

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT8 – Informação e Tecnologia

DESAFIOS DA GESTÃO DE DADOS NA ERA DO *BIG DATA*: PERSPECTIVAS PROFISSIONAIS

Adriana Alves Rodrigues – (Universidade Federal da Paraíba UFPB)

Emeide Nóbrega - (Universidade Federal da Paraíba UFPB)

Guilherme Ataíde Dias - (Universidade Federal da Paraíba UFPB)

DATA MANAGEMENT CHALLENGES IN THE BIG DATA AGE: PROFESSIONAL PERSPECTIVES

Modalidade da Apresentação: Comunicação Oral

Resumo: Investiga os desafios da gestão de dados na era do *Big Data*, bem como da avaliação das potencialidades, riscos e oportunidades na sociedade contemporânea, em especial para a Ciência da Informação, a partir do impacto da expansão dos dados. Nos últimos dez anos tem ocorrido transformações estruturais no campo informacional em decorrência da velocidade e do grande volume de dados permeando a sociedade e com implicações na dimensão econômica, tecnológica, social e informacional apontando para novas tendências. No contexto, reflete-se sobre o perfil profissional do cientista de dados e as habilidades que este deve ter diante de um ambiente de mega dados. Neste sentido, busca-se a compreensão do que são esses dados, como utilizá-los estrategicamente e da complexidade envolvida no tocante a gestão de dados. As estratégias metodológicas envolvem pesquisa exploratória, descritiva e bibliográfica na exploração teórico-conceitual. Os resultados apresentam o delineamento de um novo perfil profissional para o contexto de gestão de dados na era do *Big Data* em que características interdisciplinares se evidenciam como dinâmicas no contexto de perspectiva profissional. Concluímos que é urgente endereçar e ampliar pesquisas sobre o fenômeno dentro da Ciência da Informação.

Palavras-chave: Big Data; Gestão de dados; Cientista de dados.

Abstract: Investigates the data management challenges in the Big Data era, as well as the assessment of potentialities, risks and opportunities in contemporary society, especially in the Information Science field, from the impact of data expansion. In the last ten years there have been structural transformations in the information field due to the speed and the large volume of data permeating society with implications in the economic, technological, social and informational dimension pointing to new trends. In this context, we reflected on the data scientist professional profile and the skills

that it should have in face of the Big data environment. In this sense, we seek to understand what these data are, how to use them strategically and the complexity involved in data management. The methodological strategies involve exploratory, descriptive and bibliographic research in theoretical-conceptual exploration. The results present an outline of a new professional profile for the data management context in the Big Data era in which interdisciplinary characteristics are evidenced as dynamics in the context of professional perspective. We conclude that it is urgent to address and expand research on the phenomenon within Information Science field.

Keywords: Big Data; Data Management; Data Scientist.

1 INTRODUÇÃO

O fenômeno do *Big Data* evidencia um ambiente em mutação em termos culturais, sociais, políticos e econômicos e, sobretudo, na Ciência da Informação como uma preocupação central com a gestão de dados. Assim, *este se* revela um conceito e uma prática em *continuum* como problemática quanto à capacidade de recuperação e processamento de dados dentro desse contexto multifacetado em que o fenômeno se entrecruza considerando sua natureza interdisciplinar (Computação, Ciência da Informação, Administração, entre outras). Em específico para a Ciência da Informação o conceito expõe desdobramentos que exigem estudos e pesquisas para uma melhor compreensão de suas dimensões no tocante aos mega dados¹ e os desafios da gestão.

A Ciência da Informação vivenciou a explosão informacional no período pós-guerra cujo contexto modificou o *status quo* da informação, bem como introduziu novos modos de recuperação e processamento dos documentos armazenados (BARRETO, 2014, *online*) e ao longo do tempo foram se complexificando até à atualidade *com o Big Data*, em que os fluxos informacionais “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”. Trata-se, portanto, de uma área que está diretamente vinculada com os estudos da informação e tecnologia e, neste sentido, “a Ciência da Informação é o campo de estudos apropriado para desenvolver “meios para o fornecimento de informações relevantes para os indivíduos, grupos e organizações envolvidas com a ciência e tecnologia” (SARACEVIC, 1996, p. 43). Além dessa reconfiguração no tratamento informacional, parte das pesquisas em Ciência da Informação (NASSIF; SANTOS, 2009;

¹ Em português, *Big Data* pode ser traduzido como Mega dados. Para fins dessa investigação, utilizaremos os dois termos, simultaneamente, como sinônimos em distintos momentos.

SANTOS, 2016) tem buscado refletir sobre o perfil do profissional que lida com a gestão dos dados dentro do contexto das Tecnologias Digitais da Informação e Comunicação (TDICs).

