz respeito à coleta, tratamento, armazenamento e disseminação dessas informações, trata-se de uma forma efetiva para alertá-los sobre os seus direitos. Por outro lado, no contexto das organizações, alertá-los sobre os seus deveres éticos. De mesmo modo, através do levantamento dessas discussões, cabe uma reflexão acerca dos profissionais da Ciência da Informação e o papel que desempenham perante esta nova realidade. Portanto, o presente estudo se justifica pelos interesses que contemplam o desenvolvimento científico, pessoal e social.

### 1.4 Estrutura do trabalho

Na Introdução, aborda-se os atributos iniciais que contextualizam as temáticas centrais da pesquisa: *Big Data*, metadados, privacidade e a relação deste fenômeno com a Biblioteconomia e a Ciência da Informação. O capítulo também apresenta a definição do problema de pesquisa, os objetivos gerais e específicos, a justificativa para o desenvolvimento da pesquisa e a estrutura do trabalho.

O capítulo 2 dispõe dos conceitos encontrados na literatura acerca das temáticas centrais desta pesquisa. Dessa forma, o capítulo está dividido em tópicos, no qual apresenta-se:



embasamentos sobre o *Big Data*; conceitos e funções dos metadados e aporte teórico sobre a privacidade na *Web*.

O capítulo 3 dispõe as classificações deste estudo, como sua fundamentação, natureza e tipologia. O capítulo também descreve a metodologia utilizada ao longo do desenvolvimento da pesquisa.

O capítulo 4 apresenta o levantamento do estado da arte sobre o fenômeno do *Big Data* e da privacidade informacional no cenário da CI. O capítulo também caracteriza uma análise acerca da coleta de informações da empresa *Amazon*, através dos *Cookies*, formulários e sua Política de Privacidade.

O capítulo 5 apresenta as considerações finais englobando as temáticas desta pesquisa: *Big Data*, Privacidade e Metadados.

O capítulo 6 apresenta as referências bibliográficas dos artigos científicos, teses, livros, sites e demais fontes de informações que fundamentaram a presente investigação.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1 Embasamentos sobre o *Big Data*

Diante do avanço tecnológico e da expansão da *Internet*, intensificou-se a produção e troca de informações dos indivíduos e organizações. Nessa conjuntura, surge o *Big Data*, termo este que está em evidência na Ciência da Informação, Ciência da Computação, Tecnologia da Informação e demais áreas correlatas.

Segundo Chen, Mao e Liu (2014) o *Big Data* trata-se de conceito abstrato, mas que pode ser definido como um "conjunto de dados que não poderia ser percebido, adquirido, gerenciado e processado por ferramentas tradicionais de TI e *software/hardware* dentro um tempo tolerável"<sup>1</sup> (CHEN; MAO; LIU, p. 173, 2014, tradução nossa).

Esses conjuntos de dados são medidos em *exabytes*<sup>2</sup>, derivam-se de inúmeras fontes heterogêneas e autônomas, com relacionamentos complexos e estão em constante expansão. O tamanho e a complexidade tornam o processamento difícil e seu formato não se limita apenas a números, como também podem ser fotos, vídeos, mapas, palavras e frases (GIRI; LONE, 2014).

Grande parte desses dados são gerados em atividades na *Web*, através de ações em sites e redes sociais, como os fluxos de cliques, *posts*, compartilhamentos, *tweets*, entre outros. Dados de transações bancárias e dados gerados pelos indivíduos, como documentos eletrônicos, registros médicos e ligações telefônicas, também são alguns exemplos de ações que produzem uma grande quantidade de dados.

De acordo com Giri e Lone (2014) no *Big Data* existem três tipos de dados: estruturados, semiestruturados e não estruturados. Os dados estruturados são agrupados em linhas e colunas, são dados fáceis de analisar, como palavras e números. Geralmente são gerados em dispositivos eletrônicos, como smartphones, dispositivos de sistema de posicionamento global (GPS), entre outros. Os dados semiestruturados não seguem um esquema fixo, mas contém *tags* e marcadores que definem uma hierarquia aos dados. O XML (*Extensible Markup Language*) é um exemplo de linguagem de marcação utilizada para estruturá-los. Por fim, os dados não estruturados, que são dados complexos, os quais não podem ser separados em categorias ou analisados

---

<sup>1</sup>"In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time".

<sup>2</sup> Um *exabyte* é formado por 10<sup>18</sup> bytes. Os *exabytes* são utilizados para medir a soma de várias redes de armazenamento.

numericamente de forma fácil. Imagens, vídeos, arquivos de áudio, comentários em redes sociais, são alguns exemplos de dados não estruturados.

Atualmente, o *Big Data* pode ser considerado como minas de ouro para as empresas, visto que ao explorarem esses dados, é possível extrair informações altamente relevantes para os movimentos estratégicos das organizações. No entanto, para algumas empresas a implementação dessas tecnologias ainda são vistas como um desafio e para aquelas que já utilizam, ainda são de forma restrita e limitada. Bhadani e Jothimani (2016) expõem em sua obra alguns dos tipos de dados que são gerados nos diferentes setores, os quais são apresentados no **Quadro 1**.

**Quadro 1** – Diferentes fontes de dados

Setor	Dados produzidos	Uso
Astronomia	Movimento de estrelas, satélites, etc.	Para monitorar as atividades de corpos de asteroides e satélites.
Financeiro	Conteúdos de notícias por vídeo, áudio, <i>Twitter</i> e reportagem.	Para tomar decisões comerciais.
Assistência médica	Registros médicos eletrônicos e imagens.	Para ajudar na saúde pública em curto prazo e longo prazo para programas de pesquisas epidemiológicas.
<i>Internet</i> das coisas	Sensor de dados.	Para monitorar diversas atividades em cidades inteligentes.
Ciências da vida	Sequências de genes.	Para analisar variações genéticas e potencial eficácia dos tratamentos.
Mídia / Entretenimento	Conteúdos e visualizações dos usuários.	Para captar mais espectadores.
Mídia Social	Posts de <i>blogs</i> , <i>tweets</i> , redes sociais, detalhes de registros.	Para analisar padrões de comportamentos dos clientes.
Telecomunicações	Registros de detalhes das chamadas.	Gestão dos clientes.
Transporte, Logística, Varejo, Serviços de Utilidade Pública	Dados de sensor gerados da frota, leitores de tag RFID e medidores inteligentes.	Para otimizar operações.
Videovigilância	Gravação de CCTV para IPTV câmeras e sistema de gravação.	Para analisar padrões de comportamentos e para melhoria dos serviços de segurança.

**Fonte:** Adaptado de Bhadani, 2016.

A partir das informações apresentadas por Bhadani e Jothimani (2016), é possível ter uma percepção da variedade de setores em que se há produção de dados e algumas de suas finalidades.

No setor astronômico, os dados produzidos, processados, armazenados e disseminados, são utilizados para monitoramento dos corpos celestes e para o funcionamento de satélites, equipamentos estes essenciais para a comunicação, ciência e desenvolvimento tecnológico. No setor financeiro, os dados são produzidos através de conteúdos de notícias e reportagens, sendo estes utilizados para decisões comerciais. Em assistência médica, os dados provêm de inúmeras fontes, como prontuários dos pacientes, prescrições e anotações, imagens médicas, dados clínicos e dados gerados a partir dos aparelhos de exames e auxiliam em questões de saúde pública. Nos setores de comunicação, mídia e entretenimento, os dados são gerados através de ações na *Web*, como o fluxo de cliques, visualizações, publicações, compartilhamentos, registros de ligações e são utilizados para diversos fins, dentre estes, a análise de comportamento dos usuários - ação essa muito utilizada para estratégias de *Marketing*. Nos setores de logística, transportes, serviços e videovigilância, os dados gerados são utilizados para otimizar operações e serviços de segurança. Dessa forma, observa-se que essa produção de dados não se limita apenas ao ambiente *Web*, mas contempla diversas outras Tecnologias da Informação e Comunicação (TIC), como telecomunicações, videovigilância e até mesmo em setores de Ciências da Vida.

Doug Laney (2001) utilizou três atributos para definir o *Big Data*, o volume, a velocidade e a variedade dos dados, os quais ficaram conhecidos como os “Três V’s do *Big Data*”. No entanto, como o *Big Data* está em constante expansão, ao longo dos anos novos atributos foram selecionados para caracterizá-lo. Atualmente, podemos destacar sete atributos principais, que são: volume, velocidade, variedade, veracidade, variabilidade, volatilidade e valor.

O volume está relacionado com o tamanho dos dados que são gerados continuamente em diversas fontes, como textos, arquivos de áudio, vídeos, estudos de pesquisas, imagens de satélites, ações em redes sociais, conteúdos em páginas da *Web*, transações financeiras e inúmeras outras (GUPTA; KHAN; UDDIN, 2014). O volume desses dados representa desafios para as estruturas de TI convencionais, dado que, muitas empresas possuem grandes quantidades de dados arquivados, mas não têm capacidade de processá-los (ISWARAPPA; ANURADHA, 2015).

A velocidade se refere à velocidade em que esses dados são criados, por consequência, a velocidade em que esses dados podem ser processados, armazenados e analisados por bancos de dados relacionais (ISWARAPPA; ANURADHA, 2015). Gupta, Khan e Uddin (2014) afirmam que a alta velocidade de produção desses dados está diretamente relacionada com o volume e que para armazenar dados de entrada tão rápidos, torna-se necessário efetuar uma análise em tempo real.

A variedade representa os diversos formatos em que os dados são gerados, sejam eles estruturados, como dados numéricos em bases de dados tradicionais ou não estruturados, como as ações nas redes sociais. Iswarappa e Anuradha (2015) apontam que quando se trata de *Big Data*, a variedade também é um grande desafio para os analistas, os quais precisam conhecer a categoria dos dados em que estão lidando. De acordo com os autores, grande parte dos dados gerados não estão estruturados, o que aumenta a complexidade do armazenamento e da análise dos mesmos dados.

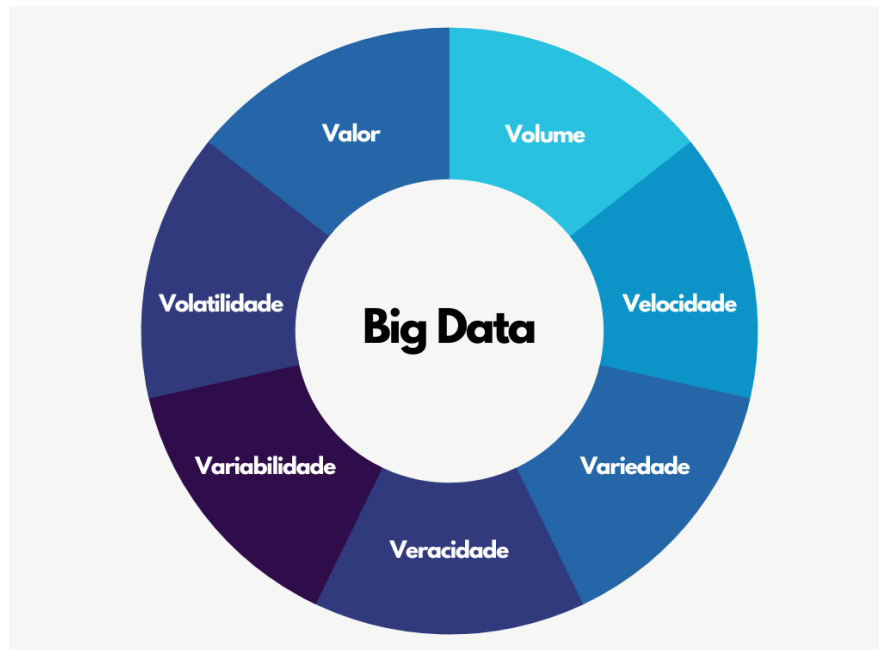
A veracidade representa a necessidade de garantir que os dados coletados sejam autênticos e verdadeiros no momento em questão (MACHADO, 2018). Quando se está lidando com um alto volume de dados, é quase impossível que todos os dados estejam totalmente corretos, o que faz com que a qualidade dos dados varie, a precisão dos dados da análise depende da veracidade dos dados de origem (ISWARAPPA; ANURADHA, 2015).

