

## Big Data: os novos desafios para o profissional da informação

**Claudio José Silva Ribeiro**

*Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Brasil. E-mail: claudio.ribeiro@unirio.br*

### Resumo

O tema Big Data tem despertado interesse nos profissionais que trabalham com a Gestão da Informação. Este artigo apresenta o tema e explora os fundamentos que auxiliam no entendimento da abordagem de Big Data. Discute a explosão informacional e a avalanche de dados, chegando aos elementos que compõem o tema. Aborda os 4 Vs do Big Data e as fases de *Discovery*, *Data Preparation*, *Model Planning* e *Analytics*. Ao final, apresenta alguns aspectos ligados ao perfil do profissional que está participando destes projetos.

**Palavras-chave:** Big Data. Gestão da Informação. Análise de Dados. Profissional de Informação.

### 1 Introdução

O tema Big Data desperta, na atualidade, o interesse, e até mesmo o fascínio, para todas as pessoas que tem algum envolvimento com atividades para Gestão da Informação (HUWE, 2012).

A recente reportagem publicada no Rio de Janeiro pelo Jornal O Globo, sob título “Big Data: enxurrada de dados emerge como novo termômetro da economia” suscitou um debate interessante sobre novas formas para interação entre a sociedade, governos e serviços em geral, pois estas formas promovem uma nova dinâmica para fluxos informacionais (SETTI, 2014). A publicação desta reportagem por um veículo de grande circulação, demonstra que este assunto já é de interesse para a sociedade, indo além das fronteiras acadêmicas e da realização de negócios.

Verifica-se também um movimento na área de Ciência da Informação neste tema, quando o Prof<sup>o</sup>. Dr. Aldo Barreto, pesquisador na nossa área, faz em seu blog algumas reflexões sobre os três tempos da ciência da informação. Ele traz uma consideração sobre os estoques disponíveis de forma online, onde afirma que:

Hoje, com a condição online os estoques e os fluxos de informação, renomeados para “Big Data”, são

multidirecionados e levam condições virtuais em seu desatamento, quando o tempo se aproxima de zero, a velocidade se acerca do infinito e os espaços são de vivência pela não presença (BARRETO, 2014, *online*).

Outra contribuição para este debate foi levantada pelo Prof<sup>o</sup>. Dr. Marcos Cavalcanti, pesquisador do CRIE/COPPE, que afirmou em recente publicação que “Ao contrário do que muita gente pensa, Big Data não é uma ‘nova tecnologia’” (CAVALCANTI, 2014, *online*).

Fazendo uma análise preliminar destes movimentos, é lícito supor que o profissional de informação deve refletir um pouco sobre como poderá se envolver nas discussões sobre o tema Big Data, pois afinal, o uso de dados e informação sempre foi objeto de estudo para a Ciência da Informação.

Vale lembrar que não está se propondo algo totalmente novo com o tema Big Data, pois o uso de informação para a obtenção de resultados não é uma coisa nova. 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) (CIUPAK, BOSCARIOLI, CATARINO, 2013; HOFFMANN, 2011; REIS, ANGELONI, SERRA, 2010; VELHO, GOMES, 2009), além de informação para planejamento estratégico, gestão de recursos informacionais e ativos de informação na Web (BETTENCOURT, CIANCONI, 2013; GUEDES, 2012; BARBOSA, NACIF, 2012; RIBEIRO, 2008; dentre outros), foram abordagens exploradas nos últimos anos pela Ciência da Informação.

Mas afinal, o que é Big Data? É uma tecnologia? Uma ferramenta? Uma metodologia nova? Como o profissional da informação pode se inserir na discussão deste tema?

Para esclarecer estas questões é preciso, antes de tudo, observar algumas considerações sobre os aspectos norteadores para o tema Big Data.

A primeira consideração diz respeito ao tratamento de Big Data como uma tecnologia, pois o tema de alto volume de dados e informação há muito se faz presente nas pesquisas sobre processos de Gestão da Informação. O impulso dado pela tecnologia, principalmente pelo incremento do uso dos dispositivos móveis, trouxe um forte aumento no volume de dados. Mas, o debate sobre temas como o crescimento exponencial da informação e explosão informacional, originados pelas pesquisas pós segunda Guerra Mundial, já se fazia presente nas discussões e pesquisas na área de Ciência da Informação (SARACEVIC, 1996).