De fato, a gestão de dados se complexificou com a emergência do *Big Data*, que aponta para novos horizontes científicos e pragmáticos conduzindo para a exigência de novas habilidades e conhecimento para os profissionais da área visando o enquadramento no perfil para a gerência dos dados. À medida que o tratamento dados se intensifica e indica desafios, competitividade e necessidade de inovação, na mesma proporção, exige-se um profissional que esteja apto a gerenciar informação neste novo contexto. Nesta perspectiva, “um indivíduo considerado competente em informação reconhece que a informação é o insumo essencial para a tomada inteligente de decisões no âmbito das suas diversas atividades, para perceber quando há necessidade de informação” (DUARTE et al, 2013, p. 556). Sendo assim, no panorama dos grandes volumes de dados, o cientista de dados desponta como uma das profissões promissoras e desafiantes da atualidade diante do fenômeno do *Big Data* e da gestão desses mega dados no contexto da Ciência da Informação. Partindo dessa situação-problema, o estudo é norteado pela questão: Quais habilidades e características o cientista de dados deve adquirir para a gestão de dados no cenário de *Big Data*?

Numa abordagem organizacional, o estudo objetiva apreender o efeito trazido pelo *Big Data* para a gestão de grande quantidade de dados e a ênfase na capacitação profissional na contemporaneidade, compreendendo o perfil que se adere a esta conjuntura de complexidade da gestão de dados. Por meio de pesquisa exploratória e bibliográfica, a investigação procura caracterizar as habilidades e perfil do cientista de dados na relação com a gestão de dados, a partir de duas perspectivas. Neste sentido, o referencial teórico explora o conceito de *Big Data* com vistas ao estado da arte do fenômeno, bem como fornece um quadro de definições para entender a gama conceitual em torno do objeto. Um segundo momento refere-se à apresentação dos resultados da pesquisa a partir do escrutínio exploratório do perfil profissional de cientista de dados voltado para a gestão de dados em rede.

2 A GESTÃO DE DADOS NA ERA DO *BIG DATA* E DO CONHECIMENTO

A explosão de dados conduziu a outras questões desafiadoras de investigação das Tecnologias Digitais da Informação e Comunicação (TDICs) no tocante à otimização eficaz para a grande massa de dados de modo a identificar mecanismos de análise. Acrescente-se a esse *rol* de possibilidades o desafio de transformar o volume de dados em informação, e como consequência, em vantagem competitiva nas organizações (FERNANDES; LIMA-MARQUES, 2015). Tais autores posicionam o *Big Data* para além de convergência tecnológica, e que pode oferecer às organizações um terreno propício à inovação visando mudanças no setor produtivo ou na ordem jurídica, social, econômica, política e cultural. Fernandes e Lima-Marques (2015) entendem o fenômeno como o grande armazenamento e velocidade de dados estruturados e não-estruturados, oriundos principalmente das redes digitais e de bancos de dados.

Partindo do entendimento de que *o Big Data* é uma consequência da sociedade da informação (BARRETO, 2004), um fenômeno cultural, tecnológico e mitológico (BOYD; CRAWFORD, 2000), que desafia o modo como vivemos e interagimos no mundo (SCHONBERGER; CUKIER, 2014), é a nova fronteira para inovação, competitividade e produtividade (MANIYKA et al, 2011) e um paradigma na era da informação (DI MARTINO, 2014), extrapola a capacidade de armazenamento dos bancos de dados (DUMBILL, 2015), deve-se questionar qual o significado dos dados digitais e como gerenciar toda a avalanche de dados, que são produzidos através do aparato tecnológico. A partir desta percepção, deve-se observar o direcionamento para desempenhar um papel econômico significativo, aumentando a produtividade das organizações e interações que resultam desse movimento, assim como a inovação para construir um capital competitivo na era dos dados. São reconfigurações advindas da explosão informacional vivenciada ainda na década de 1940 que revelam não só o poder dos *petabytes* (EINSENSTEIN, 2015) e de seus processos técnicos, mas um campo de desafios e complexidade no manejo dos dados e no modo como valoramos e transformamos em conhecimento.

Para Coneglian, Santarém Segundo e Sant’Anna (2017, p. 62), *o Big Data* “está causando uma revolução em como empresas, governos e organizações coletam e analisam os dados para a tomada de decisão, tanto no âmbito governamental quanto no empresarial”, que, ainda segundo os autores, o conceito está vinculado a três questões: heterogeneidade, rápida geração e processamento de grandes quantidades de dados disponíveis em ambientes digitais.

Quanto às definições, há uma gama de endereçamentos conceituais sobre o fenômeno e presente nas mais variadas áreas de conhecimento. Cabe às áreas como a astronomia e a genômica, advindos de um contexto de explosão informacional a partir dos anos 2000, parte da sistematização para o termo em uma escala maior de visibilidade (SCHÖNBERGER-MAYER; CUKIER, 2013). O conceito em questão ainda está longe de um consenso e, inclusive, está envolta com controvérsias e contestações tendo em vista a classificação pelos autores de enquadrar como mais de uma estratégia de marketing que um fenômeno para estudo. Contudo, é crescente a literatura sobre *Big Data* e sua consolidação para os estudos dos grandes volumes de dados. Elencamos algumas definições (Quadro 1) para auxiliar na compreensão do fenômeno e das concepções que ele adquire em campos distintos como Computação, Economia, Ciência da Informação.