A variabilidade é o atributo que representa as instabilidades do fluxo de dados, devido à velocidade e variedade dos dados. Tem-se como exemplo, o repositório do *Twitter* que armazena e gere diversos tipos de dados, os quais podem ter vários sentidos e significados. Para que um *tweet* que tem a mesma palavra com vários significados passe por uma análise adequada, os algoritmos devem ser capazes de entender o contexto e decifrar o significado exato da palavra nesse contexto (SILVA, 2020).

A volatilidade está relacionada com o período em que os dados estarão disponíveis. Gupta, Khan e Uddin (2014) afirmam que a volatilidade se torna significativa por conta do volume, variedade e velocidade dos dados que compõem o *Big Data*.

O valor é um dos principais atributos do *Big Data*. Segundo Iswarappa e Anuradha (2015), o *Big Data* possui um valor potencial enorme e os benefícios de acesso, mas ao menos que seja possível transformá-lo em valor, ele se tornará inútil. Este é inclusive, um dos atributos mais relevantes para as empresas, pois extrair o máximo de valor desses dados é o resultado esperado com o investimento nessa tecnologia. A **Figura 1** apresenta os sete V's do *Big Data*.

**Figura 1** - Sete V's do *Big Data*



**Fonte:** Elaborado pela autora.

Nesse viés, para o processamento e armazenamento do imenso fluxo de informações gerados pelo *Big Data*, torna-se necessário o uso de tecnologias específicas. Em suma, existem diversas tecnologias para tal finalidade, mas a escolha da tecnologia depende de múltiplos fatores, como tipologia dos dados, formatos de origem, fontes de informação, entre outros. *Hadoop*, *NoSQL* e *Data Mining* são algumas das tecnologias que serão abordadas a seguir.

Quando o assunto é tecnologia que envolve o *Big Data*, o *Apache Hadoop* é sempre mencionado. De modo geral, o *Hadoop* trata-se de uma estrutura que permite o processamento distribuído de grandes conjuntos de dados em *clusters* de computadores, utilizando modelos de programação simples (APACHE HADOOP, 2021). Dentre as vantagens do *Hadoop*, está a sua alta flexibilidade, capacidade, baixo custo e confiabilidade no gerenciamento e processamento de dados estruturados e não estruturados (DA SILVA, 2018).

A segunda tecnologia do *Big Data* a ser destacada, é o *NoSQL* (*Not Only SQL*) que consiste em bancos de dados não relacionais e que acomodam dados não estruturados. A diferença entre os bancos de dados relacionais (*SQL*) com os bancos de dados não relacionais (*NoSQL*) está em seus esquemas. No primeiro caso, os esquemas são rígidos e no segundo caso, são mais flexíveis (BHOGAL; CHOKSI, 2015). Os bancos de dados *NoSQL* oferecem suporte para aplicativos interativos, dispositivos móveis, organizações em setores *on-line*, soluções corporativas em nuvem, redes sociais, entre outros (BHOGAL; CHOKSI, 2015).

O *Data Mining* (Mineração de Dados) é o processo analítico no qual os dados são explorados com intuito de identificar padrões relevantes, através de algoritmos matemáticos. Segundo Han, Kamber e Pei (2011, p. 21), o *Data Mining* incorporou muitas técnicas, como estatísticas, linguagens de máquinas, reconhecimentos de padrões, bancos de dados, algoritmos, entre outros. A interdisciplinaridade de pesquisa e desenvolvimento do *Data Mining* contribui para o sucesso da mineração dos dados e suas extensas aplicações.

O *Big Data* requer tecnologias inovadoras, coleta, armazenamento e processamento rápido e eficiente. A depender dos objetivos de cada organização, os tipos de dados a serem trabalhados, os formatos de origem e as fontes de informação, é necessário recorrer até mesmo ao uso de mais de uma tecnologia.

Dentre os benefícios do *Big Data*, pode-se destacar o seu valor em diferentes formas, sendo algumas delas: criar transparência ao facilitar o acesso aos dados às partes interessadas, permitir a experimentação, descoberta de necessidades e propor melhorias de desempenhos para níveis mais elevados; segmentar públicos para personalizar ações (abordagem muito conhecida em *Marketing*, mas que também pode ser revolucionária em outros lugares); substituir ou apoiar a tomada de decisão humana com a análise de dados; entre muitos outros (MANYIKA et al., 2011). O *Big Data* tem o poder de oferecer um grande potencial para empresas de criação de novos negócios e desenvolvimento de novos produtos e serviços, através da melhoria de suas operações (LEE, 2017).

Entretanto, assim como ocorre com qualquer inovação disruptiva, o *Big Data* também apresenta muitos desafios em sua adoção (LEE, 2017). Giri e Lone (2014) destacam alguns desafios, como a necessidade de filtrar os dados, visto que nem todos os dados gerados são úteis; a propriedade dos dados; a complexidade por conta dos formatos variados; o consumo de energia e a segurança.

Em relação ao *Big Data*, a segurança e a privacidade dos dados é o aspecto que mais causa preocupação entre as pessoas. Giri e Lone (2014) ressaltam que “uma vez que os dados tenham sido coletados, a questão é como proteger e controlar esses dados, ou seja, como manter esses dados seguros. Porque os dados contêm informações confidenciais, como registros financeiros”<sup>3</sup> (GIRI; LONE, p. 528, 2014, tradução nossa). Para manter e analisar esses conjuntos de dados, as organizações necessitam de profissionais e ferramentas próprias para a análise, o

---

<sup>3</sup> “Once the data has been collected the question is how to protect and control this data i.e. how to secure this data. Because the data contains sensitive information like financial records”.

que aumenta o potencial de riscos de segurança, em virtude de que os dados são entregues para terceiros (CHEN; MAO; LIU, 2014).

O *Big Data* funciona como uma ferramenta estratégica muito eficiente para as organizações. No entanto, para implementar suas tecnologias, é imprescindível que as organizações invistam em uma infraestrutura de excelência, em mão de obra qualificada e desenvolvam uma política de segurança informacional.

## 2.2 Os conceitos e funções dos metadados

Diante ao desenvolvimento das TIC's e do grande fluxo informacional, coube aos profissionais da Ciência da Informação a adaptação às novas formas de organização da informação e do conhecimento registrado. Dentre as adaptações ocorridas, pode-se destacar a utilização de metadados em objetos digitais. Popularmente conhecido como “dados sobre dados”, as definições podem variar de acordo com a área no qual se insere, de modo geral, os metadados são atributos que representam uma entidade, através de padrões específicos.

Para Dempsey et al. (1997, p. 5):

Metadados são dados que descrevem os atributos de um recurso. Normalmente, ele suporta várias funções: localização, descoberta, documentação, avaliação, seleção e outros. Essas atividades podem ser realizadas por usuários humanos ou seus agentes (humanos ou automatizados)<sup>4</sup> (DEMPSEY et al., p. 5, 1997, tradução nossa).

Dessa forma, os metadados não se limitam apenas em ambientes digitais, dado que eles são utilizados há anos por bibliotecários na descrição de recursos, através dos catálogos. Em um catálogo é possível encontrar informações que descrevem o recurso como o número de chamada, nome do autor, data e local de publicação, entre outros. Metadados esses que facilitam o acesso dos usuários ao material desejado (POMERANTZ, 2015).

Pereira (2005) define que “a principal finalidade dos metadados é documentar e organizar de forma estruturada os dados nas Unidades de Informação (UI), com o objetivo de minimizar a duplicação de esforços e facilitar a manutenção dos dados”.

Reis e Sá (2020) afirmam que:

A proliferação de formatos de arquivos digitais e a explosão informacional vivida atualmente, com a produção de conteúdo digital por smartphones e outros dispositivos conectados à *Internet*, geram uma grande quantidade de

---

<sup>4</sup> “*Metadata is data which describes attributes of a resource. Typically, it supports a number of functions: location, discovery, documentation, evaluation, selection and others. These activities may be carried out by human endusers or their (human or automated) agents*”.



metadados. A quantidade disponível e sua relação indissociável do cotidiano do indivíduo tornam os metadados a mais rica fonte para a análise e extração de informações sobre consumidores e seus hábitos, despertando o interesse de empresas que buscam atingir vantagem competitiva através de ferramentas de análise de dados. Os metadados são, portanto, o principal insumo do *Big Data* (REIS; SÁ, p. 237, 2020).

Como apontado pelos autores, os metadados podem ser considerados como o principal insumo do *Big Data*. Desta maneira, nos tópicos seguintes serão abordadas algumas classificações dos Metadados e o papel crucial que desempenham no *Big Data*.

Segundo o que Trust (2008) expõe em sua obra, os metadados podem ser divididos em categorias de acordo com as funções dos dados, como é possível visualizar no **Quadro 2**:

**Quadro 2** - Tipos de metadados e suas funções

Tipo	Definição	Exemplos
Administrativo	Metadados usados no gerenciamento e administração de coleções e recursos de informação	Informações de aquisição Rastreamento de direitos e reprodução Documentação dos requisitos legais de acesso Informações de localização Critérios de seleção para digitalização
Descritivo	Metadados usados para identificar e descrever coleções e recursos de informação relacionados	Registros de catalogação Encontrar auxílios Diferenciações entre versões Índices especializados Informação curatorial Relações com hiperlink entre recursos Anotações de criadores e usuários
Preservação	Metadados relacionados à preservação e gerenciamento de coleções e recursos de informação	Documentação da condição física dos recursos Documentação das ações tomadas para preservar as versões físicas e digitais dos recursos, por exemplo, atualização e migração de dados Documentação de quaisquer alterações que ocorram durante a digitalização ou preservação
Técnico	Metadados relacionados às funções do sistema ou metadados se comportam	Documentação de <i>hardware</i> e <i>software</i> Informações técnicas de digitalização, por exemplo, formatos, taxas de compressão,

		dimensionamento rotinas Rastreamento dos tempos de resposta do sistema Dados de autenticação e segurança, por exemplo, chaves de criptografia, senhas
Uso	Metadados relacionados ao nível e tipo de uso de coleções e recursos de informação	Registros de circulação Registros de exposições físicas e digitais Rastreamento de uso e usuário Reutilização de conteúdo e informações de várias versões Registros de pesquisa Metadados de direitos

**Fonte:** TRUST, 2008 - Página 09

No **Quadro 2** é possível visualizar os tipos de metadados, as suas definições e exemplos de suas aplicações nas unidades de informação. Apesar de suas diferenças, os metadados possuem como principal função auxiliar na descrição e identificação dos objetos, sejam eles físicos ou digitais, facilitando a gestão e a recuperação dele.

Em especial, para este trabalho, os metadados de uso serão evidenciados. Assim, os metadados classificados como “Metadados de Uso” referem-se aos dados coletados de acordo com a atividade nos ambientes digitais e são relacionados ao nível e tipo de uso dos recursos de informação. Além das informações sobre os indivíduos, os metadados podem fornecer dados valiosos sobre redes sociais e as conexões entre indivíduos, locais e organizações (POMERANTZ, 2015). Tais informações são criadas automaticamente e armazenadas pelo sistema e são valiosas para empresas de *Marketing* e entidades de negócios, visto que contém diversos dados de seus usuários.

Pomerantz (2015) afirma que:

A *Amazon*, por exemplo, coleta uma grande quantidade de metadados sobre seus usuários. Para comprar qualquer coisa na *Amazon*, você deve criar um perfil que inclua no mínimo um número de cartão de crédito e um endereço para o qual enviar os itens. Ela captura dados adicionais sobre você: quais itens você compra, quais itens você analisa, quaisquer comentários que você escreve etc. Isso não é exclusivo da *Amazon*, é claro; todos os fornecedores *on-line* coletam tipos semelhantes de dados<sup>5</sup>(POMERANTZ, p. 125, 2015, tradução nossa).