Uma segunda consideração diz respeito à variedade de dados disponíveis. A profusão de informações na internet, originadas pelas diferentes fontes de dados, ocasionam uma sobrecarga de dados e informação disponíveis para a sociedade. Cabe registrar que apenas 1% destes dados é efetivamente analisado (BREITMAN, 2014) e, conseqüentemente, pode-se deduzir que existe um grande campo de atuação para os profissionais que atuam na disciplina de gestão da informação.

Dentro deste contexto, este relato introduz a discussão dentro do projeto de pesquisa do autor, que tem por objetivo investigar métodos e abordagens para gerir recursos de informação residentes em ambientes digitais.

## **2 Da Explosão Informacional à Avalanche de Dados**

O tratamento e uso da informação pela sociedade têm se modificado nas últimas décadas como consequência do surgimento de novos modelos sociais, econômicos ou tecnológicos. Estes modelos promoveram uma mudança de paradigma tão importante quanto à invenção da imprensa, ou ainda, quanto à própria revolução industrial. A crescente utilização de meios de comunicação com alto grau de mobilidade e o uso cada vez maior da Internet, definem outros espaços e demarcam novas fronteiras para a sociedade contemporânea (RIBEIRO, 2008).

A quantidade de informação disponível cresce vertiginosamente e surgem novos comportamentos e sentimentos decorrentes deste crescimento. Termos como Ansiedade da Informação, cunhado por Wurman (2005), e Explosão Informacional cunhado por Gopinath e Das (1997) corroboram este estado de insatisfação dos usuários. Além disto, apesar do frequente lançamento de novas tecnologias e ferramentas para resolver os problemas dos usuários, a cada dia que passa, eles (os usuários) têm mais dificuldades em solucioná-los. Para Wurman:

Cada nova tecnologia não acaba com o resto, mas termina somando-se às outras. Disseram que o computador tornaria o papel obsoleto; aconteceu justamente o contrário: graças às impressoras e copiadoras. Acreditava-se que o vídeo fosse dar fim ao cinema, mas hoje há mais filmes do que nunca. [...] Estamos cercados por versões alternativas, imitações e originais, tudo em números incontáveis de exemplares. (WURMAN, 2005, p.4-5)

Heath e Bizer (2011) reforçam que na atualidade estamos cercados por uma grande quantidade de dados e informação. São registros sobre o cotidiano – desempenho da educação, produção de bens e serviços, investimentos e impostos governamentais, estatísticas sobre a economia e dados sobre o consumo - que nos ajudam a tomar decisões e gerar conhecimento.

Verifica-se também que existe uma retomada de debates para organizar a informação, que estão sendo discutidos sob a disciplina de ERM (Electronic Resource Management), pois este tema tem surgido com frequência nos projetos sobre o gerenciamento de ativos e/ou recursos de informação digitais (SPREHE, 2005).

Ribeiro (2008) convalida esta percepção e observa que:

[...] o processo de estruturação de dados e informações carece de maior instrumentação, pois a ótica utilizada na atualidade está mais concentrada em aspectos tecnológicos do que nas questões de organização das informações, deixando em segundo plano as indagações ligadas à gestão da informação (RIBEIRO, 2008, p. 18).

Mas afinal, o que está impulsionando esta avalanche? Conforme mencionado na introdução deste relato, uma diferença observada é que com a evolução da tecnologia o cotidiano ficou repleto de dados e informação, só que agora ao alcance dos nossos dedos. O avanço do uso de dispositivos móveis, o uso de sensores industriais e biomédicos, fotos, vídeos, e-mails, redes sociais, além do comércio eletrônico, interações via call centers, dispositivos móveis, dados públicos imagens médicas e outros dados científicos, câmeras para monitoramento, medidores inteligentes, GPS, aplicativos para troca de mensagens, aplicações que nos ajudam a pegar táxis, outras que nos ajudam na locomoção urbana evitando engarrafamentos, ou ainda no monitoramento de ônibus e até de aviões, são exemplos concretos desta avalanche.

Complementarmente, é possível perceber também uma mudança no funcionamento das aplicações de comércio eletrônico. A ampliação do uso de sistemas de recomendação<sup>1</sup> na Web, permite que sejam indicados dezenas de opções de compras aos clientes usuários destes serviços.

