

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

GT-8 – Informação e Tecnologia

POTENCIAL DE INOVAÇÃO ATRAVÉS DO BIG DATA ANALYTICS: PANORAMA DAS ÚLTIMAS PESQUISAS

Letícia de Castro Peixoto (Universidade Federal de Minas Gerais - UFMG)

Ricardo Rodrigues Barbosa (Universidade Federal de Minas Gerais - UFMG)

INNOVATION POTENTIAL THROUGH BIG DATA ANALYTICS: BIG PICTURE OF RECENT RESEARCHES

Modalidade da Apresentação: Comunicação Oral

Resumo: A nova economia de dados inclui uma variedade de ferramentas tecnológicas e metodológicas para melhorar a análise dos dados e informações disponíveis. Neste contexto, um dos destaques do *Big Data Analytics* (BDA) em organizações tem sido o potencial para promover grandes inovações. Este artigo pretende apresentar uma análise das pesquisas recentes sobre este tema. Realizou-se uma Revisão de Literatura Sistemática por meio da qual foram investigadas as publicações desde 2014 sobre este tema para possibilitar um melhor entendimento das potencialidades das aplicações BDA para criar inovação em produtos e processos. A análise indica que os estudos têm enfatizado desenvolvimentos em análises preditivas e gestão do conhecimento, os países com maior número de publicações são Estados Unidos, Alemanha e Reino Unido e parte relevante dos resultados destes estudos tem sido apresentado como *frameworks*. A estrutura deste artigo é a seguinte: primeiro é apresentado cenário contextual do tema e sua relevância; A seção que se segue descreve brevemente BDA e seus potenciais pensáveis; A Seção 3 descreve a metodologia e resumos das tipologias de artefatos usadas; A seção 4 mapeia os resultados e apresenta um panorama; E, finalmente, considerações e limitações de pesquisa são apresentadas.

Palavras-Chave: *Big Data Analytics*; Tomada de Decisão; Inovação.

Abstract: The new data economy includes a diversity of technological and methodological tools to improve the analyses of the data and information available. This paper aims to present a perspective of recent researches in this theme. A peculiar Systematic Literature Review was performed investigating publications since 2014 to produce a big-picture of these recent researches in Big Data Analytics that aims to better understand the potential of Big Data Analytics applications to create innovation in products and process. The analysis indicates these studies have emphasized developments in predictive analysis and knowledge management, the countries with the largest number of publications are United States, Germany and UK and relevant part of the results of these studies has been presented as frameworks. The structure of this paper is as follows: first is presented contextual scenario of theme and its relevance; the ensuing section briefly describes about Big Data Analytics and its thinkable potentials; section 3 describes the methodology and summaries artifact typologies used; section 4 maps the results and presents a panorama; and finally, considerations and research limitations are presented.

Keywords: Big Data Analytic, Decision, Innovation

1 INTRODUÇÃO

As organizações contemporâneas têm-se deparado com significativos desafios no processo de aperfeiçoamento do uso da informação atualmente disseminada. Informação oriunda de dados que se apresentam em enorme Volume, Variedade, Velocidade (LANEY, 2001), Veracidade e Valor (ZIKOPOULOS; EATON et al, 2011) denominados recentemente de Big Data (BD). Acredita-se que o grande interesse dos mundos acadêmico e da prática profissional em melhor entender Big Data está no seu potencial de transformar radicalmente os processos de negócio (WAMBA; AKTER et al, 2015; JIN; WAH et al, 2015), gerando inovação e estimulando o processo decisório orientado a dados ou *data-driven* (KLATT; SCHLAEFKE, 2011). A análise destes volumosos dados, recém-denominada de *Big Data Analytics* (BDA), é tipicamente caracterizada pelo foco em dados não estruturados, de grande volume e rápido movimento (DAVENPORT, 2014, apud LOEBBECKE & PICOT, 2015). Estes dados não estruturados recentemente têm sido, inclusive, considerados “o novo óleo” desta geração, segundo Actio e Khatri (2014).

Considerando o novo cenário acima apresentado, tornou-se relevante apreender como o BDA permitirá às organizações melhor entenderem seus clientes, seus produtos, serviços e processos e procurarem antever necessidades para se encontrar soluções em tempo hábil. Descobrir conhecimento em uma grande massa que possam ser usadas para tomar decisões, fazer previsões e inovar, despertou grande interesse no mundo organizacional. Porém, sem o adequado tratamento e análise dos dados coletados, não há geração de métricas e, como não se mede o que não se vê, não é possível gerenciar o que não se mede. Portanto, prosperar na nova economia de dados exige repensar em como a análise de dados pode criar valor e ampliar possibilidades de incrementos para empresas e seus *stakeholders*, entendendo suas preferências e tendências de arranjos organizacionais.

Neste contexto, investigar como os recentes estudos sobre Big Data Analytics (BDA), estão sendo conduzidos servirá de alicerce para desenvolvimento do conhecimento na área, tanto no meio acadêmico quanto na prática. Pretende-se, assim, neste estudo exploratório, sondar como este tema tem sido abordado no meio acadêmico, que países representam polos de pesquisa nesta área e quais os seus enfoques e perspectivas.

A seguir, apresenta-se a metodologia aplicada na busca dos estudos recentes de BDA e seu potencial inovador, e, posteriormente, expõem-se os achados desta pesquisa exploratória e suas implicações.

2 OCEANO DE DADOS NO MUNDO DOS NEGÓCIOS

A facilidade com que dados e informações são, hoje, gerados, coletados e disseminados tem difundido, a falsa sensação de detenção da íntegra da informação ou de conhecimento pleno dos fatos e dados a cerca de um evento. No entanto, em grande parte das vezes, não são feitas suficientes análises e avaliações dos dados para averiguar a veracidade e relevância da informação ou ainda sua correlação com outros eventos relacionados ao contexto que se investiga.

A conectividade recente da humanidade possibilita o monitoramento da dinâmica dos eventos e tendências em tempo real, sejam eles gerados pelo homem ou pelas “coisas” feitas por ele (denominada de “internet das coisas”). No entanto, o enorme volume de dados atualmente produzido, quando não se conhecem suas causas, efeitos ou relações, não produz informação relevante ou conhecimento reconhecidamente diferenciado ou inovador. Como consequência, em contextos organizacionais, esse bombardeio de dados dificulta a visualização das conexões entre dados, informação, conhecimento e decisões, comprometendo assim a qualidade do processo decisório (DAVENPORT & HARRIS, 2010). Assim sendo, a vantagem competitiva sustentável de uma organização está fortemente relacionada a esta capacidade interpretativa e antecipatória, mais recentemente denominada de Inteligência Analítica por Davenport e Haris (2010). Este novo ecossistema de dados e métodos aproveita, então, as análises de padrões de informações para tomar decisões inteligentes oportuniza, portanto, a proposições de valor e busca de inovação.

