



UNIVERSIDADE DE BRASÍLIA  
Faculdade de Ciência da Informação  
Curso de Graduação em Biblioteconomia

O PAPEL DO PROFISSIONAL DA INFORMAÇÃO EM AMBIENTES DE BIG  
DATA E CIÊNCIA DE DADOS

Pollyana Diniz Dos Santos  
Orientador: Prof. Dr. Marcio Victorino

Brasília  
2021

Pollyana Diniz Dos Santos

O PAPEL DO PROFISSIONAL DA INFORMAÇÃO EM AMBIENTES DE BIG  
DATA E CIÊNCIA DE DADOS

Monografia apresentada como  
parte das exigências para obtenção do  
título de Bacharel em Biblioteconomia  
pela Faculdade de Ciência da Informação

Orientador: Prof. Dr. Marcio Victorino

Brasília

2021

N754i

SANTOS, Pollyana Diniz Dos.

O papel do profissional da informação em ambientes de big data e ciência de dados / Pollyana Diniz. – Brasília, 2021. 104 f.

Orientação: Prof. Dr. Marcio Victorino

Monografia (Bacharelado em Biblioteconomia) – Universidade de Brasília, Faculdade de Ciência da Informação, Curso de Biblioteconomia, 2021.

1. Big Data. 2. Ciência da Informação. 3. Ciência de Dados. 4. Profissional da Informação. 5. Bibliotecário de Dados.

I. Título.

**FOLHA DE APROVAÇÃO**

**Título:** O papel do profissional da informação em ambientes de big data e ciência de dados.

**Autor(a):** Pollyana Diniz Dos Santos

Monografia apresentada remotamente em **04 de maio de 2021** à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Orientador(a) (FCI/UnB): Prof. Dr. Marcio de Carvalho Victorino

Membro Interno (FCI/UnB): Prof. Dr. André Luiz Appel

Membro Externo (CEF): José Marcelo Schiessl

Em 06/05/2021.



Documento assinado eletronicamente por **André Luiz Appel, Usuário Externo**, em 06/05/2021, às 14:41, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Marcio de Carvalho Victorino, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 06/05/2021, às 14:53, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

Documento assinado eletronicamente por **José Marcelo Schiessl, Usuário Externo**, em 06/05/2021, às 16:43, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da



Universidade de Brasília.



Documento assinado eletronicamente por **Pollyana Diniz dos Santos, Usuário Externo**, em 07/05/2021, às 10:19, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site [http://sei.unb.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **6638655** e o código CRC **49478699**.

Referência: Processo nº 23106.050563/2021-07

SEI nº 6638655

## DEDICATÓRIA

*“Dedico este trabalho a Deus. Sem Ele nada seria possível.”*

## **AGRADECIMENTOS**

A Deus antes de tudo, por ter me dado a vida e a capacidade de realizar este trabalho. Aos meus pais, pelo apoio e dedicação em todos os momentos da minha vida. Aos professores que passaram ao longo da minha jornada acadêmica, verdadeiros mestres. A minha mãe e melhor amiga, Udineide Diniz, por estar sempre ao meu lado me apoiando e partilhando alegrias, conquistas e desafios.

Aos meus amigos do curso de Biblioteconomia, pelo incentivo e pela troca de conhecimento.

Agradeço às entidades onde reproduzi atividades de estágio, Biblioteca do Senado Federal – Setor de Jornais – e Centro de Documentação e Informação (CEDI) – no TECAD Tesouro da Câmara dos Deputados – por terem desenvolvido capacidades pessoais e profissionais que hoje apresentam-se relevantes para preparar esse estudo e que levarei por toda minha vida.

Agradeço às adversidades que apareceram na minha vida, pois elas me ensinaram a tolerância, a simpatia, o autocontrole, a perseverança e outras qualidades que, sem essas adversidades, eu jamais conheceria. Sou muito grata, à Deus, por nascer em uma época a qual não sou privada de obter e me enriquecer de conhecimento.

Agradeço a Cristo, que me ensinou e veem me ensinando que a minha leve e momentânea tribulação produz para mim um peso eterno de glória e mui excelente; não me atentando nas coisas que se veem, mas nas que não se veem; por que as que se veem são temporais, e as que não se veem são eternas.

*“Os céus declaram a glória de Deus e o firmamento anuncia a obra das suas mãos”.*  
*Bíblia Sagrada - Salmos 19:1*

## RESUMO

As mudanças de fatores sociais e econômicos instauradas por volta do ano de 1900, incentivaram o desenvolvimento do arquétipo tecnológico transformando a sociedade. Na denominada sociedade da informação – a qual surgiu no século XX – que reúne a grande inserção das emergentes tecnologias e o raciocínio coerente de redes dominantes, os dados são gerados em volumes desmedido e se convertem no fundamental insumo dessa conjuntura econômica. Diante desse cenário, o Big Data comparece para prover a premência de extrair qualidade e utilidade desses dados, os quais são criados a cada instante como resultado de qualquer ato executado manuseando dispositivos digitais. Dessa forma, o Big Data gerou uma nova área de performance para os profissionais de ciência da informação, o bibliotecário. O exposto estudo analisa – mediante uma revisão da literatura vigente, e da pesquisa bibliográfica de objetivo exploratório e qualitativo – as conceptualizações de Ciência da Informação, Big Data, Ciência de Dados suas técnicas e ferramentas de análise e, dessa maneira, procura demarcar a descrição do profissional do campo estudado, apontando suas competências e habilidades na sociedade e suas atribuições no contexto do profissional da informação e bibliotecário de dados. Revelando os inter-relacionamentos entre esses conceitos. Além de apresentar as funções de Cientista da Informação atrelado ao bibliotecário de dados. Ao relacionar as competências dos profissionais de Big Data e dos bibliotecários o estudo exhibe a era Big Data como uma nova oportunidade de atuação para aqueles bibliotecários que desejem se capacitar tecnicamente no campo de análise de dados.

**Palavras-chave:** Big Data. Ciência da Informação. Ciência de Dados. Profissional da Informação. Bibliotecário de Dados.

## ABSTRACT

The socioeconomic changes initiated in the second half of the 20th century encouraged the evolution of the technological paradigm by reformulating society. In the so-called information society, which brings together the great insertion of new technologies and the predominant network logic, data is generated in exorbitant volumes and becomes the main raw material in this economic context. In this context, Big Data arrives to supply the need to extract value from these data that are generated at all times as a result of each action performed using digital devices. In this way, Big Data formed a new field of action for information science professionals, the librarian. The present work analyzes, through a review of the existing literature, and bibliographic research of exploratory and qualitative scope, the concepts of Information Science, Big Data, Data Science, its analysis tools and, from that, seeks to demarcate the profile of area, identifying their competencies and skills in society and their duties in the context of the information professional and data librarian. Revealing the interrelationships between these concepts. In addition to presenting the functions of Information Scientist linked to the data librarian. When relating the skills of Big Data professionals and librarians, the study shows the Big Data era as a new opportunity for those librarians who wish to be technically trained in the field of data analysis.

**Keywords:** Big Data. Information Science. Data Science. Information Professional. Data Librarian.

## LISTA DE QUADROS

Quadro 1 – Definições de Buckland.....	20
Quadro 2 – Definições de Laney.....	33
Quadro 3 – Visão geral das tecnologias de big data.....	36
Quadro 4 – Categoria de gerenciadores de banco de dados NoSQL.....	42
Quadro 5 – Dados estruturados.....	46
Quadro 6 – Habilidades exigidas aos bibliotecários de dados.....	51

## **LISTA DE ABREVIATURAS E SIGLAS**

<b>CRISP-DM</b>	Cross Industry Standard Process for Data Mining
<b>CI</b>	Ciência da Informação
<b>IA</b>	Inteligência da Informação
<b>IBICT</b>	Instituto Brasileiro de Informação em Ciência da Informação
<b>IOT</b>	Internet of Things
<b>LISA</b>	Library and Information Science Abstracts
<b>TIC</b>	Tecnologias da Informação e Comunicação
<b>TI</b>	Tecnologia da Informação
<b>UNB</b>	Universidade de Brasília

## SUMÁRIO

1	INTRODUÇÃO .....	16
2	PROBLEMA DE PESQUISA.....	18
2.1	Justificativa .....	18
3	OBJETIVOS.....	19
3.1	Objetivo geral .....	19
3.2	Objetivos específicos .....	19
4	REVISÃO DE LITERATURA .....	20
4.1	O CONCEITO DE INFORMAÇÃO .....	20
4.1.1	CIÊNCIA DA INFORMAÇÃO .....	22
4.1.2	BIBLIOTECONOMIA .....	26
4.1.3	O PROFISSIONAL DA INFORMAÇÃO.....	30
4.2	BIG DATA .....	32
4.2.1	BIG DATA ANALYTICS.....	39
4.2.2	MAP-REDUCE .....	40
4.2.3	NOSQL .....	43
4.3	O CONCEITO DE DADOS .....	45
4.3.1	BIBLIOTECONOMIA DE DADOS .....	48
4.3.2	BIBLIOTECÁRIO DE DADOS.....	50
4.3.3	CIÊNCIA DE DADOS .....	54
4.3.4	CIENTISTA DE DADOS.....	57
4.4	MINERAÇÃO DE DADOS .....	60
4.4.1	CRISP-DM .....	64
5	METODOLOGIA .....	67
6	APRESENTAÇÃO E ANÁLISE.....	68

7	CONSIDERAÇÕES FINAIS .....	75
	REFERÊNCIAS .....	78



## 1 INTRODUÇÃO

O avanço da tecnologia e a atual estrutura da sociedade, observados nos últimos anos, aliado com a popularização da Internet e com o aumento na quantidade e complexidade dos serviços oferecidos por ela, contribuíram de forma significativa à geração massiva de dados. Dessa forma, a quantidade de informação que foi gerada com o avanço da tecnologia aumentou de uma forma rápida e desordenada o uso e a produção de dados na última década.

Por isso, a manipulação e a análise de forma inteligente desse grande volume de dados têm se tornado um dos grandes desafios da atualidade além de ter uma maneira eficiente de filtrar tais informações. Castells (1999 apud WERTHEIN 2000), descreve as características essenciais dessa comunidade, “a informação como insumo, alta penetrabilidade dos impactos das novas tecnologias, predominância da lógica de redes, flexibilidade e próspero consonância de tecnologias”. Tais atributos são espontaneamente percebidas no presente contexto social com o surgimento de tecnologias, como os smartphones e o avanço do renome de redes sociais –*Facebook, Instagram* e aplicativos de mensagem instantânea como o *Whatsapp* (NAISMITH; et al, 2004).

Khupat e Jain (2014) observam que o grande volume de dados presente em nossos dias demandam novas formas de registrar, organizar, indexar, analisar e disponibilizar esses dados. Esse contexto apresenta novos desafios para a CI e particularmente na contribuição de quatro ciências que foram fundamentais desde os primórdios desta ciência: a biblioteconomia, a inteligência artificial, a ciência da computação e a tecnologia da informação, conforme pode-se verificar em Souza, Almeida e Baracho (2013) e Saracevic (1996).

É nesse cenário de grande volume e variedade de dados que se ergue o Big Data, para Manyika (2011, p. 1) “Big data se descreve ao complexo de dados cujo tamanho está além da capacidade típica dos softwares de bancos de dados de capturar, armazenar, gerenciar e analisar” (REIS; SÁ, 2020).

Segundo Pjipers (2010), é exatamente a exorbitância de dados que se converte na maior dificuldade para a recuperação da informação fundamental, visto que é imprescindível perder demasiado tempo procurando essa informação, o que diminui o período a disposição para a análise e proveito da informação enquanto ela é valiosa. Isso é provocado pela desatualização tecnológica entre os recursos de produção e distribuição da informação e os métodos de organização e recuperação dela (PJIPERS, 2010).

Diante do fenômeno Big Data, é possível admitir que a CI esteja diante da mesma necessidade que determinou o seu surgimento: expandir a capacidade de guardar e recuperar informações que extrapolam as capacidades humanas de processamento (SARACEVIC, 1996).

Isto posto, ciência moderna e aberta à interdisciplinaridade, a CI tem sua origem na resposta à necessidade de gestão, tratamento, organização e disponibilização da informação e do conhecimento. Diante disso, Souto (2006) declara, que, “existe um novo perfil de competências e habilidades para o profissional da informação no contexto social atual. Devendo este, ser um sujeito com constante capacidade de aprendizagem e de adaptação às mudanças”.

Em síntese, a recuperação de informação não é mais uma aflição somente de especialistas e pesquisadores, a recuperação da informação se transformou em uma necessidade corrente. Diversas pessoas de níveis intelectuais distintos realizam pesquisas por informação através da web a todo instante, à vista disso, o universo está carregado de informações e muito mais do que isso, ele está rodeado de necessidades informacionais (SILVA; MOMM; BENKENDORF, 2018).

Se antigamente pesquisadores ocuparam-se com a questão da explosão informacional, levantando respostas para organizar e auxiliar a recuperação das informações relevantes – sobretudo para a evolução científica – atualmente se faz indispensável que os profissionais, dos mais variados campos, encontrem novos meios para o controle e tratamento dessa profusão de dados produzidos em ambientes digitais, com a intenção de extrair informação pertinente para variados objetivos, surgindo, assim, a Ciência de Dados.

Assim sendo, a Ciência da Informação tem potencial de executar uma atividade essencial nesse contexto, pois é necessário realizar a seleção e a avaliação da qualidade dos dados para gerar uma Ciência de Dados evoluída. Ou seja, as concepções da ciência da informação caracterizam-se como notáveis aspirantes para alicerçar o âmbito da competência e qualidade na Ciência de Dados (WANG, 2018).

Em correspondência ao advento da explosão informacional – o qual evidenciamos no âmbito do big data – manifesta-se, diante desse cenário, o Bibliotecário de Dados o qual se emprega no ambiente de dados, junto com os princípios, teorias, conhecimentos e habilidades da Ciência da Informação que são utilizáveis e benéficos na gestão e no auxílio a Ciência de Dados. Dessa forma, o que os bibliotecários de dados executam abarca todos os ciclos de vida da gestão de dados (WANG, 2018).

Em suma, compreende o bibliotecário o qual, habitualmente, exerce suas tarefas alicerçado na armazenagem, circulação, preservação e na gestão de acervos documentais. Dedicar-se a se importar com as aprendizagens a respeito dos dados no âmbito da

biblioteconomia. Considera-se que este consiga empregar suas habilidades costumeiras para apresentar serviços e produtos de dados de pesquisa, como no cenário referente às tecnologias, tais como o Big Data e Ciência de Dados e repositórios digitais.

## **2 PROBLEMA DE PESQUISA**

De que maneira o profissional da informação pode exercer seu papel em ambientes de big data e ciência de dados?

### **2.1 Justificativa**

De acordo com perspectiva de Ribeiro (2014), “[...] o big data provoca, na atualidade, o interesse, e até mesmo a admiração, para todas os indivíduos que têm alguma ligação com ofícios para a Gestão da Informação”. Assim sendo, a atual abundância de pesquisas e artigos que dissertam a respeito do conteúdo são indício de que esse campo já excedeu uma categoria de inclinação a tecnológica para se concretizar como campo de relevância dos negócios. Todavia, esse campo ainda necessita de profissionais qualificados para realizar suas atividades que demandam aprimorados conhecimentos técnicos relacionados ao negócio em que estão introduzidos. Dessarte, a afirmação de tecnologias de coleta, manipulação, análise e visualização de dados vem a ser um serviço notável e vigente para a indústria e para a ciência, logo, para os negócios; fazendo-se uma mutável universal a grande parte das teorias e práticas atuais (KOLTAY, 2017). Portanto, de acordo com Semeler e Pinto (2019), “os dados de pesquisa têm de a ser identificáveis, citáveis, visíveis, recuperáveis, interpretáveis, contextualizáveis, interoperáveis e reutilizáveis onde quesitos de consistência e procedência são considerados”. Em síntese, os dados de pesquisa são artigos digitais e não digitais – como documentos, questionários, avaliações, registros de casos, protocolos de estudo, planilhas, notas de laboratório, notas de campo, diários, filmes, imagens, arquivos digitais de áudio e vídeo, sequências genéticas, coordenadas geográficas, banco de dados, algoritmos, metodologias, protocolos, entre outros tipos de manifestação de pesquisa. (DUDZIAK, 2016; HENDERSON, 2017). Diante desse cenário, engloba o bibliotecário que, habitualmente, exerce seus deveres fundamentado no armazenamento, na circulação, na preservação e na gestão de acervos documentais. Semeando os estudos a respeito dos dados no âmbito da biblioteconomia. Depreende-se que este profissional seja capaz de empregar suas habilidades convencionais para disponibilizar serviços e produtos de dados de pesquisa, conforme no contexto referente às

tecnologias, como o Big Data ou aos repositórios digitais (SALES, 2015). Dessa forma, o bibliotecário de dados precisa apurar os sistemas e as normas que envolvem a pesquisa com dados de pesquisa em repositórios digitais, conforme um profissional que têm de procurar conhecimentos sobre coleta, manipulação, análise e visualização de dados de pesquisa para ofertar serviços e produtos relacionados à gestão e curadoria de dados de pesquisa (SEMELER; PINTO, 2019). Em virtude disso, o bibliotecário de dados necessita adquirir as habilidades essenciais para trabalhar com todo tipo de dados – sejam eles observacionais, computacionais ou experimentais. Além de entender como as pesquisas sobre tecnologia de dados mudam as práticas e as teorias que baseiam a biblioteconomia tradicional (CUNHA, 2010). Enfim, os profissionais bibliotecários precisam concentrar seus esforços para os dados de pesquisa digitais, pois cada vez mais eles aumentam devido aos avanços tecnológicos da informação e da computação ligados a fazer ciência, sendo que o crescente uso da tecnologia é uma das principais fontes geradoras de dados de pesquisa (PRYOR, 2012). O presente estudo se direciona pela temática de qual maneira os bibliotecários são capazes de se qualificar para exercer funções profissionais na esfera de análise de dados e que competências e habilidades são essenciais para este feito. Logo, procura explorar as concepções de Big Data, Ciência de Dados, Mineração de Dados, suas ferramentas, tarefas e atividades e suas aplicações na conjuntura do profissional da informação, o bibliotecário.

### **3 OBJETIVOS**

#### **3.1 Objetivo geral**

Apresentar de que maneira o profissional da informação é capaz atuar no cenário de big data, caracterizado por um grande volume, variedade e velocidade de dados.

#### **3.2 Objetivos específicos**

Apresentar as concepções de:

- O conceito de Informação
- Ciência da Informação
- Biblioteconomia
- Profissional da Informação
- Big Data
- Big Data Analytics
- MapReduce

- NOSQL
- O conceito de Dados
- Biblioteconomia de Dados
- Bibliotecário de Dados
- Ciência de Dados
- Cientista de Dados
- Mineração de Dados
- CRISP-DM

Revelando os inter-relacionamentos entre esses conceitos. Além de apresentar as funções de Profissional da Informação atrelado ao Bibliotecário de Dados.

## 4 REVISÃO DE LITERATURA

### 4.1 O CONCEITO DE INFORMAÇÃO

A noção do conceito de informação é essencial para se entender a ciência da informação enquanto disciplina científica. Assim sendo, o âmbito epistemológico da informação origina-se à luz de quem a estuda, além de sua aplicação conceitual que depende, por vezes, do contexto específico (EDIVANIO; EDUARDO, 2011). Ademais, Silva e Ribeiro (2012) revelam um conceito de informação para ser usado junto a disciplinas de Documentação, como a Biblioteconomia. Para esses autores, “a informação é um conjunto estruturado de representações mentais e emocionais codificadas — sinais e símbolos — que são desenvolvidas pela interação social”. Logo, estende-se que a informação é algo capaz de ser registrada em toda forma de suporte, seja ele físico ou digital. Portanto, transmitida de uma forma assíncrona e multidirecional de maneira interativa (SILVA; RIBEIRO, 2012).

Seguindo a concepção de Rafael Capurro (2007), “recupera-se a ideia do termo desde a etimologia da palavra *infomatio* (dar forma) até os usos nas atuais abordagens que utilizam a tecnologia como base”. De acordo com o autor:

O que torna a informação importante nos dias de hoje é a sua natureza digital, e é de extrema relevância a interpretação de suas particularidades, dessa forma, a CI reflete uma interpretação do conceito de informação tornando necessária a realização de uma abordagem investigativa de dupla via, objetiva e subjetiva, sendo o conceito

de informação envolvido em interpretações diversas, configurando-se como um elo entre reflexões pragmáticas e filosóficas sobre uma visão da informação como fenômeno relacionada a ambientes tecnológicos (CAPURRO, 2007, p. 259).

Luciano Floridi (2010), empenha-se em retratar a classificação da informação como algo ligado ao impacto social da cultura tecnológica. A obra *Luciano Floridi's Philosophy of Technology*, faz Silva e Ribeiro (2012), citarem que “a revolução tecnológica das últimas décadas, aumentou a participação da sociedade no que tange aos fenômenos informacionais ligados aos meios digitais”. Esse contexto ocasiona profundas alterações na área da CI e de disciplinas relacionadas, como a Biblioteconomia (SILVA; RIBEIRO, 2012).

Existem diversas vertentes, conceitos e definições do termo informação (MACHLUP; SCHLADER; WELLISCH; NATO; WERSING; NEVELING, 1975).

Porém, de acordo com Buckland (1991), as principais utilizações do termo informação em CI podem ser identificados, classificados e caracterizados em:

<b>Definições de Buckland</b>		
<b>Informação-como-processo</b>	<b>Informação-como-conhecimento</b>	<b>Informação-como-coisa</b>
Refere-se ao ato de informar, quando alguém é informado, aquilo que o indivíduo sabe é modificado.	É aquilo que é percebido na informação-como-processo, ou seja, é a informação que é assimilada, compreendida. É intangível porque é algo subjetivo (convicções, opiniões) que pertence ao indivíduo, não pode ser tocado ou medido.	É a informação registrada. Por isso é tangível, porque é algo expresso, descrito ou representado de alguma forma física.

Quadro 1: Definições de Buckland  
(BUCKLAND adaptado, 1991,)

Isto posto, o estudo desse atual contexto de enorme aumento de dados na era big data – com início na ciência da informação – emerge de maneira que, a disciplina composta de potencialidades e capacidades para o profissional, o qual tem como agente de estudo a informação. Além de assumir-se como uma fonte para o progresso da pesquisa, a relação intrínseca existente entre dado, informação e conhecimento, estando um persuadindo a edificação dos demais (ANGELONI, 2003).

Com início nos anos de 1960, as temáticas bem como os conteúdos de Ciência da Informação foram introduzidas nos cursos de biblioteconomia (ARAÚJO, 2013, p. 6). Diante desse perspectiva, compreende-se a capacidade de relacionar o reconhecimento da ciência da informação, com a biblioteconomia, como uma das orientações e raízes da CI – ligada ao comportamento informacional e a memória intelectual da civilização, além do exercício do papel social e a preocupação comum com os impasses da eficiência na utilização dos registros humanos (CAPURRO, 2003, cap. II; PINHEIRO, 1997, p. 245; SARACEVIC, 2009, p. 15)

#### **4.1.1 CIÊNCIA DA INFORMAÇÃO**

A Ciência da Informação é uma área do conhecimento que sendo uma disciplina jovem, foi datada pela primeira vez em 1958 (Institut or Information Science – UK). A ampliação da quantidade de pesquisadores e, por conseguinte, da produção científica, simultaneamente com o desenvolvimento tecnológico, apareceram originando o advento da Ciência da Informação (MEADOWS, 1999). Segundo Miranda (2002), a Ciência da Informação (CI) trouxe sua origem no contexto da revolução científica/técnica que se observou na Segunda Guerra Mundial desencadeando o fenômeno da explosão informacional, além do empenho consecutivo do controle bibliográfico e do tratamento da documentação implícita no processo.

