

**UNIVERSIDADE DE CAXIAS DO SUL  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA  
CURSO DE MESTRADO**

**PRISCILA NESELLO**

**IMPLICAÇÕES DO FENÔMENO *BIG DATA* NA ANÁLISE PARA INTELIGÊNCIA  
ESTRATÉGICA**

**CAXIAS DO SUL  
2014**

**PRISCILA NESELLO**

**IMPLICAÇÕES DO FENÔMENO *BIG DATA* NA ANÁLISE PARA INTELIGÊNCIA  
ESTRATÉGICA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como requisito à obtenção do grau de Mestre em Administração.

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Cristina Fachinelli

**CAXIAS DO SUL  
2014**

Dados Internacionais de Catalogação na Publicação (CIP)  
Universidade de Caxias do Sul  
UCS - BICE - Processamento Técnico

N459i Nesello, Priscila

Implicações do fenômeno big data na análise para inteligência estratégica / Priscila Nesello. - 2014.

185 f. : il. ; 30 cm

Apresenta bibliografia.

Dissertação (Mestrado) – Universidade de Caxias do Sul, Programa de Pós-Graduação em Administração. 2014.

“Orientação: Profª. Drª. Ana Cristina Fachinelli”

1. Inteligência estratégica. 2. Big data. 3. Administração. I. Título.

CDU 2.ed. : **658.012.2**

Índice para o catálogo sistemático:

1. Inteligência estratégica	<b>658.012.2</b>
2. Big data	658.012.4
3. Administração	658

Catalogação na fonte elaborada pela bibliotecária  
Carolina Machado Quadros – CRB 10/2236.

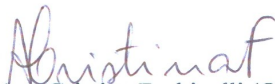
**“Implicações do Fenômeno Big Data na Análise para Inteligência Estratégica”**

Priscila Nesello

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como parte dos requisitos necessários para a obtenção do título de Mestre em Administração, Área de Concentração: Administração da Produção

Caxias do Sul, 10 de abril de 2014.

Banca Examinadora:



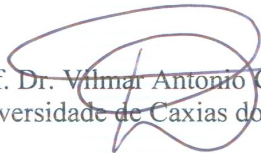
Profa. Dra. Ana Cristina Fachinelli (Orientadora)  
Universidade de Caxias do Sul



Prof. Dr. Carlos Alberto Costa  
Universidade de Caxias do Sul



Prof. Dr. Marcelo Machado Barbosa Pinto  
Universidade Federal de Santa Maria



Prof. Dr. Vilmar Antonio Gonçalves Tondolo  
Universidade de Caxias do Sul

## **DEDICATÓRIA**

Aos meus pais, Vanda e Valdemar, por me encorajarem na busca de novos desafios. Ao meu marido Luiz Marcelo Ruziska, por estar ao meu lado em todos os momentos.

## AGRADECIMENTOS

Agradeço a Professora Ana Cristina Fachinelli pela grande oportunidade que foi a mim proporcionada, em pesquisar um tema tão novo e desafiador como *big data*.

A empresa Maius Tecnologia Ltda. e equipe do PSI, por cederem o tempo e suportarem as responsabilidades rotineiras, enquanto eu estava ausente.

Ao Sr. Fabio Gandour, por dispor de sua atenção em prol de minha pesquisa, oferecendo suas contribuições e incentivo para trabalhos futuros.

Aos Professores Carlos Alberto Costa, Rolando Vargas Vallejos e Vilmar Antonio Gonçalves Tondolo, demais professores e funcionários do PPGA.

*"O valor de todo o conhecimento está no seu vínculo com as nossas necessidades, aspirações e ações; de outra forma, o conhecimento torna-se um simples lastro de memória, capaz apenas - como um navio que navega com demasiado peso - de diminuir a oscilação da vida quotidiana."*

**Vassili Kliutchevski**

## RESUMO

Uma grande quantidade de dados é produzida diariamente por operações comerciais e financeiras, mídias sociais e dispositivos móveis, sensores e outros equipamentos inseridos no mundo físico. Este fenômeno deu origem ao termo *big data*, e seus efeitos podem ser percebidos por empresas, ciência e governo. Entretanto, é a inteligência estratégica, não a informação, que auxilia gerentes a extrair valor dos grandes volumes de dados. Para isto, é necessário transformar a informação dispersa no ambiente em conhecimento estruturado e útil à tomada de decisão nas organizações. Este é um processo complexo, pois apesar das ferramentas e técnicas disponíveis é indispensável que o profissional em inteligência saiba lidar com a complexidade cognitiva inerente ao processo de análise. Neste contexto, o objetivo deste trabalho foi o de examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica. A pesquisa tratou de como o fenômeno *big data* é percebido pelos entrevistados em suas atividades analíticas em inteligência estratégica e propôs uma análise acerca de suas implicações. Para isso, foi realizado um estudo exploratório qualitativo. Foram entrevistados profissionais brasileiros, residentes nos Estados do Rio Grande do Sul, Rio de Janeiro, Distrito Federal e São Paulo. Estes entrevistados foram selecionados por meio de agentes com atuação, conhecimento e trânsito nos campos de inteligência estratégica e/ou *big data*. O roteiro que serviu de base para estas entrevistas foi estruturado a partir das dimensões do fenômeno *big data* e seus efeitos nas atividades analíticas no processo de inteligência estratégica. A técnica utilizada para análise dos dados foi análise de conteúdo. Os resultados indicam que o volume de *big data* contribui para a compreensão dos métodos de coleta, entretanto prejudica o domínio da matéria. Outras descobertas revelam que para alguns entrevistados *big data* já integra as práticas profissionais na realização de análises mais elaboradas e no desenvolvimento de projetos específicos. Entretanto para outros, *big data* ainda não é uma realidade, não sendo percebida a necessidade de utilizar grandes volumes de dados nas análises. Este fato também denota um paradoxo entre a caracterização da produção de conhecimento no campo do *big data* e o seu uso no campo profissional de inteligência estratégica. Por um lado o maior volume de trabalhos sobre o *big data* está no campo profissional das organizações produtivas e não na academia e por outro os profissionais em inteligência ainda não percebem o valor do fenômeno para sua atuação profissional.

**Palavras-chave:** Análise, *Big data*, Informação, Inteligência Estratégica.



## ABSTRACT

A considerable amount of data is daily produced by business and financial operations, social media, mobile devices, sensors as well as other gadgets available in the world. This phenomenon gave rise to the big data term whose effects can be perceived by companies, science and governments. However, the strategic intelligence, not the information itself, supports managers eliciting values from big volume of data. For this purpose, transforming the dispersed information in the environment into structure knowledge is necessary and useful for organizations' decisions. This is a complex process, and despite the tools and available techniques, it is indispensable that the professional in intelligence knows how to deal with inherit cognitive complexity in the courses of analyses. In this context, the objective of the present work was examining how the phenomenon big data affected the course of analyses in the strategic intelligence activity. The research was addressed on how the phenomenon big data was perceived by the interviewers in their analytics activities in strategic intelligence. It also proposed analysis based on its implications. In order to achieve this, a qualitative exploratory study was conducted. Several Brazilian professionals were interviewed, including residents in states such as Rio Grande do Sul, Rio de Janeiro, Distrito Federal and São Paulo. Those interviewed were chosen by agents with experience and knowledge in the strategic intelligence field and/or big data. The guide used for the interviews was structured from dimensions of big data phenomenon and its effects on the course of analyses in the strategic intelligence activity. The technique used for analyzing data was through content review. Results indicate that the volume of big data contributes to the comprehension of collection methods even though it eventually might debilitate the ability to grasp the topic. Other discoveries show that for some of those interviewed, big data has already integrated professional practices on not only performing a more detailed analyses but also developing specific projects. Nevertheless, big data is not a reality yet for others since the necessity of utilizing big volume of data for analysis is not really being perceived. This fact also denotes a paradox between the characterization of production knowledge within big data field and its use in the professional area of strategic intelligence. On the one hand, the great workload about big data is located in the professional area of productive organizations. Not in the academy, though. After all, intelligence professionals have not realized yet the real value of big data phenomenon for their professional performance.

**Key-words:** Analysis, Big data, Information, Strategic Intelligence.

## LISTA DE FIGURAS

Figura 01 - Relação entre temas abordados.....	20
Figura 02 - Conjunto teórico. ....	26
Figura 03 - Ecossistema de <i>big data</i> .....	30
Figura 04 - Linha do tempo dos recentes desenvolvimentos tecnológicos. ....	42
Figura 05 - Número de publicações/ano.....	47
Figura 06 - Tipo de publicação.....	48
Figura 07 - Número de artigos por áreas de publicação Scopus. ....	48
Figura 08 - Número de artigos por áreas de publicação WoS.....	49
Figura 09 - Publicações área ciência da computação. ....	50
Figura 10 - Publicações área ciência da computação e engenharia.....	51
Figura 11 - Publicações área ciência da computação, engenharia e ciência dos materiais. ....	52
Figura 12 - Publicações áreas de ciências agrárias e biológicas e biotecnologia e microbiologia aplicada. ....	53
Figura 13 - Publicações área multidisciplinar. ....	54
Figura 14 - Dispersão Q1. ....	55
Figura 15 - Scopus h=7.....	55
Figura 16 - WoS h=4.....	56
Figura 17 - Passos para inteligência. ....	62
Figura 18 - Ciclo da inteligência. ....	63
Figura 19 - Mudanças de mentalidade associadas ao <i>big data</i> .....	73
Figura 20 - Dimensões de big data relacionadas à eficiência e eficácia.....	81
Figura 21 - Evolução da ciência de dados. ....	83
Figura 22 - Habilidades sobrepostas de um cientista de dados. ....	90
Figura 23- Matriz de análise dos atributos de conteúdo.....	91
Figura 24- Ligação de dados na produção de conteúdo. ....	92
Figura 25 - Cadeia de referência do grupo. ....	103
Figura 26 - Estrutura da HU. ....	107
Figura 27 - Categorias contextuais e resultantes. ....	108
Figura 28 - Conjunto metodológico. ....	109
Figura 29 - UF dos entrevistados.....	110
Figura 30- Nível de formação.....	111

Figura 31 - Tempo de experiência.....	111
Figura 32 - Setores de atuação.....	112
Figura 33 - Disposição dos dados.....	113
Figura 34 - Estrutura dos memorandos. ....	114
Figura 35 - Grupos de entrevistas.....	115
Figura 36 - Influência do <i>big data</i> na atividade de IE: conceito. ....	118
Figura 37 - Influência do <i>big data</i> na atividade de IE: diversos. ....	120
Figura 38 - Influência do <i>big data</i> na atividade de IE: viés negativo.....	122
Figura 39 - Influência do <i>big data</i> na atividade de IE: diversos. ....	124
Figura 40 - Influência do <i>big data</i> na atividade de IE: relação com IE.....	126
Figura 41 - Influência do <i>big data</i> na atividade de IE: importância do analista.....	128
Figura 42 - Influência do <i>big data</i> na atividade de IE: experiência. ....	130
Figura 43 - Domínio da matéria e <i>big data</i> . ....	132
Figura 44 - Compreensão de métodos de investigação e <i>big data</i> . ....	132
Figura 45 - Imaginação e rigor científico e <i>big data</i> . ....	133
Figura 46 - Compreensão dos métodos de coleta e <i>big data</i> . ....	134
Figura 47 - Consciência das próprias potencialidades e <i>big data</i> .....	134
Figura 48 - Mentalidade aberta e <i>big data</i> .....	135
Figura 49 - Autoconfiança e <i>big data</i> .....	136
Figura 50 - Dimensões que estão sendo afetadas pelo fenômeno <i>big data</i> . ....	136
Figura 51 - Interface entre <i>big data</i> e atividade analítica. ....	138
Figura 52 - Outras dimensões de <i>big data</i> . ....	162
Figura 53 - Veracidade de <i>big data</i> . ....	163
Figura 54 - Viscosidade, Complexidade e Viralidade de <i>big data</i> . ....	165
Figura 55 - Efeitos do fenômeno <i>big data</i> . ....	166
Figura 56 - Esquema do Estudo.....	168

## LISTA DE TABELAS

Tabela 01 - Evolução da utilização dos dados e informação.....	27
Tabela 02 - Definições de <i>big data</i> .....	31
Tabela 03 - Volume de informações geradas. ....	36
Tabela 04 - Velocidade em que os dados são processados.....	37
Tabela 05 - De onde vem a informação a cada dia.....	40
Tabela 06 - Resumo dos atributos de qualidade da informação. ....	41
Tabela 07 - Publicações índice H - Área geral. ....	56
Tabela 08 - Publicações índice H - Área ciências sociais. ....	58
Tabela 09 - Definições análise.....	64
Tabela 10 - Exemplos para utilização de <i>big data</i> .....	71
Tabela 11 - Amostragem aleatória e <i>big data</i> .....	74
Tabela 12 - Análise big data.....	78
Tabela 13 - Tarefas e desafios no processo de análise de <i>big data</i> . ....	79
Tabela 14 - Tarefas e desafios no processo de análise de <i>big data</i> . ....	84
Tabela 15 - Resumo dos atributos de qualidade da informação. ....	89
Tabela 16 - Organização dos dados da entrevista.....	106

## LISTA DE SIGLAS

ABRAIC	Associação Brasileira dos Analistas de Inteligência Competitiva
BI	<i>Business Intelligence</i>
CDC	<i>Centers for Disease Control</i>
CRM	<i>Customer Relationship Management</i>
CTO	<i>Chief Technology Officer</i>
EDI	<i>Electronic Data Interchange</i>
ERP	<i>Enterprise Resource Planning</i>
HDFS	<i>Hadoop Distributed File System</i>
HMR	<i>Hadoop Map Reduce</i>
HU	<i>Hermeneutic Unit</i>
IDC	<i>International Data Corporation</i>
IE	Inteligência Estratégica
IQPC	<i>Internacional Quality &amp; Productivity Center</i>
JCR	<i>Journal Citation Report</i>
MIT	<i>Massachusetts Institute of Technology</i>
NoSQL	<i>Not Only Structured Query Language</i>
PLC	<i>Power Line Communication</i>
Q1	<i>High Quality Publications</i>
RF	<i>Radio Frequency</i>
SAS	<i>Statistical Analysis System</i>
SBGC	Sociedade Brasileira de Gestão do Conhecimento
SJR	<i>Scimago Journal Rank</i>
SPSS	<i>Statistical Package for the Social Sciences</i>
SQL	<i>Structured Query Language</i>
SWOT	<i>Strengths, Weaknesses, Opportunities, Threats</i>
TI	Tecnologia da Informação
Três Vs	Volume, Velocidade e Variedade
UCS	Universidade de Caxias Do Sul
UF	Unidade Federativa
W3C	<i>World Wide Consortium</i>
WoS	<i>Web of Science</i>
WWW	<i>World Wide Web</i>

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>15</b>
1.2 OBJETIVOS .....	21
1.3 JUSTIFICATIVAS E RELEVÂNCIA DO TEMA.....	21
1.4 DELIMITAÇÃO DO TEMA .....	25
<b>2 REFERENCIAL TEÓRICO .....</b>	<b>26</b>
2.1 <i>BIG DATA</i> .....	27
<b>2.1.1 Definições para o fenômeno .....</b>	<b>31</b>
<b>2.1.2 Características de <i>big data</i> .....</b>	<b>34</b>
2.1.2.1 Volume .....	35
2.1.2.2 Velocidade .....	37
2.1.2.3 Variedade.....	38
2.1.2.4 Outras características de <i>big data</i> .....	40
<b>2.1.3 Tecnologia da informação aplicada a <i>big data</i> .....</b>	<b>41</b>
<b>2.1.4 Pesquisas em <i>big data</i> e análise.....</b>	<b>47</b>
2.2 INTELIGÊNCIA ESTRATÉGICA.....	59
<b>2.2.1 Aspectos metodológicos e cognitivos do processo de análise .....</b>	<b>63</b>
<b>2.2.2 Ciência de dados .....</b>	<b>81</b>
<b>2.2.3 Aspectos humanos no processo de análise com <i>big data</i> .....</b>	<b>86</b>
2.3. VALOR DE <i>BIG DATA</i> PARA ORGANIZAÇÃO .....	91
<b>3 MÉTODO DA PESQUISA .....</b>	<b>100</b>
3.1 PARTICIPANTES DA PESQUISA.....	101
3.2 TÉCNICA DE ANÁLISE .....	103
3.3 COLETA E ORGANIZAÇÃO DOS DADOS .....	105
<b>4 RESULTADOS E ANÁLISE.....</b>	<b>110</b>
4.1 PROCESSO DE ANÁLISE DO ESTUDO .....	114
<b>4.1.1 Principais etapas das práticas de análise dos Profissionais de IE.....</b>	<b>115</b>
<b>4.1.2 Percepções da influência do <i>big data</i> na atividade de IE .....</b>	<b>117</b>
<b>4.1.3 Dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno <i>big data</i> .....</b>	<b>131</b>
<b>4.1.4 Interfaces entre <i>big data</i> e atividade analítica no processo de interpretação de informações para a atividade de IE .....</b>	<b>137</b>

<b>4.1.5 Efeitos fenômeno <i>big data</i> no processo de análise.....</b>	<b>139</b>
4.1.5.1 Volume/Velocidade/Variedade contribuem para o domínio da matéria .....	139
4.1.5.2 Volume/Velocidade/Variedade prejudicam o domínio da matéria .....	140
4.1.5.3 Volume/Velocidade/Variedade não afetam o domínio da matéria.....	141
4.1.5.4 Volume/Velocidade/Variedade contribuem para a compreensão de métodos de investigação .....	143
4.1.5.5 Volume/Velocidade/Variedade prejudicam a compreensão de métodos de investigação .....	144
4.1.5.6 Volume/Velocidade/Variedade não afetam a compreensão de métodos de investigação .....	145
4.1.5.7 Volume/Velocidade/Variedade contribuem para a imaginação e rigor científico ....	146
4.1.5.8 Volume/Velocidade/Variedade prejudicam a imaginação e rigor científico.....	147
4.1.5.9 Volume/Velocidade/Variedade não afetam a imaginação e rigor científico.....	148
4.1.5.10 Volume/Velocidade/Variedade contribuem para a compreensão dos métodos de coleta.....	149
4.1.5.11 Volume/Velocidade/Variedade prejudicam a compreensão dos métodos de coleta	150
4.1.5.12 Volume/Velocidade/Variedade não afetam a compreensão dos métodos de coleta	152
4.1.5.13 Volume/Velocidade/Variedade contribuem para a consciência das próprias potencialidades e outras influencias cognitivas.....	152
4.1.5.14 Volume/Velocidade/Variedade prejudicam a consciência das próprias potencialidades e outras influências cognitivas.....	154
4.1.5.15 Volume/Velocidade/Variedade não afetam a consciência das próprias potencialidades e outras influências cognitivas.....	155
4.1.5.16 Volume/Velocidade/Variedade contribuem para a mentalidade aberta a opiniões contrárias ou modelos alternativos .....	156
4.1.5.17 Volume/Velocidade/Variedade prejudicam a mentalidade aberta a opiniões contrárias ou modelos alternativos .....	157
4.1.5.18 Volume/Velocidade/Variedade não afetam a mentalidade aberta a opiniões contrárias ou modelos alternativos .....	157
4.1.5.19 Volume/Velocidade/Variedade contribuem na autoconfiança para admitir e aprender com os erros analíticos .....	159
4.1.5.20 Volume/Velocidade/Variedade prejudicam a autoconfiança para admitir e aprender com os erros analíticos .....	159

4.1.4.21 Volume/Velocidade/Variedade não afetam a autoconfiança para admitir e aprender com os erros analíticos .....	160
4.1.5.22 Outras dimensões de <i>big data</i> .....	162
4.2 DISCUSSÃO .....	167
<b>CONSIDERAÇÕES FINAIS.....</b>	<b>172</b>
<b>REFERÊNCIAS .....</b>	<b>177</b>
<b>APÊNDICE A - Detalhamento dos respondentes.....</b>	<b>181</b>
<b>APÊNDICE B - Roteiro de base para as entrevistas.....</b>	<b>184</b>



## 1 INTRODUÇÃO

Meio século depois de os computadores passarem a fazer parte da rotina de pessoas e empresas, as informações começaram a se acumular. De acordo com as previsões de Paquet (2010), da Gartner, os dados gerados por empresas, em todos os formatos, irão crescer 650% até 2015. De acordo com *International Data Corporation* (IDC), o volume mundial de dados dobra a cada 18 meses. A informação digital está dobrando a cada 1,5 anos e vai exceder 1000 *exabytes* no próximo ano, de acordo com o Centro de Pesquisa do *Massachusetts Institute of Technology* (MIT) Digital (AKERKAR, 2014).

Segundo a lei Moore, o aumento do poder de processamento dos computadores dobra a cada dois anos. Esta evolução é necessária, para que seja possível suportar o atual volume de dados e informações gerados por operações comerciais e financeiras, provenientes de mídias sociais, dispositivos móveis, sensores e outros equipamentos inseridos no mundo físico, como rodovias, automóveis, aeronaves e máquinas robóticas. Esse dilúvio de dados, denominado *big data*, pioneiramente apresentado por Cukier (2010) em relatório publicado no periódico britânico *The Economist*, cria um desafio para a comunidade empresarial e os cientistas de dados.

Assim, o mundo não apenas está mais cheio de informação como também a informação está se acumulando com mais rapidez. Tecnólogos acreditam que a linhagem do *big data* remonta à época da revolução do silício. Certamente os sistemas modernos de TI possibilitaram o *big data*, mas em essência, este avanço é uma continuação da antiga busca da humanidade por medir, registrar e analisar o mundo. Esta evolução em termos de análise remonta aos primórdios na antiga Mesopotâmia. Teve seu início com a era analógica, onde a análise de dados era cara e demorada, passando pela digitalização, onde as informações analógicas se tornaram compreensíveis à computadores e chegando à origem do que atualmente se chama de dataficação, que consiste na coleta de tudo que existe, e na transformação disto em dados quantificáveis (SCHÖNBERGER-MAYER E CUKIER, 2013).

A definição tomada como base para este estudo é a de Manyika et al. (2011), segundo o autor *big data* é o conjunto de dados cujo tamanho vai além da capacidade para capturar, armazenar, gerenciar e analisar de ferramentas de *software* de banco de dados típicos. Existem organismos internacionais como *Boston Consulting Group*, *UN Global Pulse*, *Global Viral Forecasting Initiative*, *Ushahidi*, *McKinsey Global Institute*, *World Bank*, *Vodafone*, *World*

*Economic Forum's ICT Global Agenda Council*, entre outros, que estão se dedicando a pesquisar e melhor entender como extrair valor dos grandes volumes de dados.

De acordo com Mineli, Chambers e Dhiraj (2013), *big data* tem o potencial para ser diferente das outras épocas em função da possibilidade de analisar os dados em sua forma original, não estruturada; e, analisar não apenas o que houve no passado, mas sim prever o que irá acontecer ao redor do mundo, com riqueza de detalhes. Esta mudança não envolve apenas a quantidade de dados, mas também a frequência com que os dados chegam e a diversidade de fontes de onde eles chegam, os diferentes tipos de dados e sua disponibilidade (BETSER E BELANGER, 2013). Isto compõe os 3 Vs de *big data*, o volume, a velocidade e a variedade. De acordo com McAfee e Brynjolfsson (2012), estes são os atuais elementos de diferenciação deste fenômeno, em comparação a outras épocas.

Em termos tecnológicos, para que seja possível tratar com volume, velocidade e variedade, é necessário que haja uma evolução tecnológica (TAURION, 2013). As tecnologias que sustentam *big data* podem ser analisadas sob duas óticas: as tecnologias de infraestrutura, como banco de dados *NoSQL*<sup>1</sup>, que armazenam e processam *petabytes* de dados, e as envolvidas com análise, como *Hadoop* e *Map Reduce*<sup>2</sup>. Entretanto, para que o investimento em tecnologias possa proporcionar à empresa o retorno esperado, é necessário que haja um alinhamento ao modelo de negócio existente e que seja desenvolvido um processo de transformação sincronizado entre pessoas, processos e tecnologia.

De acordo com Akerkar (2014), um fato interessante sobre *big data* é que nem tudo o que é considerado grande volume de dados é, de fato, *big data*. É preciso explorar em profundidade os aspectos científicos, tais como análise, processamento e armazenamento de grandes volumes de dados. Essa é a única maneira de usar as ferramentas de forma eficaz. Desenvolvedores de dados/cientistas precisam saber sobre processos analíticos, estatísticas e aprendizagem de máquina. O núcleo é o lado analítico, mas eles também precisam de fundo científico e profundo conhecimento técnico das ferramentas com que trabalham, a fim de ganhar o controle de *big data*. Não há ferramenta que possa realizar isto.

Akerkar (2014) explica que *big data* tem seus impactos tanto na ciência, quanto nos negócios. De acordo com Dumbill (2012) o valor de *big data* para uma organização se divide em duas categorias: de uso analítico e permitindo a criação de novos produtos. Atualmente a análise de *big data* pode revelar percepções ocultas, pois antes os dados eram muito caros

---

<sup>1</sup> *NoSQL* refere-se à tecnologias de infraestrutura de bancos de dados não-relacionais (SCHÖNBERGER-MAYER E CUKIER, 2013).

<sup>2</sup> *Hadoop* e *Map Reduce* são tecnologias de processamento que permitem o gerenciamento de grandes volumes de dados (SCHÖNBERGER-MAYER E CUKIER, 2013).

para o processo. Ser capaz de processar cada item de dados em tempo razoável elimina a necessidade problemática pela amostragem e promove uma abordagem investigativa de dados, em contraste com a natureza um tanto estática de execução de relatórios pré-definidos. Assim, *big data* está intimamente ligado ao surgimento da ciência de dados, uma disciplina que combina matemática, programação e instinto científico.