Estudos recentes examinam e refletem sobre as possíveis transformações do modelo de negócios com a rápida taxa de desenvolvimentos das iniciativas digitais, incluindo o *Big DataAnalytics*, que vem sendo considerado o novo “oceano azul” na criação de oportunidades de negócios (KWON; LEE, 2014). O modo como o *Big Data* e a necessidade de análises avançadas em decisões estratégicas informa e reforma (*design* e implementação) os Sistemas de Gestão do Conhecimento (SGC) foi recentemente avaliado por Intezari e Gressel (2017). Esses autores ressaltam que a crescente demanda por análises em tempo real, e em fluxo contínuo, promove reações mais rápidas e flexíveis, o que intensifica esta visão reformista.

Portanto, o surgimento do *Big Data* requer ajustes inevitáveis nos SGCs para permitir que as organizações e os gestores integrem esta nova perspectiva em seu conhecimento e visão especializada para tomar decisões estratégicas mais eficazes. Chen et al (2012) e Loebbeck (2015) consideram que o BDA é uma evolução do *Business Intelligence* (BI), uma vez que amplia o escopo do BI, foca na integração e comunicação de dados estruturados e semiestruturados, tratando também os dados não estruturados, que constituem 95% do *Big Data* (GANDOMI;HAIRDER, 2015). Estes dados são originários de fontes como a web, dispositivos móveis e redes de sensores que são externas à empresa.

Os reais desafios desta explosão de fontes de informação e, possivelmente, de conhecimento e inovação, não estão exatamente no volume de dados a serem tratados, mas na diversidade de tipos de dados (variedade), nos requerimentos de respostas imediatas (velocidade) e na incerteza destes dados (veracidade). Zikopoulos e Eaton (2011) enfatizam que a dimensão “valor” como a capacidade de tais fontes em trazer respostas aos complexos problemas atuais e, desta forma, aprimorar os resultados organizacionais. Para tanto, a qualidade destes dados está intimamente ligada ao valor que podem gerar, uma vez que dados pobres podem ter impacto direto nas decisões de negócio e ocasionar perdas tangíveis e intangíveis (HAZEN; BOONE, 2014; ISASI et al, 2015). Assim sendo, diversos estudos têm buscado conceber e implementar técnicas, métodos, modelos, guias, tecnologias, sistemas e aplicativos para melhor análise e gestão dos dados e informações. Por outro lado, essa busca de soluções tem tornado se cada vez mais desafiadora devido à necessidade de se explorar todo o potencial dos dados disponíveis e, assim, contribuir para o desenvolvimento de inovações e modelos de gestão (LOEBBECKE; PICOT, 2015) mais aderentes às demandas de uma gestão baseada em objetivos (KLATT; SCHLAEFKE, 2011).

Encorajado pela necessidade de análise em massa de uma escalabilidade exponencial quase em tempo real dos dados gerados pela humanidade (incluindo a internet das coisas), o BDA está sendo definido como uma evolução dos conceitos e terminologias como “apoio a decisão”, “*on line analytical processing*” e “*Business Intelligence*” (LOEBBECKE; PICOT, 2015). Vários autores já mesclam ou aglutinam BI e Business Analytics (BA) (CHEN et al, 2012; FOSHAY et al, 2015; KOWALCZYK; BUXMANN, 2015), como uma nova tendência. Chen et al (2015) acreditam que *Big Data Analytics* é um campo que oferece novas direções para *Business Intelligence and Analytics* (BI&A), sendo BA uma combinação de BI e análises mais sofisticadas, com o amplo uso de dados, análise quantitativa e sofisticação matemática para

conduzir a uma melhor tomada de decisão dentro das organizações (LAURSEN; THORLUND, 2010; SCHLAEFKE; SILVI, 2013). A relevância do BDA também se sustenta na acentuada demanda por soluções inovadoras, e consequentes investimentos feitos pelas organizações nesta busca, inclusive em ferramentas de BI e BA, que se têm tornado essenciais na formulação e execução estratégica das organizações. Fato este também comprovado pela existência dos cargos de *Chief Information Officer* (CIO) e pelos novos cargos de *Chief Data Officer* (CDO) (GALBRAITH, 2014) e *Chief Technology Officer* (CTO), tão ventilados ultimamente. Cargos com títulos como *Chief Digital Officer* ou *Chief Web Officer* já podem ser encontrados no LinkedIn, dadas proporções de investimentos nestas áreas. Estas empresas que almejam considerar Big Data em suas estratégias de gestão terão que inovar na administração de dados, informação e processos de decisão para sustentar as inovações estratégicas que certamente determinarão novas estruturas e processos organizacionais (GALBRAITH, 2014; ACTIO; KHATRI, 2014).

Na perspectiva da indústria de informação, Big Data apresenta-se como um forte impulso para a próxima geração da indústria de TI, - que apresenta-se muito referenciada no Big Data, computação em nuvem, internet móvel e negócios sociais (JIN; WAH et al, 2015). O instituto Internacional Data Corporation (IDC), em seu relatório anual de 2015 “*Big Data and Analytics Spending Guide*”, prevê que até 2020 a dimensão do mercado desta terceira plataforma TI chegará a US\$ 5,3 trilhões, correspondendo a 90% do crescimento no setor de TI. Ainda estima-se que a grande maioria dos dados armazenados no mundo foi produzida apenas nos últimos dois anos anteriores ao ano de divulgação do relatório e os rastros desses dados continuam duplicando a cada ano.

No contexto da BDA, a análise e interpretação dos dados incluem análises de predição, que buscam descobrir padrões através da captura de relacionamentos de dados e, segundo Gandomi (2015), são baseadas em métodos estatísticos, e se dividem em dois subgrupos: técnicas de regressão e técnicas de aprendizagem de máquina ou *machine learning*. Estas técnicas, apoiadas tanto pela aprendizagem de máquina quanto pela regressão estatística, buscam efetuar previsões ou antecipações assertivas a cerca do futuro a partir dos dados, estruturados (tabelas) ou não estruturados (texto e imagens).

Considerando os elementos acima apresentados, por meio do presente estudo busca-se desenhar um panorama do que tem se pesquisado sobre as potencialidades do BDA em contribuir para a inovação de processos, produtos e serviços organizacionais.

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

3 METODOLOGIA

Este artigo apresenta os resultados de uma Revisão Sistemática da Literatura (RSL) com o objetivo de sumarizar as evidências existentes, os benefícios e limitações dos estudos sobre BDA (KITCHENHAM; BRERETON, 2013), identificar as tendências das pesquisas mais recentes e ainda descrever o perfil dos estudos acadêmicos de BDA no intuito de incentivar o desenvolvimento de novas atividades de pesquisa. A RSL é uma técnica de pesquisa passível de ser adotada para se explorar e deduzir evidências pertinentes a uma determinada questão de pesquisa (SIVARAJAH et al, 2017). No contexto restrito deste artigo, a RSL objetivou aparelhar o panorama delineado pelas recentes pesquisas do tema. Para tanto, alguns critérios de seleção foram aplicados na recuperação das publicações.