Porém, embora alguns cientistas alegarem que a expressão Ciência da Informação apenas foi preferida pela primeira vez no ano de 1960 – com início nos estudos e reflexões realizados em volta da produção, tratamento e uso da informação no âmbito humano – é com a solução dos confrontos entre as nações, ocasionando enorme evolução científica e tecnológica, que sucede enorme aumento da informação no contexto social, a denominada explosão informacional, o que, no que lhe concerne, serve como início para o surgimento da CI (HEILPRIN, 1989 apud PINHEIRO; LOUREIRO, 1995).

Segundo Pinheiro (1997, p. 18), “a CI situa-se em fases diferenciadas: a bibliografia/documentação e a recuperação da informação. Porém, o estímulo ao seu desenvolvimento se dá a partir dos estudos e reflexões voltados à busca de soluções para problemas informacionais acentuados a partir do término da II Guerra Mundial”.

Considerando-se o contexto da explosão informacional que existia desde a década de 1940 – na controversa a respeito da recuperação da informação – manifesta-se a apreensão em incentivar a discussão referente as superiores e mais apropriadas soluções, testificando a recuperação de informações propícias. Convém frisar que a expressão recuperação da informação, de acordo com Saracevic (1996, p. 44), foi marcado a princípio por Mooers (1951),

em que “compreende as características intelectuais da definição de informação e suas particularidades para a pesquisa, além de sistemas, técnicas ou automação utilizadas para a execução da atividade”.

Em seu estudo, Saracevic (1996) argumenta exatamente isso, “a necessidade de se refletir em técnicas de recuperação da informação tal como equipamentos essenciais para a evolução do campo da ciência da informação”. Saracevic (1996), argumenta que, “embora a recuperação da informação não tenha sido a única responsável pela evolução da CI, é admitido declarar como a fundamental geradora deste realidade” (SARACEVIC, 2016).

Assim sendo, o método de estudar a recuperação da informação foi certamente, fundamental para a e da eclosão da indústria informacional. Esse seguimento de eclosão de novas áreas e de aperfeiçoamento e permuta de junções multidisciplinares das velhas áreas, de maneira nenhuma está acabado, como comprova a ocorrência, na década passada, da ciência cognitiva, esse cenário observado – de enorme quantidade de informação, apresentou-se ser simultaneamente, algo benéfico, visto que, moldava-se de input no desenvolvimento gradativamente maior da ciência e da tecnologia, bem como algo negativo, em virtude da complexidade de recuperação da informação produzida (SARACEVIC, 1996).

Não importa qual arranjo venha a ser empregada para definir a ciência da informação, faz-se importante entender que os desafios informacionais são, em seu âmago, internacionais ou globais. Visto que, o progresso da ciência da informação, no mundo todo, conduziu dispares acontecimento e prioridades, entretanto a justificativa e os conceitos básicos são os mesmos universalmente. Em vista disso, cada doutrina de concepções sobre a CI declara uma característica importante e singular, já que sugere definições precisas e fundadas na prática dos cientistas da informação de cada país em específico (SARACEVIC, 1996; PINHEIRO; LOUREIRO, 1998).

Lena Vânia Pinheiro e José Mauro Matheus Loureiro (1998) apresentam “a ideia de que a CI se desenvolve em diferentes territórios: científicos, tecnológicos, educacionais, sociais, artísticos e culturais; portanto, múltiplos contextos e condições experimentais”.

Uma das primeiras definições acerca da CI, é a de Taylor (1966), posteriormente sintetizada e reelaborada por Borko (1968, p.3), em definitivo artigo em torno do que seria Ciência da Informação:

[..] Ciência da Informação é a disciplina que investiga as propriedades e o comportamento informacional, as forças que governam os fluxos de informação e os

significados do processamento da informação, visando à acessibilidade e à usabilidade ótima. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, à coleção, à organização, ao armazenamento, à recuperação, à interpretação, à transmissão, à transformação e à utilização da informação. Isso inclui a pesquisa sobre a representação da informação em ambos os sistemas, tanto naturais quanto artificiais, o uso de códigos para a transmissão eficiente da mensagem, bem como o estudo do processamento e de técnicas aplicadas aos computadores e a seus sistemas de programação, [...] tem ambos os componentes de ciência pura, visto que investiga seu objeto sem considerar sua aplicação, e um componente de ciência aplicada, visto que desenvolve serviços e produtos (BORKO, 1968, p. 3).

Saracevic (2010, p. 2570) também propõe mais um conceito para a Ciência da Informação, indicando a tecnologia de forma relacionada a serviços de informação:

[...] a ciência e a prática de lidar com a efetiva coleta, o armazenamento, a recuperação e o uso de informação. Está preocupada com informação registrada e com os conhecimentos, as tecnologias e os serviços que facilitem a sua gestão e utilização (SARACEVIC, 2010, p. 2570)

Ademais, Tefko Saracevic (1996, p. 42) declara os essenciais atributos da área da CI: a interdisciplinaridade, a enorme junção com a tecnologia bem com os desenvolvimentos da Sociedade da Informação:

[...] Primeiro, a CI é, por natureza, interdisciplinar, embora suas relações com outras disciplinas estejam mudando. A evolução interdisciplinar está longe de ser completada. Segundo, a CI está inexoravelmente ligada à tecnologia da informação. O imperativo tecnológico determina a CI, como ocorre também em outros campos. Em sentido amplo, o imperativo tecnológico está impondo a transformação da sociedade moderna em sociedade da informação [...]. Terceiro, a CI é, juntamente com muitas outras disciplinas, uma participante ativa e deliberada na evolução da sociedade da informação. A CI teve e tem um importante papel a desempenhar por

sua forte dimensão social e humana, que ultrapassa a tecnologia (SARACEVIC, 1996, p. 42).

Investigando a literatura da área, parece não existir dúvida entre os principais autores de que a Ciência da Informação (CI) é um campo recente, e que, segundo Shera e Cleveland (1977 apud BRAGA, 1995, p. 4), surgiu da demanda social pela otimização dos processos de coleta, armazenamento, recuperação e disseminação da informação científica e tecnológica, cuja produção apresentava um crescimento enorme ao final da década de 50, caracterizada como crise da informação. Além do aspecto interdisciplinar da CI, que se mostra recorrente no consenso entre os principais autores.

Japiassu (1976, p. 74), afirma que a função interdisciplinar acarreta necessariamente um trabalho conjunto ordenado, havendo enriquecimento e/ou modificação das disciplinas envolvidas, com a finalidade de estudar um objeto sob diferentes ângulos, a partir de acordo prévio sobre os métodos a seguir ou sobre os conceitos a serem utilizados. Diz ele que, “Interdisciplinaridade, portanto, se caracterizaria pelas trocas de conhecimento e pelo grau de integração entre disciplinas relacionadas, definidas por uma inegável comum, o que introduz a noção de finalidade, apresentando um conjunto de níveis e de objetivos diversos”.

Segundo Severino (1995, p. 11), a interdisciplinaridade busca observar o todo, deixando que o pensamento ocorra com ênfase no diálogo entre as diversas áreas do saber. O autor argumenta que, “A conceituação de interdisciplinaridade é, sem dúvida, uma tarefa inacabada: até hoje não conseguimos definir com precisão o que vem a ser essa “vinculação, essa reciprocidade, essa interação, essa comunidade de sentido ou essa complementaridade entre várias disciplinas”. Tal argumento para essa dificuldade estaria na in experiência prática e vivencial do nosso meio acadêmico.

Isto posto, Wersig e Nevelling (1975 apud FREIRE, 2003), expõe que a Ciência da Informação alcançou em seu universo multidisciplinar uma forma de se evoluir, avançando em direção à sua independência, através de duas linhas supostamente turbulentas, o dever de consolidar sua essência teórica e, simultaneamente, de relacionar-se com demais áreas.

Essa evolução ocorre perante resistente atuação das incontáveis ferramentas tecnológicas atuais de informação e comunicação, e tem de exercer, ao fim, a sua função de ciência social, que emprega os conhecimentos que tece em contextos sociais, culturais, econômicos e políticos (WERSIG; NEVELLING, 1975 apud FREIRE, 2003).

Segundo Pinheiro (1997), a Ciência da Informação seria, desta forma, aquela disciplina científica voltada para o estudo da informação em suas diferentes manifestações e fenômenos, no contexto social, por meio da interface com diferentes campos e domínios do saber, desde as das ciências exatas percorrendo pelas ciências humanas, chegando aos domínios extra científicos, tais como a filosofia e a arte, por exemplo.

Foskett (1980, p. 51) em suas definições, salienta as possíveis relações interdisciplinares, pois segundo ele:

[...] Ciência da Informação é uma disciplina que surge de uma fertilização cruzada de ideias que incluem a velha arte da Biblioteconomia, a nova arte da Computação, as artes dos novos meios de comunicação, e aquelas ciências como a Psicologia e Linguística, que em suas formas modernas têm a ver diretamente com todos os problemas da comunicação, a transferência do pensamento organizado (FOSKETT, 1980, p. 51).

Já segundo Pinheiro (1987), a CI estabelece uma rede de interdisciplinaridade, de grande alcance teórico, abrangendo disciplinas das Ciências Naturais, da Filosofia, Ciências sociais/humanas, e tecnologia – incluindo a inteligência artificial. A interdisciplinaridade da Ciência da Informação foi introduzida por diferentes experiências profissionais, mas nem todas trazem uma contribuição igualmente relevante (SARACEVIC, 1992, p. 12), como por exemplo, a biblioteconomia.

#### **4.1.2 BIBLIOTECONOMIA**

Observou-se, antigamente, que a Biblioteconomia era considerada uma das mais antigas disciplinas que se ocupa do acesso à informação e de sua transmissão, pois está intrinsecamente ligada ao surgimento da biblioteca e ambientes designados para manter as tábuas de argilas, consideradas as categorias mais primitivas de livros (LE COADIC 1996).

De acordo com Ortega (2004), as primeiras coleções organizadas de documentos – primeira biblioteca primitiva – é a Biblioteca de Ebla, na Síria, cuja existência foi comprovada, datada no terceiro milênio a.C. A qual Ortega (2004), destaca que “apresentava um acervo criteriosamente disposto em estantes, segundo o assunto. A organização encontrada nessa

biblioteca, vem sendo considerada a origem dos princípios da Biblioteconomia” (ORTEGA, 2004).

Por conseguinte, a criação da imprensa – em meados do ano 1440, por Gutenberg – o poderio da Igreja em relação aos livros e as bibliotecas foi suspenso. Sucederam, dessa forma, os livros, a serem fabricados em grande escala, o que acarretou uma enorme disponibilidade no mercado. Diante disso, em virtude das pressões sociais, surgiram na Europa as primeiras bibliotecas públicas (ORTEGA, 2004; LEMOS, 2008).

Haja vista, segundo Ortega (2004), sucedeu o crescimento da produção de livros e o aumento das bibliotecas, disparando a emergência de propostas que pretendessem a organização desses ambientes. Isto posto, a biblioteca pública deu-se a retratar a atualidade, em contrapartida às bibliotecas da antiguidade e da idade medieval que a precederam (ORTEGA, 2004).

No compromisso do advento da biblioteca pública e do aumento de periódicos, a Biblioteconomia sucedeu-se a traçar novas trajetórias. A exigência de organizar, conservar e disseminar os documentos – desde os primórdios da escrita até a atual era – conduziu as bibliotecas a conceberem uma cadeia de procedimentos de elevadas características técnicas, planejando a solução de desafios práticos, abarcando o costume dessas técnicas que, no decorrer do tempo, se conceberam na estrutura da disciplina póstera Biblioteconomia (SIQUEIRA, 2010).

Encontra-se diferentes conceitos pertencentes a palavra Biblioteconomia. Para Le Coadic (2004) apresenta a finalidade de organização, gestão e administração de bibliotecas. Ele destaca, “a Biblioteconomia sendo uma das disciplinas que antecedem a Ciência da Informação da qual concentra no suporte da informação e por essa razão, atenta-se com a prática de organização de bibliotecas cercado-se com problemas referentes ao acervo e a seu público-alvo”.

O surgimento das bibliotecas, de acordo com Siqueira (2010), recebeu contribuições de muitos pesquisadores que serviriam mais tarde para formar as bases da Biblioteconomia. Assim, o crescimento da produção bibliográfica – efeito da tecnologia e prensa gráfica – ocasionou a utilidade de inovações, no que tange as ferramentas de organização, preservação e recuperação das coleções, isso desencadeou diversos pesquisadores a se inclinarem para essas questões. Dentre os estudiosos, ressalta-se: aquele que criou o sistema de Classificação Decimal de Dewey (CDD) – Melvil Dewey. Paul Otlet e Henri La Fontaine criaram a Classificação Decimal Universal (CDU), alicerçado na CDD (BARBOSA, 1969).

De acordo com Martins et al. (2016), no que tange à catalogação, se faz presente a elaboração de Anthony Panizzi, criando as 91 regras de catalogação – *Rules for the Compilation of the Catalog: Catalogue of Printed Books in British Museum*. Os autores Paul Otlet e Henri La Fontaine criaram um padrão de tamanho das fichas catalográficas para descrição dos materiais, e a Tabela de Cutter, publicada em 1876, denominada *Rules for a Printed Dictionary Catalog* (SIQUEIRA, 2010).

Enfim, um dos maiores bibliotecários e o qual deixou muitos legados para a Biblioteconomia, Ranganathan. Com a criação de um sistema de classificação facetado e contribuições para a área de gestão de bibliotecas. Ranganathan é dirigente por descrever as cinco leis da Biblioteconomia (RANGANATHAN, 2009; SANTOS, RODRIGUES, 2013).

Ademais, a expansão da elaboração de bibliografia, da pesquisa científica e a manifestação de novos suportes, concedeu primícias à necessidade do desenvolvimento de diferentes técnicas de tratamento das informações, para organizar e administrar os conteúdos bibliográficos. Essa necessidade provocou uma ruptura entre a Biblioteconomia e a Documentação. Isto posto, Santos e Rodrigues (2013), destacam “que a Documentação proporcionou possibilidades e a Biblioteconomia traçou novos sentidos, realizando reflexão a respeito de sua cientificidade enquanto ambiente produtor e sistematizador de conhecimentos teóricos para basear os procedimentos usados pelas bibliotecas”.

Nesse contexto, Saracevic (1996, p. 45) discorre sobre a ciência da informação, suas origens e desenvolvimento:

A CI surge da interseção entre a Recuperação de Informação, a Teoria da Informação, a Organização da Informação e a Cibernética. As origens dos estudos sobre a recuperação de informação em registros documentais impulsionam o desenvolvimento de pesquisa em CI desde os anos de 1950. [...] o trabalho com a recuperação de informação foi responsável pelo desenvolvimento de inúmeras aplicações bem-sucedidas (produtos, sistemas, redes, serviços). Mas, também, foi o responsável por duas outras coisas: primeiro, pelo desenvolvimento da CI como um campo onde se interpenetram os componentes científicos e profissionais. [...] segundo, a recuperação da informação influenciou a emergência, a forma e a evolução da indústria informacional (SARACEVIC, 1996, p. 45).

Segundo Bugembe (2016), “o conceito de dados aponta uma forte tendência de estudos e práticas profissionais que está se desenvolvendo no cerne das investigações realizadas por bibliotecários e cientistas da informação”.

Ademais, as novas inclinações de pesquisa em Ciência da Informação no âmbito de big data e ciência de dados apresentam conhecimentos a respeito da origem, as melhores práxis, o planejamento, a curadoria dentre diversas práticas científicas sobre a teoria e a utilização com dados de pesquisa; primordialmente quando esses dados se relacionam com os processos de pesquisa científica em meio digital (SEMELER, 2017).

De acordo com Bugembe (2016), “o campo da big data e ciência de dados está relacionado com a coleta, a curadoria, o processamento, o armazenamento, a análise e a visualização de dados de pesquisa”. Assim, as bibliotecas, diante deste cenário, têm de estar comprometidas no decorrer de todo o procedimento de pesquisa e requerem estar dinamicamente dedicadas a curadoria, aconselhamento e preservação dos desfechos da pesquisa. Algumas funções complementares sugeridas para as bibliotecas são expandir o conhecimento dos dados dos pesquisadores, adotar uma função de armazenamento e preservação de dados e treinar bibliotecários de dados (RIBEIRO, 2014).

Em suma, a Biblioteconomia vem desenvolvendo-se em seus conhecimentos e buscando ajustar-se à nova realidade que o surgimento tecnológico expõe – a organização da informação no âmbito digital. Tal imposição tem orientado os profissionais desse campo a cogitarem em um protótipo que auxilia não apenas os documentos físicos, porém aos documentos digitais juntamente (VALENTIM, 2004). Destarte, finda-se que a Biblioteconomia encaminha-se aperfeiçoando nas técnicas de organização do conhecimento, assim como no desenvolvimento de seus modelos de representação e recuperação da informação.

Diversos conteúdos nessa vertente, foram apresentados na Federação Internacional das Associações e Instituições Bibliotecárias (IFLA). A International Federation of Library Associations (IFLA) apontou algumas tendências que irão alterar os ambientes de informação, e que estão inevitavelmente relacionadas ao contexto atual, que se mostra diretamente ligado ao Big Data. Dentre essas tendências, a IFLA aponta que a economia global será alterada pelas novas tecnologias, além de haver tanto uma expansão quanto uma redução de quem tem acesso às informações (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2016).

As questões levantadas pela IFLA inserem o profissional da informação como agente central das mudanças que estão ocorrendo atualmente, sendo necessário que esse profissional possa contribuir nessa ruptura. Dessa forma, as análises de dados estão sendo fundamentais para as transformações que estão ocorrendo, desafiando o profissional da informação a ser um

ator ativo nesse processo, que possa tornar as evoluções atuais mais intensas e benéficas (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2016).

### 4.1.3 O PROFISSIONAL DA INFORMAÇÃO

Na bibliografia da área de Ciência da Informação, há diversos estudos sobre a temática profissional da informação e observa-se que apresentar uma definição exata para o termo é algo difícil e complexo, diante da variedade de definições existentes, algumas até contraditórias. Por isso, Mota e Oliveira (2005, p. 99), sustentam, que, “o conceito de profissional da informação está em processo evolutivo e sua abrangência ainda se encontra indeterminada, suscitando vários debates em torno de quem realmente pode ser considerado tal”.

Segundo Neves (1998), a expressão “Profissional da Informação” aparece na literatura, a partir do final da década de 80 e início da década de 90, para atender a uma necessidade das unidades de informação, que trabalham, principalmente, com a realidade das novas tecnologias.

Posto isso, a Federação Internacional de Informação e Documentação (FID) criou, em 1992, o grupo *Special Interest Group/ Modern Information Professional (SIG/MIP)*, com o propósito de compreender as necessidades do mercado de trabalho em evolução e suas consequências na formação do profissional da informação (NEVES, 1998).

As novas tecnologias de informação modificaram a práxis de vários segmentos e instituições sociais e, nessa visão, também causaram impacto na forma de organização, disseminação e uso de bibliotecas e outras unidades de informação. Assim, de acordo com Marlene de Oliveira (2011), tais mudanças repercutiram na formação e no perfil de bibliotecários e outros especialistas da informação. Os profissionais da informação passaram assim a se deparar com um novo contexto que lhe exigia, e exige, não só um corpo de conhecimentos especializados, mas também conhecimento e habilidades no uso de tecnologias para organizar, processar, recuperar e disseminar informações, independentemente do suporte no qual elas estarem registradas (OLIVEIRA, 2011).

Diante disso, Souto (2006) declara, que, “há um moderno retrato de competências e habilidades para o profissional da informação no contexto social atual. Devendo este, ser um profissional que consiga frequentemente aprender e se adaptar às mudanças, levando em relevância o trabalho em equipes multidisciplinares além de ser alfabetizado digitalmente”. Outrossim, ele precisa ser capaz de estar em contínuo processo de aprendizagem e desenvolvimento para dominar e ajustar-se aos padrões atuais organizacionais e de gestão das atividades e portanto, um gestor do processo de inovação (SOUTO, 2006).

Diante disso, Segundo Oliveira (2011) observa que, “verifica-se que em diversos âmbitos profissionais estão emergindo novas exigências destacando a procura por atitudes como a receptividade diante de mudanças, capacidade de ser aliado no objetivo da missão; propósitos e valores das unidades de informação; conseguindo encarar os desafios e os compromissos assumidos de forma inovadora” (OLIVEIRA, 2011).

Logo, os profissionais da informação precisam considerar que apesar de gerenciar de maneira eficiente e eficaz um serviço de informação, também se faz necessário ir além na competência de se posicionar na essência da organização, capaz de ser um sujeito indispensável para a transformação e adaptação desses ambientes frente aos costumes atuais e expectativas da sociedade (OLIVEIRA, 2011).

Para os profissionais da informação, o bibliotecário dentre eles, as mudanças estão afetando de maneira mais complexa, seus habituais padrões de trabalhar, isto porque, o objeto de trabalho destes profissionais é a informação (TARAPANOFF, 1996, p.115). A informação, portanto, como objeto de trabalho e estudo do bibliotecário, tem sido afetada pelas tecnologias de informação, modificando seu formato, seu suporte, seu processamento e disseminação, influenciando na forma de mediação entre o bibliotecário e o usuário. Dessa forma, cabe ao profissional da informação, estar em sintonia com esta realidade e se readequar para enfrentar as mudanças cada vez maiores (TARAPANOFF, 1996).

Nesse sentido, o *Bureau of Labor Statistics – U. S. Department of Labor*, em seu *Occupational Outlook Handbook*, descreve o bibliotecário como um profissional da informação que:

A partir de uma redefinição do conceito de biblioteca, passou a redesenhar as atividades do seu cotidiano profissional à luz da inserção das novas tecnologias. A partir daí, os bibliotecários começaram a desenvolver atividades não só de caráter técnico, mas também administrativo e gerencial, passando a coordenar equipes de funcionários e a desenvolver e gerir programas e sistemas de informação, assegurando que a informação seja organizada de maneira que atenda às necessidades dos usuários (2004 apud MOTA; OLIVEIRA, 2005, p. 100).

Beraquet et al. (2006), também, deixam claro, que, “a profissão do bibliotecário exige mudanças, não em sua estrutura e concepção, porém, nos seus resultados e performance de atuação, o que expressa analisar a ênfase de suas habilidades e saberes”.

Nesse contexto, segundo Wormell (1999) “entende-se que o profissional da informação é o intercessor entre os abastecedores de informação, os usuários e as tecnologias de informação, sendo assim lhe é exigido, no desenvolver de suas tarefas, algumas condutas como flexibilidade, adaptabilidade e habilidade para recuperar, organizar e armazenar informação, tanto de fontes impressas como eletrônicas” (WORMELL, 1999, p.14).

Tais fatos, instiga amplos debates no campo da Ciência da Informação, no que fere, principalmente, a utilização favorável de dados, informação e conhecimento gerados a partir dos rastros digitais produzidos por dispositivos eletrônicos. Em vista disso, no campo da Ciência da Informação, comprova-se a necessidade de um espaço interdisciplinar para discutir questões polêmicas sobre informação, conhecimento e ação autônoma, relacionando-as com o fenômeno tecnológico denominado Big Data (EIICA, 2019).

## 4.2 BIG DATA

As evoluções tecnológicas marcaram o início do século XXI, e nos anos atuais, vivenciamos uma gradual aglomeração de dados, à disposição na gigante rede. Estima-se que do início da civilização até o ano de 2000, a raça humana tenha gerado cinco exabytes (10 bytes elevados a 18ª potência) de dados (SCHMIDT, 2010).

Hodiernamente, gera-se esse equivalente volume de dados a cada dois dias, gradativamente mais presente em diversos setores da sociedade e em quase todas as tarefas diárias realizadas pelos indivíduos, gerando, assim, quantidades de dados extraordinárias – dados gerados em tempo real e em fluxo contínuo oriundos de rastros digitais produzidos por redes sociais, smartphones, cartões de créditos, registros de transações online, dentre outros, têm se aumentado enormemente em servidores da internet (MCAFEE; BRYNJOLFSSON, 2012).