Quadro 1 – Definições e contextos de *Big Data*.

AUTOR	DEFINIÇÃO
DI MARTINO et al. (2010)	“É um campo emergente onde inovadora tecnologia oferece alternativas para resolver os problemas inerentes que surgem quando se trabalha com grandes quantidades de dados, fornecendo novas maneiras de reutilizar e extrair valor a partir de informação” (p. 5).
MANYKA, J. et al (2011)	“Refere-se a um banco de dados cujo tamanho vai além da capacidade do software de banco de dados e ferramentas típicas para capturar, armazenar, gerenciar e analisar” (p. 1)
BOYD e CRAWFORD (2012)	“Um fenômeno cultural, tecnológico, acadêmico e que repousa sobre a interação de tecnologia, análise e mitologia” (p. 663).
DUMBILL (2012)	São dados que excedem a capacidade de processamento dos sistemas de banco de dados convencionais” (<i>online</i>).
SCHÖNBERGER-MAYER E CUKIER (2013)	“Refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar novas formas de valor de maneira que alterem os mercados, as organizações, a relação entre cidadãos e governos, etc” (p. 4).
MOURA; AMORIM (2015)	“expõe uma nova geração de tecnologia e arquitetura, destinada a extrair valor de uma imensa variedade de dados permitindo alta velocidade de captura, descoberta e análise, transformando dados em informações valiosas” (p. 2).
GOULARTE; ZILBER; PEDRON (2015)	“não se trata apenas de uma ferramenta, mas é, em verdade, uma geração de novas tecnologias e arquiteturas projetadas para extrair valor econômico de grandes volumes de dados” (p. 3)

MENEZES; FREITAS;
PARPINELLI (2016)

“inúmeras bases de dados estão tendenciando a possuir grande volume, alta velocidade de crescimento e grande variedade. Esse fenômeno é conhecido como Big Data e corresponde a novos desafios para tecnologias clássicas como Sistema de Gestão de Banco de Dados Relacional” (p. 1).

Fonte: elaboração dos autores.

Nesse contexto de massa de dados em abundância, pode-se apontar atividades relacionadas à Gestão dos Dados (*data management*) que estão divididas em: a) Uso e análise dos dados: exploração das ferramentas tecnológicas para analisar os dados a serem adotados nas empresas com finalidades nos negócios e aplicações de inteligência; b) Gestão e armazenamento de dados: passa pelo entendimento de uma infraestrutura de armazenamento e recursos para manipulá-los (DI MARTINO, 2004). O gerenciamento de dados, neste contexto, é considerado o aspecto central em um ambiente permeado por dados complexos tendo em vista que requer acesso para alavancar o conjunto de funcionalidades intrínsecas ao fenômeno análise transversal dos dados (SPECHT *et al.* 2015). Para eles, gerenciamento de dados é o desenvolvimento e implementação de políticas, planos e processos que gerenciam estes dados para manter a integridade, de segurança e de utilização dos dados. O resultado é ideal para os dados a serem auto-descritos para que outros possam descobrir e reutilizá-lo de forma eficaz. Desta maneira, é necessária para solucionar problemas complexos de equipes multidisciplinares capazes de manejar a abundância de dados (SPECHT *et al.* 2015).

A Gestão de dados é um conceito amplo que atua em três aspectos: operacional, tático e estratégico (LOPES, 2012, p. 9). Além disso, trata-se de uma disciplina que tem como finalidade “definir, planejar, implantar e executar: estratégias, procedimentos e práticas necessárias para gerenciar de forma efetiva os recursos de dados e informações das organizações, incluindo planos para sua definição, padronização, organização, proteção e utilização.” Apesar de todo potencial tecnológico que rege o fenômeno para a gestão de dados, a quantidade e variedade de dados não asseguram a melhor decisão, já que “a demora na análise dos dados pode nos levar a um atraso na tomada de decisão. Então, ou você perde a oportunidade da travessia, ou age com atraso, o que pode ser fatal” (COSTA; CAVALCANTI, 2014, *online*).

A gestão de dados no contexto do *Big Data* acarreta também impactos que estão alicerçados em quatro pilares: técnico, social, político e legal. Autores destacam dois

aspectos que merecem um olhar mais crítico quanto a benefícios que pode ocasionar: 1) Em função dos grandes volumes e diversidade de fontes, torna-se fácil perder-se no oceano dos dados e não obter resultados significativos. Para um resultado mais produtivo, deve-se tornar um imperativo saber o que e onde buscar tais informações. Os mapas, nesta seara, são relevantes para uma compreensão geográfica dos dados; e 2) Dificuldade de minerar dados algoritmicamente na busca por padrões originais em um ambiente de complexidade. “Os resultados que podem ser obtidos com algoritmos rígidos serão tão menos significativos quanto maior for a necessidade de integrar dados de fontes e formatos diferentes” (FERNANDES; LIMA-MARQUES, 2015, p. 69-70).