<sup>5</sup> “*Amazon, for example, collects a great deal of metadata about its users. In order to purchase anything on Amazon, you must create a profile, which at a minimum includes a credit card number and an address to ship items to. Amazon then captures additional data about you: what items you buy, what items you look at, any reviews that you write, etc. This is not unique to Amazon, of course; all online vendors collect similar types of data*”.

Como visto anteriormente, dentre os desafios do *Big Data*, o que se sobressai são as questões de privacidade e segurança dos dados (e metadados) que são recolhidos continuamente. Portanto, nos tópicos seguintes são descritas algumas das informações que são coletadas no ambiente *Web*, bem como, a maneira em que essas informações são coletadas e tudo o que norteia a privacidade e segurança desses dados.

### 2.3 Aporte teórico sobre a privacidade na *Web*

Antes de mais nada, é preciso refletir acerca do conceito de privacidade na *Web*, tema este que está em proeminência em razão da enorme quantidade de informações pessoais que estão sendo coletadas, armazenadas e difundidas através da *Internet*.

Peixoto e Júnior (2018) promovem uma contextualização no que diz respeito à privacidade e suas possíveis formas de ofensas. Segundo os autores, a privacidade é uma das palavras mais multifacetadas que se pode trazer ao debate, devido aos seus sentidos atribuídos ao longo dos tempos e culturas. No que tange a compreensão da privacidade, os autores afirmam que a melhor forma para a entender, é quando ela é dividida em três dimensões distintas, entretanto, complementares que são: dimensão decisional, dimensão espacial e dimensão informacional.

A dimensão decisional da privacidade está relacionada com as particularidades do indivíduo, como suas escolhas, gostos, projetos e características, de forma com que se tenha autonomia individual sem que haja interferência alheia. A dimensão espacial é a mais tradicional dimensão da privacidade, a qual está relacionada com um determinado lugar físico, como os cômodos da casa. A dimensão informacional da privacidade, como o próprio nome diz, está relacionada com as informações do indivíduo. Compreender essas dimensões possibilita identificar os problemas relativos à elas (PEIXOTO; JÚNIOR, 2018).

Embora as dimensões propostas pelos autores possam coexistir, no que permeia o ambiente *Web*, a considerada "dimensão informacional da privacidade" se mostra em evidência. A violação à privacidade na *Web* pode ocorrer através da divulgação, venda e/ou coleta de dados pessoais, sem que o indivíduo esteja ciente dessas ações. Entende-se por dados pessoais, as informações que permitem identificar um indivíduo, direta ou indiretamente, como nome, Registro Geral (RG), Cadastro de Pessoa Física (CPF), endereço, comportamento na *Web*, hábitos de consumo, entre muitos outros. Mesmo informações pessoais não explícitas, como os

hábitos de consumo, podem causar impactos para o indivíduo, haja vista que com a tecnologia específica, cruzar os dados e identificar a identidade do usuário pode não ser uma tarefa difícil.

Os sites coletam uma diversidade de informações dos indivíduos, como por exemplo, as que se dizem respeito à vida dos mesmos. Conhecidas como *PII (Personally identifiable information)* são informações relativas a uma determinada pessoa, desde características físicas até hábitos variados, de modo em que o cruzamento desses dados torna possível traçar um verdadeiro perfil do indivíduo (MATOS, 2005). No **Quadro 3** é possível visualizar as classificações das informações que são coletadas:

**Quadro 3** - Classificações das informações coletadas *on-line*

Categoria	Descrição
Informações anônimas	Informações coletadas sobre visitas à página, sem o uso de tecnologias invasivas, normalmente são informações padrão enviadas em qualquer solicitação da <i>Web</i> ou <i>Internet</i> . Essas informações incluem o endereço IP da máquina, tipo de domínio, versão e tipo do navegador, sistema operacional, idioma do navegador, tempo e local.
Informações pessoais não identificáveis	Informações que, consideradas de forma isolada, não podem ser utilizadas para identificar ou localizar um indivíduo. Informações como idade, data de nascimento, gênero, ocupação, escolaridade, renda, CEP sem endereço, interesses e <i>hobbies</i> . Os consumidores divulgam a maioria dessas informações por meio de botões e menus nas páginas da <i>Web</i> . Além disso, também há tecnologias sofisticadas para a coleta dessas informações, como por exemplo, os <i>cookies</i> . Embora essas tecnologias não identifiquem o usuário individualmente, permitem traçar o seu perfil de forma eficaz.
Informações pessoais identificáveis	Informações que permitam identificar ou localizar o indivíduo. Essas informações incluem nome, endereço, número de telefone, número de cartão de crédito, número de seguro social, endereço de e-mail, etc. Essas informações são quase sempre coletadas de forma explícita, quando os usuários se cadastram nos sites ou participam de transações financeiras.

**Fonte:** Adaptado de Chellappa e Sin (2005)

Observando o **Quadro 3**, é possível inferir que para qualquer ação que o usuário tenha no ambiente da *Web*, seja para realizar uma compra, acessar uma rede social ou simplesmente utilizar o buscador, ele necessita fornecer ao menos uma das informações categorizadas por Chellappa e Sin (2005). Para a coleta de informações na *Web*, utiliza-se técnicas e tecnologias específicas para tal finalidade. Conforme apresentado por Silva (2020), no **Quadro 4** é possível

visualizar alguns dos mecanismos que são utilizados para a recolha de informações dos consumidores *on-line*.

**Quadro 4** – Mecanismo de recolha de informações dos consumidores *on-line*

Mecanismos	Impacto na privacidade
<i>Smartphones e apps</i>	Usado para rastrear a localização e compartilhar fotos, endereços, números de telefone, e outros comportamentos.
Redes de Publicidade	Usado para rastrear indivíduos que se movimentam entre milhares de sites.
Redes Sociais	Usado para recolher informações sobre conteúdo fornecido pelo utilizador, como livros, música, amigos e outros interesses, preferências e estilos de vida.
<i>Cookies</i> e Supercookies	Usado para rastrear indivíduos num único local. Supercookies são quase impossíveis de identificar ou remover.
<i>Spyware</i>	Pode ser usado para registar toda a atividade de teclado de um utilizador, incluindo sites visitados e códigos de segurança usados; também usado para exibir anúncios a utilizadores com base nas suas pesquisas ou outro tipo de comportamento.
<i>Search engine behavioural targeting</i>	Utiliza o histórico de pesquisas anteriores, dados demográficos, interesses expressos, geográficos ou outros dados inseridos pelo utilizador para segmentar a publicidade.
Carrinho de Compras	Pode ser usado para recolher informações detalhadas de pagamento e compra.
Formulários	Formulários <i>on-line</i> que os utilizadores preenchem voluntariamente em troca de um benefício ou recompensa prometidos que são vinculados ao <i>clickstream</i> ou a outros dados comportamentais para criar um perfil pessoal.
<i>Logs</i> de transações do site	Pode ser usado para recolher e analisar informações detalhadas sobre o conteúdo da página visualizada pelos utilizadores.
Motores de busca	Pode ser usado para rastrear declarações e opiniões que os utilizadores manifestam na <i>Web</i> .

**Fonte:** Laudon e Traver (2019) apud. Silva (2020)

O **Quadro 4** apresenta alguns dos mecanismos que são utilizados para a coleta de informações *on-line*. Nos tópicos seguintes, contextualiza-se de maneira detalhada como funciona a recolha de informações através dos formulários *on-line* e dos *cookies*. Para a escolha destes, levou-se em consideração que esses métodos serão a principal fonte de análise para o desenvolvimento desta pesquisa.

Os formulários são meios explícitos para a coleta de dados pessoais dos indivíduos. Neste modelo, os visitantes oferecem de maneira espontânea ao site, os dados que estão sendo solicitados, que variam entre os mais básicos como nome, endereço, profissão, aos mais íntimos, como áreas do seu interesse. Embora essas informações aparentem inofensivas, se comercializadas ou cruzadas com outras *PIIs*, podem compor um dossiê físico-psíquico do indivíduo (MATOS, 2005).

Os formulários são muito utilizados para cadastros em redes sociais, *blogs* e principalmente, sites de comércio eletrônico. Em sua maioria, para que o usuário possa usufruir dos recursos que deseja, como por exemplo, efetuar compras, ele necessita realizar o cadastro.

A **Figura 2** apresenta o formulário de cadastro do *Facebook*:

**Figura 2** - Formulário de cadastro do *Facebook*

The image shows a screenshot of the Facebook registration form, titled "Cadastro-se" (Sign up). The form is overlaid on a blurred background of the Facebook homepage. The form includes the following fields and options:

- Nome** (Name) and **Sobrenome** (Surname) text input fields.
- Celular ou email** (Cell or email) text input field.
- Nova senha** (New password) text input field with a strength indicator icon.
- Data de nascimento** (Date of birth) section with three dropdown menus for day (10), month (Set), and year (2022).
- Gênero** (Gender) section with three radio button options: **Feminino** (Female), **Masculino** (Male), and **Personalizado** (Custom).
- A green **Cadastre-se** (Sign up) button at the bottom.

Below the form, there is a small disclaimer: "As pessoas que usam nosso serviço podem ter carregado suas informações de contato no Facebook. Saiba mais." (People who use our service may have loaded their contact information on Facebook. Learn more.) and a link to the Terms, Privacy Policy, and Cookies Policy: "Ao clicar em Cadastre-se, você concorda com nossos Termos, Política de Privacidade e Política de Cookies. Você poderá receber notificações por SMS e cancelar isso quando quiser." (By clicking Sign up, you agree to our Terms, Privacy Policy, and Cookies Policy. You may receive notifications by SMS and cancel this when you wish.)

**Fonte:** Elaborado pela autora.

Na Figura 2, é possível visualizar que para o uso da rede social é necessário fornecer uma série de informações que se classificam como “Informações pessoais identificáveis”, ou seja, o nome, sobrenome, celular/e-mail que possibilitam a identificação do usuário. Além

disso, também é necessário identificar o “Gênero” do usuário, o que se enquadra como “Informações pessoais não identificáveis”.

Os *cookies* são pequenos arquivos de textos criptografados que estão localizados em diretórios do navegador. Eles são utilizados por desenvolvedores da *Web* para ajudar o usuário a navegar nos sites de forma eficiente e executar determinadas funções. Os *cookies* são criados quando o usuário acessa um site, o site envia informações para o navegador e então, o navegador cria os arquivos de textos (ALL ABOUT COOKIES, 2021).

Os *cookies* permitem que o usuário faça *login* em uma página, se mova para outra página e ainda sim permaneça conectado. De mesmo modo, também permitem que o usuário defina as preferências para a exibição de uma página para que sejam lembradas na próxima vez em que for acessá-la. Além disso, os *cookies* também podem ser utilizados para monitorar as páginas em que o usuário visita entre os sites, permitindo que os anunciantes construam uma imagem dos seus interesses (COOKIEPEDIA, 2020).

