

ARTIGO

Recebido em:  
20/03/2017

Aceito em:  
26/05/2017

*Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 22, n.50, p. 128-143, set./dez., 2017. ISSN 1518-2924. DOI: 10.5007/1518-2924.2017v22n50p128

## O Profissional da Informação na Era do Big Data

### *The Information Professional in the Age of Big Data*

Caio Saraiva CONEGLIAN ([caio.coneglian@gmail.com](mailto:caio.coneglian@gmail.com))\*

Paula Regina Ventura Amorim GONÇALVES ([paulaventuramorim@gmail.com](mailto:paulaventuramorim@gmail.com))\*\*

José Eduardo Santarém SEGUNDO ([santarem@usp.br](mailto:santarem@usp.br))\*\*\*

\* Doutorando do Programa de Pós-Graduação em Ciência da Informação – UNESP.

\*\* Professor do Departamento em Ciência da Informação – UEL.

\*\*\* Professor do Programa de Pós-Graduação em Ciência da Informação – UNESP. Professor do Departamento de Educação, Informação e Comunicação, da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – USP.

#### Resumo

Os profissionais da informação são desafiados a trabalhar com a rapidez imposta tanto pela geração de informação e de dados, quanto pela variedade de fontes em que elas se encontram; assim, sua atuação se apresenta como fundamental ao buscar a sinergia entre as necessidades de gestores e a expertise dos profissionais da computação. Nesse cenário o objetivo deste trabalho é identificar e discutir a intersecção entre as funções executadas pelos administradores (gestores), cientistas da computação e dos profissionais da informação em ambientes de *Big Data Analytics*. A metodologia utilizada é de caráter qualitativo, sendo a pesquisa exploratória, visto que partiu dos estudos da Ciência da Informação, do Profissional da Informação e do Big Data, para estabelecer um elo entre tais conceitos. Como resultado demonstra-se e descreve-se a atuação do profissional da informação em todas as fases da análise dos dados. Verificou-se uma intersecção entre a proposta de relacionamento entre os atores “administrador”, “profissional da informação” e “profissional da computação”, com o chamado cientista de dados, em que devido a diferentes competências exigidas do cientista, conclui-se que tanto o profissional da informação quanto o profissional da computação podem se especializar nesta área, visto que, atuando em campos diferentes, mas complementares, fornecerão informações em cenários de *Big Data* auxiliando a tomada de decisões e agregando valores para as organizações.

**Palavras-chave:** Profissional da informação. Cientista de dados. Big data. Big data analytics.

#### Abstract

Information professionals are challenged to work with the speed imposed both by the generation of information and data, as well as by the variety of sources in which they are found, thus their performance is fundamental in seeking the synergy between the needs of Managers and the expertise of computer professionals. In this scenario, the objective of this work is to identify and discuss the intersection between the functions performed by administrators (managers), computer scientists and information professionals in Big Data Analytics environments. The methodology used is of a qualitative nature, being the exploratory research, since, starting from the studies of Information Science, Information Professional and Big Data, to establish a link between these concepts. As a result, the performance of the information professional is demonstrated and described in all phases of data analysis. There was an intersection between the proposal of relationship between the actors "administrator", "information professional" and "computer professional", with the so-called data scientist, where due to the different skills required of the scientist, it is concluded that both the information professional and the computer professional can specialize in this area, since, working in different but complementary fields, will provide information in scenarios of Big Data helping decision making and adding values to the organizations.

**Keywords:** Information professional. Data scientist. Big data. Big data analytics.



v. 22, n. 50, 2017.  
p. 128-143  
ISSN 1518-2924



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

## 1 INTRODUÇÃO

O Big Data caracteriza-se essencialmente pelo volume dos dados gerados e disponíveis, pela velocidade com que estes devem ser tratados e apresentados e pela variedade de fontes onde eles se encontram. Essa conjuntura tem provocado uma revolução no modo como as análises de dados estão ocorrendo, tanto no âmbito organizacional, quanto no acadêmico. Segundo Mayer-Schönberger e Cukier (2013) o *Big Data* representa “[...] uma nova fonte de valor econômico e informação”, visto que os dados têm se tornado uma das principais fontes de rendas e informações das organizações.

Nesse contexto, profissionais estão se especializando em análises de dados no contexto do *Big Data*, chamado de *Big Data Analytics*, em busca de atuarem nesse novo segmento. Os profissionais, conhecidos como cientistas de dados, estão gradualmente se tornando mais importantes nas empresas, nos governos e nas instituições de ensino, visto que proveem o fornecimento de uma perspectiva mais exata, como, por exemplo, no que diz respeito à concorrência e aos padrões de consumo.

Os cientistas de dados são em sua maioria profissionais oriundos das Ciências Matemáticas e Computacionais, preocupados em sua essência na criação de algoritmos e de técnicas computacionais de mineração e de análises de dados. No entanto, esses profissionais, em sua maioria, não estão capacitados para ter uma perspectiva focada nas necessidades informacionais de gestores e administradores.

Na busca de preencher esse *gap* existente nos atuais cientistas de dados, os profissionais da informação ampliam sua área de atuação, ao se apresentarem como fundamentais no *Big Data Analytics*, buscando a sinergia entre as necessidades dos gestores e a *expertise* dos profissionais da computação.

Frente a essa demanda, a *International Federation of Library Associations* (IFLA) apontou algumas tendências que irão alterar os ambientes de informação, e que estão inevitavelmente relacionadas ao contexto atual, que se mostra diretamente ligado ao *Big Data*. Dentre essas tendências, a IFLA aponta que a economia global será alterada pelas novas tecnologias, além de haver tanto uma expansão quanto uma redução de quem tem acesso às informações. (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2016).

As questões levantadas pela IFLA inserem o profissional da informação como agente central das mudanças que estão ocorrendo atualmente, sendo necessário que esse profissional possa contribuir nessa disruptura. Dessa forma, as análises de dados estão sendo fundamentais para as transformações que estão ocorrendo, desafiando o profissional da informação a ser um ator ativo nesse processo, que possa tornar as evoluções atuais mais intensas e benéficas.

Assim, o presente trabalho tem como objetivo identificar e discutir a intersecção entre as funções executadas pelos administradores (gestores), cientistas da computação e profissionais da informação em ambientes de *Big Data Analytics*, buscando a compreensão sobre o papel do profissional da informação nesse novo campo de estudo.

## 2 O PROFISSIONAL DA INFORMAÇÃO E A CIÊNCIA DA INFORMAÇÃO

Com o desenvolvimento tecnológico tornou-se possível que imagem e som, ou a combinação dos dois, resultassem em meios para que transmissão de ideias, conhecimentos e informações ocorresse mais facilmente. Esse foi o cenário que permitiu o surgimento dos chamados “meios de comunicação de massa” ou simplesmente *mass media*. O processo evolutivo causado a partir dessa transformação pôde ser percebido com os progressos ocorridos durante passagem da comunicação oral para o registro do conhecimento em diferentes suportes

informativos (comunicação escrita), que tiveram como consequência a imprescindibilidade do uso das tecnologias nos procedimentos de comunicação.