Deste modo, atributos como competência e experiência são os principais combustíveis para o aprendizado e inovação nas organizações em ambientes complexos (CHOO, 2003). Sendo assim, as organizações/empresas devem estar inseridas numa logística contínua e mutável, que renove o estoque de conhecimento e pratica o processo de informação com vistas à tomada de decisões (CHOO, 2003). Neste sentido de ambiente dinâmico e inovador, a globalização é apontada como indutor do crescimento explícito em que as organizações devem adaptar-se a ambientes cada vez mais complexos, como a possibilidade de expansão dos conhecimentos e capacidades para aquisição de novos conhecimentos, estabelecimento de conexões de conhecimento (TARAPANOFF; ALVARES, 2015).

Importante ressaltar que a gestão de dados está diretamente vinculada à gestão de conhecimento, em que características como criar significados, construir conhecimentos e tomar decisões revelam como usos estratégicos para as empresas se adaptarem às mudanças (CHOO, 2003). O autor enfatiza que as novas formas de conhecimento e de ação surgem quando a tensão entre o velho e o novo, entre mudança e estabilidade são confrontadas, interpretadas e resolvidas da melhor forma, uma vez que a construção do conhecimento depende de experiências coletivas. Chyi Lee e Yang (2000) consideram que o conhecimento está contido no padrão geral da organização em rede, numa dimensão interativa e compartilhada e que apenas confere sentido quando os profissionais envolvidos leem, entendem, interpretam e aplicam esse conhecimento para funções específicas de trabalho. Assim, ainda segundo as autoras, “conhecimento torna-se visível quando pessoas

experientes colocam em prática as lições aprendidas muitas vezes” (CHYI LEE; YANG, 2000, p.2)².

Entretanto, no ambiente permeado por dados com vistas à inovação e gestão do conhecimento, faz-se necessária a diferenciação entre dado, informação e conhecimento, que são os que sustentam a atividade de gestão do conhecimento, uma vez que “a *Web* é um espaço distribuído de informação que provê uma abundância de fontes de informação heterogêneas e em muitos formatos” (FERREIRA; SANTOS, 2013, p. 13). De acordo com Gutiérrez (2008), **dado** é visto como suporte físico da informação. “São fatos físicos que não contém um significado inerente, não incluem necessariamente interpretações ou opiniões, e não levam a nenhum significado que possa revelar sua importância ou relevância” (p.37). Já **informação** é definida como “o conteúdo semântico dos dados. Neste sentido, a informação possui uma natureza física ou material, mas uma natureza conceitual, pertencente ao território do conceitual” (p. 39). Ele afirma que a informação é o conteúdo semântico do dado derivado de uma chave de codificação (dado+ chave de codificação)” (p. 40). **Conhecimento** é determinado como “estados mentais de um indivíduo construído a partir da assimilação da informação e que guiam as ações do próprio sujeito” (p. 42). Assim, cria-se um novo conhecimento a partir da obtenção de uma nova informação. Para ele, é imprescindível que esse conhecimento seja valorizado nas organizações.

Nesta perspectiva, o gerenciamento de dados no *Big Data* é uma nova fonte de valor econômico e inovação. “O predomínio do Big Data representa mudanças na forma como analisamos informações que transformam a maneira como entendemos e organizamos a sociedade” (SCHÖNBERGER-MAYER; CUKIER, 2013, p. 8). A partir dessa constatação, parte-se, neste artigo, para a exploração do cenário da Gestão de Dados através da inserção do *Big Data* enquanto fenômeno catalizador de transformações estruturais no contexto da Ciência da Informação e das interfaces oriundas desta dimensão dos dados na sociedade e nas organizações.

3 PROCEDIMENTOS METODOLÓGICOS

² No original: “Knowledge becomes visible when experienced people put the lessons learned into practice oftentimes”.

Para essa investigação, realizou-se uma pesquisa exploratória-descritiva, adotando a abordagem qualitativa acerca da temática para tentar delinear o perfil profissional em torno dos cientistas de dados como ator emergente e apontar as habilidades que estes devem possuir cujo objetivo principal é fornecer um panorama geral que descreve as características de determinado fenômeno. Para Gil (2008, p. 46) “As pesquisas descritivas são, juntamente com as exploratórias, as que habitualmente realizam os pesquisadores sociais preocupados com a atuação prática”. Nessa linha de raciocínio, ainda segundo o autor, “são também as mais solicitadas por organizações como instituições educacionais, empresas comerciais, partidos políticos etc”.

Para a pesquisa, adotou-se um levantamento bibliográfico e documental que possibilitou elaborar as características do cientista de dados na era do *Big Data* a partir da literatura e do estado da arte. A pesquisa bibliográfica seguiu alguns passos de acordo com Lima e Mioto (2007, *online*), a saber:

- a) Etapa 1: **Definição do percurso metodológico**: pesquisa exploratória-descritiva e documental;
- b) Etapa 2: **Levantamento bibliográfico**: que adota os seguintes passos (Leitura de reconhecimento do material bibliográfico; Leitura exploratória; Leitura seletiva; Leitura reflexiva ou crítica e Leitura interpretativa).
- c) Etapa 3: **Coleta de Dados**: Foram adotados os seguintes passos: o **parâmetro temático** – as obras vinculadas ao objeto de estudo, em conformidade com os temas; b) o **parâmetro lingüístico** – obras nos idiomas português, inglês, espanhol, etc.; c) as **principais fontes** que se pretendem consultar – livros, periódicos, teses, dissertações, coletâneas de textos, etc.;
- d) Etapa 4: **Análise explicativa das soluções**: se constrói através dos dados obtidos no material bibliográfico escolhido, como também fundamentado no referencial teórico empreendido. Nesta direção, a exposição é realizada considerando os procedimentos determinados em conformidade com as obras pesquisadas.