Contudo, de acordo com Fuld (1995) é a inteligência estratégica (IE), não a informação, que auxilia gerentes a responderem corretamente questões específicas e tomarem decisões a longo prazo. Atualmente, todas as companhias, grandes ou pequenas, têm o mesmo acesso à informação. Assim, é necessário que sejam feitas as questões certas, e que seja feita a análise nos resultados das respostas, para que as informações disponíveis possam ser convertidas em inteligência. Em meio às pesquisas que estão sendo desenvolvidas sobre *big data* não foram identificados estudos estabeleçam a relação do fenômeno no contexto da IE, de acordo com análise das bases de dados realizada pela autora. No entanto, a análise que de acordo com Schönberger-Mayer e Cukier (2013) é o núcleo para utilização do *big data*, compõe o que na visão de Fleisher e Bensoussan (2007) é a etapa do processo de inteligência em que maior valor é gerado.

De acordo com Fuld (1995), o processo de IE inicia com a descoberta dos dados, passando pela análise e por último a transformação dos dados em inteligência. Para Fleisher e Bensoussan (2007), o trabalho da análise é interativo, com todos os sub processos de classificação e síntese de dados para produzir mapas, gráficos e outros recursos de comunicação para disseminação. A análise, ainda leva em consideração aspectos metodológicos e cognitivos para executar a interpretação da informação. Os aspectos metodológicos consistem em técnicas que auxiliam o trabalho dos profissionais envolvidos, entretanto os cognitivos são os que parecem se colocar como obstáculo na transformação da informação em inteligência (HEUER, 1999).

Isto porque, o profissional em inteligência está sujeito a fraquezas e predisposições inerentes aos processos do pensamento humano. De acordo com o autor, os aspectos cognitivos do processo de análise podem ser vistos como perceber o mundo através de uma lente ou tela onde canais são focados, podendo assim, distorcer as imagens que são vistas. Assim, os profissionais em inteligência precisam mais do que a informação, eles também precisam entender as lentes pelas quais esta informação passa. Dentro do contexto de *big data*, a este profissional foi atribuída a designação "cientista de dados" (DAVENPORT E PATIL, 2012).

Assim, com a combinação de *big data* com a análise inteligente é possível que sejam descobertos novos padrões (SCHÖNBERGER-MAYER E CUKIER, 2013). Neste sentido é possível extrair valor de *big data*, mas a figura do profissional que irá realizar a transformação dos dados em inteligência é o ponto essencial em que se deve focar. Isto se deve ao fato de que as ferramentas tecnológicas disponíveis possibilitam a manipulação de grandes volumes de dados, mas a transformação é realizada pelo profissional. Este ainda pode contar com ferramentas técnicas para realizar a análise dos dados, mas é nos aspectos cognitivos que está a chave para construir sentido e agregar valor ao processo, em prol de ganhos científicos, empresariais e sociais.

### 1.1 PROBLEMATIZAÇÃO

Cukier (2010) explica que os negócios em gestão da informação estão contribuindo para facilitar o entendimento dos grandes volumes de dados. Empresas como Oracle, IBM, Microsoft e SAP têm realizado nos últimos anos investimentos na ordem de US\$ 15 bilhões em *softwares* especializados de gestão e análise. Estima-se que o valor desta indústria esteja em torno dos US\$ 100 bilhões com crescimento de quase 10% ao ano, o que representa aproximadamente duas vezes mais do que o negócio de *software* como um todo.

O potencial de agregação de valor de *big data* nos negócios, se reflete em uma melhor satisfação dos clientes, melhores serviços e na contribuição para criação e manutenção de um negócio bem sucedido. Neste sentido, *startups* de sucesso da *web* da última década são exemplos de *big data* utilizado como facilitador de novos produtos e serviços. Além disso, outros setores da indústria, como bancos, saúde, seguros, manufatura, varejo, atacado, transporte, comunicações, construção, educação e serviços públicos são setores da economia que *big data* está transformando (AKERKAR, 2014).

Na ciência, *big data* também demonstra sua presença, no sentido de oferecer um novo olhar à pesquisa científica, pelo fato de atualmente expressarmos mais aspectos da realidade na forma de dados. Assim, a dataficação<sup>3</sup> permite que dados quantificados possam ser utilizados de novas maneiras, como análises preditivas, revelando o valor latente e implícito das informações e a mudança da causalidade pela probabilidade e correlação. Para isso, a mentalidade sobre como os dados podem ser utilizados é uma mudança mais importante e sutil do que as tecnológicas, apesar destas terem sido fator crítico que possibilitou o

---

<sup>3</sup> Dataficação um fenômeno é colocá-lo num formato quantificado de modo que possa ser tabulado e analisado (SCHÖNBERGER-MAYER E CUKIER, 2013).

surgimento de vários negócios. Essa mudança subverte séculos de práticas consagradas e desafia nossa compreensão mais básica de como tomamos decisões e compreendemos a realidade (SCHÖNBERGER-MAYER E CUKIER, 2013).

De acordo com Mineli, Chambers e Dhiraj (2013), nos negócios, cada vez mais a necessidade de análises tem se mostrado necessária. Contudo, esta é uma disciplina em evolução e o que ocorre, em função disso, é que as empresas estão investindo em análises que muitas vezes não são utilizadas no negócio. Para isto, é necessário que se invista corretamente em termos das dimensões de processo, tecnologia e pessoas para que possa ser preenchida a lacuna entre a criação e o consumo da análise.

Entretanto, Manyika et al. (2011) expõem que o cenário que se apresenta é de escassez de talento analítico e gerencial para transformar o conhecimento em melhoria do desempenho organizacional. Os Estados Unidos sozinhos deverão enfrentar até o ano de 2018, uma escassez de 140.000 a 190.000 pessoas com habilidades analíticas. Isso representa um *gap* de 50% a 60% na demanda. Também haverá falta de 1,5 milhões de gerentes e analistas para analisar os grandes volumes de dados e tomar decisões com base em suas descobertas. Assim, embora a análise quantitativa seja com base nos Estados Unidos, o relatório declara que a restrição sobre este tipo de talento será global. De acordo com Manyika et al. (2011), este é um desafio importante e urgente que as empresas e formuladores de políticas devem resolver no curto prazo: o desenvolvimento da capacidade de atrair funcionários com habilidades críticas para o mundo dos grandes volumes de dados.

Por outro lado, as contribuições de *big data* são possíveis de serem realizadas desde que a matéria-prima informacional, gerada pela exaustão dos dados<sup>4</sup>, seja transformada em inteligência, por meio do processo de análise. A análise em IE, na abordagem de Fleisher e Bensoussan (2007), é definida como a aplicação hábil de métodos e processos científicos e não científicos pelos quais os indivíduos interpretam dados ou informações para produzir resultados de inteligência perspicazes e recomendações viáveis para os tomadores de decisão.

Fleisher e Bensoussan (2007) esclarecem ainda que os analistas são tão importantes atualmente, por causa do aumento do reconhecimento e do valor das boas habilidades de pensamento. A análise não pode ser conduzida na ausência de pensar. Sem isso, seria uma escolha aleatória e de sorte. Isto não anula o valor do instinto, mas recomenda que o processo seja conduzido paralelamente com métodos de análise fidedignos e testados. Como o acesso

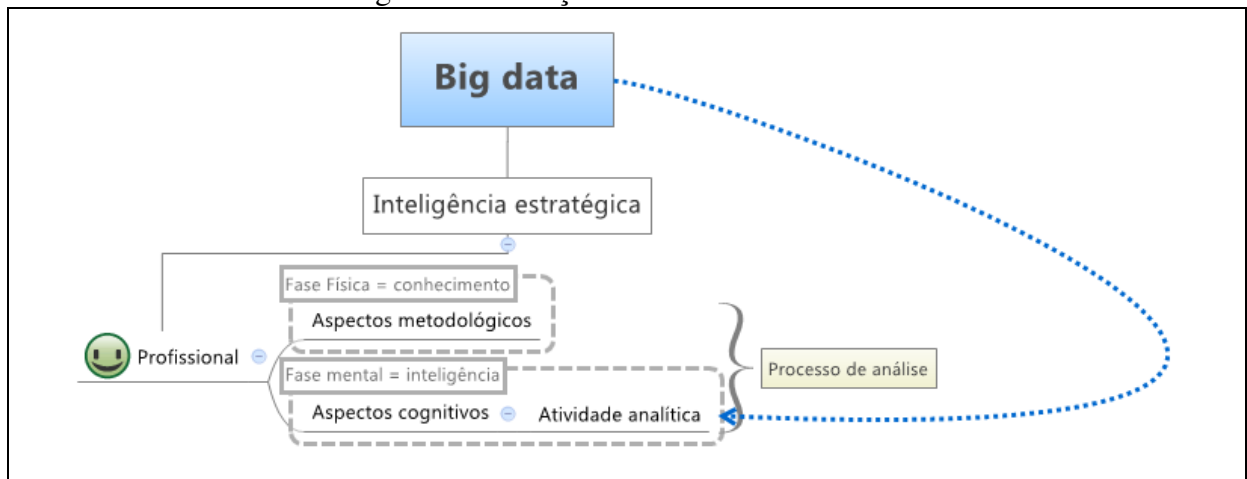
---

<sup>4</sup> Exaustão dos dados foi um termo que surgiu para descrever a trilha digital que as pessoas deixam, que se refere a dados colhidos como subprodutos das ações e dos movimentos das pessoas (SCHÖNBERGER-MAYER E CUKIER, 2013).

aos dados ou informações tem aumentado, o maior valor é agora colocado em não apenas obter os dados apropriados, mas mais importante ainda, fazer com que eles façam sentido.

Isso ocorre em meio ao processo de análise, com a geração de conhecimento. Nesta fase dados são agregados, relacionados entre si e passam a fazer sentido. Dentre os métodos disponíveis para o processo de análise, Fleisher e Bensoussan (2003) expõem cinco categorias que são as técnicas de análise estratégica, análise da competência e do consumidor, análise de retorno, análise evolutiva e análise financeira. Esta é a chamada fase física, apontada por Palop e Vicente (1999). A fase mental ocorre na sequência, e é quando diferentes significados podem ser atribuídos para o mesmo conjunto de informações. Assim, segundo Palop e Vicente (1999) aos dados é atribuído significado genérico o qual ambientado ao contexto e com influência de fatores pessoais, adquire significado e pode então ser chamado de inteligência. Na Figura 01 é possível visualizar os temas abordados:

Figura 01 - Relação entre temas abordados.



Fonte: elaborado pelo próprio autor.

A Figura 01 descreve este processo, onde o conhecimento é ambientado ao contexto da organização, por meio de técnicas, capacidade analítica e criatividade do profissional envolvido (atividade analítica). Assim, o processo de IE é o caminho pelo qual o conhecimento percorre até ser transformado em inteligência. A inteligência, por sua vez, é a implementação do conhecimento transformado em ações que levem a organização a obter algum tipo de vantagem competitiva frente aos seus concorrentes. Com base nestes pressupostos, a questão que emerge e que orienta a presente pesquisa é como o fenômeno *big data* afeta o processo de análise de informações na atividade de inteligência estratégica?

## 1.2 OBJETIVOS

O tema que trata esta dissertação diz respeito ao entendimento das implicações do fenômeno *big data* na análise para inteligência estratégica. Com isto o objetivo proposto para o estudo é examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica.

Assim, foram determinados os seguintes objetivos específicos:

- a) identificar as principais etapas das práticas de análise dos profissionais de IE;
- b) verificar a percepção da influência do *big data* na atividade de IE;
- c) identificar as dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data*;
- d) analisar as interfaces entre *big data* e atividade analítica no processo de interpretação de informações para a atividade de IE.

## 1.3 JUSTIFICATIVAS E RELEVÂNCIA DO TEMA

Letouzé (2012) aponta que a quantidade de dados digitais disponíveis em nível global cresceu de 150 *exabytes* em 2005 para 1200 *exabytes* em 2010. Ela é projetada para aumentar em 40% ao ano para os próximos anos. Isso representa cerca de 40 vezes o debatido crescimento da população mundial. Esta taxa de crescimento significa que o estoque de dados digitais deve aumentar 44 vezes entre 2007 e 2020, dobrando a cada 20 meses. Segundo Rapporteur (2010), a explosão das redes móveis, a computação em nuvem e novas tecnologias deram origem a um número incompreensivelmente grande de informações. A oportunidade está no fato de que usando técnicas de correlação avançada, analistas podem peneirar densos trechos de dados para prever condições, comportamentos e eventos antes que eles aconteçam.

Rapporteur (2010) relata exemplos como o da multinacional de serviços *online* e *software* dos Estados Unidos, Google, que estuda a hora e o local em que o motor de pesquisa irá realizar consultas para prever surtos de gripe e desemprego, entre outros. Ela capta tendências antes que estas sejam publicadas em estatísticas oficiais do governo. De forma semelhante, empresas de cartão de crédito utilizam informação financeira e pessoal para tentar detectar fraudes e identificar consumidores perseguindo tendências. Da mesma forma,

médicos pesquisadores vasculham os registros de saúde de milhares de pessoas para tentar identificar correlações úteis entre os tratamentos médicos e resultados de saúde.

Em termos de tecnologia, Cukier (2010) coloca que os negócios em gestão da informação estão contribuindo para facilitar o processo de construção de sentido nos grandes volumes de dados. Assim, ainda segundo o autor, o fenômeno *big data*, tem seus efeitos percebidos nas empresas, na ciência e no governo. Suas características/dimensões são apresentadas com mais volume, velocidade e variedade - os três Vs - de dados que chegaram de forma inédita. De acordo com Taurion (2013), as tecnologias que sustentam *big data* podem ser analisadas sob duas óticas: as tecnologias de infraestrutura, como bancos de dados, *NoSQL*, que armazenam e processam *petabytes* de dados e as envolvidas com análise, como *Hadoop* e *MapReduce*.

Contudo, atualmente nenhum sistema de computador pode traduzir de forma confiável linguagem não estruturada para formatos de bancos de dados estruturados. Portanto, sua atuação se restringe apenas ao processamento de dados. Ainda assim, a tecnologia de *big data* é a parte fácil - a parte mais difícil é descobrir para onde ir com a análise de dados e certificar-se de que a empresa tem as pessoas e os processos prontos antes de se comprometer a comprar a tecnologia (MINELI, CHAMBERS E DHIRAJ, 2013).

Assim, os investimentos em novas tecnologias devem estar alinhados aos modelos de negócio existentes - em termos de pessoas e processos, para que se possa ter o retorno esperado das iniciativas em *big data*. Por isso, mesmo com suporte de ferramentas de *software* é o processo de tomada de decisão que possibilitará às organizações realizarem as mudanças necessárias em busca de um posicionamento estratégico que leve à vantagem competitiva (CUKIER, 2010). Mineli, Chambers e Dhiraj (2013) confirmam este pensamento quando colocam que as empresas percebem que um dos muitos fatores que separam o fracasso do sucesso nos negócios, é a capacidade de utilizar eficazmente a análise, que é parte do processo de inteligência estratégica, para tomar melhores decisões.

De acordo com os autores, o que se apresentava no passado e que ainda permanece em termos de análise é a utilização das disciplinas de negócios com a matemática, permitindo a utilização de heurísticas e criatividade para tornar persuasivos os argumentos na sala de reuniões, dando origem ao negócio da consultoria. As disciplinas de negócios, combinadas com a tecnologia, por sua vez, contribuindo para a automatização de tarefas algorítmicas e melhorando a produtividade e eficiência dos processos, dando origem ao negócio da TI. A utilização das disciplinas de tecnologia, com matemática, inspirando o desenvolvimento de



produtos de *software* que ajudariam a lidar com uma ampla gama de problemas de negócios e operar de forma proativa, com antecipação.

Nestes termos, dadas as características apresentadas pelo fenômeno *big data*, o que se apresenta em termos de análise é a utilização das disciplinas de negócios, combinadas com a matemática e tecnologia, se unindo com ciências comportamentais. Estas irão permitir conectar pontos entre as interações e o desenvolvimento de uma compreensão mais profunda do talento humano. Esta disciplina emergente, denominada ciência de dados, permitirá que sejam criados os incentivos adequados para conduzir comportamentos que se alinham aos objetivos de negócios (MINELI, CHAMBERS E DHIRAJ, 2013).

A importância deste tema é mencionada, inclusive, em periódicos de negócios de circulação nacional. Um exemplo é o do *expert* Peter Hirshberg, ex-diretor de marketing da Apple, conselheiro do laboratório de cidades digitais do MIT e membro do Global Pulse, que em entrevista concedida a Revista Exame, diz que o fator humano é fundamental para se chegar a conclusões corretas em meio ao *big data*. Hirshberg (2013) coloca que é preciso ter profissionais que saibam ler os dados e transformá-los em ações práticas e que infelizmente o caminho é curto para àquelas empresas que querem extrair valor de *big data*, a partir apenas de consultorias e *softwares*.

Algumas universidades na América do Norte e da Europa estão fazendo sua parte para alimentar a necessidade de habilidades atuais de análise. Nos últimos anos, eles introduziram grau de mestrado em análise de *big data*, ciência de dados e análise de negócios (AKERKAR, 2014). Assim, o número de cursos tende a aumentar em todo o mundo, pois de acordo com Manyika et al. (2011), *big data* está se tornando uma base essencial para a concorrência, novas ondas de crescimento, produtividade, inovação e excedente do consumidor.

Segundo o Gartner (2013), 4,4 milhões de postos de trabalho em *big data* serão criados em 2015, e apenas um terço deles será preenchido. Desta maneira, a previsão para o talento em *big data*, será formada de organizações que proporcionarão ofertas de emprego lucrativas na área. Esse cenário promete grandes oportunidades para analistas de dados, em curto prazo, mas obscurece os desafios de longo prazo das organizações que esperam se beneficiar de estratégias em *big data*.

Neste sentido, a contratação de cientistas de dados será a parte mais fácil. O verdadeiro desafio será integrar esse talento, recém adquirido, nas estruturas organizacionais existentes e inventar novas estruturas que permitam a estes profissionais gerar valor real para as organizações (BARLOW, 2013). Assim, a contribuição apresentada por este trabalho está

no entendimento das implicações de *big data* na análise para inteligência estratégica, considerando as características do fenômeno e as atividades do profissional em inteligência estratégica.

Para levantar subsídios que justificassem este trabalho, do ponto de vista teórico, a autora realizou uma pesquisa bibliométrica nas bases de dados dos indexadores *Scopus* e *Web of Science* (WoS). A coleta dos dados foi realizada no mês de dezembro de 2013 e foram pesquisadas publicações que empregassem os termos *big data* e análise, nos campos "título do artigo", "resumo" e "palavras-chave", no caso do Scopus e no campo "tópico", no caso da WoS. Os dados foram coletados desde 2010, onde o termo foi primeiramente citado por Cukier (2010), até dezembro de 2013.

Como resultados, foram identificadas 325 publicações na base Scopus e 113 na WoS. Pode-se observar que o número de publicações nestes temas apresenta um crescimento médio de mais de 200% ao ano. Os tipos de publicação que mais se sobressaem neste total são as do tipo conferência, caracterizando a contemporaneidade do tema. Tomados os artigos e livros lançados, neste total foi possível observar que a maior área de concentração dos focos de publicação é a ciência da computação. As ciências sociais ocupam posições de menor expressão.

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

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

Assim, embora algumas fontes de alto fator de impacto estejam publicando sobre o tema, ainda há muito a ser explorado em termos de *big data* e análise (AKERKAR, 2014). Entretanto, na pesquisa às bases de dados não foram identificados artigos publicados com os temas *big data* e análise, relacionados à inteligência estratégica. Sendo assim, fica evidenciado um campo aberto em termos acadêmicos o que motiva o desenvolvimento deste trabalho.

#### 1.4 DELIMITAÇÃO DO TEMA

Os dois temas abordados neste estudo, do ponto de vista teórico são: *big data* e processo de análise para inteligência estratégica. Considerando a influência do *big data* no processo de análise, e sendo este ponto crítico para a geração de valor, o presente estudo examina como ocorre esta relação nas práticas de IE, este é o objeto de estudo desta pesquisa.

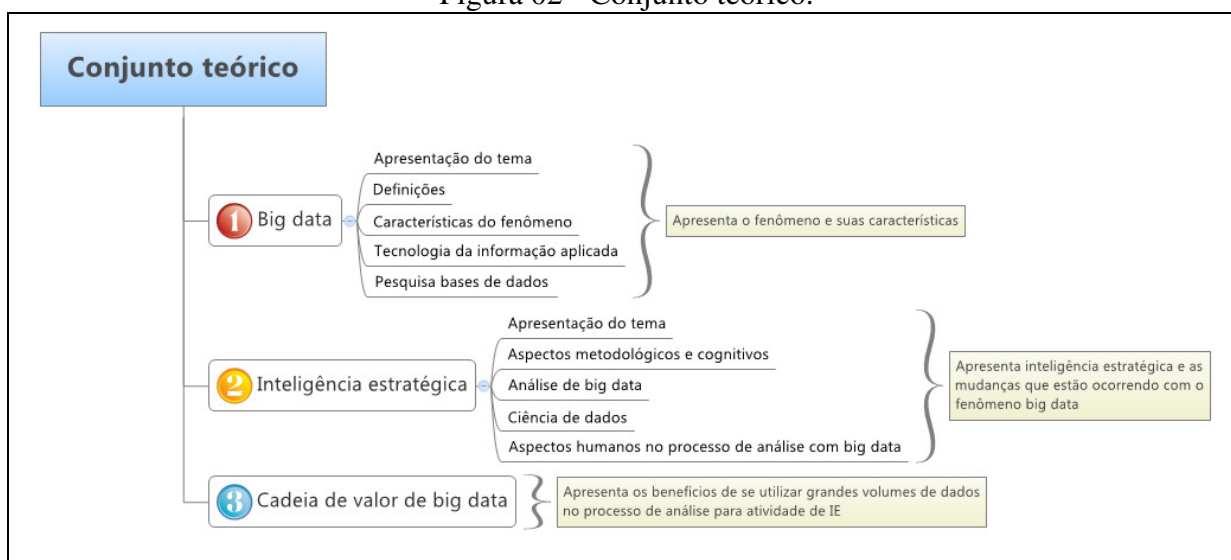
Ao profissional que realiza a análise de dados podem ser atribuídas diferentes denominações. Na análise da literatura pode-se encontrar termos como desenvolvedor de dados, analista de dados e mais recentemente, cientista de dados (DAVENPORT E PATIL, 2012). Por este trabalho ter um viés do processo de análise em inteligência estratégica, este profissional foi denominado "profissional em inteligência estratégica". Entretanto, os demais termos foram citados, no sentido preservar os trabalhos que vem sendo desenvolvidos em *big data*.

Para efeitos deste estudo, foram considerados o volume, a velocidade e a variedade de dados e informações como dimensões atribuídas ao fenômeno do *big data*. Outros autores sugerem diferentes características ou dimensões, que serão citadas ao longo do trabalho, entretanto estas três são as que predominam na bibliografia utilizada. As atividades do profissional em inteligência estratégica consideradas como dimensões para este estudo foram aquelas definidas por Fachinelli, Rech e Bertolini (2011), a partir dos trabalhos de Heuer (1999) e Bruce e George (2008). Estas foram consideradas por serem uma síntese, com abrangência metodológica e cognitiva, de outras capacidades que serão citadas ao longo da pesquisa. Este estudo possui caráter exploratório, portanto não propõe novas ferramentas ou mecanismos, mas sim descreve as percepções dos profissionais entrevistados sob a luz do eixo teórico acerca dos temas *big data* e análise em inteligência estratégica.

## 2 REFERENCIAL TEÓRICO

Este capítulo apresentará o referencial teórico da dissertação com os *temas big data* e inteligência estratégica. A pesquisa acadêmica realizada reflete estudos empresariais do campo da gestão e por isto pode ser interpretada como tendo um discurso fragmentado e voltado para empresas. A Figura 02 apresenta o conjunto teórico tomado como base para realização desta pesquisa:

Figura 02 - Conjunto teórico.



Fonte: elaborado pelo próprio autor.

De acordo com a Figura 02, a seção de *big data* subdivide-se na apresentação do tema, definições, características do fenômeno e tecnologia da informação aplicada a *big data*. Esta seção será finalizada com a apresentação de uma pesquisa realizada às bases de dados *Scopus* e *Web of Science*, realizada pela autora com o objetivo de identificar o número e tipos de publicação que estão sendo desenvolvidas sobre os temas *big data* e análise. Isto foi feito para tornar evidente a carência de publicações acadêmicas sobre os temas e apresentar que não foram identificados estudos que mencionassem a questão da inteligência estratégica relacionada à *big data*.

Outra seção irá tratar da inteligência estratégica, apresentação do tema e aspectos metodológicos e cognitivos. Esta mesma seção também irá apresentar a análise de *big data*, ciência de dados e aspectos humanos no processo de análise com *big data*. Isto foi incluído na seção que fala de inteligência, pois representa as mudanças que estão ocorrendo no contexto

de IE em função do fenômeno *big data*. Posteriormente isto será utilizado para subsidiar a seção de discussão dos resultados encontrados no estudo.

A última seção apresentada no capítulo irá tratar da agregação de valor do fenômeno *big data* ao processo de inteligência e às organizações. Nela serão expostos aspectos referentes ao potencial de geração de valor de *big data*. Esta seção também irá subsidiar uma outra de discussão dos resultados.

## 2.1 BIG DATA

De acordo com Akerkar (2014), o surgimento de *big data* se deve à rápida expansão do universo de máquinas e a usuários altamente qualificados. Como resultado, a composição tradicional de dados duráveis e estruturados, servindo a um propósito específico, passa a se modificar. Este fato, aliado ao uso de ferramentas avançadas de análise exploratória de dados, mineração de dados aprendizagem/máquina e visualização de dados, oferecem uma nova forma de compreender o mundo. Com isso, os dados que antes eram considerados estáticos e banais se tornaram matéria-prima dos negócios.