A coleta de dados foi realizada na base de dados Scopus durante os últimos 42 meses (anos de 2014, 2015, 2016 e 2017 até o mês de junho). O foco se restringiu ao período mais recente dado a rápida evolução da tecnologia, que aumenta a obsolescência da ciência, e a novidade no tema como campo de pesquisa. A base Scopus usada é o maior banco de dados de resumo e citações de literatura revisada por pares, com ferramentas inteligentes para rastrear, analisar e visualizar pesquisas (BAR-ILAN, 2008). Possui mais de 21.500 títulos de mais de 5.000 editores internacionais, fornecendo, portanto, uma visão abrangente do resultado da pesquisa mundial nos campos da ciência, tecnologia e ciências sociais que interessam a esta pesquisa (SCOPUS CONTENT COVERAGE GUIDE, 2016). Segundo Simões (2015), as revistas de resumos assumiram importante papel possibilitando a divulgação e acesso aos conteúdos substanciais dos documentos primários, fato este que corroborou o aumento considerável do número de revistas de resumos e a alterações nos seus conteúdos e processos editoriais. Adicionalmente, como apontado por Souza, Almeida e Baracho (2013), foram realizados recortes nos resultados das buscas devido ao grande volume de publicações e diversidade de fontes.

Os argumentos de busca dos artigos e trabalhos apresentados em eventos foram (*"big data analytics" OR "big data analysis"*) AND (*insight* OR innovation**), restringindo-se aos documentos publicados na língua inglesa, na subárea Business, Management e Accounting, nos anos de 2014 a 2017. O resultado desta seleção feita no texto dos resumos, títulos e palavras chaves recuperou 362 publicações científicas.

4 RESULTADOS

4.1 Artefatos produzidos

O foco da busca foi identificar publicações que apresentaram artefatos como produtos das pesquisas conduzidas. Este critério se justifica pela visão pragmática imposta a este estudo. Desta forma, nesta seção faz-se uma breve explanação sobre as definições dos possíveis artefatos gerados, baseados na tipologia construída por Herver (2004), que normalmente resultam de *Design Science Researches*. Esta distinção se faz relevante para análise de quais destes artefatos estão sendo mais aplicados na descoberta do potencial inovador do BDA. A *Design Science* (DS) é o estudo de artefatos em um contexto (WEIRINGA, 2012) e representa o desejo de fazer pesquisa com resultados mais relevantes, através da adição do ciclo de resolução de problemas ao ciclo de construção de teorias. Posto isto, acredita-se que este critério abarcaria as pesquisas mais criteriosas e com resultados mais robustos.

Os tipos de artefatos descritos por Herver (2004) são: constructos, modelos, métodos e instanciações. Estas tipologias são normalmente difíceis de operacionalizar devido a uma relativa confusão entre os conceitos. Objetivando reduzir este possível interlace entre os conceitos, as definições foram emprestadas do estudo já realizado por Mwilu et al. (2016). Neste estudo, o autor reúne a definição de vários autores, aqui não relacionados, e define as tipologias como demonstrado no quadro 1.

Com base nas tipologias descritas por Herver (2004), das 362 publicações recuperadas, 179 não continham resumos científicos ou descritivos com menção ao desenvolvimento de novos artefatos como os mencionados na figura 1 ou ainda mencionavam BDA de forma secundária e não como foco de pesquisa. No entanto, 28 dessas 179 publicações, apresentaram achados em seus resumos como análises de *surveys*, estudos de caso e revisões sistemáticas de literatura sobre BDA, embora não tenham mencionado exatamente os artefatos descritos na figura 1. Este grupo de publicações foi categorizado pelo autor como “Analítico”. Desta forma, dos 151 dos 362 artigos (41%) foram eliminadas por possuírem resumos científicos sem descrição dos resultados da pesquisa ou por não apresentarem resumos.

Quadro 1: Tipologias organizadas por Mwilu, Comyn-Wattiaua e Prat (2016)

TIPOLOGIAS DE ARTEFATOS	SUB- CATEGORIAS & CONCEITOS
<p align="center">Constructo</p> <p>Nesta categoria se inserem: linguagem, meta-modelo e conceito</p>	<p>Linguagem: um conjunto de conceitos, símbolos, combinações de regras (sintaxe) para interpretação de símbolos (semântica).</p>
	<p>Meta-modelo: um conjunto de conceitos representados em notação gráfica, com regras para combinar os conceitos.</p>
	<p>Conceito: uma nova construção adicionada a uma linguagem existente ou meta-modelo.</p>
<p align="center">Modelos</p> <p>Sendo (Mwilu, Comyn-Wattiaua, & Prat, 2016), quando constructos são desenvolvidos para estruturar objetos, tem-se modelos. E estes poderiam ser separados em: Design de Sistemas, Ontologias, Taxonomias e Arquiteturas.</p>	<p>Design de Sistemas: uma estrutura ou descrição de comportamento usando alguma notação gráfica e possivelmente textual.</p>
	<p>Ontologias: uma especificação formal explícita de uma conceituação compartilhada.</p>
	<p>Taxonomias: uma classificação de objetos em um domínio de interesse, baseada em características comuns.</p>
	<p>Framework: uma estrutura lógica para organizar informações complexas.</p>
	<p>Arquiteturas: um plano que representa a organização fundamental de um sistema envolvendo seus componentes, suas inter-relações e o ambiente.</p>
	<p>Requerimentos: uma condição ou capacidade que deve ser atendida ou processada por um sistema</p>
<p align="center">Métodos</p> <p>A proposta selecionada por Mwilu estabelece que o a categoria de modelo agrupa dinamicamente os artefatos. Deste modo seria: metodologia, guias, algoritmos, fragmentos de métodos e métricas.</p>	<p>Metodologias: um conjunto predefinido de etapas e diretrizes, com técnicas e ferramentas.</p>
	<p>Guias ou Diretrizes: uma sugestão sobre o comportamento em uma situação particular. Exemplos: diretrizes de projeto, orientações, regras.</p>
	<p>Algoritmos: uma sequência executável de operações para executar uma tarefa específica.</p>
	<p>Fragmentos de métodos: um componente do método que pode ser tratado como unidade separada e reutilizados em diferentes contextos.</p>
	<p>Métricas: função que atribui um número ou símbolo a uma entidade a fim de caracterizar um atributo ou um grupo de atributos.</p>
<p align="center">Instanciações</p> <p>São definidos como artefatos específicos para acessar a viabilidade de algum outro constructo, método ou modelo, como os citados anteriormente. Nesta categoria caberiam várias possibilidades.</p>	<p>Sistemas implementados: software ou hardware implementado. Em trabalhos de pesquisa, os sistemas implementados podem ser protótipos ou a implementação de ferramentas finalizadas.</p>
	<p>Outros Exemplos: qualquer outra concretização de um artefato abstrato (constructo, modelo ou método).</p>

Fonte: Mwilu; Comyn-Wattiaua; Prat, 2016.

Os resumos dos 211 trabalhos que relatam o desenvolvimento de algum tipo de artefatos em seus achados ou que podem ser considerados com estudos “analíticos” foram lidos na íntegra, com o intuito de se realizar a seguinte análise: a) definir qual dos artefatos melhor representa os achados de cada publicação para tipificá-los; b) buscar, dentre as palavras-chaves dos autores, aquelas que não estavam em duplicidade com as palavras de busca já utilizadas na base Scopus (*big data analytics, big data analysis, insight, innovation*); c) agrupar as palavras chaves selecionadas anteriormente por grupos de assuntos mais

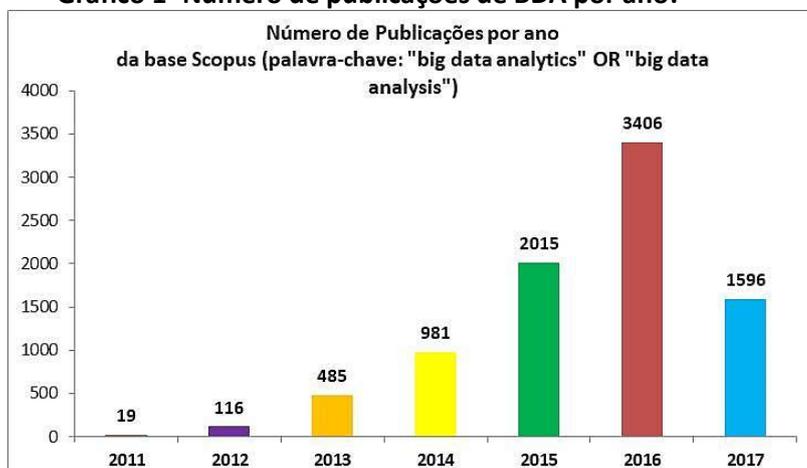
abrangentes; e d) agrupar as publicações em categorias dos diversos setores da economia em que se enquadram, com base na taxonomia criada pelo Departamento de Assuntos Econômicos e Sociais das Organizações Unidas (*International Standard Industrial Classification of All Economic Activities*).

A base de dados, resultante dos procedimentos acima descritos, foi alimentada na ferramenta de *Business Intelligence* da Microsoft, chamada *Power BI*, e produzidos resultados analítico-descritivos. A base completa extraída, com as 362 publicações, também foi utilizada para produção de mapas bibliográficos no software VOSviewer e ambas são apresentados na seção a seguir.

4.2 Publicações e palavras chaves

Previamente à apresentação das análises das 211 publicações selecionadas, confirmou-se que intervalo de período analisado (2014 a 2017) refere-se ao intervalo mais representativo de publicações na área. O gráfico 1 mostra o número de publicações (artigos e conferências) em língua inglesa contidas na base Scopus recuperadas usando somente as palavras chaves (“big data analytics” OR “big data analysis”). Este conjunto de dados não sofreu qualquer restrição de data ou subárea. O número total de publicações desta visão mais ampla é de 8.619. Percebe-se que em 2012 iniciaram-se as publicações sobre o BDA com 116 documentos, seguido de 485 em 2013. Fato este que comprova a concentração das publicações a partir de 2014, com 93% de todas as publicações que utilizam as palavras chaves (“big data analytics” OR “big data analysis”) em seus títulos, resumos ou palavras chaves.

Gráfico 1- Número de publicações de BDA por ano.

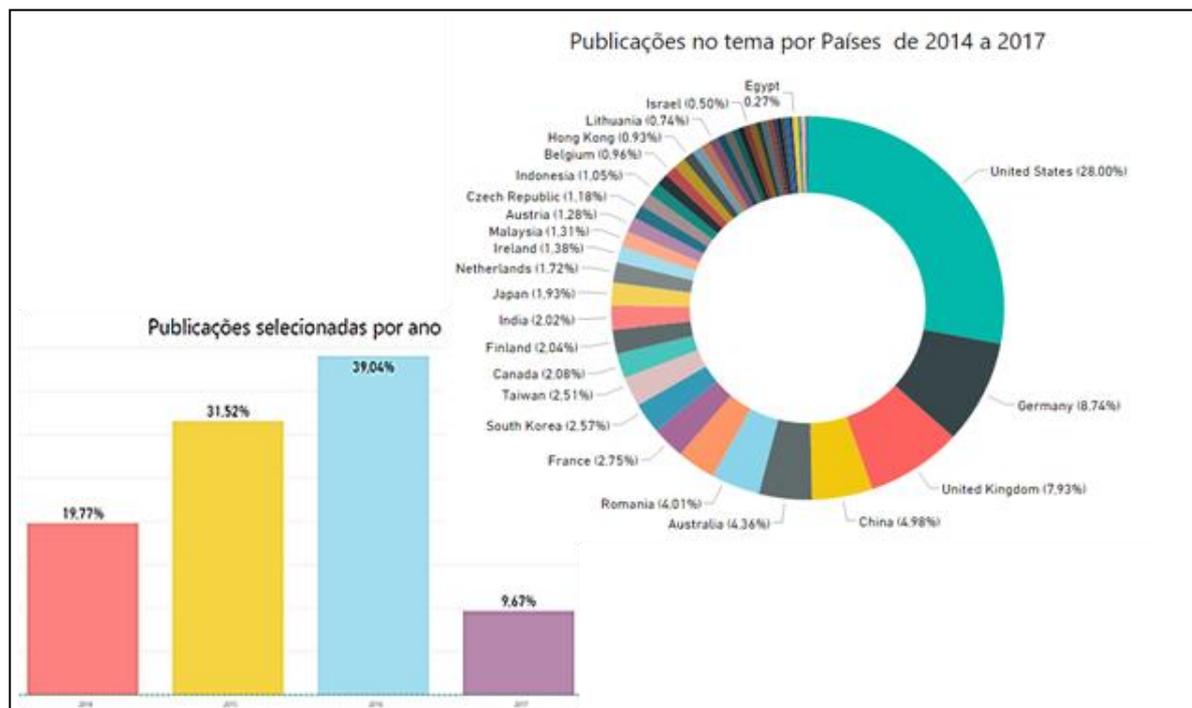


Fonte: Base de Dados Scopus utilizando Power BI.

Esta análise possivelmente infere que muito pouco ainda é estudado sobre o potencial de geração de inovações do BDA, visto que somente 362 das 8.619 publicações (4% do universo recuperado) mencionam inovações ou insights em seus textos.

As análises descritivas das 211 publicações selecionadas, apresentada na figura 1, mostram que as pesquisas neste tema têm crescido a cada ano e que de 2014 a 2016, por exemplo, o número de publicações praticamente dobrou (de 20% para 39% do total das 211 publicações), sendo grande parte delas em forma de artigos científicos (68%). Os países com maior número de publicações em conferências e revistas científicas são os Estados Unidos da América, a Alemanha, Reino Unido, a China e a Austrália que representam juntos mais da metade (54%) do volume de publicações. A figura 1 mostra os 10 países no topo desta lista, que juntos somam 67% das publicações.

Figura 1: Visualizações das publicações por ano, países e tipos de documentos.

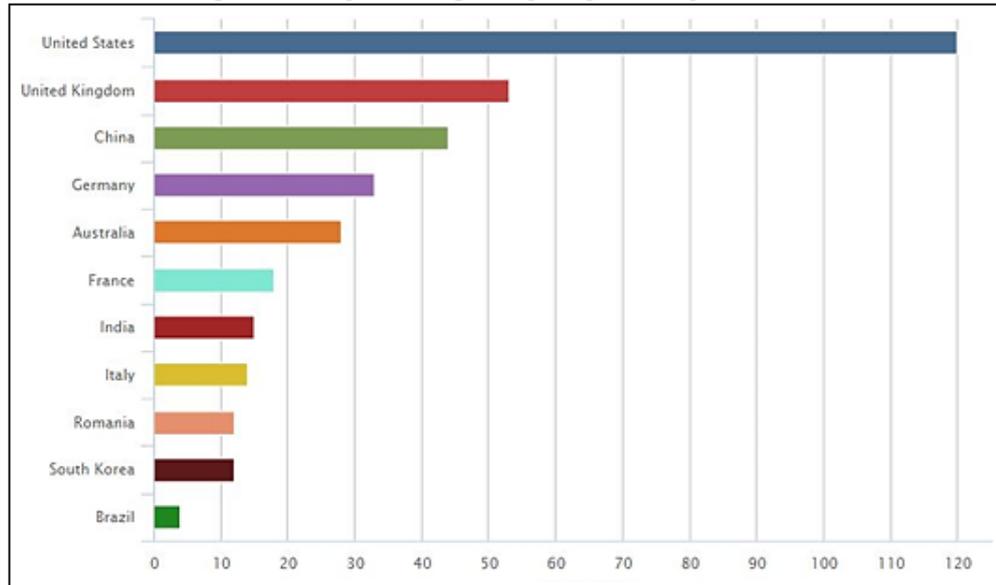


Fonte: Base de Dados Scopus utilizando Power BI.

O Brasil representa uma fatia pouco significativa neste contexto, com apenas 4 publicações dentre as publicações selecionadas primeiramente pela Scopus (1,1% das 362 publicações inicialmente recuperadas) quanto nas selecionadas após aplicação dos critérios, como mostra a figura 2. Para o contexto da pesquisa ampla mostrada no gráfico 1, com 8.619 publicações (desconsiderando as restrições período e subárea e considerando na busca somente as palavras chaves (“big data analytics” AND big data analysis”). Nesta visão ampla, o

Brasil apresentou 119 publicações (1,4% do total de 8.619 publicações). Mostrando que em ambos os cenários a participação do Brasil é próximo de 1%.

Figura 2 – Visualizações das publicações por países para a amostra de 362 publicações

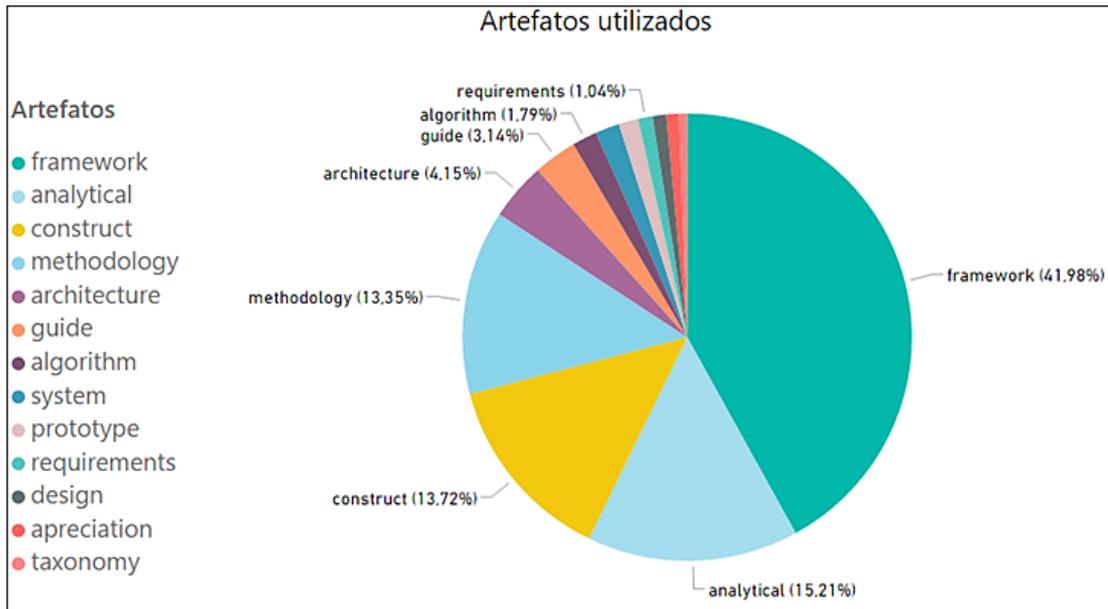


Fonte: Base de Dados Scopus.

Quanto aos artefatos utilizados para desenvolvimento das pesquisas em âmbito de análise, o gráfico 2 mostra que frameworks estão presentes em 42% dos resultados apresentados nas pesquisas. Se adicionarmos os 4% de desenvolvimento de arquiteturas, tem-se 46% de artefatos de modelos. Isto mostra uma tendência comum em áreas novas de conhecimento, onde a academia busca dar corpo e criar uma estrutura lógica para organizar informações complexas gerando modelos conceituais do tema. As pesquisas que não apresentam artefatos de forma explícita, mas que avaliam casos específicos conduzem RSL e analisam estatisticamente ou não particularidades do tema (denominadas de “analytical” pelo autor) representam 15 %, seguidos dos constructos com 14%.

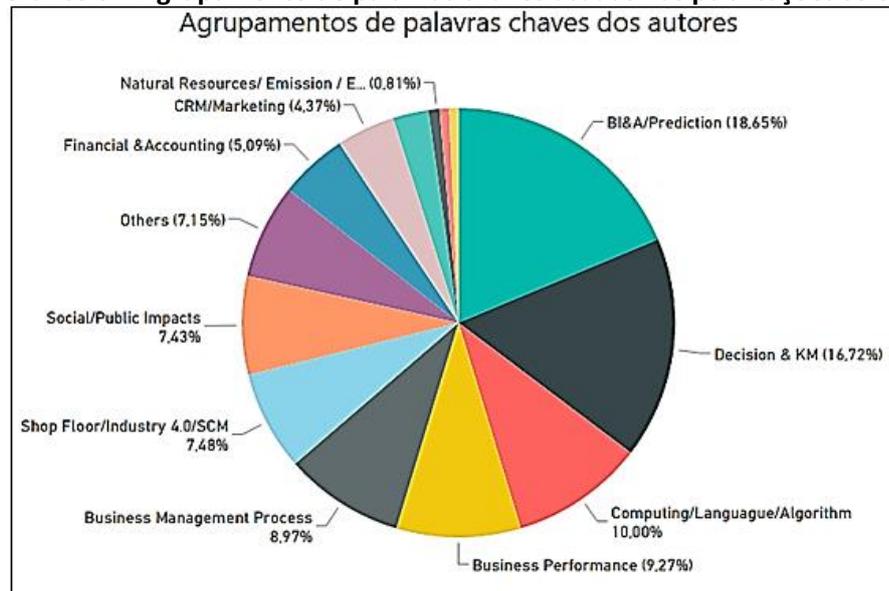
O gráfico 3 mostra a concentração das pesquisas por grupamento de palavras chaves, confirmando que análises avançadas (18,65%) e tentativas de incrementar o processo de tomada de decisão e aquisição de conhecimento (16,72%) tem sido mais exploradas no contexto de BDA. São seguidas por palavras chaves relacionadas à ciência da computação e investigações de performance empresarial.

Gráfico 2 – Distribuição da utilização de artefatos nas publicações selecionadas.



Fonte: Base de Dados Scopus utilizando Power BI.

Gráfico 3 – Agrupamento de palavras chaves usadas nas publicações selecionadas.



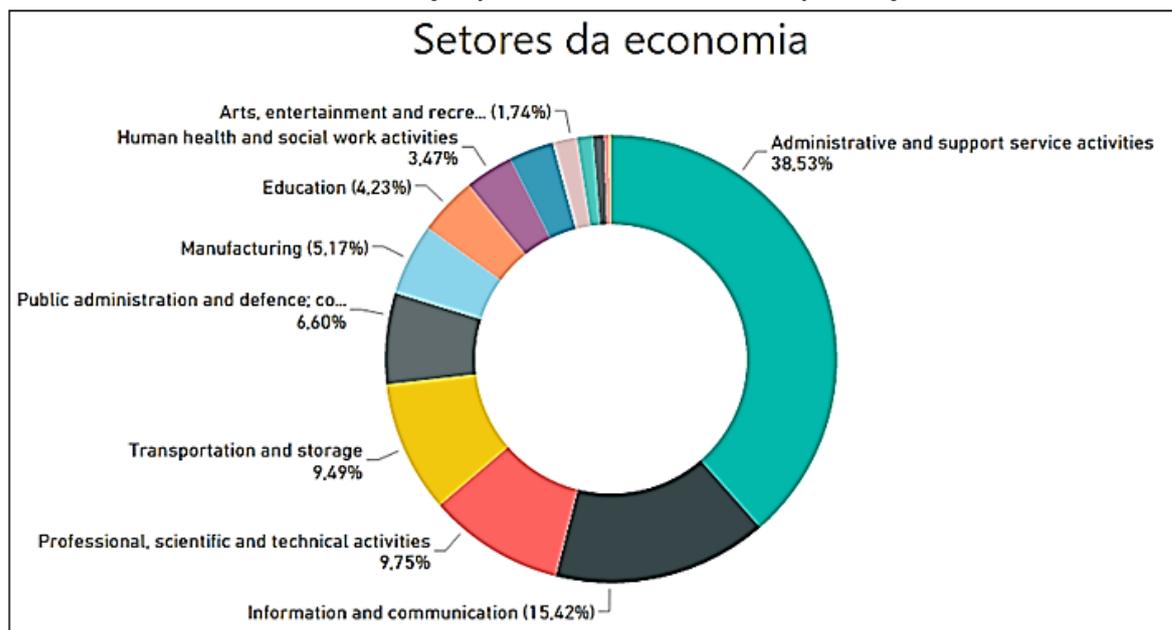
Fonte: Base de Dados Scopus utilizando Power BI.

Os dados acima mostram que a gestão dos dados e informações tem-se voltado para análises preditivas (classe de algoritmos que procuram através de padrões de dados passados prever o futuro), inteligência coletiva (usa inputs de grandes grupos na tentativa de criar comportamento inteligente assemelhado) e, ainda, aprendizado da máquina (programas que aprendem com a experiência e refina seus algoritmos baseado em novas informações). Todo este contexto, segundo Waller e Fawcett (2013 apud ISASI et al, 2015) de desenvolvimento da

ciência dos dados, análises preditivas e *big data* são partes emergentes de uma nova área competitiva que irá transformar a maneira como os negócios são gerenciados e desenhados.

Observa-se também que os temas abordados nos trabalhos tangenciam outros campos da ciência na análise das palavras chaves dos autores. Este resultado corrobora Saracevic (1996, 1999, 2009), que caracteriza o campo da Ciência da Informação (CI) como ciência interdisciplinar com marcante influência da Recuperação da Informação (RI) e da conexão com a Ciência da Computação. Além destas, esse autor destaca fortes influências das áreas de cognição (representada também pela Inteligência Artificial) na CI. Outra análise que fortalece a multidisciplinaridade é o agrupamento das publicações por setores da economia reflete a permeabilidade do tema em vários os setores, mesmo com filtro de restrição na busca no Scopus pela subárea Business, Management e Accounting. O gráfico 4 mostra esta estratificação.

Gráfico 4 – Estratificação por setor da economia das publicações selecionadas.



Fonte: Base de Dados Scopus utilizando Power BI.

Avaliando o agrupamento de palavras chaves dos autores nos países de maior expressão, percebe-se que a Austrália tem dirigido suas pesquisas para o setor social e público, enquanto os Estados Unidos e Reino Unido focam em pesquisas ligadas ao processo de tomada de decisão e à gestão do conhecimento. Na China, os três temas de maior interesse se equiparam em termos percentuais e na Alemanha percebe-se maior interesse nas análises preditivas e metodologias analíticas (figura 3).

Figura 3 – Visão de uso de palavras chaves por países em número de ocorrências.



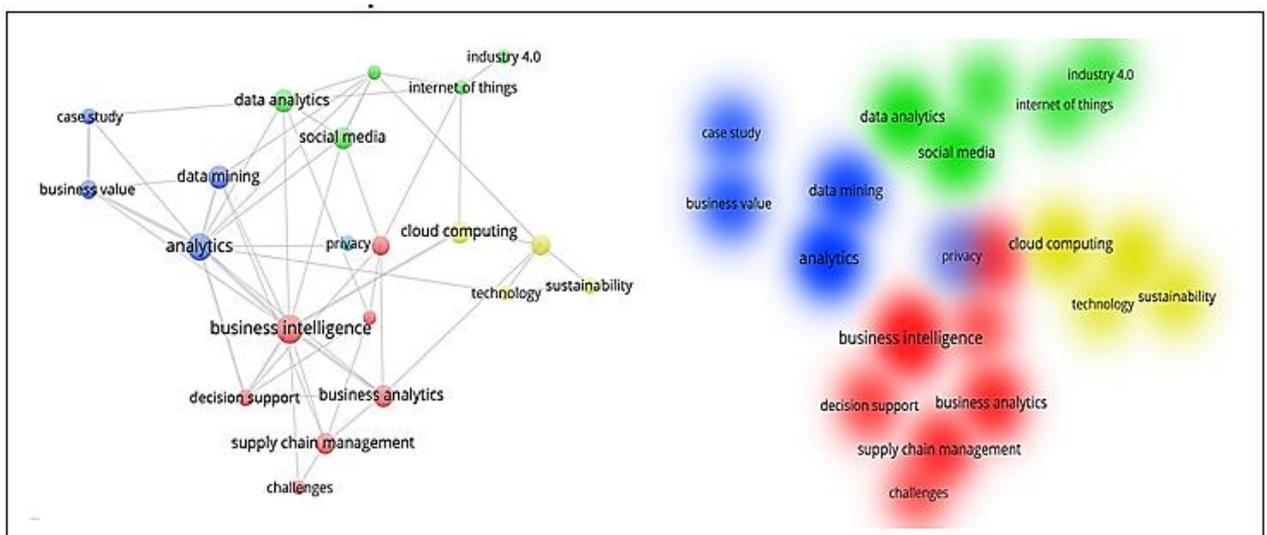
Fonte: Base de Dados Scopus utilizando Power BI.