Por outro lado, a previsão da expansão das fontes de dados é de aproximadamente 50 vezes nos próximos 10 anos. Segundo previsões apresentadas pela empresa EMC, instituição especializada em armazenamento de dados, o crescimento de dados e informações digitais no mercado brasileiro crescerá de 212 Exabytes<sup>2</sup> em 2014, alcançando a marca de 1.6 Zettabytes (1.600 Exabytes) em 2020 (EMC, 2014).

Fruto deste cenário, rico em volume e variedade de fontes, tem surgido uma nova disciplina que, apesar de não ser apenas um tema essencialmente tecnológico, vem sendo impulsionado pelos projetos de tecnologia: a vertente de Big Data.

### 3 A visão do Big Data

Fox e Hendler (2011) também anteciparam que estamos vivendo com uma nova abordagem chamada de “Big Data”. Esta abordagem está surgindo em decorrência da geração, e, conseqüentemente, da necessidade da coleta de grande volume de dados com formatos variados. Ademais, estes dados ainda precisam ser geridos e, neste sentido, Hendler e Fox continuam e observam que a gestão destes recursos possibilitará a resolução de problemas que nem sabíamos que existiam. No entanto, vale ressaltar

<sup>1</sup> Em um sistema de recomendação, parte-se de perfis de usuários específicos, que podem ser agrupados e relacionados a outros perfis que, quando incrementados com seus respectivos históricos de compras e com os dados originados pelas redes sociais, possibilitam a descoberta de produtos a serem ofertados (FLORISSI, 2014).

<sup>2</sup> A unidade Exabyte é equivalente, em números aproximados, a 1.000 Petabytes, ou a 1.000.000 de Terabytes, ou ainda a 1.000.000.000 de Gigabytes.

que não podemos prescindir de ferramentas, pois a capacidade do ser humano de analisar dados e informações com múltiplas facetas é limitada. Logo, são necessários alguns instrumentos que nos auxiliem a executar estas tarefas.

A necessidade de vencer o desafio, reunindo e analisando fontes de diversas naturezas, deu origem a pesquisas que nos levaram ao tema “Big Data”. Estas pesquisas foram desenhadas a partir de três aspectos iniciais (DAVENPORT, 2014):

1. A múltipla natureza dos dados – aspecto relacionado com as diferentes fontes disponíveis.
2. O uso de processamento em nuvem – aspecto relacionado ao uso ilimitado de recursos computacionais e com processamento em larga escala, com a possibilidade de redução de custos (economia de escala – é o aspecto econômico-financeiro).
3. Uso de tecnologias específicas, tais como processamento de rotinas em paralelo e ferramentas para otimização como *Hadoop*<sup>3</sup> e *MapReduce*<sup>4</sup>, *HDFS*<sup>5</sup>, além de abordagens de *Machine Learning*<sup>6</sup> e *Analytics*<sup>7</sup>.

---

<sup>3</sup> Hadoop é tecnologia *open source* desenvolvida pela Google e Yahoo para processar muitos dados em servidores, usando a noção de processamento em paralelo e uso de *clusters* (conjuntos) de computadores no processamento. Pode ser chamado de Apache Hadoop e também foi desenvolvido e customizado por outros fabricantes (EMC, Intel, Microsoft, dentre outros).

<sup>4</sup> *MapReduce* é o framework arquitetural que deu origem à tecnologia de *Hadoop*. Usa a estratégia de dividir para conquistar, ou seja, distribui e aloca um problema muito grande em *clusters* de armazenamento, usando registros serializáveis do tipo <chave, valor>.

<sup>5</sup> HDFS é a sigla de *Hadoop File System*. É uma estrutura de armazenamento de arquivos que utiliza blocos de 128 Mbytes, que são muito menores do que os blocos de particionamento tradicionais, utilizados em dispositivos de armazenamento.

<sup>6</sup> *Machine Learning* trata o uso de algoritmos que identificam o melhor modelo para ser aplicado ao conjunto de dados.

<sup>7</sup> *Analytics* é a essência de Big Data. Trata a análise dos dados e será apresentada mais à frente neste relato.

A abordagem de Big Data está apoiada em quatro outros fatores de sustentação, conhecidos como os 4 Vs do Big Data: Volume, Variedade, Velocidade e Veracidade (DUMBILL, 2012). A seguir será apresentado um breve esclarecimento do papel de cada um desses componentes:

O primeiro V é de Volume e está ligado ao grande quantitativo de dados e informações que nos cercam no cotidiano. Já o segundo V está ligado à variedade destes recursos.