Segundo Meira (2016), ocorre que, diante do grande crescimento do número de pessoas e dispositivos conectados à rede mundial de computadores e do exponencial barateamento dos custos para armazenamento de dados, a quantidade de informação gerada e armazenada não para de crescer a cada dia. Além de pessoas criando conteúdo e refletindo em dados disponíveis na internet, vivemos, como conceitua Ashton (2015), a era da Internet das Coisas, ou *Internet of Things*, o que significa dizer que sensores de todas as formas se encontram espalhados no

ambiente, captando grande quantidade de dados como imagens, temperatura, localização, dentre outros (TECHOPEDIA, 2015).

Segundo Cavalli (2016), o que hoje é chamado de Internet das Coisas (*Internet of Things*) é um conjunto de tecnologias e protocolos associados que permitem que objetos se conectem a uma rede de comunicações e são identificados e controlados através desta conexão de rede. Há grandes divergências envolvendo o conceito de IoT, não havendo, dessa forma, um único conceito que possa ser considerado unânime.

Assim sendo, pode ser entendido como um ambiente de objetos físicos interconectados com a internet por meio de sensores pequenos e embutidos, criando um ecossistema de computação onipresente, voltado para o auxílio do cotidiano das pessoas, introduzindo soluções funcionais nos processos do dia a dia e os facilitando (CAVALLI, 2016).

Em suma, o que todas as definições de IoT tem em comum é que elas se concentram em como computadores, sensores e objetos interagem uns com os outros e processam informações/dados em um contexto de hiper conectividade (BREWSTER, 2014). Há, nesse contexto, um fluxo contínuo de informações e massiva produção de dados. O tempo inteiro, “coisas” se conectam à internet com capacidade para compartilhar, processar, armazenar e analisar um volume enorme de dados entre si (TECHTARGET ANZ STAFF, 2007).

Essa prática é o que une o conceito de IoT ao de *big data*. A primeira propriedade envolvendo *big data* consiste no volume crescente de dados. Outra propriedade envolve a alta velocidade com que os dados são produzidos, analisados e visualizados. Além disso, a variedade de formatos de dados representa um desafio adicional. Essa característica é potencializada pelos diferentes dispositivos responsáveis por coletar e produzir dados em diversos âmbitos (RIJMENAM; MARK VAN, 2015).

O conceito de *big data* pode implicar, junto com o de *Data Science*, a capacidade de transformar dados brutos em gráficos e tabelas que permitam a compreensão do fenômeno a ser demonstrado (CAVALCANTI; JOSÉ CARLOS, 2016).

À vista disso, percebe-se que a difusão em larga escala de dispositivos conectados pela internet, captando, analisando e armazenando grande quantidade de dados, estabelecendo conexões entre si e realizando tarefas de maneira autônoma, deu início a uma verdadeira revolução digital e comportamental, abrindo novas oportunidades em diversos ramos (PEPPET, 2014). Esta explosão de informações, sem qualquer precedente na história, vem gerando interesse de governos e empresas na utilização destes dados disponíveis para os mais variados fins, especialmente diante da tecnologia ubíqua que as coleta, armazena e analisa.

Captar, armazenar e analisar dados não é algo novo. Aliás, historiadores, como, Kramer (1963), defendem que a escrita foi inventada pelos Sumérios por volta do ano de 3.000 A.C. com o objetivo de guardar e controlar informações. A coleta de informações e seu armazenamento se transformaram ao longo da história. Na idade contemporânea, empresas e governos dedicam-se cada vez mais a gerar técnicas e métodos com este intento. Afirma-se, portanto, que vivemos uma verdadeira era, a era do *big data*, onde a sua utilização é cada vez mais ampla (WEISER, 1991).

A definição de big data surge, dessa forma, conforme um retrato de representação dos aspectos observados no contexto de grande profusão de dados. Com base na inclusão de computadores no meio social, os dados admitem-se a ser aglomerados, possibilitando o manifestação de algo atual. O universo percorre a não estar somente cheio de informação, porém a informação principia a ser concentrada com mais velocidade (MAYER-SCHONBERGER; CUKIER, 2013).

Mayer-Schonberger (2013), completa, destacando que o big data se caracteriza como uma nova forma de retratar, captar, analisar, armazenar, extrair valor de grande quantidade de informações, possibilitando, dentre outros, a tomada de decisões automatizadas, aumento na eficiência empresarial e governamental, criando modelos de negócios e gerando substancial riqueza, além de economia de preciosos recursos. O acúmulo de conhecimento e informação, que um dia significou estudar, conhecer e compreender o passado, está se transformando, significando, com o *big data*, a habilidade de prever o futuro (MAYER-SCHONBERGER, 2013).

Pries (2015), frisa que o simples ato, atualmente corriqueiro, de utilizar o telefone celular para realizar compras online, conversar com amigos e parentes através de rede sociais, realizar pesquisas em sites de busca, representam exemplos tangíveis de interação com o universo do big data, culminando na melhora da funcionalidade de aplicativos em dispositivos eletrônicos, tornando mais simples, fácil e eficiente a utilização destes pelos usuários.

Há, desse modo, segundo Loukides (2012), “uma progressiva demanda de análise de dados por meio de uma abordagem holística e interdisciplinar, que considere a integração e a combinação de dados provenientes de diferentes fontes”. A grande dificuldade, portanto, não é o de localizar dados, mas sim esclarecer o que realizar com eles e como usá-los de maneira relevante, e empregando sua enorme capacidade de aplicação (LOUKIDES, 2012).

O Big Data anunciava três características essenciais dos dados, denominadas por Laney (2001) como 3Vs:

<b>3 Vs – Três características essenciais dos dados</b>
---

<b>Volume</b>	<b>Velocidade</b>	<b>Variedade</b>
Grandes volumes de dados são gerados através do uso abundante de recursos computacionais. Com a evolução das mídias sociais e outros recursos e serviços da Internet, as pessoas produzem cada vez mais conteúdo, vídeos, fotos, tweets, entre outros tipos de dados.	Os dados são gerados em grande velocidade, à medida que os recursos computacionais têm sua capacidade de produção, captura e processamento de dados aumentada.	Os dados advêm de variadas fontes (sistemas legados, e-mails, posts em mídias sociais, arquivos de vídeo/áudio, gráficos, dispositivos ou sensores), as quais implementam tecnologias distintas para representação e armazenamento de recursos digitais.

Quadro 2: Definições de Laney  
(LANEY, 2001, adaptado)

Posteriormente surgiram novas obras sobre o tema big data e conseqüentemente acabaram por dar novas teorias tornando-se estabelecidas com base nos 3 Vs iniciais de Laney. Gandomi e Haider (2015, p. 139) mais tarde, adicionam mais três dimensões: *veracidade*, *variabilidade* e *valor*.

De acordo com Gandomi e Haider (2015), “a veracidade está ligada com a instabilidade específica que algumas fontes de dados dispõem, isto é, tratar dados indefinidos talvez seja mais um aspecto do big data no que lhe concerne, a variabilidade está vinculada com a oscilação nas porcentagens de fluxo dos dados”. Este aspecto associa-se com a velocidade de produção de dados, sendo capaz de ocorrer alta e baixa velocidade (variabilidade), por fim, o atributo valor é julgada como uma característica definidora do big data. Os dados adquiridos em seu aspecto original frequentemente obtêm pequeno valor referente ao seu volume, sendo capaz de gerar alto valor com início na análise de grandes volumes desses dados (GANDOMI; HAIDER, 2015).

Apesar de haver outras facetas associadas ao big data, conforme a perspectiva exclusiva de autores diversos, vale ressaltar que cada atributo não é independente em relação aos demais. No instante que um atributo altera, há enorme possibilidade de isso se refletir nas outras características, validando grande relação entre as dimensões ligadas ao contexto big data (GANDOMI; HAIDER, 2015).

Alguns autores ressaltam referindo-se ao propósito máximo da análise de grandes volumes de dados, a era de informações valiosas, como autores Moura e Amorim (2015) afirmam:

Expõe uma nova geração de tecnologia e arquitetura, destinada a extrair valor de uma imensa variedade de dados permitindo alta velocidade de captura, descoberta e análise, transformando dados em informações valiosas (Moura e Amorim, 2015, p. 2).

Mayer-Schonberger e Cukier destacam os conceitos atuais que irão ajudar as organizações na hora da tomada de decisão:

Refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar formas de valor de maneira que alterem os mercados, as organizações, a relação entre cidadãos e governos etc. (Mayer-Schonberger e Cukier, 2013, p. 4).

Dumbill (2012) evidencia as dificuldades provenientes dessa atual existência de multiplicação de dados, retratando desta forma, a superação da capacidade de processamento de tecnologias tradicionais. Os autores Goularte, Zilber e Pedron (2015), observam e demonstram o progresso de uma geração de novas tecnologias e estruturas, designada a potencializar a metodologia de captura, descoberta e análise de grande volume de dados:

Não se trata apenas de uma ferramenta, mas é, em verdade, uma geração de novas tecnologias e arquiteturas projetadas para extrair valor econômico de grandes volumes de dados (Goularte, Zilber e Pedron 2015, p. 3).

Contudo, vale destacar que não se pode desconsiderar de ferramentas, pois a capacidade do ser humano de analisar dados e informações com múltiplas facetas é limitada. Logo, são necessários alguns instrumentos que auxiliem a executar estas tarefas. A necessidade de vencer o desafio, reunindo e analisando fontes de diversas naturezas, deu origem a pesquisas que nos levaram ao tema *Big Data*. Estas pesquisas foram delineadas a partir de aspectos iniciais (DAVENPORT, 2014):

1. A múltipla natureza dos dados – aspecto relacionado com as diferentes fontes disponíveis.
2. O uso de processamento em nuvem – aspecto relacionado ao uso ilimitado de recursos computacionais e com processamento em larga escala, com a possibilidade de redução de custos (economia de escala – é o aspecto econômico-financeiro).
3. Uso de tecnologias específicas, tais como processamento de rotinas em paralelo e ferramentas para otimização como Hadoop e MapReduce
4. HDFS, além de abordagens de Machine Learning e Analytics.

Segundo Davenport (2014), “as dificuldades concebidas para examinar a grande parcela de dados, exibidas previamente, são capazes de se apresentar de diversas formas”. Em certa ocasião, os métodos tradicionais usados para manusear dados não alcançam mais acompanhar o ritmo de produção desses dados, disponíveis em vários formatos (DAVENPORT, 2014).

Isto posto, como dissertam Victorino et al. (2017), “muitas pesquisas aparecem determinadas pela procura em desenvolver novas tecnologias que possam resolver os desafios de armazenamento e processamento desse enorme volume de dados, produzidos em grande velocidade e de maneira diversa”. O big data, assim sendo, é mais que apenas um grande volume de dados não estruturados. Ele também inclui as tecnologias que possibilitam seu processamento e análise. Tecnologias específicas de big data têm a aptidão de analisar conteúdo de texto, vídeo e áudio.

No contexto de big data de fluxo rápido, tecnologias como o aprendizado de máquina possibilitam a rápida criação de padrões estatísticos que se encaixam aos dados, os otimizam e os preveem (DAVENPORT, 2013). Em seguida, serão exibidos alguns elementos fundamentais dentro do âmbito big data, bem como tecnologias cruciais para auxiliar a realização da análise e interpretação desses dados.

<b>Perspectiva integral das tecnologias de big data</b>	
<b>Tecnologia</b>	<b>Definição</b>
Hadoop	Software de código aberto para o processamento de big data em uma série de servidores paralelos.
MapReduce	Um framework arquitetônico no qual o Hadoop se baseia.
Linguagens de script	Linguagem de programação que suporta scripts, programas escritos para um sistema de tempo de execução

	especial que automatiza a execução de tarefas que poderiam alternativamente ser executadas uma por vez por um operador humano. Linguagens de script são frequentemente interpretadas (ao invés de compiladas).
Processamento de linguagem natural (PLN)	Subárea da Inteligência Artificial (IA) que estuda a capacidade e as limitações de uma máquina em entender a linguagem dos seres humanos. O objetivo do PLN é fornecer aos computadores a capacidade de entender e compor textos.
Aprendizado de máquina	Método de análise de dados que automatiza a construção de modelos analíticos. É um ramo da inteligência artificial baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana.
Visual analytics	A análise visual é uma análise de dados caracterizada por insights que são exibidos de maneira gráfica e interativa. É especialmente útil na análise de conjuntos de dados grandes e complexos.
In-memory analytics	Processamento de big data na memória do computador para obter maior velocidade.

Quadro 3: Perspectiva integral das tecnologias de big data  
(DAVENPORT, 2013, adaptado)

Dessarte, o termo big data leva a definição do conjunto de informações online e offline presentes e disponíveis na rede mundial de computadores e nos bancos de dados de servidores e empresas os quais são capazes de serem acessados remotamente e dispõem de interligações entre si. Isto é, é um acervo de informações de fontes tradicionais e digitais internos e externos de uma instituição e/ou banco de dados e/ou outros tipos de fontes de informação, que retratam uma fonte de descoberta e análise constantes.

Por conseguinte, o Big Dada Analytics é definido como a análise desses exorbitantes volumes de dados, os quais requerem a utilização de ferramentas e ações adequadas para extrair, cruzar dados e apresentar informações que são excelentes fontes para alicerçar estratégias empresariais. Mediante a definição de big data que vimos outrora, o que o big data analytics realiza é recolher todas as fontes de dados das quais foram mencionadas – matéria-prima de todo

e qualquer dado – e gerar as correlações entre dados, inferências e vínculos entre as informações.

Dessa forma, é admissível encontrar as métricas relevantes para determinada empresa, utilizando algoritmos para tomar decisões e criar estratégias mais eficientes, como por exemplo, insights referentes à descoberta de padrões, as tendências de mercado, comportamento dos consumidores e suas expectativas, dentre diversos outros benefícios aplicáveis à diversas áreas de estudo.

#### 4.2.1 BIG DATA ANALYTICS

Antes da existência do termo Big Data, as análises de dados eram discutidas e estudadas dentro do campo da inteligência artificial (IA) desde a década de 50, sendo que a abordagem do tema se tornou mais popular na década de 90, com a inserção destas análises dentro do campo de tecnologia da informação como Business Intelligence (CHEN; CHIANG; STOREY, 2012).

Russom (2011) enfatiza que o Big Data Analytics é constituído por técnicas analíticas avançadas, exercendo operações sobre grandes conjuntos de dados. O autor, aponta ainda, as razões que unem *big data* com a análise de dados, destacando, dentre tantos motivos: a geração de amostras estatísticas colossais; dados desorganizados contendo uma grande carga de informações importantes; e processos analíticos em grandes conjuntos de dados apresentando resultados com grande impacto para as organizações (RUSSOM, 2011).

Assim como no meio acadêmico, quanto entre as organizações, ainda existem questionamentos acerca do quanto as técnicas e ferramentas de analytics devem substituir os modelos convencionais de análise de dados. No Brasil, por exemplo, 93% de uma amostragem de CEOs acredita que o uso de analytics em Big Data serão essenciais em suas tomadas de decisão (PWC 2016). Big Data não é apenas uma enorme suporte de dados, mas também um objeto de conhecimento capaz de fornecer dados que ajudam a melhor entender os seres humanos e suas interações (BOYD; CRAWFORD, 2012).

Sanders (2016), depreende que a habilidade de capturar, analisar e gerar valor a partir dos dados extraídos é cada vez mais valorizada no mercado de trabalho, além de alegar que o arranjo de Big Data com as ferramentas estatísticas de Analytics faz com que ambos tenham um valor associado para tomadas de decisões para os gerentes e executivos (SANDERS, 2016).

Ainda segundo Sanders (2016), houve uma *disruptura* tecnológica a partir do uso de Big Data Analytics partindo de 3 alterações:

1. Oportunidade de investigação, onde é possível entender a recíproca humana e eventos de maior complexidade em tempo real;
2. Natureza da investigação onde, graças à automatização, é possível coletar e analisar dados para hipóteses ainda não formuladas;
3. Natureza do experimento, através da habilidade de fazer testes de larga escala utilizando dados estruturados, semiestruturados e não estruturados em tempo real, transformando dados em ideias e inputs para melhorias e tomadas de decisão.

Apesar da tendência cada vez maior do uso de analytics em Big Data, modelos heterogêneos – que combinam análise convencional e analytics têm demonstrado melhores resultados e apresentado oportunidades para novas pesquisas na área de Data Science (CHAI; SHIH, 2017).

A dificuldade para as ferramentas de Big Data, é entre outras, a manipulação de dados semiestruturados e não estruturados com o objetivo de extrair valor destes dados, mediante às correlações e outros processamentos de análise e assim entendê-los para que produzam valor ao estipulado âmbito de aplicação.

Cezar Taurion (2013), argumenta que “as ferramentas de Big Data, retratam para as empresas e para a sociedade semelhante relevância que o microscópio equivale para a medicina”. Tal ferramenta de análise na qual é capaz de extrair informações, prever incidentes e ter a capacidade de corrigi-los quando existentes, ou até mesmo evitá-los (TAURION, 2013).

Hodiernamente, tendências de processamento intensivo de dados e de computação paralela apoiada na infraestrutura do Hadoop e no paradigma de programação MapReduce têm se apresentado como soluções para esses e outros tipos de problemas advindos da necessidade de processamento de grandes volumes de dados. A utilização dos conceitos de programação MapReduce com o objetivo de tornar executável o processamento dos grandes volumes de dados (PAIVA, 2016).

#### **4.2.2 MAP-REDUCE**

Em suma, verificou-se que um dos grandes desafios computacionais da atualidade é armazenar, manipular e analisar, de forma inteligente, a grande quantidade de dados existente. Sistemas corporativos, serviços e sistemas *web*, como as mídias/redes sociais e buscadores com acesso a bases de dados de imagens e mapas, além de diversas fontes diversificadas, produzem

juntos um volume de dados acima da capacidade de processamento realizado por tecnologias tradicionais de banco de dados relacional (WHITE, 2010).

Ademais, há uma grande riqueza de informações nesses dados e muitas empresas não sabem como obter valor a partir deles. A maior parte desses dados é armazenada em uma forma não-estruturada e em sistemas, linguagens e formatos bem diferentes e, em muitos casos, incompatíveis entre si. Esses enormes conjuntos de dados têm se tornando uma valiosa fonte de informação. Isso porque, as detentoras desses dados passam a ser avaliadas não somente pela inovação de suas aplicações, mas também pelos dados que elas mantêm e principalmente pela potencialidade que eles podem eventualmente trazer (GOLDMAN et al., 2012).

Um exemplo é a empresa Google, a qual não só possui um alto valor agregado por seu poderoso algoritmo de busca de páginas Web e seus inúmeros serviços disponíveis, mas também por manter um grande volume de dados oriundos de seus usuários. Dados esses que, ao passarem por análises, tendem a se tornar valiosos, permitindo a criação de soluções inteligentes e bem direcionadas aos seus usuários (GOLDMAN et al., 2012).

Diante desse desafio, na primeira década de 2000, alguns pesquisadores e engenheiros da computação aprimoraram e construíram soluções para tornar mais eficiente o acesso e a análise ao *big data*. Entre esses cientistas, dois engenheiros do Google, Jeffrey Dean e Sanjay Ghemawat, desenvolveram a tecnologia MapReduce, que permitiu otimizar a indexação e a catalogação dos dados sobre as páginas *web* e suas ligações (LAMMEL, 2008).

O MapReduce permite dividir um grande problema em vários pedaços e distribuí-los em diversos computadores de baixo custo. Essa técnica deixou o sistema de buscas do Google mais rápido mesmo sendo executado em computadores convencionais e menos confiáveis, reduzindo os investimentos ligados à infraestrutura (DEAN; GHEMAWAT, 2004). Esse modelo destina-se à implementação de aplicações voltadas ao processamento e à geração de grandes bases de dados em plataformas compostas por milhares de processadores.

Os proponentes do modelo argumentam que, apesar da complexidade da maioria das aplicações paralelas, diversas delas podem ser implementadas de forma mais simples e com pouco esforço (VENKATESH, 2010).

O modelo MapReduce consegue abstrair a complexidade de paralelismo e, por meio de uma interface simples, provê mecanismos de tolerância a falhas, a balanceamento de cargas e a distribuição de dados (WHITE, 2009; DEAN; GHEMAWAT, 2008). Assim, os programadores não precisam tratar tais questões diretamente e podem manter seu foco nas aplicações. Essa vantagem tem favorecido a popularização de aplicações MapReduce que tratam do

processamento offline de grandes volumes de dados, tais como o processamento de documentos e de logs de requisições da web (VENKATESH, 2010).

O modelo do MapReduce foi implementado em diferentes linguagens de programação (ISARD et al., 2007; HE et al., 2008; RANGER et al., 2007). Dentre as implementações existentes, uma das primeiras foi o Hadoop (SHVACHKO et al., 2010), desenvolvida pela Apache.

Vance (2009), cita que, a utilização de MapReduce tornou possível a melhoria de produtos como o Google Maps e o Google Earth, sendo utilizado para manipular e montar imagens compostas partindo de grandes quantidades de imagens provenientes de satélites. No mecanismo de busca Google, MapReduce tornou possível criar perfis de busca pela identificação de relacionamentos entre diferentes grupos de sites, imagens e documentos.

À vista disso, MapReduce é o processo analítico do Hadoop desenvolvido para operar com grandes volumes de dados. É o princípio da localidade no qual o código é enviado para o local onde os dados estão para ser processado. O processamento analítico é distribuído em vários servidores, dos quais se deseja tirar informação. Nesse processo, o MapReduce constitui a separação dos dados em frações, mapeia as atividades em cada local e duplica em ambientes e seguidamente faz as reduções.

#### **4.2.2.1 HADOOP**

Uma vez que MapReduce é a filosofia, Hadoop é uma implementação popular do modelo MapReduce, mantida pela fundação Apache. Embora ela possa ser utilizada na sua forma originária, existem ainda diversas implementações comerciais do Hadoop, mantidas e desenvolvidas por empresas como Cloudera, Hortonworks e Microsoft (DAVENPORT, 2014).

Além da implementação MapReduce, existe um ecossistema de aplicações que orbitam o Hadoop, oferecendo funcionalidade adicional ao modelo MapReduce funcional, como por exemplo, Sqoop (integração com sistemas relacionais); Ambari (administração web); Avro (serialização); Cassandra (banco de dados); Chukwa (integração e coleta de dados); Hbase (banco de dados distribuídos); Mahout (machine learning), Spark (processamento de grandes volumes de dados), entre outros (DAVENPORT, 2014).

A abundância de dispositivos junto com os diversos formatos de registros, e a primordialidade da extrair de valor deles, expôs a limitação dos modelos relacionais, que auxiliavam devidamente o tratamento de dados estruturados, porém não viabilizavam o tratamento de dados não estruturados. Tal fato foi um dos essenciais causadores da procura de

ferramentas NOSQL, que trabalham com bancos de dados não relacionais. Além da exorbitante quantidade de dados, sistemas NOSQL são aptos a empenhar-se e trabalhar em sistemas instáveis.

Isto posto, no modelo NOSQL, os dados provenientes de diversos dispositivos – desde smartphones até servidores – são replicados em clusters onde são processados através de ferramentas Analytics, em seguida visualizados através de gráficos, dashboards, entre outras ferramentas de análise (COELHO, 2004).

### 4.2.3 NOSQL

Os tempos atuais trouxeram novos requisitos de gerenciamento de dados. Aplicações para agregar grandes volumes de dados, prontuário eletrônico, gestão de documentos, análises de séries temporais, dentre muitos outros. O modelo relacional preza pela normalização, integridade e não redundância de dados, estas novas aplicações priorizam maior escalabilidade, volume e processamento. A solução veio através de vários produtos construídos sem as características principais presentes no modelo relacional: dados não normalizados, poucas restrições de integridade e controle mínimo de transações. Como consequência, estes produtos se tornaram mais escaláveis, mais flexíveis e com menores custos (DAVENPORT, 2014).

