



IV SINGEP

Simposio Internacional de Gest3o de Projetos, Inova3o e Sustentabilidade
International Symposium on Project Management, Innovation and Sustainability

ISSN: 2317 - 8302

O BIG DATA E SEU USO CORPORATIVO: UMA REVIS3O DE LITERATURA

M3RCIO SILVEIRA

UNINOVE – Universidade Nove de Julho
silveiramarcio2015@hotmail.com

CARLA BONATO MARCOLIN

Universidade Federal do Rio Grando do Sul
cbmarcolin@gmail.com

HENRIQUE MELLO RODRIGUES DE FREITAS

UNINOVE – Universidade Nove de Julho
freitas138@gmail.com



O BIG DATA E SEU USO CORPORATIVO: UMA REVISÃO DE LITERATURA

Resumo

O Big Data é um advento tecnológico de processamento de grandes volumes de dados que vem ganhando notoriedade por conta de oportunidades e desafios em torno de sua utilidade no apoio aos negócios. Diante disso, este artigo procura, através de uma revisão sistemática da literatura, identificar como estão ligados o Big Data e o mundo corporativo. Para isso, são investigados 444 artigos, em termos de tipo de publicação, evolução anual da produção, principais autores e instituições. Como resultado foi identificado que existe um crescente interesse pelo tema Big Data ligado aos negócios, tanto nas produções ligadas às instituições científicas, quanto àquelas ligadas às empresas. Também foram observadas ligações do tema a uma gama de negócios bastante ampla, do marketing à saúde, ou do transporte público à educação, sempre ligados à uma melhor tomada de decisão.

Palavras-chave: Big Data, Negócios, Tecnologia da Informação

Abstract

The Big Data is a technological advent of processing large volumes of data that has gained notoriety because of opportunities and challenges around their usefulness in supporting the business. Therefore, this article, through a systematic literature review to identify how they are connected to Big Data and the corporate world. For this, 444 papers are investigated in terms of type of publication, annual trends in production, leading authors and institutions. As a result it was identified that there is a growing interest in the topic Big Data connected to business, both in the productions linked to scientific institutions, as those linked to companies. Were also observed theme leads to a very wide range of businesses, health marketing, or public transportation to education, always linked to better decision making.

Keywords: Big Data, Business, Information Technology



1 INTRODUÇÃO

Na última década, o termo Big Data passou a fazer parte dos repositórios científicos. Neste curto espaço de tempo vem ganhando notoriedade na mídia e em algumas empresas, causando, ora fascínio (Huwe, 2012), ora desconfiança (Davenport, 2014) em pessoas ligadas à gestão da informação, assim como em gestores corporativos. Novas e sofisticadas ferramentas de processamento de dados surgem a cada dia, melhorando o desempenho em diversos aspectos, especialmente em volume e em tempo de processamento (Goldman, Kon, Junior, Polato, & de Fátima Pereira, 2012).

A proposta de utilização de grande quantidade de informações não é novidade nem na ciência da informação, nem na administração e remonta do início da maior utilização de sistemas de informação pelas empresas, a partir dos anos de 1950 (H. Chen, Chiang, & Storey, 2012). Um fenômeno de destaque é o crescente aumento de fontes de informações exógenas às empresas, captadas por redes sociais, sensores em diversos produtos ou registros de tráfego de internet, entre outros, caracterizado muitas vezes pela expressão Big Data.

Conforme Demchenko, Grosso, De Laat, & Membrey (2013), a quantidade de dados cresce a cada dia, assim como a possibilidade para sua utilização. Neste sentido, um dos maiores desafios está na análise destes dados e na compreensão de como eles podem melhorar os negócios. As principais aplicações focam em dados captados internamente que sejam insuficientes ou que possam ser completados pelas informações externas provenientes do Big Data. A promessa é que esta análise de dados pode ser usada para melhor compreender clientes, mercados, concorrentes, produtos, ambiente de mercado, impacto de tecnologias ou até mesmo fornecedores (Marchand & Peppard, 2013; Mayer-Schonberger & Cukier, 2013).

Diante de múltiplas possibilidades de utilização do Big Data citadas, parece que a emergência deste advento de dados tem ligações com o mundo dos negócios. Mas, como vem sendo abordada a ligação entre o Big Data e os negócios nas produções ligadas à comunidade científica e às empresas? Em busca desta resposta, este estudo procura sintetizar os principais conceitos sobre Big Data, assim como suas relações com o mundo corporativo em termos de progressão das publicações, vínculo institucional com empresas privadas ou com instituições de ensino e pesquisa, países em que se concentram as produções, assim como os principais autores. Com um melhor entendimento científico desta questão, espera-se contribuir para o alinhamento acerca do Big Data que, muitas vezes, se apresenta fluido e incógnito em relação à sua validade de utilização pelas empresas.

Para isso, na seção 2 é apresentado um panorama sobre o uso de sistemas de informações nas empresas, que antecederam o advento do Big Data, seguido por sua conceituação e características principais. Na seção 3 estão descritos os métodos de trabalho deste artigo para que os objetivos expostos fossem atingidos. Na seção 4 são apresentados os resultados obtidos e, na seção 5, foram relatadas as conclusões do estudo, em conjunto com propostas para estudos futuros.

2 BIG DATA: HISTÓRICO, CONCEITOS E CARACTERÍSTICAS

Nesta seção é descrito um breve histórico do uso de sistemas de informação nas empresas, dos primórdios à sua evolução. Nos últimos anos, com o surgimento de novas fontes e ferramentas surgiu o termo Big Data, que designa uma série de características que foram condensadas através da visão de alguns autores, para a apresentação neste artigo.

2.1 Histórico da utilização de Sistemas de Informação nas empresas

Alguns autores como Chen, Chiang & Storey (2012) e Davenport (2014) distinguem três períodos do uso de sistemas de informação para a análise de dados como apoio aos negócios. Estes períodos estão ligados ao surgimento de inovações tecnológicas que alteraram



as fontes de dados ou as formas de seu processamento. Na Figura 1 estão descritas as nomenclaturas destes períodos, segundo três autores.