Devido a forte relação entre Volume e Variedade, estes fatores serão comentados em conjunto. A multiplicidade de dispositivos e a capacidade destes dispositivos interagirem em rede está promovendo a verdadeira inundação de dados. Cada um de nós carrega junto de si um celular, que agindo como um sensor, pode enviar informação de localização das pessoas e permitir a realização de negócios direcionados<sup>8</sup>. Ao levarmos em consideração que o mundo tem cerca de 7 bilhões de habitantes (WIKIPEDIA, 2014) e que aproximadamente 6 bilhões possuem celulares (ONUBR, 2013), pensem no volume e na variedade de dados que pode ser gerado, captado, processado, (re)utilizado e entregue.

As cidades estão repletas de câmeras de monitoramento nos prédios, lojas, ruas e avenidas. Qualquer cidadão pode gravar e postar um vídeo em mídias sociais ou no Youtube. Estima-se que a quantidade de vídeos produzidos diariamente ultrapassa a produção dos primeiros 50 anos de televisão (DAVENPORT, 2014).

Saindo do cotidiano e observando o ambiente de ciência e tecnologia, temos muitos outros exemplos. Os projetos de pesquisa de perfuração de petróleo em águas profundas, incluindo o pré-sal (CIARINI, 2013; SANTOS, 2014), além de projetos de pesquisa em astronomia,

---

<sup>8</sup> Um exemplo para a prática de negócios direcionados é o *e-couponing*. Esta prática possibilita o envio de cupons de desconto em tempo real para os usuários, quando os mesmos estão nas proximidades das lojas, utilizando-se as coordenadas GPS dos celulares dos usuários (FLORISSI, 2014).

estão impulsionando o uso da abordagem de Big Data (PORTO, 2013).

Adiciona-se a esse cenário, uma vasta coleção de outras fontes e formas para geração de unidades documentárias. O crescimento do uso de documentos digitais e páginas Web nas organizações, recursos estes estruturados por meio de ferramentas para Gestão de Conteúdo (RIBEIRO, 2012), bem como o desenvolvimento de propostas de uso da Web of Data e Linked Data (RIBEIRO, ALMEIDA, 2011; RIBEIRO, PEREIRA, 2014) também têm contribuído para um aumento em Volume e Variedade de dados e informação.

Voltando aos 4 Vs do Big Data, chega-se agora ao terceiro V, de velocidade. A melhoria dos canais de transmissão, com redes em fibra ótica e emissores de sinais de alta capacidade, o uso de satélites, o uso de outras bandas para a telefonia celular, as comunicações em tempo real para controle de processos na internet, os workflows científicos com processamento paralelo e *cluster* de processamentos vem possibilitando atingir uma maior velocidade para troca de dados e informação (MATTOSO, 2013). Ademais, é possível afirmar que a velocidade continuará crescendo, pois o desenvolvimento da tecnologia de processadores, dos canais e do hardware para armazenamento (discos rígidos e memória rápida – *flash memory*), duplica o seu poder a cada período de 2 anos (FLORISSI, 2012).

O quarto V é de Veracidade. A qualidade dos dados e informação é característica essencial para que os usuários interessados (executivos, gestores públicos e a sociedade em geral) usem e (re)usem os dados de maneira apropriada e real, gerando informações críveis para eles mesmos.

Para concluir a noção de Big Data ainda vale explorar um componente que faz parte do terceiro aspecto relacionado anteriormente por Davenport. A discussão sobre o trabalho de análise dos dados,

entendidos pela noção de Big Data *Analytics*.

#### **4 O esforço para análise: o trabalho com Big Data Analytics**

O objetivo da tarefa de *Analytics* é executar a análise preditiva dos dados por meio da execução de *mining* (minerações)<sup>9</sup>. Segundo os autores Oliveira (2013) e Tavares (2014), inicialmente, serão tratados os dados com o uso de técnicas estatísticas, para separação e reunião de conjuntos (denominado de fase de *discovery*).

Adicionalmente, para executar a tarefa também pode-se fazer uso de técnicas para categorização, limpeza e transformação dos dados, utilizando, inclusive, a visão da proveniência (fontes de origem) dos dados para auxiliar no processo de categorização. Ao final desta fase é possível chegar à definição e preparação de modelos (fase de *data preparation e model planning*) que serão úteis na construção do grande conjunto de dados, chamado de lago de dados (*data lake*).

A carga de dados (denominada fase de *ingest*) ocorrerá em seguida e será realizada para povoar o lago de dados. No lago estarão reunidos todos os dados que serão alvo de análise. Por fim, os resultados que serão obtidos a partir do tratamento e análise do conteúdo do lago serão apresentados com uso de ferramentas de visualização e deverão estar associados ao contexto de negócios (OLIVEIRA, 2013; TAVARES, 2014).

A análise de dados que atendem aos requisitos descritos anteriormente (lembrem-se dos 4 Vs), precisará ser desenvolvida segundo uma nova arquitetura de análise, onde dados serão obtidos de múltiplas fontes e em tecnologias diversas. O ponto central desta análise está ligado à capacidade de

---

<sup>9</sup> A noção de *mining* de dados passa pela extração e análise de grandes volumes de informação em busca de padrões e comportamentos.

correlacionar dados, pois, como já observado, o ser humano possui limitações para fazer análises associadas a múltiplas dimensões. Em essência, quando temos uma pequena quantidade de dados (*little data*) não temos muita dificuldade de correlacioná-los, pois existem poucas inter-relações. Mas, com uma grande quantidade (*big data*), temos muitos dados sendo gerados em paralelo, logo, surge a dificuldade para correlacioná-los (SEYMOUR, 2014).

Então, decorrente deste cenário, chegamos não apenas a um novo conjunto de passos para análise, mas também a um outro perfil profissional atuando neste mercado. Na visão de Sathi (2013), a vertente de *Analytics* começa a se integrar aos processos de negócio das empresas, tendo em vista a mudança do comportamento nos executivos e na nova ótica de produção de bens e serviços que está influenciando estas organizações.

## 5 O profissional do Analytics

Fruto do trabalho com *Analytics* cunhou-se um novo perfil profissional. Este perfil passou a ser denominado de Cientista de Dados (*Data Scientist*). A característica principal deste profissional é ter a capacidade de aplicar ferramentas analíticas e algoritmos para gerar previsões sobre produtos e serviços (DAVENPORT; PATIL, 2012). Oliveira (2013) complementa e detalha que este perfil deve ter forte conhecimento em disciplinas como a matemática e a estatística, com treinamento avançado em estratégias para tratamento de grandes conjuntos de dados, fazendo uso de modelos matemáticos, formulação de hipóteses e técnicas de regressão. Já Brietman (2013) observa que o Cientista de Dados deve ter capacidade de levantar requisitos dos usuários, buscando não apenas nas necessidades destes usuários, mas também nos outros envolvidos no ambiente sob análise (clientes, parceiros de negócio, informações de mercado, feeds de notícias, redes sociais, blogs, dentre outros).

Para Oliveira (2013), o cientista de dados deve ser um técnico cético, curioso, criativo, comunicativo e deve saber trabalhar em colaboração. Ademais, o cientista de dados deve sempre (re)avaliar questões durante as primeiras fases do desenvolvimento do trabalho.

Oliveira continua e apresenta questões que podem auxiliar na revisão destas fases (OLIVEIRA, 2013):

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?

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, 2013).

### 5.1 Uma proposta de atuação para o profissional da informação

A Ciência da Informação é o campo de estudos apropriado para desenvolver novos estudos acerca de dados e informação. Segundo Saracevic (1996, p.43) O papel da CI é “prover os meios para o fornecimento de informações relevantes para indivíduos, grupos e organizações envolvidas com a ciência e tecnologia.”

Tálamo e Smit (2007, p. 39, grifo nosso, p.48-49) convalidam este argumento quando observam que “a área se relaciona tanto à organização de práticas científicas e profissionais, quanto ao acompanhamento das mudanças nas práticas culturais e nas modalidades de difusão e aquisições de conhecimentos.”

Este acompanhamento se desenvolve apoiado em práticas informacionais que buscam: ter forte relação com as ações da sociedade, por meio de serviços para atender as atividades (sejam elas públicas e/ou privadas); articulações entre campos de estudo e a disponibilidade de correlacionar problemáticas oriundas de correntes distintas (TALAMO, SMIT; 2007).

Saracevic também observa que são várias as teorias e disciplinas que auxiliam na formalização da investigação em Ciência da Informação, dentre elas: teoria da informação, teoria das decisões, ciência cognitiva, lógica, análises em bibliometria, além da quantificação das estruturas do conhecimento e de seus efeitos. Assim, é possível propor que o profissional da informação acompanhe as pesquisas em Big Data, especialmente nos temas iniciais desenvolvidos na fase de *Discovery, Data Preparation e Modeling Planning*.

