

BIG DATA: O NOVO DESAFIO PARA GESTÃOPriscila Nesello¹
Ana Cristina Fachinelli²**RESUMO**

A explosão informacional nos anos 2000, deu origem ao fenômeno definido como *big data*. Este fenômeno se refere a terceira época da era da informação e difere das outras épocas pois atualmente a análise pode ser realizada com dados em sua forma original, não estruturada, aumentando os níveis de previsibilidade de diferentes eventos. Este é um tema emergente, e para analisar o seu crescimento em termos de relevância, o presente artigo apresenta um estudo bibliométrico cujo objetivo foi de analisar o crescimento das publicações sobre *big data* e análise em IE nas bases de dados *Scopus* e *Web of Science*, desde o ano de 2010. Para isso foi analisada a totalidade de artigos referentes aos temas, os tipos de publicação, as fontes que estão publicando por área e seu fator de impacto, bem como o índice H. Os resultados indicam que embora algumas fontes de alto fator de impacto estejam publicando sobre o tema, a carência de estudos em *big data* e análise, fica evidenciada, principalmente no que tange às ciências sociais.

Palavras-Chave: *Big Data*. Análise. Bibliometria. Inteligência Estratégica.

1 INTRODUÇÃO

Big data se refere à terceira época da era da informação (MINELI, CHAMBERS E DHIRAJ, 2013). A primeira foi em 1954, com a implementação dos sistemas de informação nas organizações. Após 35 anos, o foco em processos internos, cedeu lugar para as interações externas. Com isto teve início a época da rede, que juntamente com o incremento da globalização tornou o ambiente empresarial ainda mais complexo. Atualmente, o fenômeno *big data* surge viabilizado pelo aumento do poder de processamento que de acordo com a lei de Moore, dobra a cada dois anos.

Letouzé (2012) indica que a quantidade de dados digitais disponíveis em nível global cresceu de 150 *exabytes* em 2005 para 1200 *exabytes* em 2010 e deve aumentar em 40% ao

¹ Mestranda em Administração da Universidade de Caxias do Sul - pri.nesello@gmail.com

² Doutora em Ciências da Informação e da Comunicação pela Université de Poitiers, França. Professora pesquisadora do Centro de Ciências da Comunicação e do Programa de Pós Graduação em Administração da Universidade de Caxias do Sul - afachinelli@gmail.com

ano para os próximos anos. Esta taxa de crescimento significa que o estoque de dados digitais deve aumentar 44 vezes entre 2007 e 2020, dobrando a cada 20 meses. Segundo Rapporteur (2010), a explosão das redes móveis, a computação em nuvem e as novas tecnologias deram origem a um número incompreensivelmente grande de informações. Para a IE, a oportunidade está no fato de que usando técnicas de correlação avançada, analistas podem peneirar densos trechos de dados para prever condições, comportamentos e eventos antes que eles aconteçam.

Em termos de tecnologia, Cukier (2010) indica que os negócios em gestão da informação estão contribuindo para facilitar o processo de construção de sentido nos grandes volumes de dados. Por outro lado, as contribuições de *big data* são possíveis de serem realizadas desde que a matéria-prima informacional, gerada pela exploração dos dados, seja transformada em inteligência, por meio do processo de análise. A análise em IE, na abordagem de Fleisher e Bensoussan (2007), é definida como a aplicação hábil de métodos e processos científicos e não científicos pelos quais os indivíduos interpretam dados ou informações para produzir resultados de inteligência perspicazes e recomendações viáveis para os tomadores de decisão.

Por outro lado, e acordo com Cukier (2010), os efeitos do fenômeno *big data* vem sendo percebidos em todos os lugares, nas empresas, na ciência e no governo. Seu conceito está migrando para todos os campos do conhecimento humano pois em essência seu avanço é uma continuação da antiga busca da humanidade em medir, registrar e analisar o mundo (SCHÖNBERGER-MAYER E CUKIER, 2013).

Considerando esse cenário, o foco deste trabalho está no efeito do *big data* na ciência, mais especificamente na evolução da produção científica sobre *big data* e análise. Assim, o objetivo do presente estudo foi de analisar o crescimento das publicações sobre *big data* e análise em IE.

2 O IMPACTO DO *BIG DATA* NA ANÁLISE DE DADOS

Ciências como a astronomia e a genômica, que vivenciaram uma explosão informacional nos anos 2000, cunharam o termo *big data* (SCHÖNBERGER-MAYER E CUKIER, 2013). Este termo, relacionado aos grandes volumes de dados, foi primeiramente citado no relatório "*Data, data, everywhere: a special report on managing information*", do periódico britânico The Economist (CUKIER, 2010). Entretanto, durante estes anos o termo

foi sendo utilizado e relacionado a *datawarehouses* ou soluções de *business intelligence* (BI), com *data sets* de *terabytes* de dados. O fato é que *big data* representa muito mais que isto (TAURION, 2013) e atualmente, o conceito está migrando para todos os campos do conhecimento humano (SCHÖNBERGER-MAYER V.; CUKIER K., 2013).

No relatório "*Data: the next frontier for innovation, competition, and productivity*" é apresentada a definição de *big data*, como "o conjunto de dados cujo tamanho vai além da capacidade para capturar, armazenar, gerenciar e analisar de ferramentas de *software* de banco de dados típicos" (MANYIKA ET AL., 2011). Outras definições para o termo são apresentadas na Tabela 3, porém sem rigidez conceitual:

Tabela 1 - Definições de *big data*.

| Autor | Definição |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AKERKAR (2014) | <i>Big Data</i> refere-se a conjuntos de dados, cujo tamanho está além das capacidades da tecnologia de banco de dados atual. É um campo emergente onde a tecnologia inovadora oferece alternativas para resolver os problemas inerentes que aparecem quando se trabalha com dados massivos, oferecendo novas maneiras de reutilizar e extrair valor a partir de informações. |
| DUMBILL (2012) | <i>Big data</i> são os dados que excedem a capacidade de processamento de dados de sistemas convencionais. |
| LOHR (2012) | <i>Big data</i> é um termo de marketing, mas também um atalho para o avanço de tendências em tecnologia que abrem a porta a uma nova abordagem para a compreensão do mundo e da tomada de decisões. |
| MINELI, CHAMBERS E DHIRAJ (2013) | <i>Big data</i> é a próxima geração de <i>data warehousing</i> e análise de negócios e está pronta para entregar receitas economicamente eficientes para as empresas. Este fenômeno se deve, em maior parte, ao rápido ritmo de inovação e mudança que estamos vivenciando hoje. |
| SATHI (2012) | Existem duas fontes comuns de dados agrupados sob a bandeira <i>do big data</i> . A primeira são os dados internos (dados estruturados, não estruturados ou semiestruturados) da organização que, graças à automação e acesso estão sendo cada vez mais compartilhados. A segunda são os dados de fora da organização, como as informações disponíveis em sites de mídia social, literatura do produto distribuído livremente pelos concorrentes, hierarquias organizacionais dos clientes corporativos, dicas úteis disponíveis a partir de terceiros e reclamações de clientes postados em sites de regulamentação. |
| SCHÖNBERGER-MAYER E CUKIER (2013) | <i>Big data</i> é a capacidade de uma sociedade de obter informações de maneiras novas a fim de gerar ideias úteis e bens e serviços de valor significativo. Assim, a verdadeira revolução não está nas máquinas que calculam dados, e sim nos dados em si e na maneira como usamos. |
| TAURION (2013) | Resumindo o que é <i>big data</i> em uma simples fórmula para conceitualizá-lo: <i>Big Data</i> = volume + variedade + velocidade + veracidade, tudo agregado + valor. |
| ZHAO (2013) | <i>Big data</i> é um sonho tornado realidade para os cientistas de dados, pois se pode ter tudo para obter <i>insights</i> interessantes, que não seriam possíveis antes. <i>Big data</i> não transforma informação e conhecimento sem análises detalhadas. Requer soluções de armazenamento grandes e escaláveis, bem como capacidades e aplicações de análise escaláveis. Análise não significa que se pode jogar dados em alguma <i>machine-learning</i> e algoritmos estatísticos, tais como redes neurais, árvores de |

| | |
|--|-----------------------------------------------------------------------------------------------------|
| | decisão, máquinas de apoio vetor, e assim por diante e esperar ter bons resultados automaticamente. |
|--|-----------------------------------------------------------------------------------------------------|

Fonte: elaborado pelo próprio autor.