No que se refere à realização da análise empreendida foram selecionados artigos de periódicos, trabalhos publicados em anais de eventos técnico-científico, livros específicos da temática abordada, tanto nas plataformas brasileiras quanto nas estrangeiras sobre o tema em questão nessa área de conhecimento. Ainda como estratégias de coleta de dados,

explorou-se o Portal de Periódico Capes, SciELO, Anais de eventos do Enancib. Sendo assim, foram elencados na busca os seguintes termos, em dois idiomas, inglês e português: “big data”, “gestão de dados”, “cientista de dados”, “profissional da informação”, “desafios na gestão de dados”, “data scientist”, “big data challenges”, “Data management on big data”, “Professional Profile”, que foram coletados nos campos do assunto, resumo, título e palavras-chave como fontes de coleta de dados. O período da coleta de dados foi procedido entre 1 a 30 de março de 2017.

4 DISCUSSÃO SOBRE PERFIL DO CIENTISTA DE DADOS

Os resultados mostram que, ao mesmo tempo em que os desafios estão postos aos profissionais de modelar o perfil para trabalhar grandes volumes de dados, os mesmos estão direcionados ao mercado de trabalho e demais organizações em encontrar o perfil adequado para atuar diretamente com os dados (Quadro 2). O monitoramento de palavras-chaves e de termos sobre a gestão de dados e cientista de dados nos revela, ainda de forma preliminar, que o *Big data* trouxe implicações para o perfil dos profissionais que lidam com mega dados. Embora a profissão esteja em alta devido às potencialidades trazidas pelo *Big Data* há, contudo, um movimento inverso, uma escassez desses profissionais que estejam com o currículo em sintonia com o mercado de Tecnologia da Informação (DAVENPORT; PATIL, 2013). A partir dessa discussão empreendida à luz da literatura e da exploração do cenário, traçamos um perfil com habilidades e característica que os profissionais devem adquirir ou possuir para atuar na área de *Big Data* voltado para a gestão de dados fazendo comparação entre perfil desejado pelo mercado e a efetivação do profissional.

Quadro 2 – Comparativo dos desafios do cientista de dados em duas perspectivas: mercado de trabalho e profissionais.

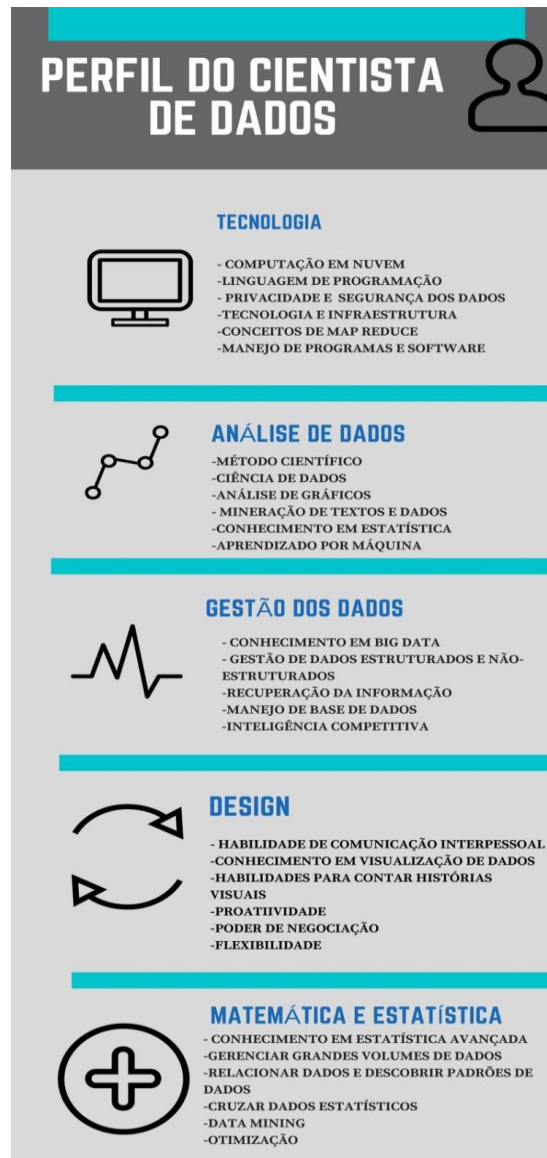
Mercado de trabalho	Profissionais
Localizar os cientistas de dados e entender qual sua função na empresa	Capacidade de estruturar grandes volumes de dados amorfos
Checar a qualificação profissional que precisar ter ao lidar com grandes volumes de dados	Tornar os dados possíveis para análise
Procurar profissionais que domine as novas tecnologias e métodos para trabalhar com dados	Identificar fonte de grandes volumes de dados e cruzar com outras fontes