A Ciência da Informação nasceu e foi gestada com o objetivo maior de apresentar solução para problemas ligados ao uso de dados e informação e como tal,

tem um importante papel nos estudos que envolvem o tema Big Data. Versig (*apud* Pinheiro e Loureiro, 1995, p.4) observa que, em função da interdisciplinaridade da nossa área, o cientista da informação é obrigado a lidar “com dados fragmentados de natureza empírica e teórica.” Além disto, Versig continua e complementa com a ideia de reformulação constante da Ciência, quando observa que:

[...] se a ciência da informação existe, qualquer que seja a denominação dada a esse campo, ela não possuirá uma teoria, mas uma estrutura proveniente de um amplo conceito científico ou modelos e conceitos reformulados. Esses serão intertecidos a partir de seu desenvolvimento e do problema do uso do conhecimento nas condições pós-modernas de informatização. Havendo uma interconexão entre tudo, ciência da informação deve desenvolver um sistema de navegação conceitual (VERSIG *apud* PINHEIRO e LOUREIRO, 1995, p.4).

## 6 Considerações Finais

O que vivemos é um momento singular, pois a contínua redução do custo dos equipamentos, levando este insumo ao patamar de *comodities*, os novos processos de gestão de dados e informação, além de novos softwares e ferramentas para apoiar o processo de análise de dados (*analytics*), tem contribuído para um momento especial no tratamento da informação (MINELLI, CHAMBERS, DHIRAJ, 2013).

As reflexões apresentadas aqui ainda estão em fase de gestação, pois o próprio desenvolvimento do tema Big Data ainda é bastante recente. Muitas dúvidas ainda cercam os projetos, especialmente nas fases iniciais, pois conhecer o ambiente de dados e informação, efetuar a organização de conjunto de dados (categorizá-los?), realizar entrevistas e desenvolver os modelos (tanto estruturais quanto matemáticos), ainda carecem de melhor experimentação nos diferentes domínios de conhecimento.

Espera-se que este movimento de pesquisa sobre o tema Big Data na área da

Ciência da Informação, ilumine o caminho a ser trilhado e possibilite que outros pesquisadores interessados possam se

engajar nesta discussão, levando este tema para além da tecnologia.

---

## ***Big Data: new challenges for the information professional***

### ***Abstract***

*The theme Big Data has been raising interest on professionals who work with Information Management. This article not only presents this theme and explores the fundamentals that help in understanding Big Data's approach, but also discusses the information explosion and data deluge, reaching elements that make up the theme. Moreover, it covers the 4 Vs of Big Data and the phases of Discovery, Data Preparation, Model Planning and Analytics. At it's end, this paper points out some aspects related to the professional who takes part in these projects.*

**Key-words:** *Big Data. Information Management. Data Analysis. Information Professional.*

---

### **Referências**

BARBOSA, R. R.; NASSIF, M. E. PRÁTICAS DE GESTÃO E DE TECNOLOGIA DA INFORMAÇÃO E SEU RELACIONAMENTO COM O DESEMPENHO ORGANIZACIONAL.

#### **Perspectivas em Gestão &**

**Conhecimento**, João Pessoa, v. 2, n. esp., 2012.

BARRETO, A. Os três tempos da Ciência da Informação. 16 maio 2014. Disponível em: <<http://aldobarreto.wordpress.com>>. Acesso em: 12 maio 2014.

BETTENCOURT, M. P. da L.; CIANCONI, R. de B. GESTÃO DO CONHECIMENTO: um olhar sob a perspectiva da ciência da informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 5, n. 1, 2013.

BRIETMAN, K. Big Data Overview. Palestra apresentada no 1º. **EMC Summer School on Big Data**. EMC/NCE/UFRJ. Rio de Janeiro. 2013.

\_\_\_\_\_. Big Data Seen from the Clouds. Palestra apresentada no 2º. **EMC Summer School on Big Data**. EMC/NCE/UFRJ. Rio de Janeiro. 2014. Disponível em: <[http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Big\\_Data\\_Summer\\_School\\_Karin.pdf](http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Big_Data_Summer_School_Karin.pdf)>. Acesso em: 30 Maio 2014.