De acordo com Mineli., Chambers e Dhiraj (2013), *big data* se refere a terceira época da era informação. A primeira foi em 1954, com a implementação de um sistema de folha de pagamento pela General Electric Corporation, por Joe Glickauf e Arthur Andersen em um computador eletrônico digital. Foi então introduzida a época computacional da era da informação nas corporações Americanas. Em meio a década de 1950 outras corporações rapidamente adotaram sistemas para servir a um amplo espectro de processos corporativos. Nesta época também tiveram início as empresas de consultoria em TI.

Há aproximadamente trinta anos atrás, Leonard Kleinrock, Lawrence Roberts, Robert Kahn e Vint Cerf inventaram a internet e a segunda época da era da informação, a era da rede (MINELI, CHAMBERS E DHIRAJ, 2013). Os primeiros 35 anos da digitalização tinham foco em processos internos, posteriormente o foco ficou concentrado mais em interações externas. Como um tipo de evolução, as organizações passaram a ver mais, predizer mais profundamente o futuro, e dar respostas rápidas.

A complexidade do ambiente das corporações foi incrementado com a globalização, tornou-se ágil e orientado a rede e as organizações puderam passar a pesquisar com inteligência suas bases de clientes. Estas organizações tiveram que organizar suas bases, analisar padrões de e-mail, gravações telefônicas, mensagens instantâneas e outras evidências para determinar o modelo organizacional emergente. Assim, a própria forma da empresa começou a se modificar, habilitada pela tecnologia e estimulada pela necessidade de complexidade.

Esta combinação da *internet* e o *WWW* em meados dos anos 1990, no entanto, sinalizaram uma mudança radical não só na quantidade de dados - o volume, mas também na taxa com que os dados chegavam - a velocidade; na diversidade de fontes de onde eles chegavam e nos tipos de dados e sua disponibilidade - a variedade (BETSER E BELANGER, 2013) - os três Vs de *big data*. Assim, teve início mais uma era da informação a época do *big data*. Entretanto, *big data* não é *business intelligence* (BI), com "grandes" dados. Por 55 anos as corporações de tecnologia da informação dominaram uma estrutura transacional de mundo, com o foco da tecnologia da informação na automatização, eficiência e produtividade. Agora

big data representa uma transição em termos de armazenamento e análise (MINELI, CHAMBERS E DHIRAJ, 2013).

De acordo com os autores, *big data* tem o potencial para ser diferente das outras épocas por dois motivos: primeiramente, os dados podem ser analisados em sua forma original, não estruturada; e, a possibilidade de analisar não apenas o que houve no passado, mas sim prever o que irá acontecer ao redor do mundo, com riqueza de detalhes. Agora pode-se pensar em *logs* na *web*, vídeo *clips*, gravações de voz, documentos de repositórios como *share point*, dados sociais, dados abertos do governo e outros que irão compor o corpo analítico. O termo "ciência de dados" refere-se a utilização dos dados aplicados ao método científico e aos negócios.

Este fenômeno surge viabilizado pelo aumento do poder de processamento que de acordo com a lei de Moore, dobra a cada dois anos - quantidade de transistores num *chip*. Esta contínua melhoria tornou os computadores mais rápidos, e a memória mais profusa. O desempenho dos algoritmos também aumentou, segundo conselho de Ciência e Tecnologia da Presidência dos Estados Unidos. Entretanto, muitos dos ganhos com *big data*, acontecem não por causa de *chips* mais rápidos ou melhores algoritmos, mas sim pela existência de mais dados (SCHÖNBERGER-MAYER E CUKIER, 2013). Mineli, Chambers e Dhiraj (2013) exploram melhor a questão e identificam as razões para esta nova era:

1. Tempestade perfeita da computação: *big data* é o resultado natural das quatro maiores tendências globais: a lei de Moore, computação móvel (com *smartphones* e *tables*), redes sociais (Facebook, Four Square e outros), e a computação em nuvem (*cloud computing* - possibilidade de alugar ou arrendar *hardware* ou *software* para utilização);

2. Tempestade perfeita de dados: volumes de dados transacionais foram por décadas utilizados pela maioria das grandes empresas, mas o que se apresenta atualmente é mais volume, velocidade e variedade - os três Vs - de dados que chegaram de forma inédita. Esta tempestade perfeita dos três Vs, torna extremamente complexo e pesado o gerenciamento de dados atual e análise de tecnologias e práticas;

3. Tempestade perfeita de convergência: gerenciamento de dados tradicionais, análise de *software*, tecnologias de *hardware*, tecnologia de código aberto e *commodities* de *hardware* estão se fundindo para criar novas alternativas para TI e executivos de negócios para enfrentar *big data*.

Assim, os dados que antes eram considerados estáticos e banais, com término de sua utilidade depois que o objetivo da coleta era alcançado, se tornaram matéria-prima dos negócios, um recurso econômico vital, usado para criar uma nova forma de valor econômico.

A princípio, a ideia era a de que o volume de informação crescera tanto que a quantidade examinada já não cabia na memória de processamento dos computadores, por isso os engenheiros tiveram de aprimorar os instrumentos que utilizaram para fazer a análise. Esta é a origem de novas tecnologias de processamento como a MapReduce da Google e sua equivalente de código aberto, Hadoop, lançada pela Yahoo. Elas permitem que se gerenciem muito mais dados que antes, e os dados não precisam ser alocados em fileiras ou nas clássicas tabelas (SCHÖNBERGER-MAYER E CUKIER, 2013).

Meio século depois de os computadores entrarem no meio social os dados começaram a se acumular. Assim, o mundo não apenas está mais cheio de informação como também a informação está se acumulando com mais rapidez. Tecnólogos acreditam que a linhagem do *big data* remonta à época da revolução do silício. Certamente os sistemas modernos de TI possibilitaram o *big data*, mas em essência o avanço rumo ao *big data* é uma continuação da antiga busca da humanidade por medir, registrar e analisar o mundo (SCHÖNBERGER-MAYER V.; CUKIER K., 2013). A Tabela 3 mostra a evolução da utilização dos dados e informação desde sua origem até os dias atuais:

Tabela 2 - Evolução da utilização dos dados e informação.

