

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (UFRJ)  
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS (CCJE)  
FACULDADE DE ADMINISTRAÇÃO E CIÊNCIAS CONTÁBEIS (FACC)  
CURSO DE BIBLIOTECONOMIA E GESTÃO DE UNIDADE DE INFORMAÇÃO (CBG)

**LUIZ CLAUDIO REZENDE REIS**

BIG DATA: PERFIS PROFISSIONAIS E COMO O BIBLIOTECÁRIO PODE SE INSERIR  
NA ÁREA

Rio de Janeiro

2017

LUIZ CLAUDIO REZENDE REIS

**BIG DATA: UM NOVO CAMPO DE ATUAÇÃO PARA BIBLIOTECÁRIOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Biblioteconomia e Gestão de Unidades de Informação da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de bacharel em Biblioteconomia.

Orientadora: Profa. Dra. Maria Irene da Fonseca e Sá

Rio de Janeiro

2017

## Ficha catalográfica

R375 Reis, Luiz Claudio Rezende.

Big data: um novo campo de atuação para bibliotecários /  
Luiz Claudio Rezende Reis. - Rio de Janeiro, 2017

46f : il.

Orientadora: Maria Irene da Fonseca e Sá.

Monografia (Trabalho de Conclusão de Curso) – Curso de  
Biblioteconomia e Gestão de Unidades de Informação, Universidade  
Federal do Rio de Janeiro.

1.Big Data 2. Bibliotecário 3. Biblioteconomia 4.  
Profissional da Informação. I. Sá, Maria Irene da Fonseca e. II.  
Universidade Federal do Rio de Janeiro. III. Título.

CDD: 027.004.6

**LUIZ CLAUDIO REZENDE REIS**

**BIG DATA: UM NOVO CAMPO DE ATUAÇÃO PARA BIBLIOTECÁRIOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Biblioteconomia e Gestão de Unidades de Informação da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de bacharel em Biblioteconomia.

Rio de Janeiro, 18 de Dezembro de 2017.

---

Profa. Dra. Maria Irene da Fonseca e Sá – UFRJ  
Orientadora

---

Profa. Dra. Maria de Fátima Sousa de Oliveira Barbosa – UFRJ  
Membro interno

---

Prof. Me. Nikiforos Joannis Philyppis Junior – UFRJ  
Membro interno

Dedico este trabalho às minhas 4 mães:  
Claudia, Claudete, Terezinha e Maria Teresa.

## **AGRADECIMENTOS**

Agradeço à minha família, por todo o apoio que sempre me deram e por todo apoio que sei que sempre terei.

Agradeço à minha orientadora Maria Irene, pela paciência e ajuda no desenvolvimento desse trabalho.

Agradeço aos meus amigos de turma: Beatriz, Camila, Gabriel, Marvin e Rafael por terem tornado os anos de graduação em memórias que sempre apreciarei.

Agradeço aos meus amigos de fora da faculdade: Mateus, Tadeu, Tiago e Paulo Nobre por todas as vezes em que me ajudaram a segurar as pontas.

“A estrada do excesso leva ao palácio da  
sabedoria” (William Blake)

## RESUMO

As transformações socioeconômicas iniciadas na segunda metade do século XX estimularam a evolução do paradigma tecnológico reconfigurando a sociedade. Na chamada sociedade da informação, que alia a alta penetrabilidade das novas tecnologias e a lógica de redes predominante, os dados são produzidos em volumes exorbitantes e se tornam a principal matéria-prima desse contexto econômico. Nesse cenário, o Big Data chega para suprir a necessidade de extrair valor desses dados que são gerados a todo momento como consequência de cada ação realizada utilizando dispositivos digitais. A convergência tecnológica permitiu o desenvolvimento de hardware e software capazes de armazenar e processar quantidades exponenciais de dados com rapidez, tornando possível o cruzamento dos mesmos e sua análise transformando dados não-estruturados em informações úteis e em tempo real para as empresas e governos. Assim, o Big Data formou um novo campo de atuação para os profissionais de ciência da informação. O presente trabalho observa através de uma revisão da literatura existente, e da pesquisa bibliográfica de escopo exploratório e qualitativo as conceituações de Big Data, suas ferramentas de análise e, a partir disso, busca delimitar o perfil do profissional da área, identificando os cargos criados no mercado de trabalho e suas atribuições. Ao relacionar as competências dos profissionais de Big Data e dos bibliotecários o trabalho apresenta o campo de 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. Análise de Dados. Bibliotecário. Ciência da Informação. Mercado de Trabalho.



## ABSTRACT

The socioeconomic changes started at the second half of the twentieth century encouraged the evolution of the technological paradigm reshaping society. On the so called information society, that allies the deep penetrability of new technologies with the predominant network logic, data are produced in exorbitant volume and becomes the feedstock of the economical context. In this scenario, Big Data comes up to supply the necessity of extracting value from the everlasting data production that occurs as consequence to every action made with a digital device. The technological convergence allowed the development of hardware and software able to store and process exponential quantities of data with speed, enabling data crossing and its subsequent analysis turning non-structured data into useful information in real-time for companies and governments. So, Big Data became a new field of performance to Information Science professionals. This paper observes through literature review, and bibliographic research with an exploratory and qualitative scope Big Data concepts, its analysis' tools and from this seeks to define a profile of its field professional, identifying jobs created in the market and its assignments. Relating Big Data workers' skills with those of librarians the paper presents Big Data as a new opportunity to those librarians who wish to capacitate themselves in the data analysis field.

**Keywords:** Big Data. Data Analysis. Librarian. Information Science. Job Market.

**LISTA DE ILUSTRAÇÕES**

|   |    |
|---|----|
| <b>Figura 1</b> - Quantidade de dados gerados em um minuto na internet.....                       | 14 |
| <b>Quadro 1</b> - Dado, informação e conhecimento segundo Davenport.....                          | 23 |
| <b>Figura 2</b> - Fórmula do <i>Big Data</i> por Taurion.....                                     | 25 |
| <b>Figura 3</b> - Principais características dos 5 V's do <i>Big Data</i> .....                   | 26 |
| <b>Quadro 2</b> - O <i>big data</i> e o <i>analytics</i> tradicional.....                         | 27 |
| <b>Quadro 3</b> - Definições das tecnologias de <i>Big Data</i> por Davenport.....                | 29 |
| <b>Quadro 4</b> - Produtos de <i>NoSQL</i> .....  | 30 |
| <b>Quadro 5</b> - Diferenças entre Analistas de BI e Cientistas de Dados.....                     | 32 |
| <b>Quadro 6</b> - Novos cargos criados pelo <i>Big Data</i> e suas atribuições profissionais..... | 35 |
| <b>Quadro 7</b> - Principais cargos de Ciência de Dados e suas competências.....                  | 36 |
| <b>Figura 4</b> - Etapas da análise de dados e atribuição de cada cargo.....                      | 36 |
| <b>Quadro 8</b> - Competências e habilidades de Graduados em Biblioteconomia.....                 | 38 |

**LISTA DE ABREVIATURAS E SIGLAS**

|       |   |
|-------|---|
| ASCII | <i>American Standard Code For Information Interchange</i>   |
| BI    | <i>Business Intelligence</i>                                |
| CBG   | Curso de Biblioteconomia e Gestão de Unidades de Informação |
| FGV   | Fundação Getúlio Vargas                                     |
| GPS   | <i>Global Positioning System</i>                            |
| HMR   | <i>Hadoop MapReduce</i>                                     |
| IDC   | <i>International Data Corporation</i>                       |
| ISACA | <i>Information Systems Audit and Control Association</i>    |
| MARC  | <i>Machine Readable Cataloging</i>                          |
| NoSQL | <i>Not Only Structured Query Language</i>                   |
| TI    | Tecnologia da Informação                                    |
| TIC   | Tecnologia da Informação e Comunicação                      |
| UFRJ  | Universidade Federal do Rio de Janeiro                      |

## SUMÁRIO

|              |   |           |
|--------------|---|-----------|
| <b>1</b>     | <b>INTRODUÇÃO.....</b>                      | <b>12</b> |
| 1.1          | JUSTIFICATIVA.....                          | 16        |
| 1.2          | OBJETIVOS.....                              | 17        |
| <b>1.1.2</b> | <b>Objetivo Geral.....</b>                  | <b>18</b> |
| <b>1.1.2</b> | <b>Objetivos Específicos.....</b>           | <b>18</b> |
| <b>2</b>     | <b>METODOLOGIA.....</b>                     | <b>19</b> |
| <b>3</b>     | <b>REFERENCIAL TEÓRICO.....</b>             | <b>22</b> |
| 3.1          | CONCEITUANDO DADOS.....                     | 22        |
| <b>3.1.1</b> | <b>Dado, Informação e Conhecimento.....</b> | <b>22</b> |
| <b>3.1.2</b> | <b>Grandes Dados.....</b>                   | <b>24</b> |
| 3.2          | FERRAMENTAS DE ANÁLISE.....                 | 28        |
| <b>4</b>     | <b>PERFIL DE ATUAÇÃO PROFISSIONAL.....</b>  | <b>32</b> |
| 4.1          | CARREIRAS EM <i>BIG DATA</i> .....          | 34        |
| 4.2          | ATUAÇÃO DE BIBLIOTECÁRIOS.....              | 37        |
| <b>5</b>     | <b>CONSIDERAÇÕES FINAIS.....</b>            | <b>41</b> |
|              | <b>REFERÊNCIAS.....</b>                     | <b>43</b> |

## 1 INTRODUÇÃO

A estrutura atual da sociedade tem sido alterada intrinsecamente pela evolução tecnológica subsequente ao final da Segunda Guerra Mundial, quando inovações nos campos de telecomunicações e microeletrônica se tornaram o foco de investimento econômico para as nações que competiam pela liderança política mundial durante o período da Guerra Fria. Esse cenário político e econômico propiciou o barateamento da produção e difusão do acesso às novas tecnologias de armazenamento e transmissão de dados no decorrer da segunda metade do século XX. Por consequência, a evolução das tecnologias de informação e comunicação (TICs) reconfigurou a forma de organização social em rede, redefinindo seu alcance e suas fronteiras e assim transformando todas as dinâmicas da sociedade moderna em processos altamente conectados.