CAVALCANTI, M. Big Data não é tecnologia. 23 de março de 2014. **O Globo**. Disponível em: <<http://oglobo.globo.com/blogs/inteligenciaempresarial/posts/2014/03/23/big-data-nao-tecnologia-528599.asp>>. Acesso em: 30 Abril 2014.

CIARINI, A. E. M. Research on Big Data and Opportunities. Palestra apresentada no 1º. **EMC Summer School on Big Data**. EMC/NCE/UFRJ. Rio de Janeiro. 2013.

CIUPAK, L. F.; BOSCARIOLI, C.; CATARINO, M. E. AN ANALYSIS OF THE USE OF BUSINESS INTELLIGENCE TECHNOLOGIES AS FACILITATORS OF UNIVERSITY MANAGEMENT. **Brazilian Journal of Information Science**, Marília, v. 7, n. esp., 2013.

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

DAVENPORT, T. H.; PATIL, D.J. "Data Scientist: The Sexiest Job of the 21st Century." **Harvard Business Review** 90, no. 10, October, p.70–76, 2012.

DUMBILL, E. What is Big Data? In: O'Reilly Media Inc. **Big Data Now: current perspectives**. O'Reilly Media:California. 2012. Disponível em: <<http://www.oreilly.com/data/free/files/big-data-now-2012.pdf>>. Acesso em: 03 maio 2014.

EMC. Brazil country brief. The Digital Universe of opportunities. 2014.

Disponível em:

<<http://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014-brazil.pdf>>. Acesso em: 12 maio 2014.

FLORISSI, P. Big Data: Challenges and Opportunities. Palestra apresentada no 2º. **EMC Summer School on Big Data**.

EMC/NCE/UFRJ. Rio de Janeiro. 2014.

FLORISSI, P. EMC On Big Data. 2012.

Disponível em:

<<https://www.carecorenational.com/healthcaresummit/powerpoints/PatriciaFlorissiPhD.pdf>>. Acesso em: 15 maio 2014.

FOX, P.; HENDLER, J. Changing the Equation on Scientific Data Visualization.

*Science* **331**, 705 (2011). Disponível em:

<[http://data2discovery.org/dev/wp-content/uploads/2013/05/Fox-and-Hendler\\_Visualization\\_Science-2011-Fox-705-8.pdf](http://data2discovery.org/dev/wp-content/uploads/2013/05/Fox-and-Hendler_Visualization_Science-2011-Fox-705-8.pdf)>. Acesso em: 15 maio 2014.

GOPINATH, M. A.; DAS, P.

Classification and representation of Knowledge. **Library Science with a Slant to Documentation and Information Studies**, v. 34, n. 2, p. 85-90, 1997.

GUEDES, V. L. da S. A BIBLIOMETRIA E A GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO CIENTÍFICO E TECNOLÓGICO: uma revisão da literatura. **Ponto de Acesso**, Salvador, v. 6, n. 2, 2012.

HEATH T.; BIZER C. **Linked Data: Evolving the Web into a Global Data Space** (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2011.

HOFFMANN, W. A. M. Monitoramento da informação e inteligência competitiva: realidade organizacional. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 2, n. 2, p. 125-144, 2011.

HUWE, T. K. Big Data, Big Future.

**Computers in libraries**, v. 32, n. 5, p.20-22, 2012.

MARCHAND, D. A.; PEPPARD, J. Why IT Fumbles Analytics. **Harvard Business Review**, jan-fev. 2013.

MARSHALL, Cathy. Big Data, the crowd and me. **Information Services & Use**, IOS Press, 2012, p. 213-224.

MATTOSO, M. Scientific Workflows and Big Data. Palestra apresentada no 1º. **EMC Summer School on Big Data**.

EMC/NCE/UFRJ. Rio de Janeiro. 2013.

MINELLI, M.; CHAMBERS, M.; DHIRAJ, A. **Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses**. Wiley CIO Series. 2013.

OLIVEIRA, A. Data Science and Data Analytics. 2013. Palestra apresentada no 1º. **EMC Summer School on Big Data**. EMC/NCE/UFRJ. Rio de Janeiro. 2013.

**ONUBR**. ONU: Dos 7 bilhões de habitantes do mundo ..... 22 de março de 2013. Disponível em <http://www.onu.org.br/onu-dos-7-bilhoes-de-habitantes-do-mundo-6-bi-tem-celulares-mas-25-bi-nao-tem-banheiros/> Acesso em 10 maio 2014.