O termo NoSQL não significa exatamente que não se possa usar uma linguagem declarativa para consultar dados em um gerenciador NoSQL, mas sim que são bancos de dados que não estão apenas baseados no modelo relacional. Existem várias “famílias” de produtos NoSQL, cada família compartilha de um mesmo modelo de armazenamento. O modelo mais tradicional, em vez de possuir uma tabela com um número fixo de colunas tipadas, trabalha com o conceito chave-valor, ou KVS, acrônimo para Key-Value Store. Isso significa que em vez de incluir um conjunto de atributos, a operação insere apenas uma chave e um valor, nada mais. Algumas implementações do tipo KVS são Couchbase (anteriormente, Membase), Kyoto Cabinet, Redis e DynamoDB da Amazon (DAVENPORT, 2014).

Indrawan-Santiago (2012), expõe a categoria de gerenciadores de banco de dados NoSQL, baseado no modelo de dados aplicado:

<b>Categoria de gerenciadores de banco de dados NoSQL</b>	
<b>Gerenciadores</b>	<b>Descrição</b>
Chave-valor	Os dados são armazenados conforme o aspecto de pares chave/valor, de maneira estruturado ou não estruturado. Cada uma das chaves é única, e os clientes

	atribuem ou solicitam os valores para cada chave.
Família de colunas	Os bancos NoSQL colunares, ou orientados à coluna (família de colunas) armazenam e processam os dados por coluna ao invés de linhas, como sucede no banco de dados relacional.
Orientado a documentos	Um banco NoSQL orientado a documentos concebe a utilização da definição de par chave/valor para o armazenamento de dados. Porém, é exigida certa estrutura em de que maneira o valor é armazenado. Relativamente com os bancos de dados chave-valor, isso fornece mais informação sobre a estrutura, além de suportar estruturas mais complexas.
Orientado a grafos	Nesta categoria de banco NoSQL, os dados são retratados como uma rede de nós conectados por arestas, o que possibilita a definição e qualificação da conectividade entre as entidades.

Quadro 4: Categoria de gerenciadores de banco de dados NoSQL  
(INDRAWAN-SANTIAGO, 2012)

Ainda segundo Indrawan-Santiago (2012), “há várias espécies de banco de dados NoSQL que apresenta suas essenciais vantagens e desvantagens quando comparados aos relacionais”. As fundamentais vantagens indicadas por Indrawan-Santiago (2012) são: “a flexibilidade de suas estruturas, a alta escalabilidade horizontal, o suporte a dados não estruturados e o processamento distribuído”. Dessa forma, essas propriedades fazem dos bancos de dados NoSQL excelentes dispositivos de empenho para o âmbito de Big Data (VICTORINO et al., 2017).

Uma perspectiva dos autores, Kimball e Ross (2013), declaram que “bancos de dados relacionais e a linguagem de consulta SQL, bastante usados em DWs convencionais, não conseguem analisar a enorme quantidade de dados de um Big Data”. No entanto, Indrawan-Santiago (2012), destaca que “os bancos de dados NoSQL, demasiadamente empregados em esferas de Big Data, foram gerados para disponibilizar suporte a processos operativos que manipulam um grande volume de dados em vários formatos em um período razoável, e não para realizar análises mais desenvolvidas, com propósito de possibilitar auxílio à decisão” (INDRAWAN-SANTIAGO, 2012).

Uma nova perspectiva relevante que tem de ser analisada é que há escassas ferramentas de consultas analíticas a repositórios de dados disponíveis para o Hadoop o qual é o fundamental framework usado em ambiente de Big Data (VICTORINO et al., 2017).

Aliado ao surgimento do Big Data, tem-se o advento da Ciência de Dados como um campo de atuação de competências interdisciplinares em ascensão.

Dessarte, big data define-se como um exorbitante volume de dados não estruturados e estruturados, que, em sua captura, processamento, análise e armazenamento, são gerados de forma rápida e diversificada. Os dados gerados no mundo digital crescem celeremente. Tais informações e dados necessitam de tecnologias inovadoras e meios econômicos para que sejam processados e organizados.

### 4.3 O CONCEITO DE DADOS

Segundo Koltay (2015), a palavra dados apareceu no ano 1700 d.C. significando uma única peça de informação. A definição de dados no *Oxford English Dictionary* online, denomina o termo como “fatos e estatísticas reunidas para referência ou análise de algo, as quantidades (informação estatística), os caracteres e os símbolos de operações lógicas executadas por um computador (software)” (KOLTAY, 2015; NSB, 2005; OED, 2015).

Ademais, o termo dados também desempenha o registro dessas operações – armazenando, preservando e transmitindo sob forma digital, o qual é capaz de ser gravado em todo tipo de mídia magnética, óptica ou de hardware (SEMIDÃO, 2014). Portanto, o termo “dados” deve-se a uma ordem de artefatos que se derivam de ambientes digitais – tais como textos, números, imagens, vídeos, áudios, softwares, algoritmos, equações, animações, modelos e simulações. Em virtude do progresso das tecnologias digitais, os objetos do mundo orgânico são cada vez mais armazenados na forma de dados e compartilhados em redes, frequentemente a Internet. (NATIONAL SCIENCE BOARD, 2011; KOLTAY, 2017; OED, 2015).

Dessa forma, os dados são introduzidos, gerados e criados no computador de diversas formas que os convertem gradualmente variados, complexos e fora do controle humano. Portanto, os dados no ciberespaço — são virtuais e mostram características de um mundo independente — essas características, podem ser semelhantes ou diferentes daquelas encontradas nos dados gerados para representar o mundo natural (ZHU; XIONG, 2015).

Na atualidade, os dados são armazenados em uma condição sempre ativa e podem ser acessados globalmente a qualquer momento, por qualquer usuário. O crescimento exponencial da geração de dados está relacionado a tudo o que usamos durante nossa rotina diária. Segundo

Van der Aalst (2014), “[...] a sociedade, as organizações e as pessoas estão enfrentando um momento em que os dados são coletados sobre qualquer coisa, a qualquer hora e em qualquer lugar”.

Por conseguinte, o uso intensivo de dados estimula um ambiente tecnológico de informações de dados, ideal para pesquisas voltadas para modelos, técnicas e tecnologias usadas para gerenciamento e curadoria de dados. Assim sendo, esse ambiente de dados deve ter competências relacionadas à criação, gerenciamento e preservação de dados (AALST, 2014).

Os dados podem representar objetos no mundo natural, um exemplo são informações pessoais que representam as características de uma pessoa. Por outro lado, os dados podem não representar objetos no mundo natural. Com o aumento das tecnologias digitais, os objetos do universo orgânico são paulatinamente armazenados em redes – web – na forma de dados (ZHU; XIONG, 2015).

A definição de dados muda entre diferentes disciplinas e pesquisadores. Por isso, pode-se inferir que o significado concedido ao conceito de dados necessita da circunstância em que os dados são gerados e da fase do ciclo de vida dos dados em que as informações estarão presentes. Assim, enfatiza a importância de entender o conceito de dados no contexto da biblioteconomia de dados (TIBBO; JONES, 2017).

A biblioteconomia de dados preocupa-se com a representação, organização e disseminação de dados e o uso de tecnologias para projetar serviços de gerenciamento e serviços de dados de pesquisa. Dados de exploração que são matéria-prima para pesquisa científica são objetos e itens heterogêneos utilizados e contextualizados, dependendo da disciplina acadêmica de origem (IASSIST, 2016; KOLTAY, 2015; XIA e WANG, 2014).

Os dados são a matéria-prima da pesquisa, produzida através de qualquer coleta sistemática de informações para fins de análise (HENDERSON, 2017; KELLAM; TOMPSON, 2016). Segundo Tenopir et al. (2017, p. 25), o gerenciamento de dados de pesquisa é um papel importante para as bibliotecas acadêmicas:

[...] Os dados de pesquisa são uma parte essencial do registro acadêmico, e o gerenciamento de dados de pesquisa é cada vez mais visto como uma tarefa importante para as bibliotecas acadêmicas. [...] o bom gerenciamento de dados de pesquisa é essencial para garantir a transparência da pesquisa científica, preservar os dados e permitir a reutilização e reanálise dos dados e o avanço do conhecimento. [...] os dados da pesquisa são cada vez mais vistos como parte essencial do registro

acadêmico. Como as bibliotecas acadêmicas tradicionalmente têm um papel no fornecimento de acesso ao registro acadêmico de várias formas, não é de surpreender que o gerenciamento de dados de pesquisa seja um problema global para as bibliotecas acadêmicas (TENOPIR, 2017).

O enredamento da concepção de dados aponta que a ciência de dados é uma teoria e prática necessária para todos os profissionais que buscam entender as várias dimensões dos dados. Em suma, os dados são demasiadamente difíceis e correntes que potencialmente qualquer tipo de informação digital é capaz de ser algum tipo específico de dados. É viável se referir a tudo em meio digital como dados (MANTRA; 2012-13; RICE; SOUTHALL, 2016).

Em conclusão, dados são símbolos que representam propriedades dos objetos ou eventos e, que podem ser coletados de diferentes fontes, sejam através de meios digitais como sensores de automóveis, de temperaturas e de aviões, câmeras de vigilância, satélites, dispositivos móveis, computadores etc. ou meios não digitais, como prescrições médicas, livros, jornais, relatórios e notas-fiscais impressas, entre outros (SIEGEL, 2018).

Os dados são o insumo principal durante todo o processo de Ciência de Dados e estão classificados em duas categorias: estruturados e não-estruturados, como explicita o quadro a seguir (FAYYAD, 1996):

<b>Dados estruturados</b>	<b>Dados não-estruturados</b>
Os dados estruturados possuem características que podem ser comparadas às tabelas de banco de dados relacionais, onde os registros estão organizados em linhas e cada dado possui uma denominação que são as colunas. Uma nota fiscal, por exemplo, é um dado estruturado pois possui campos com seus respectivos valores, o que torna fácil a identificação de cada item. Geralmente o armazenamento desses dados estão nos bancos de dados relacionais, portanto, os métodos de consultas são simplificados e podem ser realizadas através de linguagem SQL.	Já os dados não-estruturados não possuem uma estrutura lógica de fácil identificação como campos e colunas, ou seja, não existe um modelo pré-definido de organização. Um documento no formato PDF, vídeo, imagem, mensagens em redes sociais são exemplos de dados não-estruturados. Devido a essa diversidade, uma nova abordagem para seu armazenamento foi criada, são os chamados bancos de dados NoSQL (Not only SQL), tradução livre “Não somente SQL”, ou seja, que possibilita também o armazenamento de tipos complexos de dados e que não são suportados pelos bancos de dados tradicionais.

Quadro 5: Dados estruturados e Dados não-estruturados (FAYYAD, 1996)

Isto posto, a ciência com a utilização abundante de dados estabelece uma nova perspectiva para os pesquisadores – o universo dos dados – provocando uma inovação no modo de pensamento científico, a qual embasa teorias e práticas científicas que se empregam para as múltiplas perspectivas do conceito de dados. Considera-se que elas simbolizam um novo desafio de aprendizagem e que se descrevem como uma nova oportunidade aos profissionais bibliotecários, cientistas da informação e bibliotecários de dados.

#### **4.3.1 BIBLIOTECONOMIA DE DADOS**

Segundo Koltay (2017), a biblioteconomia direcionada a dados – Data Librarianship, em inglês – refere-se a vantagem de bibliotecários em entender o trabalho relacionado à gestão e à curadoria de qualquer tipo de dados, tornando sua estrutura e concentração o tratamento, a gestão e a curadoria de dados de pesquisas em toda disciplina científica (KOLTAY, 2017; RICE, 2016; RICE; SOUTHALL, 2016).

A biblioteconomia de dados pesquisa dispor de procedimentos e infraestruturas para o apoio a coleções de dados de pesquisa, analisando dilemas usuais referente ao armazenamento, exploração e ao acesso a dados de pesquisa. Na atualidade, é uma propensão de estudos desenvolvida em bibliotecas acadêmicas, principalmente, na América do Norte (SEMELER; PINTO, 2019).

Os primórdios da biblioteconomia de dados estão no advento das coleções de dados de pesquisa provenientes da aplicação de métodos quantitativos de pesquisa em ciências sociais, principalmente de dados públicos gerados por entidades governamentais, por exemplo, os censos eleitorais e os dados sobre a economia (RICE; SOUTHALL, 2016).

A biblioteconomia de dados não é uma nova área da biblioteconomia, ela consiste em uma variedade de habilidades e competências integradas e aprendidas pelos bibliotecários. Habilidades habituais são reintegradas com a finalidade de serem usadas em conjuntos de dados (RICE; SOUTHALL, 2016). Kellam e Thompson (2016), citam caso parecido, “a catalogação e a organização de materiais, a preservação e a curadoria, os serviços de usuários e de referência, a consultoria e os treinamentos são reconsiderados para serem aplicados junto às novas práticas – que envolvem o gerenciamento, a curadoria e o compartilhamento de dados”.

Porém, a biblioteconomia orientada a dados não ajuda somente à descoberta de novas habilidades em biblioteconomia, mas também rodeia-se com o desenvolvimento e a preservação de dados de pesquisa, empenhando-se com quase todas as funções tradicionais das atividades do profissional bibliotecário – como aquisição e desenvolvimento de coleções,

organização, catalogação de acervos e implementação de serviços de referência aos usuários (RICE; SOUTHALL, 2016; KELLAM; THOMPSON, 2016, SEMELER, PINTO, ROZADOS, 2017).

Segundo Gold (2007), uma das indispensáveis funções da biblioteconomia de dados é ser *upstream*, termo que qualifica essas áreas que se situam distante da zona de conforto habitual dos bibliotecários acadêmicos, como trabalhando a montante no processo de pesquisa, antes do auge de publicação (GOLD, 2007). Isto posto, significa dizer que a biblioteconomia de dados não tem de somente se atentar com informação publicada, porém também com fontes potenciais de dados, sendo o seu foco entender como os diversos tipos de dados são capazes de criar informação útil e relevante para a pesquisa (XIA; WANG, 2014; KOLTAY, 2017).

Rice e Southall (2016), no trabalho *The Data Librarian's Handbook*, discutem o conceito de Biblioteconomia de dados e alegam que esta nova configuração da área alia habilidades tradicionais, auxiliando pesquisadores na descoberta de recursos informacionais, ao mesmo tempo que fornece informações em demandas peculiares de modelos de dados ou obsolescência digital. Assim, essa atual definição requer que o profissional bibliotecário obtenha entendimento qualificado e esmiuçado a respeito dos acervos, assim como consiga conduzir usuários referente as demandas que abrangem direito autoral, gestão de dados, atuação complementar do processo da pesquisa científica e utilização de ferramentas de análise de dados mais adequadas para cada usuário acolhido (RICE; SOUTHALL, 2016).

Em vista disso, o bibliotecário de dados está intimamente ligado ao suprimento de informações importantes para as instituições e empresas bem como para a sociedade, o que o incorpora no âmbito das tecnologias da informação e comunicação e, portanto, no universo do Big Data (RIBEIRO, 2014).

Nesse esfera, o bibliotecário de dados está diretamente introduzido, uma vez que, de acordo com Costa (2010), “seu instrumento de trabalho é a informação e é classificado um gerenciador desta, organizando e disponibilizando de maneira compreensível e adequada”. Assim, esse profissional tem capacidade para a curadoria, isto é, para gerenciar os dados a partir de sua origem até sua utilização, tanto no presente ou no futuro, sustentando sua reprodutibilidade, reutilização e agregando valor a eles (HARVEY, 2010 apud COSTA; CUNHA, 2014).

Ademais, este profissional dispõe de competências e habilidades para operar com Big Data, extraindo valor dos dados na recuperação da informação e os transformando em informação para o usuário, através de uma boa indexação desses dados. O bibliotecário ainda

pode atuar no âmbito de registro e preservação dos dados estruturados ou não estruturados, contribuindo numa adequada recuperação da informação no futuro (PEREIRA, 2013).

Desse modo, compete aos profissionais da informação compreenderem suas funções nesse primordial ambiente de transformações. Pois, pesquisas necessitam ser evoluídas para tornar claro qual a importância de informação científica e tecnológica que disponibilize o avanço tecnológico do Brasil. Logo, os profissionais da informação devem aprender esse novo ambiente de abundância de dados e como são capazes de colaborar para um gerenciamento eficaz da informação (COSTA; CUNHA, 2014).

#### **4.3.2 BIBLIOTECÁRIO DE DADOS**

Os bibliotecários de dados tornam-se, assim, mestres em informação científica, aptos a agregar e auxiliar no progresso de acervos de conjuntos de dados de pesquisa, utilização e preservação dos dados, assim como no acesso a dados antecipadamente arquivados ou na elaboração de novos dados, como componente de seu exercício de pesquisa (RICE; SOUTHALL, 2016).

À vista disso, sendo este uma nova concepção do campo, são raros os conhecimentos destinados ao assunto. Na Ciência da Informação do Brasil, é carente a elaboração de pesquisas que examinem dados de pesquisa e repositórios de dados de pesquisa como âmbito de pesquisa, sendo capaz de mencionar os estudos concebidos por Sayão e Sales (2016), Bertin, Visoli e Drucker (2017), Rocha, Sales e Sayão (2017), Cavalcanti e Sales (2017), Cavalcanti, Sales e Pimenta (2018), Santos e Rockembach (2018), Sayão e Sales (2018), Semeler e Pinto (2019), entre outros (TARTAROTTI; DAL 'EVEDOVE; FUJITA, 2019).

Sales e Sayão (2017), expõe a respeito do significativo desempenho do bibliotecário como conciliador da preservação e do autônomo acesso aos dados de pesquisa, isto é, operando na curadoria das informações elaboradas e consignadas através dos pesquisadores na gestão da pesquisa científica. No meio dos trabalhos constatados na literatura nacional, a orientação comandada por esse autores, é a que mais se associa do prisma da Biblioteconomia de dados, segundo revela a atuação bibliotecária no procedimento. Em linha semelhante de pensamento, a investigação exposta por Semeler e Pinto (2019) declara “as diversas concepções de dados de pesquisa como forma de certificar o surgimento de uma Biblioteconomia conduzida a utilização de dados, dissertando a respeito da essencial atuação de bibliotecários e cientistas da informação nesta nova perspectiva” (TARTAROTTI; DALEVEDOVE; FUJITA, 2019).

Segundo Henderson (2017, p. 16), o bibliotecário de dados é por vezes competente em suas habilidades em demasiadas instituições:

[...] um bibliotecário que trabalha todo o ou parte do tempo com dados. [...] na maior parte das vezes, esse bibliotecário trabalhará em algum tipo de biblioteca institucional, seja ela acadêmica, escolar, pública ou de negócios. Existem bibliotecários que trabalham com dados fora das bibliotecas, por exemplo, criando taxonomias corporativas ou trabalhando com pesquisadores fora da biblioteca e é provável que haja mais empregos fora das bibliotecas, já que os empregadores percebem como um bibliotecário bem treinado pode ajudar a encontrar, a organizar e analisar dados e/ou informações (HENDERSON, 2017, p. 16).

Logo, da mesma maneira que os artefatos convencionais de uma biblioteca física necessitam ser coletados, descritos e organizados de tal forma que sejam capazes de serem utilizados pelo usuário, os dados também precisam. Os usuários de dados, normalmente carecem encontrar dados que enquadrem informações aptas às suas dúvidas de busca, necessitando de acesso aos dados brutos e aos dados secundários, além de precisarem poder citar esses dados como citam outros documentos (HENDERSON, 2017, p. 34-35).

Diante do exposto, um bibliotecário de dados tem potencial de dedicar-se junto a pesquisadores e a estudantes ajudando-os ao longo do manuseio de seus dados de pesquisas, participando da coleta, da organização, da classificação e da preservação dos dados, realizando dessa ação uma extensão de suas atividades corriqueiras (HENDERSON, 2017).

Assim como, poder operar os conceitos indispensáveis da Biblioteconomia, para: adquirir, manter e gerenciar coleções de arquivos de dados usados para análise primária e secundária. Disponibilizando serviços de referência e assistência técnica, assegurando o acesso aos dados e a toda uma coletividade de usuários, além de propiciar instrução referente à entrada e à utilização dos conjuntos de dados, tal qual elemento de ligação para áreas afins sujeitas às Ciência da Informação (INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICES AND TECHNOLOGY, 2016).

Dentre diversas habilidades de um bibliotecário de dados, Semeler e Pinto (2019), destacam “se faz presente se atentar com a análise e contenção de dados, assim como a promoção de dados, marketing de dados, conscientização de dados, coordenação de práticas com dados em toda unidade e instituição, habilidades de negociação e habilidades de

gerenciamento de expectativas de dados”. Além disso, o profissional bibliotecário de dados compartilhará as mesmas habilidades e competências de um gestor de dados, em relação à preservação e à avaliação dos valores econômicos atribuídos aos dados (SEMELER; PINTO, 2019).

Segundo Xia e Wang (2014), sobre a aplicação de tecnologias, o bibliotecário de dados têm de a dispor da habilidade *hacker* de um cientista de dados. O termo indica um profissional introduzido em circunstâncias de processamento de grande desempenho e criatividade realçando as competências criativas dos especialistas das linguagens de programação. Enfim, um hacker é um programador habilidoso para isso deve conhecer uma variedade de linguagens de programação (DAVENPORT, 2014).

Portanto, segundo Dekker e Lackie (2016), “o bibliotecário de dados tem que possuir as habilidades necessárias para trabalhar com qualquer tipo de dados – sejam eles observacionais, computacionais ou experimentais. Além de perceber como as indagações a respeito dos dados transformam a práxis e os conceitos que baseiam a biblioteconomia”.

E por fim, de impedir a função de auxiliar pesquisadores a coletar, localizar e recomendar artifícios de dados adequados, de maneira que a utilização de suprimentos de serviços de dados de pesquisa, os bibliotecários de dados são capazes de proporcionar serviços de consultoria referente à organização e a aquisição de dados para investigação de problemas de pesquisa em disciplinas inerentes (DEKKER; LACKIE, 2016).

Para terminar, o ofício de um bibliotecário de dados consegue estar relacionados à dominação de dados, entretanto também unindo com a função humana de interpretação, tomada de decisão e comunicação de dados (SEMELER; PINTO, 2019). McCaffrey e Giesbrecht (2016), citam certas habilidades e competências essenciais de um bibliotecário de dados, eles apontam três campos distintos para o exercício desse profissional – gestão e curadoria de dados, visualização de dados e representação de dados geoespaciais além de serviços avançados de referência a dados de pesquisa (MCCAFFREY; GIESBRECHT, 2016).

Kennan (2016, p. 1-10), expõe uma pesquisa referente as habilidades exigidas por aqueles que empregam bibliotecários de dados. Ela lista as habilidades requerentes por instituições de pesquisa e universidades nos Estados Unidos, Reino Unido, Canadá e Austrália, conforme o quadro 6 apresenta:

<b>Habilidades exigidas aos bibliotecários de dados</b>	
<b>Habilidades</b>	<b>Descrição</b>

Interpessoais e características comportamentais	Competência referente à comunicação formal, como a escrita de documentação técnica e estudos de caso. Além de ser capaz de adaptar-se às inovações e experiências e estar incessantemente à procura de inovação.
Conhecimento contextual sobre ambientes institucionais	Habilidade da qual abrange ideias sobre as políticas de financiamento concedidas por agências que promovem a pesquisa científica. Necessita de conhecimento de técnicas éticas do processo científico, métodos de pesquisa disciplinar, formas de comunicação científica, propriedade intelectual, formas de acesso, marcos legais e regulatórios ( <i>copyright e Creative Commons</i> ) e políticas de direito autoral.
Aplicação de dados	Compreende o entendimento de diversos dados – quantitativos e qualitativos – além de modelos e esquemas de metadados, ( <i>Dublin Core, RDF</i> ), bem como indagações sobre os identificadores únicos ( <i>Digital Object Identifiers</i> ) e preservação de dados digitais.
Compreensão referente às tecnologias de manipulação de dados	Os estudos sobre tecnologia mais essenciais aos bibliotecários de dados contêm as linguagens de programação ( <i>Python, SQL, Java, XML</i> ), design e estrutura de bases de dados, APIs de recuperação de dados, o design focado no usuário, ferramentas de processamento de linguagem natural, <i>internet of things</i> e Big Data.