Certificar se o profissional é capaz de encontrar boas histórias em meio à uma avalanche de dados e apresentar uma narrativa coerente	Criar ferramentas e analisar grande quantidade de dados
Capacidade de comunicar-se através de cifras, visual e verbalmente	Domínio de ferramentas que deem conta do volume de dados (Hadoop, por exemplo)
Buscar profissionais que saibam escrever códigos	Formação em qualquer área, desde que tenha foco em dados e na computação

Fonte: Elaboração dos autores com base em Davenport e Patil (2013).

Observe que as habilidades se apresentam incluem, entre outros, manuseio das ferramentas de código aberto, *cloud computing* (computação em nuvem) e visualização de dados (DAVENPORT; PATIL, 2013). Contudo, embora o profissional deva ter o domínio da parte operacional de gerenciar esses dados e escrever códigos (mas que este pode enfrentar dificuldades técnicas), mais relevante “será a necessidade de esse profissional se comunicar em linguagem que toda parte interessada possa entender — e de exibir particular habilidade para contar histórias com dados de forma verbal ou visual (ou, de preferência, ambas)”(DAVENPORT; PATIL, 2013, *online*). Para Gittlen (2015), o que diferencia os cientistas de dados além de outros que lidam com o volume de dados é sua capacidade de criar lógica por trás dos dados que levam a decisões de negócios ou da gestão em si dos dados para fomentar novas possibilidades de uso. A preocupação em compreender os desafios e cuidados em gerenciar os volumes de dados tem recebido uma atenção de especialistas na área com o intuito de oferecer possíveis soluções neste ambiente novo e de competitividade. Nesta direção, e através de nosso estudo, foi possível traçar o perfil profissional que um cientista de dados na contemporaneidade deve conter para estar apto a administrar grandes volumes de dados (figura 1).

Figura 1 – Delineamento do perfil do cientista de dados.



Fonte: Elaboração dos autores a partir dos resultados obtidos.

Nestas atribuições do perfil do profissional que, em tese, está apto a trabalhar com grandes volumes de dados, imerso em um universo da cultura do *Big Data* (BARLOW, 2013), essa construção está diretamente vinculada ao modo de como a informação é tratada, refinada, processada, gerando algum tipo de valor, conhecimento e inovação para as organizações. Nesta conjectura, os cientistas de dados, munidos com suas habilidades, podem tornar-se peças centrais para a gestão do conhecimento, partilha, e competitividade na “exaustão de dados”. A importância atribuída ao perfilar este tipo de profissional reside na necessidade da emergência de um cenário complexo. Gehl (2015) usa como sinônimo de cientista de dados, os “trabalhadores do conhecimento” aos que se destinam a se integrar no universo não apenas tratando dados, mas compartilhando conhecimento com

perspectivas de ganhos sociais, econômicos e políticos, em um sistema comunicacional de ubiquidade dos dados.

Deste modo, a pesquisa exploratória permitiu apontar quatro áreas que fundamentam os aspectos do cientista de dados:

a) Tecnologia: Dentre o arsenal que compõem o rol de habilidades que o profissional deve ter, maior manejo e familiaridade no aparato tecnológico, mais vantagem terá na hora de gerenciar dados. Domínios em computação em nuvem, linguagem de programação, segurança e privacidade dos dados, tecnologia e infraestrutura conceitos de *Map Reduce*, manejo de programas e *software*;

b) Análise de dados: Domínios em: método científico, conhecimento em Ciência de Dados (*Data Science*), análise de gráficos, mineração de dados e textos, conhecimento em estatística e aprendizado por máquina;

c) Gestão de dados: Domínios em: *Big Data (Hadoop, SQL, Spark)*, gestão de dados estruturados (RDB, XML), gestão de dados não-estruturados (bancos de dados *NoSQL*), recuperação dos dados, inteligência competitiva;

d) Design e Comunicação: Domínios em: comunicação interpessoal, conhecimento em visualização de dados, habilidades de contar histórias visuais, proatividade, poder de negociação, flexibilidade;

e) Matemática e estatística: Domínios em: Conhecimento em estatística avançada, gerenciamento de grandes volumes de dados, relacionar dados e descobrir padrões, cruzar dados estatísticos, *data mining*, otimização.