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

PORTO, Fabio. Big Data in Astronomy: The LIneA-DEXL case

2013. Palestra apresentada no 1º. **EMC Summer School on Big Data**.

EMC/NCE/UFRJ. Rio de Janeiro. 2013.

REIS, E. S.; ANGELONI, M. T.; SERRA, F. R. Business Intelligence como Tecnologia de Suporte a Definição de Estratégias para a Melhoria da Qualidade de Ensino. **Informação & Sociedade: Estudos**, João Pessoa, v. 20, n. 3, p. 157-167, set./dez. 2010.

RIBEIRO, C. J. S. Diretrizes para o projeto de portais de informação: uma proposta interdisciplinar baseada na Análise de Domínio e Arquitetura da Informação. 2008. 298 f. Tese (Doutorado em Ciência da Informação) – Convênio UFF/IBICT, Rio de Janeiro.

\_\_\_\_\_. Projeto de serviços para Gestão Documental com uso de abordagem tecnológica SOA (Arquitetura Orientada a Serviços). **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 2, 2012.

RIBEIRO, C. J. S.; ALMEIDA, R. F. Dados Abertos Governamentais (Open Government Data): Instrumento para Exercício de Cidadania pela Sociedade. In: **Elmira Simeão, Jorge Henrique Cabral Fernandes, Isa Maria Freire. (Org.). XII Enancib - Políticas de Informação para a Sociedade - Anais**. Brasília: Thesaurus, 2011.

RIBEIRO, C. J. S.; PEREIRA, D. V. El proceso de publicación e los datos gubernamentales abiertos acerca de la clase de la Seguridad Social Brasileña de Vocabulario Controlado del Gobierno Electrónico (VCGE). In: **Simposio Latinoamericano Acceso a la Información Gubernamental, 2014**, Cidade do México. Anais do Simposio Latinoamericano Acceso a la Información Gubernamental, 2014.

SANTOS, I. H. R. BigData Research and Development at Petrobras. 13 de maio de 2014. Palestra apresentada no **2º. EMC Summer School on Big Data**. EMC/NCE/UFRJ. Rio de Janeiro, 2014. Disponível em: <[http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Ismael\\_BigDataTOOL\\_SummerSchool\\_v2.pdf](http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Ismael_BigDataTOOL_SummerSchool_v2.pdf)>. Acesso em: 20 maio 2014.

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

SATHI, A. **Big Data Analytics: Disruptive Technologies for Changing the Game**. Mc Press. 2013.

SETTI, R. Enxurrada de dados emerge como novo termômetro da economia. 22 de março de 2014. **O Globo**. Disponível em:

<<http://oglobo.globo.com/sociedade/tecnologia/big-data-enxurrada-de-dados-emerge-como-novo-termometro-da-economia-11959190>>. Acesso em: 10 maio 2014.

SEYMOUR, C. The State of Big Data. **EContentMag.com**, jan-feb, p. 26-27. 2014

TÁLAMO, M. de F. G. M.; SMIT, J. W. CIÊNCIA DA INFORMAÇÃO: PENSAMENTO INFORMACIONAL E INTEGRAÇÃO DISCIPLINAR. *Brazilian Journal of Information Science*, v.1, n.1, jan./jun. 2007. Disponível em: <<http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/30/54>>. Acesso em: 15 junho 2014.

TAVARES, E. BIG DATA in Business. 12 de maio de 2014. Palestra apresentada no **2º. EMC Summer School on Big Data**. EMC/NCE/UFRJ. Rio de Janeiro. 2014. Disponível em: <[http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao\\_Elaine\\_Tavares.pdf](http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao_Elaine_Tavares.pdf)>. Acesso em: 30 maio 2014.

VELHO, A. C. M.; GOMES, S. L. R. A tomada de decisão na Previdência social: transpondo o uso das tecnologias de Informação na DATAPREV. **Informação & sociedade (UFPB. Online)**, v. 19, p. 55-68, 2009.

WIKIPEDIA. A Enciclopédia Livre. Disponível em: <[http://pt.wikipedia.org/wiki/Popula%C3%A7%C3%A3o\\_mundial](http://pt.wikipedia.org/wiki/Popula%C3%A7%C3%A3o_mundial)>. Acesso em: 10 maio 2014.

WURMAN, R. S. **Ansiedade de Informação 2**. São Paulo: Editora de Cultura, 2005.