Essas mudanças são acompanhadas também por uma nova estruturação da economia que segundo Castells (1999, p. 20) passa a ser conhecida como economia em rede e é "uma nova e eficiente forma de organização da produção, distribuição e gestão" que se torna a base do aumento da taxa de crescimento da produtividade nos países que adotaram essa nova forma de organização econômica. É esse recente paradigma técnico-econômico pelo qual as sociedades contemporâneas se configuram e que vem sendo chamado de "sociedade da informação". Também é Castells (1999 apud WERTHEIN 2000) que identifica as características fundamentais dessa sociedade: a informação como matéria-prima, alta penetrabilidade dos efeitos das novas tecnologias, predomínio da lógica de redes, flexibilidade e crescente convergência de tecnologias. Essas características são facilmente observáveis no atual contexto social com o advento de tecnologias como *smartphones* e a ascensão da popularidade de redes sociais como *Facebook* e aplicativos de mensagem instantânea como o *Whatsapp*.

O avanço tecnológico catapultado por essa reconfiguração socioeconômica gerou um aumento da velocidade em que os novos produtos de tecnologia são pesquisados e desenvolvidos, acarretando também uma diminuição de seus custos. Tornando-os, portanto, mais acessíveis para a população em geral. Atualmente as TIC têm uma acentuada penetrabilidade na sociedade, só no Brasil há mais de 152 milhões de computadores em uso segundo o 26º Relatório Anual de Tecnologia da Informação da Fundação Getúlio Vargas (FGV, 2015), somados a eles há também 154 milhões de *smartphones* em uso por uma população de 206 milhões de habitantes, um alcance de 3 dispositivos para cada 2 habitantes.

A *internet* é usada regularmente por 48% dos brasileiros segundo a Pesquisa Brasileira de Mídia 2015, e o uso de *smartphones* já ultrapassou o de microcomputadores como forma de acesso à internet. Os *smartphones* são um aparelho fruto de uma convergência tecnológica que aglutinou telefonia, fotografia, geolocalização e navegação na internet causando uma revolução na dinâmica de produção e consumo de informações, uma vez que agora todos os tipos de conteúdo textuais e audiovisuais podem ser produzidos e transmitidos de forma móvel. Isso gerou uma quebra do antigo paradigma de produção e disseminação de informação, antes restrito aos tradicionais veículos de mídia. Agora todo indivíduo que possui um *smartphone* deixa de ser apenas um consumidor para potencialmente se tornar também um produtor massivo de dados.

Fotos postadas no *Facebook*, *tweets*, mensagens instantâneas, buscas no *Google*, a rota utilizada por seu Global Positioning System (GPS), todos os tipos de ações realizadas online produzem dados e metadados, que se acumulam e formam assim um rastro digital. Segundo Eric Schmidt, então presidente do *Google*, diariamente eram criados 5 *exabytes* de dados no mundo. Um número que se equipara a quantidade total de dados criados desde o começo da civilização até 2003 (EIU, 2012). Taurion (2013) estimou que no ano de 2015 tenham sido alcançados 7,9 *zettabytes* de dados, um número de potência 3 vezes maior que o *exabyte*. Para dar a devida dimensão a esse número e entender sua escala, 1 *Byte* é o tamanho da quantidade de valores binários necessária para representar um caractere no código da *American Standard Code for Information Interchange* (ASCII). Um *zettabyte* é um sextilhão de bytes, que representado numericamente dá uma noção mais clara de sua grandeza: 1.000.000.000.000.000.000.000 *bytes*.

A empresa *Penny Stocks Lab* mantinha em sua página na *internet* um infográfico atualizado em tempo real quantificando os dados gerados nos principais serviços *online*. Em um minuto, aproximadamente 1.354.440 *Gigabytes* de dados eram transferidos pela internet. E essa quantidade de dados provém de múltiplas fontes, podendo ser mensagens trocadas no *Whatsapp*, arquivos armazenados no *Dropbox*, e-mails enviados ou mesmo o equivalente a 5.785 dias de vídeos assistidos através do *Youtube*. (PENNY STOCKS LAB, 2015)

Figura 1 – Quantidade de dados gerados em um minuto na internet



Fonte: Penny Stock Labs (2015)

Esses dados que formam o rastro digital são de fácil recuperação, porém devido a dificuldade do processamento em massa e o fato de serem essencialmente dados não-estruturados, o potencial da exploração desse volume de dados ainda não é plenamente aproveitado. Embora sua análise possa gerar informações significativas para empresas e governos estima-se que apenas 1% desses dados seja efetivamente analisado (BREITMAN, 2014).

É nesse contexto de volume e variedade de dados que surge o *Big Data*, termo que ainda não possui um consenso quanto ao seu conceito, mas que é alvo de diversos estudos e tentativas de definição. No geral, se refere a uma abordagem de armazenamento e tratamento de grandes conjuntos de dados com objetivo de extração de valor desses dados no sentido de tomada de decisões que esbarra, porém, nas dificuldades de trabalhar com uma amostra de dados tão extensa e variada. Para Manyika (2011, p. 1) “*Big data* se refere a conjuntos de dados cujo tamanho está além da capacidade típica dos softwares de bancos de dados de capturar, armazenar, gerenciar e analisar”, semelhante ao conceito mais técnico de Vieira (2012) que também problematiza a capacidade de trabalhar esses dados utilizando os sistemas tradicionais:

*Big Data* pode ser resumidamente definido como uma coleção de bases de dados tão complexa e volumosa que se torna muito difícil (ou impossível) e complexa fazer algumas operações simples (e.g., remoção, ordenação,

sumarização) de forma eficiente utilizando Sistemas Gerenciadores de Bases de Dados (SGBD) tradicionais. (VIEIRA et. Al., 2012, p. 2)

Outra definição de *Big Data* que reitera as dificuldades de se processar esses dados é a feita pela *International Business Machines* (IBM):

*Big Data* é um termo utilizado para descrever grandes volumes de dados e que ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas a cada dia. As dificuldades em armazenar, analisar e utilizar grandes conjuntos de dados têm sido um considerável gargalo para as companhias (IBM, 2015, p. 1)

Pjipers (2010) afirma que justamente a quantidade de dados se torna o maior empecilho para a recuperação da informação crucial, uma vez que é necessário gastar muito tempo buscando essa informação, o que reduz o tempo disponível para a análise e utilização da informação enquanto ela é relevante. Isso é causado pela defasagem tecnológica entre os meios de produção e distribuição da informação e os processos de organização e recuperação dela.

Para reduzir essa distância entre a capacidade e velocidade de ser gerar dados e os recupera-los de forma ágil dentro de um volume gigantesco de dados variados estão sendo desenvolvidas novas ferramentas baseadas em bancos de dados *NoSQL*, capazes de armazenar e processar *petabytes* de dados estruturados e não-estruturados, com destaque para o *Apache Hadoop*, utilizado por empresas como *IBM, Google, Twitter, Yahoo!, Netflix, Facebook*. (VIEIRA et al., 2012, p. 4)

O mercado de *Big Data* tem crescido em um ritmo análogo ao do volume de dados, uma previsão do *International Data Corporation* (IDC) aponta uma taxa de crescimento anual de 26.24% até 2018 para o mercado de tecnologia e serviços de *Big Data*, alcançando um valor total de mais de 41 bilhões de dólares. Como um campo novo e próspero no âmbito da análise de dados, tem atraído a atenção dos profissionais em ciência da informação e áreas afins.

Assim, o presente trabalho busca em primeiro momento estabelecer em sua introdução o contexto de seu tema, discorrendo sobre a sociedade da informação, o volume de dados gerado por ela, o advento das iniciativas de *Big Data* como forma de aproveitar esses dados, e o tamanho de seu mercado. A seguir é feita a justificativa pela escolha do tema, indicando seu problema de pesquisa e os objetivos gerais e específicos. Na parte seguinte, são apresentados seus procedimentos metodológicos e fontes de pesquisa consultadas. Na sua terceira seção, é apresentado o referencial teórico da pesquisa. São apresentados os conceitos de *Big Data*,



suas aplicações e ferramentas para melhor entendimento do tema. No capítulo seguinte é feita uma apuração do perfil profissional e uma análise dos cargos criados na área, para então relacionar a atuação na área de Big Data como uma opção para Bibliotecários e outros profissionais de ciência da informação. Por fim, as considerações finais são apresentadas respondendo o problema da pesquisa e propondo mais questões sobre a atuação profissional de bibliotecários e cientistas de dados em Big Data para trabalhos futuros.