Schönberger-Mayer e Cukier (2013) complementam com a constatação de que o volume de informação cresceu tanto que passou a não caber na memória de processamento dos computadores, por isso os engenheiros tiveram de aprimorar os instrumentos utilizados para fazer a análise. Esta é a origem de novas tecnologias de processamento como a *MapReduce* da Google e sua equivalente de código aberto, *Hadoop*, lançada pela Yahoo. Elas permitem que se gerenciem muito mais dados que antes, e os dados não precisam ser alocados em fileiras ou nas clássicas tabelas.

Ocorre que meio século depois de os computadores entrarem no meio social os dados começaram a se acumular. A Tabela 01 mostra a evolução da utilização dos dados e informação desde sua origem até os dias atuais:

Tabela 01 - Evolução da utilização dos dados e informação.

(continua)

Ano ou Período	Fato
8000 a.C.	Comerciantes sumérios usavam tábuas e fichas de barro para denotar os bens comercializados.
1086	Livro Domesday, um dos mais venerados tesouros britânicos, foi uma contagem abrangente - e sem precedentes - dos ingleses, suas terras e propriedades.
1439	Primeira revolução da informação: Impressora de Gutenberg.

(conclusão)

Ano ou Período	Fato
1453 e 1503	De acordo com a historiadora Elizabeth Eisenstein, oito milhões de livros foram impressos. <b>Volume de informações dobra na Europa, em 50 anos.</b>
1662	O comerciante britânico chamado John Graunt queria saber a população de Londres na época da peste negra. Em vez de contar as pessoas, ele inventou uma abordagem - que hoje chamamos de "estatística"- que lhe permitiu estimar a população.
1880	O Census Bureau dos Estados Unidos contratou o inventor Herman Hollerith, que desenvolveu um sistema de cartões perfurados e máquinas de tabulação para o censo de 1890. Ele conseguiu diminuir o tempo de tabulação do censo de oito anos para um.*
1920	Descobertas da mecânica quântica abalaram para sempre o sonho das medições abrangentes e perfeitas.
1934	Jerzy Neyman, estatístico polonês, demonstrou que a abordagem da amostragem levava a erros e que para isto deveria ser considerada a aleatoriedade na escolha da amostra. A amostragem tornou-se a solução para o problema da profusão de dados.**
1950-1960	Um sistema de folha de pagamento foi implementado na General Electric Corporation, por Joe Glickauf e Arthur Andersen em um computador eletrônico digital. Início das ideias de "Revolução da informação" e "Era Digital" ou " <b>Era da informação</b> ".
1980-1990	Nesta época, Leonard Kleinrock, Lawrence Roberts, Robert Kahn, and Vint Cerf inventaram a <i>internet</i> . 40% de toda a potencia computacional do mundo existia na forma de calculadoras de bolso. " <b>Era da rede</b> ".
2000	Dados: 75% dados analógicos (papel, filme, vinil, fitas magnéticas, livros, fotografia); 25% dados digitais.
2007	Dados: 7% dados analógicos; 93% dados digitais.
2010	Ciências como a astronomia e a genômica, vivenciaram uma explosão informacional e cunharam o termo <i>big data</i> . Este fenômeno, representa uma transição em termos de armazenamento e análise. " <b>Era big data</b> ".
2013	Dados: 2% dados analógicos; 98% dados digitais - 1200 <i>exabytes</i> . <b>Volume de informações dobra a cada 3 anos.</b>

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

\* Marco do início do processamento automatizado de dados: a quantidade de informações coletadas emperrava os instrumentos usados para processá-la, e eram necessárias novas técnicas.

\*\* A amostragem aleatório foi um sucesso e é a base da medição moderna em escala.

Desta forma, a Tabela 01 mostra que o mundo não apenas está mais cheio de informação como também a informação está se acumulando com mais rapidez. Tecnólogos acreditam que a linhagem do *big data* remonta à época da revolução do silício. Certamente os sistemas modernos de TI possibilitaram o *big data*, mas em essência este avanço é uma continuação da antiga busca da humanidade por medir, registrar e analisar o mundo (SCHÖNBERGER-MAYER E CUKIER, 2013).

A maneira para pensar a questão, proposta por Schönberger-Mayer e Cukier (2013), é que *big data* se refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar novas formas de valor de maneira que se alterem os mercados, as organizações, a relação entre cidadãos, governos e outros. Então, a sociedade precisará conter um pouco a obseção pela causalidade e trocá-la pela correlação simples: sem saber o porquê, apenas o quê. Essa mudança subverte séculos de práticas consagradas e desafia a compreensão mais básica de como tomar decisões e compreender a realidade.

Entretanto, Akerkar (2014) expõe que essa é uma visão otimista para *big data* e que esses conjuntos de dados, onde a semântica é ignorada em favor de correlações, irá expirar cedo, sem utilidade. No entanto, de acordo com o autor a integração holística de algoritmos em evolução, máquinas, pessoas e esforço de pesquisa em vários domínios vai garantir a exigida aptidão de *big data*, garantindo qualidade adequada no momento certo.

De acordo com Schönberger-Mayer e Cukier (2013), os frutos da sociedade da informação, como celulares e computadores, deram origem ao termo "exaustão de dados". O termo descreve a trilha digital que as pessoas deixam, que se refere a dados colhidos como subprodutos das ações e dos movimentos das pessoas. Estes elementos são facilmente identificados em todos os lugares. Entretanto, a informação em si não é facilmente percebida, pois demanda uma transformação no conhecimento.

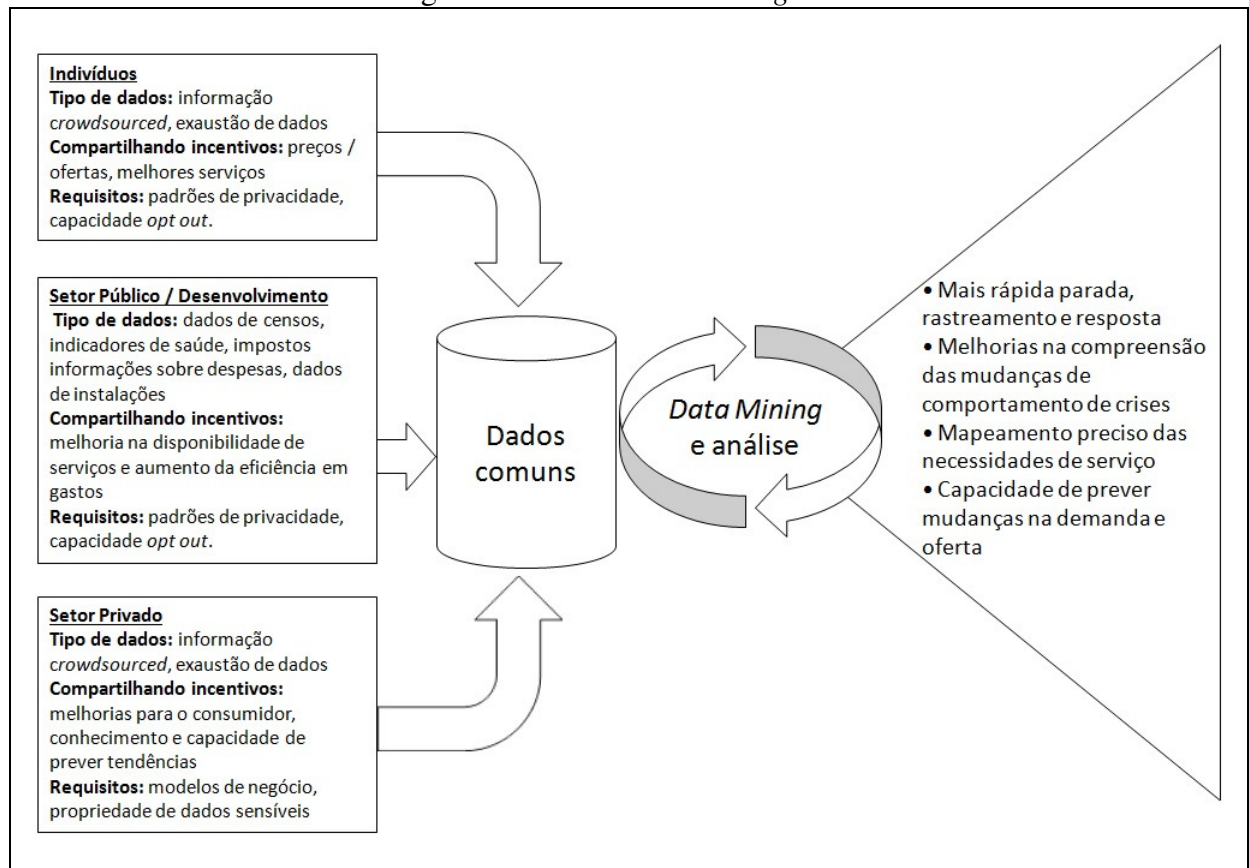
Na visão do autor, essa mudança quantitativa advinda de *big data* tem gerado uma mudança qualitativa em termos de resultados do processamento de dados. Betsler e Belanger (2013) complementam que a mudança qualitativa refere-se a quantidade de detalhes que são apurados e mantidos por esses bancos de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

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

Assim, o objetivo proposto para *big data* nos negócios e na TI será a aplicação de dados e analítica para incrementar a inteligência corporativa (MINELI, CHAMBERS E DHIRAJ, 2013). Esta é uma meta completamente diferente de enquadramento para a tecnologia e vai significar novas formas de organizar e conceituar como ela é financiada e entregue atualmente.

De acordo com relatório publicado pelo *World Economic Forum* (2012), para transformar a mobilidade de dados gerada em uma ferramenta de desenvolvimento econômico, elementos de um ecossistema devem estar presentes e serem tratados. A Figura 03 ilustra os vários tipos de dados, incentivos, e os requisitos de atores no contexto denominado pelo *World Economic Forum* como ecossistema de dados:

Figura 03 - Ecossistema de *big data*.



Fonte: *World Economic Forum* (2012).

De acordo com a Figura 03, para aqueles indivíduos que geram os dados, mecanismos devem ser desenvolvidos para garantir privacidade e segurança. Ao mesmo tempo, modelos de negócios devem ser criados para fornecer os incentivos adequados aos atores do setor privado, para que estes possam compartilhar e utilizar dados para o benefício da sociedade.



O setor público na maioria dos países também mantém enormes conjuntos de dados na forma de dados de censos, indicadores de saúde, impostos e informações sobre despesas. Na *internet* a revolução móvel adicionou ainda outra fonte: dados cedidos por indivíduos voluntariamente ou por meio de *crowdsourcing* - indivíduos contribuem para o processo de coleta de informações, tornando-o mais democrático e transparente. O setor privado, mantém vastas coleções de dados transacionais, muitos dos quais dados criados como um subproduto de outras transações. Com o uso de telefones celulares, grande parte destes dados pode ser associada aos indivíduos e sua localização.

### 2.1.1 Definições para o fenômeno

Ciências como a astronomia e a genômica, que vivenciaram uma explosão informacional nos anos 2000, cunharam o termo *big data* (SCHÖNBERGER-MAYER E CUKIER, 2013). Este termo, relacionado aos grandes volumes de dados, foi primeiramente citado no relatório "*Data, data, everywhere: a special report on managing information*", do periódico britânico *The Economist* (CUKIER, 2010). Entretanto, durante esses anos o termo foi sendo utilizado e relacionado a *datawarehouses* ou soluções de *business intelligence* (BI), com *data sets* de *terabytes* de dados. O fato é que *big data* representa muito mais que isto e atualmente, o conceito está migrando para todos os campos do conhecimento humano (SCHÖNBERGER-MAYER E CUKIER, 2013).

No relatório "*Data: the next frontier for innovation, competition, and productivity*" é apresentada a definição de *big data*, como "o conjunto de dados cujo tamanho vai além da capacidade para capturar, armazenar, gerenciar e analisar de ferramentas de *software* de banco de dados típicos" (MANYIKA ET AL., 2011). Outras definições para o termo são apresentadas na Tabela 02:

Tabela 02 - Definições de *big data*.

(continua)

Autor	Definição
AKERKAR (2014)	<i>Big data</i> refere-se a conjuntos de dados, cujo tamanho está além das capacidades da tecnologia de bancos de dados atuais. É um campo emergente onde a tecnologia inovadora oferece alternativas para resolver os problemas inerentes que aparecem quando se trabalha com dados massivos, oferecendo novas maneiras de reutilizar e extrair valor a partir de informações.

(conclusão)

Autor	Definição
DUMBILL (2012)	<i>Big data</i> são os dados que excedem a capacidade de processamento de dados de sistemas convencionais.
LOHR (2012)	<i>Big data</i> é um termo de <i>marketing</i> , mas também um atalho para o avanço de tendências em tecnologia que abrem a porta a uma nova abordagem para a compreensão do mundo e da tomada de decisões.
MINELI, CHAMBERS E DHIRAJ (2013)	<i>Big data</i> é a próxima geração de <i>data warehousing</i> e análise de negócios e está pronta para entregar receitas economicamente eficientes para as empresas. Este fenômeno se deve, em maior parte, ao rápido ritmo de inovação e mudança que estamos vivenciando hoje.
SATHI (2012)	Existem duas fontes comuns de dados agrupados sob a bandeira <i>do big data</i> . A primeira são os dados internos (dados estruturados, não estruturados ou semiestruturados) da organização que, graças à automação e acesso estão sendo cada vez mais compartilhados. A segunda são os dados de fora da organização, como as informações disponíveis em sites de mídia social, literatura do produto distribuído livremente pelos concorrentes, hierarquias organizacionais dos clientes corporativos, dicas úteis disponíveis a partir de terceiros e reclamações de clientes postados em sites de regulamentação.
SCHÖNBERGER-MAYER E CUKIER (2013)	<i>Big data</i> é a capacidade de uma sociedade de obter informações de maneiras novas a fim de gerar ideias úteis e bens e serviços de valor significativo. Assim, a verdadeira revolução não está nas máquinas que calculam dados, e sim nos dados em si e na maneira como usamos.
ZHAO (2013)	<i>Big data</i> é um sonho tornado realidade para os cientistas de dados, pois se pode ter tudo para obter <i>insights</i> interessantes, que não seriam possíveis antes. <i>Big data</i> não transforma informação e conhecimento sem análises detalhadas. Requer soluções de armazenamento grandes e escaláveis, bem como capacidades e aplicações de análise escaláveis. Análise não significa que se possa jogar dados em alguma <i>machine-learning</i> ou algoritmos estatísticos, tais como redes neurais, árvores de decisão, máquinas de apoio vetor, e assim por diante e esperar ter bons resultados automaticamente.

Fonte: elaborado pelo próprio autor, tradução nossa.

Em meio as definições apresentadas na Tabela 02, pode-se observar algumas similaridades, no que tange aos limitadores tecnológicos de capacidade e processamento e ao surgimento de outros tipos de dados, não estruturados e semiestruturados, que irão subsidiar mudanças no processo de tomada de decisão. Também foi levantado o fato de que apenas os aspectos tecnológicos não serão suficientes para geração de bons resultados.

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

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

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

Atualmente, se inicia mais uma era da informação: a época do *big data*. Entretanto, *big data* não é BI, com "grandes" dados. Por 55 anos as corporações de tecnologia da informação (TI) dominaram uma estrutura transacional de mundo, com o foco na automatização, eficiência e produtividade. Agora *big data* representa uma transição em termos de armazenamento e análise (MINELI, CHAMBERS E DHIRAJ, 2013).

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

Este fenômeno surge viabilizado pelo aumento do poder de processamento que de acordo com a lei de Moore, dobra a cada dois anos. Esta contínua melhoria tornou os computadores mais rápidos, e a memória mais profusa. Entretanto, muitos dos ganhos com *big data*, acontecem não por causa de *chips* mais rápidos ou melhores algoritmos, mas sim pela existência de mais dados (SCHÖNBERGER-MAYER E CUKIER, 2013). Mineli, Chambers e Dhiraj (2013) compartilham este pensamento e exploram mais profundamente a questão, ao investigar as razões para o surgimento desta nova era: computação, dados e convergência.

Os autores explicam que em termos de computação, *big data* é o resultado natural das quatro maiores tendências globais: a lei de Moore, computação móvel (com *smartphones* e *tablets*), redes sociais (Facebook, FourSquare e outros), e a computação em nuvem (*cloud computing* - possibilidade de alugar ou arrendar *hardware* ou *software* para utilização). Em termos de dados, os volumes de dados transacionais, que foram por décadas utilizados pela maioria das grandes empresas, se apresentam atualmente com mais volume, velocidade e variedade - os três Vs - de dados que chegaram de forma inédita. Esta tempestade perfeita dos três Vs, torna extremamente complexo e pesado o gerenciamento de dados atual e análise de tecnologias e práticas.

Assim é consenso entre os autores o fato de que a evolução tecnológica em termos de processamento e a existência de mais dados, tenham sido elementos fundamentais para o surgimento de *big data*. Mineli, Chambers e Dhiraj (2013) ainda complementam o cenário abordando a questão da convergência. O gerenciamento de dados tradicionais, análise de *software*, tecnologias de *hardware*, tecnologia de código aberto e *commodities* de *hardware* estão se fundindo para criar novas alternativas para TI e executivos de negócios.

Sathi (2012) também menciona os aspectos tecnológicos e dá ênfase a dois outros fatores: consumidores mais analíticos e conectados e monetização, ou seja, a utilização da análise grandes volumes de dados externos, onde as informações dos clientes são recolhidas, trocadas e vendidas. Estes são eventos não vistos em épocas anteriores e fornecem suporte ao atual volume, velocidade e variedade, características de *big data*.

### **2.1.2 Características de *big data***

Conforme mencionado anteriormente, desde o início da aplicação de computadores para problemas de negócios em 1950, e metade da década de 1960 quando os primeiros

sistemas de gerenciamento de banco de dados bem sucedidos surgiram, tem havido um aumento constante na quantidade de dados armazenados e o reconhecimento do valor de que os dados vão além da simples informatização das tarefas de rotina. Assim, o volume de dados tem aumentado de forma constante e substancial até o presente momento. Segundo Betser e Belanger (2013), a combinação da *internet* e a *World Wide Web* (WWW) em meados dos anos 1990, sinalizaram uma mudança radical não só na quantidade de dados (o volume), mas também na taxa com que os dados chegavam (a velocidade), na diversidade de fontes de onde eles chegavam e nos tipos de dados e sua disponibilidade (a variedade).

De acordo com Mineli, Chambers e Dhiraj (2013), essas são dimensões razoáveis para quantificar *big data* e levam em conta as medidas típicas em torno do volume e variedade, introduzindo a dimensão de velocidade, que é um fator chave de composição. Em artigo publicado por McAfee e Brynjolfsson (2012) é realizada uma comparação entre *big data* e analítica. Os autores esclarecem que de fato o *big data* assim como a analítica antes dele, tem foco na extração de informações em grandes volumes de dados e na transformação destes em vantagem competitiva para as organizações. Entretanto, volume, velocidade e variedade são apontados como os atuais elementos de diferenciação.

#### 2.1.2.1 Volume

De acordo com Dumbill (2012), volume é a dimensão mais facilmente perceptível de *big data*, devido à grande quantidade de dados continuamente recolhida e armazenada para diferentes usos e efeitos. O principal desafio relacionado ao grande volume de dados, considerando mecanismos eficazes de armazenamentos de dados, é a escalabilidade. Ou seja, a capacidade que um sistema possui de se adequar e atender de forma eficaz um aumento específico de demanda. As decisões de armazenamento influenciam a recuperação de dados, o objetivo final para o usuário, que espera que ela seja realizada o mais rápido possível, especialmente em sistemas de tempo real.

Mineli, Chambers e Dhiraj (2013) elucidam que o volume de dados pode ser medido pela grande quantidade de transações, eventos ou histórias e é muitas vezes agravado pelos atributos, dimensões ou variáveis preditivas. Normalmente, a análise tem usado conjuntos de dados menores - amostras, para criar modelos preditivos. Muitas vezes, os negócios usam casos ou conhecimento preditivo já que o volume de dados tem sido limitado, devido ao armazenamento ou restrições de processamento computacional.

A atual exaustão de dados, de acordo com McAfee e Brynjolfsson (2012), tem gerado cerca de 2,5 *exabytes* de dados criados diariamente, e este número dobra a cada 40 meses. Assim, o volume de dados que circula a cada segundo pela *internet* é maior que toda a informação armazenada na rede 20 anos atrás. Radfahrer (2012) complementa que existem cerca de 2 bilhões de pessoas *on-line*. Serviços de telefonia celular atingem a marca de 5,3 bilhões de usuários, em um mercado ainda em crescimento.

A Amazon.com, maior livraria do mundo, vende 180 *e-books* para cada livro em capa dura. O *site* de serviços de rede Facebook, se fosse um país, seria atualmente o terceiro mais importante do mundo em termos de população. Duhigg (2012) traz informações de que, em quinze minutos, a humanidade gera o triplo de informações disponíveis no acervo da Biblioteca do Congresso Americano, a maior do mundo. A Tabela 03 demonstra a evolução ocorrida em termos do volume de dados gerados desde o surgimento dos primeiros computadores, no início dos anos 1940:

Tabela 03 - Volume de informações geradas.

(continua)

<i>Bit</i>	Abreviação do dígito binário, a linguagem dos computadores: sequência de números 0 e 1 que guarda informações codificadas em HDs de computador.
<i>Byte</i>	Equivale a 8 <i>bits</i> e é o suficiente para armazenar um caractere de texto no PC.
<i>Kilobyte</i> (1.000 bytes)	A informação contida em uma página de livro.
<i>Megabyte</i> (1.000.000 de bytes)	Armazena um quinto de toda a obra de William Shakespeare.
<i>Gigabyte</i> (1.000.000.000 de bytes)	Uma hora de vídeo em baixa resolução.
<i>Terabyte</i> (1.000.000.000.000 de bytes)	385 <i>terabytes</i> guardam todo o catálogo da Biblioteca do Congresso americano, a maior do mundo.
<i>Petabyte</i> (1.000.000.000.000.000 de bytes)	1,5 <i>petabyte</i> armazena todas as músicas á criadas pela humanidade.
<i>Exabyte</i> (1.000.000.000.000.000.000 de bytes)	3 <i>exabytes</i> é tudo o que a humanidade conseguia guardar em 1986 - hoje produzimos quase o dobro disso em dois dias.
<i>Zettabyte</i> (1.000.000.000.000.000.000.000 de bytes)	48 bilhões de iPads, que montam um muro de 17 metros de altura com 6.400 quilômetros de extensão (o raio da Terra); 1,8 <i>zettabyte</i> armazena todos os dados acumulados pela civilização em um ano.

(conclusão)

<i>Yottabyte</i> (1.000.000.000.000.000.000.000.000.000.000 de <i>bytes</i> )	Encheria a Grande Pirâmide Gizé com <i>memory cards</i> (minúsculos HDs) de 64 <i>gibabytes</i> de capacidade.
--	--

Fonte: Adaptado de Petry (2013).

Assim, eliminando-se a restrição do volume de dados e utilizando-se conjuntos de dados maiores, as empresas podem descobrir padrões sutis que podem levar à decisões, ou podem levar em consideração mais observações ou variáveis em previsões, aumentando a precisão dos modelos de previsão. Além disso, pela liberação dos dados, as empresas podem olhar para os dados ao longo de um período maior de tempo para criar previsões mais precisas que refletem às complexidades do mundo real de *bits* inter-relacionados de informações MINELI, CHAMBERS E DHIRAJ (2013).

#### 2.1.2.2 Velocidade

De acordo com Dumbill (2012), velocidade descreve o fluxo de dados, a taxas elevadas, em um cenário cada vez mais distribuído. Atualmente, a velocidade aumenta de forma semelhante ao volume. *StreamProcessing* é o principal desafio relacionado com esta dimensão, porque o armazenamento seletivo é obrigatório para prática de gestão de volume, mas também para a resposta em tempo real.

Mineli, Chambers e Dhiraj (2013) apontam que a velocidade está relacionada à rapidez com a qual os dados são criados, acumulados, internalizados e processados. O ritmo crescente do mundo colocou demandas de negócios para processar a informação e dar respostas em tempo real. Assim, os dados são processados enquanto ocorre o fluxo de tomada de decisões ou são processados em lote para produzir decisões mais oportunas.

McAfee e Brynjolfsson (2012) esclarecem que a informação em tempo real, ou quase, permite à empresa ser muito mais ágil do que a concorrência. Na Tabela 04, pode-se observar a evolução em termos da velocidade com que os dados são processados:

Tabela 04 - Velocidade em que os dados são processados.

(continua)

<i>Flops</i>	Unidade para calcular a velocidade de processamento de computadores: equivale à capacidade de realizar um cálculo matemático simples em um segundo.
<i>Kiloflops</i> (1.000 <i>flops</i> )	Processamento de um supercomputador em 1951.

(conclusão)

<i>Megaflops</i> (1.000.000 de <i>flops</i> )	Velocidade alcançada por computadores pessoais no início dos anos 90.
<i>Gigaflops</i> (1.000.000.000 de <i>flops</i> )	É nesta faixa que operam os PCs mais comuns, residenciais.
<i>Teraflops</i> (1.000.000.000.000 de <i>flops</i> )	Capacidade da próxima geração de videogames. O PlayStation 4 e o Xbox 720, que serão lançados no ano de 2013.
<i>Petaflops</i> (1.000.000.000.000.000 de <i>flops</i> )	Apenas supercomputadores chegam a esse patamar: o mais poderoso deles, o americano Titan, roda com 27 <i>petaflops</i> .
<i>Exaflops</i> (1.000.000.000.000.000.000 de <i>flops</i> )	Pelo mais conhecido parâmetro da computação, a Lei de Moore, que estima que a capacidade de processamento dobra a cada dois anos, um supercomputador deve alcançar essa velocidade na década de 2020.

Fonte: Adaptado de Petry (2013).