Existem diversos tipos e inúmeras formas para utilizar os *cookies*, a Cookiepedia (2020) disponibiliza algumas classificações que estão resumidas no **Quadro 5**:

**Quadro 5** - Tipos de *Cookies*

Classificação	Definição
<i>Cookies</i> Primários	São definidos ou recuperados pelo usuário apenas durante a visita, portanto, não podem ser utilizados para passar dados de um site para o outro. Entretanto, o proprietário do site ainda pode coletar os dados, utilizá-los ou vendê-los para outras organizações. Caso o faça, deve ser explicado na política de privacidade.
<i>Cookies</i> de Terceiros	São colocados em um site por meio de <i>Scripts</i> ou <i>Tags</i> adicionados à página da <i>Web</i> . Por exemplo, se um usuário visita um site que possui um vídeo do <i>YouTube</i> , que foi incluído pelo proprietário do site com um código fornecido pelo <i>YouTube</i> , o <i>YouTube</i> será capaz de definir <i>cookies</i> por meio desse código e saber que o usuário assistiu ou visitou a página em que o vídeo está. É o <i>cookie</i> mais utilizado pela publicidade <i>on-line</i> , dado que com o rastreamento do usuário (ou seu dispositivo) é possível construir um perfil comportamental do indivíduo.
<i>Cookies</i> de Sessão	São armazenados de forma temporária na

	memória do navegador e são destruídos quando o navegador é fechado.
<i>Cookies</i> Persistentes	São salvos no computador, de forma com que ao fechar e reiniciar, o <i>cookie</i> ainda estará lá. Eles são criados com uma data de expiração, ao atingir essa data, o <i>cookie</i> será automaticamente destruído pelo computador. No entanto, não há limites para a data.
<i>Cookies</i> Seguros	São transmitidos apenas via HTTPS, que geralmente se encontra em sites de compras <i>on-line</i> . Isso garante que os dados do <i>cookie</i> são criptografados.

**Fonte:** Adaptado de Cookipedia (2020)

Os *cookies* são ferramentas essenciais para a navegação *on-line*. Portanto, ao acessar uma página na *Web*, é certo que ao menos um dos *cookies* classificados no **Quadro 5** será utilizado. Como é possível visualizar no quadro, cada *cookie* dispõe de características específicas, mas de modo geral, todos possuem a função de armazenar e transferir informações. Levando em consideração a privacidade na *Web*, os “*Cookies* de Terceiros” são mais suscetíveis a problemas, dado que geralmente pertencem a outras páginas além da que está sendo acessada - e que por vezes, o usuário não possui o conhecimento a quem se refere.

Sendo assim, a partir do que foi exposto, é possível perceber que a abundância de dados e metadados que dão luz ao *Big Data*, são gerados e processados a partir de diversas fontes e possuem extrema relevância em múltiplos setores da sociedade. Contudo, apesar dos benefícios, a sua implementação ainda possui inúmeros desafios, como o alto investimento em sua infraestrutura, a necessidade de mão de obra qualificada e desenvolvimento de políticas de segurança informacional, sendo estas extremamente necessárias para garantir o direito à proteção nas práticas de coletas e tratamento de dados pessoais. Nesse contexto, conforme apontado por Reis e SÁ (2020), o mercado do *Big Data* se mostra um novo campo de atuação para Bibliotecários e Cientistas da Informação, tendo em vista que os conhecimentos de classificação, análise, indexação e recuperação da informação, são algumas das atribuições dos profissionais que atuam com o *Big Data*. Com base nisto, é válido destacar que os profissionais da informação, além da atuação direta com o *Big Data* e suas tecnologias, também podem assumir funções em outras frentes relacionadas, como no suporte ao desenvolvimento das Políticas de Segurança e Privacidade Informacional e no suporte à adequação às normas, como a LGPD.



### 3. PROCEDIMENTOS METODOLÓGICOS

De caráter investigativo, o presente estudo se configura como abordagem exploratória, de natureza qualitativa, cunho teórico e documental. A exploração da literatura existente acerca da temática, tem por objetivo estruturar um conhecimento teórico a respeito do fenômeno do *Big Data*, os Metadados e questões que implicam a Privacidade mediante ao desenvolvimento dessa tecnologia.

Dessa maneira, o estudo partiu do princípio de realizar uma investigação no que diz respeito ao *Big Data*, visando compreender suas tecnologias, benefícios, desafios e convergências com a privacidade na *Web*. Nesse sentido, com o objetivo de contextualizar a temática, realizou-se um levantamento da literatura disponível em fontes bibliográficas, contemplando autores da Ciência da Informação e áreas correlatas.

Martins e Theóphilo (2016, p.52) definem que “uma pesquisa bibliográfica procura explicar e discutir um assunto, tema ou problema com base em referências publicadas em livros, periódicos, revistas, enciclopédias, dicionários, jornais, sites, CDs, anais de congressos etc.” (apud SOARES; PICOLLI; CASAGRANDE, 2018). Portanto, como procedimento técnico, este estudo se classifica como uma pesquisa bibliográfica.

Em primeiro momento, o desenvolvimento deste estudo partiu da investigação teórica acerca das temáticas centrais desta pesquisa: *Big Data*, Metadados e Privacidade. Em segundo momento, a investigação se direcionou para uma análise empírica no site da *Amazon*, com o propósito de correlacionar as definições encontradas em fontes documentais. Em terceiro momento, a investigação se dirigiu ao levantamento do estado da arte sobre o fenômeno do *Big Data* no cenário da CI. Em quarto momento, após o levantamento e análise das informações, a investigação se direcionou para as considerações finais.

Neste escopo, a execução deste estudo ocorreu através de 6 etapas, as quais podem ser visualizadas no **Quadro 6**:

**Quadro 6** - Etapas do desenvolvimento da pesquisa

Etapas	Procedimento
1 <sup>a</sup>	Realizou-se um levantamento bibliográfico em nível nacional e internacional para a coleta de materiais acerca da temática central deste estudo. Para o levantamento, utilizou-se os termos descritores “ <i>Big Data</i> ”, “Privacidade”, Metadados, "Privacy" e “Metadata” nas seguintes bases de dados: Base de Dados em Ciência da Informação (BRAPCI), <i>Scientific Electronic Library Online</i> (SciELO) e Google

	Acadêmico;
2 <sup>a</sup>	Leitura e interpretação da síntese dos materiais recuperados, a fim de filtrar e selecionar materiais pertinentes ao desenvolvimento deste estudo;
3 <sup>a</sup>	Estruturação da fundamentação teórica com base na literatura recuperada;
4 <sup>a</sup>	Levantamento do estado da arte sobre o fenômeno do <i>Big Data</i> e da privacidade informacional no cenário da CI. Nesta etapa apenas a literatura recuperada na BRAPCI foi levada em consideração;
5 <sup>a</sup>	Análise e estruturação da caracterização da Política Corporativa da empresa <i>Amazon</i> a partir de fontes documentais;
6 <sup>a</sup>	Estruturação das considerações finais com base na fundamentação teórica e a análise das temáticas que compõem este estudo.

**Fonte:** elaborado pela autora

Além desse protocolo, com o objetivo de uma análise ainda mais enriquecedora, a escolha se deu pela *Amazon*, uma empresa multinacional de tecnologia norte-americana que atualmente se enquadra como um dos *e-commerces* mais influentes e valiosos do mundo, ao lado de outras empresas como o *Google*, *Facebook* e a *Apple*. A escolha pela *Amazon* se deu não apenas por ser uma das empresas mais influentes da atualidade, mas devido às suas estratégias que envolvem a coleta e análise de dados referente ao comportamento do consumidor.

## 4. RESULTADOS

### 4.1 Os temas *Big Data*, privacidade e metadados de uso na Ciência da Informação

Visto os conceitos que norteiam o *Big Data*, os metadados e a privacidade na *Web*, é possível ter uma perspectiva de sua relação com a Ciência da Informação e o papel dos profissionais deste campo perante o fenômeno. Com base nisto, realizou-se um levantamento bibliográfico na Base de Dados em Ciência da Informação (BRAPCI), a fim de visualizar por meio da literatura, alguns dos estudos realizados acerca destas temáticas.

Para o levantamento bibliográfico na BRAPCI, utilizou-se a variável booleana “AND”, “aspas” e os termos centrais deste estudo “*Big Data*” e “Privacidade”. Além disso, utilizou-se como delimitação para a busca, os artigos publicados entre os anos de 2010 a 2022. Ao todo foram recuperados onze trabalhos (artigos), dos quais por questões não identificáveis, apenas sete estavam disponíveis para download.

Os autores Szinvelski, Arceno e Francisco publicaram em 2019 o artigo intitulado “Perspectivas jurídicas da relação entre a *Big Data* e a proteção de dados”, o qual tem como objetivo discutir a harmonização entre o *Big Data* e a proteção dos dados pessoais. Como resultados, os autores chegaram à conclusão de que a mundialização proporcionada pela *Internet* gerou a mutação do que se conhece como “intimidade”, de forma com que os dados pessoais ocupam a centralidade das inovações - incluso o *Big Data*, um produto delas. Dessa forma, por se tratar de uma aplicação de uso inevitável pelas organizações públicas e privadas, os autores consideraram fundamental um alerta sobre três paradoxos do *Big Data*: a transparência, a identidade e o poder. Como perspectiva jurídica, cabe ao Direito atuar como mediador no relacionamento entre o *Big Data* e a proteção dos dados pessoais.

Os autores Lott e Cianconi publicaram em 2018 o artigo “Vigilância e privacidade, no contexto do *Big Data* e dados pessoais: análise da produção da Ciência da Informação no Brasil” com o objetivo de realizar um levantamento sobre os estudos acerca da vigilância e privacidade associadas ao *Big Data* e dados pessoais em pesquisas acadêmicas da ciência da informação nos últimos dez anos. A partir do levantamento, os autores concluíram que em 2018 a “vigilância” e a “privacidade” eram temáticas trabalhadas no meio acadêmico, no qual encontraram-se sete amostras para o termo “vigilância” e seis amostras para o termo “privacidade”. Ao realizar uma busca para relacionar os termos “*Big Data*” e “dados pessoais”,

os autores identificaram menos ocorrências. Além disso, como resultados os autores evidenciaram a importância de se discutir outras temáticas, como a carência de legislação para a proteção dos dados pessoais, a exploração desses dados e a ética na captação dessas informações, que até a publicação do artigo em 2018, eram pautas pouco discutidas no campo da Ciência da Informação.

O autor M. Pimenta publicou em 2013 o artigo denominado “*Big Data* e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado”. O artigo é produto de um outro projeto do mesmo autor, denominado “Dimensões tecnopolíticas do esquecimento: usos e abusos da memória na era digital”, o qual tem como objetivo entender a memória no *World Wide Web* como ação política da sociedade. Em seu segundo artigo, M. Pimenta dispõe alguns dos resultados da revisão de literatura acerca do *Big Data* e sua relação com a memória na era digital. O autor concluiu que se há um obstáculo para a privacidade na *Internet*, há também um desafio ao “segredo” dado que as empresas detêm todas as informações e não há garantias que esses dados são, de fato, deletados ao excluir um perfil nas redes sociais. Perante a isso, refletir sobre os desdobramentos dessa nova tecnologia, torna-se fundamental.

Rockembach publicou em 2020 o artigo “Estudos de usuários de arquivo e os desafios da Lei Geral de Proteção de Dados” com o objetivo de destacar alguns dos conceitos referentes aos usuários, ao comportamento informacional e aos desafios da implementação da LGPD. Como considerações finais, o autor concluiu que a LGPD veio para garantir os direitos sobre a informação quanto à anonimização e ao controle na exigência excessiva de dados por parte das organizações que realizam a coleta deles. Além disso, o autor afirma que os princípios éticos no meio digital são uma extensão da ética no mundo real, portanto faz-se necessário representar e colocar em prática as virtudes também nos ambientes digitais.

A autora Serra publicou em 2015 o artigo intitulado “Empréstimo de livros eletrônicos, redes sociais e a proteção de dados dos usuários” com intuito de discorrer sobre questões de privacidade e proteção de dados de usuários de bibliotecas, a partir do empréstimo de livros eletrônicos pela *OverDrive* em parceria com a *Amazon* e o uso de redes sociais. A motivação da autora por essa análise, se deu pela exposição de dados históricos e coleta de informações dos usuários das bibliotecas norte-americanas ao realizarem empréstimos de livros eletrônicos nessas plataformas. Como resultados, a autora concluiu que as bibliotecas podem utilizar informações dos usuários para oferecer sugestões personalizadas de acordo com cada perfil, no entanto, é essencial que se preocupem com a preservação e não exposição dos dados dos usuários, garantindo a sua confidencialidade.