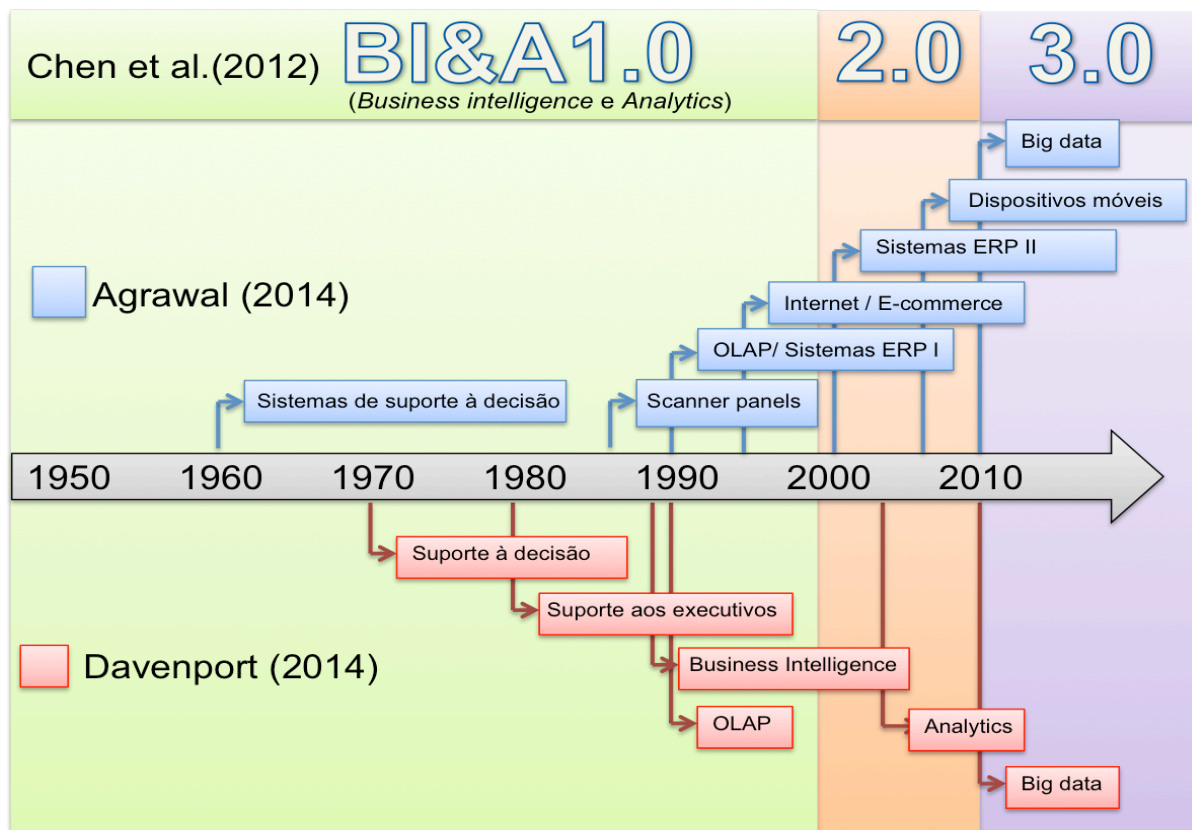


Figura 1: Linha do tempo da evolução dos sistemas de informação nas empresas

Fonte: Adaptado de Chen et al.(2012), Agrawal (2014) e Davenport (2014).

O uso de ferramentas de tecnologia de informação pelas empresas é relatado a partir do final da última metade do século passado. Chen et al.(2012b), diz que a inteligência artificial para uso corporativo já era tema de estudos desde o início dos anos de 1950. Nesta época, os dados passam a ser recolhidos através de sistemas legados em formato estruturado de dados relacionais. Inicialmente, a tecnologia desenvolvida para ajudar a tomada de decisões nos negócios se resumia a processadores que tornaram eletrônicos os processos que antes eram mecânicos (Agrawal, 2014).

Nos anos 90, outras aplicações como o ERP (*Enterprise Resource Planning*) ganharam força na estruturação dos negócios, como maneira de organizar as informações da empresa. Esta aplicação contribuiu para a estruturação do modelo do negócio. (Agrawal, 2014) e a partir deste momento foram aprimoradas as ferramentas de consultas aos bancos de dados, como o processamento analítico online (OLAP) e o *Business Performance Management* (BPM), quando se popularizou o termo *Business Intelligence* (BI), com diversas aplicações de suporte a tomada de decisão nas empresas. Esta fase, que se estende até meados do ano 2000, foi denominada de “*Business Intelligence and Analytics 1.0*” (BI&A 1.0) (Chen et al., 2012b).

Na chegada aos anos 2000, Davenport (2012) e Chen et al. (2012) citam o advento da internet como fonte de mudança de paradigma. A partir deste, uma imensa quantidade de dados passou a ser gerada instantaneamente. Padrões de navegação, fluxo de cliques, compras realizadas, além de conteúdo gerado pelo próprio usuário em mídias sociais, blogs e outros são alguns dos exemplos de atividades que passaram a canalizar um fluxo crescente de dados.



Esta capacidade de produção de dados proporcionada pelas atividades conectadas à Internet foram identificadas por Chen *et al.* (2012b) e denominadas como a versão 2.0 do uso do *Business Intelligence and Analytics* (BI&A 2.0). Agrawal (2014) e Davenport (2012) também abordam estes advenços como alavancas para a produção de dados que podem ser analisados e aplicados aos negócios.

Um novo salto na produção de dados vem se consolidando através da intensa massificação de dispositivos móveis, aliado a uma gama crescente de aplicativos que potencializaram os efeitos do BI&A 2.0, justificando sua reformulação para BI&A 3.0 (Chen *et al.* 2012). As transformações advindas da ubiquidade dos aparelhos móveis, da maciça utilização da computação em nuvem, assim como da conexão de diversos equipamentos cotidianos (“internet das coisas”) (Chen, Mao, & Liu, 2014), canalizam a produção de dados para um fluxo contínuo.

2.2 Caracterização e definições

A nomenclatura Big Data é um conceito abstrato que surgiu em meados de 2010 para designar a tendência tecnológica de gerar grandes quantidades de dados, de diferentes origens e formatos (Chen *et al.*, 2014). Autores como Davenport e Chow-White *et al.*(2013) retratam o uso do Big Data como ferramenta emergente para a tomada de decisão corporativa. De fato, o grande volume de dados é a principal característica já estampada na nomenclatura “big”. De acordo com Davenport (2014), somente no ano de 2012, foram gerados no mundo cerca de 2,8 trilhões de *gigabytes* em dados. Esta medida em *bytes*, demonstra a escalada da geração transmissão de dados pelas redes informatizadas, já que 2,8 *exabytes* correspondem a uma quantidade mais de um bilhão de vezes maior daquilo que era a referência na transação de dados de uma década atrás (McAfee & Brynjolfsson, 2012). Além do volume de dados, outras características também fazem parte da literatura do Big Data: variedade e a velocidade (Manyika *et al.*, 2011; McAfee & Brynjolfsson, 2012).

A variedade de fontes de dados está relacionada às diversas possibilidades de equipamentos ou aplicações envolvidas na geração e captura de dados, provenientes, especialmente, da *Web 2.0*, além dos sistemas convencionais. Muitas vezes, estes dados são captados em formato não estruturado, ou seja, não estão em tabelas de banco de dados convencionais, como imagens, sinais de geo-localização, sons, entre outras possibilidades. Este é um grande desafio quando se trata do Big Data (Davenport, 2014).

Já a velocidade, em muitos casos, é uma característica importante, uma vez que transformar os dados em informações para a tomada de decisões em tempo real é um desafio e uma fonte de diferencial competitivo. Como estes dados, muitas vezes, são provenientes de fontes contínuas, sua análise deve ocorrer em tempo hábil para proporcionar decisões contemporizadas e válidas (McAfee & Brynjolfsson, 2012).

Demchenko *et al.* (2013) agregam outros dois desafios para o trabalho com Big Data: a veracidade e o valor. Para o autor, a veracidade trata da necessidade de que os dados coletados tenham origem comprovada, ou seja, dados confiáveis. O valor é trabalhar com dados que de fato sejam fontes de valor agregado para a tomada de decisão. As principais características citadas por autores estão descritas no Quadro 1.

Características	Autores	Descrição
Volume	(Demchenko <i>et al.</i> , 2013)	Relacionado ao tamanho e quantidade de dados
	(Davenport <i>et al.</i> , 2012)	Centenas de Terabytes ou Petabytes
	(Agrawal, 2014)	Grande quantidade e complexidade de dados
	(McAfee & Brynjolfsson, 2012)	Grande quantidade de dados criados por dia
Velocidade	(Goldman <i>et al.</i> , 2012)	Necessidade de respostas em tempo real
	(Demchenko <i>et al.</i> , 2013)	Dinâmica de crescimento e processamento de dados
	(McAfee & Brynjolfsson, 2012)	Dados capturados e processados quase em <i>real time</i>



	(Zikopoulos et al., 2013)	Velocidade de captura e análise de dados
Variedade	(Demchenko et al., 2013)	Diversidade de origens, formas e formatos de dados
	(McAfee & Brynjolfsson, 2012)	Grande variedade de fontes e formas de dados com o desafio de encontrar os padrões de dados úteis para os negócios
Veracidade	(Demchenko et al., 2013)	Autenticidade, reputação de origem e confiabilidade dos dados
Valor	(Demchenko et al., 2013)	Dados com significado para os negócios, que contribuam com valor agregado

Quadro 1: Características do Big Data

Fonte: Adaptado de Demchenko et al. (2013), Davenport et al. (2012), Agrawal (2014), McAfee & Brynjolfsson (2012), Goldman et al. (2012), Zikopoulos et al. (2013).

3 MÉTODO

O uso de revisões sistemáticas é uma importante ferramenta de análise. Este tipo de pesquisa difere das tradicionais revisões por adotar um processo científico e replicável com o objetivo de minimizar o viés de pesquisas não estruturadas. Como resultado, evidencia-se o caminho científico para o tema em termos de teorias, procedimentos e conclusões (Tranfield, Denyer, & Smart, 2003). Uma revisão sistemática bem conduzida serve de base para o avanço do conhecimento, delimitando as áreas de pesquisa já abordadas, além de revelar áreas em que cabe maior investigação (Webster & Watson, 2002).

Neste trabalho foi realizada uma revisão sistemática da literatura que permite ao pesquisador melhor compreensão sobre um tema específico servindo ao objetivo de mapear trabalhos publicados no tema de pesquisa, para que o pesquisador possa compreender o “estado da arte” do assunto e direcionar estudos futuros (Bereton et al. 2007). Neste sentido, este método atende a esta proposta.

Foi seguido o processo descrito por Trafield et al. (2003), iniciando pelo planejamento no qual o pesquisador pode considerar uma breve descrição histórica e metodológica acerca da área de estudo e suas conexões, além de determinar a questão de pesquisa, que serve de base para as próximas etapas. As etapas seguintes são condução e comunicação da pesquisa, respectivamente. Na condução da pesquisa, são identificados as palavras-chave e termos de busca tendo como resultado uma lista completa de trabalhos científicos que atendam aos requisitos desejados. Em seguida, é realizada avaliação mediante leitura, para que o pesquisador possa decidir por incluir outras buscas ou excluir estudos que estejam fora do escopo. Após, são elaboradas as sínteses que respondem as questões de pesquisa do ponto de vista de cada autor para apresentação ou confronto. Por fim, a comunicação é a etapa em que o pesquisador categoriza os resultados pelos critérios necessários para a compreensão do objetivo da pesquisa (Tranfield et al., 2003).

Para tal, foi definido uma questão norteadora para busca e seleção dos trabalhos. Dado que o objetivo deste estudo foi sistematizar as buscas por produção científica referente ao uso do Big Data nas organizações, a questão para a qual se busca respostas é: como vem se desenvolvendo os estudos de Big Data nos negócios?

A busca inicial foi feita pelo indexador Scopus, uma base de dados bibliográfica de trabalhos científicos. A escolha pela utilização do Scopus foi dada por se tratar de uma das maiores bases de dados de trabalhos científicos multidisciplinares (Björk et al., 2010). Os dados foram extraídos em junho de 2015. Foi utilizado o termo “big data” como critério de seleção constante no título, palavras-chave ou resumo. Outras delimitações da pesquisa foram para o idioma inglês, assim como para a área temática de “negócios, gestão e contabilidade”. Em relação ao aspecto temporal, foram delimitadas as datas entre os anos de 2010 e 2014, concluindo 5 anos para as análises. Os trabalhos do ano de 2015 foram rejeitados em razão de



ajuste da comparação das produções em anos completos. Resultaram 439 estudos nesta amostra inicial. Este montante foi trabalhado em planilha Excel para que pudessem ser gerados gráficos e quadros comparativos para as análises propostas. Também foram geradas nuvens de palavras e expressões que permitissem verificar visualmente os termos emergentes. Além disso, foram acrescentados 5 artigos de autores brasileiros que têm artigos publicados nos anais de todas as edições do CONTECSI (Conferência Internacional de Sistemas de Informação e Gerenciamento de Tecnologia), com o objetivo de incluir pontos de vista de pesquisadores brasileiros acerca do Big Data e as empresas. Por fim, são apresentadas algumas oportunidades e desafios identificados em relação ao Big Data e a área de negócios.

4 RESULTADOS

O intuito deste estudo é de verificar na literatura científica a forma com que vem sendo abordada a ligação entre o Big Data e os negócios. O uso de informações para subsidiar a tomada de decisões nas empresas têm, por si só, estreita ligação (Becker, Freitas, Hoppen, & Kladis, 1997) na qual o Big Data insere novas características de captação e processamento destes dados. Os resultados desta revisão de literatura apontam a evolução dos estudos, tipos de publicações, autores e instituições, que clareiam o direcionamento científico do tema.

A partir das palavras-chave de todos os 439 artigos, foi produzida uma nuvem de palavras ou expressões. O resultado demonstra as áreas de negócios às quais o tema Big Data está relacionado nos artigos, conforme Figura 2. A nuvem, com tamanho das letras proporcionais ao número de vezes que apareceram nas palavras-chave, revela conexões com “negócios”, “gestão da informação”, “tomada de decisão”, entre outras.



Figura 2: Nuvem de palavras das palavras-chave dos artigos selecionados

Fonte: Elaborado pelos Autores

Como efeito da busca de artigos realizada, a nuvem de palavras apresenta uma variedade de conceitos relacionados ao Big Data. Estas relações envolvem bases tecnológicas de sistemas de informação, assim como áreas de negócios onde esta tecnologia se insere. Trata-se, portanto, de dois grupos de assuntos: Big Data e negócios, abordados nos artigos selecionados, evidenciando o método utilizado e já direcionando para os subtemas emergentes na atualidade. Os principais subtemas envolvem suas fontes conectadas às mídias e redes sociais; sua busca e processamento através de *data mining* e *data warehouse*; nas premissas de qualidade dos dados, apontada por diversos autores como Davenport (2014a); ou nos desafios, como os ligados às pessoas e seus conhecimentos. Por fim, podemos observar a ligação com atividades organizacionais. Big Data e negócios se interligam neste estudo através da gestão da informação para a tomada de decisão, assim como já tratado por Becker et al. (1997), com a inclusão de novas abordagens de gerenciamento de dados, como a



mineração de dados, ou com suas ferramentas propriamente ditas: Hadoop, Analytics, entre outros. Davenport (2014) diz que há um crescente interesse por parte de profissionais corporativos e acadêmicos no tema Big Data. A Figura 3 mostra, ano a ano, um aumento de produções: o número de artigos em 2014 foi 101,5% superior às produções do ano anterior.

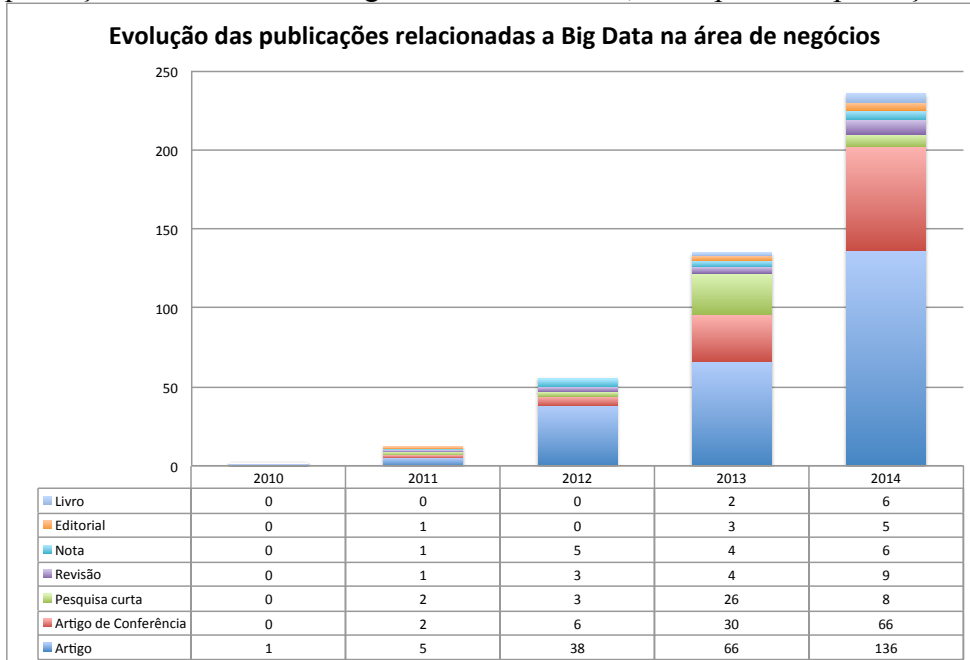


Figura 3: Evolução anual das publicações relacionadas a Big Data na área de negócios

Fonte: Elaborado pelos Autores

No ano de 2010 apenas um artigo foi publicado relacionando Big Data na área de negócios. Nos anos seguintes podemos observar crescimento, com destaque para a publicação de artigos. A única queda significativa evidenciada foi na produção de pesquisas curtas entre os anos de 2013 e 2014 e está relacionada à queda de publicação da *IBM Data Management*.

Também foram distinguidas as produções científicas oriundas de instituições de ensino e pesquisa, das produções de empresas, dentre as quais há publicações de cunho comercial. Na Figura 4 é demonstrada esta distinção, com 47,4% das produções advindas de instituições de ensino e pesquisa, 35,3% de empresas, enquanto que 17,3% das produções são independentes de instituições que não estão identificadas nos repositórios das publicações.

Vínculo Institucional e tipo das publicações

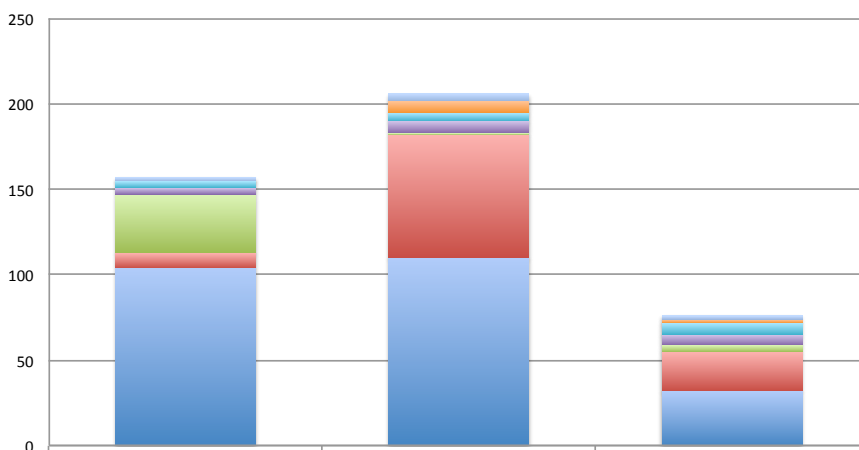


Figura 4: Vínculo institucional e tipo das publicações

Fonte: Elaborado pelos Autores



Dentre as empresas com produções publicadas no âmbito desta pesquisa, predomina a IBM com o maior volume: 73,9% das publicações, sendo 80 artigos e 34 pesquisas curtas. Outras empresas participam com quantidades de produção reduzida em relação a IBM. Uma possível explicação para este fato é que entre as cinco empresas com maior volume de produção, apenas duas, a IBM e a ComScore são fornecedoras de ferramentas de processamento de dados nos parâmetros de Big Data, sendo a IBM um dos principais *players* deste mercado. Os objetivos comerciais podem fundamentar esta concentração.

Já em relação às instituições de ensino e pesquisa que publicam estudos relacionados ao Big Data nos negócios, não há uma polarização tal qual evidenciado entre as empresas. Prova disso pode ser verificada no Quadro 2, em que a instituição com maior número de publicações, o MIT, participa com 3,4% do total de publicações. A lista é seguida por instituições como a Escola de Negócios da Harvard e a Universidade Britânica de Southampton.

Um único estudo oriundo de uma instituição brasileira figura nesta relação. Trata-se de um artigo escrito por Inácio e Dantas, da Universidade Federal de Santa Catarina, abordando a eficiência energética como meio para a evolução da computação de alto desempenho a serviço dos negócios. Na Quadro 3, este artigo está em destaque.