Assim, a evolução em termos de rapidez na disponibilidade da informação, apresentada na Tabela 04, pode auxiliar organizações em termos competitivos. McAfee e Brynjolfsson (2012) citam o exemplo de um experimento realizado pelo MIT *Media Lab*, que usou dados de localização de celulares para inferir quanta gente havia nos estacionamentos da loja de departamentos americana Macy's durante a chamada *Black Friday*, que inaugura a temporada natalina de compras nos Estados Unidos. Com a utilização destes dados, foi possível estimar as vendas da varejista naquele dia, mesmo antes destas serem digitadas.

### 2.1.2.3 Variedade

Dumbill (2012) coloca que variedade refere-se a vários graus de estrutura (ou a falta dela), podendo vir de várias origens (por exemplo, ciências, política, economia, redes sociais ou *logs* de servidor *web*, entre outros) e cada um descreve sua própria semântica, portanto, os dados seguem uma modelagem estrutural específica. O principal desafio da grande variedade de dados é atingir um mecanismo eficaz para a ligação de diversas classes de dados diferentes na estrutura interna.

Mineli, Chambers e Dhiraj (2013) esclarecem que tradicionalmente os dados, especialmente os operacionais, são estruturados como em um banco de dados com base em seu tipo. Entretanto, ao longo das duas últimas décadas, os dados estão se tornando não estruturados e as fontes de dados têm se proliferado além das aplicações operacionais. Muitas vezes, texto, áudio, vídeo, imagem, dados geoespaciais, e os dados de *internet* são



considerados dados não estruturados. No entanto, uma vez que muitas das fontes desses dados são programas, os dados são na realidade semiestruturados. Estes podem ser uma combinação de diferentes tipos de dados que tem um padrão ou estrutura que não é estritamente definida como estruturada.

Assim, a principal fonte de dados utilizada no processamento analítico eram os dados tradicionalmente focados em sistemas operacionais, como ERP e CRM. No entanto, a fim de aumentar o conhecimento e a consciência, a complexidade das fontes de dados que alimentam os processos de análise estão crescendo rapidamente para incluir uma ampla variedade, tais como:

- Dados na *internet* (fluxo de cliques, mídias sociais, redes sociais, links de redes sociais);
- Pesquisa primária (pesquisas, experimentos, observações);
- Pesquisa secundária (dados da concorrência e do mercado, relatórios da indústria, dados de consumo, dados de negócio);
- Dados de localização (dados de dispositivo móveis, dados geoespaciais);
- Dados de imagem (vídeo, imagem de satélite, vigilância);
- Dados da cadeia de suprimentos (EDI, catálogos de fornecedores e preços, informação de qualidade);
- Dados do dispositivo (sensores, PLCs, dispositivos de RF, telemetria).

McAfee e Brynjolfsson (2012) expõe que muitas das fontes mais importantes dessa profusão de dados são relativamente novas: os *sites* de serviços de rede sociais como Facebook e Twitter são de 2004 e 2006; e, dispositivos como iPhone e iPad são de 2008 e 2010. De acordo com Mineli, Chambers, Dhiraj (2013), esta grande variedade de dados conduz a complexidades na entrada e armazenamento dos dados. A variedade também torna complicada a mudança de dados para um formato que possa ser usado no processamento e análise.

Isto porque, bancos de dados que antes armazenavam informações da base transacional das empresas, hoje são inadequados para armazenar e guardar *big data*. Entretanto, Radfaher (2012) estima que a capacidade computacional atual cresça cerca de um milhão de vezes nos próximos trinta anos. Na Tabela 05 são expostas as principais fontes de onde podem provir informações:

Tabela 05 - De onde vem a informação a cada dia.

Humanidade	2,5 <i>exabytes</i> de informação são produzidos pela humanidade a cada dia
Famílias	375 <i>megabytes</i> de dados são acumulados por cada família a cada dia
Site do Google	24 <i>petabytes</i> são processados pelo site do Google a cada dia
<i>Smartphones</i> e <i>tablets</i>	43 <i>petabytes</i> de dados são trocados por <i>smartphones</i> e <i>tablets</i> conectados à <i>internet</i> a cada dia
E-mails enviados	10 <i>petabytes</i> correspondem aos e-mails enviados a cada dia

Fonte: Adaptado de Petry (2013).

Adicionalmente, McAfee e Brynjolfsson (2012) esclarecem que com mais atividades empresariais sendo digitalizadas, novas fontes de informação e equipamentos cada vez mais baratos irão se unir para inaugurar uma nova era de grandes volumes de informação digital, sobre praticamente todos os assuntos.

#### 2.1.2.4 Outras características de *big data*

Uma das características mais perceptíveis de *big data* é o volume sempre em expansão de muitos *terabytes*, ou mesmo *petabytes* de dados. Entretanto, Beyer et al. (2011) esclarece que o desafio empresarial completo vai muito além do volume e é definido como *extreme information* (informação extrema). Neste contexto, os autores propõem o gerenciamento de *big data* não como um silo em separado, mas de forma holística, em conjunto com o aumento da variedade, da complexidade e da velocidade de informação.

Desta forma, Beyer et al. (2011) vem a confirmar as características antes citadas de velocidade e variedade e acrescenta a dimensão complexidade a este processo. Nestes termos, complexidade tem seu significado nas relações entre as fontes de informação da empresa. Consiste na capacidade de integrar *big data* com outras fontes de informação relevantes por meio de relações complexas de dados.

Sathi (2012) também agrega a dimensão veracidade ao contexto de *big data*. Na definição do autor, veracidade representa tanto a credibilidade das fontes de dados, bem como a adequação dos dados para o público-alvo. Assim, ao contrário dos dados internos da organização, cuidadosamente regulados, a grande maioria de dados de *big data* é proveniente de fontes fora de controle e, portanto, com problemas significativos de correção ou de precisão.

### 2.1.3 Tecnologia da informação aplicada a *big data*

De acordo com O'Brien (2010) a transformação de dados em informação se dá por meio de atividades de processamento como cálculo, comparação, separação, classificação e resumo. Essas atividades organizam, analisam e manipulam dados, convertendo-os em informação para os usuários finais. Assim, de acordo com o autor, a informação é transmitida de várias formas e torna-se meta dos sistemas produzir a informação apropriada para os usuários finais.

Para O'Brien (2010), a qualidade da informação deve ser dotada de três dimensões: tempo, conteúdo e forma. A Tabela 06 resume os atributos importantes da informação e os agrupa nessas três dimensões:

Tabela 06 - Resumo dos atributos de qualidade da informação.

<b>Dimensão do tempo</b>	
Prontidão	A informação deve ser fornecida quando for necessária
Aceitação	A informação deve estar atualizada quando for fornecida
Frequência	A informação deve ser fornecida tantas vezes quantas forem necessárias
Período	A informação pode ser fornecida sobre períodos passados, presentes e futuros
<b>Dimensão do Conteúdo</b>	
Precisão	A informação deve estar isenta de erros
Relevância	A informação deve estar relacionada às necessidades de informação de um receptor específico para uma situação específica
Integridade	Toda a informação que for necessária deve ser fornecida
Concisão	Apenas a informação que for necessária deve ser fornecida
Amplitude	A informação pode ter um alcance amplo ou estreito, ou um foco interno ou externo
Desempenho	A informação pode revelar desempenho pela mensuração das atividades concluídas, do progresso realizado ou dos recursos acumulados
<b>Dimensão da Forma</b>	
Clareza	A informação deve ser fornecida de uma forma que seja fácil de compreender
Detalhe	A informação pode ser fornecida em forma detalhada ou resumida
Ordem	A informação pode ser organizada em uma sequência predeterminada
Apresentação	A informação pode ser apresentada em forma narrativa, numérica ou gráfica
Mídia	A informação pode ser fornecida na forma de documentos em papel impresso, monitores de vídeo ou outras mídias

Fonte: O'Brien (2010).

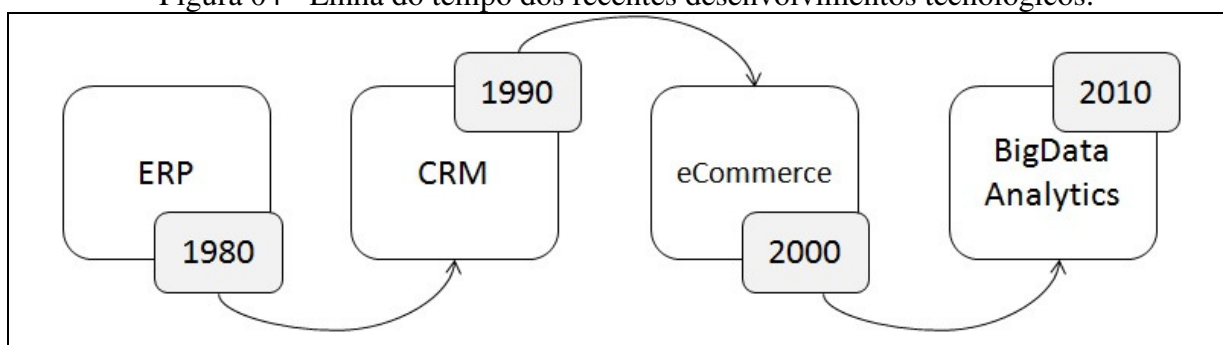
Assim, há três papéis vitais que os sistemas de informação podem desempenhar em uma empresa: (1) suporte de seus processos e operações; (2) suporte na tomada de decisões de seus funcionários e gerentes; e, (3) suporte em suas estratégias em busca de vantagem competitiva (O'BRIEN, 2010). Para isso, há um número crescente de tecnologias usadas para

agregar, manipular, gerenciar e analisar grandes bases de dados (MANYIKA ET AL., 2011). Mineli, Chambers e Dhiraj (2013) definem três estágios para a evolução dos sistemas de base de dados: dependente (primeiros tempos), independente (últimos anos) e interdependente (era *big data*).

O estágio "dependente" refere-se à época em que os sistemas de informação eram bastante novos e os usuários não sabiam muito bem o que queriam, fazendo com que áreas de TI desenvolvessem as soluções de forma isolada. Na época do estágio "independente" os usuários passaram a trabalhar junto com a TI no desenvolvimento de uma plataforma analítica alinhada às necessidades de negócio e uma abordagem que gerasse *insights* para a empresa. Atualmente a era *big data*, "interdependente" caracteriza o estágio interacional entre várias empresas, criando mais colaborações sociais que vão além de sua estrutura física.

Mineli, Chambers e Dhiraj (2013) colocam alguns marcos importantes com relação à evolução dos sistemas de informação ao longo das décadas, que podem ser visualizados na Figura 04:

Figura 04 - Linha do tempo dos recentes desenvolvimentos tecnológicos.



Fonte: Mineli, Chambers e Dhiraj, 2013.

A Figura 04 expõe uma mudança fundamental nos tipos de dados utilizados pelos diferentes tipos de sistemas de informação. Com os sistemas ERP, em 1980, as organizações estavam mais focadas em seus dados internos e, a medida que estes sistemas avançam na linha do tempo, pode-se observar uma maior utilização de dados externos.

De acordo com Manyika et al. (2011), com o fenômeno *big data* as organizações passaram a capturar trilhões de *bytes* de informações sobre seus clientes, fornecedores e operações por meio de sistemas digitais. Milhões de sensores de rede embutidos em telefones celulares, automóveis e outros produtos estão continuamente criando e comunicando dados. O resultado projetado é de 40% no crescimento anual do volume de dados gerados. Estes dados podem ser estruturados, não estruturados e semiestruturados. Os estruturados são de fácil

definição, armazenamento e análise, e podem ser processados por computador (MINELI, CHAMBERS E DHIRAJ, 2013).

Os dados não estruturados representam as informações que não têm um modelo de dados pré-definido e/ou não se encaixam bem em um banco de dados relacional. Informações não estruturadas são tipicamente representadas por texto, mas podem conter dados como datas, números e fatos que as pessoas leem, escrevem e falam. Os dados semiestruturados não se encaixam em uma estrutura formal de modelos de dados. No entanto, contêm *tags* que separam elementos semânticos, e possibilitam a construção de hierarquias dentro dos dados (MINELI, CHAMBERS E DHIRAJ, 2013).

Infelizmente, nenhum sistema de computador hoje pode traduzir de forma confiável linguagem não estruturada para os formatos de bancos de dados estruturados, planilhas, bem como a *web* semântica. Mas de acordo com Majumdar e Sowa (2013), eles podem trabalhar no processamento destes dados, e com isto estão se tornando mais versáteis.

Quando proposto pelo pesquisador da IBM, Edgar F. Codd, em 1969, o modelo relacional foi muito bem sucedido, tanto que é utilizado até hoje. Entretanto, ele foi desenhado para dados estruturados, gerados pelos sistemas internos das organizações. Para tratar com dados não estruturados em volume, velocidade e variedade, é necessário que haja uma evolução tecnológica (TAURION, 2013). As tecnologias que sustentam *big data* podem ser analisadas sob duas óticas: a) as tecnologias de infraestrutura, como bancos de dados *NoSQL*, que armazenam e processam *petabytes* de dados e b) as envolvidas com análise, como *Hadoop* e *MapReduce*.

Sob o ponto de vista de infraestrutura, a linguagem mais comum para acessar bancos de dados, há muito tempo é a SQL, ou "linguagem de consulta estruturada". Mas em função do *big data*, tem surgido o *NoSQL*, que não requer uma estrutura pré-determinada para funcionar. Ele aceita dados de vários tipos e tamanhos e permite que estes sejam vasculhados com sucesso. Em troca da confusão estrutural, esses bancos de dados requerem mais recursos de processamento e armazenagem, que por sua vez, representam custos cada vez menores (SCHÖNBERGER-MAYER E CUKIER, 2013).

Em termos de análise, o *Hadoop* é um projeto da comunidade Apache, criado pelo Yahoo em 2005. Consiste em uma combinação de dois projetos separados, que são o *Hadoop MapReduce* (HMR), que é um *framework* para processamento paralelo e o *Hadoop Distributed File System* (HDFS). O HMR é um *spinoff* do *Map Reduce*, *software* que o Google usa para acelerar as pesquisas endereçadas ao seu buscador. O HDFS é um sistema de

arquivos distribuídos e otimizados para atuar em dados não estruturados e é também baseado na tecnologia do Google, ou seja, é o sistema que armazena os dados para o *Hadoop* (TAURION, 2013).

Desta forma, os dados HDFS são divididos em pequenos pedaços, chamados de blocos e distribuídos por diversos servidores. Isto faz com que o processamento seja acelerado, pois ao invés de uma pesquisa sequencial, os dados são pesquisados de forma simultânea. O outro componente do *Hadoop* é o *Map Reduce*. Este paradigma de programação é o coração do *Hadoop* e permite escalabilidade paralela em centenas ou milhares de servidores. A primeira tarefa executada pelo *Hadoop* com *Map Reduce* é o mapeamento dos dados, ou seja, o acesso a um conjunto de dados e a sua conversão em outro conjunto, onde os elementos individuais são quebrados em tuplas<sup>5</sup> (pares chave/valor). A tarefa *Reduce* pega o resultado do mapeamento e combina estas tuplas em um conjunto menor, obtendo o resultado (TAURION, 2013).

O *StreamProcessing* é outra tecnologia que merece destaque em *big data*. O conceito foi implementado por uma solução da IBM chamada *InfoSphereStreams*. Trata-se de um *middleware* ou uma infraestrutura de *software*, para o desenvolvimento de aplicações que processam dados que são gerados continuamente. Ao contrário do *Hadoop*, que utiliza processamento paralelo, o *StreamProcessing* analisa dados em movimento. Isto quer dizer, análise de imensos volumes de dados no momento em que estão sendo gerados, como uma corrente de dados que passa pelas regras de negócio de modo que se possa analisar, entender em tempo real, e definir ações em resposta (TAURION, 2013).

Assim, segundo Taurion (2013), no modelo de *data mining* tradicional, uma empresa filtra dados dos seus vários sistemas e após criar um *data warehouse*, dispara *queries*. Na prática, faz-se a garimpagem em cima de dados estáticos, com *StreamProcessing*, isto ocorre em tempo real. Coloca-se uma corrente contínua de dados (*streaming data*) atravessando um conjunto de *queries*. Um exemplo é o controle automático de tráfego. Nele é possível coletar dados de sensores, de câmeras de vídeo, dados históricos e alertas de acordo com um determinado algoritmo, identificando a probabilidade de um futuro congestionamento. Com isto, pode-se atuar de forma preditiva, alterando a temporização dos semáforos e desviando o fluxo de veículos para rotas alternativas. Este é um caso real, desenvolvido pela IBM para a Autoridade de Trânsito de Singapura (TAURION, 2013).

---

<sup>5</sup> Tuplas ou registros representam Cada linha formada por uma lista ordenada de colunas. (TAURION, 2013)

Outra tecnologia que vai avançar muito nos próximos anos será a das técnicas de visualização e animação de dados. A visualização permite potencializar a apropriação da informação pelo usuário, por meio de recursos gráficos. Entre as diversas técnicas, pode-se citar as *tag clouds*, que é a visualização em forma de lista visual ponderada, na qual as palavras que aparecem com maior frequência são visualizadas em caracteres maiores e destacados (TAURION, 2013).

Também, pode-se citar o *clustergram*, usado para visualizar análises de *cluster* (*cluster analysis*) ou agrupamentos. Esta técnica classifica os objetos em diferentes grupos, cada um dos quais contendo os objetos semelhantes segundo alguma função de distância estatística. Outra técnica é a chamada *history flow*, que mostra a evolução de um documento à medida que ele é modificado por seus diversos colaboradores. A técnica *spatial information flow* visualiza de forma espacial informações específicas.

Da mesma forma, o campo de interfaces de voz deve avançar nos próximos anos, com o usuário perguntando e o computador respondendo em viva voz. Um exemplo disto é o supercomputador Watson, apresentado pela IBM em 2011. Este supercomputador foi concebido para entender o sentido da linguagem humana de acordo com o seu contexto, com o objetivo de encontrar a resposta precisa para perguntas complexas (TAURION, 2013).

Betser e Belanger (2013) complementam este cenário, citando outras orientações tecnológicas em *big data: Cloud* - estrutura técnica e econômica de serviços em nuvem (*as a service*), tornaram possível para muitas organizações usar e publicar grandes quantidades de dados e análises; *Bandwidth* - capacidade de mover grandes quantidades de dados, especialmente vídeos. É dependente do aumento de largura de banda disponível; *Mobilidade/Wireless*: acesso móvel a todas as formas de dados, com alta largura de banda, em qualquer lugar, a qualquer hora. Aumenta drasticamente a oferta e procura de dados e informações. Além disso, a comunicação sem fio faz com que muitas das comunicações *machine-to-machine* sejam possíveis.

Assim, a proliferação de dispositivos móveis, plataformas de mídia social e a *Internet das Coisas* - relação simbiótica entre o mundo físico e o mundo digital - desencadearam um aumento no volume de dados gerados, fazendo com que outros modelos fossem desenvolvidos. Isto não significa eliminar os tradicionais sistemas de BI que existem atualmente, mas os arquitetos e os sistemas devem estar preparados para trabalhar com diversos formatos de dados, de relacional a *NoSQL*, bem como integrar tecnologias legadas

equipadas com sistemas que usam bancos de dados relacionais à novos sistemas que usarão tecnologias ainda estranhas à TI, como *Hadoop* (TAURION, 2013).

Akerkar (2014) esclarece que enquanto o volume e a velocidade são preocupações físicas, variedade refere-se a uma questão lógica, principalmente relacionada com a forma como os dados são modelados para permitir a integração efetiva. Assim, quanto mais dados são integrados, mais conhecimento interessante pode ser gerado, aumentando o valor do conjunto de dados resultante. Assim, um dos principais objetivos em processamento de *big data* é aumentar o valor de dados, tanto quanto possível, abordando diretamente a grande variedade de dados com o auxílio de tecnologias semânticas.

Dados semânticos têm sido tradicionalmente relacionados com o conceito de *web* semântica. A *web* semântica é a extensão da WWW, incorporando semânticas processáveis por máquina para seus objetos de informação (páginas, serviços, fontes de dados, etc.). Seus objetivos podem ser resumidos em: dar significado semântico às informações sobre a WWW e tornar os dados semânticos na WWW processáveis por máquinas.

Desta forma, a diferença entre a abordagem de técnicas de recuperação de informação (que atualmente dominam a WWW de processamento de informações) e os bancos de dados é que, no último os dados são estruturados por meio de esquemas que são essencialmente de metadados. Metadados dá o significado (a semântica) de dados, permitindo a consulta estruturada, isto é, a consulta de dados com significado e precisão lógica.

Na WWW a semântica dos dados é dada por seres humanos (ou diretamente durante a navegação manual e pesquisa, ou indiretamente, através de algoritmos de recuperação de informação que usam o *feedback* humano introduzido através de *links* estáticos ou registros de interações. Este processo, embora bem sucedido tem suas limitações e é crucial para *big data*, para automatizar o processo de entendimento (dar sentido) dos dados sobre a WWW.

Assim, a *Web Semantic Community* e o *World Wide Consortium* (W3C) têm se dedicado a desenvolver modelos e linguagens para representar a semântica e os protocolos e linguagens para consultá-la (AKERKAR, 2014). Contudo, a tecnologia de *big data* é a parte fácil - a parte mais difícil é descobrir para onde ir com a análise de dados e certificar-se de que a empresa tem as pessoas e os processos prontos antes de se comprometer a comprar a tecnologia (MINELI, CHAMBERS E DHIRAJ, 2013).

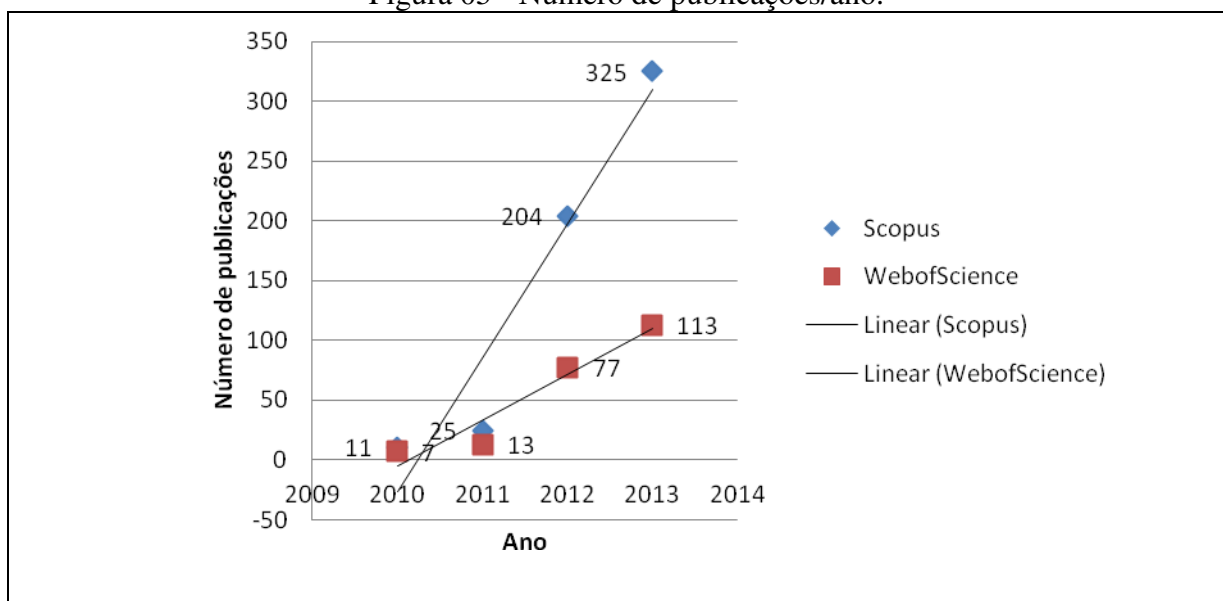


### 2.1.4 Pesquisas em *big data* e análise

No sentido de identificar o número e tipos de publicação que estão sendo desenvolvidas sobre os temas *big data* e análise, foi realizada uma pesquisa nas bases de dados bibliométricos dos indexadores Scopus e *Web of Science* (WoS), desde o ano de 2010 até dezembro de 2013. Nesta análise bibliométrica o ano de 2010 foi considerado o marco do surgimento do termo *big data*, sua relação com grandes volumes de dados e o processo de análise, considerando Cukier (2010) em relatório publicado no periódico *The Economist*.

A coleta de dados realizada durante o mês de dezembro de 2013, foi feita pela busca de estudos que empregassem os termos *big data* e análise, nos campos "título do artigo, resumo e palavras-chave" no caso do Scopus e no campo "tópico", no caso do WoS. Como resultados deste processo, pode-se identificar um total de 325 publicações na base Scopus e 113 na base WoS, de acordo com a Figura 05:

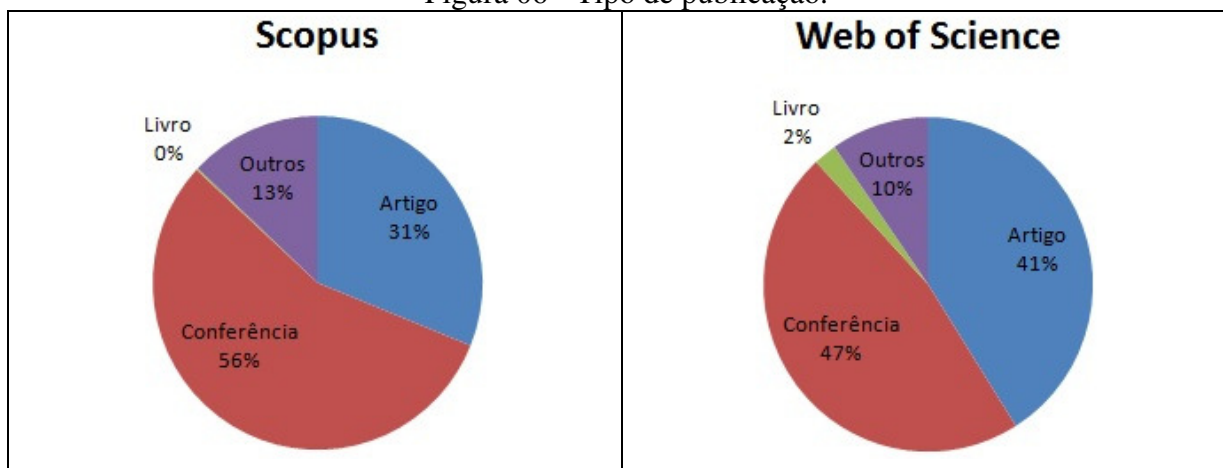
Figura 05 - Número de publicações/ano.



Fonte: elaborado pelo próprio autor.

A Figura 05 mostra um crescimento médio de mais de 200% ao ano em publicações no período analisado. A Figura 06 apresenta os tipos de publicação que estão sendo desenvolvidas em ambos indexadores:

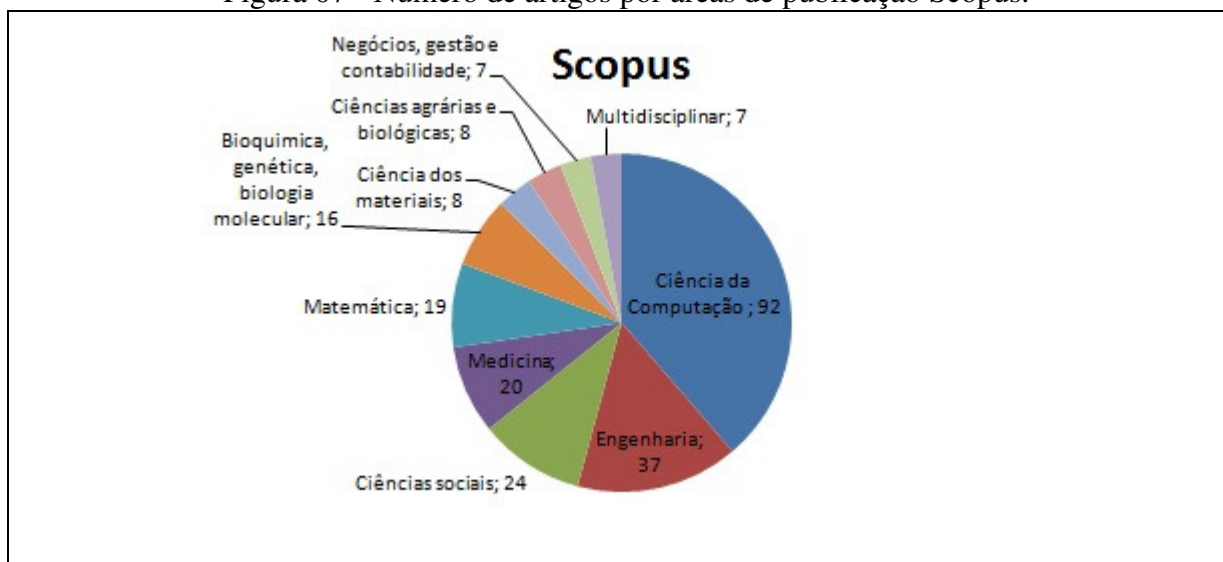
Figura 06 - Tipo de publicação.



Fonte: elaborado pelo próprio autor.

De acordo com a Figura 06, pode-se perceber um número mais elevado de publicações do tipo conferência. Isto caracteriza que os estudos que estão sendo conduzidos para o tema ainda estão em suas fases iniciais de desenvolvimento. Assim, em um refinamento da análise, a Figura 07 apresenta o número de artigos e livros lançados no período por área de publicação do indexador Scopus:

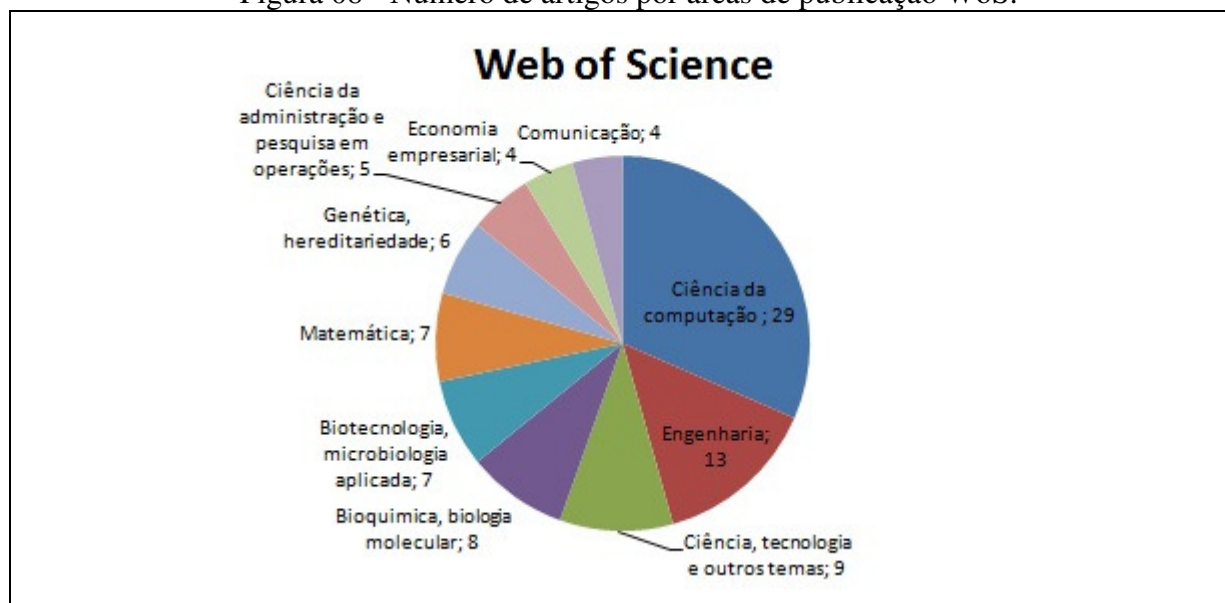
Figura 07 - Número de artigos por áreas de publicação Scopus.



Fonte: elaborado pelo próprio autor.

Os resultados apresentam que 39% das publicações estão concentradas na área de ciência da computação, seguido de engenharia. A Figura 08 apresenta o número de artigos e livros, por área de publicação do indexador WoS, lançados no período analisado:

Figura 08 - Número de artigos por áreas de publicação WoS.



Fonte: elaborado pelo próprio autor.

Nesta análise há uma concentração de mais de 32%, em artigos e livros da área de ciência da computação, seguida da engenharia. Este resultado é muito semelhante ao do indexador Scopus. Em ambas análises, as ciências sociais ocupam posições de menor expressão no contexto analisado, refletindo a carência de pesquisas em *big data* nesta área.

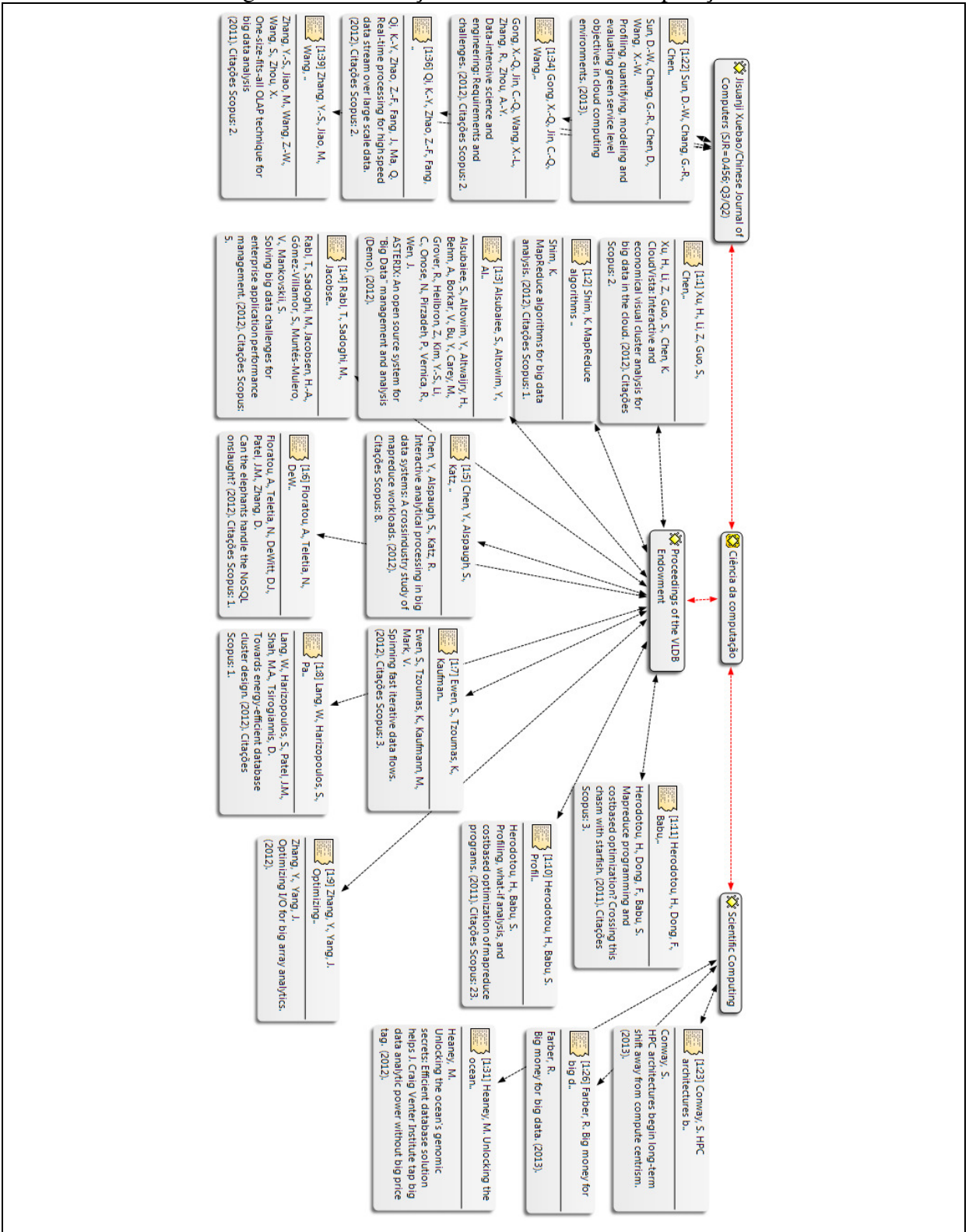
Posteriormente, as publicações identificadas foram analisadas com relação a área, grau de relevância - índices *SCImago Journal Rank* (SJR) da Elsevier, indexador Scopus e *Journal Citation Report* (JCR) da Thompson e Reuters, indexador WoS - *High Quality Publications* (Q1) e índice H. Ainda foi realizada uma análise específica das publicações da área ciências sociais com foco em administração.

O índice SJR, é uma medida de impacto, influência ou prestígio da revista. Expressa o número médio de citações ponderadas recebidas no ano escolhido, pelos documentos publicados nos três anos anteriores. O JCR mede o impacto científico de uma média de artigos publicados em uma revista. O Q1 avalia a revista com relação ao *ranking* de fator de impacto, dentro de determinada área. Assim, revistas consideradas Q1 são aquelas classificadas no primeiro quartil (25%) em suas áreas de publicação<sup>6</sup>.

Com relação ao impacto, influência ou prestígio das fontes analisadas, foram considerados para análise, primeiramente as fontes com mais de duas publicações, totalizando 9 fontes. Assim, a área que mais se destacou foi a de ciência da computação, com 18 publicações. A Figura 09 apresenta as publicações desta área:

6 (<<http://www.scimagojr.com/index.php>>. Acessado em 19 de dezembro de 2013).

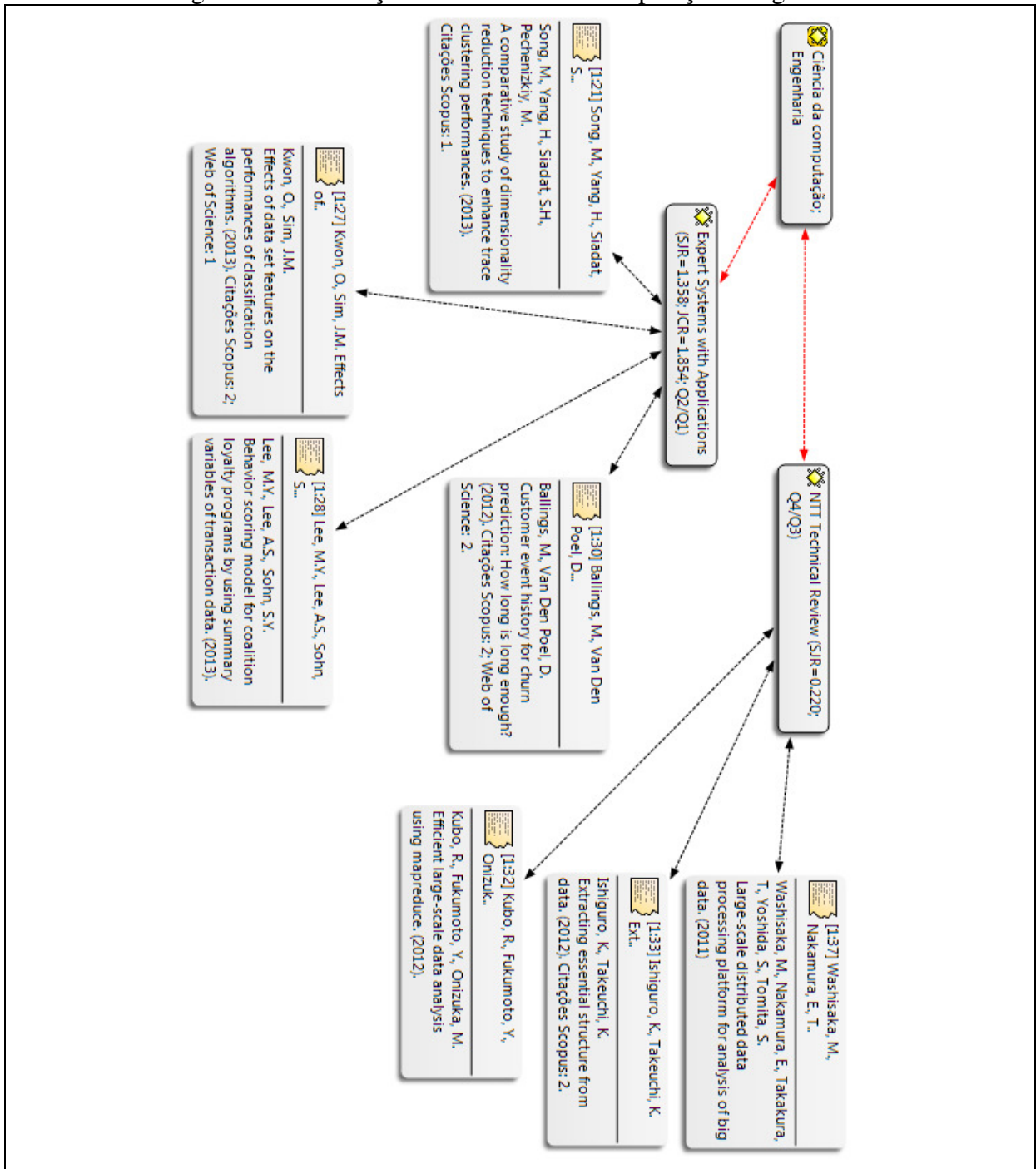
Figura 09 - Publicações área ciência da computação.



Fonte: elaborado pelo próprio autor.

De acordo com a Figura 09, na área de ciência da computação, os *journals* que mais se destacaram foram: *Jisuanji Xuebao/Chinese Journal of Computers*, *Proceedings of the VLDB Endowment* e *Scientific Computing*. Em segundo lugar, apresentou-se a área de ciência da computação e engenharia com 7 publicações. A Figura 10 ilustra as publicações na área:

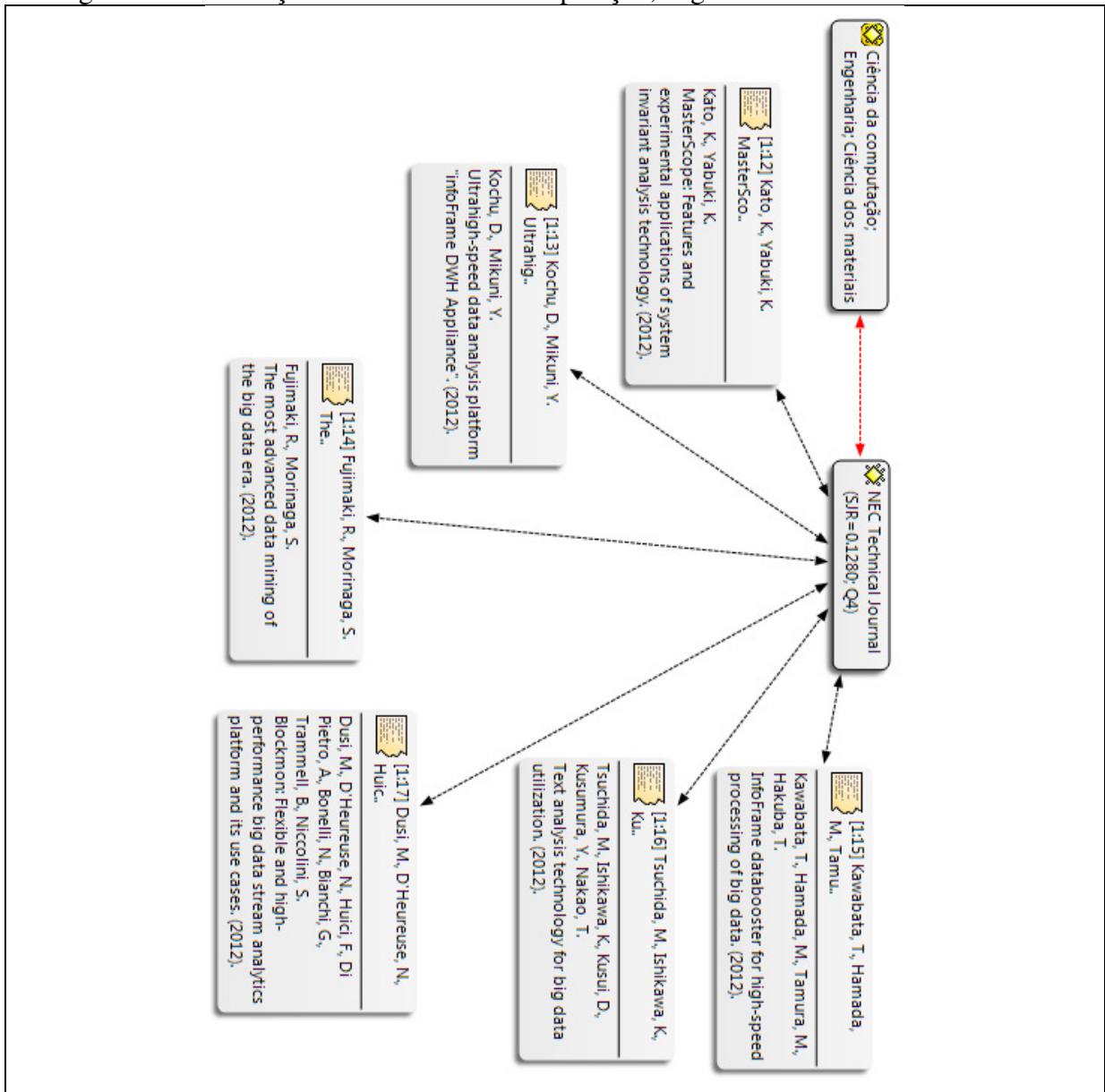
Figura 10 - Publicações área ciência da computação e engenharia.



Fonte: elaborado pelo próprio autor.

Conforme a Figura 10, na área de ciência da computação e engenharia os *journals* que tiveram destaque foram: *Expert Systems with Applications* e *NTT Technical Review*. A próxima área identificada foi a de ciência da computação, engenharia e ciência dos materiais com 6 publicações. A Figura 11 ilustra estas publicações:

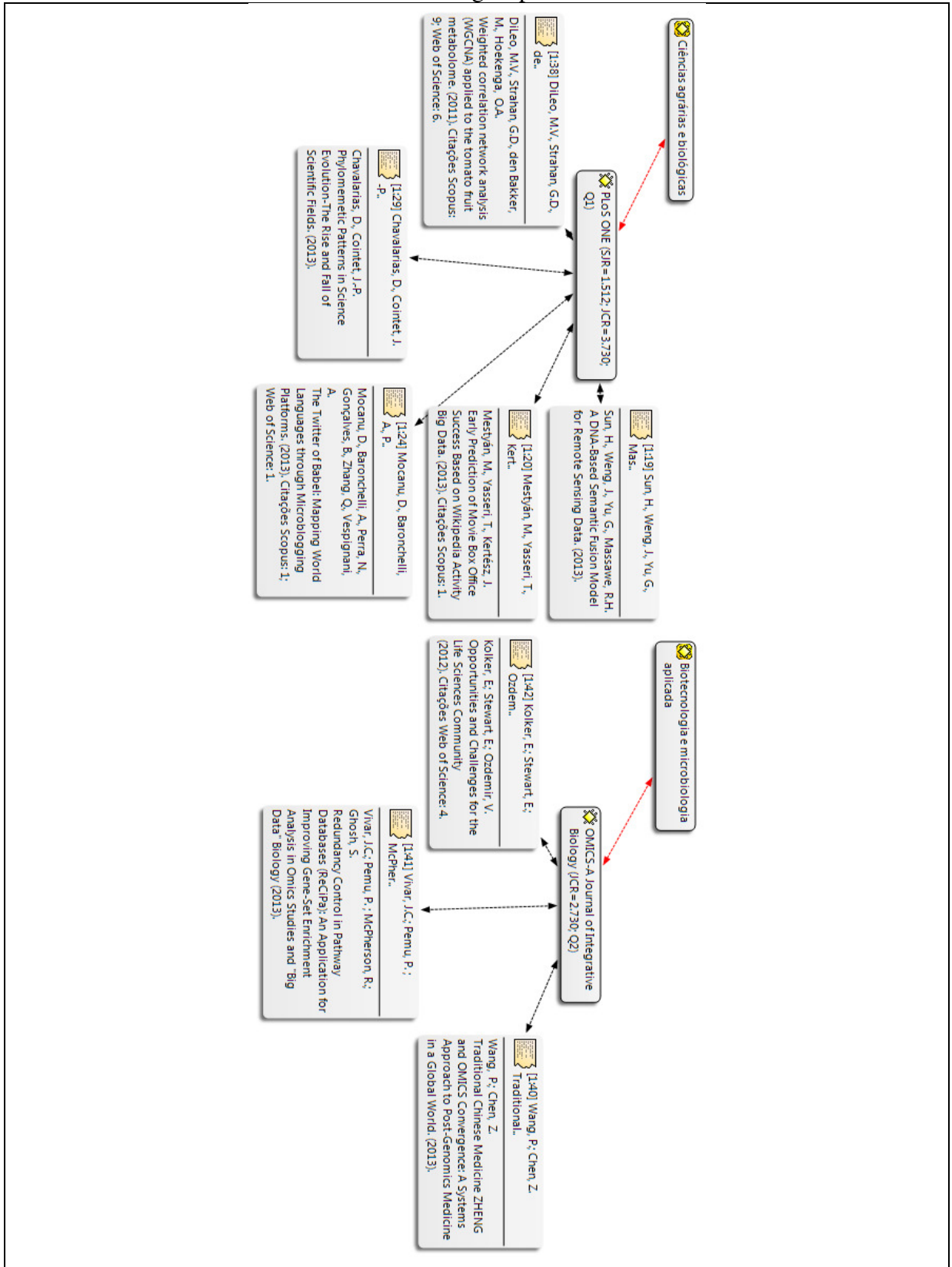
Figura 11 - Publicações área ciência da computação, engenharia e ciência dos materiais.



Fonte: elaborado pelo próprio autor.

Nestas áreas o *journal* que teve destaque foi: *NEC Technical Journal*. Também foram consideradas as áreas de ciências agrárias e biológicas com 5 publicações e biotecnologia e microbiologia aplicada com 3 publicações. A Figura 12 ilustra as publicações nestas áreas:

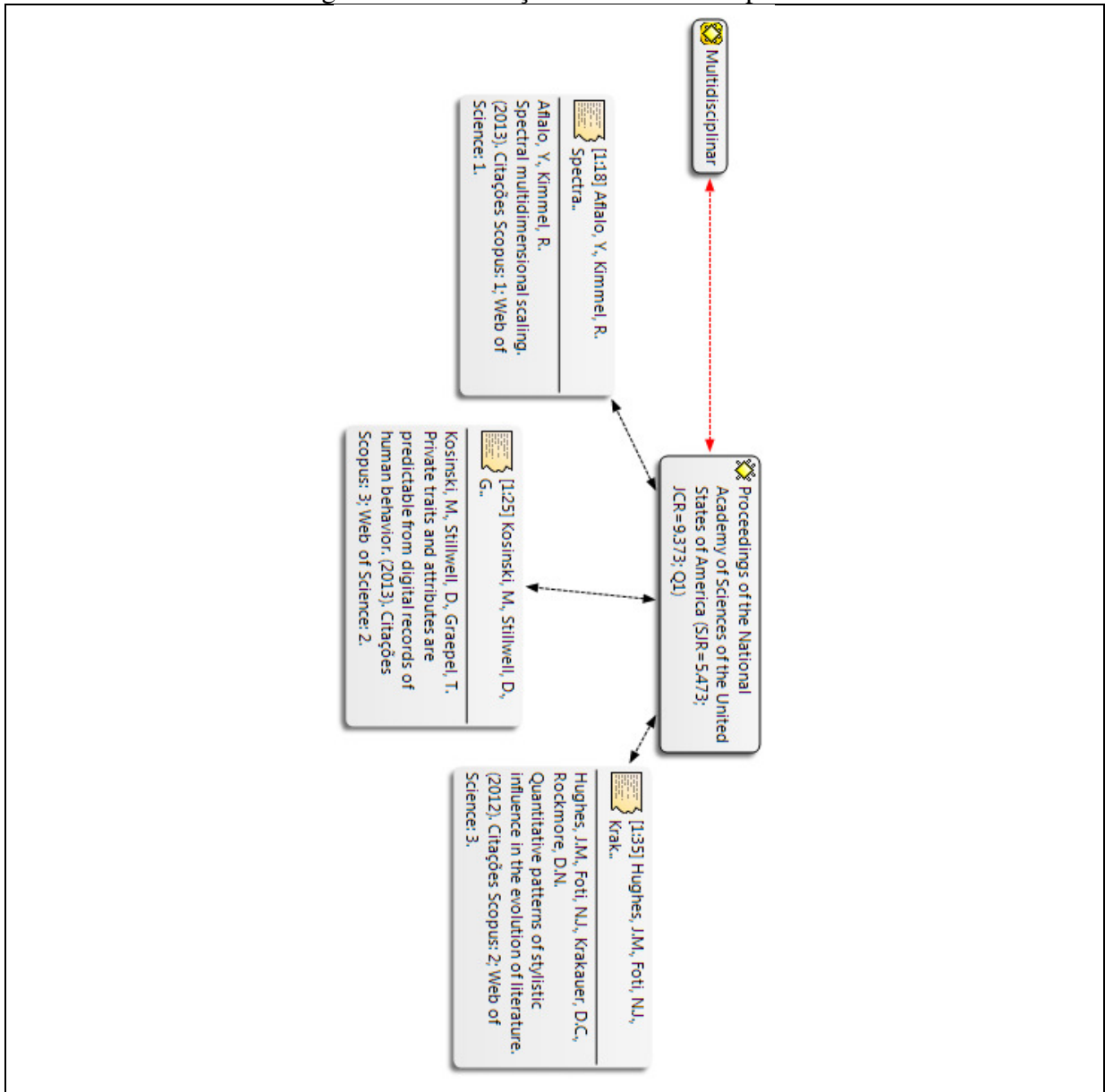
Figura 12 - Publicações áreas de ciências agrárias e biológicas e biotecnologia e microbiologia aplicada.



Fonte: elaborado pelo próprio autor.

Nestas áreas, os *journals* que se destacaram foram: *PloS ONE* e *OMICS-A Journal of Integrative Biology*. Por fim foi analisada a área multidisciplinar, que teve 3 publicações, de acordo com a Figura 13:

Figura 13 - Publicações área multidisciplinar.

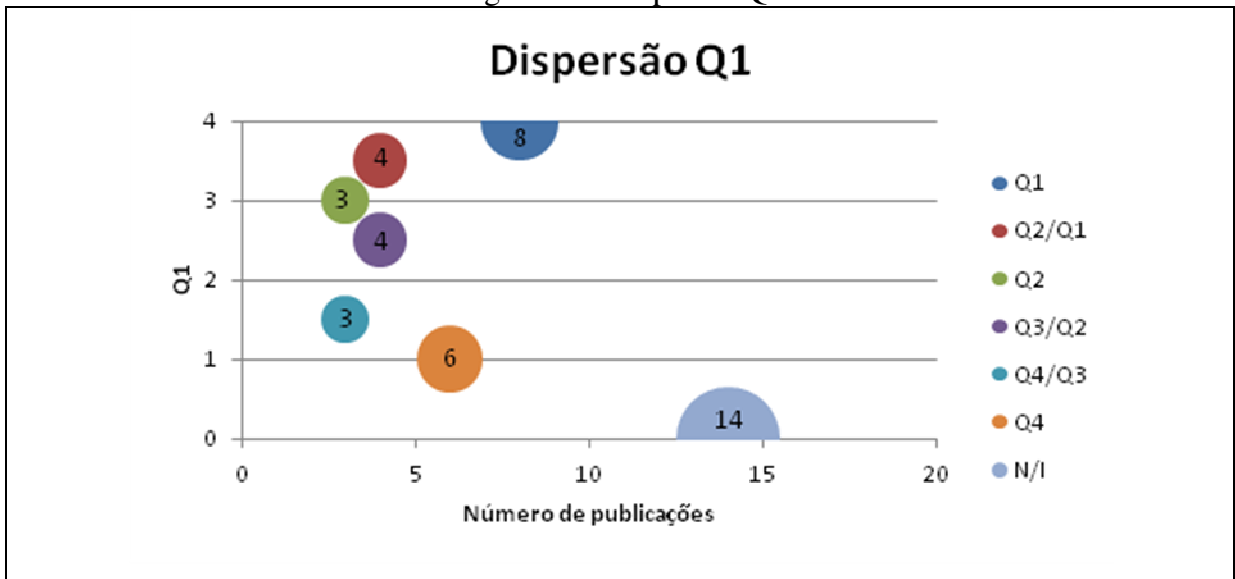


Fonte: elaborado pelo próprio autor.

Na área multidisciplinar, o *journal* que teve destaque foi: *Proceedings of the National Academy of Sciences of the United States*. A Figura 14 apresenta um gráfico de dispersão, onde é possível analisar as fontes de publicação de acordo com o índice Q1, que avalia a revista com relação ao *ranking* de fator de impacto dentro de determinada área:



Figura 14 - Dispersão Q1.

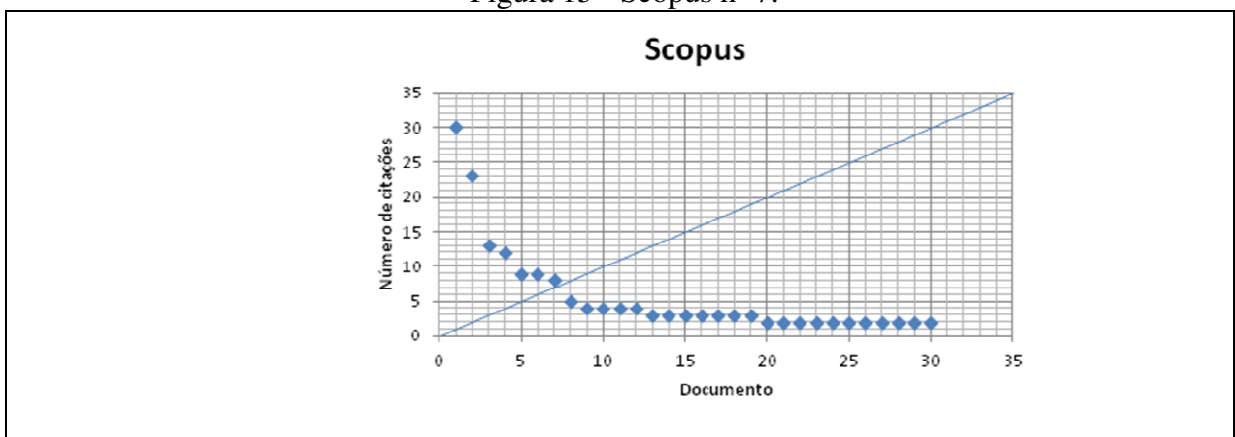


Fonte: elaborado pelo próprio autor.

Assim, pode-se observar que 33% das publicações analisadas não apresentam o índice, enquanto as demais se dividem em publicações do tipo Q1, ou seja, estão entre as 25% com melhor índice em suas áreas e as demais estão em Q4, classificadas no último quartil.

Outra análise realizada foi com relação do índice H, que mede o número de artigos da revista que tenham registro de ao mínimo "h" citações durante o período analisado<sup>7</sup>. A Figura 15 apresenta os dados do indexador Scopus:

Figura 15 - Scopus h=7.

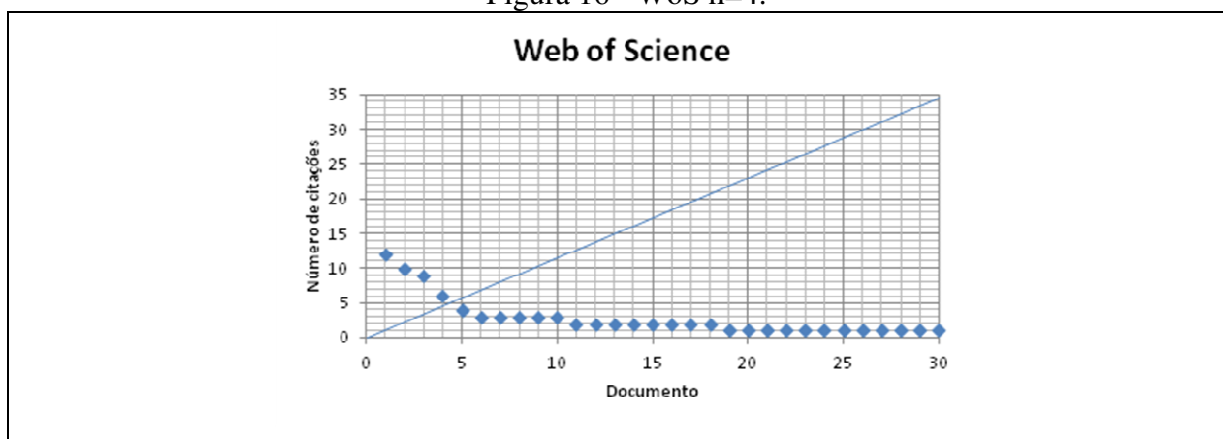


Fonte: elaborado pelo próprio autor.

<sup>7</sup> (<<http://www.scimagojr.com/index.php>>. Acessado em 19 de dezembro de 2013).

Com relação a este índice, o indexador Scopus apresentou um índice  $h=7$ , que significa que houveram 7 artigos com mais de 7 citações. A Figura 16 apresenta os dados do indexador WoS:

Figura 16 - WoS  $h=4$ .



Fonte: elaborado pelo próprio autor.

De acordo com a Figura 16, o indexador WoS apresentou um índice  $h=4$ , ou seja, 4 artigos com mais de 4 citações. Com relação às fontes selecionadas pelo índice H, pode-se observar que estas estão com um melhor fator de impacto do que àquelas analisadas apenas com relação ao número de publicações. A Tabela 07 apresenta as publicações de acordo com o índice H, com  $h=7$ :

Tabela 07 - Publicações índice H - Área geral.

(continua)

Scopus		Web of Science	
Obras	Citações	Obras	Citações
Boyd D., Crawford K. Critical questions for <i>big data</i> : Provocations for a cultural, technological, and scholarly phenomenon. (2012) Information Communication and Society (Ciências Sociais: Comunicação-SJR=0.799; Q1)	30	Aronova, E.; Baker, K. S.; Oreskes, N. Big Science and <i>Big data</i> in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957-Present. (2010) Historical Studies in the Natural Sciences (Artes e Humanidades: História e Filosofia da Ciência-SJR= 0.593; JCR=1.087; Q1)	12

(continuação)

Scopus		Web of Science	
Obras	Citações	Obras	Citações
Herodotou H., Babu S. Profiling, what-if analysis, and costbased optimization of mapreduce programs. (2011) Proceedings of the VLDB Endowment (Ciência da computação; Engenharia; Ciência dos materiais)	23	Giannotti, F.; Pedreschi, D.; Pentland, A.; et al. A planetary nervous system for social mining and collective awareness. (2012) European Physical Journal-Special Topics (Ciência e Tecnologia-JCR=1.796; Q2)	10
Bhatarai B., Gramatica P. Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals. (2011) Environmental Science and Technology (Ciências Ambientais: Química Ambiental-SJR= 2.665; Q1)	13	Boyd, D.; Crawford, K. Critical questions for <i>big data</i> provocations for a cultural, technological, and scholarly phenomenon. (2012) Information Communication and Society (Ciências Sociais: Comunicação-SJR=0.799; Q1)	9
Giannotti F., Pedreschi D., Pentland A., Lukowicz P., Kossmann D., Crowley J., Helbing D. A planetary nervous system for social mining and collective awareness. (2012) European Physical Journal: Special Topics. Information Communication and Society (Ciências Sociais: Comunicação-SJR=0.799; Q1)	12	DiLeo, M. V.; Strahan, G. D.; den Bakker, M.; et al. Weighted Correlation Network Analysis (WGCNA) Applied to the Tomato Fruit Metabolome. (2011) PLoS ONE (Ciências Agrárias e Biológicas-SJR=1.512; JCR=3.730; Q1)	6
DiLeo M.V., Strahan G.D., den Bakker M., Hoekenga O.A. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. (2011) PLoS ONE (Ciências Agrárias e Biológicas-SJR=1.512; JCR=3.730; Q1)	9		
Fisher D., DeLine R., Czerwinski M., Drucker S. Interactions with <i>big data</i> analytics. (2012) Interactions (Ciência da Computação: Interação Humano-Computador)	9		

(conclusão)

Scopus		Web of Science	
Obras	Citações	Obras	Citações
Chen Y., Alspaugh S., Katz R. Interactive analytical processing in <i>big data</i> systems: A crossindustry study of mapreduce workloads. (2012) Proceedings of the VLDB Endowment (Ciência da computação; Engenharia; Ciência dos materiais)	8	/	

Fonte: elaborado pelo próprio autor.

De acordo com este índice, apresentado na Tabela 07, mais de 70% das publicações estão enquadradas com relação ao índice Q1, em Q1 ou Q2. A Tabela 08 mostra as publicações na categoria de ciências sociais - áreas de: negócios, gestão e contabilidade; ciência da decisão, economia, econometria e finanças; ciências sociais, com o número de citações:

Tabela 08 - Publicações índice H - Área ciências sociais.

(continua)

Scopus		Web of Science	
Obras	Citações	Obras	Citações
Hamel, M.-P., Marguerit, D. Opportunities for improving public teleservices through " <i>big data</i> " analysis [Quelles possibilités offertes par l'analyse des <i>big data</i> pour améliorer les téléservices publics?] (2013) Revue Francaise d'Administration Publique (Administração pública; SJR=0.197; Q3)	30	Davenport, T. H.; Barth, P.; Bean, R. How ' <i>Big data</i> ' Is Different. (2012). MIT Sloan Management Review (Negócios, gestão e contabilidade; ciência da decisão-SJR=0,879; Q1/Q2)	1
Shen, Y., Varvel, V.E. Developing data management services at the Johns Hopkins University. (2013) Journal of Academic Librarianship (Educação; SJR=1.577; Q1)	3	/	

(conclusão)

Scopus		Web of Science	
Obras	Citações	Obras	Citações
Valentine, E.L.H., Stewart, G. The emerging role of the Board of Directors in enterprise business technology governance (2013) International Journal of Disclosure and Governance (Negócios, gestão e contabilidade-SJR=0.241; Q3)	3		

Fonte: elaborado pelo próprio autor.

Nesta categoria, no indexador Scopus, pode-se identificar 31 publicações e índice  $h=3$ . No indexador WoS consta índice  $h=1$ . Os artigos identificados tratam sobre a utilização de *big data* para a melhoria dos serviços disponíveis e desempenho das administrações públicas; o exame de modelos de serviços de gestão de dados; a necessidade de melhoria da governança de tecnologia de negócios da empresa; e, as organizações que estão aprendendo a tirar proveito de *big data* com o entendimento do ambiente de negócios em um nível mais granular, a criação de novos produtos e serviços, e a rapidez na resposta às mudanças.

Assim, na pesquisa às bases de dados bibliométricos realizada foi possível avaliar que nenhum dos artigos publicados em *big data* e análise leva o contexto dos profissionais e da inteligência estratégica em seu escopo. Assim, embora algumas fontes de alto fator de impacto estejam publicando sobre o tema, ainda há muito a ser explorado em termos de *big data* e análise. Sendo assim, fica evidenciada uma lacuna em termos acadêmicos, já exposta na justificativa deste trabalho.

## 2.2 INTELIGÊNCIA ESTRATÉGICA

De acordo com Fuld (1995), é a inteligência estratégica (IE), não a informação, que auxilia gerentes a responderem corretamente questões específicas e tomarem decisões a longo prazo. A análise pode terminar em um relatório extenso ou com uma breve declaração, entretanto atenção deve ser dada em não confundir volume com valor. Assim, por meio da busca ativa de IE e aprendendo como usá-la, a organização pode transformar a informação em um elemento fundamental que irá contribuir para a obtenção de vantagens competitivas.

Na abordagem de Fleisher e Bensoussan (2003), a ideia de gerar inteligência estratégica não é nova. Entretanto, as graves consequências de se tomar decisões baseadas em

informações que são incompletas e imprecisas têm recebido atenção significativa apenas nas últimas décadas, por meio do desenvolvimento de efetivos processos de IE.

Assim é pela atividade de IE que as organizações reúnem informações práticas sobre os concorrentes e o ambiente competitivo e as aplicam ao processo de planejamento e tomada de decisões, a fim de melhorar o seu desempenho. De acordo com os autores, a IE liga sinais, eventos, percepções e dados em padrões perceptíveis e tendências sobre os negócios e ambiente competitivo. Fuld (1995) corrobora com esta visão ao propor que as decisões de negócios são baseadas em uma combinação de experiência, bom senso e inteligência, sendo a última a mais importante para o processo.

Atualmente, a informação é um bem relativamente barato e fácil de obter. Segundo Akerkar (2014), empresas e agências governamentais agregam dados de várias fontes privadas e/ou públicas. Os dados privados são informações que estão disponíveis para uma organização exclusiva, tais como dados de funcionários, dados de clientes e dados da máquina (por exemplo, transações de usuário e comportamento do cliente). Dados públicos são informações que estão disponíveis ao público mediante pagamento de uma taxa ou sem custo, tais como avaliações de crédito e conteúdos de mídia social (por exemplo, LinkedIn, Facebook e Twitter).

Para Fleisher e Bensoussan (2003), IE não é espionagem, é ética, legal, legítima e essencial. Os profissionais também podem usar informações públicas, mas não necessariamente publicadas. Em outras palavras, a informação que os profissionais procuram está prontamente disponível e identificada através de meios legais de fontes abertas, como documentos públicos, entrevistas, anúncios e documentos internos. Na visão dos autores, um caminho para entender inteligência é visualizá-la como um processo de entradas de matérias brutas e saídas de matérias acabadas.

Nesta perspectiva, a IE começa com pedaços espalhados de dados brutos, básicos. Essas matérias-primas são então organizadas pelos profissionais e tornam-se informação. Esta por sua vez será inteligência quando colocada em um formato útil para necessidades únicas ou críticas de mercado. A inteligência é, portanto, informação que é analisada, interpretada e infundida com implicações desenvolvidas. É o produto refinado por um analista que irá atender às necessidades únicas de um tomador de decisão para entender um aspecto competitivo do ambiente interno ou externo. Assim, a IE eficaz ajuda o tomador de decisão a tomar uma melhor decisão (FLEISHER E BENSOUSSAN, 2003).

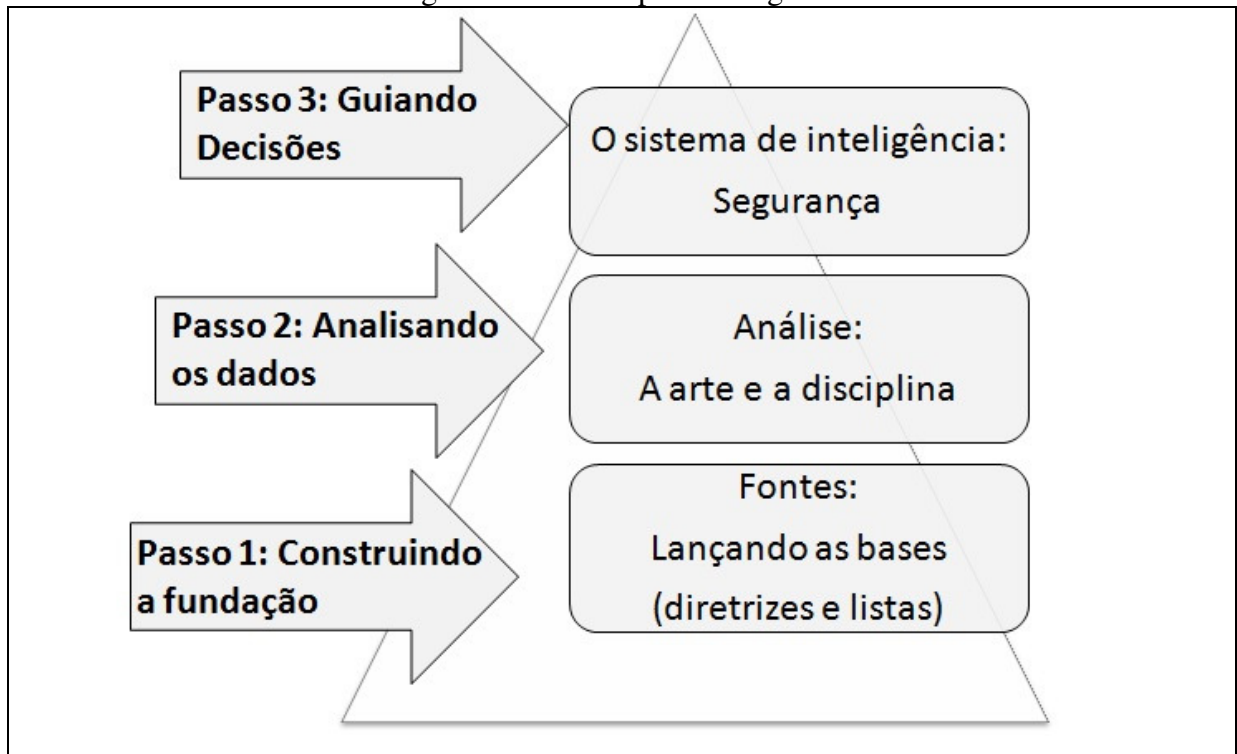
Fuld (1995) complementa que é necessário que se tenham as questões certas, e que seja feita uma análise nos resultados do processo para que se obtenha valor. Da mesma forma é necessário ter rapidez no seu uso, pois o valor da IE declina rapidamente. Assim, a disciplina do "quebra-cabeças" da inteligência, representa a habilidade de interpretar o que na superfície parecem ser peças independentes de informação e encaixá-las para formar uma imagem completa.

Para Fleisher e Bensoussan (2003), a inteligência estratégica também pode ser vista como uma função organizacional, variando em escopo entre a área mais ampla de inteligência de negócios e a versão mais pontual da análise da concorrência. A função IE fornece a base sobre a qual estratégia e tática são construídas, avaliadas e modificadas. Como uma função *staff-oriented*, IE vai cortar transversalmente e sobrepor outras funções, em particular, aquelas associadas com o *marketing*, planejamento e estratégia. Programas de IE têm objetivos como detectar proativamente oportunidades e ameaças, eliminar ou reduzir pontos cegos, riscos e surpresas, e o tempo de reação à mudanças na concorrência e no mercado. Os programas de IE tentam garantir que os tomadores de decisão tenham informações precisas e atualizadas sobre o ambiente competitivo da organização, e um plano para utilizar essas informações.

De acordo com Fuld (1995), a atividade de IE inicia com a descoberta dos dados, passando pela análise e por último a transformação dos dados em inteligência. Assim, no primeiro passo o analista deve captar as fontes de informação, com as quais irá conduzir o processo de análise. No segundo passo, ocorre a análise, propriamente dita, mas para este processo não existe um modelo ou uma única forma de fazer. Algumas pessoas têm a habilidade de rapidamente assimilar e analisar grandes quantidades de informação. Elas intuitivamente compreendem o processo de análise e têm disciplina para constantemente analisar a informação que recebem.

O pensamento de análise também evoca uma variedade de equações, análises de regressão, e outras técnicas estatísticas. Para Fuld (1995), a análise não precisa ser complicada, porém deve ser completa e precisa. A Figura 17 ilustra os passos do processo de análise para atividade de inteligência, na visão do autor:

Figura 17 - Passos para inteligência.



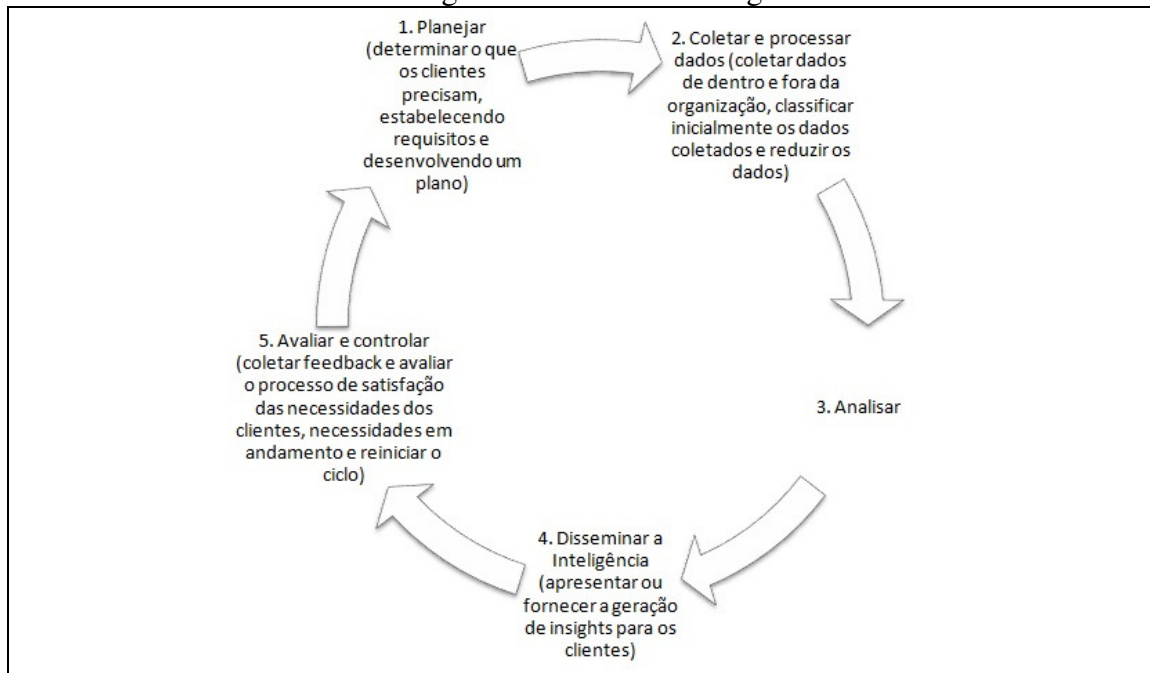
Fonte: Fuld (2012).

De acordo com a Figura 17, este pode ser simplesmente um processo de peneirar a informação inútil e encontrar um modelo que agregue valor. Por fim, o produto da análise será transformado em inteligência se for possível fazer uso do conhecimento gerado. O terceiro passo colocado pelo autor é a segurança, que envolve as ações a serem tomadas para evitar o vazamento de informações. De acordo com Fleisher e Bensoussan (2003), o processo de análise é a porção da atividade de inteligência em que o maior valor é gerado. Os autores sugerem que ela pode também ser mais difícil de fazer, do que as atividades de drenagem de dados e esforços de coleta. Além disso, análise serve a uma variedade de importantes necessidades a longo prazo para as organizações.

Desta maneira, análise é um passo importante do processo de desenvolvimento da inteligência em uma organização, mas existem outras. O "ciclo da inteligência", conforme exposto na Figura 18 apresenta uma série de 4 a 6 etapas:



Figura 18 - Ciclo da inteligência.



Fonte: Fleisher e Bensoussan, 2003.

Na Figura 18, os autores colocam como passos do ciclo da inteligência: planejar, coletar e processar dados, analisar, disseminar a inteligência e avaliar e controlar se as necessidades dos clientes foram aceitas. Fleisher e Bensoussan (2003) ainda esclarecem que na atividade de inteligência, a análise é sempre específica. Isto porque ela não ocorre dentro de um vácuo, uma condição estática ou sob condições experimentais. A análise é realizada por profissionais em organizações ativas, e em condições competitivas importantes. O processo de análise neste ciclo envolve as próprias subtarefas que são conduzidas em ordem para gerar saídas e resultados.

Trata-se de um trabalho interativo com todos os subprocessos de classificação e síntese de dados para produzir mapas, gráficos, tabelas, textos, resumos, ou outros recursos de comunicação apropriados para disseminação. Assim, para que seja possível executar a interpretação da informação é necessário que o profissional leve em consideração aspectos metodológicos e cognitivos no processo de análise.

### 2.2.1 Aspectos metodológicos e cognitivos do processo de análise

A eficácia do processo de análise depende da obtenção de interpretações e conclusões que farão parte do processo de tomada de decisões, como resultado da atividade de IE. Este processo é uma parte central do ciclo de inteligência, tendo profundo impacto no resultado do

processo. A Tabela 09 expõe algumas definições para o termo, usadas em contextos de inteligência:

Tabela 09 - Definições análise.

<b>Definição</b>	<b>Fonte</b>
A aplicação do senso comum e experiência na informação bruta.	FULD, 1995
Trata-se de um exame aprofundado de itens relacionados de informações, para determinar a extensão em que eles confirmam, complementam ou contradizem uns aos outros para, assim, estabelecer probabilidades e relacionamentos.	MATHAMS, 1995
A saída primária da fase de processamento do ciclo de inteligência é o processo humano de sintetizar pedaços de informação em inteligência.	PROCYCSHN, 1998
O coração do processo de inteligência pelo qual o significado é derivado de dados.	IALEA, 2001
O uso de alguma metodologia ou técnica para, em primeiro lugar, encontrar relações entre diferentes peças de informações e, em seguida, fazer inferências a partir das relações.	HALLIMAN, 2003
A aplicação de métodos cognitivos individuais e coletivos para pesar os dados e testar hipóteses.	JOHNSTON, 2005
Aplicação hábil de métodos e processos científicos e não-científicos pelos quais os indivíduos interpretam dados ou informações para produzir resultados de inteligência perspicazes e recomendações viáveis para os tomadores de decisões	FLEISHER E BENSOUSSAN, 2007

Fonte: Fleisher e Bensoussan (2007).

De acordo com as definições apresentadas na Tabela 09, algumas semelhanças são encontradas entre os autores, no que tange a intervenção humana para atribuir significado à informação bruta, por meio de métodos ou técnicas. Assim, o processo de análise na atividade de inteligência estratégica leva em consideração aspectos metodológicos e cognitivos para que seja possível executar a interpretação da informação. De acordo com Palop e Vicente (1999), os aspectos metodológicos são chamados de fase física da análise. É onde dados são agregados e relacionados entre si e passam a fazer sentido, gerando como produto o conhecimento.

Para Fleisher e Bensoussan (2003), os aspectos metodológicos compreendem as técnicas de análise estratégica; análise da competência e do consumidor; análise de retorno; análise evolutiva, e; análise financeira. De acordo com Gomes e Braga (2006), neste contexto, algumas ferramentas e técnicas podem contribuir para facilitar o trabalho de análise. Entre elas pode-se citar:

- Modelo das 5 forças de Michael Porter: auxilia na análise do ambiente competitivo com a identificação da vantagem potencial de um setor, verificando as suas forças e antecipando proativamente mudanças na sua estrutura. Utilizado para analisar informações relativas aos concorrentes;

- Fatores críticos de sucesso: utilizado para avaliar dois tipos de atributos: atratividade da indústria e força do negócio. O analista lança vários fatores que afetam a indústria e o negócio, atribuindo a cada um deles um valor crescente de 1 a 5. Calcula-se a média dos valores e monta-se uma matriz. Avalia as forças do negócio e compara concorrentes em uma indústria;

- Cenários: método utilizado para fazer previsões (situação provável, positiva e negativa) de tendências para variáveis do ambiente externo que exercem forte impacto sobre o desempenho da organização;

- Perfil do concorrente: fornece um quadro das forças e fraquezas dos potenciais e atuais concorrentes. Tem quatro grandes objetivos: (1) identificar os planos estratégicos dos concorrentes; (2) prever quais serão as reações dos concorrentes frente às iniciativas competitivas; (3) determinar como melhor agrupar as estratégias dos atuais competidos; e, (4) entender as fraquezas dos concorrentes;

- *Benchmarking*: é um método para avaliar quão bem as funções e habilidades de uma organização estão sendo desempenhadas em comparação com outra organização ou um conjunto de organizações. Seu objetivo principal é entender completamente as “melhores práticas” de outras organizações e empreender ações para atingir tanto o melhor desempenho quanto os menores custos;

- SWOT: analisa a capacidade interna (forças e fraquezas) e sua posição em relação aos concorrentes (ameaças e oportunidades). Com esta ferramenta é possível identificar na análise, quais fatores devem ser atacados para melhorar nosso negócio.

Uma das importantes contribuições de Fleisher e Bensoussan (2003) neste contexto é a abordagem FAROUT. Ela apoia os analistas na escolha da técnica ou combinação de técnicas para melhor atender às situações específicas. FAROUT é baseado na premissa de que para a saída analítica ser perspicaz, inteligente e útil aos tomadores de decisão, ela precisa atender a uma série de características comuns. A saída deve ser: (F) *Future-oriented* (orientada para o futuro); (A) *Accurate* (precisa); (R) *Resource efficient* (eficiente em recursos); (O) *Objective* (objetiva); (U) *Useful* (útil); e, (T) *Timely* (oportuna). De acordo com os autores cada técnica é avaliada contra os seis elementos FAROUT e, o não cumprimento de todos estes critérios

em um nível satisfatório, resultará na saída analítica ser abaixo do ideal e de menor valor para o tomador de decisão.

De acordo com Fachinelli, Rech e Bertolini (2011), os aspectos metodológicos deste processo já estão em um bom nível de desenvolvimento, ao contrário dos cognitivos que ainda não estão epistemológica nem pragmaticamente consolidados. Os aspectos cognitivos atuam com a atribuição de sentido e são denominados por Palop e Vicente (1999) como fase mental. Nela ocorre a interpretação da informação, por meio da atividade analítica. É quando diferentes significados podem ser atribuídos para o mesmo conjunto de informações. Assim, aos dados é atribuído significado genérico o qual ambientado ao contexto e com influência de fatores pessoais, adquire significado e pode então ser chamado de inteligência.

Como em muitos campos de investigação em desenvolvimento, negócios e análise competitiva não são puramente arte ou ciência, mas uma combinação de porções substanciais dos dois. Assim como na pesquisa formal, o processo de análise também pode ser visto como tendo muitos aspectos em comum com o método científico. Desta forma, o profissional em meio ao processo de análise observará certos eventos, pessoas, ou ações, desenvolvendo uma proposição ou hipóteses que irão descrever ou explorar o que se têm observado. Estas hipóteses serão então usadas para produzir previsões sobre o que pode ocorrer na sequência. As previsões podem ser avaliadas através de observações adicionais de dados e hipóteses que podem ser modificadas com base nos resultados.

Entretanto, Fleisher e Bensoussan (2007) esclarecem que este processo de análise aplicado em teoria, se torna complexo por fatores presentes no mundo real dos negócios e concorrência de mercado, bem como políticas sociais e natureza dos tomadores de decisão e das empresas. A análise frequentemente trabalha em grupos ou times e se beneficia da partilha de conhecimentos. Hipóteses não são sempre desenvolvidas, testadas, ou reformuladas, mas são frequentemente modificadas em tempo real pela evidência que é adquirida. Estes fatores apontam para análise de negócios e competitiva como sendo mais uma perseguição das ciências sociais, do que da ciência física ou pura.

Assim, o mundo real tende a dar ênfase ao aspecto da arte da análise. Por exemplo, ao realizar um experimento usando grupos de controle, alguma percentagem de análises de um grupo de tratamento pode não confirmar as análises do grupo de controle. Analistas nestes dois grupos vão pesar os mesmos dados ou informações de forma diferente, com base em esquemas mentais que podem ser difíceis de reproduzir com exatidão. Mesmo quando os analistas não substantivamente concordam sobre a natureza dos problemas que estão sendo

abordados, eles ainda podem, posteriormente, discordar sobre o curso de ação proposto. Consequentemente, os analistas podem nunca ser capazes de provar que estavam certos. São esses tipos de experiências que tornam a análise, em algumas vezes mais semelhante a arte do que a ciência.

De acordo com Fleisher e Bensoussan (2007), analistas de negócios e competitivos são encarregados de fazer com que assuntos muitas vezes ambíguos, complexos e desafiadores, façam sentido para os tomadores de decisão. Assim eles também têm que ponderar as probabilidades, trabalhar através dos cenários, descobrir algo que seus concorrentes não saibam, e tomar medidas. Eles têm que encontrar sentido, ou criar significado a partir de uma amostra normalmente limitada de dados e informações. Em uma paisagem muitas vezes confusa e que se move rapidamente, eles tentam responder às três questões críticas: O que? Então, o que? Agora o que? Qualquer uma destas três perguntas pode ser respondida em uma variedade de formas construtivas utilizando procedimentos replicáveis e métodos (ciência), bem como intuição ou criatividade (arte).

Assim, os profissionais que estão envolvidos no processo de análise podem contar com metodologias que irão auxiliar no seu trabalho. Entretanto são os aspectos cognitivos que parecem se colocar como obstáculos na transformação da informação em inteligência pelo processo de análise. Isso se deve ao fato de que o profissional em inteligência estará sujeito a fraquezas e predisposições inerentes aos processos do pensamento humano. Estas fraquezas podem se reduzir pela aplicação de dispositivos e técnicas que devem estar no conjunto de ferramentas cognitivas de todos os profissionais em inteligência (HEUER, 1999).

De acordo com Sfez (1993), as ciências cognitivas têm se desenvolvido a partir de muitas disciplinas tradicionais: psicologia experimental, inteligência artificial, linguística e neurociência. A convergência destas áreas deu origem ao conceito da cognição, o estudo das estruturas e dos processos conscientes e inconscientes que permite aos seres humanos produzir representações de mundo que podem ser verdadeiras ou falsas.

Acompanhando este pensamento, Heuer (1999) no relatório *Psychology of intelligence analysis*, do *Center of the Study of Intelligence da Central Intelligence Agency*, aponta os três pontos fundamentais sobre os desafios enfrentados pelos profissionais em inteligência. Primeiramente a mente humana não funciona bem para lidar com a incerteza inerente, ou a incerteza natural que envolve as questões de inteligência, e com a incerteza induzida, que é gerada pela negação e engano nas operações.

Outro ponto refere-se ao fato de que mesmo uma maior consciência cognitiva, desprovida de preconceitos como a tendência em ver a informação se confirmar ao invés de negá-la, por si só é pouco para ajudar os profissionais em inteligência a lidar eficazmente com a incerteza. Por fim, ferramentas e técnicas aplicadas pela mente do profissional, no sentido de ampliar o pensamento crítico, podem contribuir para melhorar substancialmente a análise de questões complexas em que a informação é incompleta, ambígua, e muitas vezes distorcida. Exemplos importantes de tais dispositivos intelectuais são: utilizar informações técnicas para estruturação, desafiar suposições e explorar interpretações alternativas.

Ainda segundo o mesmo relatório, Heuer (1999) destaca a importância dos modelos mentais para os aspectos cognitivos do processo de análise. O autor coloca que os profissionais constroem a sua própria versão de realidade nas bases de informações fornecidas pelos sentidos. Isso é mediado por complexos processos mentais que determinam qual informação é percebida, como é organizada, e qual o significado que lhe é atribuído. O que as pessoas percebem, quão facilmente percebem, como processam essas informações após recebê-la são todos elementos fortemente influenciados pela experiência do passado, a educação, os valores culturais, requisitos, e normas organizacionais, bem como pelas especificidades da informação recebida.

Este processo pode ser visualizado como perceber o mundo através de uma lente ou uma tela onde canais são focados, podendo assim, distorcer as imagens que são vistas. Assim, para atingir a imagem mais clara possível os profissionais em inteligência precisam mais do que a informação, eles também precisam entender as lentes pelas quais esta informação passa.

Segundo Bruce e George (2008), este profissional deve ser um *expert* em como utilizar a coleção das atividades de inteligência. Este profissional deve ser imaginativo e rigoroso tendo em conta as explicações dos dados perdidos, confusos e, frequentemente contraditórios. Além disso, deve ser capaz de ser autocrítico das próprias predisposições e expectativas sobre o que a série de dados apresenta e, o mais importante, deve estar aberto em mudar sua forma de pensar e modificar a forma de analisar o problema.

No relatório *Psychology of intelligence analysis*, Heuer (1999) ainda aponta aos líderes, gerentes e analistas que, para garantir a melhoria sustentada na avaliação de questões complexas, a análise deve ser tratada como mais do que um processo efetivo e organizacional. Atenção também deve ser dada às técnicas e ferramentas para lidar com as limitações inerentes aos modelos mentais do profissional em inteligência. Neste sentido devem-se tomar medidas para estabelecer um ambiente organizacional que promova e recompense a defesa do

pensamento crítico e que instigue a investigação com análise em profundidade, ao invés da conformidade com a análise superficial.

Este ambiente também deve estar empenhado em expandir o financiamento para a investigação sobre o papel dos processos mentais no desempenho da definição de juízos analíticos. Uma agência que se baseia em desempenho cognitivo acentuado por seus profissionais deve ficar a par de estudos sobre como a mente funciona, ou seja, em como profissionais em inteligência chegam aos julgamentos no processo de análise. Outro ponto a ser desenvolvido no ambiente organizacional refere-se a fomentar o desenvolvimento de ferramentas para auxiliar os profissionais na avaliação de informações. Em questões difíceis, eles precisam de ajuda para melhorar seus modelos mentais derivados de conclusões incisivas a partir de informações de que já possuem.

Segundo Bruce e George (2008), o papel fundamental do analista está em transformar informações fornecidas por vários sistemas de coleta em julgamento e percepção para o cliente. Entretanto esta é uma visão linear, na realidade a parte cognitiva da análise é mais parecida com um modelo computacional quando usada para organizar e interpretar dados de entrada e constantemente reavaliar como novos dados podem mudar não só os resultados, mas também a técnica que está sendo usada para organizar e interpretar os dados.

Assim, o processo analítico pode ser entendido como mais do que apenas uma pessoa bem educada (inteligente) que pode escrever de forma concisa. O analista de inteligência completo deve combinar as habilidades de historiador, jornalista, pesquisador, gerente de coleta e cético profissional. Isto é, no mínimo, ele ou ela deve demonstrar um conjunto muito original de atividades analíticas (BRUCE E GEORGE, 2008):

- Domínio do assunto, bem como as políticas relacionadas;
- Compreensão dos métodos de pesquisa para organizar e avaliar os dados;
- Auto-consciência de vieses cognitivos e outras influências cognitivas em análise;
- Imaginação e rigor científico gerado, bem como testar hipóteses;
- Compreensão das tendências exclusivas e outras influências cognitivas em análise;
- Mente aberta para opiniões contrárias ou modelos alternativos que se ajustam os dados; e,
- Auto-confiança para admitir e aprender com os erros analíticos.

De acordo com os autores, o que distingue um analista de inteligência de um especialista de fora da comunidade de inteligência, não são as três primeiras características, que são compartilhadas com muitos especialistas em diversas áreas, embora esses atributos

sejam especialmente importantes em inteligência. O analista completo deve ser um especialista em como usar os recursos de coleta de inteligência, ser ao mesmo tempo imaginativo e rigoroso na análise com explicações incompletas, confusas, e dados, muitas vezes contraditórios, ao passo que é capaz de ter auto crítica de suas próprias tendências e expectativas sobre o que os dados mostram, e mais importante, estar aberto para mudar a mente e conscientemente questionar-se: "Se eu estiver errado, como eu poderia modificar a maneira que eu estou analisando o problema?" (BRUCE E GEORGE, 2008).

Fachinelli, Rech e Bertolini (2011) ao revisar o trabalho de Heuer (1999) e Bruce e George (2008) identificaram que os autores propõem estas categorias, mesmo em estudos anteriores. Entretanto, não foram identificados estudos empíricos derivados desta base conceitual.

### **2.2.1.1 Análise de *big data***

A humanidade tem evoluído em termos de análise, desde seus primórdios na antiga Mesopotâmia, passando pelo início da era analógica, onde a coleta e análise de dados era cara e demorada. Decorrente disto, surgiu a necessidade de um gerenciamento mais eficiente com a digitalização, que tornou as informações analógicas compreensíveis a computadores, facilitando e barateando o armazenamento e processamento, e finalmente, o momento que se tem presenciado atualmente, com a dataficação.

Schönberger-Mayer e Cukier (2013) atribuem ao termo dataficação a coleta de informações de tudo o que existe e a transformação disso em dados que possam ser quantificados. Isto permite que sejam usadas informações de novas maneiras, como análises preditivas, revelando o valor latente e implícito das informações e a mudança da causalidade pela probabilidade e correlação. Akerkar (2014) contrapõe este pensamento quando coloca que correlações, embora muito úteis, fornecem uma resposta para "o que ", mas não "por que". Para Schönberger-Mayer e Cukier (2013), isto muitas vezes basta, porém Akerkar (2014) complementa com a colocação de que é necessária uma abordagem científica para inferir as verdadeiras razões, e que este trabalho é feito atualmente, predominantemente pela análise humana.

Para Barlow (2013), pessoas e processos são parte integrante e fundamental para o sucesso de iniciativas que envolvam a tecnologia. Esse fato tem sido demonstrado repetidamente, ao longo das últimas seis décadas de evolução tecnológica. Cada uma dessas



revoluções foi seguida por um período de intensa adaptação cultural, com indivíduos e organizações lutando para capitalizar seus benefícios. Assim, parece improvável que *big data* siga uma trajetória diferente.

A mentalidade sobre como os dados podem ser usados é uma mudança mais importante e sutil do que as tecnológicas, neste processo. Um exemplo disto é o de Oren Etzioni, que entre outros feitos, mesmo antes do fenômeno dos grandes volumes de dados, possibilitou a criação da Farecast, empresa ícone em *big data*. Por meio de um modelo que se destinava a prever se e quanto o preço de uma passagem aérea aumentaria ou diminuiria, as probabilidades obtidas a partir de dados de outros voos trouxeram importantes benefícios em termos de economia aos passageiros (SCHÖNBERGER-MAYER E CUKIER, 2013). Outros exemplos de novos negócios surgiram a partir desta mudança na mentalidade para o uso de dados, mesmo antes do fenômeno. Alguns deles podem ser visualizados na Tabela 10:

Tabela 10 - Exemplos para utilização de *big data*.

(continua)

Ano	Empresa	Área	Problema	Aplicação <i>Big data</i>
1994	Infospace	Computação	Realizar buscas na <i>internet</i> .	Foi criado por Oren Etzioni - Mestre em Ciência da Computação pela Harvard em 1986 - um dos primeiros mecanismos de busca na <i>internet</i> : MetaCrawler.
1997	Excite	Marketing	Comparar de preços pela <i>internet</i> .	Foi criado por Etzioni o primeiro grande site de comparação de preços pela <i>internet</i> : Netbot.
1998	Reuters	Computação	Extrair significado de documentos de texto.	Foi criada por Etzioni a primeira <i>startup</i> para extrair significado de documentos de texto: ClearForest.
2000	Governo Estados Unidos	Astronomia	Explorar a estrutura e composição estelar de galáxias.	Foram coletados, no início da Sloan Digital Sky Survey, mais dados nas primeiras semanas que em toda a história da astronomia. Os lançamentos de dados ficam disponíveis na <i>internet</i> e o SkyServer oferece uma gama de interfaces utilizando o Microsoft SQL Server. Os dados ficam disponíveis para uso não comercial.

(conclusão)

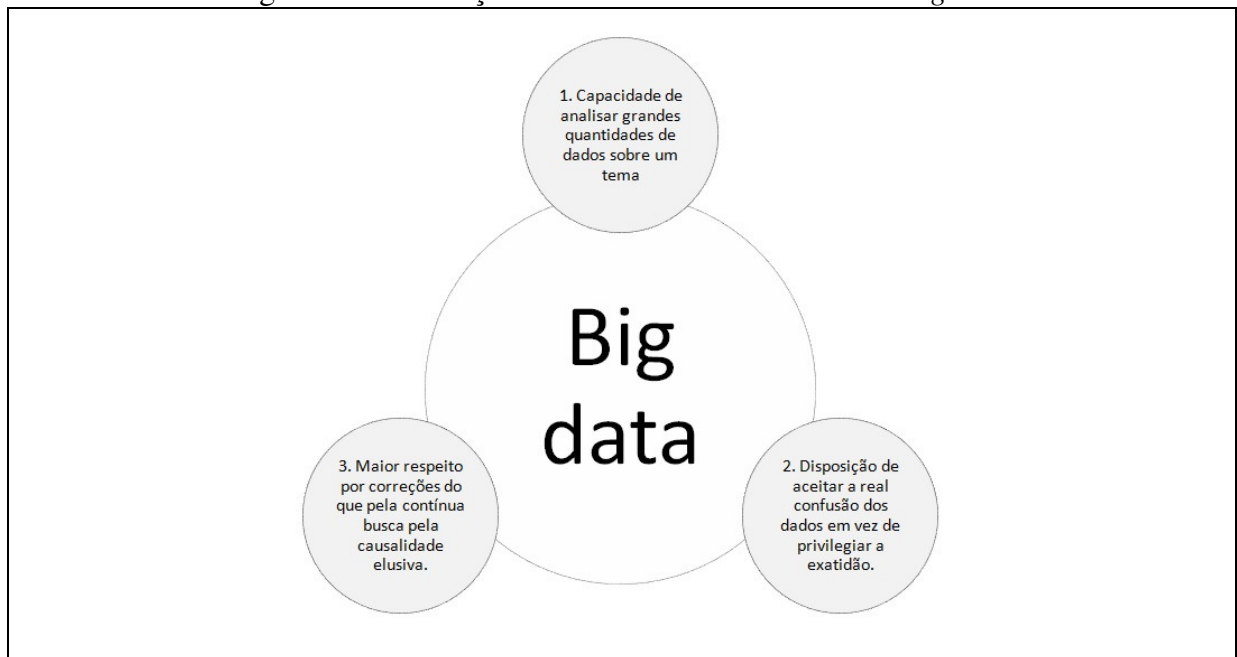
Ano	Empresa	Área	Problema	Aplicação <i>Big data</i>
2008	Microsoft	Marketing	Descobrir uma maneira de as pessoas saberem se o preço das passagens on-line é ou não um bom negócio.	Foi pega uma amostragem de 12 mil preços obtidos em um <i>website</i> de viagens durante 41 dias, e Etzioni criou um modelo de previsão que proporcionou grande economia aos passageiros simulados. O modelo não entendia o <i>porquê</i> , apenas o <i>quê</i> . O sistema baseava suas previsões no desconhecido: probabilidades obtidas a partir de dados de outros voos. O projeto foi chamado pesquisa de Hamlet e deu origem a uma <i>startup</i> chamada Farecast, que mais tarde foi comprada pela Microsoft por cerca de US\$ 110 milhões e integrada ao sistema de buscas da Bing.
2009	Google	Saúde Pública	Desacelerar a propagação do vírus H1N1, que foi comparado a gripe espanhola de 1918. Necessidade de descobrir onde estava a doença.	Foram pegos os 50 milhões de termos de busca mais comuns dos americanos, comparados a lista com os dados do CDC ( <i>Centers for Disease Control</i> ) sobre eliminação da gripe entre 2003 e 2008. A ideia era identificar pessoas infectadas pelo vírus de acordo como que pesquisavam na <i>internet</i> . Com isso a Google pode "prever" a disseminação da gripe de inverno nos Estados Unidos, não apenas nacionalmente, mas em regiões específicas e até mesmo em estados.
2009	Flight Caster	Serviços	Prever se determinado voo nos Estados Unidos provavelmente atrasaria.	Para isto, analisavam todos os voos dos 10 anos anteriores e comparavam com dados antigos e atuais sobre o clima. As previsões eram tão precisas que até mesmo os funcionários das companhias aéreas começaram usá-las. Isto porque, as empresas só queriam anunciar atrasos no último minuto, ou seja, embora fossem a principal fonte de informação, não eram a melhor.