Os autores Lima e Presser publicaram em 2022 o artigo “A Lei Geral de Proteção de Dados e os desafios para a gestão nas organizações brasileiras na era do Big Data” com o objetivo de analisar as ações que devem ser executadas pelas organizações ao gerir e proteger os dados pessoais no Brasil e o papel dos profissionais da Ciência da Informação nessas ações. Como considerações finais, os autores destacaram que a explosão de dados e informações criaram desafios para que as organizações consigam gerenciá-los e que embora exista a LGPD, as organizações não estão preparadas para este momento de produção massiva e dessa forma, devem reformular sua gestão técnica e administrativa. Acerca dos profissionais da Ciência da Informação, os autores destacaram se configurar como um promotor de mudanças muito positivas para a gestão das organizações, mas que ainda possui uma participação escassa nesses ambientes.

O autor Sant’ana publicou em 2016 o artigo intitulado “Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação” com o objetivo de propor um novo olhar ao Ciclo de Vida dos Dados, se amparando nos conceitos que a Ciência da Informação proporciona. Em sua pesquisa, o autor conclui que a Ciência da Informação oferece um arcabouço teórico para todo o processo de acesso aos dados, ainda que em conjunto com outras áreas, como a Ciência da Computação e a Matemática. No entanto, de acordo com o autor, os estudos da Ciência da Informação possuem maior aderência nas fases de “coleta” e “recuperação” dos dados, mas que apesar disto, não diminui sua importância nas demais fases.

No **Quadro 7** é possível visualizar de forma resumida os conceitos utilizados pelos autores como a definição do *Big Data* e palavras-chave que representam suas pesquisas. Além disso, apresenta-se parâmetros bibliométricos como o título, identificação dos autores, ano de publicação, título do periódico e temas:

**Quadro 7** - Conceitos de *Big Data*

<b>Título</b>	<b>Autores</b>	<b>Ano de publicação</b>	<b>Título Periódico</b>	<b>Temas</b>	<b>Conceito de <i>Big Data</i></b>	<b>Palavras-chave</b>
Perspectivas jurídicas da relação entre <i>Big Data</i> e proteção de dados	SZINVELS KI, Martín Marks; ARCEÑO, Taynara Silva; FRANCISCO, Lucas Baratieri	2019	Perspectivas em Ciência da Informação	Ciência da Informação; Direito Privado; <i>Big Data</i> ; Privacidade; Proteção de Dado Pessoal e Tecnologia	Conjuntos de dados variados e coletados em grande quantidade, deu forma em que as técnicas de análises	Ciência da Informação. Direito Privado. <i>Big Data</i> . Privacidade. Proteção de Dado Pessoal. Tecnologia.

					tradicionais se demonstram insuficientes	
Vigilância e privacidade, no contexto do <i>Big Data</i> e dados pessoais: análise da produção da Ciência da Informação no Brasil	LOTT, Yuri Monnerat; CIANCONI, Regina de Barros; CIANCONI, Regina de Barros	2018	Perspectivas em Ciência da Informação	Ciência da Informação; Vigilância; Privacidade; <i>Big Data</i> ; Dado Pessoal e ENANCIB	Conjuntos de dados massivos originários de situações reais, extremamente complexos e impossíveis de serem interpretados pela mente humana	Ciência da Informação. Vigilância. Privacidade. <i>Big Data</i> . Dado Pessoal.
<i>Big Data</i> e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado	PIMENTA, Ricardo Medeiros	2013	Tendências da Pesquisa Brasileira em Ciência da Informação	-	Conjunto complexo de dados em volume e velocidade extraordinários	-
Estudos de usuários de arquivo e os desafios da Lei Geral de Proteção de Dados	ROCKEMBAACH, Moisés	2020	Acervo - Revista do Arquivo Nacional	-	É provido de vários tipos de dados e de diversas fontes heterogêneas	-
Empréstimo de livros eletrônicos, redes sociais e a proteção de dados dos usuários	SERRA, Liliana Giusti	2015	Revista Digital de Biblioteconomia & Ciência da Informação	Biblioteconomia; Ciência da Informação; Livro Eletrônico; Proteção de Dado do Usuário;	Armazenamento de dados e processamento em alta velocidade	Biblioteconomia. Ciência da Informação. Livro Eletrônico. Proteção de Dado do Usuário. Rede

				Rede Social; Privacidade ; <i>Big Data</i> ; Livros Eletrônicos; Proteção de Dados dos Usuários; Redes Sociais		Social. Privacidade . <i>Big Data</i> . Livros Eletrônicos. Proteção de Dados dos Usuários. Redes Sociais.
A Lei Geral de Proteção de Dados e os desafios para a gestão nas organizações brasileiras na era do <i>Big Data</i>	LIMA, Paulo Ricardo da Silva; PRESSER, Nadi Helena	2022	Revista P2P e INOVAÇÃO	Cientista da Informação; Proteção de Dado; Gestão de Dado; Information Scientist; Data Protection; Data Management	Produção de dados em massa	Cientista da Informação. Proteção de Dado. Gestão de Dado.
Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação	SANTANA, Ricardo César Gonçalves; SANTANA, Ricardo César Gonçalves	2016	Informação & Informação	Ciclo de Vida do Dado; Ciência da Informação; <i>Big Data</i> ; Ciencia de la Información ; Ciência da Informação; Ciclo de Vida de lo Datos; Data Life Cycle; Information Science	Grandes volumes de dados, com grande variedade e em velocidades	Ciclo de Vida do Dado. 1 Ciência da Informação. <i>Big Data</i> .

**Fonte:** Elaborado pela autora

A partir do levantamento bibliográfico realizado na BRAPCI, é possível identificar que os autores utilizam dois atributos principais para definir o *Big Data*: o volume e a velocidade dos dados. Por meio da análise das palavras-chave utilizadas para registro na base, observou-se

que todos os autores utilizaram os termos centrais “*Big Data*”, “Privacidade”. Além destes, os que se demonstraram em evidência são termos relacionados aos dados pessoais, tal como “Proteção de Dado”, “Proteção de Dado Pessoal” e “Dado pessoal”.

## 4.2 Um exemplo da coleta de metadados pela *Amazon*

De origem norte-americana, a *Amazon* é considerada um dos maiores e mais valiosos *e-commerces* da atualidade. Fundada em 1994 pelo empresário Jeff Bezos, a empresa iniciou suas atividades como uma livraria digital, contudo, ao longo dos anos se expandiu e atualmente comercializa uma variedade de produtos, como eletrônicos, alimentos, vestuários, brinquedos, acessórios e até mesmo um serviço de *streaming* (CANALTECH, 2022).

No que diz respeito ao uso das tecnologias do *Big Data* como estratégia para tomada de decisão e desenvolvimento, a *Amazon* também pode ser considerada como uma das maiores empresas que investem nesse quesito. O seu *e-commerce* é líder na coleta, armazenamento e processamento de dados e metadados dos seus clientes.

Segundo Bernard Marr (2015), entre as maiores inovações da *Amazon*, pode-se destacar o seu sistema de recomendação personalizado, construído através da coleta de milhões de transações dos clientes. Esse sistema está entre suas maiores inovações, pois ao colocar algo que alguém possa gostar em sua frente, pode fazer com que esse usuário seja dominado pelo desejo de comprá-lo. Portanto, ao invés de uma abordagem dispersa, a *Amazon* utilizou das informações dos seus clientes para desenvolver um sistema de vendas certeiro.

O sucesso da *Amazon* provém, além de tudo, do seu caráter inovador ao utilizar, desde a sua fundação, a tecnologia como base para o seu desenvolvimento e a satisfação do seu cliente como uma aliada neste processo.

O *e-commerce* da *Amazon* é referência na experiência do cliente e isso inclui todo o processo de venda. A arquitetura do seu site e aplicativo, a divisão dos produtos, a compra facilitada e a sugestão de produtos com base no histórico dos usuários, são alguns aspectos que interferem na decisão de compra dos clientes (AMAZON, 2022).

Dessa forma, a metodologia utilizada pela empresa, tem como foco promover ao cliente uma experiência personalizada, ao longo de todo o seu processo de compra, sendo este um fator crucial para a venda dos seus produtos e a qualidade dos seus serviços.

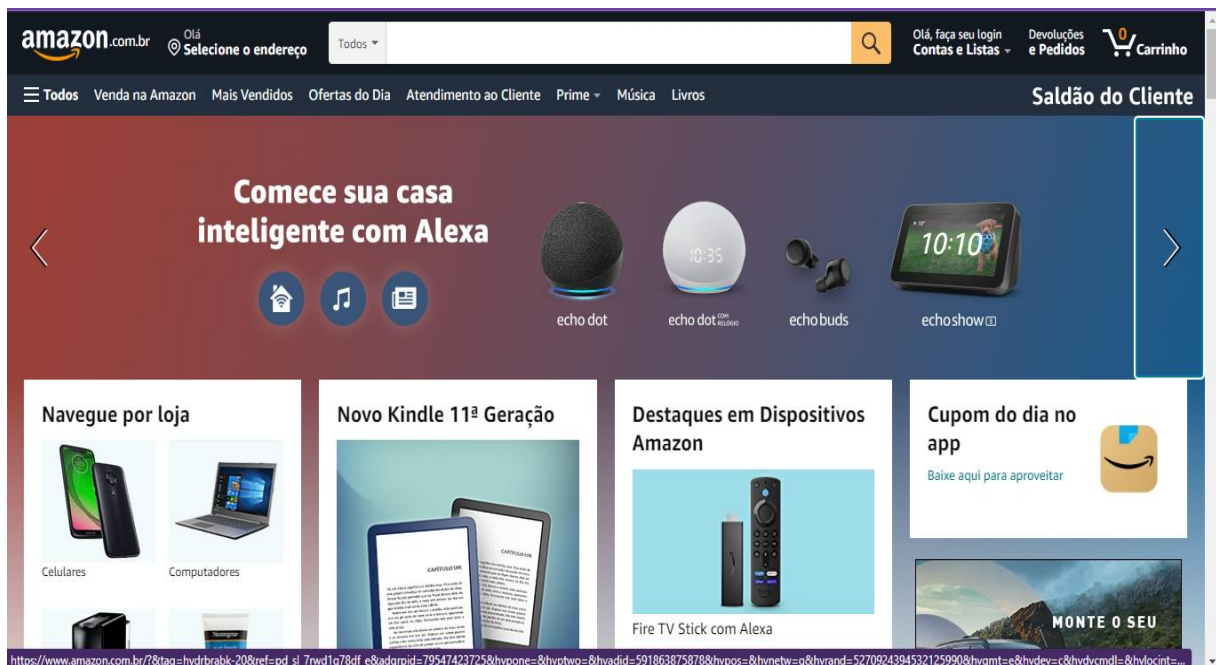
Com o objetivo de compreender como é realizada a coleta de informações na *Amazon* e correlacionar com os conceitos abordados neste estudo, realizou-se uma análise na estrutura da



página inicial do seu site<sup>6</sup>. A análise também contempla as informações disponibilizadas em suas Políticas de Privacidade e *Cookies*.