	Artigo	Livro	Artigo de Conferência	Editorial	Nota	Revisão	Pesquisa curta	Total
Empresa	104	2	9	0	4	4	34	157
IBM	80	0	2	0	0	1	33	116
McKinsey	6	0	0	0	1	0	0	7
Fortune	1	0	0	0	1	0	0	2
ComScore	2	0	0	0	0	0	0	2
Korea Technology	2	0	0	0	0	0	0	2
Outros	13	2	7	0	2	3	1	28
Instituição de ensino e pesquisa	110	4	72	7	5	7	1	206
MIT Sloan	5	0	0	1	1	0	0	7
Harvard Business School	4	0	0	0	2	0	0	6
University of Southampton	6	0	0	0	0	0	0	6
Beihang University	1	0	1	0	0	0	0	2
Brunel University	0	0	2	0	0	0	0	2
Indiana University	0	0	1	1	0	0	0	2
Northwestern University	1	0	0	0	0	0	1	2
Shanghai University	1	0	1	0	0	0	0	2
Singapore Management University	1	0	0	0	0	1	0	2
University of Calabria	0	0	2	0	0	0	0	2
University of California	2	0	0	0	0	0	0	2
University of Maryland	1	0	1	0	0	0	0	2
University of Oxford	1	0	0	0	0	1	0	2
University of Saskatchewan	2	0	0	0	0	0	0	2
Weber State University	2	0	0	0	0	0	0	2
Zhejiang University	0	0	1	1	0	0	0	2
London Business School	1	0	0	0	1	0	0	2
Queensland University of Technology	2	0	0	0	0	0	0	2
Technische Universität	0	0	2	0	0	0	0	2
Universidade Federal de Santa Catarina	0	0	0	0	0	1	0	1
Outras Instituições	80	4	61	4	1	4	0	154
Não identificado	32	2	23	2	7	6	4	76
Total	246	8	104	9	16	17	39	439

Quadro 2: Instituição, Vínculo institucional e tipo das publicações

Fonte: Elaborado pelos Autores

Em relação aos países de origem no desenvolvimento dos estudos, os Estados Unidos preponderam em número, com 50,1% das publicações, conforme a Figura 5. Entre as produções de empresas americanas, o percentual é ainda maior, representando 89,8% das produções, número também influenciado pelo grande número de publicações geradas pela IBM. Por sua vez, as instituições de ensino e pesquisa deste país contam com 34,9% de trabalhos relacionando Big Data a negócios. Os países que seguem nesta lista são o Reino Unido, China, Índia e Alemanha, além de outros 27 países em todos os continentes. Uma parcela dos estudos, correspondente a 17,5% do total, não consta o país da publicação, impedindo a verificação da existência de outros polos de produção de estudos na área.

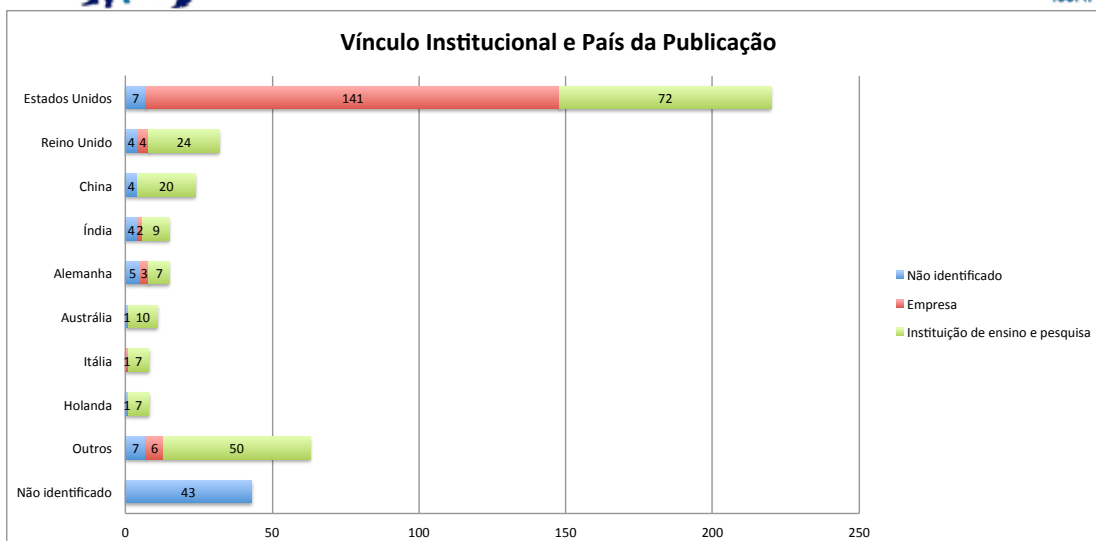


Figura 5: Vínculo institucional e país da publicação

Fonte: Elaborado pelos Autores

Em relação aos autores, aqueles com maior número de publicações são os que aparecem em destaque na nuvem de palavras expressa na Figura 6. Esta nuvem foi formada por todos os autores das 439 publicações pesquisadas com dois ou mais estudos publicados. A mesma característica de concentração de produções em relação a uma única empresa, enquanto que se mantém pulverizada quando se trata de instituições de ensino e pesquisa também foi verificada quando são analisados os autores. A maior quantidade de autores em destaque são afiliados à IBM, como Deutsch T., Kobielus J., Soares S., e Sigmon P.N. Os dois autores com maior número de produções dentre as instituições de ensino foram James R., da Universidade de Southampton e Fawcett S.E., da Universidade Weber State, com 6 e 4 publicações, respectivamente.



Figura 6: Nuvem de palavras das palavras-chave dos principais autores

Fonte: Elaborado pelos Autores

4.1 Produções brasileiras

Com o intuito de identificar alguma produção científica brasileira, foram realizadas buscas aos anais do CONTECSI (Conferência Internacional em Sistemas de Informação e Gerenciamento de Tecnologias), utilizando como único critério de busca o termo “Big Data”. O resultado obtido nesta busca foi de um número pequeno de produções: um artigo no ano de 2013 e outros 4 artigos no ano de 2014, totalizando 5 artigos disponíveis nos anais desta



Conferência. Esta diminuta participação nacional nas produções relacionadas a Big Data foram constatadas anteriormente por Park e Leydesdorff.

No Quadro 3, estão expostos os títulos, autores e uma breve descrição do método e objetivo das pesquisas dos estudos publicados no CONTECSI. Nota-se que são abordagens preliminares, descritas por Luvisan et al. (2014) como evidências de que os estudos científicos sobre Big Data no Brasil ainda não conquistaram o mesmo grau de adesão, tal qual a internacional.

Título	Autor(es)	Abordagem
Big Data: trazendo novas oportunidades e desafios	Breternitz & da Silva (2013)	Ensaio exploratório com o objetivo de abordar os conceitos, oportunidades e desafios do Big Data.
Big Data: evolução das publicações e oportunidades de pesquisa	Luvizan, Meirelles & Diniz (2014)	Estudo bibliométrico com o intuito de identificar o interesse acadêmico e não-acadêmico em torno do Big Data.
Novas perspectivas utilizando Big Data: um estudo bibliométrico	Silva & de Campos, (2014)	Bibliometria acerca do termo Big Data.
O uso do big data no sistema de informação de marketing	Salvador & Ikeda, (2014)	Revisão de literatura acerca do uso do big data no contexto do sistema de informações de marketing e sua utilidade nas decisões.
Critérios para Seleção de SGBD NoSQL: uma Aplicação da Técnica Delphi	Souza, Prado, Peres, & Sun, (2014)	Revisão de literatura com objetivo de identificar os principais critérios para a seleção de dados não estruturados em organizações privadas.

Quadro 3: Títulos, autores e abordagens das produções brasileiras do Contecsi

Fonte: Elaborado pelos Autores

Nos estudos brasileiros do Quadro 4, o Big Data é apresentado como área de interesse acadêmico e empresarial. Desafios e oportunidades são apresentados sem um contexto específico, exceto no estudo de Salvador e Ikeda (2014), que tratam da tomada de decisão na área de marketing. Não se revelam, nestes casos, peculiaridades que se apresentam no Brasil, já que as abordagens são pouco contextualizadas e de caráter exploratório.

