

## **BIG DATA E INOVAÇÃO EM NEGÓCIOS: A GERAÇÃO DE VALOR E INOVAÇÃO NOS NEGÓCIOS POR MEIO DE BIG DATA**

**MARCOS ROGÉRIO MAZIERI**

UNINOVE – Universidade Nove de Julho  
m\_mazzieri@hotmail.com

**EDUARDO DANTAS SOARES**

UNINOVE – Universidade Nove de Julho  
edudantasp@hotmail.com

**LUC QUONIAN**

USP - Universidade de São Paulo  
quonian@usp.br

**GUSTAVO VIEGAS RODRIGUES**

UNINOVE – Universidade Nove de Julho  
gusviegas@yahoo.com

**LEONARDO VILS**

UNINOVE – Universidade Nove de Julho  
leovils@me.com

Área Temática: Gestão da Inovação – Organização e  
Processos para Inovação

**Título: Big Data e Inovação em Negócios: A geração de  
valor e inovação nos negócios por meio de Big Data**

**Resumo**

O artigo empírico, descritivo, indutivo usou de métodos mistos como a bibliometria, análises quantitativas, qualitativas e de conteúdo para propor uma nova definição conceitual para Big Data, além de apresentar caminhos para gerar valor e inovação por meio da análise de Big Data, encontrando que, processos, características de modelos de negócios e arranjos organizacionais podem ser os pontos mais relevantes a serem investigados dentro do Big Data, com maiores chances de gerar valor e inovações.

Palavra chave: Big Data, Inovação, Geração de Valor

**Abstract**

The empirical, descriptive, inductive article used mixed methods as bibliometrics, quantitative, qualitative analysis and content to propose a new conceptual definition for Big Data and presents ways to generate value and innovation through Big Data analysis and found that, processes, business models of features and organizational arrangements may be the most relevant points to be investigated within the big data, most likely to create value and innovation.

Keyword: Big Data, Innovation, Value

## 1.Introdução

O aumento significativo da capacidade de criação e armazenagem de dados, propiciado pelo início da era digital no ano 2000, proporcionou as primeiras especulações sobre um fenômeno conhecido por Big Data (Hilbert & López, 2011). A descrição dos elementos que compõe os Big Data, são comuns na literatura e normalmente estão explicados por ativos formados pela informação. Caracterizam-se por apresentar um grande volume de dados que os formam e a velocidade com que são formados e a variedade que adquirem exigem tecnologias específicas, além de métodos analíticos próprios para a sua transformação em valor (Andrea De Mauro, 2014). A composição de ativos formados por informações de alto volume, velocidade e variedade também é parte da origem do termo Big Data, em 2001 (Laney, 2001). Após isso, surgiram algumas outras definições do termo (ver Beyer, 2011; Hilbert & López, 2011; Taurion, 2013; Buhl, 2013). Porém, verificamos que parece haver carência de uma definição mais robusta, que possa propiciar melhor entendimento do conceito, o que abre a possibilidade de classificar, erroneamente, qualquer grande banco de dados como Big Data. Tal imprecisão pode representar problemas no reconhecimento do valor desses dados e, conseqüentemente, prejudicar a criação de valor e inovação para os negócios, argumentação que sustentamos e que poderá ser lida na seção “discussão”, nesse artigo. Quando observamos os trabalhos que descreveram o fenômeno Big Data, especificamente na literatura acadêmica, identificamos que esses trabalhos são construídos de forma a favorecer a descrição dos elementos componentes do Big Data. Os trabalhos normalmente atêm-se, portanto, à sua descrição estrutural, aos métodos que envolvem o tratamento do fenômeno, à descrição de seu potencial, descrição arquitetural e funcional. Não identificamos, contudo, uma definição robusta, que permita circunscrever, com certa assertividade, o fenômeno.

Neste estudo empírico, também descritivo e indutivo, além de predominantemente quantitativo, sustenta-se uma nova proposta para a definição de Big Data, com base num estudo bibliométrico e análise de conteúdo, por meio dos quais, pretendemos propor algumas discussões teóricas a respeito da geração de valor e inovação para negócios usando a garimpagem nos Big Data. Concluímos que Big Data são dados heterogêneos, não agrupados, que apresentam dinâmica incremental quanto a sua quantidade, variedade e origem e cujo crescimento torna-se mais rápido à medida em que o tempo passa.

Para Buhl et al. (2013), Big Data pode ser apenas um modismo, ou uma campanha publicitária em curso para estimular o mercado dos fornecedores de tecnologia, pois analisar dados é algo realizado desde a década de 1960, não denotando novidade alguma. O autor destaca ainda a existência de leis nacionais, que protegem a privacidade dos consumidores, argumentando que esse aspecto não tem sido discutido com a profundidade necessária. Suas afirmações são de que não se trata de algo novo e, ainda assim, o potencial esperado não poderá ser automaticamente alcançado. Buhl et al. (2013), ratifica que, para explorar os Big Data, devem ser considerados os desafios do volume, velocidade, variedade, veracidade e privacidade. Esse último desafio, a privacidade, foi apresentado como o desafio que pode inviabilizar a realização do potencial esperado para os Big Data, especialmente dificultando a integração entre os Big Data e os modelos de negócios. Ainda que concordemos com as colocações de Buhl (2013), nosso artigo argumenta que as conclusões desse autor consideraram apenas o nível de análise operacional que, mesmo sendo importante, pode ser reinterpretado apropriar-se de maior aprofundamento conceitual, como o proposto neste artigo.

A seguir, apresentamos a revisão da literatura, fundamentada em estudo bibliométrico. O próximo capítulo trata da descrição dos procedimentos metodológicos, seguido pelos resultados alcançados, especialmente focados na análise de conteúdo dos artigos avaliados e dos experimentos propostos. Por fim, passamos para a discussão dos resultados, tanto da análise

de conteúdo, quanto do experimento, que fundamentaram a definição conceitual do Big Data e forneceram visões sobre possíveis estudos futuros na conclusão.

## 2. Bibliometria Big Data

O fato de estarmos buscando uma definição conceitual dos Big Data, além de definições operacionais ou mesmo ferramentais, nos motivou a realizar uma bibliometria prévia à revisão da literatura. Conforme nos ensinou Buhl et al. (2013), o que se tem visto na literatura está mais direcionada a aplicação gerencial, o que explica, inclusive, o objetivo do trabalho de Buhl et al. (2013). O artigo de Buhl estimula os pesquisadores a pensar se Big Data é apenas um modismo, ou algo que existe há muito tempo, mas agora foi rebatizado, com novo formato para vender hardware e software. Nós sustentamos que não se trata de modismo, ou seja, na nossa pesquisa evidenciamos que Big Data não é algo que em breve não será mais visto no cenário da inovação mundial.

Reservamos essa seção para dar indicações de como fizemos a bibliometria, quais os softwares usados para automatizá-la, alguns detalhes da configuração desses softwares e os principais achados, que explicam porque algumas correntes de pensamento entendem o Big Data como um modismo ou ação mercadológica. A seguir passaremos a explicar os procedimentos metodológicos especificamente bibliométricos e a apresentação de achados com comentários que julgamos pertinentes.

### 2.1. Bases de dados de textos sobre Big Data

Existem, atualmente, pelo menos 20 importantes bases de dados científicas e escolhemos para este estudo o exame de uma base que não requer cadastro prévio, login ou custos para seu acesso. Trata-se do Google Scholar, pela grande popularidade que essa base de dados vem ganhando ao longo dos últimos cinco anos e também pela abrangência desse repositório de documentos. A acessibilidade da base, considerando a possibilidade das buscas descritas serem realizadas por outros interessados, também foi fator importante para nossa decisão. Tecnicamente, trata-se de uma base de dados que rastreia os web sites e identifica os textos existentes e suas principais características como título, autores e citações, que não tem viés, a priori, quanto ao tipo de resultado que irá retornar para as buscas realizadas.

Usando a palavra chave “Big Data”, direcionando as buscas para o título ou o resumo, extraímos 1000 (mil) publicações, devido a limitações do próprio sistema de buscas do Google Scholar. A extração dos dados foi realizada usando o software “publish or perish”, um software livre que extrai publicações da base Google Scholar. Os resultados da busca foram organizados a partir dos mais artigos recentes e com mais citações, sendo os 10 primeiros classificados, apresentados na tabela 1.

Tabela 1: Publicações indexadas por ano de publicações e maior quantidade de citações

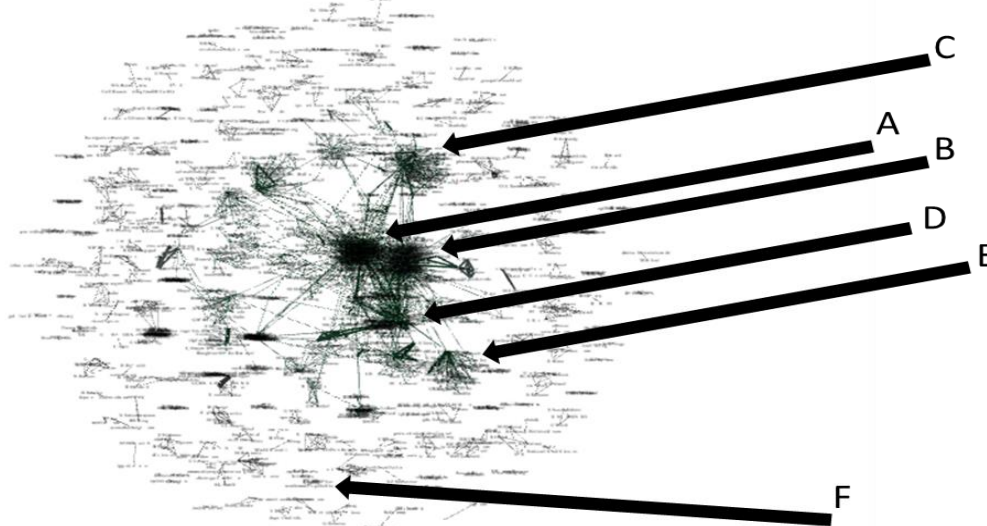
Citações	Autores	Títulos	Ano	Editora	Área
84	N Marz, J Warren	Big Data: Principles and best practices of scalable realtime data systems	2015	dl.acm.org	GR
8	V López, S del Río, JM Benitez, F Herrera	Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data	2015	Elsevier	GR
5	F Fischer, J Fuchs, F Mansmann...	BANKSAFE: Visual analytics for big data in large-scale computer networks	2015	ivi.sagepub.com	GR
134	DM Lazer, R Kennedy, G King, A Vespignani	The parable of Google Flu: traps in big data analysis	2014	dash.harvard.edu	
104	X Wu, X Zhu, GQ Wu, W Ding	Data mining with big data	2014	ieeexplore.ieee.org	GR
53	HR Varian	Big data: New tricks for econometrics	2014	JSTOR	
51	CLP Chen, CY Zhang	Data-intensive applications, challenges, techniques and technologies: A survey on Big Data	2014	Elsevier	
44	M Chen, S Mao, Y Liu	Big data: A survey	2014	Springer	
40	R Kitchin	The real-time city? Big data and smart urbanism	2014	Springer	GR
37	L Wang, J Zhan, C Luo, Y Zhu, Q Yang...	Bigdatabench: a big data benchmark suite from internet services	2014	ieeexplore.ieee.org	GR

Fonte: Autores, baseado no Google Scholar, 2015

Das dez publicações mais recentes sobre o tema Big Data, seis são dedicadas a aplicações gerenciais. Das três publicações de 2015 encontradas, todas são dedicadas a área gerencial. Tal achado, nos leva a depreender que o tema está mais ligado ao universo mercadológico do que acadêmico. Da mesma forma, a predominância de publicações voltadas a aplicação gerencial, especialmente na área de tecnologia da informação, traz ao tema caráter altamente prático e aplicado, reduzindo a importância do desenvolvimento de definições conceituais, como a que se propõe neste artigo. O problema dessa aparente superficialidade é a dificuldade na apropriação do conceito para fins de gerar valor e inovação para os negócios, o que suscita dúvidas quanto ao potencial dos Big Data.

Após extrair as publicações, optamos em mostrar tais informações em forma de rede, para conseguir mostrar a composição do universo pesquisado e realizar algumas observações importantes, por meio do destaque de algumas partições, o que está apresentado na figura 2.

Figura 1: Rede de editoras e autores que tratam de Big Data



Os pontos escuros são as editoras, formando nuvens a partir dos autores que publicam por meio delas. Podemos observar que existem duas grandes concentrações A e B e três concentrações menos intensas C, D e E. Podemos observar ainda que as publicações na área são mais distribuídas do que concentradas (Exemplo F) demonstrando que o tema Big Data ainda não está consolidado, e isso explica o porque de haver apenas definições operacionais e modelos de aplicação.

Fonte: Autores, 2015

O núcleo “A” é a Organização IEEE ([ieeexplore.ieee.org](http://ieeexplore.ieee.org)), órgão regulamentador de padrões de engenharia eletrônica e elétrica americano, o núcleo “B” é o ACM ([dl.acm.org](http://dl.acm.org)), órgão responsável por regulamentação eletrônica na Europa, o núcleo “C” é a editora Elsevier, o núcleo “D” é a editora Springer, o núcleo “E” é o website [online.libertpub.com](http://online.libertpub.com).

Dentro dos núcleos de concentração de autores ligados ao tema Big Data com objetivo científico, podemos identificar a editora Elsevier e a editora Springer. Ao verificarmos a tabela 1, podemos identificar que os 4 núcleos de concentração de edição, nesse caso as editoras, estão presentes, quando o critério é recência da publicação versus a maior quantidade de citações.

### 3.Revisão da Literatura

O aumento significativo da capacidade de criação e armazenagem de dados, propiciado pelo início da era digital no ano 2000, foi um dos grandes impulsionadores da criação do termo

Big Data (Hilbert & López, 2011). Atualmente a definição do termo Big Data entende que este termo representa os ativos de informação, caracterizado por um elevado volume, em que a velocidade e a variedade devem exigir uma tecnologia específica e métodos analíticos para a sua transformação em valor (Andrea De Mauro, 2014). O termo surgiu inicialmente no relatório de 2001 do Grupo META (empresa Gartner), de autoria do analista Doug Laney, que definiu os desafios de crescimento de dados e oportunidades como sendo tridimensional, ou seja, o aumento do volume (quantidade de dados), velocidade (velocidade de dados para dentro e para fora), e as variedades (gama ou tipos de dados e as fontes de dados), (Laney, 2001). Ainda que o relatório tenha demonstrado as dimensões que caracterizam os Big Data, os pesquisadores continuaram a procurar uma definição que melhor explique esse fenômeno. Contudo, uma das definições mais usadas para Big Data é justamente a definida pelo Gartner Group, que o entende como ativos formados por informações de alto volume, velocidade e variedade. Essas características exigem debates a respeito do custo-benefício envolvido na sua garimpagem para fins de geração de valor e inovação nos diversos negócios. As formas inovadoras de processamento de informações para maior visibilidade e tomada de decisão também tornaram-se preocupações dos pesquisadores do tema conforme Beyer (2011). A definição do IDC – Internacional Data Corporation é também bastante usada, no entanto, apresenta-se de forma mais instrumental e operacional do que conceitual: “...as tecnologias de Big Data descrevem uma nova geração de tecnologias e arquiteturas projetadas para extrair economicamente o valor de volumes muito grandes e de uma variedade de dados, permitindo alta velocidade de captura, descoberta, e/ou análise” (Hilbert & López, 2011). Tal conceito, foi ratificado no trabalho de Taurion, chamado Big Data (Taurion, 2013).

O consenso entre os pesquisadores do tema, é que o uso do Big Data se volta para sua análise, porém essa análise é muito mais poderosa do que as análises do passado (McAfee & Brynjolfsson, 2012). Esses autores explicam que hoje a informação útil para as organizações pode vir de redes sociais, imagens, sensores, web ou outras fontes não estruturadas. Isso tende a aumentar significativamente os desafios gerenciais, obrigando os tomadores de decisão a aprenderem a fazer as perguntas certas para tomarem sua decisão baseada em evidências. Para conseguir analisar e extrair esses dados as organizações precisam contratar cientistas que podem encontrar padrões em grandes conjuntos de dados e traduzi-los em informações úteis para seus negócios (McAfee & Brynjolfsson, 2012). Para abordar esse tema em específico, criamos o subtítulo de Big Data e geração de valor, que vem a seguir

### **3.1. Big Data e geração de valor**

O Big Data pode contribuir grandemente para o futuro das corporações, que precisam decidir seus processos a partir de informações de qualidade num menor tempo possível, garantindo competitividade no âmbito que estão inseridas (Volpato, Rufino, & Dias, 2014).

A análise de dados das bases de Big Data refere-se diretamente às teorias sobre BI (*Business Intelligence*) & *Analytics*, uma vez que suas tecnologias são fundamentadas principalmente em mineração de dados e análise estatística. A maioria das técnicas de busca de dados profundos conta com tecnologias comerciais para realizar este procedimento (Chaudhuri, Dayal, & Narasayya, 2011). Desde o final da década de 1980, vários algoritmos de mineração de dados têm sido desenvolvidos por pesquisadores de inteligência artificial, organizados em comunidades que debatem e desenvolvem soluções para bancos de dados. No IEEE 2006 Conferência Internacional sobre Mineração de Dados (ICDM), a maioria dos algoritmos de mineração de dados influentes foram identificados com base em nomeações de peritos, contagens de citações e em comunidade de pesquisa. Estes algoritmos procuram cobrir a classificação, agrupamentos, regressão, análise de associação e análise de rede. A maioria desses algoritmos para minerações de dados populares foram incorporadas no meio comercial

e com código aberto sobre os sistemas de mineração de dados, democratizando de certa forma o acesso (Witten et al. 2011).

*Business intelligence e analytics* (BI & A) e os respectivos campos de grandes análises de dados tornaram-se cada vez mais importante em ambas as comunidades, a acadêmica e a de negócios, ao longo das últimas duas décadas. Por exemplo, com base em uma pesquisa com mais de 4.000 profissionais de tecnologia da informação (TI), em 93 países, de 25 indústrias diferentes, a IBM Tech Trends Report identificou em 2011 a análise de negócios como um dos quatro principais tendências tecnológicas na década de 2010 (Chen, Chiang, & Storey, 2012). Em um levantamento do estado da análise de negócios, feito pela Bloomberg Businessweek, também no ano de 2011, descobriu-se que 97% das empresas com receitas superiores a 100 milhões dólares foram identificadas como usuárias de alguma forma de análise em profundidade de negócios (Chen et al., 2012). O relatório do McKinsey Global Institute (Manyika et al., 2013) previu que até 2018 os Estados Unidos terão de enfrentar uma mudança de 140 mil para 190 mil pessoas com habilidades analíticas profundas, porém haverá ainda um déficit de 1,5 milhões de gestores de base de dados com a habilidade de analisar Big Data para tomar decisões eficazes. Chen et al. (2011) alerta que Hal Varian, economista-chefe da Google e emérito professor da Universidade da Califórnia, em Berkeley, comentou em 2011 sobre as oportunidades emergentes para os profissionais de TI e estudantes em análise de dados.

As oportunidades associadas com os dados e as análises em organizações diferentes têm ajudado a gerar interesse significativo em BI & A, que é muitas vezes referida como as técnicas, tecnologias, sistemas, práticas, metodologias e aplicações usadas para analisar os dados críticos de negócios para ajudar uma empresa a compreender melhor o seu negócio e o seu mercado, fazendo negócios com agilidade e tomando decisões mais consistentes. Além do processamento de dados subjacente e tecnologias analíticas, BI & A inclui práticas de *business-centric* e metodologias que podem ser aplicadas às várias aplicações de alto impacto, como e-commerce, mercado de inteligência, *e-government*, saúde e segurança (Chen et al. 2011). Por meio das iniciativas do BI & A 1.0, as empresas e organizações de todos os setores começaram a ganhar uma visão crítica dos dados estruturados recolhidos em vários sistemas corporativos e analisados pela administração de banco de dados relacionados aos sistemas comerciais. Ao longo dos últimos anos, a inteligência da web, web analytics, web 2.0 e a capacidade de conteúdo gerada pelo usuário minerador de dados não estruturados marcaram o início de uma nova era de BI & A. Esse fato gerou a pesquisa 2.0, levando a inteligência à níveis sem precedentes, envolvendo a opinião do consumidor, as necessidades dos clientes e reconhecendo novas oportunidades de negócios.

Agora, nessa época de análise de Big Data, mesmo enquanto o BI & A 2.0 ainda está amadurecendo, encontramos-nos equilibrados à beira da BI & A 3.0, com toda a incerteza advinda das tecnologias novas, potencialmente revolucionárias e com todas as vantagens e problemas que elas podem nos trazer.

A investigação destes dados, até então invisíveis aos usuários comuns, se destina a servir, em parte, como uma plataforma e guia de conversação para examinar como a disciplina analítica pode melhor atender às necessidades dos tomadores de decisões de negócios, tendo em conta a maturação das tecnologias emergentes de BI & A, a onipresença de Big Data e as falhas provenientes de gerentes de dados, mesmo quando experientes e da falta de profissionais com habilidades analíticas profundas, que apresentam racionalidade limitada.

Alguns exemplos sobre como a análise de Big Data também pode ser aplicada, são explicitados no artigo de Melo et al. (2014) chamado *Is Big Data the Next Big Thing in Performance Measurement Systems?* Segundo os autores, essa análise pode ser aplicada em processos de monitorização contínua para detectar situações como alterações no sentimento do consumidor; para explorar as relações de rede, como amigos sugeridos no *LinkedIn* e *Facebook*; para identificar uma fraude em tempo real; para entender por que as taxas de defeitos das

montadoras aumentaram repentinamente; para digitalizar continuamente e intervir nas práticas dos cuidados de saúde; e para melhor antecipar as vendas on-line com base em um conjunto de dados de características de um determinado produto (Demirkan & Delen, 2013). Tais percepções parecem indicar uma relação interessante entre Big Data e PMSS Tien (2013) propõe uma abordagem diferente para a análise de Big Data, considerada a base para a personalização, podendo gerar a terceira revolução industrial, de acordo com o autor. Essa revolução seria baseada na confluência de grandes análises de dados, serviços de adaptação e fabricação digital, centralizando-se na integração de serviços e / ou produtos e estaria começando na segunda década do século 21.

Os investimentos em grandes dados apontam para essa realidade, haja visto os investimentos nessa área. Em 2012, a Universidade de Boston recebeu US \$ 15 milhões para criar o Instituto Rafik Hariri B. de Informática e Ciência da Computação e Engenharia, um centro de pesquisa interdisciplinar para descobertas através do uso de abordagens computacionais e orientadas a dados e para avanços na ciência da computação (Chen, 2011).

O Big Data poderia ser uma alternativa para as falhas do PMSS tradicional, pois somente quando definimos os dados, informações e as análises é que podemos ver que os mecanismos de medição tradicionais não funcionam de forma eficiente. As organizações podem se preocupar com o serviço de análise e sua precisão, qualidade, além do custo e tempo de entrega dessas análises. Por isso, as organizações precisam considerar o valor de precisão e qualidade de serviço para além do custo e duração dos serviços prestados (Demirkan & Delen, 2013).

As organizações precisam fazer mudanças em cinco áreas para colherem os benefícios reais de Big Data. Eles precisam mudar a liderança, que deve definir metas claras e definir o que se parece com o sucesso buscado pela organização, pois os dados são muito importantes, mas o *insight* humano é fundamental. Gestão de talentos também é importante, pois é imperativo para contratar as pessoas certas. É necessário investir em novas tecnologias, incluindo o software apropriado. Finalmente, a cultura da empresa deve reforçar o uso de dados na tomada de decisões mais do que os próprios gerentes (Waller & Fawcett, 2013).

A exploração adequada de *Big Data* pode ajudar a melhorar o desempenho organizacional, mas primeiro é necessário transformar a cultura corporativa das organizações e suas capacidades. Deve haver um esforço para tecer grandes dados no tecido das operações diárias (Waller & Fawcett, 2013).

Alguns benefícios do uso dos grandes dados são: melhor integração e análise de dados quantitativos e qualitativos; previsões mais precisas (Waller & Fawcett, 2013); e os processos de tomada de decisões mais eficientes e eficazes, que fazem as empresas mais ágil e eficiente (Demirkan & Delen, 2013). As empresas que estão na parte superior da sua indústria e usam o recurso de tomada de decisões com base em dados foram, em média, 5% mais produtivas e 6% mais rentáveis do que seus concorrentes que não utilizam esse recurso (McAfee & Brynjolfsson, 2012).

### **3.2. Big Data, Informação, Inovação e Valor**

Dados são símbolos, que representam determinadas circunstâncias, fenômenos, fatos, sejam eles físicos, químicos ou comportamentais. Ao atribuir determinada ordem e códigos a esses símbolos, tem-se a informação. Não se discute aqui o aspecto apenas formal da informação, mas sim o sentido mais amplo, ou seja, coletar, organizar, analisar e interpretar as informações. De acordo com Bregonje (2005), quando a informação é aplicada, ela transforma-se em conhecimento. Pode-se então fazer a analogia de que Informação aplicada as ações de solução de um problema produzem a descoberta ou a invenção, esses por sua vez, antecedentes da Inovação. A partir dessas evidências, pode-se inferir que a informação pode moderar as relações entre pesquisa básica e uma descoberta científica ou invenção, como pode moderar as



relações entre pesquisa aplicada e uma invenção e por fim moderar relações entre uma invenção e uma inovação.

Considerando os desafios para o tratamento do Big Data, podemos entender que os processos envolvidos em seu armazenamento e tratamento, também conhecidos pelo termo em inglês, *Analytics*, apresenta potencial para contribuir com a inovação, pela própria definição de Inovação. Inovação é a disponibilização de algo substancialmente novo em diferentes graus de novidade ou tipo que gera valor econômico ou financeiro (OECD & Eurostat, 2005). Quando Bregonje (2005) argumenta que a informação aplicada se transforma em conhecimento, podemos entender que o Big Data armazena dados que, quando codificados, transformam-se em informações e que essas informações, quando aplicadas, transformam-se em conhecimento. O conhecimento é o elemento que agrega valor, ou seja, que promove a proposta de valor nos modelos de negócio (Osterwalder, 2009; Enzmann & Schomer, 2013; Flor & Maglio, 2004). Considerando que algo só pode ser considerado inovação quando gera valor econômico ou financeiro e que o conhecimento é o elemento que promove a proposta de valor nos negócios, podemos compreender que há proximidade entre o elemento conhecimento e o valor para o negócio. A inovação, está nesse caso, intrínseca ao processo de construção do conhecimento, bem como presente no próprio conhecimento.

Num raciocínio inverso, pensar em conhecimento como sendo a aplicação de informações e essas informações provenientes da codificação dos dados, nos remete a compreender a conexão entre o Big Data a Inovação e o Valor para os negócios. Obviamente, precisamos considerar apenas o encadeamento teórico como proposição, guardadas todas as dificuldades, obstáculos e desafios, especialmente colocados por Buhl et al. (2013). Alguns dos desafios vem sendo enfrentados, tanto pelos acadêmicos, quanto pelos praticantes da inovação baseada em Big Data. Uma parte dos casos serão abordados nas seções seguintes como por exemplo a estrutura dos dados, o processo de recuperação de dados (*Crawling and Data Mining*), o tratamento dos dados (KDD) e algumas formas mais comuns de análise.

### **3.3. Informações Estruturadas e Não estruturadas**

No universo dos dados, quando armazenados em determinados repositórios, há dois grandes grupos, chamados de dados estruturados e os dados não estruturados. Entende-se por dados estruturados aqueles que são esperados ou conhecidos em sua forma desde a entrada, armazenamento e saída sendo normalmente dados completos. São normalmente armazenados em bancos de dados que utilizam a linguagem SQL (Structured Query Language). Tem como principais características a previsibilidade da forma como os dados entrarão e serão armazenados pelo banco de dados, favorecendo a criação de índices de busca, o que implica em maior performance na recuperação dos dados quando necessário. O aspecto da segurança também é incrementado devido a estrutura relacional. Acrescentar um dado aleatório, seja por tentativa de fraude ou por erro sistêmico em estruturas de dados relacionais não são tarefas simples.

Os dados não estruturados, são aqueles que podem se originar em fontes diversas, em formatos diversos, cuja entrada pode ser realizada de forma completa ou incompleta. Os bancos de dados são chamados de NON-SQL, como sendo bancos de dados que não usam a linguagem estruturada SQL. São os bancos de dados que permitem o armazenamento de dados que a princípio não eram esperados, como opiniões, sentimentos declarados em redes sociais, dados oriundos de sensores e de outras máquinas. As questões de segurança, podem ser apontadas como o *trade off* a ser pensado. Uma vez que não há a estrutura relacional, a inserção de dados que não fazem parte do contexto de armazenamento podem ocorrer com maior facilidade.

### **3.4. Processos de análise de Big Data**

A literatura apresenta várias formas de acessar e aproveitar o Big Data, normalmente na literatura de inteligência competitiva., No entanto, processos de análise de Big Data são praticamente incontáveis, pois dependerá de cada banco de dados ou da associação de bancos de dados, do contexto e interesse da análise, da perenidade ou precibilidade dos dados, dos recursos econômicos, tecnológicos e humanos a disposição, descrito por Fleming et al. (2001), Machado (2004) e Wu et al. (2014), (Fleming & Sorenson, 2001; Machado, 2004; Xindong Wu, Xingquan Zhu, Gong-Qing Wu, & Wei Ding, 2014).

Há etapas ou sub processos comuns nos processos de análise de dados, descritos na literatura revisada, que serão abordados resumidamente na próxima seção.

### **3.5. *Crawling, Mining, Analyse (KDD)***

Simplificadamente podemos dizer que, para aproveitar o potencial do Big Data, precisamos extrair dados (*Crawling*), separar dados que interessam (*Data Mining*) e analisar esses dados (*Analyse*). Se chamarmos essas etapas de sub processos, a associação delas forma o processo chamado de descoberta de conhecimento em base de dados (*Knowledge Discovery in Data Base – KDD*), (Tian, Zhiping, & Zhengyin, 2013).

A etapa da extração ou recuperação de dados, consiste em identificar a maneira mais apropriada de recuperar os dados que queremos, evitando ruídos. Ruídos são dados extraídos juntamente com a massa de dados recuperada, mas que de fato não fazem parte dos dados que se buscam. O ruído não pode ser totalmente eliminado, mas deve ser minimizado e para isso, os *crawlers* são compostos por algoritmos de validação dos dados. Tais algoritmos são os que tratam de responder as expressões de busca. Expressões de busca são palavras ou termos utilizados como entrada de uma busca. Popularizou-se o uso do Google, e a maioria das pessoas, usa palavras e termos para buscar respostas no buscador do Google. As palavras e termos que digitadas no Google para obter determinada resposta, são as expressões de buscas e o *crawling* é realizado pelo próprio Google que apresenta os resultados recuperados para aquela expressão de busca na tela.

*Data Mining* ou mineração de dados é o procedimento de separação dos dados recuperados em dados úteis e inúteis para aquela expressão de busca. Nessa etapa, ainda há alguns ruídos, que os algoritmos do *crawler* não trataram e também conteúdos que atendem a expressão de busca, mas que não representam os resultados pretendidos *ex ante*. Tal fato, apresenta-se como um dos desafios apontados por Buhl (2013), a veracidade. Veracidade, nesse estudo, está relacionada com a identidade do dado recuperado, a decisão se, de fato, ele pertence ao contexto declarado pela expressão de busca. Os campos de estudo que tratam da Web Semântica têm preocupações como essa, buscando contribuir tanto no sub processo de *Data Mining*, quanto no sub processo de *Analyse*.

O sub processo de *analyse* pode ser definido como as iterações envolvendo a codificação do dado para que seja transformado em informação. Tratando-se de processo iterativo, não podemos definir, observando a literatura, sua extensão em volume de dados analisados ou sua velocidade de análise, uma vez que parece ser um sub processo não linear (Martínez-Román, Gamero, & Tamayo, 2011).

As iterações necessárias, sendo não lineares estão mais relacionadas as teorias de sistemas complexos em parte. O fato de partir da complexidade para definição de modelos simplificados com graus significantes de explicação do todo, para após isso, envolver novamente a complexidade, iterando nesse sentido, nos remete a pensar em um processo em forma de espiral (Fleming & Sorenson, 2001; Prior, 2013). O espiral, mostra que a cada iteração, ainda que passe na mesma posição horizontal, há a mudança na posição vertical e vice-versa. Outra forma proposta na literatura para conceber modelos de explicação de sistemas

complexos é em forma de rede. Redes podem apresentar formas não lineares e representar alguma dimensão da complexidade que se pretende compreender. Numa rede, a direção e a intensidade com que os elementos ou nós (nodes) se constituem não são previsíveis (Yugue & Maximiano, 2013).

Dessa forma, analisar grandes volumes de dados, de forma não linear, não estruturados ou relacionados, com processos iterativos, evidenciam os outros desafios propostos por Buhl (2013), quanto ao volume, a velocidade e a variedade.

Conforme sustentado até aqui, o sub processo de análise de dados é o que os codificará, transformando-os em informações. Para tanto, métodos usados na pesquisa científica, especificamente os qualitativos e quantitativos podem contribuir com esse sub processo do KDD e, portanto, serão apresentados na próxima seção, no contexto da análise de Big Data.

#### **4.Procedimentos Metodológicos**

Por tratar-se de uma nova proposta para a definição dos Big Data, foram escolhidos métodos mistos, envolvendo a bibliometria e a análise de conteúdo. A bibliometria foi desenvolvida para levantar a literatura principal, valendo-se para tanto da base de dados Scopus e Google Scholar, cujos procedimentos foram detalhados na seção bibliometria desse artigo. A análise de conteúdo foi usada para reconhecer padrões e categorias como recomendado por Bardin (1999) e foram selecionados três dos principais artigos científicos encontrados a respeito de Big Data, publicados pelo MIT Sloan Review, proporcionando a fundamentação da nossa definição conceitual para Big Data. A análise de conteúdo que estamos propondo divide-se em análise quantitativa, envolvendo as frequências de ocorrência de palavras no conjunto de textos extraídos dos artigos científicos selecionados (*corpus*), combinada a análise fatorial confirmatória e ao método de agrupamento de Réinert (1998). A segunda parte da análise é qualitativa e foi realizada sobre os resultados alcançados da análise de conteúdo, resultando no mapeamento da ontologia encontrada sobre os Big Data. Após essa breve explicação do conteúdo dessa seção, passaremos a detalhar cada uma das etapas descritas nas sub seções a seguir.

Os textos escolhidos para realização das análises de conteúdo foram identificados na bibliometria, preferindo artigos que fossem mais novos e com maiores citações. Além disso, foram identificados os núcleos de especialização menores em termos de concentração de autores, fora das áreas de concentração, descritas na Figura 1, mas que embora publiquem de forma mais isolada, apresentam juntas 7080 citações distribuídas em 20 artigos. Esse é o caso do Sloam MIT Review, e os artigos que atenderam ais critérios estabelecidos foram *How 'big data'is diferente* de TH Davenport, P Barth, R Bean (2013) com 96 citações, *Big data, analytics and the path from insights to value* de S LaValle, E Lesser, R Shockley (2013) com 208 citações e *Strategic outsourcing: leveraging knowledge capabilities* de JB Quinn, ES Strategy (2013) com 2125 citações.

#### **4.3.Análise Quantitativa**

Dependendo da natureza da pesquisa e dos dados disponíveis, pode ser mais contributivo usar métodos quantitativos. Normalmente quando se dispõe de respostas a questionários, ou informações de forma mais estruturada, a análise quantitativa pode ser complementar ao processo de análise dos dados. Os testes mais comuns são os testes de média e hipóteses como testes paramétricos bivariados e suas derivações como as regressões. Dados multivariados são usados normalmente as análises fatoriais, análise de *cluster* e equações estruturais. A escolha do teste dependerá da natureza da expressão de busca e dos dados à

disposição, considerando sempre os tamanhos mínimos de amostras e demais parâmetros definidos pelo método quantitativo escolhido.

Tais métodos de análise, tem caráter instrumental, visando aplicar a informações para obtenção de conhecimento e de valor. As estratégias competitivas, apoiadas pela inteligência competitiva, podem receber contribuições do contexto do Big Data e tais aspectos serão apresentados na seção a seguir.

#### **4.2.Método de Réinert e a Análise Fatorial Confirmatória (AFC) em análises qualitativas**

Em toda língua existe a representação léxica, que é o conjunto de códigos usados para representar os “signos” ou significados das palavras, permitindo que haja comunicação entre as pessoas. A comunicação verbal usa os sons, no caso da fala e também usa códigos, no caso da comunicação escrita. O Big Data não precisa estar formado apenas por comunicação verbal, como sons e dados escritos, mas também com comunicações não verbais como imagens e símbolos, os quais não serão abordados nesse estudo. O foco será nas comunicações verbais escritas.

A análise de conteúdo é o método qualitativo utilizado para identificar a essência de um *corpus* de informação verbal escrita. Tal método pode ser subdividido em análise de frequência, classificação hierárquica de Réinert e Análise Fatorial Confirmatória, não exaustivamente. Para tratamento da linguagem verbal escrita, podemos usar o método de Réinert (1989), que destaca a proximidade léxica entre as palavras, produzindo um mapa de classes de palavras que compõe determinado banco de dados de textos. Esses textos podem ser entrevistas com opiniões de consumidores, reclamações, denúncias, sugestões entre outras. Analisar 30 ou 40 sugestões poderia ser realizado manualmente, mas analisar 10 mil, 100 mil ou 1 milhão de sugestões, normalmente não é possível de forma manual. Em casos como esses, aplicando o método de Réinert (1989), automatizado através de softwares, pode permitir analisar 1 milhão de textos de sugestões ou reclamações de clientes em algumas horas.

Nessa mesma linha de raciocínio, avaliando informações qualitativamente, podemos realizar a análise fatorial confirmatória (AFC) considerando a distância léxica das palavras. Percebamos que estamos usando uma técnica quantitativa AFC, mas a natureza da investigação é em aprofundar-se em dados não estruturados que pertencem a determinado contexto, sem interesse na extrapolação ou inferência. A partir da AFC, são construídos os fatores e as cargas fatoriais do conjunto de palavras relacionadas. A AFC separa as palavras pela proximidade léxica, mostrando ao analista, quais os temas presentes no banco de dados e como eles se relacionam. A interpretação de tais dados, auxiliados pelos métodos citados, dará aos dados a codificação e o significado necessário para que passe a ser chamado de informação, útil para aplicação e criação de valor.

#### **4.1.Análises Qualitativas**

Quando precisamos conhecer de forma mais aprofundada determinado tema de pesquisa, pode-se recorrer a análise qualitativa. Identificar os motivos, e as justificativas da ocorrência de determinados fenômenos, representados nos dados, podem conduzir a ideias a respeito de como tomar decisões, sobre quais produtos investir, serviços a serem desenvolvidos, processos a serem propostos ou melhorados, arranjos organizacionais e possíveis planos de marketing a construir.

O presente estudo visa apenas oferecer de forma resumida a visão geral sobre o universo de obtenção de valor a partir da análise do Big Data, portanto, detalhes sobre metodologias qualitativas e quantitativas não serão desenvolvidas, apresentando apenas as informações entendidas como relevantes para introdução de tais métodos no contexto da análise de Big Data.

#### 4.4. Estratégias de Inovação usando Big Data e a Inteligência Competitiva

Os autores Demirkan e Delen (2013) procuram orientar sobre algumas estratégias que as organizações podem usar para obter vantagem competitiva por meio do uso de Big Data. Sua indicação remete ao uso de sistemas de apoio à decisão orientada a serviços (DOS em nuvem). Segundo os autores, essa é uma das principais tendências para muitas organizações, na esperança de se tornarem mais ágeis. Para que isso ocorra, eles propõem uma estrutura conceitual para DOS. Esse, define a perspectiva sobre a forma de alinhar o ambiente DOS orientado para o produto, e demonstrar as oportunidades e desafios da engenharia orientada a serviços DOS em nuvem, como um ciclo que se retroalimenta.

Ao definir os dados, informações e análises como serviços, percebeu-se que os mecanismos tradicionais de medição, que são principalmente o tempo e o custo, não funcionam bem. Assim as organizações precisariam considerar o valor do nível de serviço e qualidade, além de o custo e a duração dos serviços prestados. O DOS em nuvem permite a elaboração de escala, alcance e velocidade, gerando economia para as organizações, amarrando as perspectivas de estratégia de tecnologia da informação com as perspectivas de banco de dados, gerando uma estratégia de inovação contínua de processos dentro das organizações (Demirkan e Delen, 2013). Pode-se ainda analisar o papel do ambiente de decisão para explicitar como os recursos de business intelligence (BI) são aproveitados como estratégia para alcançar o sucesso. Seu procedimento consistiu em analisar o ambiente de decisão em termos dos tipos de decisões tomadas e as necessidades da organização, quanto ao seu processamento de informações, apoiadas nos melhores resultados obtidos das decisões anteriores (Işık, Jones, & Sidorova, 2013). Os autores contribuem com o campo de estudos do Big Data, propondo que as capacidades tecnológicas, como a qualidade dos dados, o acesso do usuário e a integração do BI com outros sistemas são necessários para o sucesso do BI, independentemente do ambiente de decisão, porém o ambiente de decisão influenciou a relação entre o sucesso e as capacidades dinâmicas da organização, tais como a extensão em que BI suportou, a flexibilidade e o risco na tomada de decisões quando essa tornou-se de fato, uma estratégia da organização (Işık et al., 2013).

Devido ao objetivo desse artigo, que é o de propor uma definição conceitual para Big Data e ainda identificar formas de gerar valor e inovação para os negócios, as análises descritas foram usadas para identificar as categorias de palavras mais frequentes e a carga fatorial dos agrupamentos de palavras encontrados nos artigos, para poder comparar com a proposição da ontologia de modelo de negócio de Osterwalder (2009). A partir desse mecanismo, foi possível comparar os padrões presentes nos artigos científicos que tratam de Big Data e as possíveis contribuições com a ontologia de modelo de negócios de Osterwalder (2009).

#### 5. Resultados

Ao analisar os 3 textos sobre Big Data, realizou-se a análise de Réinert e a análise fatorial confirmatória em relação ao ranking de palavras classificadas. As cargas fatoriais encontradas são as descritas na tabela 2

Tabela 2: Cargas fatoriais da análise fatorial confirmatória das frequências

	Carga Fatorial	Porcentagem
Fator 1	0,35192	30,16
Fator 2	0,31543	27,03
Fator 3	0,2784	23,86
Fator 4	0,22094	18,93



informações. De acordo com Bregonje (2005), quando a informação é aplicada, ela transforma-se em conhecimento. Pode-se então fazer a analogia de que Informação aplicada as ações de solução de um problema, produzem a descoberta ou a invenção, esses por sua vez, antecedentes da Inovação. À partir das evidências de Bregonje (2005), pode-se inferir que, a informação pode moderar as relações entre pesquisa básica e uma descoberta científica ou invenção, como pode moderar as relações entre pesquisa aplicada e uma invenção e por fim moderar relações entre uma invenção e uma inovação.

Considerando os desafios para o tratamento do Big Data, podemos entender que os processos envolvidos em seu armazenamento e tratamento, também conhecidos pelo termo em inglês, *Analytics*, apresenta potencial para contribuir com a inovação, pela própria definição de Inovação. Inovação é a disponibilização de algo substancialmente novo em diferentes graus de novidade ou tipo que gera valor econômico ou financeiro (OECD & Eurostat, 2005). Quando Bregonje (2005) argumenta que a informação aplicada se transforma em conhecimento, podemos entender que o Big Data, armazena dados que quando codificados, transformam-se em informações e que essas informações, quando aplicadas, transformam-se em conhecimento.

O conhecimento é o elemento que agrega valor, ou seja, que promove a proposta de valor nos modelos de negócio (Osterwalder, 2009; Enzmann & Schomer, 2013; Flor & Maglio, 2004). Considerando que algo só pode ser considerado inovação quando gera valor econômico ou financeiro e que o conhecimento é o elemento que promove a proposta de valor nos negócios, podemos compreender que há proximidade entre o elemento conhecimento e o valor para o negócio. A inovação, está nesse caso, intrínseca ao processo de construção do conhecimento, bem como presente no próprio conhecimento. Num raciocínio inverso, pensar em conhecimento como sendo a aplicação de informações e essas informações provenientes da codificação dos dados, nos remete a compreender a conexão entre o Big Data a Inovação e o Valor para os negócios. Obviamente, precisamos considerar, apenas o encadeamento teórico como proposição, guardadas todas as dificuldades, obstáculos e desafios, especialmente colocados por Buhl et al. (2013). Alguns dos desafios vem sendo enfrentados, tanto pelos acadêmicos, quanto pelos praticantes da inovação baseada em Big Data. Uma parte dos casos serão abordados nas seções seguintes como por exemplo a estrutura dos dados, o processo de recuperação de dados (*Crawling and Data Mining*), o tratamento dos dados (KDD) e algumas formas mais comuns de análise.

Compreendemos as limitações metodológicas desse estudo, especialmente o uso de três artigos na análise de conteúdo, que poderá ser inclusive alvo de estudos futuros, avaliando se, numa amostra maior de artigos, o comportamento dos fatores e a hierarquia de palavras se mantem. Se as palavras de maior frequência e se as similitudes serão parecidas com por exemplo um *corpus* de 100 textos. Nesse caso, há que se avançar no processo que automatiza a análise para realizar com volumes de textos maiores, e produzir comparações, entre categorias de textos, envolvendo demais áreas da administração, finanças, contabilidade, recursos humanos, estratégia, marketing, não exaustivamente, comparando as similitudes inclusive entre áreas.

Foi possível evidenciar, nesse artigo, que embora haja publicações com objetivo gerencial, existem ao menos duas concentrações de autores publicando através da editora Elsevier e da Springer, notadamente publicações científicas. Evidencia-se que há muitas dúvidas sobre como os Big Data podem de fato contribuir com a geração de valor, que se explica pela fragmentação das publicações e de certa forma, pela falta de cooperação entre os autores que tratam do tema, como pudemos verificar na rede de editoras e autores apresentada na figura 1. A presença de bolsões de conhecimento, evidenciadas pela grande quantidade de citações de suas publicações, como o MIT Sloan Management Review, com 20 artigos e 7080 citações distribuídas entre eles, denota que há ainda muito conhecimento já desenvolvido a ser explorado do ponto de vista teórico a respeito da geração de valor por meio de Big Data.

Concluimos que Big Data, são dados heterogêneos, não agrupados, que apresentam dinâmica incremental quanto a sua quantidade, variedade e origem, cujo crescimento torna-se mais rápido a medida em que o tempo passa. O agrupamento, ainda que necessário para realização da garimpagem, aparenta ser o primeiro passo para que a massa de dados em questão comece a deixar de ser Big Data, uma vez que a estruturação, organização ou mesmo delimitação da massa de dados, pode torná-los menos dinâmicos em termos de movimento incremental, perdendo uma de suas características principais.

## 8.Referências

- Andrea De Mauro, M. G. (2014). What is Big Data? A Consensual Definition and a Review of Key Research Topics. <http://doi.org/10.13140/2.1.2341.5048>
- Beyer, M. (2011). Gartner says solving “big data” challenge involves more than just managing volumes of data. *Gartner. Archived from the Original on, 10*.
- Bregonje, M. (2005). Patents: A unique source for scientific technical information in chemistry related industry? *World Patent Information, 27*(4), 309–315.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM, 54*(8), 88–98.
- 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.
- 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.
- Enzmann, D. R., & Schomer, D. F. (2013). Analysis of Radiology Business Models. *Journal of the American College of Radiology, 10*(3), 175–180. <http://doi.org/10.1016/j.jacr.2012.09.001>
- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy, 30*(7), 1019–1039. [http://doi.org/10.1016/S0048-7333\(00\)00135-9](http://doi.org/10.1016/S0048-7333(00)00135-9)
- Flor, N. V., & Maglio, P. P. (2004). Modeling business representational activity online: A case study of a customer-centered business. *Knowledge-Based Systems, 17*(1), 39–56. <http://doi.org/10.1016/j.knosys.2003.08.011>
- Hilbert, M., & López, P. (2011). The World’s Technological Capacity to Store, Communicate, and Compute Information. *Science, 332*(6025), 60–65. <http://doi.org/10.1126/science.1200970>
- Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management, 50*(1), 13–23. <http://doi.org/10.1016/j.im.2012.12.001>
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*.
- Machado, M. P. (2004). A consistent estimator for the binomial distribution in the presence of “incidental parameters”: an application to patent data. *Journal of Econometrics, 119*(1), 73–98. [http://doi.org/10.1016/S0304-4076\(03\)00156-8](http://doi.org/10.1016/S0304-4076(03)00156-8)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & others. (2013). *Big data: The next frontier for innovation, competition, and productivity. 2011, McKinsey Global Institute*.



- Martínez-Román, J. A., Gamero, J., & Tamayo, J. A. (2011). Analysis of innovation in SMEs using an innovative capability-based non-linear model: A study in the province of Seville (Spain). *Technovation*, 31(9), 459–475. <http://doi.org/10.1016/j.technovation.2011.05.005>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, (90), 60–6.
- OECD, & Eurostat. (2005). *Oslo Manual*. OECD Publishing. Retrieved from [http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/oslo-manual\\_9789264013100-en#page1](http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/oslo-manual_9789264013100-en#page1)
- Prior, D. D. (2013). Supplier representative activities and customer perceived value in complex industrial solutions. *Industrial Marketing Management*, 42(8), 1192–1201. <http://doi.org/10.1016/j.indmarman.2013.03.015>
- Taurion, C. (2013). *Big Data*. Brasport.
- Technology, C. ©. M. I. of, & reserved, 1977-2015 All rights. (n.d.). Big Data, Analytics and the Path From Insights to Value. Retrieved from <http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>
- Tian, L., Zhiping, Y., & Zhengyin, H. (2013). The Large Aperture Optical Elements patent search system based on Domain Knowledge Organization System. *World Patent Information*, 35(3), 209–213. <http://doi.org/10.1016/j.wpi.2013.04.007>
- Tien, J. M. (2013). Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2), 127–151. <http://doi.org/10.1007/s11518-013-5219-4>
- Volpato, T., Rufino, R. R., & Dias, J. W. (2014). BIG DATA–TRANSFORMANDO DADOS EM DECISÕES. Retrieved from [http://web.unipar.br/~seinpar/2014/artigos/graduacao/Tiago\\_Volpato.pdf](http://web.unipar.br/~seinpar/2014/artigos/graduacao/Tiago_Volpato.pdf)
- Waller, M. A., & Fawcett, S. E. (2013). Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain. *Journal of Business Logistics*, 34(4), 249–252.
- Xindong Wu, Xingquan Zhu, Gong-Qing Wu, & Wei Ding. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97–107. <http://doi.org/10.1109/TKDE.2013.109>
- Yugue, R. T., & Maximiano, A. C. A. (2013). Understanding and Managing Project Complexity. *Revista de Gestão E Projetos - eISSN: 2236-0972*, 4(1), 01–22. <http://doi.org/10.5585/gep.v4i1.10>