Na Figura 3, visualiza-se o site:

**Figura 3** - Página inicial da *Amazon*

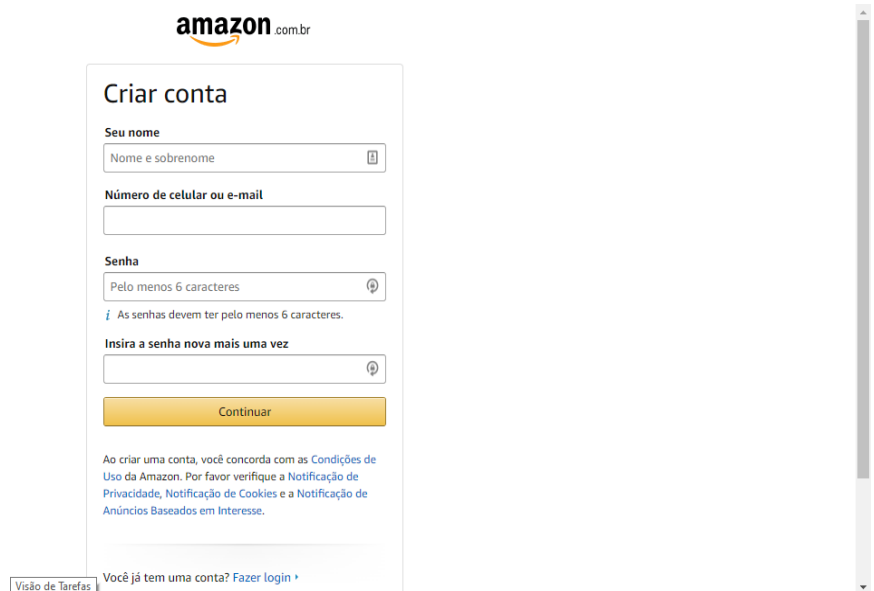


**Fonte:** Elaborado pela autora

Na página inicial do site da *Amazon* (figura 3), identifica-se de imediato a classificação dos seus produtos. Alguns dos seus produtos eletrônicos mais famosos, tal como a *Alexa* (dispositivo de assistente virtual) e o *Kindle* (dispositivo eletrônico *e-reader*) estão em destaque na parte superior do site. Além dos produtos, é possível visualizar o "menu de navegação" que contém o botão de acesso às ofertas, ao atendimento ao cliente, aos produtos mais vendidos, ao carrinho de compras e demais páginas de interesse dos usuários.

Para realizar uma compra, inserir produtos no "carrinho", "lista de compras" ou fazer uma avaliação dos produtos, é necessário criar uma conta através do Formulário de Cadastro, o qual é possível visualizar na Figura 4:

<sup>6</sup> Disponível em: <https://www.Amazon.com.br/>

**Figura 4-** Formulário de cadastro

The image shows the Amazon.com.br registration form. At the top is the Amazon logo. The form is titled "Criar conta" and contains the following fields and elements:

- Seu nome:** A text input field with the placeholder "Nome e sobrenome" and a small icon on the right.
- Número de celular ou e-mail:** A text input field.
- Senha:** A text input field with the placeholder "Pelo menos 6 caracteres" and a small icon on the right.
- Insira a senha nova mais uma vez:** A text input field with a small icon on the right.
- Continuar:** A yellow button.
- Legal notices:** A small text block below the button: "Ao criar uma conta, você concorda com as [Condições de Uso da Amazon](#). Por favor verifique a [Notificação de Privacidade](#), [Notificação de Cookies](#) e a [Notificação de Anúncios Baseados em Interesse](#)."
- Footer:** A small text block at the bottom: "Você já tem uma conta? [Fazer login](#)".

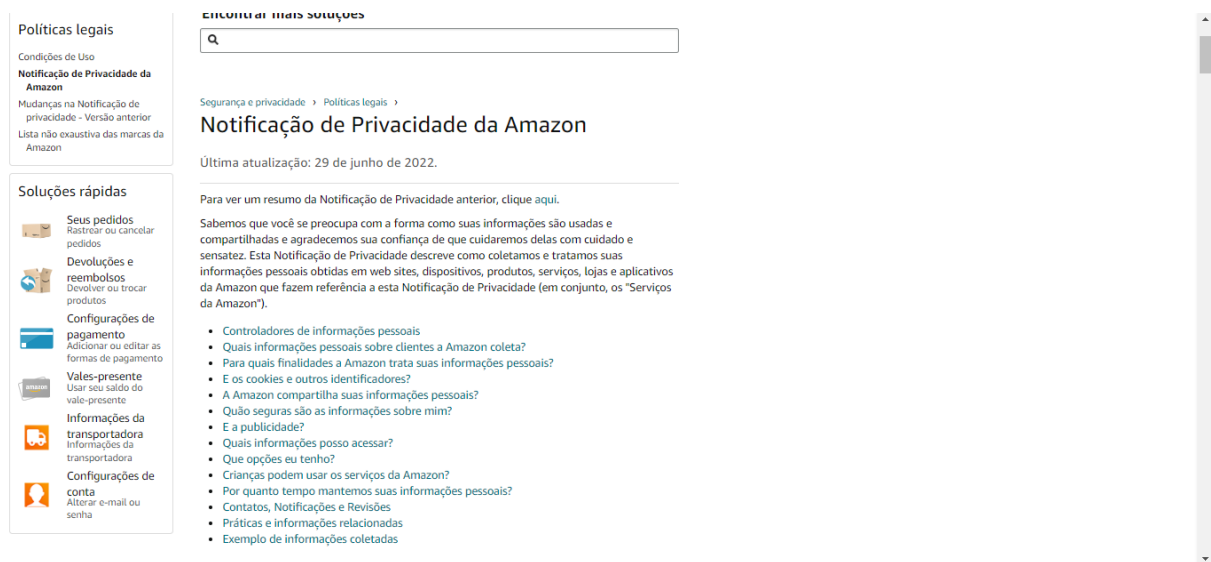
**Fonte:** Elaborado pela autora.

No Formulário de Cadastro da *Amazon* (figura 4) contém os campos, metadados, para que o usuário preencha com os seus dados e efetue o seu cadastro. Os dados solicitados se enquadram como "Informações Pessoais Identificáveis", ou seja, informações que permitem identificar o usuário, como o nome, número de celular ou o endereço de e-mail.

Assim como qualquer *e-commerce* ou rede social da *Internet*, para que o usuário possa usufruir, antes é preciso concordar com o que é denominado "Condições de Uso". As condições de uso indicam as regras e diretrizes para que os usuários utilizem a plataforma. Através do formulário, também é possível ter acesso a "Notificação de Privacidade", a qual tem por objetivo descrever como são coletadas e tratadas as informações pessoais que são obtidas pela *Amazon*.

A Figura 5 apresenta a Notificação de Privacidade da *Amazon*:

**Figura 5 - Notificação de privacidade**



**Fonte:** Elaborado pela autora

Por meio da Notificação de Privacidade, a *Amazon* destacou três tipos de informações pessoais que são coletadas e armazenadas através do uso dos seus serviços, são elas: informações fornecidas pelos próprios usuários; informações automáticas (geradas por tecnologias, como os *Cookies*) e informações provenientes de outras fontes.

As “informações fornecidas” provém da procura e compra de produtos; adição de itens no carrinho de compras e listas de desejos; por meio do download e visualização de vídeos; interações com o dispositivo de assistente virtual (*Alexa*); *upload* de contatos, imagens, vídeos e outros arquivos; avaliações dos clientes; comunicação com a empresa através do e-mail, telefone ou outra forma de contato, entre outras fontes. Como resultado dessas ações, os usuários fornecem informações pessoais identificáveis como o nome, número de telefone e endereço. Além disso, há o fornecimento de informações de pagamento, idade, localização, descrição pessoal, gravação de voz, imagens, vídeos, e uma variedade de outras informações. As “informações automáticas” são coletadas e analisadas de forma automática, tem-se como exemplo, o endereço do Protocolo de *Inter-net* (*IP*); *login*, e-mail e senha; as interações com os conteúdos (downloads, reproduções e cliques.); métricas do dispositivo e serviços da *Amazon*; histórico de uso do conteúdo e de compras; fluxo de cliques e interações com as páginas, etc. As informações também são coletadas através de identificadores de dispositivos, *cookies* e outras tecnologias. As “informações de outras fontes” incluem: atualizações de entrega e endereço das transportadoras; informações de conta, compra, resgate e visualizações de páginas de comerciantes parceiros; interações com produtos e serviços de afiliadas; resultados de buscas e links; informações de dispositivos conectados à *Internet* e *Alexa*, entre outras.

As informações pessoais coletadas e tratadas pela *Amazon* possuem como objetivo efetivar a compra e a entrega dos seus produtos e serviços; fornecer, corrigir e melhorar os serviços; promover a recomendação e personalização dos seus produtos; processar dados de voz, imagem e vídeos para prestar determinados serviços (como àqueles que envolvem sua assistente virtual; cumprir obrigações legais; se comunicar com os clientes; otimizar a publicidade; prever fraudes e riscos de créditos e também, finalidades para quais antes, são pedidos o consentimento do usuário (*AMAZON*, 2022)). Para a proteção dessas informações, a Política de Privacidade da *Amazon* (2022), descreve que algumas ações são colocadas em prática, são elas: o uso de protocolos e *softwares* de criptografia; o Padrão de Segurança de Dados da Indústria de Pagamento com Cartão (PCI DSS) - padrão utilizado para a segurança dos pagamentos realizados com cartões de crédito; Processos de segurança no que diz respeito ao fornecimento de informações pessoais dos clientes, tal como solicitar a comprovação de identidade antes de divulgar informações pessoais ao cliente e dispositivos com recursos de segurança e proteção contra o acesso não autorizado.

A partir da análise da Notificação de Privacidade da *Amazon*, é possível ter uma percepção de como uma simples ação (como inserir um item no carrinho de compras, por exemplo) necessita de outros processos, os quais, em sua totalidade, são capazes de coletar, produzir, armazenar e disseminar uma quantidade gigantesca de dados e metadados.

Os “*Cookies*” fazem parte das informações que são coletadas automaticamente pela *Amazon* e que possuem o objetivo de reconhecer o navegador ou dispositivo em que o usuário está acessando as suas plataformas. Além dessas funções, os *cookies* possibilitam a recomendação de produtos e conteúdo personalizados; acompanhar preferências dos usuários; acompanhar itens armazenados no carrinho de compras; aumentar a segurança, oferecer anúncios e conteúdos relevantes em sites da *Amazon* e de terceiros, entre outras funções (*AMAZON*, 2020).

A Figura 6 apresenta a Notificação de *Cookies* disponibilizada no site da *Amazon*:

**Figura 6** – Notificação de *Cookies*

Ajuda e Serviço de atendimento ao cliente

Todos os tópicos da Ajuda

**Segurança e privacidade**

Como a Amazon usa suas informações pessoais

Como a Amazon coleta suas informações pessoais

Como a Amazon protege suas informações pessoais

Gerenciar suas informações pessoais

Solicitar suas informações pessoais

Amazon e minhas informações pessoais

Anúncios baseados em interesses

**Cookies**

Comunicações da Amazon.com.br

Políticas legais

Sua segurança

**Soluções rápidas**

 Seus pedidos  
Rastrear ou cancelar pedidos

 Devoluções e reembolsos  
Devolver ou trocar produtos

[Configurar a sua conta](#)

**Encontrar mais soluções**

Q

Segurança e privacidade >

**Cookies**

Última atualização em 28 de julho de 2020.

Usamos cookies, pixels e outras tecnologias (coletivamente, "cookies") para reconhecer o seu navegador ou dispositivo, saber mais sobre seus interesses e oferecer a você recursos e serviços essenciais, além de finalidades adicionais, incluindo:

- Reconhecer você quando você acessa nossos serviços. Isso nos permite oferecer a você recomendações de produtos, exibir conteúdo personalizado, reconhecê-lo como membro Prime, permitir que você utilize o recurso de compra com 1-Click e fornecer outros recursos e serviços personalizados.
- Acompanhar suas preferências específicas. Isso nos permite respeitar seus gostos como se você gostaria ou não de ver anúncios baseados em interesse. Você pode configurar suas preferências em [Sua conta](#).
- Acompanhar itens armazenados em seu carrinho de compras.
- Fazer pesquisas e diagnósticos para melhorar os conteúdos, produtos e serviços da Amazon.
- Impedir atividades fraudulentas.
- Aumentar a segurança.
- Entregar conteúdo, inclusive anúncios, relevante para seus interesses em sites da Amazon e de terceiros (consulte a [notificação de Anúncios Baseados em Interesse](#) para saber como usamos cookies para exibir anúncios baseados em interesse).