| Ano ou Período | Fato |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 8000 a.C. | Comerciantes sumérios usavam tábuas e fichas de barro para denotar os bens comercializados. |
| 1086 | Livro Domesday, um dos mais venerados tesouros britânicos, foi uma contagem abrangente - e sem precedentes - dos ingleses, suas terras e propriedades. |
| 1439 | Primeira revolução da informação: Impressora de Gutenberg. |
| 1453 e 1503 | De acordo com a historiadora Elizabeth Eisenstein, oito milhões de livros foram impressos. Volume de informações dobra na Europa, em 50 anos. |
| 1662 | O comerciante britânico chamado John Graunt queria saber a população de Londres na época da peste negra. Em vez de contar as pessoas, ele inventou uma abordagem - que hoje chamamos de "estatística"- que lhe permitiu estimar a população. |
| 1880 | O Census Bureau dos Estados Unidos contratou o inventor Herman Hollerith, que desenvolveu um sistema de cartões perfurados e máquinas de tabulação para o censo de 1890. Ele conseguiu diminuir o tempo de tabulação do censo de oito anos para um. |
| 1920 | Descobertas da mecânica quântica abalaram para sempre o sonho das medições abrangentes e perfeitas. |
| 1934 | Jerzy Neyman, estatístico polonês, demonstrou que a abordagem da amostragem levava a erros e que para isto deveria ser considerada a aleatoriedade na escolha da amostra. A amostragem tornou-se a solução para o problema da profusão de dados. |
| 1950-1960 | Implementação de um sistema de folha de pagamento para General Electric Corporation, por Joe Glickauf e Arthur Andersen em um computador eletrônico digital. Início das ideias de "Revolução da informação" e "Era Digital" ou " Era da informação ". |
| 1980-1990 | Nesta época, Leonard Kleinrock, Lawrence Roberts, Robert Kahn, e Vint Cerf inventaram a internet. 40% de toda a potência computacional do mundo existia na forma de calculadoras de bolso. " Era da rede ". |
| 2000 | Dados: 75% dados analógicos (papel, filme, vinil, fitas magnéticas, livros, fotografia); 25% dados digitais. |
| 2007 | Dados: 7% dados analógicos; 93% dados digitais. |

| | |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2010 | Ciências como a astronomia e a genômica, vivenciaram uma explosão informacional, e cunharam o termo <i>big data</i> , que representa uma transição em termos de armazenamento e análise. " Era <i>big data</i> ". |
| 2013 | Dados: 2% dados analógicos; 98% dados digitais - 1200 <i>exabytes</i> . Volume de informações dobra a cada 3 anos. |

Fonte: elaborado pelo próprio autor com base em Schönberger-Mayer V.; Cukier K. (2013) e Mineli M., Chambers M., Dhiraj A. (2013).

Assim, a maneira utilizada para pensar a questão, proposta por Schönberger-Mayer e Cukier (2013) é que *big data* se 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 se alterem os mercados, as organizações, a relação entre cidadãos, governos, etc. Então, a sociedade precisará conter um pouco a obsessão pela causalidade e trocá-la pela correlação simples: sem saber o porquê, apenas o quê. Essa mudança subverte séculos de práticas consagradas e desafia a compreensão mais básica de como tomar decisões e compreender a realidade (SCHÖNBERGER-MAYER E CUKIER, 2013).

De acordo com Schönberger-Mayer e Cukier (2013), os frutos da sociedade da informação, como celulares e computadores, deram origem ao termo "exaustão de dados". O termo descreve a trilha digital que as pessoas deixam, que se refere a dados colhidos como subprodutos das ações e dos movimentos das pessoas. Estes elementos são facilmente identificados em todos os lugares, entretanto, a informação em si é discreta.

Segundo os autores, esta mudança quantitativa advinda de *big data* tem gerado uma mudança qualitativa em termos de resultados do processamento de dados. Betser e Belanger (2013) complementam que a mudança qualitativa refere-se a quantidade de detalhes que são apurados e mantidos por esses bancos de dados. Há também mudanças na tecnologia disponível para analisar e extrair informações a partir desses dados, no custo de disponibilidade, processamento, armazenamento de dados e mecanismos de origem/entrega, como *smartphones* e sensores. Estes por sua vez, têm gerado mudanças e criado oportunidades na busca por excelência na utilização de dados e informações.

Para exemplificar isto, podemos utilizar a analogia da nanotecnologia: quando se chega ao nível molecular, as propriedades físicas da matéria podem se alterar; assim, ao saber o que significam essas novas características, podem-se criar materiais e construir o que não podia ser feito antes - obter metais e cerâmicas mais flexíveis. Ou seja, quando aumentamos a escala de dados com a qual trabalhamos, ganhamos margem para inovar, o que não ocorria antes com poucos dados (SCHÖNBERGER-MAYER E CUKIER, 2013).

Assim, o objetivo proposto para *big data* nos negócios e na TI será a aplicação de dados e analítica para incrementar a inteligência corporativa (MINELI, CHAMBERS E DHIRAJ, 2013). Esta é uma meta completamente diferente de enquadramento para a tecnologia e vai significar novas formas de organizar e conceituar como ela é financiada e entregue atualmente. Ou seja, os autores confirmam os pressupostos de Schönberger-Mayer e Cukier (2013) ao afirmarem que com *big data*, cientistas de dados podem utilizar mais ou todos os dados para criar um modelo. Ao fazerem isto, é possível que sejam introduzidas variáveis de previsão adicional, a fim de aumentar seu nível de exatidão. Quando o histórico de *big data* é utilizado, o modelo pode identificar tendências que estão fora dos ciclos que foram utilizados na técnica de extração de dados históricos.

3 MÉTODO

O método da bibliometria pode ser definido como estudos que buscam quantificar os processos de comunicação escrita, em relação à análise da produção, disseminação e uso da informação na área de conhecimento (MACIAS-CHAPULA, 1998). Cooper e Lindsay (1998), relatou que a análise bibliométrica permite indicar aos pesquisadores dos temas já tratados em um campo de conhecimento, verificar inconsistências em estudos anteriores e determinar as lacunas que ainda existem em uma área ou em um tema.

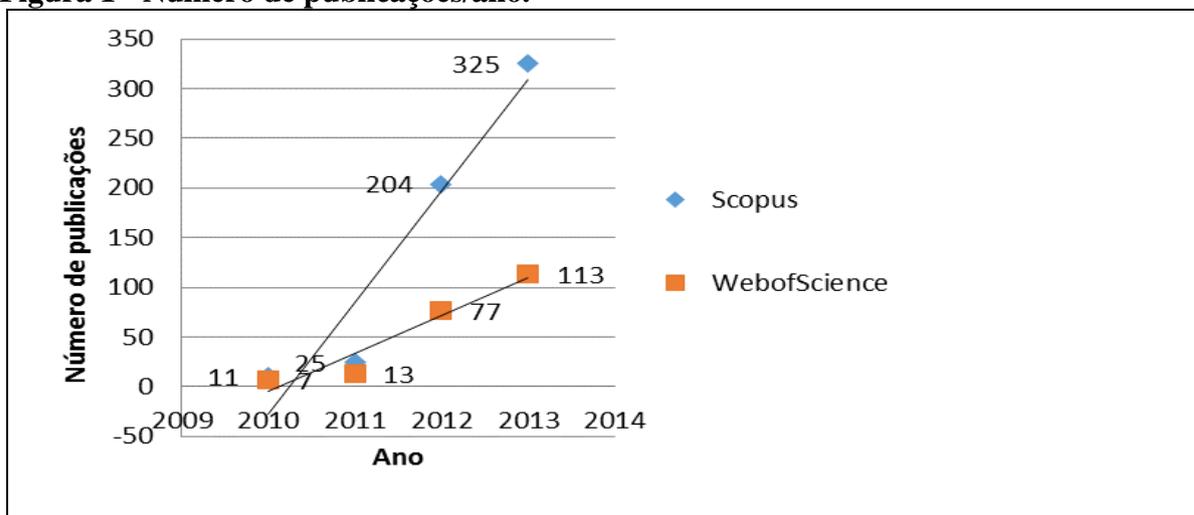
No sentido de identificar o número e tipos de publicação que estão sendo desenvolvidas sobre os temas *big data* e análise, foi realizada uma pesquisa nas bases de dados Scopus e Web of Science (WoS), desde o ano de 2010. As publicações identificadas foram analisadas com relação a área, grau de relevância - índices *SCImago Journal Rank* (SJR) da Elsevier, indexador Scopus e *Journal Citation Report* (JCR) da Thompson e Reuters, indexador WoS - *High Quality Publications* (Q1) e índice H. Posteriormente, foi realizada uma análise específica das publicações da categoria ciências sociais com foco em administração. Ou seja, o desenvolvimento do artigo se deu por meio de seis etapas: (1) coleta de artigos das bases dos indexadores; (2) análise da totalidade de artigos referentes aos temas; (3) análise dos tipos de publicação; (4) análise das fontes que estão publicando por área e seu fator de impacto, avaliado pelos índices SJR e JCR; (5) análise do número de citações, representado pelo índice H; (6) análise do número de citações, representado pelo índice H em ciências sociais.