As modificações ocorridas inseriram a Ciência da Informação (CI) no centro desse debate e, segundo Le Coadic (1996, p. 21), a CI afirma-se como uma ciência social aplicada, pois está preocupada em “[...] esclarecer um problema social concreto, o da informação.” E afirma que “[...] desenvolveu-se em função dessa necessidade e foi de certa forma dirigida, e até mesmo financiada por ela.” Ainda para o autor, esta ciência ampara-se na tecnologia e tem como objetivo o estudo da natureza, da gênese e dos efeitos da informação; para tanto se propõe a fazer, “A análise dos processos de construção, comunicação e uso da informação. E a concepção dos produtos e sistemas que permitem sua construção, comunicação, armazenamento e uso” (LE COADIC, 1996, p. 26).

No campo da informação, a biblioteconomia, a museologia, a documentação e o jornalismo foram as quatro disciplinas que surgiram tendo seus interesses focados no suporte da informação e não na informação em si (LE COADIC, 1996).

Tendo como foco a representação, a organização e o armazenamento da informação, para que se torne possível a recuperação, o acesso, o uso e reuso e a disseminação, a Ciência da Informação constatou que as Tecnologias de Informação e Comunicação (TIC) possibilitariam a concepção de sistemas e serviços que permitissem a comunicação, o uso e o armazenamento da informação.

Santos e Vidotti (2009, p.1) apontam duas tendências no estudo da natureza do tratamento e da gestão da informação e do conhecimento na Ciência da Informação:

A primeira, predominantemente na Organização da Informação, focaliza os procedimentos de análise, de síntese, de condensação, de representação e de recuperação do conteúdo informacional e a reflexão sobre organização do conhecimento, seus desdobramentos epistemológicos e instrumentais. Estes, por hipótese, constituem as bases de tratamento e da representação da informação para a recuperação. A segunda, predominantemente nas Tecnologias de Informação e Comunicação (TICs), é marcada nas estruturas e modelos de sistemas computacionais atuantes nos processos de produção, de armazenamento, de preservação, de representação, de recuperação, de acesso, de (re)uso e de disseminação de conteúdos informativos.

Diante do exposto, pode-se afirmar que a área de atuação da Ciência da Informação é ampliada para os ambientes informativos digitais, principalmente nos ambientes Web. Hoje, o contexto sociocultural é evidenciado pela rapidez com que as TIC estão se desenvolvendo, o que tem gerado diversas transformações no universo informacional que abrange todas as áreas do conhecimento. Dessa forma, o *status* alcançado pela informação tem se tornado cada vez mais valoroso.

Marcondes (2001) aponta que a informação se tornou o insumo que proporciona atividades nos processos de ensino e de aprendizagem, nos processos de aquisição e construção de conhecimento para pesquisa científica e tecnológica, nos processos culturais e ainda em setores ou atividades em que se faz necessário o conhecimento humano.

Complementarmente, Rosenfeld (2002) pontua que, no mundo dos negócios, empresários constatarem que os problemas de negócios são na verdade problemas de informação; portanto, os profissionais da informação podem colaborar na resolução desses problemas. Santos (2000, p. 103) afirma ainda que nos dias atuais “[...] o contorno da economia é definido pela quantidade de informação possuída, veiculada e disseminada, podemos identificar a informação como matéria-prima do mundo contemporâneo, juntamente com as tecnologias disponíveis.”

Como uma das principais características da Ciência da Informação é sua natureza interdisciplinar (SARACEVIC 1995; LE COADIC, 1996; BORKO, 1968), frente à vertiginosa evolução da Sociedade da Informação e a sua consolidação nos dias atuais, também foram provocadas alterações no perfil dos profissionais da informação; assim, segundo Rezende (2002, p. 120), “cada vez mais as organizações passaram a ter em seu quadro de pessoal não apenas especialistas técnicos, mas também especialistas em trabalhar a informação.”

Faria et al. (2005, p. 31) apontam o profissional da informação como sujeito “[...] ativo e agente criativo, capitalizando sua competência informacional para as estratégias da organização em que atua.” O que se presencia nos dias atuais é a necessidade de que o profissional da informação amplie sua atuação profissional para além dos muros das instituições, e verifique que na ambiência digital sua expertise e seus saberes são essenciais.

Segundo entendimento de Valentim (2004, p.140), o profissional da informação é o “[...] indivíduo que recebeu formação específica para trabalhar com dados, informação e conhecimento, bem como sua mediação”. A autora também ressalta que o tripé informação, tecnologia da informação e telecomunicações não somente muda a sociedade, como também muda suas demandas. Assim, o profissional da informação deverá assumir:

[...] uma postura investigativa e crítica, de modo que possa assumir essas mudanças sociais de forma natural. [...] a globalização, fenômeno mundial que afeta profundamente as relações sociais e de trabalho, cria novas situações para o profissional da informação que atua com dados, informação e tecnologia. (VALENTIM, 2004, p. 119)

Profissionais da informação podem estar vinculados a qualquer etapa do ciclo de informação, pois possuem capacidade para trabalhar com eficiência e eficácia no tratamento da informação em qualquer tipo de organização. (PONJUÁN DANTE, 2000).

Esses profissionais que têm a informação como insumo em seu trabalho, utilizando-a de maneira estratégica em seu ambiente, desenvolvem recursos informacionais e estabelecem serviços, colaborando com a missão e consequentemente com o crescimento da organização em que atuam. Dentre os profissionais da informação estão incluídos: bibliotecários; gestores do conhecimento; gestores web; *information brookers* e consultores (ABELS et al., 2003).

No ano de 2003, foi feita a revisão pela *Special Libraries Association – SLA* do seu relatório publicado no ano de 1996, intitulado *Competencies for Information Professionals of the 21st Century* (ABELS et al., 2003). Nesse documento, destacaram-se as principais competências, tanto profissionais quanto pessoais, que os bibliotecários, ou qualquer profissional da informação que tenha como foco atender as necessidades informacionais dos usuários, devem possuir.

Suas competências profissionais abarcam a gestão de serviços e a aplicação de ferramentas e tecnologias de informação, a gestão da informação e a gestão de recursos de informação; para tanto é essencial em seu âmbito de atuação que o profissional tenha conhecimento e domínio na utilização da tecnologia de informação e de recursos informacionais e que inclusive tenha competência para que sejam feitas avaliações e triagens, tendo conhecimento especializado dos negócios de sua organização e de seu cliente. Além disso, o profissional deve oferecer serviços de apoio e treinamento de excelência em sua instituição. Também deve utilizar a tecnologia de informação apropriada para adquirir, organizar e disseminar a informação. Ademais, esse profissional deve desenvolver produtos de informação especializados para uso interno ou externo à instituição como, por

exemplo: desenvolver bancos de dados, criar *home pages*, arquivo de texto, entre outros. Por fim, ser membro integrante da equipe gerencial e consultor da organização nos assuntos relativos à informação.