## 1.1 JUSTIFICATIVA

Ribeiro (2014) afirma que *Big Data*: “[...] desperta, na atualidade, o interesse, e até mesmo o fascínio, para todas as pessoas que tem algum envolvimento com atividades para Gestão da Informação”. A recente profusão de trabalhos e artigos que discorrem sobre o tema são indicativo de que essa área já ultrapassou uma posição de tendência tecnológica para se solidificar como área de interesse do mercado. Porém essa área ainda carece de profissionais competentes para exercer suas funções que exigem refinados conhecimentos técnicos e sobre o negócio em que estão inseridos. Taurion (2013) afirma que:

[...] nos próximos anos viveremos uma escassez de profissionais, não só no Brasil, mas no mundo todo. Esta escassez, ao mesmo tempo que abre muitas perspectivas profissionais para os que abraçarem a função, também atuará como um entrave, pois dificultará às empresas usarem *Big Data* com eficiência. Recentes pesquisas estimam que, por volta de 2015, *Big Data* demandará cerca de 4,4 milhões de profissionais em todo o mundo e que apenas 1/3 destes cargos poderá ser preenchido com as capacitações disponíveis hoje em dia. (TAURION, 2013, não paginado)

Esses números indicam um déficit de quase 3 milhões de profissionais qualificados para preencher as vagas ofertadas. Breternitz e Silva (2013) reforçam a escassez de profissionais acusada por Taurion:

Profissionais com esse perfil são muito raros e sua formação demanda muito tempo, o que torna sua utilização cara – Bertolucci (2012), citando pesquisas da consultoria *McKinsey*, diz que apenas nos Estados Unidos faltarão em 2018 aproximadamente 190.000 desses profissionais. (BRETERNITZ; SILVA, 2013)

Novos cargos no campo de análise de dados estão sendo criados e apesar de normalmente demandarem formação em Ciência da Computação ou Matemática exigem também habilidades profissionais de classificação, busca e recuperação de dados. O

profissional competente de *Big Data* deve ser capaz de escolher que tipo de dados irá analisar, qual sua procedência, sua veracidade e seu prazo de relevância. Essas preocupações já fazem parte do trabalho dos bibliotecários, que em sua formação acadêmica são instruídos a organizar, tratar e selecionar dados e informações com a finalidade de gerar conhecimento.

Tendo em vista essa capacidade o mercado de *Big Data* se configura como um novo campo de atuação para os bibliotecários, onde seus conhecimentos de classificação, análise, indexação e recuperação de dados coadunam com as atividades realizadas pelos profissionais de *Big Data*, abrindo assim oportunidades de carreira para aqueles que possuem um perfil mais inclinado para o trabalho com tecnologia. Davenport (2014) indica a dificuldade de uma pessoa reunir todas as habilidades requeridas para o trabalho com *Big Data* ressaltando que o ideal é que se formem equipes multidisciplinares. Com essa proposição o bibliotecário pode se encaixar em uma posição de apoio ao trabalho de outros profissionais agindo como um consultor ou adquirindo um perfil mais técnico em tecnologia da informação para o desenvolvimento de sistemas de tratamento e recuperação de dados tendo em vista que a graduação em biblioteconomia já possui esse perfil multidisciplinar, destacando o curso de Biblioteconomia e Gestão de Unidades de Informação (CBG) da Universidade Federal do Rio de Janeiro (UFRJ), que envolve disciplinas técnicas, disciplinas de gestão e administração e disciplinas como Arquitetura da Informação, Análise e Modelagem de Processos e Tecnologias da Informação e Comunicação, formando assim um profissional com esse viés multidisciplinar inerente. Isso provém um interesse particular em pesquisar sobre esse campo de atuação não tradicional para graduados em Biblioteconomia e suas ferramentas de gestão, análise e visualização de dados.

Assim o trabalho se norteia pela questão de como os bibliotecários podem se capacitar para desempenhar papéis profissionais no campo de análise de dados e que competências são necessárias para isso. Portanto, busca analisar os conceitos de *Big Data*, suas ferramentas, mercado de atuação e as carreiras geradas na área para então buscar traçar a relação entre as áreas.

### 1.3 OBJETIVOS

Aqui são apresentados o Objetivo Geral, que delimita o foco da pesquisa em seu tema, e os Objetivos Específicos, que definem as questões propostas a serem respondidas por esse trabalho.

### **1.3.1 Objetivo Geral**

Apresentar *Big Data* como um possível campo de atuação profissional para bibliotecários.

### **1.3.2 Objetivos Específicos**

- a) Apresentar os conceitos do termo *Big Data* dando lastro ao tema da pesquisa.
- b) Levantar perfis de cargos e competências relacionados a *Big Data*.
- c) Apresentar maneiras como Bibliotecários podem se capacitar para atuar no campo de *Big Data*.

## 2 METODOLOGIA

Inicialmente foram definidos o tema e o escopo a serem pesquisados. Como uma área de interesse recente que tem gerado novos desafios para os profissionais de informação, o tema escolhido foi *Big Data*. Para delimitar o escopo de pesquisa ao curso, o trabalho tem como objetivo geral propor o campo de atuação profissional em *Big Data* como uma área que pode ser preenchida por bibliotecários. A abordagem de pesquisa que melhor se adota para o caso do trabalho é a pesquisa qualitativa, buscando aprofundar a compreensão do tema estudado. Gerhardt e Silveira (2009) indicam que essa abordagem busca explicar o porquê das coisas, sem quantificar seus valores pois os dados analisados não são simétricos, se valendo de diferentes abordagens. Esses autores apresentam também as características dessa abordagem:

As características da pesquisa qualitativa são: objetivação do fenômeno; hierarquização das ações de descrever, compreender, explicar, precisão das relações entre o global e o local em determinado fenômeno; observância das diferenças entre o mundo social e o mundo natural; respeito ao caráter interativo entre os objetivos buscados pelos investigadores, suas orientações teóricas e seus dados empíricos; busca de resultados os mais fidedignos possíveis; oposição ao pressuposto que defende um modelo único de pesquisa para todas as ciências. (GERHARDT; SILVEIRA, 2009, p.32)

Com o objetivo e o intuito de explicitar o tema e demonstrar hipótese de ser um foco para discussão em Biblioteconomia, a pesquisa realizada nesse trabalho é uma pesquisa exploratória. Na definição de Prodanov e Freitas (2013) a pesquisa exploratória:

[...] tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar, possibilitando sua definição e seu delineamento, isto é, facilitar a delimitação do tema da pesquisa; orientar a fixação dos objetivos e a formulação das hipóteses ou descobrir um novo tipo de enfoque para o assunto. (PRODANOV; FREITAS, 2013, p.51, 52)

Quanto aos procedimentos, foi escolhida a pesquisa bibliográfica para que pudessem ser analisadas as fontes que já se propuseram a realizar estudos sobre o tema:

A pesquisa bibliográfica é feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites. Qualquer trabalho científico inicia-se com uma pesquisa bibliográfica, que permite ao pesquisador conhecer o que já se estudou sobre o assunto. Existem, porém pesquisas científicas que se baseiam unicamente na pesquisa bibliográfica, procurando referências teóricas publicadas com o objetivo de recolher informações ou

conhecimentos prévios sobre o problema a respeito do qual se procura a resposta (FONSECA, 2002, p. 32).

Esses dados, porém, não são suficientes já que *Big Data* é um tema recente e foco de pesquisas, reportagens e outros tipos de materiais que não foram tratados cientificamente. Para essas fontes foi realizada também uma pesquisa documental:

A pesquisa documental trilha os mesmos caminhos da pesquisa bibliográfica, não sendo fácil por vezes distingui-las. [...] A pesquisa documental recorre a fontes mais diversificadas e dispersas, sem tratamento analítico, tais como: tabelas estatísticas, jornais, revistas, relatórios, documentos oficiais, cartas, filmes, fotografias, pinturas, tapeçarias, relatórios de empresas, vídeos de programas de televisão, etc. (FONSECA, 2002, p. 32).

Assim foi realizada uma revisão bibliográfica foram utilizadas diversas fontes de informações sobre o assunto como livros, reportagens, notícias e postagens de *blogs*. Também foi realizada uma pesquisa em portais de busca como o *Google* e o *Google Acadêmico*, bases de dados como a *SciELO* e repositórios digitais da área de Ciência da Informação, como o do periódico *DataGramZero* com a finalidade de encontrar as produções acadêmicas e científicas que já relacionassem *Big Data* a Biblioteconomia.

Nessa busca foram utilizadas palavras chaves como: *Big Data*, profissional da informação, cientista de dados, analista de dados, bibliotecário e perfil profissional. Para uma melhor recuperação de resultados essas palavras chave foram utilizadas em combinação com operadores booleanos e também na língua inglesa, visto que é a língua em que se encontra o grosso da produção textual sobre o tema. Os resultados dessas buscas compõem o objeto de análise do trabalho, sendo formado por livros, artigos científicos, relatórios, notícias e descrições de vagas anunciadas no *LinkedIn*.

Esses dados foram submetidos à análise buscando verificar as hipóteses propostas para se responder o problema da pesquisa. A execução dessa etapa é descrita por Quivy e Campenhoudt (1998):

(...) cada hipótese elaborada na fase de construção expressa as relações que pensamos serem corretas e que devem ser confirmadas pela coleta de dados. Os resultados encontrados são os que resultam das operações precedentes. É comparando os resultados encontrados com os resultados esperados pela hipótese que poderemos tirar as conclusões. Se houver divergência entre os resultados observados e os resultados esperados, será necessário examinar de onde provém esse distanciamento e em que a realidade é diferente do que se presumia no início, elaborando novas hipóteses e, a partir de uma nova análise dos dados disponíveis, examinar em que medida elas se confirmam. (QUIVY; CAMPENHOUDT, 1998, p. 222)

Por fim, os resultados da análise do referencial teóricos são apresentados nas considerações finais, respondendo o problema da pesquisa e abrindo novas vertentes para futuros trabalhos explorando o tema.

### 3 REFERENCIAL TEÓRICO

#### 3.1 CONCEITUANDO DADOS

Para o desenvolvimento do trabalho e sua compreensão, é necessário elucidar os conceitos utilizados fazendo seu paralelo entre as áreas de Ciência da Informação e de Tecnologia da Informação (TI). Nessa parte discute-se primeiramente os conceitos de dados, informação e conhecimento para o entendimento do processo de extração de valor que é o objetivo do *Big Data*. Em sequência, apresenta-se diversos conceitos sobre o termo, relacionando autores para expandir o entendimento do mesmo e destrinchar suas características. Finalizando o capítulo, são apresentadas as ferramentas que são utilizadas para se trabalhar com *Big Data*.

##### 3.1.1 Dado, Informação, Conhecimento

Antes de entender qualquer conceito de *Big Data* é necessário o entendimento do que é Dado, a unidade básica de conteúdo trabalhada pelo *Big Data* e que posteriormente é ligado à informação e ao conhecimento.

Davenport (1998, p.19) define dados como elementos brutos, sem significado. Já Setzer (2001) os define como “uma abstração formal que pode ser representada e transformada por um computador” (SETZER 2001 apud RUSSO, 2010, p.15). Russo (2010, p.15) expande a definição que dados são “[...] sinais que não foram processados, correlacionados, integrados, avaliados ou interpretados de qualquer forma, e, por sua vez, representam a matéria prima a ser utilizada na produção de informações”. Essa definição é bastante próxima da de Miranda, que por sua vez se assemelha a como os dados são trabalhados na área de *Big Data*: “um conjunto de registros qualitativos ou quantitativos, conhecido, que organizado, agrupado, categorizado e padronizado adequadamente transforma-se em informação” (1999, p.285).

O passo seguinte é a definição de informação, Drucker define como “dados dotados de relevância e propósito” (DRUCKER 1988 apud DAVENPORT, 1998, p.18). Miranda (1999) avança o termo como “dados organizados de modo significativo, sendo subsídio útil à tomada de decisão”. Russo (2010, p.15) entende informação como “[...] dados processados e contextualizados”. Setzer (2001) considera que informação “[...] é uma abstração informal (isso é, não pode ser formalizada através de uma teoria lógica ou matemática), que está na

mente de alguém, representando algo significativo para essa pessoa”. Essa definição que aborda a significação dos dados realizada pela mente humana adianta o fator de interpretação que diferenciará informação de conhecimento.

As definições de conhecimento em geral envolvem o fator humano de processar as informações e as relacionar com outras, avaliando seus fatores externos. Para Davenport conhecimento é definido como “informações que foram analisadas e avaliadas sobre sua confiabilidade, sua relevância e sua importância” (DAVENPORT, 1998). Na definição de Angeloni (2003, p.18):

O conhecimento pode então ser considerado como a informação processada pelos indivíduos. O valor agregado à informação depende dos conhecimentos anteriores desses indivíduos. Assim sendo, adquirimos conhecimento por meio do uso da informação nas nossas ações. Desta forma, o conhecimento não pode ser desvinculado do indivíduo, ele está estritamente relacionado com a percepção do mesmo, que codifica, decodifica, distorce e usa a informação de acordo com suas características pessoais, ou seja, de acordo com seus modelos mentais. (ANGELONI, 2003, p.18)

Quadro 1 – Características de Dado, Informação e Conhecimento segundo Davenport:

| <b>Dado</b>  | <b>Informação</b>                         | <b>Conhecimento</b>   |
|--|---|---|
| Simple observações sobre o estado do mundo.                  | Dados dotados de relevância e propósito.  | Informação valiosa da mente humana. Inclui reflexão, síntese, contexto. |
| Facilmente estruturado.                                      | Requer unidade de análise.                | De difícil estruturação.  |
| Facilmente obtido por máquinas. Quantificado com frequência. | Exige consenso em relação ao significado. | De difícil captura em máquinas.   |
| Facilmente transferível.                                     | Exige necessariamente a mediação humana   | Frequentemente tácito; de difícil transferência.                        |

FONTE: DAVENPORT, 1998, p.18

Outro tipo de dados que é preciso definir para a compreensão de *Big Data* são os Metadados, que basicamente são dados sobre dados. Dziekaniak e Kirinus (2004) identificam metadados como a estrutura descritiva de um documento. São os metadados que contém as informações implícitas dos arquivos digitais, como a geolocalização de uma foto ou qual dispositivo foi utilizado para captura-la. De Souza, Catarino e Dos Santos (1997) definem como objetivo dessa forma de descrição a colaboração na orientação, no desenvolvimento e



descrição dos documentos eletrônicos, emergindo padrões, produção e manipulação da descrição por metadados. Para eles:

A catalogação dos dados propiciará a maior utilização deles por usuários com múltiplos interesses. Sem uma documentação eficiente dos dados é dificultada aos usuários a localização de dados necessários para suas aplicações. Os dados precisam conter informações que auxiliem seus usuários a tomar decisões sobre sua devida aplicação (DE SOUZA, CATARINO, DOS SANTOS, 1997, p. 94)

Existem diversos padrões de metadados, que dizem respeito aos diferentes tipos de informação. Na biblioteconomia, a aplicação de metadados é observada na catalogação. Ao atribuir informações como título, autoria, data de publicação, entre outros na catalogação de uma obra esses metadados tornam possível a pesquisa e recuperação de um item do acervo, ou mesmo pesquisas sobre quantos títulos de um mesmo autor existem no catálogo. Um desses padrões de metadados amplamente utilizado para a catalogação bibliográfica, é o *Machine Readable Cataloging* (MARC).

A proliferação de formatos de arquivos digitais e a explosão informacional vivida atualmente com a produção escalonada de conteúdo digital diariamente por *smartphones* e outros dispositivos conectados à internet geram metadados em um número exorbitante. A quantidade disponível e sua relação indissociável ao cotidiano do indivíduo tornam os metadados a mais rica fonte para a análise e extração de informações sobre consumidores e seus hábitos, despertando o interesse de empresas que buscam atingir vantagem competitiva através de ferramentas de análise de dados. Os metadados são, portanto, o principal insumo do *Big Data*.

### 3.1.2 Grandes Dados

Davenport (2014, p.114) descreve *Big Data* como um rio de informações cuja correnteza é impetuosa. Essa analogia se refere ao volume e velocidade com que as informações são produzidas e disponibilizadas. As definições de *Big Data* se assemelham em sua maioria, é comum descrever as características como a quantidade de dados trabalhados, seus diversos tipos e a acelerada frequência em que são disponibilizados. Um outro ponto comum é apontar a inadequação das ferramentas de análise já existentes, como na definição do *Information System Audit and Control Association* (ISACA):

*Big data* refere-se, principalmente, aos conjuntos de dados que são muito grandes ou com rápidas mudanças para serem analisados com técnicas de banco de dados relacionais tradicionais ou multidimensionais ou ferramentas de software comumente usadas para capturar, gerenciar e processar os dados em um tempo razoável. (ISACA, 2013, p.5)

Algumas definições se atêm a resumir *Big Data* como a análise de volumes de dados provenientes de diversas fontes como na de Vieira: “*Big Data* pode ser resumidamente definido como o processamento (eficiente e escalável) analítico de grandes volumes de dados complexos produzidos por (várias) aplicações” (VIEIRA et al., 2012, p.6).

Laney (2001 apud SIMON, 2013) estabeleceu as características mais observadas em outras definições como os 3 V's do *Big Data*: Volume, Variedade e Velocidade. Esse conceito é expandido por Taurion ao adicionar o fator humano de análise da utilidade da informação e o fator econômico de se obter vantagem competitiva:

*Big Data* não trata apenas da dimensão volume, como parece à primeira vista, mas existe também uma variedade imensa de dados, não estruturados, dentro e fora das empresas (coletados das mídias sociais, por exemplo), que precisam ser validados (terem veracidade para serem usados) e tratados em velocidade adequada para terem valor para o negócio. A fórmula é então, *Big Data* = volume + variedade + velocidade + veracidade, gerando valor. (TAURION, 2013, não paginado).

Figura 2 – Fórmula do *Big Data* por Taurion



Fonte: o autor (2017)

Figura 3 – Principais características dos 5 V's do *Big Data*:



Fonte: Silva (2017)

Taurion busca justificar cada fator dessa equação: volume a característica mais óbvia devido a quantidade de dados produzida todo dia; variedade por serem provenientes de fontes estruturadas e não-estruturadas e de natureza multimídia; velocidade devido a necessidade de extração de informação praticamente em tempo real; veracidade para se ter certeza da autenticidade dos dados e valor como contrapartida financeira para os investimentos em projetos de *Big Data*. Outra definição que relaciona essas características é a de Magalhães (2014):

*Big Data* é uma grande massa de dados/metadados aos quais geramos todos os dias, com características estruturadas (armazenados em banco de dados) e não-estruturadas (fotos, vídeos, e-mails) e que, na maioria das vezes, é analisada para atender a eventos em tempo real, buscando a partir da autenticidade dos dados dar sentido as informações relevantes passíveis de agregar valor tanto para empresas que buscam estratégias para seu negócio como para governos que buscam entender as demandas e características da população. (MAGALHÃES et al., 2014)

É preciso ressaltar o valor como característica diferencial dessas descrições, uma vez que é seu potencial econômico que torna o *Big Data* um campo de análise de dados de viabilidade mercadológica e justifica os investimentos empresariais em iniciativas de pesquisa e desenvolvimento das tecnologias utilizadas em sua aplicação. O acesso a esse conhecimento massivo no âmbito da análise de dados e *Business Intelligence* (BI) ou Inteligência Empresarial se traduz em medidas de ajuste da tomada de decisões pelos seus gestores, adaptando estratégias para obter melhores desempenhos e vantagem competitiva sobre seus concorrentes.

Davenport (2014, p.4) estabelece as diferenças concretas entre o *Big Data* e ferramentas de análise de dados tradicionais:

Quadro 2 – O *big data* e o *analytics* tradicional

|                           | <b>Big Data</b>            | <b>Ferramentas tradicionais</b>       |
|---------------------------|----------------------------|---------------------------------------|
| <b>Tipo de dados</b>      | Formatos não estruturados  | Dados formatados em linhas e colunas  |
| <b>Volume de dados</b>    | 100 terabytes a petabytes  | Dezenas de terabytes ou menos         |
| <b>Fluxo de dados</b>     | Fluxo constante de dados   | Pool estático de dados                |
| <b>Métodos de análise</b> | Aprendizado de máquina     | Baseados em hipóteses                 |
| <b>Objetivo principal</b> | Produtos baseados em dados | Suporte ao processo decisório interno |

Fonte: Davenport (2014) – Quadro produzido pelo autor

Manyika et al. (2011) apontam que “os dados podem realmente criar valor significativo para a economia mundial, potencialmente aumentando a produtividade e competitividade das empresas e a criação de um excedente econômico substancial para os consumidores e seus governos”. Magalhães et al. (2014) exemplificam diferentes formas como isso acontece:

[...] empresas de diferentes áreas vêm utilizando, constantemente, o *Big Data* para buscar dados relevantes que possam prover a estratégia do seu negócio, fazendo uso de informações de diversos usuários para melhorar um produto, criar uma estratégia de *marketing* mais eficiente, cortar gastos, evitar o desperdício de recursos, superar um concorrente disponibilizando um serviço a um cliente de maneira satisfatória. (MAGALHÃES et al, 2014)

Boyd e Crawford (2011) apontam que a característica mais relevante de *Big Data* talvez não seja seu volume, mas sim a capacidade de relacionar os dados uns aos outros. Essa afirmação evoca a ideia de que apenas o acesso a quantidades exacerbadas de dados não produz valor algum, é necessária a indicação humana de quais dados devem ser confrontados com a amostra e posterior interpretação das informações extraídas disso para que se possa produzir conhecimento.

Essa multiplicidade de definições e características às quais *Big Data* é associado coadunam com a explicação de Taurion (2013) de que “*Big Data* não é apenas um produto de software e hardware, mas um conjunto de tecnologias, processos e práticas que permitem às empresas analisarem dados a que antes não tinham acesso e tomar decisões ou mesmo gerenciar atividades de forma muito mais eficiente”.

### 3.2 FERRAMENTAS DE ANÁLISE

O avanço tecnológico de infraestrutura física, como a miniaturização de dispositivos de armazenamento de dados com o aumento da capacidade de discos rígidos e cartões de memória, aumento da banda de tráfego de dados via internet e barateamento dos componentes de processamento, possibilitou às organizações possuírem aparato de *hardware* potente suficiente para administrar e processar um fluxo enorme de dados sem um ônus financeiro demasiadamente elevado. Em paralelo a isso, as iniciativas de *software* de código aberto desenvolveram uma plataforma de programas para gerir a parte lógica desse processamento em massa. Para tal, é necessário que os programas sejam capazes de lidar com um volume grande de dados encaixando-os em modelos estatísticos que permitam otimizar sua leitura e prever padrões. Davenport (2014) aponta que a tecnologia utilizada para isso é a do Aprendizado de Máquina, processo no qual o *software* automatiza o desenvolvimento de modelos analíticos através do aprendizado com os dados, gerando assim resultados que não foram inicialmente programados para encontrar.

Pereira (2016) divide as tecnologias de *Big Data* em duas óticas:

- [...] a) as tecnologias envolvidas com análise de dados, cujos principais representantes são o *Hadoop MapReduce* (HMR);
- b) as tecnologias de infraestrutura, responsáveis pelo armazenamento e processamento dos dados. São os bancos de dados *Not Only Structured Query Language* (NoSQL). (PEREIRA, 2016)

O *framework* Apache *Hadoop*, referência no processamento de volumes massivos de dados através da distribuição do processamento em *clusters* e nós, foi desenvolvido baseado no modelo *MapReduce* de processamento dividido em duas etapas: *Map*, que mapeia e distribui os dados em diversos nós de processamento e armazenamento; *Reduce*, que agrega e processa os resultados parciais para gerar um resultado final (Vieira et al., 2012, p.4). Na definição de Da Silva (2006), cluster é a denominação dada a um agrupamento de máquinas para solução de um dado problema, em que cada um dos computadores é chamado de nó.

Composto por módulos divididos com funções de distribuir os dados em máquinas dentro do cluster, o *Hadoop* é capaz de otimizar o processamento gerenciando os recursos computacionais e prevendo falhas de algum nó. Outra característica de destaque é a capacidade de customização com a adição de pacotes de *software* extras integrantes do ecossistema *Hadoop* como: *Apache Pig*, *Apache Hive*, *Apache HBase*, *Apache Phoenix*, *Apache Spark* e outros. Em resumo “[...] um framework para distribuir os dados em vários computadores; trata-se de um ambiente unificado de armazenamento e processamento altamente escalonável para grandes e complexos volumes de dados” (DAVENPORT, 2014).

Davenport apresenta também uma visão geral das tecnologias de Big Data e suas definições no seguinte quadro.

Quadro 3 – Definições das tecnologias de *Big Data* por Davenport:

| Tecnologia                               | Definição   |
|--|---|
| <i>Hadoop</i>                            | <i>Software</i> de código aberto para o processamento de big data em uma série e servidores paralelos |
| <i>MapReduce</i>                         | Um <i>framework</i> arquitetônico no qual o <i>Hadoop</i> se baseia                                   |
| Linguagens de <i>script</i>              | Linguagens de programação adequadas ao <i>big data</i> (Exemplos: <i>Python</i> , <i>Pig Hive</i> )   |
| Aprendizado de Máquina                   | <i>Software</i> para identificar rapidamente o modelo mais adequado ao conjunto de dados              |
| <i>Visual Analytics</i>                  | Apresentação dos resultados analíticos em formatos visuais ou gráficos                                |
| Processamento de Linguagem Natural (PLN) | <i>Software</i> para análise de texto   |
| <i>In-memory analytics</i>               | Processamento de <i>big data</i> na memória do  |

|  |  |
|--|--|
|  | computador para maior velocidade de resultados |
|--|--|

Fonte: Davenport (2014, p.112)

Vieira (2012) aponta o surgimento de bancos de dados *NoSQL* no contexto de grande quantidade de dados gerados em um espaço de tempo curto, sendo necessário sistemas de grande poder de processamento de forma eficiente e escalável com suporte a dados complexos, semiestruturados ou não-estruturados. A escolha do bando de dados *NoSQL* a ser utilizado é relacionada ao modelo de representação de dados que se quer.

A tabela a seguir lista os principais produtos *NoSQL* disponíveis de acordo com seu modelo de classificação de dados:

Quadro 4 – Produtos de *NoSQL*

| <b>Modelo de Dados</b> | <b>Produto NoSQL</b>  |
|------------------------|---|
| Baseado em Coluna      | <i>Hbase, Cassandra, Hypertable, Accumulo, Amazon SimpleDB, Cloudata, Cloudera, SciDB, HPCC, Stratosphere;</i>  |
| Baseado em Documentos  | <i>MongoDB, CouchDB, BigCouch, RavenDB, Clusterpoint Server, ThruDB, TerraStore, RaptorDB, JasDB, SisoDB, SDB, SchemaFreeDB, djondb;</i>  |
| Baseado em Grafos      | <i>Neo4J, Infinite Graph, Sones, InfoGrid, HyperGraphDB, DEX, Trinity, AllegroGraph, BrightStarDB, BigData, Meronymy, OpenLink Virtuoso, VertexDB, FlockDB;</i>   |
| Baseado em Chave-Valor | <i>Dynamo, Azure Table Storage, Couchbase Server, Riak, Redis, LevelDB, Chordless, GenieDB, Scalaris, Tokyo Cabinet/Tyrant, GT.M, Scalien, Berkeley DB, Voldemort, Dynamite, KAI, MemcacheDB, Faircom C-Tree, HamsterDB, STSdb, Tarantool/Box, Maxtable, Pincaster, RaptorDB, TIBCO Active Spaces, allegro-C, nessDB, HyperDex, Mnesia, LightCloud, Hibari.</i> |

FONTE: (VIEIRA et al, 2012) Quadro produzido pelo autor

Pereira (2016) ressalta que o grande desafio para os profissionais atuando com *Big Data* é justamente a escolha do banco de dados mais adequado e qual ferramenta de análise de dados será aplicada, visto que a tecnologia escolhida é a responsável por extrair dos dados as informações que podem gerar valor para a empresa.



#### 4 PERFIL DE ATUAÇÃO PROFISSIONAL

Como a área de *Big Data* ainda é novidade em termos acadêmicos, não há um campo de formação formal para esse profissional. Os cargos da área são geralmente ocupados por pessoas oriundas das áreas de Ciência da Computação, Matemática ou Engenharia devido à sua natureza técnica. Entre as competências requeridas para os profissionais da área, Taurion destaca “[...] ter conhecimentos de estatística, matemática, entender do negócio e ter familiaridade com tecnologias e linguagens como *Hadoop* e *Pig*” (2014). Davenport e Patil (2012) indicam como a principal característica deste profissional “[...] ter a capacidade de aplicar ferramentas analíticas e algoritmos para gerar previsões sobre produtos e serviços”.

Ribeiro (2014) identifica um dos focos de aplicação de projetos de Big Data e demonstra as atribuições do profissional:

Em suma, os projetos de *Big Data* são desenvolvidos com os objetivos de criar novos produtos, compreender novas necessidades dos clientes e seus comportamentos, bem como perceber novos mercados. Para isto, é necessário desenvolver teorias para tratar com clientes e usuários construindo hipóteses e identificando dados e informações relevantes. Este processo deve ser repetido e refinado, de acordo com os experimentos realizados e as respostas obtidas. (MARCHAND; PEPPARD apud RIBEIRO 2014)

Assim o profissional de *Big Data* possui um perfil com foco maior em questões preditivas, desejando antecipar padrões e gerando previsões baseadas em fatos ao correlacionar os dados analisados. O que o difere do perfil do profissional de *Business Intelligence* (BI), mais voltado para representações gráficas sobre dados passados.

Quadro 5 - Diferenças de competências entre Analistas de BI e Cientistas de Dados:

| ANALISTA DE BI   | CIENTISTA DE DADOS  |
|--|---|
| <i>Cognos</i> , modelo relacional, banco de dados <i>SQLServr</i> , <i>Oracle</i> , <i>DB2</i> . | <i>Hadoop</i> , modelos relacionais e <i>NoSQL</i> , bancos de dados não relacionais e <i>in-memory</i> . |
| Modelagem relacional/estruturada.  | Inclui também modelagem não estruturada. Modelagem analítica é essencial.                                 |
| Desenvolve <i>queries</i> estruturadas sobre dados passados.                                     | Cria perguntas e busca relacionamentos entre fatos aparentemente desconexos.                              |

FONTE: TAURION, 2013

Ribeiro (2014) ressalta que no viés da utilização de informação para obtenção de resultados o que se propõe com Big Data não é exatamente novidade:

Ao fazer um breve retrospecto nas pesquisas desenvolvidas na área, é possível observar que o desenvolvimento de ações investigativas em Sistemas de Apoio à Decisão (EIS), uso de Armazéns de Dados (*Data Warehouses* e *Data Marts*), aplicações para melhorar o Desempenho dos Negócios (*Business Intelligence*), soluções para Mineração de Dados (*Data Mining*), além de informação para planejamento estratégico, gestão de recursos informacionais e ativos de informação na Web, foram abordagens exploradas nos últimos anos pela Ciência da Informação. (RIBEIRO, 2014, p. 96)

Essas abordagens buscavam fazer uma análise preditiva de dados por meio de sua mineração. Tavares (2014) separa as etapas desses processos, começando com a utilização de técnicas estatísticas para o tratamento dos dados, os separando e reunindo em conjuntos, etapa de *Discovery*, seguinte a essa etapa os dados também podem ser categorizados de acordo com suas fontes de origem, etapa de *data preparation*. A etapa de definição dos modelos a serem aplicados na análise é denominada *model planning* e é necessária para a construção do grande conjunto de dados. Em sequência os dados são inseridos (etapa de *ingest*) a esse grande conjunto para possam ser analisados pelos modelos definidos. Finalmente, os resultados são apresentados com a utilização de ferramentas de visualização de acordo com o contexto de negócios em que estão. Essa estrutura de análise envolvia quantidades reduzidas de dados, o que tornava a atividade humana de correlacioná-los mais fácil, na transição para as quantidades envolvidas em Big Data surgem dificuldades. Devido a isso se torna necessário um novo perfil de profissional capaz de atuar com grandes volumes de dados.

Davenport, Barth e Bean (2012) listam como características desse novo perfil que passa a ser chamado de Cientista de Dados (*Data Scientist*):

[...] precisa ter raciocínio lógico apurado, conhecimento profundo de estatística e software dessa área, modelagem, domínio de sistemas computacionais e conhecimento do negócio e do mercado em que atuará. Além disso, será necessária a capacidade de comunicar seus insights a um pessoal com uma formação não técnica. (DAVENPORT; BARTH; BEAN, 2012)

Oliveira (2013) ressalta essa necessidade de conhecimentos matemáticos, treinamento avançado em estratégias para tratamento de grandes volumes de dados utilizando modelos matemáticos, hipóteses e técnicas de regressão. Brietman (2013) adiciona que esse

profissional deve ter capacidade de levantar os requisitos dos usuários, indo além das necessidades deles e buscando também as de outros envolvidos no ambiente sob análise.

Para exemplificar o tipo de questões que o profissional encara na execução dos projetos, Oliveira (2013) demonstra indagações que ele deve se fazer ao rever as etapas do seu trabalho:

Na fase de *Discovery*: Eu possuo o conhecimento suficiente do ambiente de dados e informação? Eu tenho informação suficiente para esboçar um plano analítico e compartilhar com meus pares? Eu consigo desenvolver trabalhos para organização para tipos de problemas? Categorizações e classificações de dados? Projeto de conjuntos (*clusters*) de dados? Eu consigo esboçar e realizar entrevistas para conhecer o contexto e domínio que será trabalhado? Eu posso identificar as diferentes fontes de dados?

Na fase de *Data Preparation* e *Model Planning*: Eu tenho um conjunto de dados que seja suficiente e de boa qualidade para iniciar a construção de um modelo? Eu tenho uma boa ideia sobre o tipo de modelo que vou testar? Eu posso refinar o modelo analítico? (OLIVEIRA, 2013 apud RIBEIRO, 2014, p.101)

Assim, inicialmente a relação entre a formação em Ciências Exatas e a atuação em *Big Data* se dá mais pela proximidade da construção das aplicações do que pela atividade operacional.

#### 4.1 Carreiras em *Big Data*

Conforme aumenta a adoção de *Big Data* pelas empresas cada vez mais vagas são criadas no setor. Com isso há uma profusão de novos títulos de cargos, com definições ainda tão diversas quanto as que buscam conceituar a própria área. Muitos títulos foram cunhados relacionando nomes cargos tradicionais à função que o profissional espera desempenhar no ciclo do projeto de *Big Data*. Comum a todos esses novos cargos ainda é a falta de consenso sobre suas competências desejadas, localização na hierarquia da organização e formar de metrificar sua performance.

McKendrick (2012) identifica sete principais cargos criados em *Big Data*, cada um com um título e uma atribuição diferente de acordo com a responsabilidade exercida no projeto. Essas funções passam variam de cargos com base mais técnica, responsáveis pelo estabelecimento da infraestrutura e programação dos bancos de dados, passando por cargos operacionais até cargos menos técnicos que ficam com a função de traduzir os resultados da análise de forma clara para os outros setores da empresa.

Entre essas novas carreiras geradas pelo *Big Data*, a de maior destaque e demanda no mercado é a do Cientista de Dados, descrita por Davenport (2012) como “a profissão mais

sexy do século XXI”, apontando o Cientista de Dados como um profissional de alto ranking, com treinamento e curiosidade para realizar descobertas no mundo do *Big Data*. O termo cientista se aplica ao cargo pois nesse campo também é necessário desenvolver as próprias ferramentas, pesquisar as fontes de dados a serem utilizadas, conduzir múltiplos experimentos e ser capaz de comunicar seus resultados.

No quadro a seguir, McKendrick (2012) relaciona esses cargos e sua principal atribuição para os projetos de *Big Data*:

Quadro 6 - Novos cargos criados pelo Big Data e suas atribuições profissionais:

| <b>Cargo</b>                          | <b>Atribuição</b>   |
|---------------------------------------|---|
| <i>Data Scientist</i>                 | Processamento de dados brutos; determinar o tipo de análise aplicada.                                       |
| <i>Data Architect</i>                 | Construção de modelos de dados; planejamento de que fontes de dados e ferramentas de análise utilizar.      |
| <i>Data Visualizer</i>                | Capacidade de contextualizar os dados de forma visual, facilitando seu entendimento.                        |
| <i>Data Change Agent</i>              | Um gestor capaz de angariar os recursos para gerar inovação e novos negócios.                               |
| <i>Data Engineer</i>                  | Responsável pela infraestrutura de <i>Big Data</i> ; desenvolvimento das ferramentas e sistemas.            |
| <i>Data Steward</i>                   | Administrador das fontes de dados; garante o fluxo de dados entre o repositório e as áreas da organização.  |
| <i>Data Virtualization Specialist</i> | Construção e suporte de bancos de dados virtuais para distribuição dos dados entre as fontes e seu destino. |

FONTE: MCKENDRICK (2012). Quadro produzido pelo autor.

Ledell (2015) indica a dificuldade em encontrar candidatos que reúnam todas as competências esperadas desse profissional apesar do número de cientistas de dados no mercado ter duplicado entre 2011 e 2015. Entre as principais competências listadas pela autora para cientistas de dados estão: Análise de Dados, R (linguagem de programação), *Python* (linguagem de programação), *Data Mining* e *Machine Learning*. No quadro a seguir,

estão listadas as bases técnicas segundo Ledell (2015) das carreiras de maior destaque em *Big Data*:

Quadro 7 – Principais Cargos de Ciência de Dados e suas competências

| <b>Time de Ciência de Dados</b> |   |
|---------------------------------|---|
| Analistas de Dados              | Conhecimento amplo e capacidade de usar ferramentas de análise de dados<br>Capacidade de comunicar uma narrativa utilizando dados |
| Arquiteto de Dados              | Formação em ciência da computação ou engenharia<br>Habilidades avançadas de programação e prática de <i>DevOps</i>                |
| Cientista de Dados              | Conhecimento de matemática/estatística e programação<br>Entender algoritmos de <i>Machine Learning</i>                            |

Fonte: Ledell (2015) – Quadro produzido pelo autor

Além das diferentes competências desses cargos, o momento de sua atuação no processo de análise dos dados também difere, como observado na figura seguinte:

Figura 4 – Etapas da análise de dados e atribuição de cada cargo



Fonte: Ledell (2015) – Figura traduzida pelo autor

## 4.2 Atuação de bibliotecários

Sendo um novo campo de atuação, *Big Data* deve despertar o interesse dos bibliotecários e outros profissionais da área de ciência da informação que busquem expandir suas possibilidades de posicionamento profissional no mercado. A presente escassez de mão-de-obra competente para o preenchimento das vagas criadas em *Big Data* se torna uma oportunidade para bibliotecários com perfis mais tecnológicos se capacitarem com o objetivo de ocupar um desses cargos.

Como observado, profissionais capazes de atuar em *Big Data* devem ser altamente capacitados, e essa capacitação demanda muito tempo enquanto a necessidade de implantação dos projetos é imediata. Uma vez que somente o treinamento de funcionários se torna inviável em uma linha de tempo curta e a oferta de profissionais dos campos de formação tradicionais para a área é restrita, uma solução oportuna para o estabelecimento ágil de uma iniciativa de *Big Data* é o recrutamento de candidatos de origens acadêmicas diversas.

Davenport (2014) já indicava que a formação de uma equipe multidisciplinar era ideal para a análise de dados, visto a dificuldade de um indivíduo só reunir todas as habilidades necessárias para a função. Gondim (2002) explica a exigência de profissionais com cada vez mais competências e formação técnica:

O desenvolvimento científico e tecnológico, suporte fundamental da globalização, aumenta a complexidade do mundo e passa a exigir um profissional com competência para lidar com um número expressivo de fatores. Este perfil profissional desejável está alicerçado em três grandes grupos de habilidades: i) as cognitivas, comumente obtidas no processo de educação formal (raciocínio lógico e abstrato, resolução de problemas, criatividade, capacidade de compreensão, julgamento crítico e conhecimento geral); ii) as técnicas especializadas (informática, língua estrangeira, operação de equipamentos e processos de trabalho) e iii) as comportamentais e atitudinais - cooperação, iniciativa, empreendedorismo [...], motivação, responsabilidade, participação, disciplina, ética e a atitude permanente de aprender a aprender. (GONDIM, 2002, p.300)

Para se estabelecer como um candidato desejável para preencher essa fatia do mercado, é preciso que o profissional apresente as competências requeridas pela vaga ofertada, reunindo em sua formação acadêmica e experiência de trabalho as habilidades necessárias para o desempenho da função. De acordo com as Diretrizes Curriculares Nacionais do Ministério da Educação (MEC), o Bibliotecário possui as habilidades e competências relacionadas no quadro a seguir:

Quadro 8 – Competências e Habilidades de Graduados em Biblioteconomia

(continua)

| <b>Competências e Habilidades dos Graduados em Biblioteconomia</b> |   |
|--|---|
| <b>Gerais</b>  | <ul style="list-style-type: none"> <li>• Gerar produtos a partir dos conhecimentos adquiridos e divulgá-los;</li> <li>• Formular e executar políticas institucionais;</li> <li>• Elaborar, coordenar, executar e avaliar planos, programas e projetos;</li> <li>• Utilizar racionalmente os recursos disponíveis;</li> <li>• Desenvolver e utilizar novas tecnologias;</li> <li>• Traduzir as necessidades de indivíduos, grupos e comunidades nas respectivas áreas de atuação;</li> <li>• Desenvolver atividades profissionais autônomas, de modo a orientar, dirigir, assessorar, prestar consultoria, realizar perícias e emitir laudos e pareceres;</li> <li>• Responder a demandas sociais de informação produzidas pelas transformações que caracterizam o mundo contemporâneo;</li> </ul> |
| <b>Específicas</b>   | <ul style="list-style-type: none"> <li>• Interagir e agregar valor nos processos de geração, transferência e uso da informação, em todo e qualquer ambiente;</li> <li>• Criticar, investigar, propor, planejar, executar e avaliar recursos e produtos de informação;</li> <li>• Trabalhar com fontes de informação de qualquer natureza;</li> <li>• Processar a informação registrada em diferentes tipos de suporte, mediante a aplicação de conhecimentos teóricos e práticos de coleta, processamento, armazenamento e difusão da informação;</li> </ul>  |

|  |  |
|--|--|
|  | <ul style="list-style-type: none"> <li>• Realizar pesquisas relativas a produtos, processamento, transferência e uso da informação.</li> </ul> |
|--|--|

Fonte: Brasil (2001). Quadro produzido pelo autor

Taurion (2013) identifica três perfis básicos de profissionais de *Big Data* e para cada um deles descreve suas habilidades e atuação:

a) Cientistas de dados. Profissionais capacitados em estatística, ciência da computação e/ou matemática capazes de analisar grandes volumes de dados e extrair deles insights que criem novas oportunidades de negócio; b) Analistas de negócio que, conhecendo bem o negócio em que atuam, consigam formular as perguntas corretas. Analisar as respostas e tomar decisões estratégicas que alavanquem novos negócios ou aumentem a lucratividade da empresa. Esta função tende a ser acoplada às funções do cientista de dados; c) Profissionais de tecnologia que cuidarão da infraestrutura e seu suporte técnico para suportar *Big Data*. O aparato tecnológico de *Big Data* não é muito comum em empresas tipicamente comerciais, pois demanda *expertise* em gerenciar *hardware* em *clusters* de alta performance e pensar volumes de dados significativamente maiores e muito mais variados que comumente se usam nos sistemas tradicionais. (TAURION, 2013, não paginado)

Desses perfis o menos técnico e que por consequência mais se aproxima do perfil de profissional da informação que os bibliotecários podem assumir é o de Analista de negócios, essa descrição é bem próxima do papel que os bibliotecários já exercem nos campos de pesquisa acadêmica. Com conhecimentos da área em que atua, o profissional de informação apoia a prática dos outros profissionais e possibilita a utilização de forma mais eficaz de tudo que as informações pode-lhe oferecer (CARMO, 2014, p. 16).

Entre as habilidades gerais esperadas de um Cientista de Dados esse novo profissional deve saber interpretar dados em sistemas autômatos e ser pró-ativo na interpretação desses elementos, mas também precisa conhecer suas limitações e trabalhar em equipe (ARRUDA et al, 2000). Essas habilidades são uma sequência natural dos deveres do profissional de informação destacadas por Wormell:

[...] facilitar o uso da informação; navegar por sistemas do conhecimento e fontes de informação; consultar e assessorar sobre 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)



Pomim (2000) indica outros pontos fundamentais que o bibliotecário deve assumir para incorporar essa postura de profissional da informação:

Realidade (conhecer a realidade em que está inserido); Identidade (conhecer quem é e onde quer chegar); Foco (conhecer os clientes, os concorrentes e saber quem ele quer ser); Processos (conhecer a matéria-prima, produtos e serviços); Recursos (conhecer as tecnologias e competências necessárias para o futuro) e Perspectivas (ter perspectiva de quem quer ser e quem a sociedade será no futuro). (POMIM 2000 apud CARMO, 2014, p. 40)

Gordon-Murnane (2012) relaciona diretamente as habilidades do bibliotecário como razão para sua atuação em *Big Data*. Para a autora o bibliotecário pode facilitar e permitir a descoberta e recuperação de dados, ser responsável pela manutenção da qualidade dos dados, agregar valor aos dados através da catalogação e metadados, e servir para a autenticação, gestão, arquivamento, representação e preservação de dados.

## 4 CONCLUSÃO

Através da pesquisa realizada é possível observar que *Big Data*, apesar de ainda ser tratado pela literatura da área de ciência da informação como uma novidade no campo de análise de dados, já é um mercado de trabalho estabelecido e de crescimento comparável ao do volume de dados que visa processar. Esse mercado ainda é bastante restrito aos profissionais de áreas como ciência da computação, matemática e engenharias, porém com a criação de cargos cada vez mais diversos, o desenvolvimento contínuo das ferramentas de *software* utilizadas e a necessidade de integrar os resultados obtidos pelo setor à empresa de forma a desenvolver novos produtos ou serviços, adequar processos e permitir melhorias na tomada de decisões será necessário expandir as competências da equipe, integrando profissionais de outras áreas de conhecimento. Assim, *Big Data* se apresenta como um campo capaz de absorver bibliotecários que busquem se capacitar para a atuação nele.

A conceituação de *Big Data* não possui um consenso, como uma área em plena evolução cada autor busca complementar definições anteriores e assim o termo está geralmente relacionado ao volume, variedade e velocidade dos dados obtidos e analisados. O ponto comum às definições observadas é que *Big Data*, através de sua capacidade de relacionar dados que antes não eram analisados por não serem estruturados, chega como um elemento que pode ser decisivo para que a empresa obtenha vantagem competitiva ao traduzir o resultado da análise massiva de dados em ajustes na tomada de decisões, criação de novos produtos e adaptação de suas estratégias em tempo real.

Identificadas as competências requeridas nos cargos já existentes foi possível traçar os perfis esperados de profissionais de *Big Data*: altamente capacitado em matemática, estatística, análise de dados e de negócios, capaz de entender linguagens de programação e trabalhar com ferramentas de processamento tais como *Apache Hadoop* e bancos de dados *NoSQL*. Esse é um profissional cuja disponibilidade no mercado para a atuação na área ainda é escassa.

Com essa atual escassez de profissionais de perfil tradicional para atuação na área, já que apesar dos campos de Ciência da Computação, Matemática e Engenharias formarem profissionais com as habilidades requeridas pela área de *Big Data* esses profissionais são absorvidos por outros campos de mercado. Assim, essa fatia do mercado se torna atraente para os bibliotecários cujo perfil de trabalho tenha um viés mais técnico.

Durante sua formação acadêmica, o bibliotecário aprende a lidar com dados de uma maneira que o coloca como o facilitador da mediação entre a necessidade do usuário e as

informações que mais se adequam a ela, sendo assim um agente da proliferação do conhecimento. De determinadas maneiras o profissional de *Big Data* age de forma semelhante, sendo responsável pela identificação dos insumos informacionais que podem ser extraídos de um montante de dados, sua recuperação e análise, e por fim a apresentação das informações extraídas no processo. Essa proximidade de atuação com dados, informação e conhecimento é o ponto de partida para a integração profissional das áreas.

Observando as competências curriculares dos graduados em Biblioteconomia é possível traçar paralelos entre a atuação do bibliotecário e a de alguns cargos de *Big Data*. Desenvolver e utilizar novas tecnologias; gerar produtos através dos conhecimentos adquiridos; interagir e agregar valor no processo de geração, transferência e uso da informação; trabalhar com fontes de informação de qualquer natureza são apenas algumas das muitas habilidades em comum aos profissionais das duas áreas diretamente ligadas à sua natureza.

Portanto, o bibliotecário que busca em *Big Data* um novo campo de atuação profissional já possui, devido à sua formação, certas competências requeridas para *Data Analysts*, *Data Scientists*, *Data Architects*, entre outros cargos. Aliando esse fator a uma continuidade de seu aprendizado voltado para as áreas de Matemática, Estatística e Programação em linguagens como *Python* e *R*, esse profissional pode complementar em seu currículo as habilidades requeridas para atuar plenamente em *Big Data*.

A diversidade de cargos e conceitos empregados ao se discutir a atuação profissional em *Big Data* mantém esse campo como uma área pouco aventurada por profissionais oriundos de outras áreas de conhecimento. Novas pesquisas que busquem elucidar meticulosamente as origens e atribuições dessa profusão de cargos podem evidenciar a disponibilidade de vagas de trabalho na área e atrair a atenção de profissionais que ainda não tinham o conhecimento do campo.

Conclui-se, então, que *Big Data* é um campo de atuação interessante para bibliotecários que queiram atuar com análise de dados. Os cargos ainda requerem que esse candidato se capacite em certas habilidades específicas que estão além do que é lecionado na graduação, porém observada a evolução das tecnologias da informação e comunicação e o crescimento do mercado a tendência esperada é que as ferramentas se tornem mais acessíveis possibilitando assim uma maior penetrabilidade dos bibliotecários nesse campo.

## REFERÊNCIAS

- ANGELONI, Maria Terezinha. Elementos intervenientes na tomada de decisão. **Ci. Inf**, v. 32, n. 1, p. 18, 2003
- ARRUDA, Maria da Conceição Calmon et al. Educação, trabalho e o delineamento de novos perfis profissionais: o bibliotecário em questão. In: **Ci. Inf.**, Brasília, v. 29, n. 3, p. 14-24, set./dez. 2000.
- BOYD, Danah.; CRAWFORD, Kate. Six provocations for big data. In: **SYMPOSIUM ON THE DYNAMICS OF THE INTERNET AND SOCIETY**, 2011, New York: Oxford Internet Institute's, 2011. p. 1-17. Disponível em: <[http://papers.ssrn.com/sol3/Papers.cfm?abstract\\_id=1926431](http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1926431)>. Acesso em: 06 out. 2015
- BRASIL, Ministério da Educação. Conselho Nacional de Educação. Diretrizes curriculares para os cursos de biblioteconomia. Disponível em <<http://portal.mec.gov.br/cne/arquivos/pdf/CES0492.pdf>>. Acesso em: 07 dez. 2017
- BRASIL, Presidência da República. Secretaria de Comunicação Social. *Pesquisa brasileira de mídia 2015: hábitos de consumo de mídia pela população brasileira*. Disponível em <<http://www.secom.gov.br/atuacao/pesquisa/lista-de-pesquisas-quantitativas-e-qualitativas-de-contratos-atuais/pesquisa-brasileira-de-midia-pbm-2015.pdf>>. Acesso em: 07 dez. 2017
- BRIETMAN, K. Big Data Overview. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2013.
- BRETERNITZ, Vivaldo José. SILVA, Leandro Augusto. Big Data: um novo conceito gerando oportunidades e desafios. **Revista Eletrônica de Tecnologia e Cultura**. 56 São Paulo, n. 13, p.106-113, out. 2013. Disponível em: <<http://revista-fatecjd.com.br/retc/index.php/RETC/article/view/74>>. Acesso em: 06 out. 2015
- CARMO, Rhuama Barbosa. **O bibliotecário brasileiro está sendo preparado para lidar com o big data?** FCI, UNB, Brasília, 2014
- CASTELLS, Manuel. **A sociedade em rede**. São Paulo: Paz e Terra, v.8, 1999.
- DA SILVA, João Vanderlei. Cluster: Possibilidades de eficiência e segurança. **Campo Digital**, v. 1, n. 1, 2006.
- DAVENPORT, Thomas. **Ecologia da Informação: por que só a tecnologia não basta para o sucesso na era da informação**. São Paulo: Futura, 1998.
- \_\_\_\_\_. **Big data no trabalho: derrubando mitos e descobrindo oportunidades**. Rio de Janeiro: Elsevier, 2014.
- DAVENPORT, Thomas; BARTH, Paul; BEAN, Randy. How big data is different. **MIT Sloan Management Review**, v. 54, n. 1, 2012
- DAVENPORT, Thomas; PATIL, D.J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard Business Review**, p. 70, 2012.

DE SOUZA, Terezinha Batista; CATARINO, Maria Elizabete; DOS SANTOS, Paulo Cesar. Metadados: catalogando dados na internet. **Transinformação**, v. 9, n. 2, 2012.

DZIEKANIAK, Gisele Vasconcelos; KIRINUS, Josiane Boeira. Web semântica. **R. Eletr. Bibliotecon. Ci. Inf.** Florianópolis, n. 18, p. 20-39, jul. 2004 Disponível em <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9n18p20/5471>>. Acesso em: 15 nov. 2017

FONSECA, João José Saraiva. **Metodologia da Pesquisa Científica**. 2002

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de pesquisa**. Plageder, 2009.

GONDIM, Sônia Maria Guedes. Perfil profissional e mercado de trabalho: relação com formação acadêmica pela perspectiva de estudantes universitários. *Estudos de Psicologia*, 7, p. 299-309

GORDON-MURNANE, Laura. **Big data**: a big opportunity for librarians. Disponível em <<https://drive.google.com/file/d/0B9AibUksRnAYWTVrSU0tRmp0d28/view>>. Acesso em: 07 dez. 2017

IBM. **O que é Big Data**. Disponível em: <[http://www.ibm.com/connect/ibm/attachments/C676885M30853H76/CAMSS\\_cuadernillo\\_portugues\\_virtual.pdf](http://www.ibm.com/connect/ibm/attachments/C676885M30853H76/CAMSS_cuadernillo_portugues_virtual.pdf)>. Acesso em: 06 out. 2015

ISACA. **Big data**: impactos e benefícios. 2013. Disponível em: <[http://www.isaca.org/Knowledge-Center/Research/Documents/Big-Data\\_whp\\_Por\\_0413.pdf](http://www.isaca.org/Knowledge-Center/Research/Documents/Big-Data_whp_Por_0413.pdf)>. Acesso em: 14 nov. 2016

LEDELL, Erin. Intro to data science with h2o. Youtube. 4 dez. 2015. Disponível em <<https://www.youtube.com/watch?v=Yz5x-YU7pik>>. Acesso em: 07 dez. 2017

MAGALHÃES, V.R.V. et al. O uso do Big Data na violação da privacidade dos usuários para estratégias de negócios. 2014. Disponível em <<http://docplayer.com.br/3664523-O-uso-do-big-data-na-violacao-da-privacidade-dos-usuarios-para-estrategias-de-negocios.html>>. Acesso em: 07 dez. 2017

MANYIKA, James et al. **Big data: The next frontier for innovation, competition, and productivity**. 2011. Disponível em: <[http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)>. Acesso em: 06 out. 2015

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. **Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. Elsevier Brasil, 2013.

MCKENDRICK, Joe. 7 new types of jobs created by big data. 2012. Disponível em <<http://www.zdnet.com/article/7-new-types-of-jobs-created-by-big-data/>>. Acesso em: 07 dez. 2017

MIRANDA, R. C. da R. O uso da informação na formulação de ações estratégicas pelas empresas. **Ciência da Informação**, Brasília, DF, v. 28, n. 3, p. 284-290, set./dez. 1999.

PENNY STOCKS LABS. **The Internet in real-time**. 2015. Não paginado. Disponível em: <<http://web.archive.org/web/20150621205645/http://pennystocks.la/internet-in-real-time>>. Acesso em: 21 nov. 2017.

PEREIRA, Vanessa Alves da Silva. **Big data: um estudo em gestão empresarial**. FACC, UFRJ, Rio de Janeiro, 2016

PJIPERS, Guus. **Information overload: a system for better managing every data**. New Jersey: Wiley, 2010.

PRODANOV, Cleber Cristiano; DE FREITAS, Ernani Cesar. **Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico-2ª Edição**. Editora Feevale, 2013.

QUIVY, Raymond; VAN CAMPENHOUDT, Luc. **Manual de investigação em ciências sociais**. 1998

RIBEIRO, Cláudio José Silva. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia**, João Pessoa/Marília, v. 1, n. 1, p. 96-105, jan. 2014. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/itec/article/view/19380/11156>>. Acesso em: 06 out. 2015.

RUSSO, M. **Fundamentos de Biblioteconomia e Ciência da Informação**. Rio de Janeiro: E-papers, 2010. 177 p.

SILVA, Débora. Big data analytics: domine o conceito e aumente a eficiência da sua indústria. Disponível em: <<http://www.logiquesistemas.com.br/blog/big-data-analytics/>>. Acesso em: 07 dez. 2017

SIMON, Phill. **Too Big To Ignore: the business case of big data**. New Jersey: Wiley, 2013.

SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos; BARACHO, Renata Maria Abrantes. Ciência da Informação em transformação: Big Data, Nuvens, Redes Sociais e Web Semântica. **Ciência da Informação**, [S.l.], v. 42, n. 2, ago. 2015. ISSN 1518-8353. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2283>>. Acesso em: 06 out. 2015

TAURION, Cezar. Big data. Rio de Janeiro: **Brasport**, 2013. Não paginado.

TAVARES, Elaine. **Big data: desafios e oportunidades**. Palestra apresentada no 2º EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2014. Disponível em: <[http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao\\_Elaine\\_Tavares.pdf](http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao_Elaine_Tavares.pdf)>. Acesso em: 07 dez. 2017

VIEIRA, Marcos Rodrigues. et. al. **Bancos de dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto de Big Data**. In: Simpósio Brasileiro de Bancos de Dados, 2012, Mato Grosso. Minicurso... Mato Grosso: Universidade

Federal de Mato Grosso, 2012. Disponível em:

<[http://data.ime.usp.br/sbbd2012/artigos/pdfs/sbbd\\_min\\_01.pdf](http://data.ime.usp.br/sbbd2012/artigos/pdfs/sbbd_min_01.pdf)> Acesso em: 06 out. 2015

WARDEN, Pete. **Big data glossary**. O'Reilly Media, Inc., 2011.

WERTHEIN, Jorge. A sociedade da informação e seus desafios. **Ciência da Informação, Brasília**, v. 29, n. 2, p. 71-77, 2000.