Entretanto, é pertinente colocar que cada vez mais o trabalho com mega dados vai além do cientista de dados e envolve equipes interdisciplinares para lidar com todas essas dimensões em termos de capacidade de domínio sobre tecnologia, análise de dados, gestão de dados, design e matemática e estatística. Os riscos e desafios impostos pelo *Big Data* devem ser entendidos não apenas aos profissionais e pesquisadores interessados nos benefícios econômicos que podem trazer vantagens competitivas, mas reconhecendo o fenômeno e os impactos e oportunidades enquanto possibilidades de soluções holísticas na sociedade. De acordo com Fernandes e Lima-Marques (2015) alguns pontos precisam ser mais bem avaliados tanto do ponto de vista empresarial quanto de perfil profissional:

- a) Desafios de ordem técnica ainda a serem solucionados de forma eficiente;
- b) Interfaces interativas de baixo atrito;
- c) Técnicas e ferramentas de *backup*;
- d) Pessoal qualificado insuficiente;
- e) Crescimento acelerado de aplicações e do surgimento de novas empresas, agravando a competição tanto de mercados específicos como por mão de obra qualificada;
- f) Momento histórico de volatilidade tecnológica, com risco de surgimento e rápido desaparecimento de inúmeros produtos e, eventualmente, de suas respectivas empresas;
- g) Momento de disputa intensa entre as empresas bem-sucedidas com o lançamento de novos produtos, as quais com o tempo deverão ser incorporados por outras maiores, restando, ao final desse período de maior estabilidade, apenas as empresas que melhor conseguirem se adaptar ao novo cenário tecnológico e econômico. (FERNANDES; LIMA-MARQUES, 2015, p. 87-88).

As características para a formação de um cientista de dados é o reflexo do modo de como a sociedade contemporânea está amalgamada na cultura do *Big Data*, isto é, uma mistura de pessoas e processos e “é parte integrante e fundamental para o sucesso de qualquer nova implantação de tecnologia ou implementação desta. Da mesma maneira que uma planta precisa de água para crescer, a tecnologia precisa de pessoas e processos para prosperar e suceder” (BARLOW, 2013, p.2).

Nesta assertiva, o elemento diferenciador na cultura do *Big Data* seria o todo o processamento dos dados em ambientes interativos, alavancados pelos avanços tecnológicos. Além disso, o verdadeiro desafio será a integração das organizações existentes, recursos humanos aptos para administrar dados em abundância que possibilite o cientista de dados a gerar valor para as organizações tendo em vista que o volume de dados é crescente nos processos da sociedade.

4 CONSIDERAÇÕES FINAIS

As mudanças estruturais advindas do grande volume de dados produzem, de fato, trouxe impactos ao campo científico e nas organizações a partir da noção do conceito de *Big Data* e do desdobramento em torno da tendência de perfil profissional para cientista de dados quanto à capacidade de lidar com a gestão desses dados. O *Big Data* pode se apresentar com vistas à inovação, competitividade e tomada de decisões num universo em constante crescimento, e com demanda para novas oportunidades de atuação para profissionais de Tecnologia e Informação (TI). Por outro, há controvérsias e tensões no entorno do conceito e dos procedimentos de gestão de dados como a questão da fidelidade, de tecnologias para gerir e, principalmente, no delineamento de novo perfil profissional, ou seja, qualificação profissional para atender a um perfil que seja adequado para o gerenciamento de dados abundantes, análises de dados e indicativos de tendências.

O cientista de dados é posicionado como o profissional em ascensão no contexto gerido por grandes volumes de dados. A questão, contudo, não se limita a manejar o ferramental tecnológico e programas de *software* que deem conta do volume de dados, mas encontrar soluções estratégicas para transformações disruptivas, à medida que as empresas e organizações inserem o *Big Data* em suas decisões e administração. Tal fato requer orientações inovadoras das informações, que vão desde o planejamento à tomada de decisões. A construção de um perfil adequado para atuar no mercado de *Big Data* começa a se construir e se consolidar cada vez mais, tendo em vista a polivalência de aptidões exigidas para este tipo de profissional que resulta entrecruzamento de domínio tecnológico e capacitação adequada.

Na perspectiva da Ciência da Informação, a exploração desses elementos estruturais abordados em torno do *Big Data* gera uma nova compreensão para a complexidade dos dados em retrospectiva com o próprio surgimento da Ciência da Informação que foi criada para dar conta da quantidade de informações e documentos gerados no pós-guerra. Neste momento, essa configuração se apresenta novamente através dos dados em grande escala como um desafio que impulsiona a CI para pensar do ponto de vista metodológico e teórico sobre os desdobramentos do fenômeno na Gestão de Dados e na construção do perfil do profissional Cientista de Dados.

REFERÊNCIAS

BARLOW, Mike. **The culture of Big Data**. Sebastapol-CA: O`Reilly Media, 2013.

BARRETO, A. **Os três tempos da Ciência da Informação**. 16 maio 2014. Disponível em em: <http://aldobarreto.wordpress.com>. Acesso em: 14 dez de 2015.

BOYD, D; CRAWFORD, K. Six Provocations for Big Data). A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011. Disponível em: <https://goo.gl/SYha3> Acesso em: 9 fev 2017.

CHYI LEE, C; YANG, J. Knowledge value chain. **Journal of management development**, v. 19, n. 9, p. 783-794, 2000.

CHOO, C. W. **A organização do conhecimento**: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões tradução Eliana Rocha. - São Paulo: Editora Senac São Paulo, 2003.

CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E. ; SANT'ANA, Ricardo César Gonçalves . Big Data: fatores potencialmente discriminatórios em análise de dados. **Em Questão**, v. 23, p. 62-86, 2017.

COSTA, L. S; CAVALCANTI, M. Big Data não é tecnologia. **DataGramZero - Revista de Informação** - v.15 n.3 jun/14.

DAVENPORT , T. H., **Cientista de dados**: o profissional mais cobiçado do século 21. Havard Business Review. 2013.

DI MARTINO, B. et al. Big data (lost) in the cloud. **International Journal of Big Data Intelligence**., Vol.1, No.1/2, pp.3 – 17, 2014.

DUARTE, E. N; COSTA, L. F; SANTOS, L. F; SANTOS, J. L; MORAES, F. M. Comportamento e Competência em informação: uma experiência de extensão universitária. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v.18, n.1, p. 553-575, jan./jun., 2013. Disponível em: <https://goo.gl/MKP1VX> Acesso em: 18 jul. 2017.

DUMBILL, E. **What is big data?** An introduction to the big data landscape. O`Reilly Media, Inc., 2012. Disponível em: <https://goo.gl/ihLV94> . Acesso em 1 dez. 2016.

EINSENSTEIN, M. The Power of Petabytes. S2. **Nature**. Vol 257. Novembro, 2015.

FERREIRA, J. A. ; SANTOS, P. L. V. A. C. . O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Informação & Sociedade (UFPB. Online)**, v. 23, p. 13-23, 2013.

FERNANDES; G. L; LIMA-MARQUES, M. Competitividade na era do Big Data: uma perspectiva a partir da disciplina de arquitetura da informação. In: **Análise da Informação para a tomada de decisão**. TARAPANOFF, Kira (org.). Curitiba: Intersaberes, 2015.

FREDERIKSEN, L. (2012). Big Data. Public Services Quarterly, 8(4), 345-349.

GEHL, R. W. (2015). Sharing , knowledge management and big data : A partial genealogy of the data scientist. Disponível em: <https://goo.gl/iBSLxE> Acesso em: 7 jul 207

GIL, A. C. **Métodos e Técnicas de pesquisa social**. 6 ed. São Paulo: Atlas, 2008.

GITTLEN, S. **Could data scientist be your next job?** The definition of a data scientist varies, but no matter how it's defined, the skills are in demand. Network World 4 June 2012. Academic OneFile. Web. 7 Dec. 2015.

GUTIÉRREZ, M. P. M. Gestión del conocimiento en las organizaciones: fundamentos, metodología y). **El profesional de la información**, 2009, enero-febrero, v. 18, n. 1, pp. 111-113. Disponível em: <https://goo.gl/dohtWE> Acesso em: 8 mai. 2017.

LIMA; T.C.S; MIOTO, R.C.T. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. 2007. Disponível em: <https://goo.gl/czfNb0> Acesso em: 4 set 2017.

LOPES, B. **Métodos ágeis e Gestão de dados moderna**. Disponível em: www.bergsonlopes.com.br Acesso em: 4 dez. 2016.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. Big Data: The Next Frontier for Innovation, Competition, and Productivity. **McKinsey Global Institute**, May 2011. Disponível em: <https://goo.gl/my4utS> Acesso em: 3 nov. 2016.

MENEZES; S. R. L; FREITAS; R. S; PARPINELLI, R.S. Mineração em Grandes Massas de Dados Utilizando Hadoop MapReduce e Algoritmos Bio-inspirados: Uma Revisão Sistemática. **Rev. Inform. Teor. Apl. (Online)**. Porto Alegre: 2016.

MOURA, A. L. T.. AMORIM, D. G. Big Data: O Impacto e sua funcionalidade na sociedade tecnológica. **Revista OPARA**, 2014 .

NASSIF, M. E; SANTOS, E. L. O profissional da informação em atividades de inteligência competitiva. **Inf. Inf.**, Londrina, v. 14, n. 2, p. 21 - 37, jul./dez. 2009.

NONAKA, I; TAKEUCHI, H. **Criação de conhecimento na empresa**: como as empresas japonesas geram a dinâmica da inovação. Rio de Janeiro: Elsevier, 1997.

SANTOS, C. J. Atuação profissional da informação no processo de inteligência competitiva. **REBECIN**, v.3, n.2, p.26-50, jul./dez. 2016. Disponível em: <https://goo.gl/dFqbSL> Acesso em: 24 jul. 2017.

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.

SPECHT, A; GURU, S; HOUGHTON, L; KENIGER, L.; RITCHIE, E.G; LAI, K; TREOLAR, A. Data Management challenges in analysis and synthesis in the ecosystem sciences. **Science of the Total Environment** 534 (2015) 144- 158.

SCHÖNBERGER-MAYER, V; CUKIER, K. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

TARAPANOFF, K; ALVARES, L. Perspectivas em inteligência organizacional e competência na web: uma visao geral. In: **Análise da Informação para a tomada de decisão**. TARAPANOFF, Kira (org.). Curitiba: Intersaberes, 2015.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP