

Big Data: Um novo campo de atuação para bibliotecários

Big Data: A new field of action for librarians

Luiz Claudio Rezende Reis

Universidade Federal do Rio de Janeiro, Brasil
luizcrreis@gmail.com

Maria Irene da Fonseca e Sá

Universidade Federal do Rio de Janeiro, Brasil
mariairene@facc.ufrj.br

Resumo

Na chamada sociedade da informação, os dados são produzidos em elevados volumes e se tornam a principal matéria-prima do contexto econômico. Nesse cenário, o Big Data busca suprir a necessidade de extrair valor desses dados. O desenvolvimento de hardwares e softwares capazes de armazenar e processar altas quantidades de dados com rapidez, tornou possível o cruzamento dos mesmos e sua análise, transformando dados não-estruturados em informações úteis. Assim, o Big Data pode propiciar um novo campo de atuação para os profissionais de Ciência da Informação. O presente trabalho observa, através de revisão da literatura existente e da pesquisa bibliográfica de escopo exploratório e qualitativo, as conceituações de Big Data, suas ferramentas de análise e 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 a aplicação de Big Data como uma nova oportunidade de atuação para os bibliotecários que desejem se capacitar no campo de análise de dados.

Palavras-chave: Big Data, Análise de Dados, Bibliotecário, Ciência da Informação.

Abstract

In the so-called information society, data are produced in high volumes and become the main raw material of the economic context. In this scenario, Big Data seeks to address the need to extract value from this data. The development of hardware and software capable of storing and processing high amounts of data quickly made it possible to cross-examine and analyze them, transforming unstructured data into useful information. Thus, Big Data can provide a new field of action for information science professionals. The present study observes, through a review of the existing literature and the bibliographical research of exploratory and qualitative scope, the concepts of Big Data, its tools of analysis and the search to delimit the profile of the professional of the area, identifying the positions created in the labor market and attributions. By linking the skills of Big Data professionals and librarians, the paper presents the application of Big Data as a new opportunity for action for librarians who wish to be trained in the field of data analysis.

Keywords: Big Data, Data Analysis, Librarian, Information Science.

1. Introdução

A estrutura atual da sociedade tem sido alterada 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 (TIC) 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: 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 nesse recente paradigma técnico-econômico que as sociedades contemporâneas se configuram e que vem sendo chamado de "sociedade da informação". Também é Castells (1999, como citado em 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 o aumento da velocidade em que os novos produtos de tecnologia são pesquisados e desenvolvidos, acarretando também uma diminuição de seus custos e tornando-os mais acessíveis para a população em geral. Atualmente, as TIC têm uma acentuada penetrabilidade na sociedade. No Brasil, há mais de 152 milhões de computadores em uso segundo o 26º Relatório Anual de Tecnologia da Informação da Fundação Getúlio Vargas (Brasil, 2015). Somados a eles há também 154 milhões de *smartphones* em uso para uma população de 206 milhões de habitantes, ou seja, um alcance de 3 dispositivos para cada 2 habitantes. 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* representam a 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. Tal fato gerou a 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 um rastro digital. Segundo Eric Schmidt, então presidente do

Google, diariamente eram criados 5 *exabytes* de dados no mundo. Taurion (2013) estimou que no ano de 2015 tenham sido alcançados 7,9 *zetabytes* 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*.

Esses dados que formam o rastro digital são de fácil recuperação, porém devido à dificuldade do processamento em massa e ao 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 significantes para empresas e governos, estima-se que apenas 1% desses dados seja efetivamente analisado (Breitman, 2013).

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. Para Manyika (2011: 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 et. al (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: 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: 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. 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 gerar dados e os recuperar 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: 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% 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, *Big Data* tem atraído a atenção dos profissionais em Ciência da Informação e áreas afins.

Ribeiro (2014) afirma que *Big Data*: “[...] desperta, na atualidade, o interesse, e até mesmo o fascínio, para todas as pessoas que têm algum envolvimento com atividades para Gestão da Informação”. A recente profusão de trabalhos e artigos que discorrem sobre o tema são indicativos 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: 85) 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.

Assim, o presente trabalho busca num primeiro momento estabelecer o contexto do 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. São apresentados os conceitos de *Big Data*, suas aplicações e ferramentas para melhor entendimento do tema. Em seguida, é discutido o perfil profissional e é realizada 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.

2. Metodologia

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.

Com o objetivo e o intuito de explicitar o tema e demonstrar a 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: 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: 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: 32).

Assim, foi realizada uma revisão bibliográfica na qual foram utilizadas diversas fontes de informações sobre o assunto, como: livros, reportagens, notícias e postagens em *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* à 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. 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*.