4.2 Oportunidades corporativas relacionadas ao Big Data

Diversos autores apresentam um quadro bastante amplo de possibilidade uso do Big Data nas organizações. Davenport (2014), relaciona empresas de segmentos que, costumeiramente, utilizam-se dos dados disponíveis e tendem a avançar na vanguarda do Big Data. São empresas on-line, de bens de consumo, seguradoras, do segmento de turismo e transporte entre outras. Outros segmentos são descritos pelo autor como tendo alto potencial de uso do Big Data, como companhias bancárias, telecomunicações, mídia e entretenimento, varejo e empresas de utilidade pública. Existem ainda outros segmentos, de saúde, de empresas B2B (*business-to-business*) ou de produtos industriais, que por diversos motivos tinham menor acesso aos dados e agora podem fazer uso daqueles provenientes dos novos meios de produção e análise de informações.

Considerando o processo de tomada de decisão, há obrigação de se obter dados de maneira *on-line*, a fim de prestar apoio às decisões em tempo hábil (Tien, 2013). Neste sentido, o autor apresenta o Big Data como a "terceira revolução industrial". Huang e Van Mieghem (2014) apresentam um modelo que usa fluxo de cliques para o direcionamento da composição de estoque e redução de seus custos, uma vez que permite melhores decisões no momento da compra. Tien (2013) complementa com aplicações de Big Data na eficiência da cadeia de suprimentos, como no inventário ideal, na redução do tempo de espera e,



especialmente, na flexibilidade da cadeia de suprimentos, que permite às empresas responder melhor aos padrões de previsão de demanda.

Personalização, decisão em tempo real, estrutura de serviço, e o *service-oriented thinking* são paradigmas em crescimento. Trabalhar com estrutura de serviço e Big Data pode produzir ganho de escala na análise de dados, que é fundamental devido à quantidade de dados disponíveis. Além disso, prover agregação de contexto específico (como cuidados de saúde ou de combate ao crime, por exemplo) tem o potencial de tornar os profissionais e as decisões mais eficientes e eficazes (Demirkan & Delen, 2013).

O Quadro 4 retoma os principais aspectos sobre Big Data e suas oportunidades.

Principais oportunidades	Autores
Adquirir conhecimento de uma forma eficaz em termos de custos	Segupta (2013); Rust & Hang (2014); Mielle <i>et al</i> (2013); Klingström <i>et al</i> (2013); Tien (2012)
Otimizar o atendimento através de um quadro mais completo de informações sobre os clientes do serviço	Bartels, Barbeito & Mackensen (2011); Simpao <i>et al</i> (2014)
Permitir diagnósticos através de dados	Simpao <i>et al</i> (2014)
Melhorar o tempo de decisão	Huang & Van Mighem (2014)

Quadro 4: Principais oportunidades com o Big Data

Fonte: Elaborado pelos Autores

De múltiplas perspectivas, o Big Data tem potencial para atender diversas demandas, de melhoria na colaboração a decisões eficientes em termos de custo. No entanto, alguns aspectos sobre como torná-lo viável para a organização ainda permanecem obscuros, e o aprofundamento nas pesquisas podem contribuir com estas questões, trazendo exemplos e soluções já praticados em diferentes setores.

4.3 Desafios

Se por um lado, tem-se uma grande quantidade de dados para descobrir o que antes não era possível, por outro há desafios em lidar com esta enorme quantidade de dados. Existem algumas oportunidades que vêm junto com o Big Data, mas as perguntas em relação ao acesso, armazenamento, busca, compartilhamento, entre outros, ainda restringem e dificultam sua operacionalização.

Chen e Zhang (2014) abordam que há vários problemas que vêm junto com o Big Data. Há um custo em lidar com essa grande quantidade de informação, de forma a ganhar competitividade. No campo da pesquisa científica, reunir o conhecimento de uma produção de dados cada vez maior também é tido como um problema de Big Data, uma vez que requer ferramentas que precisam ser mais inteligentes e sofisticadas para filtrar todos os dados de pesquisa disponíveis (Demirkan & Delen, 2013).

Toda esta vasta quantidade de dados também pode ser desestruturada. Devido à sua origem variável, a nova tecnologia para lidar com eles ainda é escassa. A armazenagem de dados tradicional ainda usam ETL (Extração, Transformação e Leitura), processos que não são atualizados com dados para utilização com dados não estruturados (C. P. Chen & Zhang, 2014).

Além disso, o Big Data apresenta um outro ponto desafiador: os dados são de natureza fluida, em constante mudança. Como os *“trending topics”* no Twitter, o Big Data está continuamente se renovando. Os bancos de dados disponíveis não são capazes de armazenar toda essa informação. E ainda que fossem, análises realizadas através de fonte fixa de dados podem não trazer a resposta que a organização precisa (Davenport, Barth, & Bean, 2013).



Não obstante, alguns autores argumentam que há desafios importantes relacionados a recursos humanos, a respeito de treinamento e capacitação, que não podem ser facilmente encontrados. É preciso conseguir profissionais com conhecimentos diversos, que combinados possam obter conhecimento relevantes a partir de bases de dados de Big Data (Bartels, Barbeito, & Mackensen, 2011; Klingström et al., 2013; Mavandadi et al., 2012). ‘1

Além disso, existem mais desafios inerentes ao funcionamento do Big Data. Inconsistência e incompletude de dados, devido à falta de normalização presente nos *data warehouse* tradicionais, que restringem a análise do conjunto de dados (Blocker & Meng, 2013). Embora haja uma série de ferramentas e técnicas com objetivo de limpar os dados, às vezes não há tempo disponível para analisar conjuntos com vários gigabytes ou mais, tornando a inconsistência dos dados ainda um desafio (Chen & Zhang, 2014).

Outras restrições ao Big Data são: formação de uma cultura de tomada de decisão baseada em evidências (McAfee & Brynjolfsson, 2012), segurança e ética no uso dos dados (Davenport, 2014) e a formação do capital humano de desenvolvimento e suporte (H. Chen et al., 2012b). O "direito a ser esquecido" é um exemplo apresentado pelos autores que refletem o aspecto invasivo que acompanha o Big Data (Shum et al., 2012; Wigan & Clarke, 2013). Estes são apenas alguns exemplos de limitações que se revelam oportunidades para cientistas e profissionais, no seu entendimento e busca de soluções. Os desafios são tão grandes quanto os dados em si, e compreender como lidar com eles é importante para extrair o melhor possível do Big Data. O Quadro 5 retoma os principais aspectos sobre Big Data, considerando seus desafios.

Principais desafios	Autores
A quantidade de dados em si.	Demirkan & Delen (2013); Chen & Zhang (2014)
Fluxo constante de informações	Demirkan & Delen (2013); Chen & Zhang (2014); Fiore <i>et al</i> (2013); (Chang, Kauffman, & Kwon, 2014)
Recursos Humanos	Klingström <i>et al</i> (2013); Bartels, Barbeito & Mackensen (2011); Mavandi <i>et al</i> (2012)
Inconsistência dos dados (e potencial irrelevância)	Chen & Zhang (2014); Blocker & Meng (2013)
Privacidade	Wigan & Clarke (2013); Shum <i>et al.</i> (2012)

Quadro 5: Principais desafios com o Big Data

Fonte: Elaborado pelo Autor a partir dos autores citados

5 CONCLUSÕES

Um panorama genérico da evolução dos estudos sobre o Big Data relacionado aos negócios foi o objeto de análise desta revisão. A primeira conclusão, neste sentido, é que a partir da base de dados presente no repositório Scopus, existem diversos estudos que associam o fenômeno do Big Data com os negócios. Além disso, há uma curva com notório crescimento de interesse destes temas relacionados, assim como ocorre com o tema Big Data pesquisado isoladamente, conforme os estudos de Luvizan et al. (2014).

Este crescimento é evidenciado tanto em estudos produzidos por pesquisadores ligados às empresas, quanto os ligados às instituições de ensino e pesquisa. No segmento empresarial, há forte concentração por parte da IBM. Por vezes, estas produções revelam caráter de propaganda, inclusive pela observação dos títulos, parecidos com chamadas publicitárias, como nos exemplos: “Encarar o Big Data é a nova regra”, ou então “5 passos para a migração de dados para o IBM DB2 como acelerador BLU”. Entre as instituições de ensino e pesquisa, ao contrário, há uma relativa pulverização entre as instituições. As três instituições com maior produção deste estudo foram a Universidade de Southampton do Reino Unido, o Instituto de Tecnologia de Massachussetts (MIT) e a Universidade de



Harvard, respectivamente. No Quadro 6 encontra-se uma breve descrição da abordagem dos temas tratados nos estudos destas instituições, vislumbrando as possibilidades de uso do Big Data em diversos negócios, mas também avaliando suas limitações, dificuldades e desafios.

Outra conclusão é com relação ao país de origem dos trabalhos. Há concentração dos estudos produzidos nos Estados Unidos, enquanto que nos demais países estão 65,1% dos estudos, equivalente a 34 países diferentes, distribuídos em todos os continentes, corroborando com Woo Park (2013), quando diz que cientistas, decisores e profissionais corporativos de todo o mundo têm se atentado para as práticas corporativas do Big Data.

No Brasil, assim como evidenciado por Luvizan (2014), não são identificadas quantidades significativas de estudos sobre Big Data no período analisado. Da mesma maneira, na relação com negócios não são encontrados volumes de estudos neste sentido, assim como nos trabalhos citados na seção 4.1 que, em sua maioria, são estudos exploratórios da temática.

Os caminhos apontados nos estudos que relacionam Big Data e negócios são de áreas bastante diversas, relacionadas ao apoio à tomada de decisão (Davenport, 2014). Algumas áreas estão em evidência nos estudos aqui abordados, como as áreas de marketing e gestão do relacionamento com o cliente (Leeflang, Verhoef, Dahlström, & Freundt, 2014; Shaw, 2014), construção de modelos preditivos (Jackson, 2014), área da saúde (Sengupta, 2013) e atividades de *crowdsourcing* ou colaboração em massa (Garcia Martinez & Walton, 2014; Montejo-Ráez, Díaz-Galiano, Martínez-Santiago, & Ureña-Lopez, 2014; Petroni, Querzoni, Beraldi, & Paolucci, 2014).

O Big Data pode ainda ajudar a melhor compreender e otimizar a forma como um anestesista conduz um procedimento cirúrgico (Bartels, Barbeito, & Mackensen, 2011). Os autores citam a gestão e a análise de dados como chaves para a evolução, uma vez que todas as novas tecnologias de cuidados médicos tem criado uma grande quantidade de dados para os profissionais de saúde, citando os prontuários eletrônicos, os leitores de código de barras e os equipamentos de raios-X digitais. Simpao et al. (2014), por sua vez, afirmam que o Big Data gerado na relação com os pacientes tem potencial para lidar com a saúde de um modo mais personalizado, além de permitir o diagnóstico médico através de evidências de dados.

Conforme relatado, há uso do Big Data em áreas comuns de negócios, como no relacionamento com o cliente e demais aplicações de marketing (Davenport, 2014), nas áreas da saúde (Nash, 2014), transportes públicos (Tao, Corcoran, Mateo-Babiano, & Rohde, 2014), varejo e consumo (Pousttchi & Hufenbach, 2014), ensino e educação (Llorente & Morant, 2014) e até na área jurídica (Kemp, 2014). A variedade de áreas e temas abordados nos trabalhos colhidos para este estudo indicam a vasta amplitude do uso corporativo, atual e futuro, dos dados em volume.

Tal qual as oportunidades, os desafios também são relatados pelos autores estudados, Assuntos como a segurança cibernética (Brown, Chul, & Manyika, 2011), dos fatores éticos dos dados (Berinato, 2014), da formação do capital humano (H. Chen et al., 2012b), entre outros, são questões que fazem despertar a atenção em torno de múltiplos aspectos deste advento.

A ligação entre Big Data e negócios é bastante complexa e não se esgota neste estudo, até porque isso não seria possível dada a sua amplitude e fluidez. Neste sentido, cabem diversos estudos adicionais que evidenciem o valor deste advento tanto para a prática, como para os meios científicos nas áreas aqui citadas, mas também em outras. Estudos futuros também podem se atentar para estudos disponíveis em outros bancos de dados científicos, já que este estudo averiguou apenas aqueles presentes no repositório Scopus. Também é oportuno ampliar o estudo para o ano de 2015, não incluído aqui para fins de comparação de evolução anual, assim como a associação a outras palavras-chave. Da mesma maneira,



estudos empíricos também são maneiras indicadas de descobrir outras oportunidades e aplicações de Big Data no ambiente corporativo.

REFERÊNCIAS

- Agrawal, D. (2014). Analytics based decision making. *Journal of Indian Business Research*, 6(4), 332–340.
- Bartels, K., Barbeito, A., & Mackensen, G. B. (2011). The anesthesia team of the future. *Current Opinion in Anesthesiology*, 24(6), 687–692.
- Becker, J. L., Freitas, H., Hoppen, N., & Kladis, C. (1997). Informação e decisão: Sistemas de apoio e seu impacto. *Porto Alegre: Ortiz*, 74.
- Berinato, S. (2014). With big data comes big responsibility. *Harvard Business Review*, (November 2014).
- Bishop, N. (2014). Spotlighting big data and analytics heroes. *IBM Data Management Magazine*, (1).
- Björk, B. C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guhnason, G. (2010). Open access to the scientific journal literature: situation 2009. *PloS one*, 5(6), e11273.
- Blocker, A. W., & Meng, X.-L. (2013). The potential and perils of preprocessing: Building new foundations. *Bernoulli*, 19(4), 1176–1211.
- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, 80(4), 571–583.
- Breternitz, V. J., & da Silva, L. A. (2013). Big Data: bringing new opportunities and challenges. In 10 *CONTECSI*.
http://www.infoteca.inf.br/contecsi/smarty/templates/arquivos_template/upload_arquivos/a_cervo/docs/PDFs/157.pdf
- Brown, B., Chul, M., & Manyika, J. (2011). Are you ready for the era of “big data”? *McKinsey Quarterly*, (4), 24–27+30–35.
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67–80.
- Chen, C. P., & Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, 36(4), 1165–1188.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209.
- Chow-White, P. A., Green Jr, S., & others. (2013). Data Mining Difference in the Age of Big Data: Communication and the social shaping of genome technologies from 1998 to 2007. *International Journal of Communication*, 7, 28.
- Silva, I. M., & de Campos, F. C. ([s.d.]). NEW PERSPECTIVES USING BIG DATA: A STUDY OF BIBLIOMETRIC. Recuperado de <http://www.contecsi.fea.usp.br/envio/11contecsi-files/papers/540/submission/director/540-2359-1-DR.pdf>
- Davenport, T. H. (2014). How strategists use “big data” to support internal business decisions, discovery and production. *Strategy and Leadership*, 42(4), 45–50.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How “big data” is different. *MIT Sloan Management Review*, 54(1).



- Davenport, T. H., Barth, P., & Bean, R. (2013). How “big data” is different. *MIT Sloan Management Review*, 54(1).
- Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412–421.
- Garcia Martinez, M., & Walton, B. (2014). The wisdom of crowds: The potential of online communities as a tool for data analysis. *Technovation*, 34(4), 203–214.
- Goldman, A., Kon, F., Junior, F. P., Polato, I., & de Fátima Pereira, R. (2012). Apache Hadoop: Conceitos teóricos e práticos, evolução e novas possibilidades. *XXXI Jornadas de atualizações em informática*.
- Hayashi, A. M. (2014). Thriving in a Big Data World. *MIT Sloan Management Review*, 55(2), 35–39.
- Huang, T., & Van Mieghem, J. A. (2014). Clickstream data and inventory management: Model and empirical analysis. *Production and Operations Management*, 23(3), 333–347.
- James, R. (2014). Out of the box: Big data needs the information profession - the importance of validation. *Business Information Review*, 31(2), 118–121.
- Kemp, R. (2014). Legal aspects of managing Big Data. *Computer Law and Security Review*, 30(5), 482–491.
- Klingström, T., Soldatova, L., Stevens, R., Roos, T. E., Swertz, M. A., Müller, K. M., ... others. (2013). Workshop on laboratory protocol standards for the molecular methods database. *New biotechnology*, 30(2), 109–113.
- Leefflang, P. S. H., Verhoef, P. C., Dahlström, P., & Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European Management Journal*, 32(1), 1–12.
- Llorente, R., & Morant, M. (2014). Wearable computers and big data: Interaction paradigms for knowledge building in higher education. In *Innovation and Teaching Technologies: New Directions in Research, Practice and Policy* (p. 127–137).
- Luvizan, S., Meirelles, F., & Diniz, E. H. (2014). Big Data: publication evolution and research opportunities. In *11ª Conferência Internacional sobre Sistemas de Informação e Gestão de Tecnologia*. Recuperado de <http://www.contecsi.fea.usp.br/envio/index.php/contecsi/11contecsi/paper/view/694>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. Recuperado de <http://www.citeulike.org/group/18242/article/9341321>
- Marchand, D. A., & Peppard, J. (2013). Why IT fumbles analytics. *Harvard Business Review*, 91(1), 104–112.
- Mavandadi, S., Dimitrov, S., Feng, S., Yu, F., Yu, R., Sikora, U., & Ozcan, A. (2012). Crowd-sourced BioGames: managing the big data problem for next-generation lab-on-a-chip platforms. *Lab on a chip*, 12(20), 4102–4106.
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana* (Vol. 1). Elsevier Brasil.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 4.
- Millie, D. F., Weckman, G. R., Young, W. A., Ivey, J. E., Fries, D. P., Ardjmand, E., & Fahnenstiel, G. L. (2013). Coastal “Big Data” and nature-inspired computation: Prediction potentials, uncertainties, and knowledge derivation of neural networks for an algal metric. *Estuarine, Coastal and Shelf Science*, 125, 57–67.
- Montejo-Ráez, A., Díaz-Galiano, M. C., Martínez-Santiago, F., & Ureña-López, L. A. (2014). Crowd explicit sentiment analysis. *Knowledge-Based Systems*, 69(1), 134–139.



- Nash, D. B. (2014). Harnessing the power of big data in healthcare. *American Health and Drug Benefits*, 7(2), 69–70.
- Park, H. W., & Leydesdorff, L. (2013). Decomposing social and semantic networks in emerging “big data” research. *Journal of Informetrics*, 7(3), 756–765.
- Petroni, F., Querzoni, L., Beraldi, R., & Paolucci, M. (2014). *LCBM: Statistics-based parallel collaborative filtering* (Vol. 176 LNBIP).
- Pousttchi, K., & Hufenbach, Y. (2014). Engineering the value network of the customer interface and marketing in the data-rich retail environment. *International Journal of Electronic Commerce*, 18(4), 17–41.
- Rust, R. T., & Huang, M.-H. (2014). The service revolution and the transformation of marketing science. *Marketing Science*, 33(2), 206–221.
- Salvador, A. B., & Ikeda, A. A. (2014). O uso do big data no sistema de informação de marketing. In *11ª Conferência Internacional sobre Sistemas de Informação e Gestão de Tecnologia*.
<http://www.contecsi.fea.usp.br/envio/index.php/contecsi/11contecsi/paper/viewPaper/567>
- Sengupta, P. P. (2013). Intelligent platforms for disease assessment: novel approaches in functional echocardiography. *JACC: Cardiovascular Imaging*, 6(11), 1206–1211.
- Shaw, R. (2014). The marketing data space race. *Journal of Direct, Data and Digital Marketing Practice*, 15(4), 260–261.
- Shum, S. B., Aberer, K., Schmidt, A., Bishop, S., Lukowicz, P., Anderson, S., Dunwell, I. (2012). Towards a global participatory platform. *The European Physical Journal Special Topics*, 214(1), 109–152.
- Simpao, A. F., Ahumada, L. M., Gálvez, J. A., & Rehman, M. A. (2014). A review of analytics and clinical informatics in health care. *Journal of medical systems*, 38(4), 1–7.
- Souza, A., Prado, E., Peres, S., & Sun, V. (2014). Critérios para Seleção de SGBD NoSQL: uma Aplicação da Técnica Delphi. In *11ª Conferência Internacional sobre Sistemas de Informação e Gestão de Tecnologia*.
- Tao, S., Corcoran, J., Mateo-Babiano, I., & Rohde, D. (2014). Exploring Bus Rapid Transit passenger travel behaviour using big data. *Applied Geography*, 53, 90–104.
- Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2), 127–151.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, 14(3), 207–222.
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*, 26(2), 3.
- Wigan, M. R., & Clarke, R. (2013). Big data’s big unintended consequences. *Computer*, 46(6), 46–53.
- Zikopoulos, P., Lightstone, S., Huras, M., Sachedina, A., & Baklarz, G. (2013). New dynamic in-memory analytics for the era of big data. *IBM Data Management Magazine*, (4), 1–47.