Quadro 6: Habilidades exigidas aos bibliotecários de dados

Fonte: (KENNAN, 2016, p. 1-10).

Segundo Mccaffrey e Giesbrecht (2016), este bibliotecário de dados precisa modificar os dados de maneira reutilizável, compartilhável e conservável em futuro distante. Os autores enfatizam que “o bibliotecário de dados necessita dominar as estruturas de metadados, os repositórios de dados, os conjuntos de dados que são usados por seus usuários pesquisadores, além de ter um ótimo conhecimento a respeito das mais variadas metodologias e disciplinas científicas” (MCCAFFREY; GIESBRECHT, 2016).

Todavia, essencialmente, dispondo de habilidades e competências referentes às tarefas de coleta, de descrição, de ensino, de construção de sistemas, de elaboração de serviços de referência e de criação de políticas de gerenciamento de dados institucionais (MCCAFFREY;

GIESBRECHT, 2016; KENNAN, 2016; DEKKER, LACKIE, 2016; SEMELER, PINTO, ROZADOS, 2017).

Dessarte, a disposição de bibliotecários de dados em possuir habilidades técnicas para gestão de dados de pesquisa, comprova que a utilização de dados em biblioteconomia de dados consegue estar ligado à produção de produtos e serviços de dados vinculados ao gerenciamento de dados de pesquisa em unidades de informação. Revelando dessa forma, a função dos bibliotecários de dados no que diz respeito ao uso, à gestão e à curadoria de dados de pesquisa (SEMELER; PINTO, 2019).

Recapitulado a concepção e determinadas propriedades do conhecimento referente à dados – dados de pesquisa e biblioteconomia de dados – contempla-se que a dificuldade do conceito de dados indica a ausência de uma ciência que estude habilidades, métodos, técnicas e tecnologias sobre o uso intensivo de dados em biblioteconomia (SEMELER; PINTO, 2019).

Decerto, o bibliotecário de dados necessita aprender mais sobre ciência de dados. A tarefa fundamental dos bibliotecários de dados será a de transformar suas tarefas e atividades triviais nas bibliotecas, inserindo os dados de pesquisa no contexto da geração de produtos e/ou serviços em bibliotecas (SEMELER; PINTO; ROZADOS, 2017). Os bibliotecários de dados têm competência para auxiliar a Ciência de Dados propiciando disponibilidade à ferramentas de treinamento e instrução para ajudar a aprimorar o alicerce de conhecimento que cerca o big data.

### **4.3.3 CIÊNCIA DE DADOS**

A expressão Ciência de Dados – Data Science – foi uma área existente há alguns anos, entretanto obteve visibilidade e interesse nos últimos anos por vários motivos, como a proliferação de grandes bancos de dados. Os dados são um atuante indagador e a Ciência de Dados é a essência interdisciplinar que tem fascinado gradualmente a magnitude e conferências nas esferas de estatística, análise, computação, Ciências Sociais e demais competências e matérias científicas (CAO, 2016). Ela tem sido estudada e considerada como uma área com característica interdisciplinar por parte dos pesquisadores da temática (CONWAY, 2010; STANTON, 2012; ZHU, XIONG, 2015; STREIB, MOUTARI, DHEMER, 2016) ou multidisciplinar (TIERNEY, 2016) em sua origem.

Há distintos conceitos e definições, segundo Cao (2016), acerca da Ciência de Dados. De modo como no âmbito da ciência, que aprimora e cria metodologias, teorias, tecnologias e aplicativos de interesse para os dados. Segundo o autor, “a partir da captura, criação,

representação, armazenamento, pesquisa, compartilhamento, privacidade, segurança, modelagem, análise, aprendizagem, apresentação e visualização”. Além da inserção de processos complexos, diversificados e dependentes, para a tomada de decisões em tempo recente, contribuindo, dessa maneira, e gerando relevância e auxílio à decisão (CAO, 2016).

Como já mencionado anteriormente, acerca do conceito Ciência dos Dados, Cao (2017) expõe uma declaração que, é centrada na transdisciplinaridade: “A ciência de dados é uma nova área transdisciplinar que ergue e harmoniza diversas disciplinas e consistências de cognição valiosos, dentre elas acha-se a estatística, informática, computação, comunicação, gerenciamento e sociologia” (CAO, 2017).

Ademais, de acordo com Rice e Southall (2016), “ciência de dados é uma conjunção de análise de dados e desenvolvimento de novos produtos e serviços com base nos dados. Combinando, assim, um conjunto de técnicas, incluindo filtragem de dados, *machine learning*, design de sistema, conhecimento de domínio e visualização”. Logo, a ciência de dados é uma área do saber que combina métodos de ciência da computação e estatística, entre outras. Em síntese, os dados são demasiadamente complexos e fluidos que ciberneticamente todos as formas de informação digital podem ser algum tipo específico de dados. É plausível se mencionar a tudo em meio digital como dados (MANTRA; 2012; RICE; SOUTHALL, 2016).

Wang, definiu a ideia de Ciência de Dados através da apresentação de várias outras definições formadas por distintos pesquisadores e segundo Wang (2018), “existe aparentemente certa concordância de que a ciência de dados é uma área multidisciplinar que se refere à identificação e extração de padrões, convertendo dados em informação e conhecimento através de análise de dados e mineração” (WANG, 2018).

Logo, o objetivo da ciência de dados é divulgar as qualidades e composições de ocorrências orgânicas, humanas e sociais, aplicando dados. Por assim dizer, é possível conceituar um novo campo de estudo para bibliotecários que buscam se rodear em conteúdo de gestão e análise de dados. Zhu e Xiong (2015), na pesquisa de um status científico para a ciência de dados, descreveram os diferentes objetivos que a tornam uma disciplina científica. Os autores sugeriram que a ciência de dados é o estudo de dados científicos, o estudo de dados de negócios, a integração de estatísticas, a tecnologia computacional e inteligência artificial e, por fim, a chave para dilemas científicos e comerciais referentes à ascendência de conhecimento (ZHU; XIONG, 2015).

A Estatística vai além da mera criação de gráficos e cálculo de médias. As informações numéricas são alcançadas com o objetivo de reunir informação para a tomada de decisão. Assim sendo, é possível definir a estatística conforme um conjunto de técnicas para idealizar

experiências, captar dados e organizá-los, sintetizá-los, analisá-los, interpretá-los e extrair resultados deles. Isto posto, é possível entender a Estatística como a Ciência dos Dados, ou seja, sendo de enorme valor o controle e poder que seu âmago de conhecimento pode disponibilizar e oferecer. A Estatística se dividi em duas áreas essenciais, são ela a estatística descritiva e a estatística inferencial/indutiva (GASTÃO, 2009).

Segundo Hayashi (1996, p. 41), a ciência de dados “[...] Não é apenas um conceito sintético para unificar a análise de dados e seus métodos relacionados, mas também seus resultados”.

Em um artigo recente, Burton e Lyon (2017, p. 33) apresentaram uma definição de ciência de dados e o papel crucial dos bibliotecários em seu desenvolvimento:

Os avanços em estatística e ciência da computação, combinados com uma abundância de dados, deram origem a um novo ecossistema profissional chamado ciência de dados. [...] Trata de “gerar insight a partir dos dados para informar a tomada de decisões”. Os métodos e produtos de ciência de dados transformaram o comércio, a saúde e o governo, e continuarão a transformar outros setores. [...] à medida que a sociedade está cada vez mais infundida com dados, os bibliotecários terão um papel crucial no desenvolvimento futuro do ecossistema de ciência de dados em vários setores; o relatório reconhece que “são necessários investimentos para expandir o atual pipeline de suporte ao campo da ciência de dados” (BURTON; LYON, 2017).

A ciência de dados surgiu como um esforço multidisciplinar que combina inúmeras habilidades e vários campos de conhecimento para analisar ou prever, criativa e intuitivamente, os conjuntos de dados que são atualmente gerados. Organizações como Google, Amazon, Yahoo, Facebook, Apple, Microsoft, Netflix e Twitter estão na vanguarda da ciência de dados, pois já estabeleceram suas equipes de cientistas de dados (NIELSEN; BULINGAME, 2012; ZHU; XIONG, 2015).

Como uma nova disciplina de estudo e análise de dados, a ciência de dados requer pensamento sistemático combinado com uma conduta inovadora para deliberar impasses normalmente habituais. O planejamento é crucial para trabalhar com a ciência de dados, que

requer conhecimento de várias habilidades e uso de recursos e abordagens da tecnologia para obter os resultados desejados durante a manipulação de dados (VOULGARIS, 2014).

É imprescindível realizar a apuração e a avaliação de qualidade dos dados para elaborar uma Ciência de Dados evoluída. Pois, a ciência da informação consegue realizar uma função relevante sob este aspecto, visto que as concepções da CI se caracterizam como ótimos aspirantes para respaldar a área da habilidade na Ciência de Dados (WANG, 2018). Segundo, Wang (2018), para solucionar a explosão de informações que lidamos na era do big data, manifesta-se o bibliotecário de dados neste ambiente, o qual exerce suas funções no universo de dados, com os conceitos, conhecimentos e aptidões da Ciência da Informação que são admissíveis e benéfico no gerenciamento e no auxílio a Ciência de Dados (WANG, 2018).

De acordo com Wang (2018), “o que os bibliotecários de dados elaboram compreende todos os ciclos de vida do gerenciamento de dados”. Ainda segundo Wang, “sua função inicia-se nos projetos de gestão, coleta, preservação, curadoria, controle, acesso, metadados, documentação dos grupos de dados, compartilhamento, visualização, suporte à análise, avaliação de qualidade, referência, citação e treinamento em alfabetização de dados” (WANG, 2018).

Assim sendo, os bibliotecários de dados têm de a ter habilidade tática e tecnologia de gerenciamento de dados extraordinários, porém, por outra vertente, eles têm de a possuir um conhecimento vasto das implicações socioculturais e éticas da tecnologia de dados, em razão de que, sua compreensão dos dados é sociotécnica (WANG, 2018).

Em suma, os bibliotecários de dados necessitam compreender as técnicas, os métodos e as tecnologias utilizadas em ciência de dados para que sejam capazes de possuir as habilidades e competências atuantes indispensáveis utilizadas pelos cientistas de dados para coleta, manipulação, análise e visualização de dados de pesquisa em repositórios digitais.

#### **4.3.4 CIENTISTA DE DADOS**

O termo cientista de dados apareceu pela primeira vez em 2005, mas tornou-se popular em 2012, em um artigo intitulado "Ciência de dados: a profissão mais estimulante e atraente do século 21", publicado na *Harvard Business Review*. O economista-chefe do Google chamou o trabalho de ciência de dados de "o mais promissor para os próximos dez anos". O artigo revelou a necessidade emergente de cientistas de dados (DAVENPORT; PATIL, 2012; VAN DER AALST, 2014).

Contudo, a procura por profissionais com competências analíticas e técnicas para enfrentar enormes e diversos volumes de dados, vem repercutindo em um crescimento nas propostas de formações para a atuação deste caráter profissional. Porém, não se registra muito a respeito da literatura científica nacional e internacional destes estudos, assim como em relação as habilidades e competências almejadas dos lecionados (CURTY; SERAFIM, 2016).

A dificuldade, assim sendo, não é identificar dados, entretanto ter a sabedoria e o conhecimento adequado do que realizar com estes dados produzir de maneira apreciável o manuseio do seu imenso potencial de aplicabilidade. Há, dessa forma, um gradual consumo de análise de dados através de um comportamento abrangente e pluridisciplinar, que aprecia constituir e mesclar dados oriundos de distintas matrizes, e é exatamente esta conduta que caracteriza a ciência de dados e o cientista de dados (LOUKIDES, 2012).

Segundo Curty e Serafim (2016), “o artigo acadêmico britânico JISC, sobre as habilidades, das atividades e da carreira dos cientistas de dados, admite a complexidade de uma unanimidade referente ao conceito deste profissional, porém o conceitua de maneira abstrata como aquele que trabalha no próprio lugar onde pesquisas são feitas”. Dessa forma há uma limitação na contribuição com os pesquisadores e equipes de cientistas em centros de dados, o qual está implicado na análise de dados e em sua investigação criativa, disponibilizando recursos tecnológicos para a manipulação e utilização dos dados digitais (SWAN; BROWN, 2008).

Contudo, segundo Swan e Brown (2008), “é considerável cogitar que esta descrição é um tanto restritiva em comparação a outras encontradas na literatura há a definição de outras funções, como gerenciadores de dados e bibliotecários de dados, fragmentando suas atribuições na esfera da ciência de dados”.

A definição de cientista de dados é ampla, pois procura abranger outras profissões, como analista de inteligência de negócios e analista de dados. A distinção entre essas profissões está relacionada à capacidade dos cientistas de aplicar conhecimentos de estatística e computação de dados de forma criativa, para gerar novos produtos e serviços de dados. Assim, um cientista de dados deve ser criativo e capaz de criar soluções para as iniciativas de gerenciamento para a aplicação de tecnologias para interpretação de dados (PORTER, 2015; SAUNDERS, 2013).

De acordo com Voulgaris (2014), um cientista de dados, deve ser capaz de trabalhar de forma autônoma, ter habilidades de comunicação escrita e verbal, ser proativo em ambientes complexos, trabalhar em ritmo acelerado e observar as constantes mudanças que ocorrem em um sistema de informação. Além disso, o cientista deve ter a capacidade de consultar bancos de dados e realizar análises estatísticas, conhecer os conceitos básicos de negócios e estratégias.

Além de ser capaz de criar exemplos, protótipos e demonstrações para ajudar no entendimento dos dados de administração e trabalho. Por fim, um cientista de dados deve ter um bom entendimento dos princípios do design e arquitetura da informação (VOULGARIS, 2014).

Para Stanton et al. (2012) “a constituição do cientista de dados é amplamente multidisciplinar, pois eles são responsáveis pela identificação, coleta, tratamento, transformação, análise, visualização e curadoria de enormes blocos de dados distintos”. Além de dispor de uma extensão de conhecimento nos campos de curadoria, análise, ciberinfraestrutura e a dominação indispensável na esfera de aplicação dos dados, o que difere o cientista de dados de outras características profissionais é o destaque no auxílio às carências de dados de usuários bem como na tomada de decisão (STANTON et al., 2012).

Segundo Van Der Aalst (2014), “o cientista de dados necessita dispor clareza sobre o domínio em análise e conseguir transmitir mensagens claras e concisas através da utilização de tecnologias de dados”. Ademais ele tem de a ter conhecimento sobre o significado dos processos de coleccionar, analisar, interpretar e visualizar dados oriundos de variadas fontes (VAN DER AALST, 2014).

Como se viu, existem diversas definições sobre o perfil e as funções desenvolvidas por um cientista de dados (DAVENPORT; PATIL, 2012; VAN DER AALST, 2014). Em suma, este profissional trabalhará na busca de encontrar informações relevantes espalhadas pelo descomunal volume de dados, além de que um cientista de dados lida com conhecimento sobre tecnologias, formas de comunicação, habilidades analíticas e domínio aplicados no ciclo evolutivo dados → informação → conhecimento (SWAN; SHERIDAN, 2008).

Não obstante, a ciência de dados quer um tanto mais vasta, é possível de ser delineada como uma estratégia para atingir esse objetivo. Tendo identificado a combinação de disciplinas que compõem um cientista de dados, é importante avaliar sua influência no conceito de bibliotecário de dados. O ponto central da biblioteconomia de dados situa-se na gênese de novos ofícios de biblioteca e na transformação dos serviços de consulta de pesquisa existentes, com base em novas maneiras de gerenciar e curar dados digitais a partir de pesquisas científicas (VAN DER AALST, 2014).

As tecnologias emergentes são inovações ou aperfeiçoamento de uma tecnologia já existente das diversas áreas da tecnologia que estão em tendência. Essas evoluções trazem melhorias nos processos, ajudam na tomada de decisão e aumento da lucratividade para todos os que usufruem delas – aplicadas em uma organização ou para uso pessoal. Dentre tantas tecnologias, é possível apontar uma que utiliza a Ciência de Dados e que está em crescente uso: Data Mining (Mineração de Dados).

#### 4.4 MINERAÇÃO DE DADOS

A partir da emergência dos dispositivos eletrônicos – hardware e software – uma das cruciais finalidades das organizações vem sendo a de armazenar grandes quantidades de dados. Ademais, atualmente essa propensão se tornou cada vez mais perceptível com o declínio dos gastos para a compra de hardware, fazendo-se praticável armazenar grandes quantidades de dados (AMO, 2016).

Dessa forma, o desencadeamento do grande crescimento de volume de dados armazenados expandindo diariamente, pôs em foco o desafio do que fazer com os dados armazenados. Pois, os profissionais das empresas e organizações começaram a se preocupar com os grandes volumes de dados informáticos estocados e inutilizados dentro da empresa. As técnicas habituais e tradicionais de exploração de dados não eram mais adequadas para tratar a grande maioria dos repositórios. Com o propósito de desintrincar esse desafio, foi proposta – no final da década de 80 – a Mineração de Dados, do inglês *Data Mining* (LAROSE, 2005).

Nesta época, Mineração de Dados consistia essencialmente em extrair informação de gigantescas bases de dados da maneira mais automatizada possível. Atualmente, Mineração de dados abarca especialmente a análise dos dados após a extração, procurando, por exemplo, levantar as necessidades reais e hipotéticas de cada cliente para realizar campanhas de marketing, entre outros. Assim sendo, novas e mais complexas estruturas de armazenamento foram desenvolvidas, como banco de dados, Data Warehouse, Bibliotecas Virtuais, Web, entre outras (CIOS, K. J; PEDRYCZ, W; SWINIARSKI, R. W; KURGAN, L. A., 2007).

Isto posto, a Mineração de Dados é segundo Larose (2005), “uma das tecnologias mais promissoras da atualidade. Uma das razões deste êxito é o fato de dezenas, e centenas de milhões de reais serem gastos pelas empresas na coleta dos dados e, contudo, nenhuma informação útil é identificada” (LAROSE, 2005). Han (2006), refere-se a essa conjuntura como “rico em dados, pobre em informação”.

Em síntese, de acordo com Amo (2016), “a mineração de dados, dedica-se a extrair ou minerar conhecimento de grandes volumes de dados. A palavra mineração de dados é conceituada por vários como sinônimo de Descoberta de Conhecimento em Banco de Dados”.

Os conceitos sobre o termo Mineração de Dados diversificam-se com o campo de atuação dos atores, por ser considerada uma área multidisciplinar (CAMILO; SILVA, 2009). Destaca-se três áreas que são comumente expostas nos consensos sobre o tema e consideradas

como de maior expressão dentro da Mineração de Dados: Estatística, Aprendizado de Máquina e Banco de Dados. Zhou (2003), analisa comparativamente essas três perspectivas citadas.

Hand (2001), define por uma perspectiva estatística: "Mineração de Dados é a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados".

Pode-se entender a Mineração de Textos como uma evolução da temática de Recuperação de Informações, sendo ela um processo de descoberta de conhecimento, a qual usa técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras. Abarca a utilização de algoritmos computacionais que processam textos e identificam informações relevantes e obscuros, os quais geralmente não seria capaz de serem recuperados usando métodos convencionais de consulta, visto que, a informação compreendida nestes textos não pode ser obtida por meio direto, pois normalmente elas estão armazenadas em formato não estruturado (SILVA, 2002).

Em Cabena et al. (1998), a definição é dada por uma perspectiva de banco de dados: "Mineração de Dados é um campo interdisciplinar que junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados".

A definição pela perspectiva do aprendizado de máquina (Machine Learning) é exposta em Fayyad et al. (1996): " A mineração de dados é um avanço no sistema de exploração dos saberes que se compõe na execução da análise dos dados e na aplicação de algoritmos de descoberta que, sob certos entraves digitais, geram um grupo de padrões de determinados dados."

Embora as concepções a respeito da mineração de dados conduzir a crer que a técnica de extração de informação se faz de uma maneira plenamente automática, compreende-se atualmente que deveras isso não detém veracidade. Dessa forma, mesmo encontrando diversas ferramentas que auxiliam na execução dos algoritmos de mineração, os resultados ainda precisam de uma análise humana. Entretanto, a mineração contribui de forma significativa no processo de descoberta de conhecimento, permitindo aos especialistas concentrarem esforços apenas em partes mais relevantes dos dados (LAROSE, 2005).

A Mineração de Dados é, segundo Larose (2005), "geralmente categorizada pela sua competência em executar certas atividades, sendo importante diferenciar o que é uma tarefa e o que é uma técnica de mineração".

De acordo com Han e Kamber (2001), "a tarefa traduz-se na especificação do que está se pretendendo procurar nos dados, que categoria de padrões de constância dispomos relevância

em recuperar, ou que tipo de padrões poderiam nos vislumbrar”. Tal como, um gasto supérfluo de um cliente de cartão de crédito, fora dos padrões habituais de seus consumos (HAN; KAMBER, 2001).

As técnicas em mineração de dados consistem na caracterização de técnicas que nos assegurem de que forma encontrar os padrões que nos interessam. No meio dos métodos indispensáveis usados em mineração de dados, dispomos das técnicas de estatísticas, métodos de aprendizado de máquina dentre outras (AMO, 2016).

As principais tarefas de mineração são (BERRY; LINOFF, 1997):

- Descrição: Segundo, Berry e Linoff (1997), esta funcionalidade pode ser empregada numa análise para tornar mais clara alguma ideia que está sendo utilizada, hipóteses ou fatos observados em um banco de dados. Pode ser executada em uma análise de classificação quando as classes não estiverem bem definidas. Como exemplos, pode-se estudar a afirmação de que “mulheres tendem a votar em candidatas femininas em maior número do que os homens” ou que as características de uma pessoa que fralda cartões de crédito são do tipo “sexo masculino, idade entre 25 e 40 anos e possui nível superior” (WESTPHAL; BLAXTON, 1998).
- Associação: Um preceito de associação é um padrão da forma  $X \rightarrow Y$ , onde X e Y são conjuntos de valores – objetos adquiridos por um cliente, sintomas exibidos por um paciente etc. Imaginemos, como exemplo, um supermercado, o seguinte padrão do qual clientes que compram pão também compram manteiga, reflete uma regra de associação que repete um padrão de comportamento dos clientes do supermercado. Descobrir regras de associação entre produtos comprados por clientes numa mesma compra pode ser útil para melhorar a organização das prateleiras, facilitar (ou dificultar) as compras do usuário ou induzi-lo a comprar mais (HAN; KAMBER, 2001).
- Classificação: é o processo de encontrar um conjunto de modelos que descrevem e diferenciam classes ou conceitos, com o propósito de utilizar o modelo para indicar a classe de objetos que ainda não foram classificados. O modelo construído baseia-se na análise prévia de um conjunto de dados de amostragem ou dados de treinamento, contendo objetos corretamente classificados. Por exemplo, suponha que o gerente do supermercado está interessado em descobrir que tipo de características de seus clientes os classificam em “bom comprador” ou “mau comprador”. Um modelo de classificação poderia incluir a seguinte

regra: “Clientes da faixa econômica B, com idade entre 50 e 60 são maus compradores” (HAN; KAMBER, 2001).

- Agrupamentos (*clustering*): Esta funcionalidade visa segmentar um conjunto de dados num número de subgrupos homogêneos ou clustering. Seu objetivo é formar grupos baseados no princípio de que esses grupos devem ser o mais homogêneos em si e mais heterogêneos entre si. A diferença fundamental entre a formação de agrupamento e a classificação é que no agrupamento não existem classes predefinidas para classificar os registros em estudo. Os registros são agrupados em função de suas similaridades básicas, ou seja, quando se deseja formar agrupamentos, seleciona-se um conjunto de atributos (variáveis) e em função da similaridade desses atributos são formados os grupos (WEIS; INDURKHYA, 1999).
- Análise de Padrões Sequenciais: Um padrão sequencial é uma representação. A organização em que esses conjuntos estão colocados apresentam a ordem cronológica em que ocorrerão os fatos retratados por estes grupos. Dessa forma, tal como, clientes que comprem carro, posteriormente adquirem pneu de carro. Descobrir tais padrões sequenciais em dados temporais podem ter proveito em campanhas de marketing, por exemplo (THURASINGHAM, 1999).
- Análise de Outliers: Um banco de dados é capaz de englobar dados que não mostram o comportamento comum da maioria. Estes dados são chamados de outliers (exceções). Diversas técnicas de mineração desprezam estes outliers como sendo uma interferência rejeitada. Todavia, em certas aplicações, tais como detecção de fraudes, esses episódios inabitual podem ser mais relevantes do que fatos que ocorrem habitualmente. Tal como, conseguimos identificar a utilização fraudulenta de cartões de crédito ao descobrir que determinados clientes efetuaram compras de valor extremamente alto, fora de seu padrão habitual de gastos (ELMASRI; NAVATHE, 1999).

Dessarte, a mineração de dados é um processo complexo que requer várias ferramentas e pessoas diferentes, ou seja, o sucesso de uma mineração de dados depende de um projeto da combinação adequada de boas ferramentas e analistas qualificados. Além disso, requer uma metodologia sólida e gerenciamento eficaz de projetos. Um modelo de processo padrão pode ajudar a entender e gerenciar as interações ao longo deste processo complexo. Para o mercado, haverá muitos benefícios se um modelo de processo comum for aceito (AGRAWAL, 1999).

A Ciência de Dados utiliza das técnicas e tarefas da Mineração de Dados. Dessa forma, podemos afirmar, grosso modo, que a Ciência de Dados ampliou o escopo da Mineração de Dados.

Os analistas que executam projetos de mineração de dados também podem se beneficiar de várias maneiras. Para iniciantes, o processo desse modelo fornece orientação, ajuda a estruturar o projeto e fornece conselhos para cada tarefa do processo. Até analistas experientes podem se beneficiar de listas de verificação para cada tarefa para garantir que nada importante seja esquecido (MOORE, 1991).

Mas o papel mais importante de um modelo de processo comum é para comunicação e documentação de resultados. Ajuda a conectar as diferentes ferramentas e diferentes pessoas com habilidades e conhecimentos diversos juntos para formar um projeto eficiente e eficaz, isto posto, um dos modelos de processos mais utilizados em Mineração e Ciência de Dados é o CRISP-DM (AGRAWAL, 1999).

#### 4.4.1 CRISP-DM

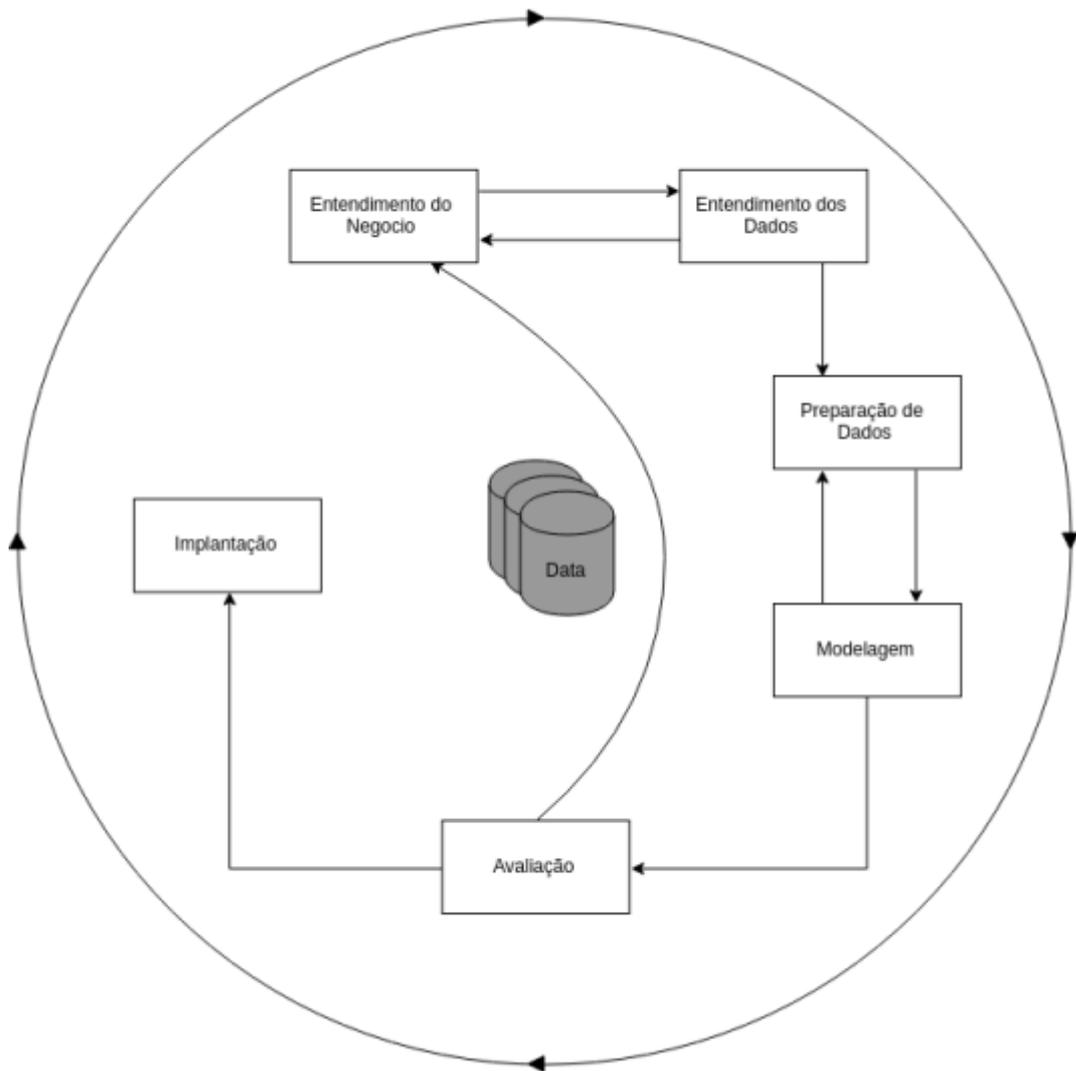
Atendendo aos objetivos e ao problema, o modelo de processo que se revela mais adequada é a metodologia CRISP-DM – a qual é mais usual em problemas que envolvam *data mining* – tendo uma taxa de preferência entre os profissionais de 42% (KDNUGGETS, 2007).

A seleção desse processo, em detrimento de outras reside no fato de esta ser mais completa e iniciar-se com o estudo do negócio, ou seja, o CRISP-DM é um processo que se centra nas necessidades dos gestores e na resolução dos seus problemas de gestão. Esse processo contempla seis fases que são flexíveis. O projeto CRISP-DM (Processo Padrão da Indústria Cross para Mineração de Dados) propôs um modelo de processo abrangente na iminência de produzir projetos de mineração de dados (CLIFTON; THURASINGHAM, 2001).

O modelo de processo CRISP-DM, visa tornar grandes concepções de mineração de dados, menos dispendiosos, mais confiáveis, repetíveis, gerenciáveis e mais rápidos. Além disso, fornece uma visão geral do ciclo de vida de um dado projeto de mineração. Ele contém as fases de um projeto, suas respectivas tarefas e seus resultados (REINARTZ; WIRTH, 1995).

O ciclo de vida de um projeto de mineração de dados é dividido em seis fases, mostradas na Figura 1, explicadas em seguida (BERRY; LINOFF, 1997)

Figura 1 – Padrão de Referência CRISP-DM



Fonte: Ajustado de Wirth e Hipp (2000)

Em seguida de modo mais específico, cada uma das seis fases expostas acima tem suas atividades, essas fases e atividades foram detalhadas por (JACKSON, 2002).

1. **Compreensão do negócio:** Essa fase incipiente busca compreender como metas, os propósitos e primordialidades em prisma de negócios e, em seguida, modificar essa compreensão, em uma atividade de mineração de dados, numa proposta preliminar de investir no problema.
2. **Compreensão dos dados:** Entende-se como a distinção da informação que venha a ser importante para o conhecimento e uma prévia adaptação com a temática, definição, qualidade e usabilidade dos dados.

3. **Preparação dos dados:** Engloba integralmente as ações para produzir o conjunto de dados definitivo – dados que serão alimentados nas ferramentas de modelagem – a partir dos dados brutos iniciais. Traduz-se em uma sequência de tarefas selecionadas a atingir um grupo final de dados, baseado do qual, será usado na modelagem. Nessa fase estão, seleção de atributos, softwares de extração, limpeza e transformação de dados.
4. **Modelagem:** No estágio de modelagem, são designadas e empregadas os métodos de exploração de cognição mais adequadas referente aos objetivos almejados. Haja vista, essa fase compõe-se em distinguir os algoritmos a serem usados e o satisfatório tratamento do modelo.
5. **Avaliação:** Essa fase representa a avaliação dos modelos, para averiguar se os resultados vão ao encontro dos objetivos do negócio, delineados a princípio. A avaliação está orientada para a conexão do recurso à necessidade do negócio e à capacidade de suprir a carência.
6. **Implantação:** A criação do modelo geralmente não é o fim do projeto. Geralmente, o conhecimento adquirido precisará ser organizado e apresentado de forma que o cliente possa usá-lo. Dependendo dos requisitos, a fase de implantação pode ser tão simples quanto gerar um relatório ou complexa como implementar um processo repetitivo de mineração de dados. Em muitos casos, pertencerá o usuário e não o analista de dados, que executará as fases de implantação. De qualquer forma, é importante entender de antemão quais ações precisarão ser realizadas para realmente fazer uso dos modelos criados.

Portanto, a utilização de ferramentas tecnológicas permite a organização, análise e a extração de conhecimento sobre grande volume de dados. Logo, considerando uma das tecnologias mais propícias, quando o contexto é buscar conhecimento em dados volumosos, podemos mencionar a Mineração de Dados (MD) ou Data Mining (DM), criada no final da década de 80, por profissionais de organizações, que dedicaram sua atenção a grandes volumes de dados armazenados, subutilizados ou ignorados pelos seus possuidores (FAYYAD, 1996).

O processo CRISP-DM ocorre em eventos distintos e reúne as melhores práticas para que o Data Mining seja o mais produtivo e eficiente possível, analisando dados financeiros, de recursos humanos, práticas dos clientes e outros, a fim de propor modelos de melhoria ou solução de problemas. O método CRISP-DM fornece as melhores práticas para gerenciar o trabalho de Big Data Analytics, facilitando a tomada de decisões e gerando uma enorme vantagem competitiva às organizações (PIATETSKY, 2014).

## 5 METODOLOGIA

A metodologia empregada nesse trabalho, detém perspectiva qualitativa, visto que de acordo com Prodanov e Freitas (2013), “acredita existir uma conexão dinâmica entre o universo tangível e o indivíduo, ou seja, uma relação intrínseco entre o universo objetivo e a subjetividade do indivíduo que não consegue ser compreendido em números” (PRODANOV; FREITAS, 2013, p. 70). Assim como também consiste, em uma pesquisa que explora definida investigação a desenvolver a essência de uma ligação entre as concepções estudadas – nessa situação, a concepção do prodígio big data, a nova área emergente ciência de dados, o profissional da informação e o bibliotecário de dados.

Ademais, analisa as habilidades e competências no qual o profissional da informação dispõe em conformidade com o cientista de dados, dando destaque para o bibliotecário de dados e as técnicas de tecnologia dentro da perspectiva da era big data. Gil (2002, p. 41) enfatiza que o escopo fundamental da pesquisa exploratória é “possibilitar maior vínculo com o problema, com intenção de torná-lo mais esclarecedor ou a estabelecer hipóteses”. O autor, da mesma forma, patenteia como sendo as técnicas de coleta de dados as mais usadas nesse categoria de metodologia, além da pesquisa bibliográfica e o estudo de caso. No concerne ao método escolhido, este trabalho aplica como método a pesquisa bibliográfica (GIL, 2002, p. 41).

A coleta dos dados da pesquisa deu-se mediante a uma pesquisa bibliográfica a respeito das expressões “ciência da informação”, “information science”, “big data”, “profissional da informação”, “information professional”, “ciência de dados”, “data science”, “bibliotecário de dados”, “cientista de dados” e “data scientist”, elaboradas junto ao Google Acadêmico e as bases de dados BDTD do IBICT, CAPES, Scielo, Lisa e Brapci, entre os dias 02/06/2019 e 20/04/2021.

Segundo Marconi e Lakatos (2003, p. 183) “a pesquisa bibliográfica, também publicada como fontes secundárias, inclina-se a explorar os estudos que já foram realizados referente a determinada área de estudo, englobando vários protótipos de publicações, tais como livros, revistas, monografias, jornais etc”. Ademais, Prodanov e Freitas (2013, p. 54) ressaltam “a necessidade da apuração dos dados que são conquistados mediante a pesquisa bibliográfica, contemplando as prováveis discordâncias e contradições expostas pelas obras verificadas”.

De acordo com Santa Anna (2017, p. 24), neste estudo foram examinados como fontes bibliográficas, artigos de periódicos e livros. No que lhe concerne, através da pesquisa bibliográfica, a escolha das matrizes usadas nesse trabalho foi elaborada com início na técnica

de leitura dinâmica, do título e do resumo dos quais, os materiais não relacionados com o tema foram descartados (ANNA, 2017).

Em seguida à pesquisa bibliográfica – para maior compreender da tônica – foi realizada pesquisa nos âmbitos da Ciência da Informação e da Biblioteconomia. Identificando no contexto do profissional da informação habilidades e competências que se ligam com o bibliotecário de dados. Essa análise foi realizada partindo do tema big data e suas vertentes, bem como a ciência de dados, suas tarefas e tecnologias.

Alicerçado ao referencial teórico constatado por meio da pesquisa bibliográfica – bem como da pesquisa de conceitos fundamentais – enfatizou-se a importância da ciência da informação no tratamento da informação e a evolução do perfil do profissional da informação no contexto do big data e da ciência de dados. Assim sendo, analisou-se um complexo de habilidades que qualifica o cientista de dados e o profissional da informação, do mesmo modo que, o bibliotecário de dados aborda algumas das habilidades do cientista de dados.

## **6 APRESENTAÇÃO E ANÁLISE**

Segundo Guinchat e Menou (1994, p. 30), o processo de tratamento da informação é chamado de “tratamento intelectual” e demanda do bibliotecário grande empenho mental, para compreender a abrangência do documento. De acordo com os autores, esse processo consiste na descrição bibliográfica, descrição de conteúdo, armazenamento, pesquisa e difusão, em que todos estes procedimentos visam atender as necessidades da comunidade usuária.

O tratamento da informação desenvolveu e permanece evoluindo, auxiliando as necessidades dos bibliotecários a fim de organizar o grande volume de informações acessíveis atualmente em vários suportes e propiciando a recuperação da informação dos usuários (GUINCHAT; MENO, 1994).

Todavia é primordial dar ênfase nos progressos que acontecem no tratamento da informação. Pois são em todo tempo conduzidas para compreender as alterações nas estruturas informacionais e nas distintas maneiras de viabilizar as informações. Dessa forma, as tecnologias da informação possibilitam que as informações sejam organizadas e tratadas de modo que facilite a recuperação da informação mais refinada, eficaz e completa, além de ser em tempo real. É capaz de se constatar que o bibliotecário desloca-se manifestando novas competências e habilidades, conduzindo as inovações e procurando gradativamente identificar o usuário (LEVACOV, 1999).

É inquestionável que uma nova dimensão da era da informação se manifesta no big data. Gerenciar esses dados de pesquisa manifesta-se como um novo item da comunicação científica que influencia pesquisadores e os âmbitos de publicação, indexação e avaliação. De outra parte, o volume de dados e as estatísticas disponíveis já estão maiores do que nunca e continuam a aumentar. Estes avanços no âmbito do conteúdo levaram à biblioteconomia e, mais especificamente, os bibliotecários a enriquecem suas funções e perfil profissional no tratamento para o armazenamento, preservação, disseminação e uso de fontes de dados (CARPENTER, 2019).

Os profissionais da informação são desafiados a trabalhar com a rapidez imposta tanto pela geração de informação e de dados, quanto pela variedade de fontes em que elas se encontram. Dessa forma, os profissionais da informação necessitam entender essa nova esfera de enorme fluxo de dados e como tem potencial para auxiliar numa gestão competente da informação.

Sayão e Sales (2012, p. 182) dedicados as amplas alterações que carregam com elas o conceito de novas funções para os profissionais que se ocupam com informação, explicam a respeito da manifestação das expressões cientista de dados e justifica “(...) embora não seja uma carreira de perfis bem delimitados e de reconhecimento explícito, a sua colaboração é essencial para uma comunicação bem-sucedida entre as partes envolvidas.”

A propagação de modelos de acervos digitais e a erupção informacional que existe atualmente, juntamente com a produção progressiva de conteúdo digital a todo instante por smartphones e diversos dispositivos eletrônicos conectados à internet produzem metadados em quantidade exorbitante (PJIPERS, 2010).

Essa quantidade acessível e sua relação inseparável com a rotina do ser humano, transforma os metadados a mais valiosa procedência para o estudo e investigação além da extração de informações referente a consumidores e seus hábitos, provocando o interesse de instituições de negócios que procuram alcançar vantagem competitiva por meio de ferramentas de análise de dados (REIS; SÁ, 2020).

Segundo Pjipers (2010), é exatamente a exorbitância de dados que se converte na maior dificuldade para a recuperação da informação fundamental, visto que é imprescindível perder demasiado tempo procurando essa informação, o que diminui o período a disposição para a análise e proveito da informação enquanto ela é valorosa. Isso é provocado pela desatualização tecnológica entre os recursos de produção e distribuição da informação e os métodos de organização e recuperação dela (PJIPERS, 2010).

Em suma, de acordo com Gondim (2012), “profissionais qualificados de trabalhar com big data precisam ser extremamente preparados, e essa preparação requer bastante investimento de tempo, ao mesmo tempo que a necessidade de implantação dos projetos é imediata”. Visto que apenas a capacitação de pessoal se torna inviável em uma linha de tempo curta e a oferta de profissionais das áreas de formação tradicionais para a área é reduzida (REIS; SÁ, 2020). Por conseguinte, uma solução pertinente para a implantação eficiente de um projeto de big data é a admissão de candidatos de origens acadêmicas diversas, tal como, os profissionais da ciência da informação, bibliotecários (GONDIM, 2012).

Segundo Davenport (2014), “a capacitação de um grupo de trabalho pluridisciplinar fazia-se essencial para a análise de dados, verificado a complexidade de um único profissional englobar todas as habilidades e competências fundamentais para o ofício”. Assim sendo, através de habilidades abrangentes almejadas de um cientista de dados esse vigente profissional tem de dominar a interpretação dados em sistemas autômatos e ser proativo na interpretação desses elementos, além de conhecer suas limitações e trabalhar em equipe (ARRUDA et al, 2000).

Tais habilidades são um desenvolvimento espontâneo das tarefas do profissional de informação evidenciadas por Wormell (2014, p. 34):

[...] facilitar o uso da informação; navegar por sistemas do conhecimento e fontes de informação; consultar e assessorar problemas de informação; gerir eficientemente os sistemas de informação; transformar os dados e o fluxo da informação entre sistemas; aliar os aspectos sociais e culturais; educar usuários; prover recursos para a alfabetização informativa; apoiar políticas de informação estratégias e de negócios (WORMELL apud CARMO, 2014, p. 34)

Gordon-Murnane (2012), associa claramente as habilidades do bibliotecário como capacidade para seu desempenho no ambiente de big data. Segundo Gordon, “o bibliotecário consegue contribuir e proporcionar a exploração e recuperação de dados, ser encarregado pela manutenção da qualidade dos dados, acrescentando valor aos dados por meio da catalogação e metadados, além de auxiliar na autenticação, gestão, arquivamento, representação e preservação de dados (GORDON-MURNANE, 2012).

No decorrer de seu desenvolvimento acadêmico, o bibliotecário instrui-se a trabalhar com dados de forma que o dado seja um auxiliador da mediação entre a necessidade do usuário

e as informações que mais se apropriam a ela, sendo assim um agente da propagação do conhecimento (CONEGLIAN; GONÇALVEZ, 2017). De certa forma o profissional de big data atua de maneira semelhante, sendo encarregado pela identificação dos insumos informacionais que podem ser extraídos de um montante de dados, sua recuperação e análise, e por último a exibição das informações extraídas no processo. Essa iminência de interpretação com dados, informação e conhecimento é início para a composição profissional das áreas (WORMELL apud CARMO, 2014).

Analisando as habilidades e competências das disciplinas dos diplomados em Biblioteconomia pode-se delinear semelhanças entre as atividades do bibliotecário e as funções de alguns ofícios de big data (CONEGLIAN; GONÇALVEZ, 2017):

- Desenvolver e utilizar tecnologias atuais;
- Criar produtos mediante os conhecimentos obtidos;
- Compartilhar e acrescentar qualidade no processo de criação, transferência e utilização da informação;
- Dedicar-se com matrizes de informação de toda origem.

Essas são somente algumas das tantas habilidades e competências em comum aos profissionais de ambas as áreas exatamente relacionadas à sua essências. Isto posto, o bibliotecário que procura em Big Data uma nova área de desempenho profissional já detêm, devido à sua formação, algumas habilidades e competências requeridas para Data Analysts, Data Scientists, Data Architects, entre outras funções (WORMELL apud CARMO, 2014).

Em síntese, big data é um âmbito de exercício relevante para bibliotecários que desejam trabalhar com análise de dados, contemplando o avanço das tecnologias da informação e o aumento das empresas, dessa forma o movimento almejado é que as ferramentas e técnicas se desenvolvam mais disponíveis facilitando dessa maneira uma grande inserção dos profissionais bibliotecários nessa esfera de atuação (SOUZA, 2006).

Conforme comentam Hey e Hey (2006) à medida que essas grandes quantidades de dados se tornam disponíveis, cientistas necessitarão do apoio dos profissionais da informação para gerenciar, pesquisar, preservar e armazenar esses novos recursos de dados. Compete, desta forma, aos cientistas da informação que se encontrem preparados para este novo cenário – da dimensão tecnológica da CI.

Segundo Semeler (2017), “a Ciência de Dados, atualmente é vista como a teoria e a prática de extrair conhecimento de dados se atentando com a criação de produtos e serviços a partir de dados”. Sendo um campo do saber que exige conhecimentos inerentes, em geral

referentes ao uso de linguagens de programação e metodologias científicas para coletar, analisar e visualizar dados. Considera-se que a ciência de dados consiga ser classificada como um dos primordiais campos de aprendizados para os bibliotecários de dados (BULINGAME; NIELSEN, 2012).

Posto isso, propõem-se que o bibliotecário de dados seja capaz de apropriar-se de determinadas competências e habilidades solicitadas aos cientistas de dados. Dessa forma, estes profissionais detêm do papel indispensável para a prosperidade da ciência de dados em unidades de informação, entretanto compete aos bibliotecários de dados empregar a ciência de desempenho carregado de dados mais de forma que ciência do que como somente uma ampliação de técnicas de manipulação e análise de dados (ZHU; XIONG, 2015; CHEN, 2015).

De acordo com Semeler (2017), “a Ciência de Dados proporciona novas habilidades, técnicas, tecnologias e métodos que são capazes de serem utilizados pelos bibliotecários de dados para que trabalhem com coleta, manipulação, análise e visualização de dados de pesquisa”. Desse modo, esta imposição defende a necessidade de se dominar o cenário internacional de repositórios de dados de pesquisa – seja para depositar dados, seja para buscar dados para realizar uma nova investigação (ZHU; XIONG, 2015; CHEN, 2015).

Isto posto, uma das indispensáveis atividades da Biblioteconomia de Dados é que ela não necessita somente se importar com informação difundida, todavia também com fontes potenciais de dados, sendo o seu propósito entender como os distintos tipos de dados conseguem formar informação proveitosa para pesquisa (XIA; WANG, 2014; KOLTAY, 2017).

Os dados de pesquisa criados pelas mais variadas ciências estão em evolução no que concerne aos conhecimentos a respeito da biblioteconomia de dados, sobretudo em aprendizagens que se importam com o ciclo de vida dos dados e com o Big Data. Diante dessa conjuntura, compreende-se o bibliotecário que, habitualmente, exerce seus ofícios fundamentado no armazenamento, na disseminação, na preservação e na gestão de acervos documentais (SEMELER; PINTO, 2019).

Necessita se envolver e empenhar-se com os conhecimentos e aprendizagens referente aos dados no âmbito da biblioteconomia de dados. Admite-se que este profissional seja capaz de empregar suas habilidades e competências habituais para proporcionar serviços e produtos de dados de pesquisa, conforme na conjuntura pertinente às tecnologias, como o Big Data ou aos repositórios digitais (KINDLING et.al., 2017).

Segundo a Gold (2007), “alinhado à noção, ao princípio e um pouco das particularidades dos estudos sobre dados e dados de pesquisa, considera a dificuldade da teoria e do conceito de

dados designa a carência de uma ciência que tenha conhecimento em habilidades, métodos, técnicas e tecnologias sobre o grande uso de dados”. Nessa perspectiva, sugere-se a Ciência de Dados como um conceito e atividade científica usada para atingir as várias dimensões do conceito de dados (GOLD, 2007).

Admita-se que a ciência de dados configura um novo estímulo de conhecimento e que se apresenta como uma inovação de oportunidade aos profissionais bibliotecários e cientistas da informação. Assim sendo, a ciência de dados possibilita técnicas atuais e atividades que possam colaborar para a investigação desses profissionais, além de ser considerada como área de estudos relacionada à gestão e à análise de dados (RICE; SOUTHALL, 2016; KELLAM; THOMPSON, 2016).

Neste contexto, a ciência de dados é uma conjunção de análise de dados e desenvolvimento de novos produtos e serviços com base nos dados. Combinando, assim, um conjunto de técnicas.

Wang delinea a concepção de ciência de dados expondo vários outras concepções desenvolvidos por diversos autores. Segundo Wang (2018), “parece que um consenso foi alcançado. Ciência de dados é um processo transversal que envolve reconhecimento e extração de padrões, e o conversão de dados em informação. Área de conhecimento. O conhecimento é obtido por meio de análise e mineração de dados” (WANG, 2018).

Logo, o projeto CRISP-DM (Processo Padrão da Indústria Cross para Mineração de Dados) propõe um modelo de processo abrangente para a execução de projetos de mineração de dados (CLIFTON; THURASINGHAM, 2001). À vista disso, a CRISP-DM é intensivamente utilizada em projetos de Ciência de Dados.

Segundo Reinartz e Wirth (1995), “o modelo de processo CRISP-DM para mineração de dados, visa tornar grandes projetos de mineração de dados, menos dispendiosos, mais confiáveis, mais repetível, mais gerenciável e mais rápido”. Além disso, fornece uma visão geral do ciclo de vida de um dado projeto de mineração. Ele contém as fases de um projeto, suas respectivas tarefas e seus resultados (REINARTZ; WIRTH, 1995).

Em virtude disso, o bibliotecário exerce e desempenha funções utilizando o processo CRISP-DM, no sentido que, o ofício de um bibliotecário de dados tem de estar relacionado à manipulação de dados, porém também se interliga com o desempenho de interpretação humana, tomada de decisão e comunicação de dados. No que tange o modelo de processo CRISP-DM, o profissional bibliotecário pode esboçar suas tarefas na compreensão do negócio, entendimento dos dados e preparação dos dados.

Na compreensão do negócio, define-se os objetivos do negócio identificando as necessidades do cliente, dessa forma o analista de mineração de dados verifica fatores relevantes que poderão influenciar os resultados. O profissional da informação desempenha papel semelhante, visto que, auxilia na pesquisa do usuário e no estudo dele, mapeando áreas de interesse e materiais mais consultados para oferecer ao leitor uma resposta ágil e veloz e propor publicações mantendo sua frequência na biblioteca.

No entendimento dos dados, discerne-se a informação que pode ser valorosa para o estudo e uma prévia clareza e percepção com o seu conteúdo, descrição, qualidade e utilidade. Além de ter como tarefa alcançar a informação com a qual se irá trabalhar, listando suas fontes, o processo de leitura e os impasses preliminares identificados. A descrição dos dados registra a forma como foram adquiridos, listando o seu formato, volume, significado e toda a informação relevante. Similarmente o profissional da informação está envolvido nessa atmosfera, trabalhando na humanização dos dados, seja na qualidade, autenticidade ou autoridade. Dispõe de habilidades para a gestão dos dados captados e sua organização (PEREIRA, 2013). Este profissional possui competência para a curadoria, isto é, para gerir os dados desde sua formação até sua utilização, seja no presente ou no futuro, patrocinando sua reprodutibilidade, reutilização e agregando valor a eles (HARVEY, 2010 apud COSTA; CUNHA, 2014).

Na preparação de dados produz-se o conjunto de dados definitivos através dos dados brutos iniciais, o profissional da informação é capaz de desempenhar uma atuação relevante no aproveitamento do poder dessa área emergente. Há na atualidade uma escassez de pessoas que são capazes de captar dados brutos e transformá-los em conhecimento. O profissional da informação pode ser extremamente útil e transformador para o âmbito da big data e ciência de dados, pois os bibliotecários possuem experiências em gerenciamento e organização de dados. Além de desempenharem um grande papel, na descoberta, compreensão e limpeza de dados para uso na análise de big data.

Assim sendo, os profissionais da informação precisam disponibilizar recursos para ajudar a impulsionar a criação de novos conhecimentos. A relação entre a biblioteca e a ciência de dados e big data é clara, visto que vem da capacidade do bibliotecário de trabalhar com o conhecimento e auxiliar os usuários da biblioteca a obter recursos para lidar com o novo conhecimento para obter percepções acionáveis.

Em suma, a ciência de dados tem potencial de ser utilizada não somente pelos usuários da biblioteca, entretanto, também nos processos internos da biblioteca. Os insights de negócios obtidos de big data podem ser utilizados na própria biblioteca. Por exemplo, os bibliotecários conseguem examinar a ciência de dados e a análise de big data na verificação de livros da

biblioteca e usar registros para determinar quais livros devem ser adicionados às coleções da biblioteca. Logo, o big data é capaz de tornar as decisões internas da biblioteca – como compras de livros e outras decisões de coleções – mais simplificadas, eficientes e melhores (COSTA, 2010).

Segundo Semeler e Pinto (2019), o bibliotecário de dados precisa pesquisar os sistemas e as normas que circunda a busca com dados de pesquisa em repositórios digitais, como um profissional que tem de a procurar domínio sobre coleta, manipulação, análise e visualização de dados de pesquisa para ofertar serviços e produtos relacionados ao gerenciamento e curadoria de dados de pesquisa (SEMELER; PINTO, 2019).

Perante a óptica de Cunha (2010); Costa, Cunha (2014), inclui-se que as atuais transformações na maneira de dirigir a pesquisa científica atribuindo oportunidades para o bibliotecário se envolver com o tratamento de dados. Em vista disso, a tarefa do bibliotecário é a gestão de dados – esta função, já discutida por Costa, Cunha (2014), está diretamente relacionada com a extração de conhecimento-informação de grandes volumes de dados de investigação científica, com o objetivo de identificar relações associativas e correlações existentes em conjuntos de dados que compõem os diversos sistemas onde a informação científica circula.

Em síntese, é necessária uma dedicação interdisciplinar de conhecimento, em união dos bibliotecários de dados, que precisam planejar as habilidades e os conhecimentos fundamentais à manipulação e ao gerenciamento de dados. Dessa forma, a função essencial dos bibliotecários de dados é procurar obter experiência técnica para aquisição, análise e interpretação de dados. Logo, o bibliotecário de dados precisa mudar os dados para um dado proveitoso, transmissível e que se consiga guardá-lo em no decorrer do tempo. Além de dominar as estruturas de metadados, bem como os repositórios de dados, os conjuntos de dados o quais são usufruídos por seus usuários pesquisadores (MCCAFFREY; GIESBRECHT, 2016; KENNAN, 2016; DEKKER, LACKIE, 2016).

## **7 CONSIDERAÇÕES FINAIS**

Através da pesquisa realizada é possível compreender o avanço Web, da Internet e das Tecnologias de Informação na forma como tem produzido uma enorme extensão de dados e informações que necessitam ser cuidadas, criando dessa maneira uma verdade desafiadora e

referente à manipulação, armazenamento e processamento desses dados, principalmente para os Profissionais da Informação.

Logo, a explosão informacional e a evolução das tecnologias permitiram um avanço na sociedade, concedendo evoluções científicas e mercantil que revolucionaram a forma como o ser humano se comunica e interage com o mundo. Naquele momento, surgia a Ciência da Informação, essa ciência interdisciplinar com a propósito de descobrir soluções ao grande volume informacional então presente.

Esse maior volume de informações produzidos pela humanidade da explosão informacional refere-se ao fenômeno big data, um complexo de aptidões tecnológicas que possibilita uma vigente perspectiva para a análise e compreensão de grandes uniões de dados. Conjunto de dados cujo tamanho está além da habilidade de ferramentas típicas de banco de dados em capturar, gerenciar e analisar (MANYIKA, 2011).

A revisão de literatura acerca do tema – no propósito da Biblioteconomia – reconheceu um crescente interesse a respeito do comportamento do bibliotecário na era do big data, lidando com dados e informação. Com o uso intensivo de dados, crê-se que constitui um grande desafio de aprender e se descrever como uma vigente chance aos bibliotecários. Contudo eles precisam adquirir maiores conhecimentos sobre os dados de pesquisa e as tecnologias e ferramentas utilizadas para sua manipulação e gestão, como MAP-REDUCE, NOSQL, Mineração de Dados e CRISP-DM, pois essas tecnologias são de suma importância para o bibliotecário de dados desempenhar e exercer sua função nos ambientes de big data e ciência de dados.

Ao longo do processo de formação acadêmica, o bibliotecário aprendeu a processar dados, tornando-se assim um mediador entre as necessidades do usuário e as informações mais adequadas para ele. De muitas maneiras, os profissionais de big data se comportam de maneira semelhante e são responsáveis por identificar entradas de informações que podem ser extraídas de grandes quantidades de dados, recuperando-os e analisando-os e encerrando a exibição das informações extraídas no processo. A estreita ligação com dados, informação e conhecimento é o ponto de partida para a integração profissional regional (BOYD; CRAWFORD, 2011).

Ao observar as habilidades e competências dos profissionais da informação, podem ser encontradas semelhanças entre o desempenho dos bibliotecários e o desempenho de várias posições de big data. Por exemplo, desenvolver e usar novas tecnologias; produzir produtos por meio do conhecimento adquirido; interagir e agregar valor no processo de geração, transmissão e uso da informação; trabalhar com qualquer tipo de fonte de informação é apenas uma questão de caráter direto nesses dois campos. das muitas habilidades compartilhadas por profissionais

relacionados. Portanto, big data é uma área de atividade interessante para bibliotecários que desejam se envolver na análise de dados (BOYD; CRAWFORD, 2011).

Dessarte, através dos pontos expostos neste trabalho, é possível identificar em quais processos esse profissional está capacitado para atuar. Assim sendo, as discussões sobre a necessidade do profissional da informação neste novo campo de trabalho podem se aprofundar, visto que há uma proposta viável a ser implementada, que se inter-relaciona com as principais contribuições teóricas e práticas existentes nas pesquisas de Big Data e Ciência de Dados.

Enfim, uma ciência fundamentada na exploração dos dados de pesquisa é capaz de ser refletida como um recente campo de estudos para bibliotecários que buscam se cercar com os problemas relacionados à coleta, manipulação, análise e visualização de dados de pesquisa. Essa corrente investigativa levantará indagações referente às práticas, modelos e metodologias usadas para se alcançar a informação proveitosa através dos dados, insights de aprendizagem mediante ambientes diferentes e complexos de dados.

Dessa forma, os direcionamentos vigentes de pesquisa em biblioteconomia de dados estão inclinados para os ensinamentos sobre a origem, as melhores práticas, o planejamento, a curadoria, dentre diversas atividades científicas referentes à teoria e à práxis com dados, essencialmente quando esses dados se ligam com os métodos de pesquisa científica em ambiente digital. Sobretudo, a área da biblioteconomia de dados está alusiva à prática, com a utilização e o consumo de dados de pesquisa.

Dessa forma, com a inserção das tecnologias de dados nas práticas bibliotecárias, compreende-se uma conversão na maneira que o bibliotecário enxerga suas próprias tarefas. As recentes alterações na maneira de se dirigir a pesquisa científica em CI disponibiliza possibilidades para os bibliotecários de dados se atrair com a ciência de uso intensivo de dados, esses profissionais necessitam especificamente obter habilidades e competências em Ciência de Dados.

Por fim, é necessário um esforço multidisciplinar de aprendizagem, por parte dos bibliotecários de dados, que devem combinar as habilidades e os conhecimentos necessários à Ciência da Informação, à Big Data e à Ciência de Dados.

## REFERÊNCIAS

ALVARENGA, Lídia. *Representação do conhecimento na perspectiva da Ciência da informação em tempo e espaços digitais*. Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. 15, 2003. Disponível em: Acesso em: 12 jan. 2020.

ANGELONI, M. T. *Technologies de transmissiun de lyinformation et consequences sur le processus des decisions des organizations*. Grenoble, France: Mémoire de DEA. ...cole Superieure des Affaires, 1992.

ASHTON, K. (2009). *That 'internet of things' thing*. RfID Journal, 22(7):97–114.

A estima de 2,5quintilhões de bytes foi retirada do site da IBM, da seção *What Is Big Data? Bringing Big Data to the Enterprise*, disponível em: [www.ibm.com](http://www.ibm.com).

AMO, Sandra de. Técnicas de Mineração de Dados. *Data Mining*, [S. l.], p. 1-43, 20 out. 2016. Disponível em: <https://www.periodicos.capes.gov.br/>. Acesso em: 7 maio 2020.

BARBOSA, Alice Príncipe. *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969.

BERAQUET, V.M.M. et al. *Desenvolvimento do profissional da informação para atuar em saúde: identificação de competências*. Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, v.3, n.2, p.1-16, jan./jun. 2006.

BUGEMBE, M. *Finding Value in Data: Determining Where Data Science has The Greatest Impact*. O'Reilly: Sebastopol, 2016.

BURTON M; LYON L (2017) *Data science in libraries. Bulletin of the Association for Information Science and Technology* 43: 33–35. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/bul2.2017.1720430409/abstract> (Acesso em Janeiro 2020).

BORKO. H. *Information Science: what is it?* American Documentation, v.19, n.1, p.3-5, Jan. 1968.

BROOKES, Bertram. C. *The foundations of Information Science. Parti: Philosophical aspects.* Journal of Information Science, v.2, p. 125-133, 1980.

BRAGA, G.M. Prefácio. In: PINHEIRO, L.V.R. (Org.). *Ciência da informação, ciências sociais e interdisciplinaridade.* Brasília: IBICT, 1999.

BUSH, Vannevar. *As we may think.* Atlantic Monthly, v.176, n.1, p.101-108, July 1945

BREWSTER, Tom. *When machines take over: our hyperconnected world.* BBC, 25 jan. 2014.

BRUNO. Aprendendo a programar em Python -Introdução. 2010. Disponível em: <<https://www.devmedia.com.br/aprendendo-a-programar-em-python-introducao/17093>>. Acesso em: 07 março 2020.

BUCKLAND, M. K. *Information as thing.* Journal of American Society for Information Science. n. 42, v.5, p. 351-360, 1991.

BOYD, Danah; CRAWFORD, Kate, *Critical question for big data: provocations for a cultural, technological, and scholarly phenomenon.* Information, Communication & Society, v. 15, n. 5, p. 662-679, 2012. Disponível em: <https://doi.org/10.1080/1369118X.2012.678878>. Acesso em: 21 de jan. 2020.

BERRY MICHAEL J. A.; LINOFF GORDON, “*Data Mining Techiques for Markenting, Sales, and Customer Support*”; JOHN WILEY&SONS. Inc., 1997.

CAVALCANTI, José Carlos. *The essential trinity in high-tech industries: ecosystem + platform + architecture.* VIII Research Workshop on Institutions and Organizations. RWIO

Center for Organization Studies — CORS. São Paulo, 2013. \_\_\_\_\_. Effects of IT on enterprise architecture, governance and grow. IGI- -Global, 2015. Disponível em: . Acesso em: 30 jan. 2020.

CASTELLS, M. *A sociedade em rede*. 4. ed. São Paulo: Paz e Terra, 2000. 617p.

CAPURRO, R.; HJORLAND, B. *O conceito de informação. Perspectivas em Ciência da Informação*, Belo Horizonte, v.12, n.1, p.148-207, abr. 2007. Disponível em: . Acesso em: 17 jan. 2020.

CAPURRO, Rafael; HJORLAND, Birger. *The Concept of Information. Annual Review of Information Science and Technology*. Ed. B. Cronin, v. 37 Chapter 8, p. 343-411.2003

CAVALLI, Olga. *Internet das coisas e inovação na América Latina*. [S.l.: s.n.], 2016. Mimeogr.

CAO, L. *Data science and analytics: a new era*. [S.l.]: Springer, 2016.

CAO, L. *Data science: Nature and pitfalls*. IEEE Intelligent Systems, IEEE, v. 31, n. 5, p. 66–75, 2016.

CAO, L. *Data science: a comprehensive overview*. ACM Computing Surveys (CSUR), ACM, v. 50, n. 3, p. 43, 2017.

CAO, L. *Data science: challenges and directions*. *Communications of the ACM*, ACM, v. 60, n. 8, p. 59–68, 2017.

CONWAY, D (2010) Venn Diagram. Available at: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram> (Acesso em Janeiro 2020)

CHEN, H., & et al. *Business Intelligence and Analytics: from Big Data to Big Impact*. *Mis Quartely*, 36(4), pp. 1165-1188. 2012.: Acesso em: 8 jan. 2020.

CHAI, Sen; SHIH, Willy. *Why big data isn't enough*. *MIT Sloan Management Review*, v. 58, n. 2, p. 57-61, 2017. Disponível em: <https://sloanreview.mit.edu/article/why-big-data-isntenough/>. Acesso em: 21 de Jan. 2020.

COSTA, M.; CUNHA, M. B. da. *O bibliotecário no tratamento de dados oriundos da e-Science: considerações iniciais*. *Perspectivas em Ciência da Informação*, v. 19, n. 3, p. 189-206, 2014.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Orientador: João Carlos da Silva. 2009. Mestrado (Ciência da Computação - INF/UFG) - Mestrando, [S. l.], 2009. Disponível em: [Mineração%20de%20dados/RT-INF\\_001-09.pdf](#). Acesso em: 7 maio 2020.

CIOS, K. J; PEDRYCZ, W; SWINIARSKI, R. W; KURGAN, L. A. *Data Mining - A Knowledge Discovery Approach*. Springer, 2007.

CABENA, P; HADJINIAN, P; STADLER, R; JAAPVERHEES; ZANASI, A. *Discovering Data Mining: From Concept to Implementation*. Prentice Hall, 1998.

CHAPMAN, P. et al, 2000. *CRISP-DM 1.0 - Step-by-step data mining guide*.

DIAS, E. W. *Biblioteconomia e ciência da informação. Perspectivas em ciência da informação*. Belo Horizonte, v. 5, n. especial, p. 60-80, jan./ jun. 2000.

DAVENPORT, T. H. *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business Review Press, 2014.

DAVENPORT, Thomas H. *Recorded Future: Analyzing Internet Ideas About What Comes Next*. Case 613-083. Boston: Harvard Business School, 2013.

DAVENPORT, Thomas H. e KIM, Jinho. *Dados demais!*. Rio de Janeiro: Elsevier, 2013. *Documentation, Information Science and Library Science in the USA*. *Information Processing & Management*, Elmsford (NY), v. 32, p. 63-76, 1996.

DAVENPORT, Thomas H. e KIM, Jinho. *Dados demais!*. Rio de Janeiro: Elsevier, 2013.

\_\_\_\_\_, T. H. *Big Data at work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press, 2014.

\_\_\_\_\_, *Estudo sobre cientista de dados* (patrocinado pela SAS e EMC): DAVENPORT, Thomas H. *The Human Side of Big Data and High-Performance Analytics*.

DEAN, J.; SANJAY, G. *MapReduce: simplified data processing on large clusters*. *Communications of the ACM*, v. 51, n. 1, p. 107-113, 2008. Disponível em: . Acesso em: 19 jan. 2020

DEKKER, H.; LACKIE, P. *Technical Data Skills for Reproducible Research*. In: KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.

EIICA (2019). *X Encontro Internacional de Informação, Conhecimento e Ação*. Marília, 2018. <http://enancib.marilia.unesp.br/index.php/EIICA/XEIICA>. (2019-02-27).

ELMASRI RAMEZ; NAVATHE SHAMKANT, “*Fundamentals of Database Systems*”; Addison-Wesley, 1999. 3rd Edition.

FOSKETT, D. J. *Informática*. In: GOMES, Hagar Espanha, org. *Ciência da Informação ou Informática?* Rio de Janeiro: Calunga, 1980. p.9-51. Artigo originalmente publicado no *Journal of Documentation*, v.26, n.4, p.340-69, Dec. 1970

FLORIDI, Luciano. *Biblioteconomia e Ciência da Informação (BCI) como Filosofia da Informação Aplicada: uma reavaliação*. In: *Revista de Ciência da Informação e Documentação*, Ribeirão Preto. v. 1, n. 2, p. 3747, jul./dez. 2010.

FLORIDI, Luciano. *Information: a very short introduction*. London: Oxford University Press, 2010

FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, 1996.

GANDOMI, Amir; HAIDER, Murtaza. *Beyond the hype: big data concepts, methods, and analytics*. *International Journal of Information Management*, [S.l.], v. 35, n. 2, p. 137-144, Apr. 2015. Disponível em: . Acesso em: 12 jan. 2020.

GOMES, H.F. *Interdisciplinaridade e ciência da informação: de característica a critério delineador de seu núcleo principal*. *DataGramaZero: Revista de Ciência da Informação*, v.24, n.4, 2001. Disponível em:. Acesso em: 16 jan. 2020.

GOMES, Hagar E. Apresentação In: *Ciência da Informação ou Informática?* Rio de Janeiro: Calunga, 1980. p.7-8.

GANDOMI, Amir; HAIDER, Murtaza. *Beyond the hype: big data concepts, methods, and analytics*. *International Journal of Information Management*, [S.l.], v. 35, n. 2, p. 137-144, Apr. 2015. Disponível em: . Acesso em: 12 abr. 2018.

GANTZ, John e REISEL, David. *Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. *International Data Corporation Digital Universe*, 1º dez. 2012.

GANTZ, John e REISEL, David. *Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. *International Data Corporation Digital Universe*, 1º dez. 2012.

IBM CORPORATION. *What is Hadoop*. Disponível em: <http://www01.ibm.com/software/data/infosphere/hadoop/>.

GASTÃO, C. *Estatística básica – arte de trabalhar com dados*. Rio de Janeiro: Elsevier, 2009.

GOLDMAN, Alfredo; KON, Fabio; JUNIOR, Francisco Pereira; POLATO, Ivanilton; PEREIRA, Rosângela de Fátima. *Apache Hadoop: conceitos teóricos e práticos, evolução e novas possibilidades*. *Apache Hadoop*, [S. l.], p. 1/49, 3 jul. 2012. DOI 8541. Disponível em: <https://www.ime.usp.br/~ipolato/JAI2012-Hadoop.pdf>. Acesso em: 21 jan. 2020.

GOLD, A. *Cyberinfrastructure, data, and libraries*, part 1: A cyberinfrastructure primer for librarians. D-Lib Magazine, v. 13, n. 9-10, 2007a. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html> .Acesso em: 31 janeiro de 2020.

HEILPRIN, Laurence B. *Foundations of information science reexamined*. Annual Review of Information Science and Technology (Arist),v. 24, p. 343-372, 1989.

HENDERSON, M. *Data Management: a practical guide for librarians*. Lanham: Rowman & Littlefield, 2017.

HAYASHI, C. *What is Data Science? Fundamental Concepts and a Heuristic Example*. In: CONFERENCE OF THE INTERNATIONAL FEDERATION OF CLASSIFICATION SOCIETIES (IFCS'96), 5th, 1996, Kobe, Japan. Proceedings. Kobe, Japan: IFCS'96, 1996.

HAN, J; KAMBER, M. *Data Mining: Concepts and Techniques*. Elsevier, 2006.

HAND, D; MANNILA, H; SMYTH, P. *Principles of Data Mining*. MIT Press, 2001.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. *Riding the Waves or Caught in the Ride? Navigating the Evolving Information Environment*. 2016. Disponível em: Acesso em: 28 jan. 2020

IBID. Ver também: MOLARO, Cristian. *Do not ignore structured data in big data analytics: the important role of structured data when gleaning information from big data*. IBM Big Data & Analytics Hub, 19 jul. 2013.

A estima de 2,5quintilhões de bytes foi retirada do site da IBM, da seção What Is Big Data? Bringing Big Data to the Enterprise, disponível em: [www.ibm.com](http://www.ibm.com).

IBM CORPORATION. *What is Hadoop*. Disponível em: <http://www-01.ibm.com/software/data/infosphere/hadoop/>.

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICES AND TECHNOLOGY (IASSIST). *Defining data librarian - call for comments*, 2016.

Disponível em: <http://www.iassistdata.org/blog/defining-data-librarian-callcomments>. Acesso em: Janeiro. 2020.

INDRAWAN-SANTIAGO, M. *Database research: are we at a crossroad?* Reflection on NoSQL. In: FIFTEENTH INTERNATIONAL CONFERENCE ON NETWORK-BASED INFORMATION SYSTEMS, p. 45-51, 2012.

JAPIASSU, Hilton. *Interdisciplinaridade e patologia do saber*. Rio de Janeiro: Imago, 1976. 221 p. (Série Logoteca).

\_\_\_\_\_. *Introdução ao pensamento epistemológico*. 2. ed. Rio de Janeiro: Francisco Alves, 1977. p. 202.

KRAMER, S. N. (1963). *The Sumerians: Their History, Culture and Character*. Chicago: University of Chicago Press.

KOBASHI, N. Y.; SMIT, Y. W.; TÁLAMO, M. F. G. M. *A função da terminologia na construção do objeto da ciência da informação*. Data grama zero - revista de Ciência da Informação, v. 2, n. 2, abr. 2001. Disponível em:. Acesso em: 16 jan. 2020.

KOLTAY, T. (2015) *Data literacy for researchers and data librarians*. *Journal of Librarianship and Information Science*. Available at: <http://lis.sagepub.com/content/early/> (Acesso: 23 Janeiro 2020).

KOLTAY, T. *Data literacy for researchers and data librarians*. *Journal of Librarianship and Information Science*, v. 49, n. 1, 2017. Disponível em: <http://journals.sagepub.com/doi/abs/10.1177/0961000615616450>. Acesso em: 4 JAN. 2020.

KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.

KENNAN, M. Data management: Knowledge and skills required in research, scientific and technical organizations. In: IFLA WLIC, 82., 2016, Columbus. *Proceedings[...]*. Columbus, 2016. Disponível em: <http://library.ifla.org/1466/1/221-kennan-en.pdf>. Acesso em: jan. 2020.

LE COADIC, Y.-F. *A ciência da informação*. Brasília: Briquet de Lemos, 2004.

LE COADIC, Y. F. *A ciência da informação*. Brasília: Briquet de Lemos Livros, 1996. 119p.

LOUKIDES, Mike. *Whatis data science?* Sebastopol, CA: O'Reilly Media, 2011.

LANEY, D. *3D Data management: controlling data volume, velocityandvariety*. Application Delivery Strategies, Stanford. 2001. Disponível em: . Acesso em: 20 jan. 2020.

LAMMEL, Ralf. *Google's MapReduce Programming Model - Revisited*. Science of Computer Programming, v.70, n. 1, p. 1-30, jan. 2008. Disponível em: <[http://web.cs.wpi.edu/~cs3013/a11/Papers/Lammel\\_MapReduce\\_Revisited.pdf](http://web.cs.wpi.edu/~cs3013/a11/Papers/Lammel_MapReduce_Revisited.pdf)>. Acesso em: 23 Jan 2020.

LAROSE, D. T. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley and Sons, Inc, 2005.

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. *Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana*. Tradução de Paulo Polzonoff Junior. Rio de Janeiro: Elsevier, 2013.

MANTRA. *Do-It-Yourself Research Data Management Training Kit for Librarians*. 2012-13. Disponível em: <http://mantra.edina.ac.uk/libtraining.html>. Acesso em: JANEIRO. 2020.

MEIRA, Silvio. *Sinais do futuro imediato, #1: internet das coisas*. Ikewai, Recife, dez. 2016.

MIRANDA, Antonio. *A Ciência da Informação e a teoria do conhecimento objetivo: um relacionamento necessário*. In: AQUINO, Miriam de Albuquerque (Org.). *O campo da Ciência da Informação: gênese, conexões e especificidades*. João Pessoa: Universitária/UFPB, 2002. p. 9-24.

MIRANDA, Antonio. *Planejamento bibliotecário no Brasil; a informação para o desenvolvimento*. Rio de Janeiro, LTC Editora S.A., Brasília, Editora da Universidade de Brasília, 1977. 135p.

MOOERS, O. N. apud SARACEVIC, Tefko. *Information science: origem, evolution and relations* opus cit.

MOTA, Francisca Rosaline Leite; OLIVEIRA, Marlene de. Formação e atuação. In: OLIVEIRA, Marlene (Org.). *Ciência da informação e biblioteconomia: novos conteúdos e espaços de atuação*. Belo Horizonte: Editora UFMG, 2005. p. 99-110.

MCAFEE, Andrew; BRYNJOLFSSON, Erik. *Big data: the management revolution*. Harvard Business Review, [S.l.], v. 90, n. 10, p. 60-68, Oct. 2012. Disponível em: . Acesso em: 12 jan. 2020.

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

MCAFEE, A.; BRYNJOLFSSON, E.. *Big Data: the management revolution*. Harvard Business Review, Brighton, v. 90, n. 10, p. 61-67, oct. 2012. Disponível em: . Acesso em: 21 jan. 2020

MCCAFFREY, M.; GIESBRECHT, W. Teaching Data Librarianship to LIS Students. In: KELLAM, L.; THOMPSON, K. *Introduction to Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: Association of College and Research Library, 2016.

Master in Information Science from the University of Brasília, Brazil Brazilian Institute of Information in Science and Technology, Brazil. ORCID <https://orcid.org/0000-0001-5815-2574>.

\_\_\_\_\_. NEVES, I. B. das. V Encontro Nacional de Ensino de Biblioteconomia e Ciência da Informação (São Carlos, SP, 6 a 8 de agosto de 1998): relatório técnico. Porto Alegre : ABEBD, 1998.

National Science Board (NSB) (2005) Cyberinfrastructure Vision for 21st Century Discovery. Available at: <http://www.nsf.gov/pubs/2007/nsf0728> (Acesso: 23 Janeiro 2020)

NATIONAL SCIENCE BOARD (NSB). *Cyberinfrastructure vision for 21st century discovery*, 2005. Disponível em: <http://www.nsf.gov/pubs/2007/nsf0728> . Acesso em: JAN. 2020.

NATIONAL SCIENCE BOARD (NSB). *Digital Research Data Sharing and Management*, 2011. Disponível em: <https://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>. Acesso em: JAN. 2020.

NIELSEN; BURLINGAM N. (2012) *A Simple Introduction to Data Science*. Wickford: New Street Communications.

ORTEGA, C. D. *Relações históricas entre Biblioteconomia, Documentação e Ciência da Informação*. Data Grama Zero – Revista de Ciência da Informação, v.5, n.5, out., 2004. Disponível em: . Acesso em: 18 jan. 2020.

OLIVEIRA, Marlene de. *Origens e evolução da Ciência da Informação*. In: OLIVEIRA, Marlene de (Org.). *Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação*. 2. ed. Belo Horizonte: ed. UFMG, 2011

OED Online (OED) (2015) datum, n. Oxford University Press. Available at: <http://www.oed.com/view/Entry/47434> (Acesso 24 Janeiro 2020).

OED Online. *Datum, n. Oxford University Press, 2015*. Disponível em: <http://www.oed.com/view/Entry/47434>. Acesso em: jan. 2020.

PEPPET, Scott R. *Regulating the internet of things: first steps toward managing discrimination, privacy, security, and consent*. Texas Law Review, v. 93, p. 117-120, 2014.

PERKEL, J. M. *Programming: pick up python*. Nature News, v. 518, n. 7537, p. 125, 2015.

PSF, P. S. F. *About the Python Software Foundation* | Python Software Foundation. 2019. (Acesso em: 07/03/2020). Disponível em: <<https://www.python.org/psf/about/#how-do-i-reach-the-psf>>

PINHEIRO, Lena Vania Ribeiro, LOUREIRO, José Mauro Mattheus. *Traçados e limites da Ciência da Informação*. Ciência da Informação, Brasília, v.24, n .1, p.42-53, jan./jul,1995.

PINHEIRO, L. V. R. *Processo evolutivo e tendências contemporâneas da ciência da informação*. Informação e sociedade: estudos, João Pessoa, v. 15, n. 1, p. 13-48, jan./jun. 2005. Disponível em:. Acesso em: 16 jan. 2020.

PINHEIRO, L.V.R. *A ciência da informação entre luz e sombra: domínio epistemológico e campo interdisciplinar*. Rio de Janeiro. 278f. Tese (Doutorado em Comunicação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1997.

PINHEIRO, L.V.R. *Campo interdisciplinar em ciência da informação: fronteiras remotas e recentes*. Investigación Bibliotecológica, v.12, n.25, p.132-163, 1998.

PINHEIRO, L.V.R. *Campo interdisciplinar da ciência da informação: fronteiras remotas e recentes*. In: PINHEIRO, L.V.R. (Org.) *Ciência da informação, ciências sociais e interdisciplinaridade*. Brasília: IBICT, 1999. p.155-182.

PINHEIRO, L.V.R. *Ciência da informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade*. In: GONZÁLEZ DE GÓMEZ, M.N.; ORRICO, E.G.D. (Org.) *Políticas de memória e informação*. Natal: EDURFN, 2006a.

PINHEIRO, L.V.R. *Movimentos interdisciplinares e rede conceitual na ciência da informação*. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 7., 2006. Marília. Anais... Marília: Unesp, 2006b

PINHEIRO, L.V.R.; LOUREIRO, J.M.M. *Traçados e limites da ciência da informação*. *Ciência da Informação*, v.24, n.1, p.42-53, 1995.

PINHEIRO, Lena Vania Ribeiro. *A Ciência da Informação entre sobra e luz: domínio epistemológico e campo interdisciplinar*. Rio de Janeiro, UFRJ/ECO, 1997. 280p. Tese (Doutorado em Comunicação e Cultura). Orientadora: Gilda Braga

PWC, Indústria 4.0: *Digitização como vantagem competitiva no Brasil*. Pesquisa Global indústria 4.0: Relatório Brasil, p. 1-38, 2016. Disponível em: <<https://www.pwc.com.br/pt/publicacoes/servicos/assets/consultoria-negocios/2016/pwcindustry-4-survey-16.pdf>> Acesso em 21 de jan. 2020.

PIATETSKY, Gregory. "CRISP-DM, still the top methodology for analytics, data mining, or data science projects." KDD News (2014).

RANGANATHAN, S. R. *As cinco leis da Biblioteconomia*. Briquet de Lemos: Brasília, DF, 2009.

RIBEIRO, C. J. S. *Big Data: os novos desafios para o profissional da informação*. *Informação & Tecnologia*, João Pessoa, v. 1, n. 1, p. 96-105, jan. 2014.

RIJMENAM, Mark van. *Why the 3 V's are not suficiente to describe big data*. Datafloq, ago. 2015.

RUSSOM, P. *Big data analytics*. TDWI Best Practices Report, Fourth Quarter, 2011. p. 1-35

RICE, R.; SOUTHALL, S. *The data librarian's handbook*. London: Facet Publishing, 2016.

SARACEVIC, Tefko. *Information Science: origin, evolution and relations*. In:

VAKKARI, Pertti, CRONIN, B. laise, ed. *Conceptions of Library and Information Science: historical, empirical and theoretical perspectives*. Proceedings of the International Conference held for the celebration of 20th Anniversary of the Department of Information Studies. University of Tampere, Finlândia, 26-28, August 1991. London, Los Angeles: Taylor Graham, 1992. p.5-27

SARACEVIC, T. *Ciência da informação: origens, evolução e relações*. *Perspectivas em ciência da informação*. Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SARACEVIC, T. *Information Science*. In: BATES, M. J. ; Maack, M. N. (Eds.) *Encyclopedia of Library and Information Science*. New York: Taylor & Francis, 2010. p. 2570-2586. Disponível em: . Acesso em: 14 jan. 2020.

SARACEVIC, T. *Information Science; origin, evolution and relations*. In: Vakkari, P. & Cronin, B. *Conceptions of Library and Information Science; proceedings of the COLIS conference, Tampere, 1991*. Los Angeles, Taylor Graham, 1992.

SIQUEIRA, Jéssica Camara. *Biblioteconomia, documentação e ciência da informação: história, sociedade, tecnologia e pós-modernidade*. *Perspectivas em Ciência da Informação*. Belo Horizonte, v.15, n. 3, p. 52-66, set./dez.2010.

SILVA, A.; RIBEIRO, F. *Information Science and Philosophy of Information: Approaches and Differences*. In: DEMIR, H. (Ed.). *Luciano Floridi's Philosophy of Technology: Critical Reflections.*, Heidelberg: Springer Science+Business Media Dordrecht, 2012. (Philosophy of Engineering and Technology 8).

SILVA, E. M. *Descoberta de conhecimento com o uso de Text Mining: Cruzando o abismo de moore*. Master's thesis, Universidade Católica de Brasília, Mestrado em Gestão do Conhecimento e da Tecnologia da Informação, 2002.

SCHMIDT, E. *Every 2 days we create as much information as we did up to 2003*. TechCrunch, 2010. Disponível em: <<http://techcrunch.com/2010/08/04/schmidt-data/>>. Acesso em: 30 jan. 2020.

SHERA, J. H.; CLEVELAND, D. B. *History and foundations of Information Science*. *Annual Review of Information Science and Technology*, v. 12, p. 248-275, 1977 apud BRAGA, Gilda Maria. *Informação, ciência da informação: breves reflexões em três tempos*. *Ciência da Informação*, Brasília, v. 1, n. 24, p.1-8, abr. 1995. Disponível em: . Acesso em: 04 jan. 2020.

SOUTO, S. M. de O. *O profissional da informação frente às tecnologias do novo milênio e as exigências do mundo do trabalho*. 2006. Disponível em: [http://www.cinform.ufba.br/iv\\_anais/artigos/TEXT016.HTM](http://www.cinform.ufba.br/iv_anais/artigos/TEXT016.HTM) Acesso em: 18 jan. 2020.

SANDERS, Nada R. *How to use big data to drive your supply chain*. *California Management Review*, v. 58, n. 3, p. 26-48, Spr. 2016. Disponível em: <https://doi.org/10.1525/cm.2016.58.3.26>. Acesso em: 27 de jan. 2020.

SEMIDÃO, R. A. M. *Dados, Informação e Conhecimento enquanto elementos de compreensão do universo conceitual da Ciência da Informação: contribuições teóricas*. 2014. 198 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília, 2014. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/110783/000799485.pdf?sequence=1>. Acesso em: Jan. 2020.

STANTON, J. *Data science: What's in it for the new librarian*. *Information Space*, v. 9, n. 20, 2012.

SEMELER, A. R.; PINTO, A. L.; ROZADOS, H.B. *Data science in data librarianship: Core competencies of a data librarian*. *Journal of Librarianship and Information Science*, v. 50, p. 096100061774246-10, 2017.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia e dados. *Ciência da Informação*, v. 48, n. 1, p. 113-129, 2019.

SEMELER, Alexandre Ribas; PINTO, Adilson Luiz. *Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados*. *Ciência da Informação*, Brasília, DF, v. 48, n. 1, p. 130-129, 1 abr. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 30 jan. 2020.

SIEGEL, IDALTCHION FABRICIO. *LINGUAGEM PYTHON E SUAS APLICAÇÕES EM CIÊNCIA DE DADOS*. Orientador: Cledson Sousa. 2018. 57 p. Trabalho acadêmico de Conclusão de Curso (Curso de Tecnologia em Sistemas de Computação) - título de Tecnólogo em Sistemas de Computação, Niterói, 2018. Disponível em: <https://app.uff.br/riuff/handle/1/8946>. Acesso em: 29 fev. 2020.

SMITH, F. Data science as an academic discipline. *Data Science Journal*, v. 5, n. 0, p. 163–164, 28 nov. 2006. Disponível em: <http://datascience.codata.org/articles/abstract/482/>. Acesso em: 20 maio 2020

ROCHA, L. L.; SALES, L. F.; SAYÃO, L. F. *Uso de cadernos eletrônicos de laboratório para as práticas de ciência aberta e preservação de dados de pesquisa*. Ponto de Acesso, v. 11, n. 3, p. 2-16, 2017.

TAYLOR, R. S. *Professional aspects of Information Science and Technology* apud BORKO, H. *Information Science: what is it?* opus cit p. 8.

TECHTARGET ANZ STAFF. *What is hyperconnectivity?* Computer Weekly, 19 fev. 2007.

TECHOPEDIA.com, '*What is the Internet of Things (IoT)?* - Definition from Techopedia', 2015. [Online]. Available: <https://www.techopedia.com/definition/28247/internetof-things-iot>.

TARAPANOFF, Kira. *O profissional da informação pensando estrategicamente*. IN: Simpósio Brasil-Sul de Informação, 27 a 30 de maio de 1996, Londrina. Anais ... Londrina: Editora UEL, 1996. p.1 15-141

THURASINGHAM B.; "*Data Mining*", CRC Press, 1999.

TIBBO, H.; JONES, S. *Research data management and sharing*, 2017. Disponível em: <https://www.coursera.org/learn/datamanagement>. Acesso em: JANEIRO. 2020.

TENOPIR, C. et al. *Research Data Services in European Academic Research Libraries*. LIBER Quarterly, v. 27, n.1, p. 23–44, 2017. Disponível em: <https://www.liberquarterly.eu/articles/10.18352/lq.10180/>. Acesso em JANEIRO. 2020.

TARTAROTTI, Roberta Cristina Dal'Evedove; DAL'EVEDOVE, Paula Regina; FUJITA, Mariângela Spotti Lopes. *BIBLIOTECONOMIA DE DADOS EM REPOSITÓRIOS DE PESQUISA: PERSPECTIVAS PARA A ATUAÇÃO BIBLIOTECÁRIA*. **Inf. Inf.**, Londrina, v. 24, n. 3, p. 207 – 226, 2 dez. 2019. DOI 10.5433/1981-8920.2019v24n3p207. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao>. Acesso em: 30 jan. 2020.

VALENTIM, M. L. P. *Equipes multidisciplinares na gestão da informação e conhecimento*. In. BAPTISTA, S. G.; MULLER, S. P. M. (Org.) *Profissional da informação: o espaço de trabalho*. Brasília: Thesaurus, 2004. Cap. 7, p. 154-176.

VICTORINO, M. C. *Organização da informação para dar suporte à arquitetura orientada a serviços: reuso da informação nas organizações*. 2011. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília. Disponível em: Acesso em: 10 jan. 2020

VICTORINO, Márcio de Carvalho et al. *Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados*. *Informação & Sociedade: Estudos*, João Pessoa, v. 27, n. 1, p. 225-242, jan./abr. 2017. Disponível em: Acesso em: 9 jan. 2020.

VENKATESH, Kirpal A.; NEELAMEGAM, K.; REVATHY, R. *Usando MapReduce e balanceamento de carga em nuvens*, out. 2010. Disponível em: <<http://www.ibm.com/developerworks/br/java/library/cl-mapreduce/#N1010A>>. Acesso em: 24 jan. 2020.

VAN DER AALST, W. M. P. *Data Scientist: The Engineer of the Future*. In: MERTINS, K. et al. (Ed.). *Enterprise Interoperability VI: Interoperability for Agility, Resilience and Plasticity of Collaborations*. Berlin: Springer, 2014.

VOULGARIS Z (2014) *Data Scientist: The Definitive Guide to Becoming a Data Scientist*. *Basking Ridge*: Technics Publications.

XIA, J.; WANG, M. *Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions*. *College & Research Libraries*, v. 75, n. 3. p. 362-388, 2014. Disponível em: <http://crl.acrl.org/content/75/3/362.abstract> . Acesso em: 24 Jan. 2020.

WANG, L. *Twinning data science with information science in schools of library and information science*. *Journal of Documentation*, Emerald Publishing Limited, v. 74, n. 6, p. 1243–1257, 2018.

WEISER, M. (1991). *The Computer for the 21st Century*. *Scientific American*. 265(3), 94-104.

WERSIG, Gernot. *Information Science: the study of postmodern knowledge usage*. *Information and management*, v.29, n.2, p.229-239, 1993.

WERSIG, Gernot; NEVELING, Ulrich. *The phenomena of interest to Information Science*. *Information Scientist*, [S.l.], v.9, n. 4, p. 127-140, Dec. 1975. Disponível em: . Acesso em: 24 jan. 2020.

WIENER, Norbert. *Cybernetics or control and communication in the animal and the machine*. New York, John Wiley, 1948. 194 p.

WORMELL, Irene. *Habilidades de gerenciamento e de empreendimento na profissão de Bibliotecário e Cientista da Informação*. *Informação & Informação*, Londrina, v. 4, n. 1, p. 7-16, jan./jun. 1999.

WHITE, T. (2010). *Hadoop: The Definitive Guide*. O'Reilly Media.

WESTPHAL CHRISTOPHER; BLAXTON TERESA, “*Data Mining Solutions – Methods and Tools for Solving real-Word Problems*”; John Wiley & Sons, Inc., 1998.

WEIS SHOLOM M.; INDURKHYA NITIM, “*Predict Data Mining*”; Morgan Kaufmann Publishers, Inc, 1999.

WEINBERG, Alvin M. Impact of large-scale Science on the United States: big science is here to stay, but we have yet to make the hard financial and educational choices it imposes. *Science*, v. 134, n. 3473, p. 161–164, 21 jul. 1961. DOI 10.1126/science.134.3473.161. Disponível em: <https://doi.org/10.1126/science.134.3473.161>. Acesso em: 4 maio 2019.

WEINBERG, ALVIN M. *Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information*. Washington, DC: Science Advisory Committee, 10 jan. 1963. Disponível em: <https://files.eric.ed.gov/fulltext/ED048894.pdf>.

ZHU, Y.; XIONG, Y. *Towards Data Science*. *Data Science Journal*, v. 14, p. 8, 2015. Disponível em: <http://doi.org/10.5334/dsj-2015-008>. Acesso em: JANEIRO. 2020.

ZHOU, Z.-H. *Three perspectives of data mining*. *Artificial Intelligence Journal*, p. 139–146, 2003.