3. Resultados

Davenport (1998: 19) define dados como elementos brutos, sem significado. Já Setzer os define como “uma abstração formal que pode ser representada e transformada por um computador” (Setzer, 2001, como citado em Russo, 2010: 15). Russo (2010: 15) expande a definição: 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: 285).

Drucker define informação como “dados dotados de relevância e propósito” (Drucker, 1988, como citado em Davenport, 1998: 18). Miranda (1999: 285) avança o termo como “dados organizados de modo significativo, sendo subsídio útil à tomada de decisão”. Russo (2010: 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 diferencia 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: 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 dele, que codifica, decodifica, distorce e usa a informação de acordo com suas características pessoais, ou seja, de acordo com seus modelos mentais. (Angeloni, 2003: 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 capturá-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: 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 de conteúdo digital por *smartphones* e outros dispositivos conectados à internet, geram uma grande quantidade de metadados. A quantidade disponível e sua relação indissociável do cotidiano do indivíduo tornam os metadados a mais rica fonte para a análise e extração de informações sobre consumidores e seus hábitos, despertando o interesse de empresas que buscam atingir vantagem competitiva através de ferramentas de análise de dados. Os metadados são, portanto, o principal insumo do *Big Data*.

Davenport (2014: 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: 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: 6).

Laney (Laney, como citado em 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).

O autor 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 definições, 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 na tomada de decisões pelos seus gestores, adaptando estratégias para obter melhores desempenhos e vantagem competitiva sobre seus concorrentes.

Davenport (2014: 4) estabelece as diferenças concretas entre o *Big Data* e ferramentas de análise de dados tradicionais, como descrito no quadro 1.

Quadro 1 – O big data e o analytics tradicional

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

Fonte: Davenport (2014) – Quadro produzido pelos autores

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”.

O avanço tecnológico de infraestrutura física, como a miniaturização de dispositivos de armazenamento de dados, o aumento da capacidade de discos rígidos e cartões de memória, o aumento da banda de tráfego de dados via internet e o barateamento dos componentes de processamento, possibilitou às organizações possuírem aparato de *hardware* suficiente para administrar e processar um fluxo enorme de dados com custo aceitável. Em paralelo, as iniciativas de *software* de código aberto desenvolveram uma plataforma de programas para gerir o processamento em massa. Para tal, é necessário que os programas sejam capazes de lidar com o alto volume de dados, encaixando-os em modelos estatísticos que permitam otimizar sua leitura e prever padrões.

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: 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 prevenindo 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 quadro 2.

Quadro 2 – 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 de 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 big data (Exemplos: Python, Pig Hive)
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: 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 necessários sistemas de grande poder de processamento de forma eficiente e escalável com suporte a dados complexos, semiestruturados ou não-estruturados. A escolha do banco de dados *NoSQL* a ser utilizado é relacionada ao modelo de representação de dados que se quer.

O quadro 3 lista os principais produtos *NoSQL* disponíveis, de acordo com seu modelo de classificação de dados.

Quadro 3 – Produtos de NoSQL

Modelo de Dados	Produto NoSQL
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, Scaliem, 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, BangDB.</i>

Fonte: (Vieira et al, 2012) - Quadro produzido pelos autores

Pereira (2016) ressalta que o grande desafio para os gestores que pretendem trabalhar 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 a responsável por extrair dos dados as informações que podem gerar valor para a empresa.

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*” (2013). 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 fala das 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 como citado em Ribeiro, 2014).

Assim, o profissional de *Big Data* (cientista de dados) 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 diferencia do perfil do profissional de *Business Intelligence* (BI), mais voltado para representações gráficas sobre dados passados, como é mostrado no quadro 4.

Quadro 4 - Diferenças de competências entre Analistas de BI e Cientistas de Dados

ANALISTA DE BI	PROFISSIONAL DE BIG DATA
<i>Cognos</i> , modelo relacional, banco de dados <i>SQL Servr</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 queries estruturados 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: 96).

Essas abordagens buscavam fazer uma análise preditiva de dados por meio de mineração. Tavares (2014) discorre sobre 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*, em seguida os dados 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*) nesse grande conjunto para que 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. Para tal 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 a necessidade de fortes conhecimentos em matemática e estatística, 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 como citado em Ribeiro, 2014: 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.

Conforme aumenta a adoção de *Big Data* pelas empresas, mais vagas são criadas no setor. Desta forma, 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 de cargos tradicionais à função que o profissional deve desempenhar no ciclo do projeto de *Big Data*. Comum a todos esses novos cargos ainda é a falta de consenso sobre as competências desejadas, localização na hierarquia da organização e formar de avaliar seu desempenho.

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. Esses cargos vão de uma 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.

No quadro 5 é descrito como McKendrick (2012) relaciona esses cargos e sua principal atribuição para os projetos de *Big Data*.

Quadro 5 - Novos cargos criados pelo Big Data e suas atribuições profissionais

Cargo	Atribuição
<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 pelos autores

Dentre essas novas carreiras geradas pelo *Big Data*, as de maior destaque e demanda no mercado são: *Data Analyst* (Analista de Dados), *Data Architect* (Arquiteto de Dados) e *Data Scientist* (Cientista de Dados). Essa última foi chamada por Davenport e Patil (2012) de “a profissão mais sexy do século XXI”, descrevendo 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.

Ledell (2015) indica a dificuldade em encontrar candidatos que reúnam todos 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 6 estão listadas as bases técnicas, segundo Ledell (2015), das carreiras de maior destaque em *Big Data*:

Quadro 6 – Principais cargos e suas competências

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

Fonte: Ledell (2015) – Quadro produzido pelos autores

Além das diferentes competências desses cargos, o momento de sua atuação no processo de análise dos dados também difere.

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.

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.

Davenport (2014) indica que a formação de uma equipe multidisciplinar é ideal para a análise de dados, dada a dificuldade de um indivíduo só reunir todas as habilidades necessárias à função. Gondim (2002) explora a exigência de profissionais com cada vez com 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: 300).

Assim, 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 7.

Quadro 7 – Competências e habilidades de graduados em Biblioteconomia

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

Fonte: Brasil (2001) - Quadro produzido pelos autores

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 alavancem 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 o que 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 podem-lhes oferecer (Carmo, 2014: 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).

Pomim (2000) indica alguns pontos fundamentais para serem incorporados pelo 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 como citado em Carmo, 2014: 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. Conclusões

Através da pesquisa, foi possível concluir 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. No entanto, a partir da criação de cargos cada vez mais diversos, do

desenvolvimento contínuo das ferramentas de *software* utilizadas, da necessidade de integrar os resultados obtidos pelo setor à empresa, de adequar processos e de 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* ainda 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, visa ser um elemento decisivo para que a empresa obtenha vantagem competitiva ao traduzir o resultado da análise de dados em ajustes na tomada de decisões, criação de novos produtos e adaptação das estratégias em tempo real.

Identificadas as competências requeridas nos cargos já existentes foi possível traçar o perfil esperado do profissional 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*.

A revisão de literatura sobre o tema, no escopo da Biblioteconomia, identificou um crescente interesse sobre a atuação do bibliotecário no contexto do *Big Data*, lidando com dados e informação.

Durante sua formação acadêmica, o bibliotecário aprende a lidar com dados de uma maneira que o coloca como facilitador da mediação entre a necessidade do usuário e as informações que mais se adequam a ela. De muitas 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 pôr 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 dos diversos 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.

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 que *Big Data* é um campo de atuação interessante para bibliotecários que queiram atuar em análise de dados. Os cargos ainda requerem que esse candidato se capacite em certas habilidades específicas que estão além do que é lecionado, atualmente, 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 Bibliográficas

- ANGELONI, Maria Terezinha. (2003). Elementos intervenientes na tomada de decisão. *Ci. Inf.*, v. 32, n. 1, p. 18.
- ARRUDA, Maria da Conceição Calmon et al. (2000). Educação, trabalho e o delineamento de novos perfis profissionais: o bibliotecário em questão. *Ci. Inf.*, Brasília, v. 29, n. 3, p. 14-24, set./dez. 2000.
- BOYD, Danah.; CRAWFORD, Kate. (2011). Six provocations for big data. *Symposium On The Dynamics Of The Internet And Society*, New York: Oxford Internet Institute's. p. 1-17. Recuperado de http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1926431
- BRASIL, Ministério da Educação. Conselho Nacional de Educação. Diretrizes curriculares para os cursos de biblioteconomia. Recuperado de <http://portal.mec.gov.br/cne/arquivos/pdf/CES0492.pdf>
- BRASIL, Presidência da República. Secretaria de Comunicação Social. Pesquisa brasileira de mídia 2015: hábitos de consumo de mídia pela população brasileira. Recuperado de <http://www.secom.gov.br/atuacao/pesquisa/lista-de-pesquisas-quantitativas-e-qualitativas-de-contratos-atuais/pesquisa-brasileira-de-midia-pbm-2015.pdf>
- BRIETMAN, K. (2013). Big Data Overview. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro.
- BRETERNITZ, Vivaldo José. SILVA, Leandro Augusto. (2013). Big Data: um novo conceito gerando oportunidades e desafios. *Revista Eletrônica de Tecnologia e Cultura*. São Paulo, n. 13, p.106-113, out. 2013. Recuperado de <http://revista-fatecjd.com.br/retc/index.php/RETC/article/view/74>
- CARMO, Rhuama Barbosa. (2014). O bibliotecário brasileiro está sendo preparado para lidar com o big data? FCI, UNB, Brasília.
- CASTELLS, Manuel. (1999). A sociedade em rede. São Paulo: Paz e Terra, v.8.
- DA SILVA, João Vanderlei. (2006). Cluster: Possibilidades de eficiência e segurança. *Campo Digital*, v. 1, n. 1.
- DAVENPORT, Thomas. (1998). *Ecologia da Informação*: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura.
- DAVENPORT, Thomas. (2014). *Big data no trabalho*: derrubando mitos e descobrindo oportunidades. Rio de Janeiro: Elsevier.
- DAVENPORT, Thomas; BARTH, Paul; BEAN, Randy. (2012). How big data is different. *MIT Sloan Management Review*, v. 54, n. 1.
- DAVENPORT, Thomas; PATIL, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, p. 70.
- DE SOUZA, Terezinha Batista; CATARINO, Maria Elizabete; DOS SANTOS, Paulo Cesar. (2012). Metadados: catalogando dados na internet. *Transinformação*, v. 9, n. 2.

- DZIEKANIAK, Gisele Vasconcelos; Kirinus, Josiane Boeira. (2004). Web semântica. *R. Eletr. Bibliotecon. Ci. Inf.* Florianópolis, n. 18, p. 20-39, jul. 2004. Recuperado de <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9n18p20/5471>
- FONSECA, João José Saraiva. (2002). *Metodologia da Pesquisa Científica*.
- GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. (2009). *Métodos de pesquisa*. Plageder.
- GONDIM, Sônia Maria Guedes. (2002). 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. (2012). Big data: a big opportunity for librarians. Recuperado de <https://drive.google.com/file/d/0B9AibUksRnAYWTVrSU0tRmp0d28/view>
- IBM. (2015). O que é Big Data. Recuperado de http://www.ibm.com/connect/ibm/attachments/C676885M30853H76/CAMSS_cuadernillo_portugues_virtual.pdf
- ISACA. (2013). Big data: impactos e benefícios. 2013. Recuperado de http://www.isaca.org/Knowledge-Center/Research/Documents/Big-Data_whp_Por_0413.pdf
- LEDELL, Erin. (2015). Intro to data science with h2o. Youtube. 4 dez. 2015. Recuperado de <https://www.youtube.com/watch?v=Yz5x-YU7pik>
- MAGALHÃES, V.R.V. et al. (2014). O uso do Big Data na violação da privacidade dos usuários para estratégias de negócios. Recuperado de <http://docplayer.com.br/3664523-O-uso-do-big-data-na-violacao-da-privacidade-dos-usuarios-para-estrategias-de-negocios.html>
- MANYIKA, James et al. (2011). Big data: The next frontier for innovation, competition, and productivity. Recuperado de http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- MCKENDRICK, Joe. (2012). 7 new types of jobs created by big data. Recuperado de <http://www.zdnet.com/article/7-new-types-of-jobs-created-by-big-data/>
- MIRANDA, R. C. da R. (1999). 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. (2015). The Internet in real-time. Recuperado de <http://web.archive.org/web/20150621205645/http://pennystocks.la/internet-in-real-time>
- PEREIRA, Vanessa Alves da Silva. (2016). *Big data: um estudo em gestão empresarial*. FACC, UFRJ, Rio de Janeiro.
- PJIPERS, Guus. (2010). *Information overload: a system for better managing every data*. New Jersey: Wiley.

- PRODANOV, Cleber Cristiano; DE FREITAS, Ernani Cesar. (2013). *Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico-2ª Edição*. Editora Feevale.
- RIBEIRO, Cláudio José Silva. (2014). Big Data: os novos desafios para o profissional da informação. *Informação & Tecnologia*, João Pessoa/Marília, v. 1, n. 1, p. 96-105, jan. 2014. Recuperado de <http://www.ies.ufpb.br/ojs2/index.php/itec/article/view/19380/11156>
- RUSSO, M. (2010). *Fundamentos de Biblioteconomia e Ciência da Informação*. Rio de Janeiro: E-papers. 177 p.
- SIMON, Phill. (2013). *Too Big To Ignore: the business case of big data*. New Jersey: Wiley.
- TAURION, Cezar. (2013). *Big data*. Brasport.
- TAVARES, Elaine. (2014). Big data: desafios e oportunidades. Palestra apresentada no 2º EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. Recuperado de http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao_Elaine_Tavares.pdf
- VIEIRA, Marcos Rodrigues. et. al. (2012). Bancos de dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto de Big Data. Simpósio Brasileiro de Bancos de Dados, Mato Grosso. Minicurso... Mato Grosso: Universidade Federal de Mato Grosso. Recuperado de http://data.ime.usp.br/sbbd2012/artigos/pdfs/sbbd_min_01.pdf
- WERTHEIN, Jorge. (2000). A sociedade da informação e seus desafios. *Ciência da Informação*, Brasília, v. 29, n. 2, p. 71-77.