**Fonte:** Elaborado pela autora

A Notificação de *Cookies* evidencia a possibilidade dos *cookies* serem recusados e/ou bloqueados pelos usuários através das configurações dos próprios dispositivos e navegadores. Contudo, determinados recursos e funcionalidades das plataformas podem não funcionar, o que causará um impacto na experiência dos usuários.

## 5. CONSIDERAÇÕES FINAIS

A expansão da *Internet* e o desenvolvimento das TIC'S deram luz ao fenômeno conhecido como *Big Data*, tecnologia esta que não possui um conceito exato, mas que pode ser definido como o grande volume de dados e metadados, os quais são coletados, gerados, processados, armazenados e disseminados de maneira veloz.

Implementado como estratégia em múltiplos setores de organizações públicas e privadas, o *Big Data* pode ser utilizado como uma ferramenta eficiente para o desenvolvimento de novos produtos e serviços, para a otimização de operação e para tomada de decisões estratégicas dessas organizações. Contudo, embora o *Big Data* tenha o poder de oferecer um grande potencial para o desenvolvimento das organizações, assim como qualquer tecnologia complexa, ele apresenta desafios para a sua adoção. Dentre os seus desafios, destaca-se: a necessidade de uma excelente infraestrutura para o processamento dos dados, profissionais qualificados para o seu manejo e o desenvolvimento de políticas que promovam a segurança informacional.

A Privacidade na *Web* é uma das pautas que vem sendo discutidas nos últimos tempos, haja vista que, com a implementação da Lei Geral de Proteção de Dados (LGPD), lei regulamentadora que visa a proteção dos dados pessoais no Brasil, as organizações se viram diante da necessidade de reformular os seus processos internos, visando estar em conformidade com a regulamentação. Com o imenso fluxo de informações coletadas e compartilhadas pelas organizações, dados pessoais sensíveis e dados que, ao serem cruzados permitem a identificação dos indivíduos, o risco de pessoas (e até mesmo organizações) mal-intencionados tenham acesso e cause danos à privacidade dos indivíduos, são altos.

Diante dessa nova realidade, surge a necessidade de levantar discussões acerca do *Big Data* e sua relação com questões que implicam a Privacidade Informacional na *Web*. A conscientização da sociedade sobre o tratamento dos seus dados é um meio de alertá-los e incentivá-los à compreensão sobre os seus direitos. Do mesmo modo, através do levantamento dessas discussões, cabe uma reflexão no que diz respeito ao papel dos profissionais da Ciência da Informação perante o *Big Data*.

Com o objetivo de compreender o *Big Data*, suas tecnologias, a coleta de informações e a Privacidade na *Web*, este estudo partiu do princípio de realizar uma investigação científica. Dessa forma, como embasamento para o desenvolvimento da fundamentação teórica, utilizou-se da literatura disponível em diversas fontes de informação, como livros, sites e projetos de

pesquisa recuperados em bases de dados, tal como BRAPCI, Google Acadêmico e SciELO. Além do levantamento teórico, realizou-se uma análise na Política de Privacidade e Política de *Cookies* da *Amazon*, com o propósito de compreender como são realizadas as coletas e tratamentos de dados pela empresa e a partir disso, correlacionar com a Privacidade da Informação no contexto da *Web*.

Mediante a todo o material bibliográfico levantado para o desenvolvimento desta investigação, foi possível concluir que o *Big Data* representa a criação de um novo paradigma, haja vista que, com suas técnicas de coleta, armazenamento e tratamento dos dados, amplia-se as fronteiras do conhecimento, no qual, a partir de dados heterogêneos e suas análises, agrega-se valor para organizações de diversos setores. Entretanto, inúmeros desafios envolvem a sua implementação: lidar com tamanha variedade de dados, proveniente de diversas fontes, requer uma infraestrutura adequada, que contempla o alto investimento em *softwares* e *hardwares* - capazes de suprir as demandas, e mão de obra qualificada - para o seu devido manejo. Outro desafio do *Big Data* se dá em relação às questões de Privacidade e suas possíveis formas de ofensas, em meio a quantidade exorbitante de dados coletados, armazenados e disseminados, mesmo aqueles que não configuram como dados pessoais sensíveis (tais como os hábitos de consumo, fluxos de cliques, visualizações e compartilhamentos), podem, de alguma forma acabar por infringir os limites da Privacidade em suas três dimensões: decisional, espacial e principalmente, informacional. Neste sentido, cabe às organizações estabelecerem Políticas de Segurança de Informação que visem a proteção e o cuidado em suas práticas de coleta, armazenamento e tratamento dos dados, garantindo além da segurança, mas em seu caráter ético e social, a transparência com os indivíduos. Com base aos aspectos que envolvem o *Big Data*, sugere-se como agenda de pesquisa temas que se voltem à essas particularidades, tais como: desenvolvimento e implementação de políticas de segurança de dados em organizações; Lei Geral de Proteção dos Dados e o papel do profissional da informação; *Big Data* e o investimento em infraestrutura; ética na coleta, tratamento e armazenamento de dados pessoais; segurança informacional em organizações públicas e privadas; *Big Data* e o impacto de sua matriz energética, o impacto ambiental do *Big Data* e suas tecnologias, entre outros.

As análises das Políticas de Privacidade e Política de *Cookies* da *Amazon*, possibilitaram uma maior percepção de quais informações - e para quais finalidades - elas são coletadas, armazenadas e tratadas pela empresa. Não se limitando, inclui-se informações pessoais identificáveis e não identificáveis, provenientes do fornecimento do próprio usuário (através dos Formulários); geradas automaticamente (por meio dos *Cookies* e outras tecnologias) e informações de outras fontes (transportadoras e demais serviços aliados à *Amazon*). Com base

nisto, é possível afirmar que qualquer ação que o indivíduo venha a ter na *Web*, ele necessita fornecer ao menos uma de suas informações e, mesmo que tenha o "poder de escolha" para determinadas situações, essa escolha se mostra limitada, pois é certo que ele acabará tendo algum impacto em sua experiência. Esse fato apenas evidencia a importância da implementação de leis que regulamentam as práticas de coleta e tratamento de dados pessoais, como o caso da LGPD no Brasil. Haja vista que, embora os usuários passem a ter maior consciência da segurança dos seus dados na *Web* e se atentem à recolha dos mesmos, é através da legislação que as organizações se desempenharão em garantir, de maneira efetiva, a integridade dessas informações.

No contexto da Biblioteconomia e Ciência da Informação, o *Big Data* se mostra um campo com muitas oportunidades de atuação e que devido ao seu caráter inovador, vem despertando o interesse dos bibliotecários e cientistas da informação. A padronização de recursos físicos e digitais através dos Metadados é um exemplo clássico de como os bibliotecários, visando uma gestão eficiente da informação, utilizam os dados em seu cotidiano. De mesmo modo, no *Big Data*, para extrair o máximo do seu valor, é imprescindível ter uma gestão eficiente de todo o fluxo de sua informação, sendo esta característica, uma porta de entrada para os profissionais da CI. Entretanto, é válido destacar que determinadas áreas de atuação no *Big Data* requerem habilidades específicas, tais como o conhecimento de linguagens de programação, análise de dados, o conhecimento das suas tecnologias (como o *Apache Hadoop* e *NoSQL*), além do conhecimento e experiência em outras áreas, como a Matemática e a Estatística. Portanto, a depender de suas competências e da área de atuação pretendida pelo profissional Bibliotecário, ele terá que se capacitar em determinadas habilidades específicas, mas, levando em consideração o perfil multidisciplinar desses profissionais, trata-se de um caminho viável e possível, com grandes oportunidades. Além da atuação direta com o *Big Data*, os Bibliotecários e Cientistas da Informação também podem atuar em outras frentes relacionadas, como no suporte ao desenvolvimento de Políticas de Segurança e Privacidade da Informação e ao suporte à adequação às normas que vêm sendo estabelecidas, como a LGPD.

Portanto, conclui-se que na “era da informação”, o *Big Data* e o seu uso como estratégia para tomada de decisões por organizações públicas e privadas, é uma realidade que tende a aumentar. Contudo, para algumas organizações ainda há limitações: o seu custo de investimento é extremamente alto e apesar dos seus benefícios, existem muitos desafios a serem enfrentados, como a necessidade de adequação no que tange a Privacidade da Informação. Além de que, o seu desenvolvimento é rápido, novos atributos surgem para defini-lo, novas tecnologias são criadas e aperfeiçoadas, na *Web* tudo é instantâneo e todas essas transformações podem acabar



sendo um empecilho para a sua adoção. Nessa maratona, as empresas líderes continuam na frente, os usuários, participantes essenciais, aos poucos conquistando os seus direitos e os profissionais da Biblioteconomia e Ciência da Informação, em sua essência multidisciplinar, se adaptando e caminhando para novos trajetos.

## REFERÊNCIAS

- ALL ABOUT *COOKIES*. **What are cookies in computers?**. Delray Beach - FL, c2022. Disponível em: [https://allaboutcookies.org/#google\\_vignette](https://allaboutcookies.org/#google_vignette). Acesso em: 20 de jun. de 2021.
- AMAZON. **E-commerce ou marketplace? Conheça suas principais diferenças**. Brasil, 2022. Disponível em: <https://venda.Amazon.com.br/sellerblog/e-commerce-ou-marketplace-conhecas-suas-principais-diferencas>. Acesso em: 14 de set. 2022.
- AMAZON. **Cookies**. Brasil, 2020. Disponível em: <https://www.Amazon.com.br/gp/help/customer/display.html?nodeId=GVASXV5UZ64R4Y25>. Acesso em: 14 de set. 2022.
- AMAZON. **Notificação de Privacidade da Amazon**. Brasil, 2022. Disponível em: [https://www.Amazon.com.br/gp/help/customer/display.html/ref=ap\\_signin\\_notification\\_privacy\\_notice?ie=UTF8&nodeId=201283950#GUID-5EE39294-A02A-4A77-A911-F1C6521D4919\\_SECTION\\_FCD8A40F699B4AAA83724625D526E8D1](https://www.Amazon.com.br/gp/help/customer/display.html/ref=ap_signin_notification_privacy_notice?ie=UTF8&nodeId=201283950#GUID-5EE39294-A02A-4A77-A911-F1C6521D4919_SECTION_FCD8A40F699B4AAA83724625D526E8D1). Acesso em: 14 de set. 2022.
- APACHE HADOOP. **Apache Hadoop**. [S.l.]. The Apache Software Foundation, c2006-2022. Disponível em: <https://hadoop.apache.org/>. Acesso em: 29 jun. 2021.
- BHADANI, Abhay Kumar; JOTHIMANI, Dhanya. *Big Data: Challenges, Opportunities, and Realities*. **Effective Big Data Management and Opportunities for Implementation**, Pennsylvania, 22 set. 2016. Disponível em: [https://www.researchgate.net/publication/316956144\\_Big\\_Data\\_Challenges\\_Opportunities\\_and\\_Realities](https://www.researchgate.net/publication/316956144_Big_Data_Challenges_Opportunities_and_Realities). Acesso em: 13 jun. 2021.
- BHOGAL, Jagdev; CHOKSI, Imran. Handling *Big Data* Using NoSQL. [S.l.]. **2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops**. p. 393-398, 2015. Disponível em: [https://www.researchgate.net/publication/307822916\\_Handling\\_Big\\_Data\\_Using\\_NoSQL](https://www.researchgate.net/publication/307822916_Handling_Big_Data_Using_NoSQL). Acesso em: 01 de jun. de 2021.
- BRASIL. Lei Nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Brasília, 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm). Acesso em: 03 de jul. de 2022.
- CANALTECH. **Tudo sobre Amazon - História e Notícias**. [s.l.], [s.d.]. Disponível em: <https://canaltech.com.br/empresa/Amazon/>. Acesso em: 14 set. 2022.
- CHELLAPPA, Ramnath K.; SIN, Raymond G. Personalization versus privacy: An empirical examination of the online consumer's dilemma. [s.l.]. *Information technology and management*, v. 6, n. 2, p. 181-202, 2005. Disponível em: [https://www.researchgate.net/publication/226310091\\_Personalization\\_versus\\_Privacy\\_An\\_Empirical\\_Examination\\_of\\_the\\_Online\\_Consumer's\\_Dilemma](https://www.researchgate.net/publication/226310091_Personalization_versus_Privacy_An_Empirical_Examination_of_the_Online_Consumer's_Dilemma). Acesso em: 20 de jun. de 2021.
- CHEN, Min; MAO, Shiwen; LIU, Yunhao. *Big Data: A Survey*. [S.l.]. **Springer Science+Business Media New York**; 22. jan. 2014. Disponível em:

<http://www2.egr.uh.edu/~zhan2/ECE6111/class/BigDataSurvey2014.pdf> Acesso em: 02 mai. 2021.

COOKIEPEDIA BY ONE TRUST. **What Are Cookies?**. [S.l.]. c2020. Disponível em: <https://cookiepedia.co.uk/all-about-cookies>. Acesso em 20 de jun. de 2021.

DA SILVA, Ticiano L. Coelho et al. *Big Data Analytics Technologies and Platforms: A Brief Review*. [S.l.] LADaS 2018 - **Latin America Data Science Workshop**. p. 25-32, 2018. Disponível em: <http://ceur-ws.org/Vol-2170/paper4.pdf>. Acesso em: 01 de jun. de 2021.

DEMPSEY, Lorcan et al. A review of metadata: a survey of current resource description format. [S.l.] **Specification for resource description methods**. v.1,1997. Disponível em: <https://archive.ifla.org/documents/libraries/cataloging/metadata/d32p1.pdf>. Acesso em: 14 de jun. de 2021.

LANEY, Doug. *Application Delivery Strategies*. Stamford: **Meta Group**.06 de fev. de 2001. Disponível em: <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...> Acesso em: 10 de mai. de 2021.

GIRI, Kaiser J.; LONE, Towseef A. *Big Data - Overview and Challenges*. [S.l.] **International Journal of Advanced Research in Computer Science and Software Engineering**. v. 4, n. 6, p.525-529, 2014. Disponível em: [https://www.researchgate.net/publication/287294077\\_Big\\_Data\\_-\\_Overview\\_and\\_Challenges](https://www.researchgate.net/publication/287294077_Big_Data_-_Overview_and_Challenges). Acesso em: 15 de jun. 2021.

GUPTA, Navarun.; KHAN, M. Ali-ud-din.; UDDIN, Muhammad Fahim. Seven V 's of *Big Data* understanding *Big Data* to extract value. [S.l.]. **Proceedings of 2014 Zone 1 Conference of the American Society for Engineering Education**, 2014. Disponível em: <https://www.researchgate.net/public>. Acesso em: 10 de mai. de 2021.

GOV.BR. **Perguntas Frequentes – ANPD**. Brasil, 2022. Disponível em: <https://www.gov.br/anpd/pt-br/aceso-a-informacao/perguntas-frequentes-2013-anpd#c2> Acesso em: 03 de jul. de 2022.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. 3a ed. United States of America: The Morgan Kaufmann, 2011. Disponível em: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>. Acesso em: 13 de jun. de 2021.

ISWARAPPA.; ANURADHA, J. A Brief Introduction on *Big Data* 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science*. v. 48, p. 319-324, 2015. [S.l.] **International Conference on Intelligent Computing, Communication & Convergence**. Disponível em: [https://www.researchgate.net/publication/282536587\\_A\\_Brief\\_Introduction\\_on\\_Big\\_Data\\_5\\_Vs\\_Characteristics\\_and\\_Hadoop\\_Technology](https://www.researchgate.net/publication/282536587_A_Brief_Introduction_on_Big_Data_5_Vs_Characteristics_and_Hadoop_Technology). Acesso em: 10 de mai. de 2021.

LEE, In. **Big data**: Dimensions, evolution, impacts, and challenges. [S.l.]. *Business Horizons*. v. 60, n. 3, p. 293-303, 2017. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0007681317300046>. Acesso em: 13 de jun. de 2021.

LIMA, Paulo Ricardo Silva; PRESSER, Nadi Helena. A lei geral de proteção de dados e os desafios para a gestão nas organizações brasileiras na era do *Big Data*. Rio de Janeiro: **Revista P2P e INOVAÇÃO**, v. 8, p. 109-120, 2022. Disponível em: <https://brapci.inf.br/index.php/res/v/194267>. Acesso em: 11 set. 2022.

LOTT, Yuri Monnerat; CIANCONI, Regina de Barros. Vigilância e privacidade, no contexto do *Big Data* e dados pessoais: análise da produção da ciência da informação no Brasil. [s.l.]. **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: 23 ago. 2022.

MACHADO, F. N. R. **Big Data: o Futuro dos Dados e Aplicações**. 1a ed. São Paulo: Érica, 2018. Disponível em: <https://ler.Amazon.com.br/kp/embed?linkCode=kpd&asin=B07DMQSG3Q&tag=mvnminer-20&AmazonDeviceType=A2CLFWBIMVSE9N&from=Bookcard&preview=newtab&reshareId=8W83WMCQB630WE6DY6TM&reshareChannel=system>. Acesso em: 11 de mai. de 2021.

MAGALHÃES, Vitória Régia V. et al. **O uso do Big Data na violação da privacidade dos usuários para estratégias de negócios**. 2014. Artigo Científico (Grupo de Pesquisa em Informática Aplicada) - Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Ceará, 2014. Disponível em: <https://docplayer.com.br/3664523-O-uso-do-big-data-na-violacao-da-privacidade-dos-usuarios-para-estrategias-de-negocios.html>. Acesso em: 27 de jun. de 2021.

MANYIKA, James et. al. **Big data: The next frontier for innovation, competition, and productivity**. [S.l.]. McKinsey Global Institute, 2011. Disponível em: [https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi\\_big\\_data\\_exec\\_summary.pdf](https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_exec_summary.pdf). Acesso em: 14 de jun. de 2021.

MARR, Bernard. **Big Data Case Study Collection**. [s.l.]. Wiley, 2015. Disponível em: [https://www.bernardmarr.com/img/bigdata-case-studybook\\_final.pdf](https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf). Acesso em: 14 de set. de 2022.

MATOS, Tiago Farina. Comércio de dados pessoais, privacidade e *Internet*. [s.l.]. **Revista de Doutrina da 4ª Região**, v. 18, n. 7, 2005. Disponível em: [https://revistadoutrina.trf4.jus.br/index.htm?https://revistadoutrina.trf4.jus.br/artigos/edicao007/tiago\\_matos.htm](https://revistadoutrina.trf4.jus.br/index.htm?https://revistadoutrina.trf4.jus.br/artigos/edicao007/tiago_matos.htm). Acesso em: 20 de jun. de 2021.

PEIXOTO, Erick Lucena Campos; JÚNIOR, Marcos Ehrhardt. Breves notas sobre a resignificação da privacidade. Belo Horizonte: **Revista Brasileira de Direito Civil – RBDCivil**, v. 16, p. 35-56, abr./jun. 2018. Disponível em: <https://rbdcivil.ibdcivil.org.br/rbdc/article/viewFile/230/212>. Acesso em: 22 de jun. de 2021.

PEREIRA, Ana Maria; RIBEIRO JÚNIOR, Divino Ignácio; NEVES, Guilherme Luiz Cintra. Metadados para a descrição de recursos da *Internet*: as novas tecnologias desenvolvidas para o padrão Dublin Core e sua utilização. [S.l.]. **Revista ACB**, v.10, n.1, p. 241-249, 2005.

Disponível em: <https://revista.acbsc.org.br/racb/article/view/414/527>. Acesso em: 17 de jun. de 2021.

PIMENTA, Ricardo. Medeiros. *Big Data* e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado. [s.l.]. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 2, 2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119514>. Acesso em: 04 set. de 2022.

POMERANTZ, Jeffrey. **Metadata**. Cambridge, Massachusetts. The MIT Press Essential Knowledge series, 2015.

REIS, Luiz Claudio Rezende; SÁ, Maria Irene da Fonseca. *Big Data*: Um novo campo de atuação para bibliotecários. [S.l.]. **PRISMA. COM**, n. 41 (2020); 231-250, v. 24, n. 2, p. 250-231, 2018. Disponível em: <https://brapci.inf.br/index.php/res/v/135715>. Acesso em: 16 de jun de 2021.

ROCKEMBACH, Moisés. Estudos de usuários de arquivo e os desafios da lei geral de proteção de dados. Rio de Janeiro: **Acervo - Revista do Arquivo Nacional**, v. 33, p. 102-115, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/145798>. Acesso em: 04 set. de 2022.

SANTANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. Londrina: **Informação & Informação**, v. 21, n. 2, p. 116-142, 2016. Disponível em: <https://brapci.inf.br/index.php/res/v/35252>. Acesso em: 04 set. de 2022.

SERRA, Liliana Giusti. Empréstimo de livros eletrônicos, redes sociais e a proteção de dados dos usuários. Campinas, São Paulo: **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 13, n. 1, p. 42-60, 2015. Disponível em: <https://brapci.inf.br/index.php/res/v/39897>. Acesso em: 04 set. de 2022.

SILVA, Joana Luís da. **Big Data no comércio eletrônico**: utilização de *cookies* e questões de privacidade na recolha de dados pessoais. 2020. Dissertação (Mestrado em Negócio Eletrônico) - Instituto Superior de Contabilidade e Administração do Porto, Porto, 2020. Disponível em: [https://recipp.ipp.pt/bitstream/10400.22/17368/1/Joana\\_Silva\\_MNE\\_2020.pdf](https://recipp.ipp.pt/bitstream/10400.22/17368/1/Joana_Silva_MNE_2020.pdf). Acesso em 29 de jun. de 2021.

SOARES, Sandro Vieira; PICOLLI, Icaro Roberto Azevedo; CASAGRANDE, Jacir Leonir. Pesquisa Bibliográfica, Pesquisa Bibliométrica, Artigo de Revisão e Ensaio Teórico em Administração e Contabilidade. Rio de Janeiro: **Administração: Ensino e Pesquisa**, v. 19, n. 2, p. 308-339, 2018. Disponível em: <https://raep.emnuvens.com.br/raep/article/view/970>. Acesso em: 18 de set. de 2022.

SZINVELSKI, Martín Marks; ARCENO, Taynara Silva; FRANCISCO, Lucas Baratieri. Perspectivas jurídicas da relação entre *Big Data* e proteção de dados. [s.l.]. **Perspectivas em Ciência da Informação**, v. 24, n. 4, p. 132-144, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/126853>. Acesso em: 23 ago. de 2022.

TAURION, Cezar. **Big data**. 1a ed. Rio de Janeiro: Brasport, 2013. Disponível em: <https://doceru.com/doc/nxve0x>. Acesso em: 27 de jun. de 2021.

TRUST, J. Paul Getty. Introduction to Metadata. 2a ed. Los Angeles, LA: Getty Publications, 2008. Disponível em:  
[https://books.google.com.br/books?id=qOdVAgAAQBAJ&printsec=frontcover&redir\\_esc=y&hl=pt-BR#v=onepage&q&f=false](https://books.google.com.br/books?id=qOdVAgAAQBAJ&printsec=frontcover&redir_esc=y&hl=pt-BR#v=onepage&q&f=false). Acesso em 14 de jun. de 2021.