A coleta de dados realizada durante o mês de dezembro de 2013, foi feita pela busca de estudos que empregassem os termos *big data* e análise, nos campos "título do artigo, resumo e palavras-chave" no caso do Scopus e no campo "tópico", no caso do WoS.

4 RESULTADOS E DISCUSSÃO

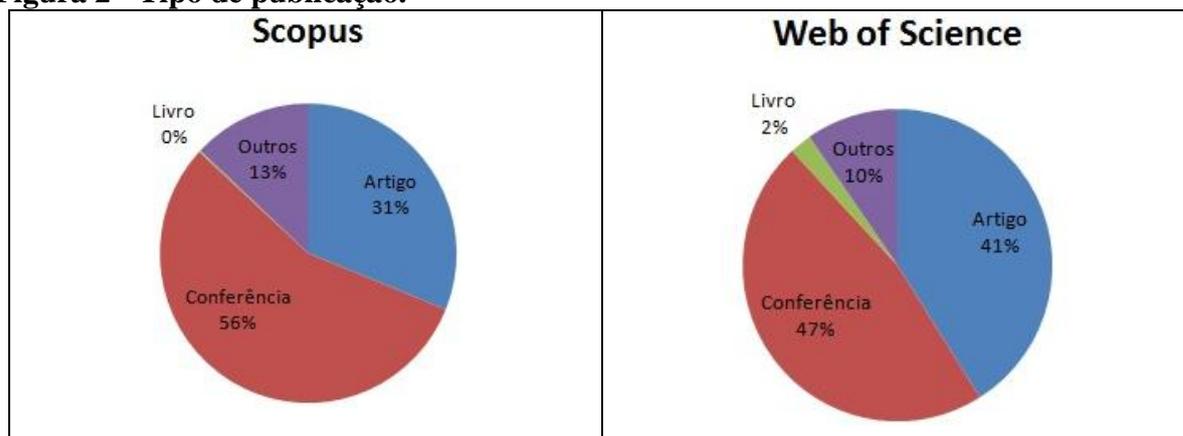
O marco do surgimento do termo *big data* e sua relação com grandes volumes de dados é o ano de 2010, considerando Cukier (2010), em relatório publicado no periódico *The Economist*. A Figura 1 mostra um crescimento médio de mais de 200% ao ano em publicações no período analisado:

Figura 1 - Número de publicações/ano.



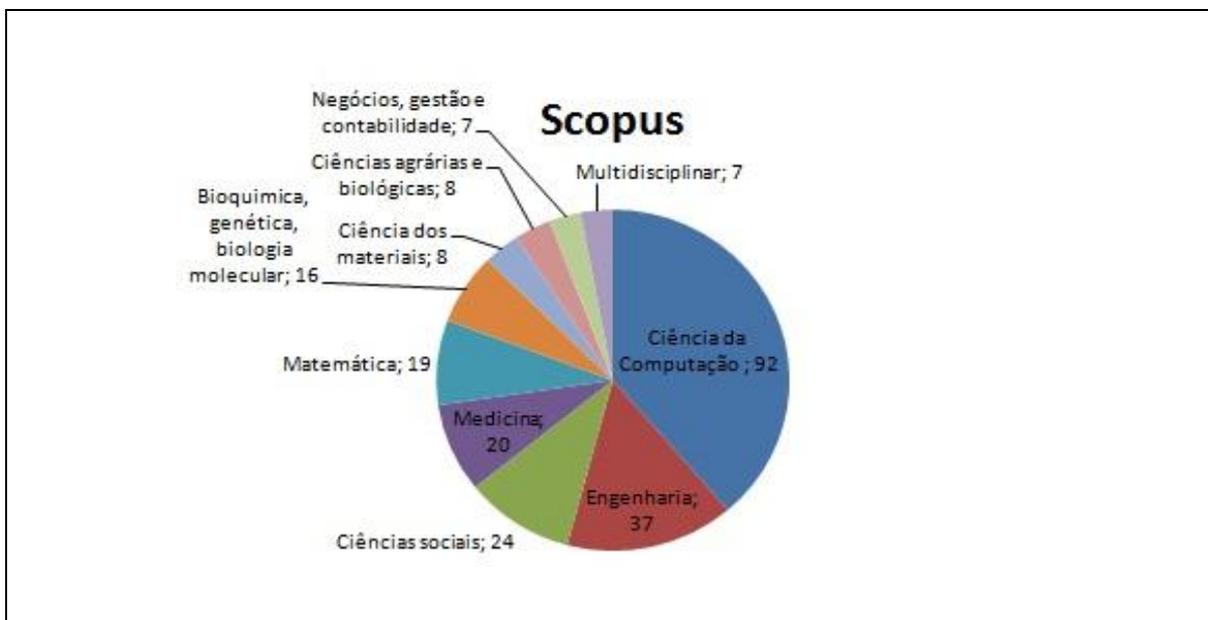
Fonte: elaborado pelo próprio autor.

A Figura 2 apresenta os tipos de publicação que estão sendo desenvolvidas em ambos indexadores. Pode-se perceber um número mais elevado de publicações do tipo conferência. Isto caracteriza que os estudos que estão sendo conduzidos para o tema ainda estão em suas fases iniciais de desenvolvimento.

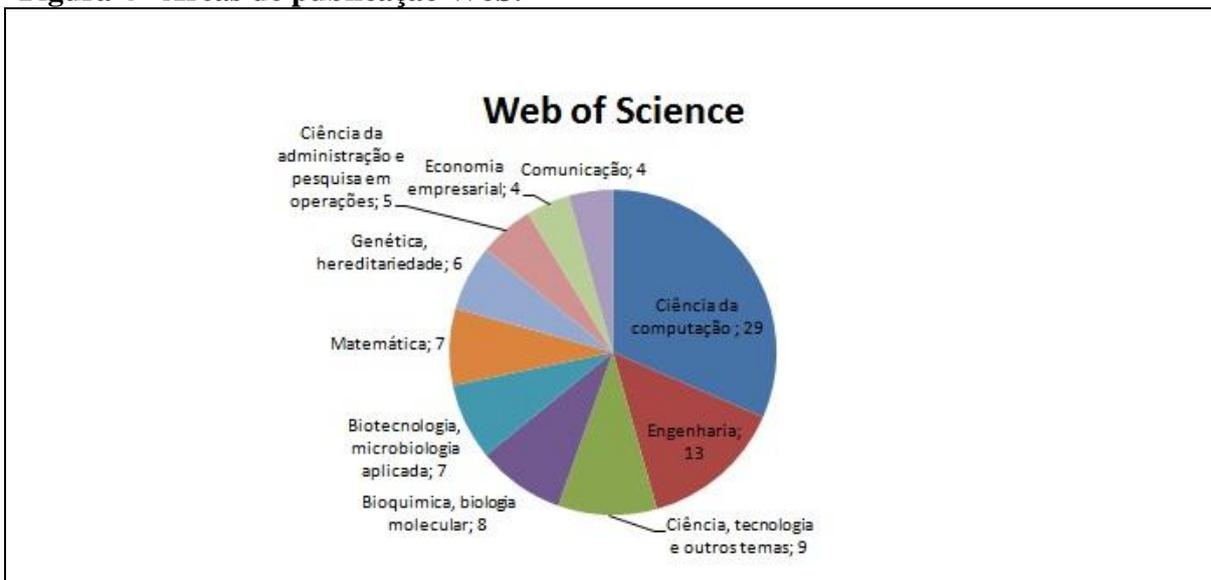
Figura 2 - Tipo de publicação.

Fonte: elaborado pelo próprio autor.

Se relacionados apenas os artigos e livros lançados, com as áreas de publicação de cada um dos indexadores, pode-se perceber uma concentração de mais de 30% na área de ciência da computação, seguido de engenharia. As ciências sociais ocupam posições de menor expressão no contexto analisado, refletindo a carência de pesquisas em *big data* nesta categoria. As Figuras 3 e 4 ilustram as áreas de publicação dos indexadores:

Figura 3 - Áreas de publicação Scopus.

Fonte: elaborado pelo próprio autor.

Figura 4 - Áreas de publicação WoS.

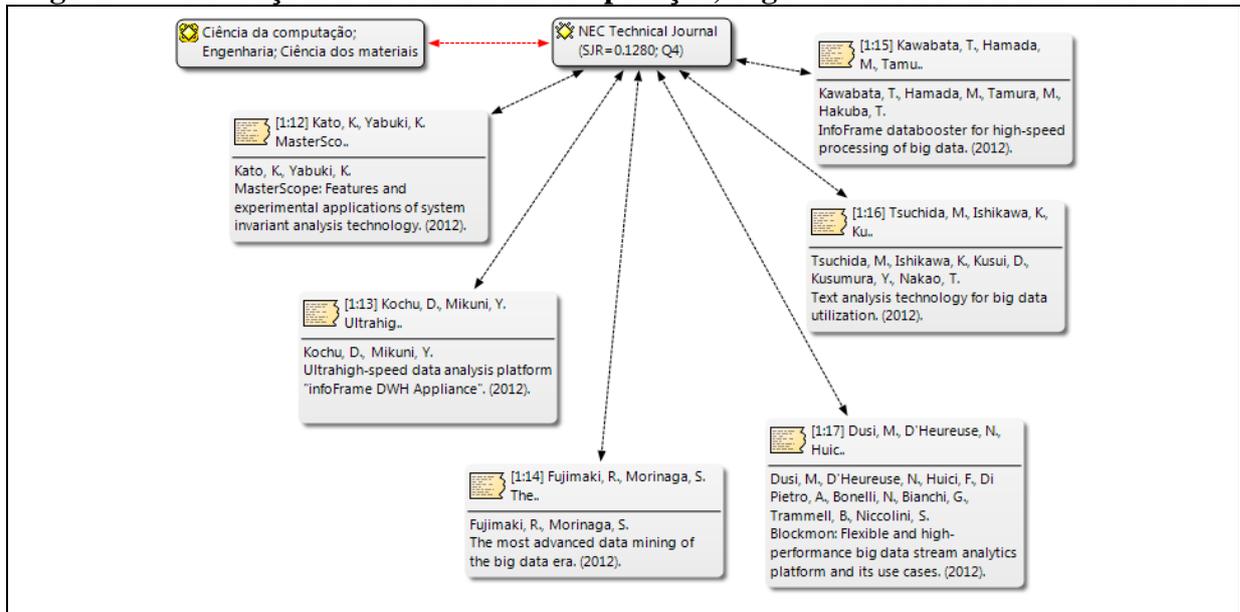
Fonte: elaborado pelo próprio autor.

No aspecto relativo ao impacto das publicações, o presente estudo considerou o índice *Scientific Journal Rankings* (SJR) que é uma medida de impacto, influência ou prestígio da revista. Expressa o número médio de citações ponderadas recebidas no ano escolhido, pelos documentos publicados nos três anos anteriores. O *Journal Citation Reports* (JCR) mede o impacto científico de uma média de artigos publicados em uma revista. O Q1 avalia a revista com relação ao ranking de fator de impacto, dentro de determinada área. Assim, revistas consideradas Q1 são aquelas classificadas no primeiro quartil (25%) em suas áreas de publicação.

Desta forma, com relação ao impacto das fontes analisadas, foram consideradas para análise as fontes com mais de duas publicações, totalizando 9 fontes. A área que mais recebeu destaque foi a de ciência da computação, com 18 publicações. A Figura 5 apresenta as publicações desta área:

A próxima área que se apresenta é a de ciência da computação para engenharia e ciência dos materiais com 6 publicações. A Figura 7 ilustra estas publicações:

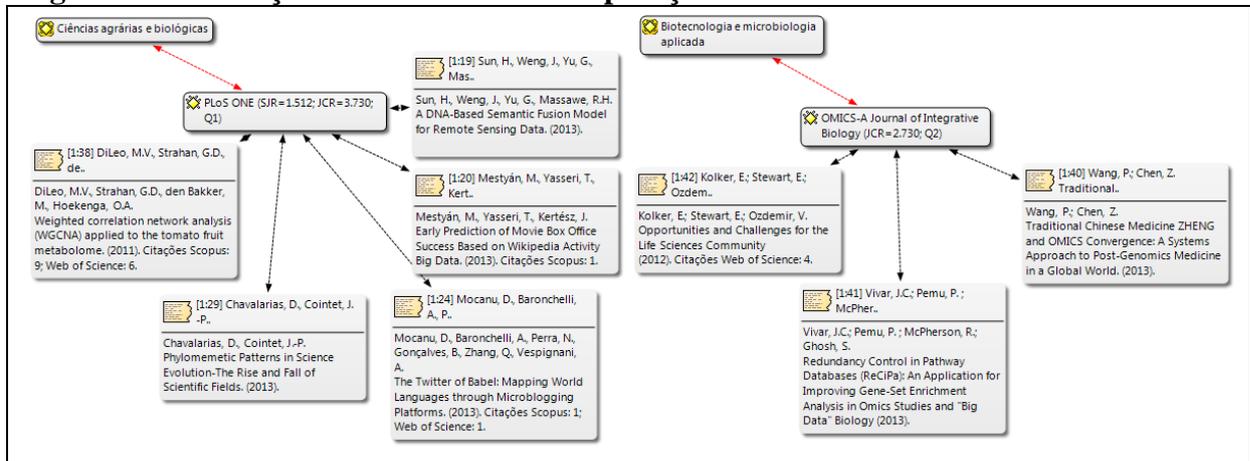
Figura 7 - Publicações área ciência da computação, engenharia e ciência dos materiais.



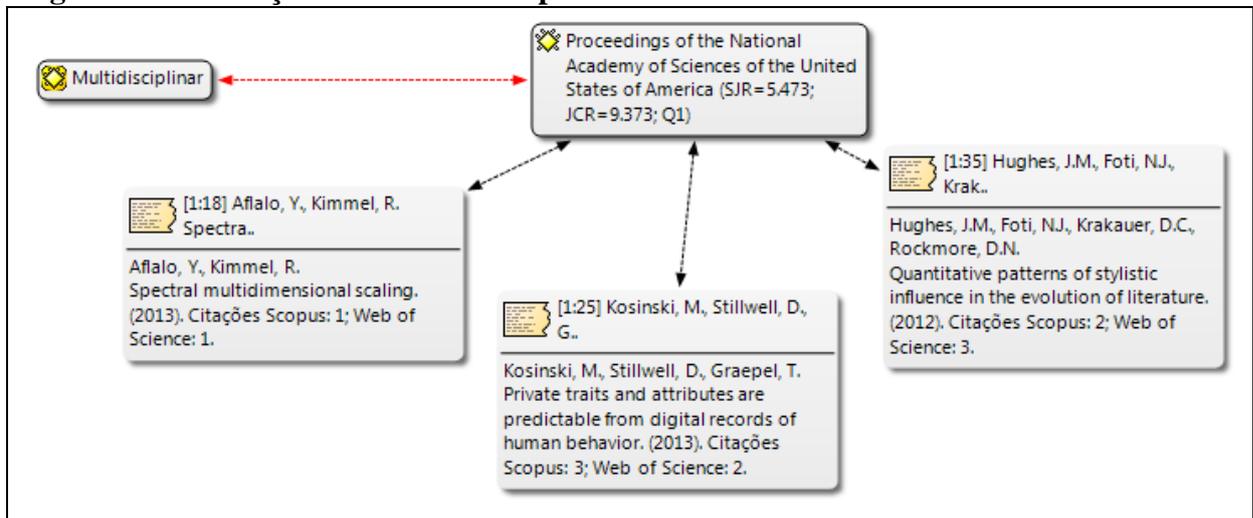
Fonte: elaborado pelo próprio autor.

Por fim, as ciências agrárias e biológicas com 5 publicações, biotecnologia e microbiologia aplicada com 3 publicações e multidisciplinar com 3 publicações. As Figuras 8 e 9, respectivamente ilustram as publicações na área:

Figura 8 - Publicações área ciência da computação.

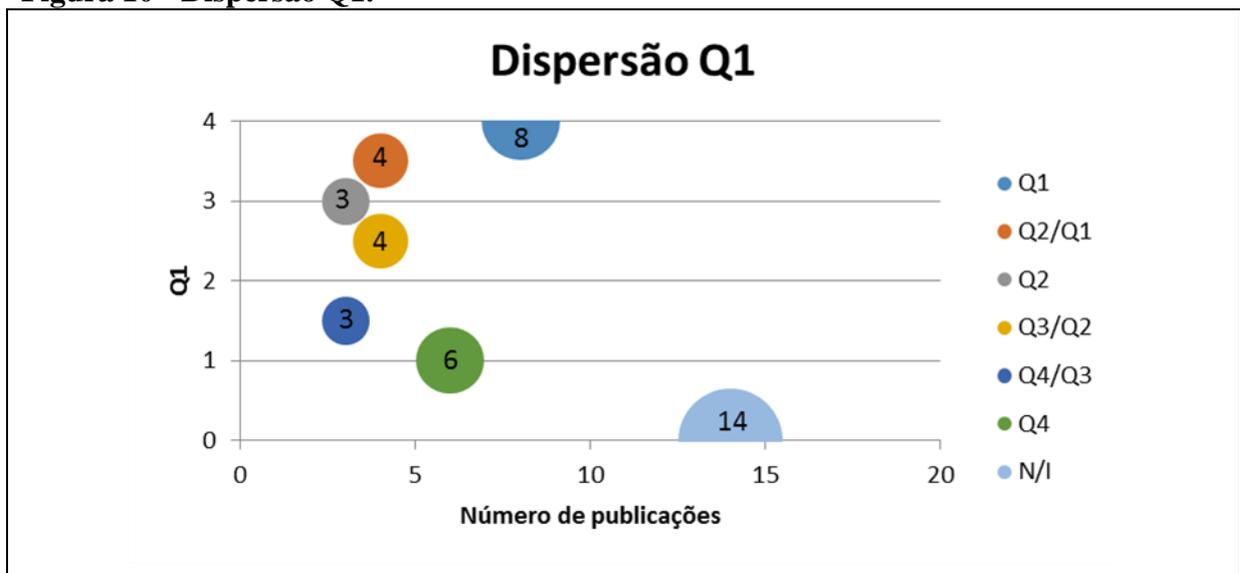


Fonte: elaborado pelo próprio autor.

Figura 9 - Publicações área multidisciplinar.

Fonte: elaborado pelo próprio autor.

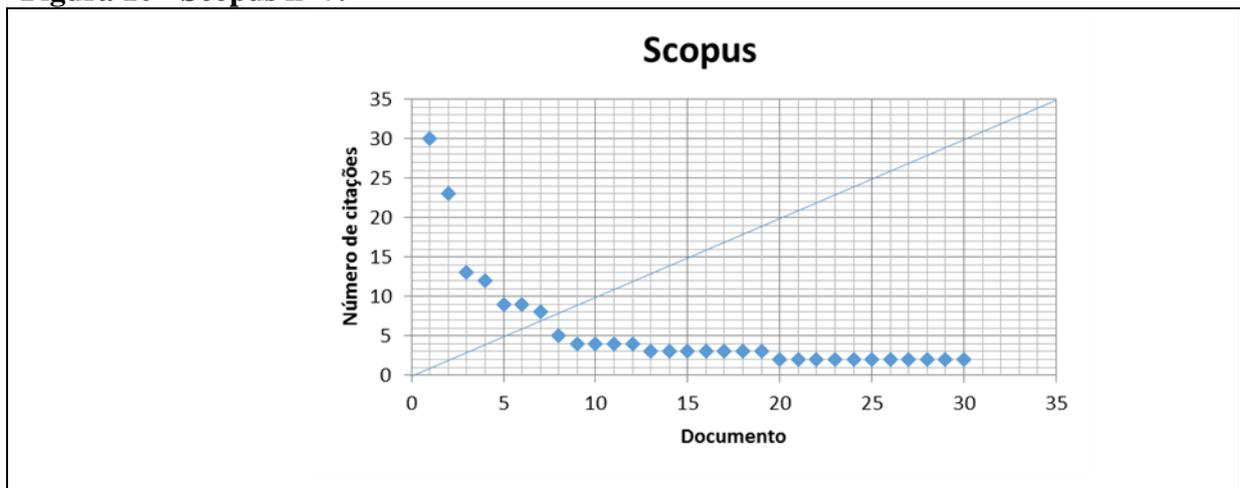
A Figura 10 apresenta um gráfico de dispersão, onde é possível analisar as fontes de publicação de acordo com o índice Q1, que avalia a revista com relação ao ranking de fator de impacto dentro de determinada área. Assim, pode-se observar que 33% das publicações analisadas não apresentam o índice, enquanto as demais se dividem em publicações do tipo Q1, ou seja estão entre as 25% com melhor índice em suas áreas e as demais estão em Q4, classificadas no último quartil.

Figura 10 - Dispersão Q1.

Fonte: elaborado pelo próprio autor.

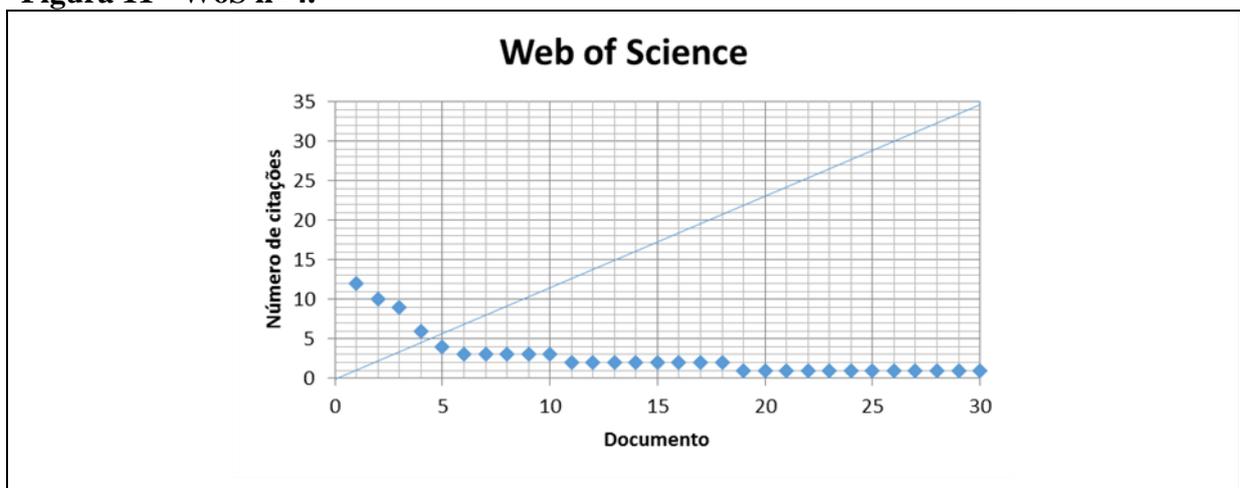
Outra análise realizada foi com relação do índice H, que mede o número de artigos da revista que tenham registro de ao mínimo "h" citações durante o período analisado. Com relação a este índice, o indexador Scopus apresentou um índice $h=7$ e o indexador WoS apresentou um índice $h=4$. As Figuras 10 e 11 representam estes dados:

Figura 10 - Scopus $h=7$.



Fonte: elaborado pelo próprio autor.

Figura 11 - WoS $h=4$.



Fonte: elaborado pelo próprio autor.

Com relação as fontes selecionadas pelo índice H, pode-se observar que estão com um melhor fator de impacto, do que aquelas analisadas apenas com relação ao número de

publicações. De acordo com este índice, mais de 70% das publicações estão enquadradas com relação ao índice Q1, em Q1 ou Q2. A Tabela 1 apresenta as publicações de acordo com o índice H, com h=7:

Tabela 1 - Publicações índice H - Área geral.

| Scopus | | Web of Science | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| Obras | Citações | Obras | Citações |
| Boyd D., Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. (2012) Information Communication and Society (Ciências Sociais: Comunicação-SJR=0.799; Q1) | 30 | Aronova, E.; Baker, K. S.; Oreskes, N. Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957-Present. (2010) Historical Studies in the Natural Sciences (Artes e Humanidades: História e Filosofia da Ciência-SJR= 0.593; JCR=1.087; Q1) | 12 |
| Herodotou H., Babu S. Profiling, what-if analysis, and costbased optimization of mapreduce programs. (2011) Proceedings of the VLDB Endowment (Ciência da computação; Engenharia; Ciência dos materiais) | 23 | Giannotti, F.; Pedreschi, D.; Pentland, A.; et al. A planetary nervous system for social mining and collective awareness. (2012) European Physical Journal-Special Topics (Ciência e Tecnologia-JCR=1.796; Q2) | 10 |
| Bhatarai B., Gramatica P. Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals. (2011) Environmental Science and Technology (Ciências Ambientais: Química Ambiental-SJR= 2.665; Q1) | 13 | Boyd, D.; Crawford, K. Critical questions for big data provocations for a cultural, technological, and scholarly phenomenon. (2012) Information Communication and Society (Ciências Sociais: Comunicação-SJR=0.799; Q1) | 9 |
| Giannotti F., Pedreschi D., Pentland A., Lukowicz P., Kossmann D., Crowley J., Helbing D. A planetary nervous system for social mining and collective awareness. (2012) European Physical Journal: Special Topics. Information Communication and Society (Ciências Sociais: Comunicação-SJR=0.799; Q1) | 12 | DiLeo, M. V.; Strahan, G. D.; den Bakker, M.; et al. Weighted Correlation Network Analysis (WGCNA) Applied to the Tomato Fruit Metabolome. (2011) PLoS ONE (Ciências Agrárias e Biológicas-SJR=1.512; JCR=3.730; Q1) | 6 |
| DiLeo M.V., Strahan G.D., den Bakker M., Hoekenga O. A. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. (2011) PLoS ONE (Ciências Agrárias e Biológicas-SJR=1.512; JCR=3.730; Q1) | 9 | | |
| Fisher D., DeLine R., Czerwinski M., Drucker S. Interactions with big data analytics. (2012) Interactions (Ciência da Computação: Interação Humano-Computador) | 9 | | |
| Chen Y., Alspaugh S., Katz R. Interactive analytical processing in big data systems: A crossindustry study of mapreduce workloads. (2012) Proceedings of the VLDB Endowment (Ciência da computação; Engenharia; Ciência dos | 8 | | |

| | | |
|------------|--|--|
| materiais) | | |
|------------|--|--|

Fonte: elaborado pelo próprio autor.

Na categoria de ciências sociais - áreas de: negócios, gestão e contabilidade; ciência da decisão, economia, econometria e finanças; ciências sociais. - do indexador Scopus, pode-se identificar 31 publicações e índice h=3. No indexador WoS consta índice h=1. A Tabela 2 mostra estas publicações, com o número de citações. A Figura 12 mostra a relação entre as citações.

Tabela 2 - Publicações índice H - Área ciências sociais.

| Scopus | | Web of Science | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| Obras | Citações | Obras | Citações |
| Hamel, M.-P., Marguerit, D. Opportunities for improving public teleservices through "big data" analysis [Quelles possibilités offertes par l'analyse des big data pour améliorer les téléseuices publics?] (2013) Revue Francaise d'Administration Publique (Administração pública; SJR= 0.197; Q3) | 30 | Davenport, T. H.; Barth, P.; Bean, R. How 'Big Data' Is Different. (2012). MIT Sloan Management Review (Negócios, gestão e contabilidade; ciência da decisão-SJR=0,879; Q1/Q2) | 1 |
| Shen, Y., Varvel, V.E. Developing data management services at the Johns Hopkins university. (2013) Journal of Academic Librarianship (Educação; SJR=1.577; Q1) | 3 | | |
| Valentine, E.L.H., Stewart, G. The emerging role of the Board of Directors in enterprise business technology governance (2013) International Journal of Disclosure and Governance (Negócios, gestão e contabilidade-SJR=0.241; Q3) | 3 | | |

Fonte: elaborado pelo próprio autor.

Assim, a pesquisa revelou que mesmo com um crescimento vertiginoso, o tema ainda pode ser considerado emergente conforme já sinalizava Akerkar (2014). Foram identificadas 325 publicações na base Scopus e 113 na WoS. Pode-se observar que o número de publicações nestes temas apresentam um crescimento médio de mais de 200% ao ano. Os tipos de publicação que mais se sobressaem neste total são as do tipo conferência, caracterizando a contemporaneidade do tema. Considerando os artigos e livros lançados, neste total foi possível observar que a maior área de concentração dos focos de publicação é a ciência da computação. As ciências sociais ocupam posições de menor expressão.

Na análise dos artigos com mais de duas publicações foi possível identificar 9 fontes - *Proceedings of the VLDB Endowment, Nec Technical Journal, Plos One, Expert Systems with applications, Jisuanji Xuebao/Chinese Journal of Computers, NTT Technical Review, Proceedings of the National Academy of Sciences of the United States, Scientific Computing and Instrumentation e Omics-a journal of integrative biology*, com total de 42 publicações. Nestas, o destaque é para ciência da computação, novamente, com 18 publicações.

Em termos de avaliação das revistas em relação ao ranking de fator de impacto dentro da área (Q1), foi possível analisar que 33% não possuem tal avaliação. Do número restante, 29% estão no quadrante Q1, contra 21% no quadrante Q4, demonstrando um equilíbrio em termos da importância das fontes no tema. Com relação ao índice H, o indexadores Scopus apresentou $h=7$ e WoS $h=4$, onde o fator de impacto Q1 e Q2 predominou em 70% das publicações. Posteriormente foi realizada uma análise apenas nas ciências sociais. Nesta área de concentração foram identificadas 31 publicações e índice $h=3$, no indexador Scopus e $h=1$ no indexador WoS.

Assim, embora algumas fontes de alto fator de impacto estejam publicando sobre o tema, ainda há muito a ser explorado em termos de *big data* e análise (AKERKAR, 2014). A pesquisa revelou também que nenhum dos artigos publicados em *big data* e análise está relacionado com o tema da IE em seu escopo. No entanto, a análise que é o núcleo para utilização do *big data* (SCHÖNBERGER-MAYER E CUKIER, 2013), compõe a etapa do processo de inteligência em que maior valor é gerado (FLEISHER E BENSOUSSAN, 2003). Também segundo Zhao (2013) *Big data* não transforma informação e conhecimento sem análises detalhadas. Requer soluções de armazenamento grandes e escaláveis, bem como capacidades e aplicações de análise. Ou seja é ainda um campo aberto para a área de IE, pois como afirma Lohr (2012) é uma nova abordagem para a compreensão do mundo e para a tomada de decisões.

5 CONSIDERAÇÕES FINAIS

Ciências como a astronomia e a genômica, que vivenciaram uma explosão informacional nos anos 2000, cunharam o termo *big data* (SCHÖNBERGER-MAYER E CUKIER, 2013). Desde então termo foi sendo utilizado e relacionado a *datawarehouses* ou soluções de *business intelligence* (BI), com *data sets* de *terabytes* de dados. Ou seja, a evolução *big data* no mundo dos negócios pode ser identificada no campo da computação. O presente estudo revelou que na produção científica também é essa a área que mais tem

produzido estudos. No entanto o *big data* representa muito mais que isto (TAURION, 2013) e atualmente, o conceito está migrando para todos os campos do conhecimento humano (SCHÖNBERGER-MAYER E CUKIER, 2013).

No campo da IE, considerando a influência do *big data* no processo de análise, e sendo este um ponto crítico para a geração de valor, o presente estudo examinou como ocorre esta relação na produção acadêmica sobre os temas. O objetivo foi de identificar e analisar o crescimento das publicações sobre *big data* e análise em IE nas bases de dados Scopus e Web of Science. Verificamos que ainda é um campo incipiente em termos de produção acadêmica não apenas para a IE mas também para as Ciências Sociais que é o campo maior no qual se situa a IE. Esse fato pode ser considerado como uma característica própria ao fenômeno já que é considerado emergente, recente e de acelerado crescimento. No entanto, pode servir também de alerta para os estudiosos em IE já que se trata de um conceito que tem a antecipação na sua essência. A análise é o que produz a inteligência de qualquer tipo de monitoramento ou tratamento de dados. Portanto não se pode prescindir de uma compreensão mais aprofundada da influência do *big data* para o campo da IE. Nesse sentido uma perspectiva para estudos futuros que emerge do presente estudo é a de analisar se no campo da atividade profissional de IE a influência do fenômeno *big data* é percebida e de que forma. A experiência dos profissionais de IE de lidar com o fenômeno do *big data* poderia gerar elementos de aprendizado que por sua vez poderiam estimular e influenciar também a produção científica sobre o tema.

BIG DATA: A NEW MANAGEMENT CHALLENGE

ABSTRACT

The information explosion in the 2000s, gave rise to the phenomenon defined as big data. This phenomenon refers to the third season of the information age. It differs from other times because nowadays the analysis can be performed with data in its original unstructured form which allow to analyze not only what happened in the past, but predict what will happen. This is an emerging issue, and in order to analyze its growth in terms of relevance, this paper presents a bibliometric study which aimed to identify and analyze the growth of publications on big data and analysis in strategic intelligence in Scopus data and Web of Science. We carried out an analysis under the totality of articles published throughout the Intelligence Strategic field considering its impact factor and the index H. An additional focus was given to the social sciences. The results indicate that although some sources of high impact factor are publishing on the topic, there is a lack of studies in big data and analysis especially in regard to the social sciences.

Keywords: *Big Data*. Analysis. Bibliometrics. Strategic Intelligence.

REFERÊNCIAS

AKERKAR, Rajendra. **Big data computing**. Boca Raton: Taylor & Francis Group, LLC, 2014.

BETSER, Joseph; BELANGER, David. Architecting the enterprise via big data analytics. In: LIEBOWITZ, Jay. **Big data and business analytics**. Boca Raton: Taylor & Francis Group, LLC, 2013.

COOPER, Harris M.; LINDSAY, James J. Research synthesis and meta-analysis. In: L. Bickman; D. J. Rog; **Handbook of applied social research methods**. p. 315-342. Thousand Oaks, CA: Sage Publications, 1998.

CUKIER, Kenneth. Data, data, everywhere: a special report on managing information. **The Economist**, v. 394, Issue 867, Feb. 2010.

DUMBILL, Edd. What is big data? An introduction to the big data landscape. **O'Reilly Media, Inc.**, 2012. Disponível em: <<http://strata.oreilly.com/2012/01/what-is-big-data.html>>. Acesso em: 5 mar. 2014.

FLEISHER, Craig S.; BENSOUSSAN, Babette. **Business and competitive analysis: effective application of new and classic methods**. New Jersey: Pearson Education, Inc., 2007. LETOUZÉ, Emmanuel. **Big data for development: challenges & opportunities**. UN Global Pulse, maio 2012.

LOHR, Steve. The age of big data. **The New York Times**, 2012. Disponível em: <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all&_r=0>. Acesso em: 5 mar. 2014.

MACIAS-CHAPULA, Cesar. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998.

MINELI, Michael; CHAMBERS, Michele; DHIRAJ Ambiga. **Big data, big analytics: emerging business intelligence and analytic trends for today's businesses**. New Jersey: John Wiley & Sons, Inc., 2013.

RAPPORTEUR, David B. The promise and peril of big data. **The Aspen Institute**, Janeiro 2010.

SATHI, Arvind. **Big data analytics**: disruptive technologies for changing the game. IBM Corporation, 2012.

SCHÖNBERGER-MAYER, Viktor; CUKIER Kenneth. Tradução Paulo Palzonoff Junior. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

TAURION, Cezar. **Big data**. Rio de Janeiro: Brasport, 2013. [ePUB].

ZHAO, Daqing. Frontiers of big data business analytics: patterns and cases in online marketing. In: LIEBOWITZ, Jay. **Big data and business analytics**. Boca Raton: Taylor & Francis Group, LLC, 2013.