Entre suas competências pessoais o profissional da informação deve entre outras coisas: ter visão globalizada; buscar desafios; investir em novas oportunidades; comunicar-se com eficácia; criar parcerias e alianças e construir um ambiente tendo como base o respeito e a confiança.

Como se pode constatar as tecnologias promoveram novas atividades, reestruturando, assim, profissões, para que estas se adequem às novas demandas da sociedade, que nesse momento transformam o conhecimento e a informação em capital.

As mudanças ocorridas universalmente nas esferas política, econômica, científica e tecnológica levaram a muitas discussões em torno da geração, armazenamento e disseminação da informação e, conseqüentemente, sobre a atuação dos profissionais que fazem da informação seu objeto de prática. Santos (2002, p. 104) faz a caracterização, no nosso atual cenário social, de seus atores:

[...] sujeitos pós modernos não tem identidade fixa, essencial ou permanente, já que estão sujeitos à formação e transformação contínuas em relação às formas em que os sistemas culturais os condicionam. Tais sujeitos encontram-se inseridos em uma rede que abrange a formação do conhecimento a partir das relações sociais, como uma teia de relações que não se inicia numa estrutura linear ou mesmo acadêmica.

Nesse contexto, informação perdeu sua localização espaço-temporal e em dias atuais passou a ser objeto de desejo de todo cidadão, isto é, ela não se restringe mais aos círculos dos intelectuais e acadêmicos. Nos dias atuais o significado de informação está relacionado, ainda que inconscientemente, à velocidade e à tecnologia. Com o aperfeiçoamento das telecomunicações e com o advento das novas tecnologias da informação e comunicação só existe o tempo real e as barreiras espaciais são rompidas.

O profissional da informação está diante do fenômeno *Big Data*, e este cenário é caracterizado por volumes de dados extremamente densos e que necessitam de competências, habilidades e ferramentas para que essa informação possa ser encontrada; para que isso seja possível, ela necessita ser tratada, analisada e disponibilizada em tempo hábil.

### **3 BIG DATA**

As evoluções tecnológicas marcaram o início do século XXI e não somente estão cada vez mais presentes em todos os setores da sociedade e em todas as tarefas diárias realizadas pelos indivíduos, como também geram quantidades de dados extraordinárias. Em um passado bastante recente, as transformações que estão sendo vivenciadas nessa nova realidade contemporânea eram inimagináveis. Estudos já demonstram que a quantidade de dados gerada por toda a humanidade até o início desse milênio, atualmente é gerada em número equivalente em poucos dias. (SAGIROGLU; SINANC, 2013).

Apesar da imprecisão da afirmação citada, fica claro que uma parcela considerável desses dados gerados deve ser armazenada, podendo ser assim recuperada, com maior ou menor dificuldade. Nesse sentido, há uma extensa seara de utilizações possíveis desses dados, seja no âmbito organizacional, na descoberta de padrões e comportamentos dos consumidores, no âmbito governamental, oferecendo serviços mais eficientes para os cidadãos e no âmbito acadêmico,

possibilitando a identificação e o estudo de fenômenos até então desconhecidos pela ciência.

Diante das circunstâncias apresentadas, cunhou-se o termo *Big Data* para descrever tal cenário. Assim, diversos autores têm se arriscado na difícil tarefa de apresentar uma definição que conceitualize de maneira consistente tal termo. Destacam-se Zikopoulos et al. (2011), que relataram sinteticamente que *Big Data* se aplica a informações que não podem ser processadas e analisadas pelos métodos tradicionais.

Taurion (2013) expande a visão apresentada, ao relatar que as diversas fontes de informações geram grandes volumes de dados, permitindo uma transformação na maneira como os negócios são realizados. Assim, tornou-se possível identificar comportamentos imperceptíveis até então, que, caso sejam explorados, poderão fornecer ferramentas estratégicas às empresas, para compreender as necessidades e os desafios que o cenário econômico apresenta.

Complementando as afirmações realizadas, destacam-se as principais características pertencentes ao fenômeno do *Big Data*, chamados de 3 Vs: **Volume**, **Velocidade** e **Variedade**. McAfee e Brynjolfsson (2012) caracterizaram o *Big Data* pelos 3 Vs, discorrendo a respeito de cada uma dessas características:

O **Volume** dos dados vem apresentando um crescimento exponencial ao passar dos anos. Conforme relatado por Taurion (2013), no ano de 2000 somente 25% dos dados se encontravam disponíveis digitalmente, enquanto atualmente a quantidade disponível está próxima dos 100%. Outro dado apontado pelo autor, que demonstra as transformações associadas ao volume das informações, é que cerca de 90% dos dados armazenados no ano de 2013 foram criados nos últimos dois anos.

Segundo McAfee e Brynjolfsson (2012), a segunda característica, **Velocidade**, configura-se pela rapidez na geração, no processamento e na análise de dados; assim, em algumas aplicações, a velocidade torna-se mais importante que o volume dos dados. Isso ocorre pelo dinamismo dos processos realizados pelas organizações, em que a tomada de decisão está vinculada diretamente com o momento e a velocidade em que esta é tomada. Nesse cenário, é fundamental que os gestores tenham acesso às informações em tempo real, efetivando as decisões de maneira precisa.

A terceira característica, **Variedade**, trata de diferentes fontes informacionais que hoje são utilizadas, como e-mail, mídias sociais, blogs, e em formatos variados, como imagens, textos e vídeos. Como consequência, os processos analíticos não podem ficar presos a dados estruturados, gerados pelos sistemas de informação tradicionais, necessitando utilizar a Web nas análises de dados.

Outros autores inserem outras duas características ao *Big Data*, o Valor e a Veracidade, denominados como os 4º e 5º Vs (KAKHANI; KAKHANI; BIRADAR, 2013; KATAL; WAZID; GOUDAR, 2013). Nessa visão, o valor é inserido como uma consequência das análises de *Big Data*, em que se torna possível adquirir informações com grande valor agregado, como por exemplo as tendências de mercado. Já a veracidade surge como um requisito para se obterem análises mais precisas, pois é necessário que sejam utilizados dados verdadeiros para que as conclusões não sejam equivocadas.

Neste estudo considerou-se como características essenciais do *Big Data*, os 3 Vs, originalmente apontados por McAfee e Brynjolfsson (2012). Tal concepção é utilizada por não se entender que veracidade e valor sejam aspectos intrínsecos do *Big Data*, e sim atributos necessários para qualquer sistema de informação que necessite realizar uma análise de dados, enquanto o volume, a velocidade e a variedade se mostram fundamentais dentro do contexto do fenômeno *Big Data*.

Um outro aspecto inerente a tal fenômeno trata de como as características discutidas anteriormente (3 Vs) podem alterar a essência dos processos de análises de dados. Mayer-Schönberger e Cukier (2013) indicam que a principal mudança

realizada nos processos analíticos se dá em função de uma mudança quantitativa, provocando, assim, uma mudança qualitativa, ou seja, o aumento da quantidade de dados faz com que as informações se tornem cada vez mais precisas, reduzindo as incertezas e auxiliando a tomada de decisões.

Esse processo pode ser melhor entendido quando é feita a comparação de como os processos de análises de informações eram realizados anteriormente ao contexto do *Big Data*, em que se utilizavam somente amostras dos dados nas análises, o que contribuía para a obtenção de resultados equivocados, caso tal amostra não representasse adequadamente a população dos dados. Deve-se ressaltar que o uso de amostras é fruto de um momento em que não havia outros meios de realizar as análises, pois tanto a captação e o processamento, quanto o armazenamento dos dados eram bastante dispendiosos, diante do problema de não haver tecnologias disponíveis que tratassem de grandes quantidades de dados.

Dentro dessa perspectiva surge um dos campos mais promissores dessa época, chamada de *Big Data Analytics*, ou análises de dados no contexto do *Big Data*. Russom (2011) destaca que o *Big Data Analytics* é composto por técnicas analíticas avançadas, operando sobre grandes conjuntos de dados. O autor aponta os motivos que unem *Big Data* com a análise de dados, destacando-se, dentre tais motivos: a geração de amostras estatísticas colossais; dados desorganizados contendo uma grande carga de informações importantes; e processos analíticos em grandes conjuntos de dados apresentando resultados com grande impacto para as organizações.

Diante do exposto, compreende-se que o fenômeno do *Big Data* está causando uma revolução em análises de dados, pois ao unir capacidade de armazenamento com processamento, torna-se possível obter informações que antes eram inalcançáveis pelos processos tradicionais. O *Big Data Analytics* é o caminho natural dessa evolução, por ser uma área que define como os processos devem ser executados.

Diversos autores buscam descrever como tais análises ocorrem em cenários de *Big Data*, relatando os passos e os modos que devem ser seguidos para que uma análise de dados gere valor para as organizações. Taurion (2013) afirma que os processos de análises de dados no contexto do *Big Data* estão divididos em quatro fases: coleta dos dados; limpeza, formatação e validação desses dados; integração e agregação dos dados obtidos em diversas fontes; e a fase analítica. Assim, é possível identificar elementos que tratam da utilização de dados confiáveis nos processos analíticos. Essa constatação corrobora reflexões em torno das fontes utilizadas nas análises, em que o descuido na escolha delas pode prejudicar os resultados obtidos.

Na busca de compreender com maior profundidade como as análises extraem, de grandes conjuntos de dados, informações com alto valor para as organizações, Mayer-Schönberger e Cukier (2013) apontam a existência de três cenários em que essa extração de valor ocorre: reutilização básica; fusão de banco de dados; e a utilização de um mesmo dado em diversos cenários. Ao descrever essas maneiras de extração de valor, os autores indicam que, para encontrar informações relevantes, há diversos fatores e cuidados que devem ser tomados, gerando conhecimento fundamental para as organizações, especialmente pelo momento histórico vivido, em que são necessárias cada vez mais informações nos processos decisórios.

Em outra perspectiva, Bugembe (2016) enumera cada fase do processo de análise dos dados, que ele aponta como as fases para obtenção do valor dos dados. Essa literatura será a base considerada neste trabalho, como as fases para a análise dos dados, que será utilizada para delinear como o profissional da informação poderá trabalhar dentro da análise de dados. Bugembe (2016) considera que há seis fases para a obtenção de valor dos dados: 1) fonte, 2) captura e armazenamento, 3) processamento e fusão, 4) acesso, 5) análise e 6) exposição.

A fase das fontes é a primeira, pois será nesse momento que serão definidas as bases de dados utilizadas durante a análise. Por tal motivo, a escolha das fontes pode levar ao sucesso ou ao fracasso de uma análise, pois não se consegue obter algum valor de dados incoerentes com o que se deseja, mesmo que os algoritmos de análises sejam eficientes. Ainda no que tange à escolha das fontes, é fundamental escolher as fontes externas com precisão, mas permitindo que em momento futuro outras bases possam ser consideradas. Isso ocorre porque não se sabe com certeza onde todos os dados necessários se encontram, mas é necessário apontar um caminho a ser seguido nas fontes que deverão ser utilizadas.

Na segunda fase, devem-se definir questões relativas à captura e ao armazenamento dos dados. Neste quesito, vale destacar que, apesar da capacidade de armazenamento atualmente ser bastante elevada, há um limite de armazenamento e de processamento, que conduz à necessidade de delimitar o que será capturado e armazenado.

Nessa fase, duas questões centrais devem centrar os processos realizados: o que capturar e quando capturar. A primeira questão é relativa a limitar e escolher o que dentre as fontes será obtido, enquanto a segunda questão trata da temporalidade em que os dados serão coletados, que pode variar de acordo com cada situação.

A fase de processamento e fusão é relativa ao momento em que os dados capturados são limpos, fundidos e transformados em dados estruturados, que permitirá a reutilização e fácil uso durante as análises. Dois processos guiam essa fase, a limpeza e a transformação dos dados. A limpeza tem a função de deixar os dados com fácil interpretação, estruturando-os. Já a transformação trata dos passos realizados para a fusão dos dados, além dos demais processos técnicos.

No passo de acesso, o objetivo é disponibilizar os dados para os gestores, administradores e marketing, visando obter retornos quanto aos dados obtidos pelos atores que são os maiores conhecedores do negócio em questão. "Em uma organização em que os usuários de negócios estão cientes de quais informações estão disponíveis, como se acessa aos dados que eles exigem e os usam para influenciar um resultado, então o valor dos dados aumenta rapidamente." (BUGEMBE, 2016, p. 62).

Na fase de análise em si, serão aplicados algoritmos e análises que levarão a resultados acerca de padrões dos dados. Essa análise reúne estatísticas, fórmulas e mineração de dados, que serão processados em sua maioria em arquiteturas complexas e distribuídas. Nesse momento, será possível realizar previsões, previsões, apontar comportamento do consumidor, entre outras questões.

Por fim, a fase de exposição, que objetiva apresentar os resultados obtidos com as análises. Essa fase apesar de trivial é essencial, pois caso a exposição não seja bem-sucedida, não tendo uma clareza do valor agregado, todos os momentos anteriores perdem a utilidade. Há uma relação desse momento com a cultura organizacional da empresa, que deverá estar aberta a receber os resultados obtidos com as análises de dados.

Em consonância aos modelos expostos, Ribeiro (2014) relata que o processo analítico de pequenos conjuntos de dados não é demasiadamente complexo para que um humano o realize, uma vez que não existem tantas correlações existentes entre os dados. Contudo, com grandes quantidades de dados (*Big Data*), o processo de correlacionamento torna-se mais complexo, pelas diversas fontes gerando dados paralelamente. O autor afirma que em função dessa nova conjuntura, surge o profissional chamado de cientista de dados (*Data Scientist*).

Há diversas definições sobre o perfil e as funções desenvolvidas por um cientista de dados (DAVENPORT; PATIL, 2012; VAN DER AALST, 2014). Em síntese, este profissional trabalhará na busca de encontrar informações relevantes espalhadas pelo descomunal volume de dados, atuando nas fases apontadas por

Taurion (2013), descritas anteriormente, mas também em processos de compreensão dos requisitos feitos pelos gestores, para criar mecanismos que atendam com eficiência as necessidades informacionais dos administradores.

A inserção dos estudos de *Big Data* no campo da Ciência da Informação pode ser compreendida quando se recorre ao próprio conceito de informação, presente na tripla classificação proposta por Buckland (1991): informação como processo, informação como conhecimento e informação como coisa. Destaca-se que neste trabalho o conceito utilizado é o de “informação como coisa”, visto que se refere aos recursos informacionais em suas características físicas, como documentos representados por alguma forma física como a imagem, o texto ou a sua representação, em que se enquadram os dados dentro do cenário do *Big Data*.

Diante disso, a bibliografia aponta o *Big Data* como uma grande evolução que, ao mesmo tempo, desafia os profissionais a se reinventarem no uso das tecnologias. Em contrapartida, verifica-se a exigência de profissionais que tenham um maior conhecimento e maiores habilidades para lidar com essas grandes quantidades de dados. Dessa forma, constatou-se que o profissional da informação encontra um novo campo de atuação, onde sua expertise se faz urgente.

#### **4 PROCEDIMENTOS METODOLÓGICOS**

A metodologia utilizada neste trabalho é de caráter qualitativo, sendo uma pesquisa exploratória, que partiu dos estudos da Ciência da Informação, do profissional da informação e do *Big Data*, para estabelecer um elo entre os conceitos.

Para o desenvolvimento dessa metodologia, primeiramente recorreu-se à literatura de profissionais da informação, para fornecer pressupostos teóricos das principais atividades desenvolvidas por esse profissional, buscando traçar um panorama das suas qualificações e expertises.

Na sequência, buscou-se subsídios teóricos visando apontar como os pesquisadores de *Big Data* estavam concebendo os principais processos da análise de dados. Nesta fase buscou-se elencar os pontos que eram considerados essenciais para as análises de dados no contexto do *Big Data*, identificando as peculiaridades deste cenário.

A partir desses elementos teóricos, sobre o profissional da informação e do *Big Data*, identificou-se em quais tópicos o profissional da informação poderia contribuir nas análises de dados no contexto do *Big Data*. Como ponto de partida, utilizou-se a proposta de Bugembe (2016), que separa em seis fases o processo de análise de dados. Em cada uma dessas fases, identificou-se como o profissional da informação pode atuar, por meio da utilização de alguns textos que relatam as atividades que esse profissional é capaz de desempenhar.

Por fim, apresenta-se uma proposta de relacionamento do profissional da informação com os demais atores das análises de dados, o profissional da computação e o gestor, e com as fontes informacionais externas e internas. Nesta proposta, busca-se demonstrar a viabilidade da inserção do profissional da informação nas análises de dados no contexto do *Big Data*, apontando que tal profissional não é um substituto do profissional da computação nessa função, mas, sim, um elemento que torna o processo mais aprimorado.

Na próxima seção, apresentam-se as discussões obtidas neste trabalho, que possibilitaram traçar relacionamentos e identificar as funções do profissional da informação no âmbito das análises de dados em cenários de *Big Data*.

#### **5 DISCUSSÕES: os desafios do profissional da informação e o *Big Data***

O surgimento do fenômeno do *Big Data* tem provocado mudanças no modo como os processos analíticos e de tomadas de decisão são concebidos,

especialmente em sistemas de informação. Identifica-se que a complexidade existente na manipulação e nas análises do grande volume de dados existente tem exigido que profissionais de diversas áreas mesquem suas habilidades, na busca de soluções mais eficazes.

Destacam-se duas carreiras que estão apresentando uma importância singular no cenário apresentado: o profissional da computação, na busca de desenvolver algoritmos e ferramentas eficientes na manipulação de grandes volumes de dados e os administradores, buscando encontrar padrões e comportamentos, bem como meios que permitam que as organizações se destaquem frente à concorrência.

Dentro da conjuntura exposta, um desafio latente que permeia o desenvolvimento de sistemas de informação é a eficiência com que eles interagem com seus usuários. Por esse motivo, bem como pela necessidade de compreender as necessidades informacionais dos gestores, profissionais da informação, com o passar dos anos, vêm se tornando cada vez mais presentes em grandes corporações, como especialistas capazes de auxiliar nessa árdua tarefa.

Ribeiro (2014) relata que o profissional da informação deve ser incluído nos processos de *Big Data*, apontando alguns momentos do processo de análises de dados em que tal profissional pode contribuir, destacando principalmente a fase de coleta dos dados.

Porém o papel do profissional da informação pode ser visualizado em outros momentos da análise de dados, quando se tem a perspectiva de que o profissional da informação pode contribuir em processos técnicos do tratamento dos dados. Dessa forma, utilizou-se a proposta de Bugembe (2016) acerca da análise de dados, que foi exposta anteriormente, para compreender o papel do profissional da informação dentro da análise de dados. Tal proposta foi utilizada pois se mostra a mais completa, e acaba por contemplar características tanto da proposta de Taurion (2013) quanto da de Mayer-Schönberger e Cukier (2013).

Nesse momento se faz importante destacar que este trabalho visa apresentar recomendações idealizadas de como o profissional da informação pode atuar no âmbito de análises de dados em cenários de *Big Data*, embasado pelas principais teorias acerca de *Big Data* e de Ciência de Dados.

Assim, a partir da proposta de Bugembe (2016), identificou-se o papel do profissional da informação em cada uma das seis fases da análise de dados. A figura 1 apresenta tais fases com o respectivo papel do profissional em cada uma delas.

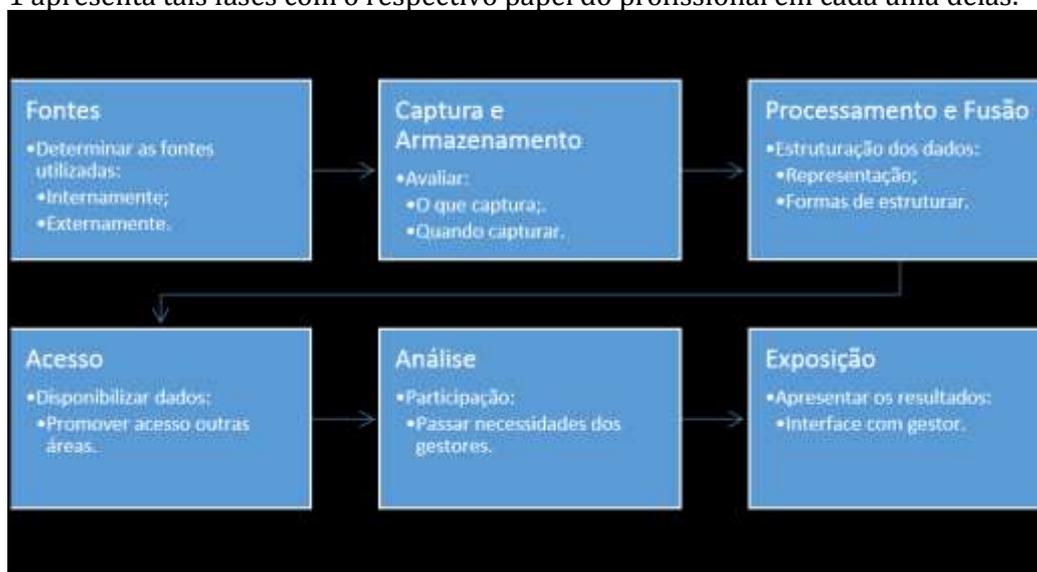


Figura 1: Papel do profissional da informação em cada momento da análise de dados.

Fonte: Elaborado pelos autores.

As funções descritas na figura 1 demonstram que o profissional da informação possui papel relevante em cada uma das fases da análise de dados. Buscando descrever como o profissional da informação pode realizar essas funções, a seguir discorre-se com detalhes sobre cada fase enumerada.

O primeiro passo da análise é relativo à definição das fontes utilizadas. Nesse momento, o profissional da informação deverá encontrar e apontar as bases de dados internas, como por exemplo os bancos de dados financeiros e de históricos de clientes, e as bases de dados externas, como informações espalhadas na Web da concorrência e dados diversos de clientes, entre outros.

O profissional da informação é capacitado para compreender o cenário e identificar as melhores fontes de informação, mostrando-se como um profissional essencial nesse ambiente, ao localizar e escolher as principais fontes para a realização das análises de dados. Nesse sentido, Barbalho e Rozados (2008), ao apontarem as competências do bibliotecário, explicitam que este profissional é capacitado para lidar, entre outras questões, com fontes de informação de qualquer natureza, bem como utilizar e disseminar fontes, produtos e recursos de informação.

Na sequência, na fase de captura e armazenamento, o profissional da informação deverá trabalhar em cooperação com o profissional da computação, para delinear os dados que serão necessários para a análise, além de ser responsável por apontar a temporalidade em que a informação deverá ser obtida. O profissional da informação, por ter um conhecimento maior das fontes, conseguirá apontar isso com mais precisão do que o profissional da computação.

A fase de processamento e de fusão contempla um importante processo computacional, em que ocorrerá uma estruturação dos dados, para possibilitar a fusão da informação. Nesse momento, o profissional da informação atua principalmente para a estruturação dos dados, isso porque ele possui capacitação para identificar estruturas capazes de converter uma massa de dados em informações estruturadas como, por exemplo, estruturas de chave-valor e metadados.

Outro ponto relativo à fase de processamento e fusão refere-se ao fato do profissional da informação aplicar conceitos de representação da informação nesses dados. Nesse sentido, a representação poderá inserir uma quantidade maior de informações para a realização da fusão dos dados. Como relatado por Alvarenga et al. (2003, 23), o profissional da informação é apontado como o responsável pela representação da informação, relatando ainda que as informações que serão representações: “[...] compreendem compactações que tentam descrever as características do documento, refletindo sua origem e conteúdo, facilitando sua recuperação.”

Já na fase de acesso, ocorre uma interação com os gestores, para possibilitar que esses profissionais contribuam na definição dos dados que serão tratados, apontando possíveis mudanças e *feedbacks* acerca dos processos realizados anteriormente.

A análise dos dados é uma fase precipuamente do profissional da computação, em que o profissional da informação poderá ser um ator secundário, com a função de participar mais para direcionar os rumos e as variáveis que estão sendo consideradas na análise. Tal fato ocorre, pois o profissional da computação é evidentemente o responsável pela construção de algoritmos e pela definição das estatísticas utilizadas.

Por fim, na fase de exposição, o profissional da informação deverá, juntamente com o profissional da computação, apontar os resultados obtidos. Como o profissional da informação se mostra como uma interface entre o profissional da computação e o gestor, é dele a principal função nessa fase, por conseguir expor com mais propriedade os resultados atingidos com a análise de dados.

A partir do olhar sobre os diversos momentos em que o profissional da informação atua, bem como das interações necessárias com os demais profissionais, verifica-se que a diversidade se faz extremamente necessária quando se trata de *Big Data Analytics*, pois requer-se habilidades que um único profissional não dispõe. Faz-se primordial que existam diversos profissionais trabalhando em conjunto; desta feita, os cientistas de dados serão oriundos das mais variadas disciplinas e áreas do conhecimento.

A partir da definição do papel do profissional da informação na análise de dados, identificou-se um cenário em que esse profissional pode estar inserido e contribuindo para que as análises de *Big Data* se tornem mais eficientes. Tal cenário é ilustrado por meio da Figura 2, em que se verificou a existência de três profissionais-chaves no processo, o administrador (ou gestor), o profissional da informação e o profissional da computação.

Cabe ressaltar que não se utiliza o termo cientista de dados para um tipo de profissional, pois se compreende que tanto o profissional da informação, quanto o profissional da computação estão exercendo as funções do cientista de dados, porém, dividindo tais funções conforme as suas habilidades.

A Figura 2 contempla os diversos atores e os ambientes informacionais envolvidos em cenários de análises de dados no contexto do *Big Data*, sintetizando o papel do profissional da informação dentro do cenário das análises de dados, no que tange à interação com os demais atores.

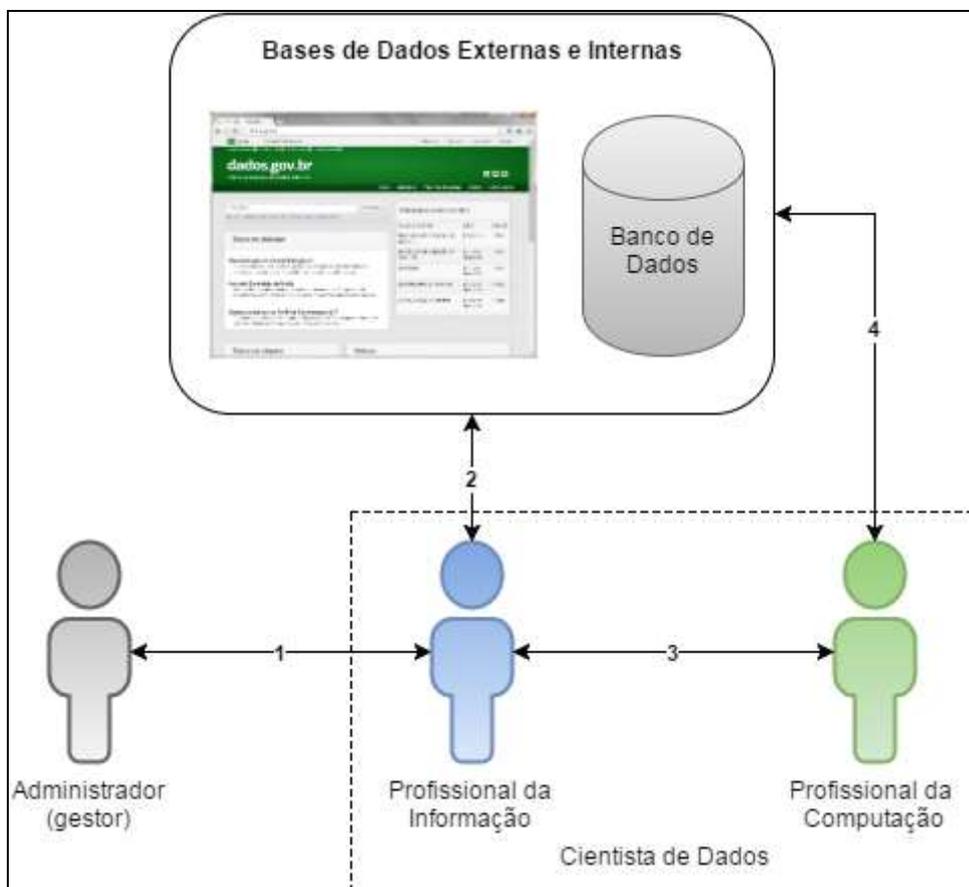


Figura 2: **Relacionamento entre os diversos atores em cenários de análise de dados no contexto do *Big Data*.**

Fonte: Elaborado pelos autores

A seguir, detalham-se as quatro fases numeradas nas setas da figura 2, explicando o relacionamento proposto.

A seta número um (1) indica o relacionamento entre o administrador e o profissional da informação. Nessa etapa, a principal tarefa do profissional da informação é compreender as necessidades informacionais do administrador, identificando quais e de que tipo são as informações de que ele necessita. A importância desse passo está em constatar as necessidades de um domínio, tarefa que por vezes um profissional da computação não está hábil para executar; nesse caso o profissional da informação deve ser inserido, para evitar que se criem ruídos que dificultem a interação entre o administrador e o profissional da computação.

Em seguida, no passo indicado pela seta dois (2), o profissional da informação necessita compreender os ambientes disponíveis que possam atender as necessidades informacionais apresentadas pelos gestores.

Posteriormente, surge o processo demonstrado por meio da seta de número três (3), que apresenta o relacionamento entre o profissional da informação e o profissional da computação, em que o segundo precisa compreender as necessidades apresentadas pelo administrador e receber as informações de quais são as fontes, identificadas pelo profissional da informação, que serão utilizadas na análise de dados.

A diferenciação das ações executadas pelos dois profissionais ocorre devido às formações distintas que o profissional da informação e o da computação possuem, em que o processo de identificar e tratar as fontes e as necessidades informacionais dos usuários fica como responsabilidade do primeiro profissional, e questões de construção de algoritmos e técnicas computacionais fica a cargo do profissional da computação.

Por fim, a seta quatro (4) indica o relacionamento entre o profissional da computação com as bases de dados, sinalizando propriamente os processos de construção dos algoritmos computacionais e de estatísticas, que buscam extrair valor nos processos de *Big Data Analytics*.

A esquematização realizada baseia-se, em essência, na interdisciplinaridade que os novos desafios exigem, fazendo com que a existência de um único profissional, que lida com todas as questões nas análises de dados em cenário de *Big Data*, perde o sentido, pelo fato de existir uma complementação entre as habilidades que os diversos profissionais possuem. Dessa forma, a divisão de tarefas deve ser distribuída de acordo com a formação de cada profissional.

Outro fator relevante se dá no relacionamento entre a esquematização e o profissional chamado de cientista de dados. Destaca-se, nessa relação, a questão do profissional cientista de dados não ser um único indivíduo que resolve todas as questões que estão envolvidas em um processo de análise de dados, e mais especificamente na questão do *Big Data Analytics*, mas sim, como relatado anteriormente, trata-se de diversos profissionais se relacionando.

Fica evidente que tanto o profissional da informação quanto o da computação deverão se capacitar para serem cientistas de dados, que, dentro de sua formação, deverão se desenvolver e se tornar especializados nas tarefas que tratam de análises de dados, sendo profissionais que irão trabalhar em conjunto nas resoluções dos problemas que envolvem *Big Data Analytics*.

## **6 CONSIDERAÇÕES FINAIS**

A evolução das tecnologias permitiu um avanço na sociedade, que era impensável há poucas décadas, trazendo evolução tanto em questões de hardware (equipamentos físicos) quanto em software (lógicas que conduzem o funcionamento dos equipamentos físicos), possibilitando evoluções científicas e comerciais que revolucionaram a maneira como o ser humano se comunica e interage com o mundo, com os outros e com ele mesmo.

Como consequência dessa evolução, nota-se que tanto as instituições quanto a academia estão tendo um novo olhar na condução de análises de dados, devido ao fenômeno conhecido como *Big Data*. Essa nova forma de realizar as análises está sendo chamada de *Big Data Analytics*, que está possibilitando a identificação de padrões e comportamentos que até então estavam imperceptíveis para as análises realizadas integralmente por humanos, ou com o auxílio de máquinas mais rudimentares, quando comparadas com as utilizadas atualmente.

Nesse contexto, verifica-se a necessidade de uma maior interdisciplinaridade na busca de obter melhores resultados quando se trata dessa área chamada *Big Data Analytics*. Tal questão se faz presente pela necessidade de haver profissionais trocando suas habilidades e competências, para atingir resultados mais significativos nas análises de dados. Cabe ressaltar que o presente trabalho verificou uma intersecção entre a proposta de relacionamento entre os atores “administrador”, “profissional da informação” e “profissional da computação”, com o chamado cientista de dados, e devido as diferentes competências exigidas do cientista, concluiu-se que tanto o profissional da informação quanto o profissional da computação podem se especializar nessa área, ainda que atuando em campos diferentes, mas complementares, conforme a proposta do presente trabalho.

A pesquisa apresentada neste artigo evidencia que o profissional da informação é elemento fundamental, assim como o profissional da computação, atuando nessa nova realidade denominada *Big Data Analytics*, em que se valerá de suas expertises e competências como parte integrante e imprescindível de sua atuação, contribuindo para o aprimoramento das análises, que fornecerão informações em cenários de *Big Data*, auxiliando a tomada de decisões e agregando valores para as organizações.

A utilização desta proposta parte de uma concepção teórica das expertises do profissional da informação frente aos processos do *Big Data*, apontando caminhos para a inserção real deste profissional no domínio em questão. Esta proposta, ainda que teórica, explícita, na prática, em quais pontos específicos o profissional da informação pode vir a contribuir efetivamente, em um momento em que se vê uma necessidade latente de especialistas no trato da informação.

Assim, por meio dos pontos apresentados neste trabalho, é possível identificar em quais processos esse profissional está capacitado para atuar. Dessa maneira, as discussões sobre a necessidade do profissional da informação neste (novo) campo de trabalho podem se aprofundar, visto que há uma proposta viável a ser implementada, que se inter-relaciona com as principais contribuições teóricas e práticas existentes nas pesquisas de *Big Data*.

## REFERÊNCIAS

ABELS, E. et al. **Competencies for Information Professionals of the 21st Century**. Special Libraries Association. 2003. Disponível em: <[http://sla.org/wp-content/uploads/2013/01/0\\_LRNCompetencies2003\\_revised.pdf](http://sla.org/wp-content/uploads/2013/01/0_LRNCompetencies2003_revised.pdf)>. Acesso em: 24 mar. 2017

ALVARENGA, L. et al. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 8, n. 15, p. 18-40, 2003. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/download/1518-2924.2003v8n15p18/5233%3E>>. Acesso em: 28 fev. 2017.

BARBALHO, C. R. S.; ROZADOS, H. B. F. Competências do profissional bibliotecário brasileiro: o olhar do Sistema CFB/CRBs. In: 2013. Encontro Nacional de Pesquisa em Ciência da Informação, IX. **Anais...** 2008. Disponível em:

<<http://repositorios.queestoemrede.uff.br/repositorios/bitstream/handle/123456789/1840/Compet%EAncias%20-%20Barbalho.pdf?sequence=1>>. Acesso em: 28 fev. 2017

BORKO, H. Information science: what is it? **American documentation** 19.1, 3-5, 1968.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**. Nova York, v.45, n.5, p. 351-360, 1991

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

DAVENPORT, T. H.; PATIL, D. J. Data scientist. **Harvard business review**, v. 90, p. 70-76, 2012.

FARIA, S. et al. Competências do profissional da informação: uma reflexão a partir da Classificação Brasileira de ocupações. **Ciência da Informação**, Brasília, v.34, n.2, p. 26-33, maio/ago. 2005.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Riding the Waves or Caught in the Ride?** Navigating the Evolving Information Environment. 2016. Disponível em: <[http://trends.ifla.org/files/trends/assets/insights-from-the-ifla-trend-report\\_v3.pdf](http://trends.ifla.org/files/trends/assets/insights-from-the-ifla-trend-report_v3.pdf)>. Acesso em: 28 fev. 2017

KAKHANI, M. K.; KAKHANI, S.; BIRADAR, S. R. Research Issues in Big Data Analytics. International Journal of Application or Innovation in Engineering & Management, Etmadpur, v. 2, n. 8, Aug. 2013. Disponível em: <<http://www.ijaiem.org/volume2issue8/IJAIEM-2013-08-29-070.pdf>>. Acesso em: 28 fev. 2017.

KATAL, A.; WAZID, M.; GOUDAR, R. H. Big data: issues, challenges, tools and good practices. In: IEEE Contemporary Computing (IC3), 6., 2013. **Proceedings... IEEE**, 2013. p. 404-409. Disponível em: <[http://www.stat.purdue.edu/~doerge/BIOINFORM.D/SPRING16/KatalWazidGoudar\\_2013.pdf](http://www.stat.purdue.edu/~doerge/BIOINFORM.D/SPRING16/KatalWazidGoudar_2013.pdf)>. Acesso em: 27 fev. 2017.

LE COADIC, Y. F. **A Ciência da Informação**. Brasília: Briquet de Lemos, 1996 (Tradução do original de 1994).

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

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, São Paulo, v. 30, n. 1, p. 61-70, jan./abr. 2001.

MCAFEE, A.; BRYNJOLFSSON, E.. Big Data: the management revolution. **Harvard Business Review**, Brighton, v. 90, n. 10, p. 61-67, oct. 2012. Disponível em: <<https://hbr.org/2012/10/big-data-the-management-revolution>>. Acesso em: 28 fev. 2017.

PONJUÁN DANTE, G. Perfil del profesional de información del nuevo milenio. In: VALENTIN, M. L. P. (Org.). **O profissional da informação**: formação, perfil e atuação profissional. São Paulo: Polis, 2000. P. 91-105.

REZENDE, Y. Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual. **Ciência da Informação**, Brasília, v.31, n. 2, p. 120-128, maio/ago. 2002.

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

ROSENFELD, L. Information architecture: looking ahead. **Journal of the American Society for Information Science and technology**, 53 (10) 874-87, 2002.

- RUSSOM, P. **Big data analytics**. TDWI Best Practices Report, Fourth Quarter, 2011. p. 1-35
- SAGIROGLU, S; SINANC, D. Big data: A review. In: Collaboration Technologies and Systems (CTS), 2013 **International Conference on**. IEEE, 2013. p. 42-47. Disponível em: <<https://xa.yimg.com/kq/groups/72986399/1585974627/name/06567202.pdf>>. Acesso em: 28 fev. 2017.
- SANTOS, P. L. V. C. As tecnologias na formação do profissional da informação. In: Valentin, M. I. P. (Org.). **Profissionais da informação: formação, perfil e atuação profissional**. São Paulo: Polis, 2000, p 103 -116.
- SANTOS, P. L. A. C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação. **DataGramaZero: revista de Ciência da Informação**. v. 10, n. 3, 2009.
- SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da Informação**. vol. 24, n. 1, 1995. Disponível em: <[http://www.brapci.ufpr.br/brapci/\\_repositorio/2010/03/pdf\\_dd085d2c4b\\_0008887.pdf](http://www.brapci.ufpr.br/brapci/_repositorio/2010/03/pdf_dd085d2c4b_0008887.pdf)>. Acesso em: 27 fev. 2017.
- TAURION, C. **Big data**. Rio de Janeiro: Brasport, 2013.
- VALENTIM, M. L. P. Equipes multidisciplinares na gestão da informação e conhecimento. In. BAPTISTA, S. G.; MULLER, S. P. M. (Org.) **Profissional da informação: o espaço de trabalho**. Brasília: Thesaurus, 2004. Cap. 7, p. 154-176.
- VAN DER AALST, W. M. P. Data scientist: The engineer of the future. In: Enterprise Interoperability VI. **Springer International Publishing**, 2014. p. 13-26.
- ZIKOPOULOS, P. et al. **Understanding big data: Analytics for enterprise class hadoop and streaming data**. New York: McGraw-Hill, 2011. Disponível em: <<http://www.bdvc.nl/images/Rapporten/ibm-understanding-big-data.pdf>>. Acesso em: 28 fev. 2016.