O software VOSViewer , que é uma ferramenta para a construção e visualização de redes bibliométricas (JAN VAN ECK; WALTMAN, 2010), foi utilizado para construção de duas análises bibliométricas de palavras chaves considerando as 362 publicações recuperadas da base Scopus: (1) uma de co-ocorrência de palavras chaves dos autores (excetuada a palavra *Big Data Analytics*) do arquivo exportado do Scopus produzindo um mapa de visualização de rede, mostrando a força de ligação entre os termos, e um de densidade, mostrando a distinção dos clusters (figura 4); (2) uma de co-ocorrência de palavras chaves dos autores e dos indexes (excetuada a palavra *Big Data Analytics*) do arquivo exportado do Scopus produzindo um mapa de visualização de rede, mostrando a força de ligação entre os termos, e um de densidade, mostrando a distinção dos clusters (figura 5).

O processo de criação do mapa na primeira análise agrupou um número mínimo de 5 ocorrência de cada palavra num universo de 1.177 palavras chaves das publicações selecionadas. Para cada uma das 20 palavras-chave, a força total dos links de co-ocorrência com outras palavras-chave foi calculado e as palavras-chave com maior força de link total foram selecionadas e apresentadas na figura 4. Identifica-se a presença de 4 clusters principais, distintos por diferentes cores: em vermelho com aspectos de *business intelligence*, *business analytics*, tomada de decisão e supply chain, em azul com *data mining*, *business value* e *analytics*, em verde com mídia social, internet das coisas e *industry 4.0*, e o menor deles, em amarelo, incluindo *cloud computing*, tecnologia e sustentabilidade.

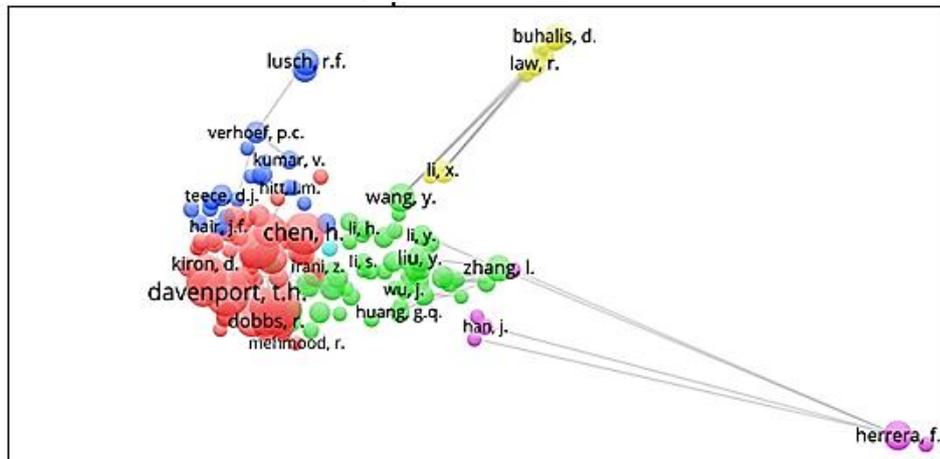
Na criação do mapa com todas as palavras chaves, ou seja, incluindo, além das palavras chaves definidas pelos autores das publicações, as palavras chaves do index, tem-se um resultado mais representativo. Este mapa agrupou um número mínimo de 5 ocorrência de cada palavra num universo de 2.389 palavras chaves das publicações selecionadas. Para cada uma das 84 palavras-chave, a força total dos links de co-ocorrência com outras palavras-chave foi calculado e as palavras-chave com maior força de link total foram selecionadas e apresentadas na figura 6. Observa-se a aparição de um novo cluster e também das palavras.

Figura 4 – Mapa visualização de rede e de densidade (clusters) criado no VOSViewer com palavras chaves dos autores da amostra.



Fonte: Base de Dados Scopus utilizando VOSViewer.

Figura 7 – Mapa visualização de rede criado no VOSViewer mostrando a força de ligação das co-citação de autores da amostra.



Fonte: Base de Dados Scopus utilizando VOSViewer.

Já a análise de citação dos autores das publicações (e não a co-citação) desta mesma amostra mostrou que as descobertas pouco se conectam, uma vez que as citações são baixas entre os autores desta amostra analisada, como mostra o mapa da figura 8.

Figura 8 – Mapa visualização de rede criado no VOSViewer mostrando a força de ligação das citações entre os autores da amostra.



Fonte: Base de Dados Scopus utilizando VOS Viewer.

5 CONSIDERAÇÕES FINAIS

As análises das publicações deste estudo específico sobre Big Data Analytics mostram que, embora o número de publicações tenha crescido bastante nos últimos anos, ainda está concentração em poucos países e em poucos periódicos. Os países com maior expressão nesta área de pesquisa (Estados Unidos, Reino Unido, China, Alemanha e Austrália) são também os países que apresentam os mais altos valores de Produto Interno Bruto (14 primeiros lugares no ranking do World Bank (2017)). O Brasil apresenta um resultado contraditório, uma vez que está entre estes países de maior PIB e, no entanto, representa pouco mais de 1% das pesquisas deste campo, como discutido anteriormente.

Percebeu-se a inexistência de inter-relação entre as pesquisas dos países envolvidos, mostrando que os esforços das investigações não convergem, visto que não existem citações expressivas entre os pesquisadores.

Observou-se ainda que as palavras-chave com ocorrência mais frequente (*business, intelligence, decision, support, management, analytics, systems*) nas pesquisas investigadas sobre este tema, pertencente ao campo da Ciência da Informação (CI), expõem alguns aspectos da tríade: informação, pessoas e tecnologia. Aspectos estes que se mantiveram estáveis ao longo da evolução da CI (SARACEVIC, 1999) preservando sua função não só tecnológica, mas social e humana.

Contudo, os resultados confirmam que dados têm sido considerados como recurso valioso a ser explorado e deve ser gerenciado como um ativo (KHATRI, 2014). Ativo este que somente possui valor se gerar conhecimento competitivo, em forma de produto, processo ou serviço. Desta forma, este artigo contribui com mais uma perspectiva de análise para este tópico de possibilidades inovativas do Big Data Analytics. É sabido que esta foi uma análise de um ponto peculiar na tão vasta rede de possibilidades de relações desta massa crescente de dados, já com autonomia própria, que inunda tudo e a todos. Percebe-se, no entanto, a tentativa conjunta de melhor entender a força das relações que ligam os dados gerados em diversos momentos, localidades do mundo, por razões distintas e por dispositivos variados.

REFERÊNCIAS

- ACTIO, F., & KHATRI, V. - **Business analytics: Why now and what next**. Business Horizons(Kelley School of Business - Indiana University), 565-570. 2014.
- BAR-ILAN, J. -Which h-index?—A comparison of WoS, Scopus and Google Scholar. **Scientometrics**, 74(2), pp. 257-271. 2008. Disponível em: <<https://link.springer.com/article/10.1007%2Fs11192-008-0216-y>>. Acesso em: 6 jun. 2017.
- CHEN, H., & et al. Business Intelligence and Analytics: from Big Data to Big Impact. **Mis Quartely**, 36(4), pp. 1165-1188. 2012. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.362.1573&rep=rep1&type=pdf>> Acesso em: 8 mai. 2017.
- DAVENPORT, T. H. - **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação**. 5a ed. São Paulo: Futura. 2002.
- DAVENPORT, T. H., & Harris, J. G. - **Inteligência analítica nos negócios: Como usar a análise de informações para obter resultados superiores**. Rio de Janeiro: Elsevier. 2010.
- FOSHAY, N., & et al. - **A Comprehensive Diagnostic Framework for Evaluating Business Intelligence and Analytics Effectiveness**. Australasian Journal of Information Systems , v.19, pp. 37-54. 2015. Disponível em: <<http://journal.acs.org.au/index.php/ajis/article/view/1178/681>>. Acesso em: 18 jun. 2016.

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

GALBRAITH, J. R. -Organization Design challenges result in form Big Data. **Journal of Organization Design**, v.3, n.1, pp. 2-13, 2014 Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2458899>. Acesso em: 18 jun. 2016.

GANDOMI, A., & Hairder, M. - Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v.35, n. 2. pp. 137-144. 2015. Disponível em: <<https://doi.org/10.1016/j.ijinfomgt.2014.10.007>>. Acesso em: 18 jun. 2016.

HAZEN, B. T., & Boone, C. A. - Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. **International Journal of Production Economics**, v. 154, pp. 72-80. 2014. Disponível em:<<https://doi.org/10.1016/j.ijpe.2014.04.018>>. Acesso em: 18 jun. 2016.

INTEZARI, A., & Gressel, S. - Information and reformation in KM systems: big data and strategic decision-making", **Journal of Knowledge Management**, v. 21n. 1, pp. 71-91. 2017. Disponível em:<<http://www.emeraldinsight.com/doi/abs/10.1108/JKM-07-2015-0293>>. Acesso em: 25 jun. 2017.

ISASI, N. K., & et al. Big Data and Business Analytics in the Supply Chain: A Review of the Literature. **IEEE**, v. 13, n.10, pp. 3382-3391. 2015. Disponível em:<<http://ieeexplore.ieee.org/abstract/document/7387245/>>. Acesso em: 20 mai. 2016.

JAN VAN ECK, N., & Waltman, L. -Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, n. 2, pp. 523–538. 2010. Disponível em:<<https://link.springer.com/article/10.1007/s11192-009-0146-3>>. Acesso em 10 jun. 2017.

JIN, X., Wah, B. W., & et al. Significance and Challenges of Big Data Research. **Big Data Research**, v. 2, n.2, pp. 49-64. 2015. Disponível em: <<https://doi.org/10.1016/j.bdr.2015.01.006>>. Acesso em: 18 mai. 2016.

KLATT, T., & Schlaefke, M. Integrating business analytics into strategic planning for better performance. **Journal of Business Strategy**, v.32, pp. 30-39. 2011. Disponível em:<<http://www.emeraldinsight.com/doi/abs/10.1108/02756661111180113>>. Acesso em: 20 jun. 2016.

KOWALCZYK, M., & Buxmann, P. An ambidextrous perspective on business intelligence and analytics support in decision processes: Insights from multiple case study. **Decision Support Systems**, v. 80, pp. 1-3. 2015. Disponível em:<<https://pdfs.semanticscholar.org/f514/d73d883c6669dab38b0957649e391460518f.pdf>>. Acesso em: 20 jun. 2016.

KWON, O., & LEE, N. Data quality management, data usage experience and acquisition intention of big data analytics. **International Journal of Information Management**, v. 34, n.3, pp. 387-394. 2014. Disponível em:<<https://doi.org/10.1016/j.ijinfomgt.2014.02.002>>. Acesso em: 20 jun. 2016.

LANEY, D. **Gartner Group Blogs**. 2001. Disponível em: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>>. Acesso em: 6 abr. 2017.

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

LEE, S. M., & Chen, J. Q. A conceptual model for executive support systems. **Logistics Information**, v. 10, n.4, pp. 154-161. 1997. Disponível em:<<http://www.emeraldinsight.com/doi/abs/10.1108/09576059710187573>>. Acesso em: 15 mai. 2016.

LOEBBECKE, C., & PICOT, A. Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. **Journal of Strategic Information Systems**, v.24, n. 3, pp. 149-157. 2015. Disponível em:<<https://doi.org/10.1016/j.jsis.2015.08.002>>. Acesso em: 15 mai. 2016.

MWILU, O. S., Comyn-Wattiaua, I., & Prat, N. Design science research contribution to business intelligence in the cloud — A systematic literature review. **Future Generation Computer Systems**, v. 63, pp. 108–122. 2016. Disponível em:<<https://doi.org/10.1016/j.future.2015.11.014>>. Acesso em: 20 jun. 2017.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. *Perspectivas da Ciência da Informação*, v. 1, n.1, pp. 41-62. 1996.

SARACEVIC, T. Information Science. **Journal of the American Society for Information Science**, v. 50, n. 12, pp. 1051–1063. 1999.

SARACEVIC, T. Information science. In: M. N. Maack, & M. Bates, **Encyclopedia of Library and Information Science**. New York: Taylor & Francis. 2009. pp. 2570-2586.

SCHLAEFKE, M., & Silvi, R. -A framework for business analytics in performance management. **International Journal of Productivity and Performance Management**, v. 62, n.1, pp. 110-122. 2012. Disponível em:<<https://doi.org/10.1108/17410401311285327>>. Acesso em: 20 jun. 2016.

SIMÕES, M. M. Resumo Documental e Literatura: origem, desenvolvimento e consolidação. **PÁGINAS a&b**, v. 3, n. 3, pp. 15-36. 2015. Disponível em:<<http://pentaho.letras.up.pt/ojs/index.php/paginasab/article/view/663/629>>. Acesso em: 10 jul. 2017.

SOUZA, R. R., Almeida, M. B., & Baracho, R. M. Ciência da informação em transformação: Big Data, nuvens, redes sociais e Web Semântica. **Ciência da Informação**, v. 42, n. 2, pp. 159-173. 2013. Disponível em:< <https://doi.org/10.18225/ci.inf.v42i2.1379>>. Acesso em: 15 out. 2016.

WAMBA, S. F., Akter, S., & et al. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. **International Journal of Production Economics**, pp. 234-246. 2015. Disponível em: < <https://doi.org/10.1016/j.ijpe.2014.12.031>>. Acesso em: 30 jun. 2016.

World Bank, **World Development Indicators database**, pp. 1-5. April de 2017. Disponível em:<<http://databank.worldbank.org/data/download/GDP.pdf>>. Acesso em 5 de setembro de 2017.

ZIKOPOULOS, P. C., EATON, C., & et al. **Understanding Big Data: Analytics for Enterprise Class, Hadoop for Streaming Data**. Mac Graw Hill. 2011.