Fonte: elaborado pelo próprio autor com base em Schönberger-Mayer e Cukier (2013).

Na Tabela 10, observa-se a criação de negócios que utilizam grandes volumes em diferentes segmentos como: computação, *marketing*, astronomia, saúde e serviços. Estes negócios foram criados mesmo com as restrições tecnológicas presentes e tornaram-se possíveis em função de uma mentalidade alinhada ao tratamento de grandes volumes de dados. Entretanto, o pensamento mais comum naquela época era considerá-los como estáticos e banais (SCHÖNBERGER-MAYER E CUKIER, 2013).

De acordo com Schönberger-Mayer e Cukier (2013), estes dados estão se tornando matéria-prima dos negócios e estão sendo usados atualmente para criar uma nova forma de valor econômico em inovação e serviços. Segundo os autores, são apresentadas na Figura 19 as três principais mudanças na forma de análise das informações:

Figura 19 - Mudanças de mentalidade associadas ao *big data*.



Fonte: elaborado pelo próprio autor com base em Schönberger-Mayer e Cukier (2013).

A primeira mudança refere-se à capacidade de analisar grandes quantidades de dados sobre um tema, ao invés de conjuntos menores de dados: ao longo de boa parte da história, coletamos poucos dados porque os instrumentos para coletar, organizar, armazenar e analisar eram ruins. As informações eram peneiradas ao mínimo, de modo que se pudesse analisar com mais facilidade. Este fato pode ser colocado como um limite artificial imposto à tecnologia da época. No passado, a coleta e análise de informações em escala massiva era tarefa que cabia a instituições poderosas como Igreja e o Estado (SCHÖNBERGER-MAYER E CUKIER, 2013).

De acordo com os autores, para que fosse possível realizar a coleta de informações foram desenvolvidas técnicas elaboradas, utilizando o menor número de dados possível. Assim, a amostragem foi a solução para o problema da profusão de informações, quando era difícil coletar e analisar dados. Entretanto, a amostragem aleatória apresenta-se como a segunda melhor alternativa para coletar e analisar todo um banco de dados, pois nesta abordagem apresentam-se alguns pontos fracos: a) sua precisão depende de garantir a aleatoriedade na coleta dos dados, b) amostragens aleatórias não são boas para incluir subcategorias e, c) amostragem requer planejamento e execução cuidadosos, geralmente não se podem fazer novas perguntas referentes a dados de amostragem se elas não tiverem sido analisadas a princípio.

Desta forma, a amostragem é resultado de uma era de limites no processamento de informações. Atualmente podemos analisar grandes quantidades de dados. A Tabela 11 aponta algumas diferenças entre a utilização da amostragem aleatória e do todo na análise de dados e informações:

Tabela 11 - Amostragem aleatória e *big data*.

<b>Amostragem aleatória - <i>Small Data</i></b>	<b>Todos os dados - <i>Big data</i>*</b>
Não podemos nos aprofundar nos dados.	Podemos nos aprofundar nos dados.
Na margem de erro de 3% são ignorados detalhes de granularidade e capacidade de analisar mais de perto os grupos.	Compreensão de padrões com dados temporais, coletados ao longo do tempo.
Baixo poder de processamento e armazenagem.	Muito poder de processamento e armazenagem e ferramentas de ponta para analisar dados, além de maneiras fáceis e baratas de coletar dados.
Análise com dados passados.	Necessidades de análise em tempo real.
Rigidez no planejamento e execução de análise com os dados coletados.	Liberdade para explorar, analisar os dados de ângulos distintos e examinar mais de perto certos aspectos.
A princípio são definidos os objetivos da pesquisa.	A princípio não se tem certeza se algo será descoberto e o que será descoberto.
É passível a tendenciosidades, por depender de uma seleção amostral.	Substitui os especialistas do passado.
Precisão nas medições.	Imprecisão nas medições com potencial de maior agregação de valor.

Fonte: elaborado pelo próprio autor com base em Schönberger-Mayer e Cukier (2013).

\* *Small e Big* são mencionados como termos relativos e não absolutos: relativamente ao amplo conjunto de dados.

A Tabela 11 apresenta como a utilização de todos os dados disponíveis possibilita novas formas para o tratamento e descoberta de valor. Assim, ao invés de se coletarem alguns

dados, coleta-se o máximo de dados, e se possível, todos os dados (SCHÖNBERGER-MAYER E CUKIER, 2013).

A segunda mudança diz respeito à disposição em aceitar a real confusão dos dados ao invés de privilegiar a exatidão: o aumento da quantidade de dados abre as portas para a inexatidão. Apesar de números errados e informações corrompidas sempre terem ocorrido em bancos de dados, eles são inevitáveis e deve-se aprender a conviver com eles. No mundo de *small data*, reduzir erros e garantir alta qualidade dos dados era um impulso natural e essencial. Entretanto, o benefício de flexibilizar padrões de erros permitidos abre a possibilidade de se coletar mais dados. O *big data* transforma os números em algo mais probabilístico que exato, ou seja, é preciso aceitar a imprecisão quando se aumenta a escala (SCHÖNBERGER-MAYER E CUKIER, 2013).

Na visão dos autores, os analistas convencionais têm dificuldade para entender a confusão, uma vez que passaram toda a vida focados em evitá-la e eliminá-la. Assim, entrar num mundo de *big data* exigirá que seja mudada a mentalidade quanto ao mérito da exatidão. Quando a quantidade de dados é enorme e de um tipo novo, em alguns casos a exatidão já não é um objetivo, desde que se possa descobrir a tendência geral. Passar a trabalhar com dados imperfeitos e imprecisos, possibilitará a geração de melhores previsões e um melhor entendimento do mundo.

Entretanto, a confusão não é inerente ao *big data*, mas uma função da imperfeição dos instrumentos que usamos para medir, registrar e analisar as informações. Contudo, grandes esforços para aumentar a precisão talvez não façam sentido em termos econômicos, já que o valor de ter quantidades maiores de dados é mais interessante. Assim, na combinação de *big data* com análise inteligente é possível que sejam descobertos padrões (SCHÖNBERGER-MAYER E CUKIER, 2013).

Segundo Schönberger-Mayer e Cukier (2013), a indústria da inteligência corporativa e dos *softwares* analíticos há muito se baseia na promessa de dar a todos que acessam os sistemas de tecnologia e informação de uma empresa os mesmos dados. Mas essa visão está sendo substituída pela dispersão, para que se possam aproveitar os benefícios dos dados em escala. Em troca do convívio com a confusão, é possível conseguir serviços extremamente valiosos, impossíveis nos métodos tradicionais. Isto porque, apenas 5% de todos os dados digitais são "estruturados", e não admitir a confusão representaria desperdiçar 95% de dados não estruturados, como sites e vídeos.

A terceira mudança é o maior respeito pelas correlações do que pela contínua busca pela causalidade elusiva. Em essência, uma correlação quantifica a relação estatística entre dois dados. As correlações permitem que analisemos um fenômeno não pelo foco nas suas engrenagens, e sim pela identificação de um substituto para elas. Entretanto, é preciso ter cuidado com o fato de que a medida que a quantidade de dados aumenta em magnitude, também encontramos mais correlações ilegítimas (SCHÖNBERGER-MAYER E CUKIER, 2013).

Assim, especialistas usavam hipóteses nascidas em teorias, e com base nelas, coletavam dados e usavam a análise das correlações para verificar se os substitutos eram adequados. Se não fossem, a tentativa era refeita até que a hipótese da qual partiram fosse confirmada verdadeira ou constatassem que a teoria era falha e precisava de emendas. O conhecimento avançou lentamente, por meio desta abordagem baseada em hipóteses, enquanto tendências individuais e coletivas poderiam estar tirando o foco das hipóteses desenvolvidas.

Na visão de Schönberger-Mayer e Cukier (2013), a tomada de decisão apenas com base em hipóteses já não é mais eficiente na era do *big data*. Os bancos de dados são grandes demais, bem como a complexidade das áreas sob investigação. Atualmente, há tantos dados disponíveis e tanto poder de processamento, que não é trabalhoso escolher um ou alguns substitutos e examiná-los um a um. Porém, encontrar substitutos em contextos sociais é apenas uma maneira pela qual as análises de *big data* podem ser empregadas.

Complementarmente a isso, correlações com novos tipos de dados podem ser utilizadas para resolver necessidades diárias. Uma dessas é chamada "análise de previsão", usada no mundo dos negócios para prever eventos. Nesta análise, o foco não está em explicar a causa do problema, mas apenas em sua existência. As correlações mostram "o que", não o "porque", o que geralmente é o suficiente. Especialistas também estão se dedicando a desenvolver os instrumentos necessários para identificar e comparar correlações não lineares. Como exemplo, pode-se citar a "análise de rede" que possibilita mapear, medir e calcular os nós e as conexões em tudo, desde amigos em redes sociais, até ligações entre telefones celulares.

Ainda assim, as análises causais são colocadas como a maneira como os seres humanos veem o mundo. Isto ocorre de duas formas: por meio da rápida e ilusória causalidade e de lentos e metódicos experimentos causais. Em geral, as causalidades são colocadas como um atalho cognitivo, que leva a compreender o mundo de uma forma

equivocada. O *big data* surge neste contexto modificando papéis, pois suas correlações serão rotineiramente utilizadas para discordar da intuição causal, mostrando que há pouco ou nenhuma conexão estatística entre o efeito e o que se supõe ser a causa (SCHÖNBERGER-MAYER E CUKIER, 2013).

Mineli, Chambers e Dhiraj (2013), confirmam os pressupostos de Schönberger-Mayer e Cukier (2013) ao colocarem que com *big data*, cientistas de dados podem utilizar mais ou todos os dados para criar um modelo. Ao fazerem isto, é possível que sejam introduzidas variáveis de previsão adicional, a fim de aumentar seu nível de exatidão. Quando o histórico de *big data* é utilizado, o modelo pode identificar tendências que estão fora dos ciclos que foram utilizados na técnica de extração de dados históricos.

A utilização de *big data* no processo de análise é um desafio a ser superado pelos profissionais em IE, pois requer uma mudança de mentalidade frente aos dados, além do desenvolvimento de processos e habilidades específicas. Já para algumas organizações, este desafio pode se apresentar na efetivação da inteligência, com a utilização do produto da análise como subsídio para o processo de tomada de decisão. Desta forma, para que fosse possível comprovar que o valor de agir com base em dados, McAfee e Brynjolfsson (2012) lideraram uma equipe no MIT *Center for Digital Business* em parceria com o departamento de tecnologia empresarial da McKinsey e outros pesquisadores.

O objetivo era testar a hipótese de que empresas movidas a dados teriam um desempenho melhor. Assim, foram realizadas entrevistas estruturadas com executivos de 330 empresas norte-americanas de capital aberto sobre suas práticas de gestão organizacional e de tecnologia, e foram reunidos dados de desempenho disponíveis em balanços anuais e fontes independentes. A pesquisa realizada trouxe a conclusão de que nem todos abraçam a tomada de decisões movida a dados. Entretanto, na análise realizada uma relação foi evidenciada: quanto mais a empresa se caracterizava como movida a dados, melhor era seu resultado em indicadores objetivos de desempenho financeiro e operacional. Em alguns casos, as empresas que utilizavam dados na tomada de decisões eram, em média, 5% mais produtivas e 6% mais rentáveis do que as concorrentes (MCAFEE E BRYNJOLFSSON, 2012).

Para que isto ocorra num contexto de grandes volumes de dados, Davenport, Barth e Bean (2012) elucidam que as organizações que capitalizam com *big data* se distinguem dos ambientes tradicionais de análise de dados de três maneiras principais: a) elas prestam atenção aos fluxos de dados ao invés de ações; b) elas contam com cientistas de dados e

desenvolvedores de produtos e processos; e c) elas estão movendo a análise para o *core business* das funções operacionais e de produção.

Assim, para captar o potencial de geração de valor de *big data* no processo de análise, as organizações precisam pensar em termos de fluxos de processos, ao invés de olharem para os dados apenas para avaliar o que ocorreu no passado. A Tabela 12 apresenta quais as questões que podem ser atendidas com a inserção de análises preditiva e prescritiva neste processo:

Tabela 12 - Análise *big data*

<b>Análise descritiva (<i>business intelligence</i>)</b>	<b>Análise preditiva</b>	<b>Análise prescritiva</b>
O que e quando isso aconteceu? Quanto é impactado e quantas vezes isso acontece? Qual é o problema?	O que é provável de acontecer em seguida? E se essas tendências continuarem? E se?	Qual é a melhor resposta? Qual é o melhor resultado dada a incerteza? Quais são as melhores e significativamente diferentes escolhas?
Estatística	<i>Data Mining</i> Modelagem preditiva Aprendizagem de máquina Previsão Simulação	Otimização baseada em restrições Otimização multiobjetivo Otimização global
Gerenciamento da informação		

Fonte: Mineli M., Chambers M., Dhiraj A. (2013).

Conforme apresentado na Tabela 12, a análise preditiva foi criada para atuar com cálculos estatísticos avançados, que são usados para prever probabilidades de resultados futuros (MINELI, CHAMBERS E DHIRAJ, 2013). Alguns setores, como bancos, companhias de seguros e publicidade digital, já utilizam fortemente a análise preditiva, mas a maioria das outras empresas ainda está nos estágios iniciais de utilização. Enquanto isto, a ênfase estará na utilização estatística descritiva na fase exploratória de *big data*, quando o valor dos dados ainda não é conhecido e se move em direção aos estágios de análise preditivos e prescritivos quando se torna conhecido ou descoberto.

De acordo com Lohr, (2012), o poder preditivo de *big data* está sendo explorado - e se mostra promissor. Desta forma, *big data* está modificando a forma de como são feitas as análises, por meio de técnicas preditivas e prescritivas. Enquanto as estatísticas descritivas se tornaram populares, inicialmente pela SAS e SPSS, relatando acontecimentos passados, a análise preditiva, usa a informação passada para prever resultados futuros com algum grau de probabilidade, com vistas aos melhores resultados. Cada uma destas técnicas têm sido usadas



por décadas, entretanto há grandes mudanças em andamento, quando estas são combinadas com *big data* (MINELI M., CHAMBERS M., DHIRAJ A., 2013). Algumas destas maiores mudanças são:

- Utilizar todos ou o maior número de dados para criar um modelo preditivo;
- Combinar vários modelos analíticos e técnicas para melhorar os resultados;
- Criar um circuito fechado onde novos aprendizados são usados para adaptar os modelos de produção;
- Utilizar os modelos de previsão o mais próximo possível do tempo real;
- Concentrar a aplicação de técnicas de modelo preditivo (algoritmos), em vez de inventar novas técnicas.

Além disso, as pessoas que trabalham com *big data* precisam de habilidades de TI substanciais e criativas. Esta combinação é muito escassa, e como resultado, alguns pioneiros de *big data* estão trabalhando para desenvolver os seus próprios talentos. E por fim, *big data* está levando as organizações a repensarem suas suposições básicas sobre a relação entre negócios e TI e seus respectivos papéis.

Em termos de operacionalização destas premissas, Kandel, Paepcke, Hellerstein, e Heer (2012) propõem cinco tarefas de alto nível que podem ser executadas em processos de análise de *big data* e seus respectivos desafios: 1) descobrir, 2) adequar, 3) desenhar, 4) modelar e, 5) reportar. A Tabela 13 apresenta estas tarefas e seus desafios no processo de análise de *big data*:

Tabela 13 - Tarefas e desafios no processo de análise de *big data*.

(continua)

<b>Tarefa</b>	<b>Descrição</b>	<b>Exemplos</b>	<b>Desafios</b>
Descobrir	Descobrir dados necessários para completar tarefa de análise	Encontrar um conjunto de dados <i>on-line</i> , localizar tabelas em um banco de dados ou planilha	Localizar dados; definições de campo
Adequar	Adequar os dados para um formato desejado	Analisar os campos dos arquivos de <i>log</i> , integrando dados de várias fontes em um único arquivo ou extraindo as entidades de documentos	Integração de dados; análise semi estruturada; agregação e filtragem avançada

(conclusão)

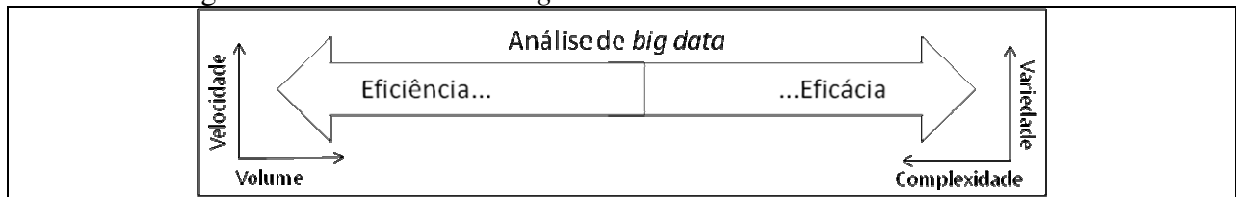
<b>Tarefa</b>	<b>Descrição</b>	<b>Exemplos</b>	<b>Desafios</b>
Desenhar	Desenhar o perfil dos dados para verificar a sua qualidade e a sua aptidão para as tarefas de análise	Inspecionar dados para identificação de <i>outliers</i> ou erros e analisar as distribuições de valores nos campos	Qualidade dos dados; verificação de hipóteses
Modelar	Modelar dados para sumarizar ou realizar previsões	Calcular estatísticas de resumo, modelos de regressão, ou executar agrupamento e classificação.	Seleção de atributos; escala; análises avançadas
Reportar	Reportar procedimentos e ideias para os consumidores da análise		Comunicação das suposições; relatórios estáticos

Fonte: Kandel, Paepcke, Hellerstein e Heer (2012).

Kandel, Paepcke, Hellerstein, e Heer (2012) esclarecem que os analistas estão envolvidos em um processo iterativo, não linear e que com isso, alternam entre as tarefas descritas na Tabela 13. Assim, também encontram desafios que passam pela migração de dados e operacionalização dos fluxos de trabalho. Na migração de dados, quando são usadas múltiplas ferramentas e bases de dados para completar as tarefas e, na operacionalização dos fluxos de trabalho, quando é gerado um número de produtos intermediários incluindo roteiros, planilhas eletrônicas, fórmulas e conjuntos de dados. Esta variedade de documentos, muitas vezes, torna difícil a composição de produtos que possam ter repetibilidade, confiabilidade e escalabilidade.

Akerkar (2014) menciona a eficácia e eficiência como padrões de desempenho na análise de *big data*. Assim, para que *big data* seja eficaz é necessário que: a) nem um único fato importante seja deixado de lado (o que significa integridade) e, b) que esses fatos sejam polidos adequadamente para posterior inferência (o que significa expressividade e granularidade). Eficiência pode ser interpretada como a razão de esforço gasto para que o resultado seja útil. Na análise de *big data*, a eficiência pode estar diretamente relacionada à pontualidade. Se um resultado não é oportuno, a sua utilidade pode ficar comprometida. A Figura 20 apresenta os padrões de desempenho de eficácia e eficiência e sua relação com as dimensões de *big data*:

Figura 20 - Dimensões de *big data* relacionadas à eficiência e eficácia.



Fonte: Akerkar (2014).

Assim, as dimensões volume e velocidade são importantes e estão relacionadas à questão da eficiência - grandes volumes de dados que podem mudar rapidamente tem que ser processados de forma rápida. Neste contexto, outras duas dimensões de *big data* tornam-se importantes. A variedade, como já mencionada, que descreve o grau de sintaxe e heterogeneidade semântica na distribuição dos módulos de dados que precisam ser integrados e harmonizados para análise. E a dimensão da complexidade, já citada por Beyer et al. (2011), que se refere a adequação das representações de dados e descrições para a análise. Estas duas dimensões estão relacionadas à questão eficácia. Entretanto, a maior complicação conceitual para análise é que eficiência está inversamente relacionada a eficácia.

### 2.2.2 Ciência de dados

Em meio à atividade de IE, a análise foi apresentada como parte central, tendo profundo impacto nos resultados. Com *big data*, este processo sofre algumas modificações, apresentadas na seção anterior por Davenport, Barth e Bean (2012) e Kandel, Paepcke, Hellerstein, e Heer (2012). Mineli, Chambers e Dhiraj (2013) também esclarecem que a capacidade de utilizar eficazmente a análise para a tomada de melhores decisões é o que irá definir o sucesso ou fracasso das organizações. Este pensamento é comprovado por meio de pesquisa realizada por McAfee e Brynjolfsson (2012).

Desta forma, mesmo com o suporte das ferramentas de *software* para tratar com *big data*, é o processo de tomada de decisão que possibilitará às organizações realizar as mudanças necessárias em busca de um posicionamento estratégico, que leve à vantagem competitiva. Dumbill (2012) coloca que o fenômeno do *big data* está intimamente ligado ao surgimento da disciplina de ciência de dados. Assim, beneficiar-se de *big data* significa investir em equipes com este conjunto de habilidades, e cercá-las com a vontade organizacional para entender e usar os dados para obter vantagem.

Em função disso, no mundo empresarial de hoje, cada função de negócios está questionando a necessidade de análises. No entanto, como esta é uma disciplina em evolução, as empresas estão investindo em análises que muitas vezes não são utilizadas no negócio. Assim, torna-se necessário que as partes interessadas invistam corretamente nas dimensões de processo, tecnologia e pessoas para que possa ser preenchida a lacuna entre a criação e o consumo da análise. Quanto mais cedo este resultado for alcançado, melhores serão as chances de alavancar o potencial de vantagem competitiva oferecido pela análise para a organização (MINELI, CHAMBERS E DHIRAJ, 2013).

Entretanto, segundo Manyika et al. (2011), o cenário que se apresenta é de escassez de talento analítico e gerencial para transformar o conhecimento em melhoria do desempenho organizacional. O mesmo relatório aponta que os Estados Unidos sozinhos deverão enfrentar até o ano de 2018, uma escassez de 140.000 a 190.000 pessoas com habilidades analíticas. Isso representa um *gap* de 50% a 60% na demanda. Também haverá falta de 1,5 milhões de gerentes e analistas para analisar os grandes volumes de dados e tomar decisões com base em suas descobertas. Para este resultado, embora a análise quantitativa seja com base nos Estados Unidos, o relatório declara que a restrição sobre este tipo de talento será global.

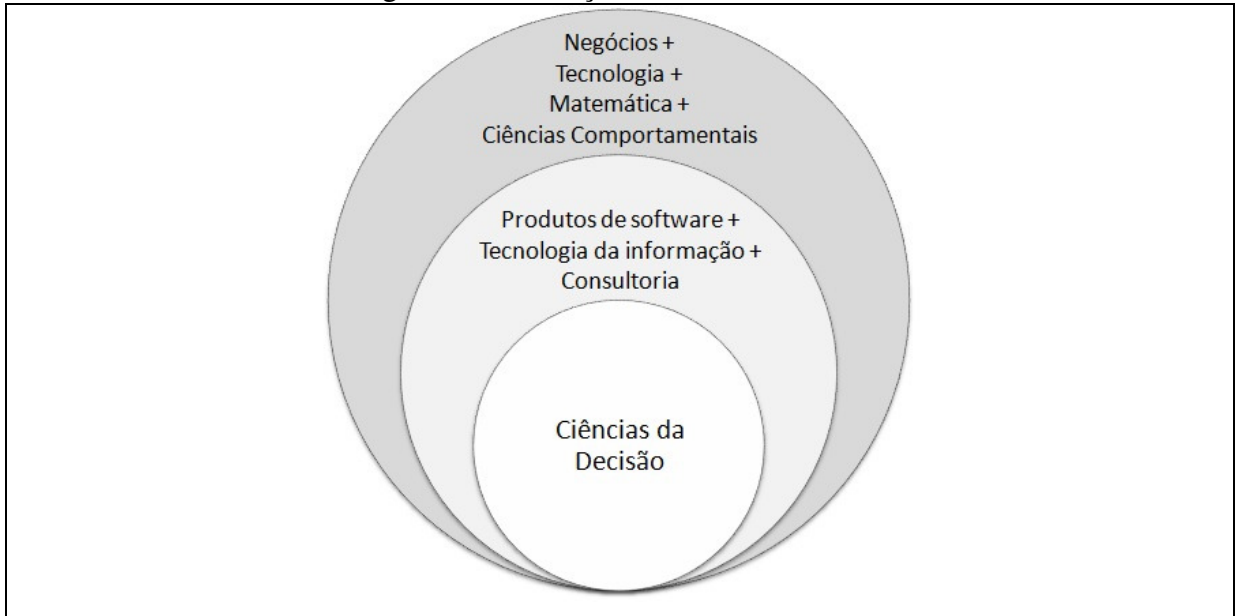
De acordo com Mineli, Chambers e Dhiraj (2013), cada tendência gera uma tendência contrária, e o campo de análise de dados não está imune a este fenômeno. Este campo tem sido beneficiado com a rápida emergência de novas ferramentas e tecnologias para explorar grandes volumes de dados. Entretanto, a disponibilidade de todos estes recursos tem contribuído em alguns aspectos, para análise de dados de parecer mais fácil e, isto pode gerar efeitos negativos ao processo.

De acordo com David Champagne, CTO (*Chief Technology Officer*) da *Revolution Analytics* que é a principal fornecedora comercial de *software* e serviços com base no projeto R - código aberto para computação estatística, o que se tem é uma nova abordagem para o método científico. Assim, na abordagem tradicional, primeiramente era feita uma pergunta, então era construída uma hipótese, e projetado um experimento. Este experimento era executado com a coleta e análise dos dados e eram tiradas conclusões. Após, os resultados eram comunicados e ocorriam as críticas.

Atualmente, em grande parte graças a todas as ferramentas e técnicas mais recentes disponíveis para manipulação de conjuntos cada vez maiores de dados, o processo se inicia com os dados. Então, modelos em torno dos dados são construídos e executados para que seja possível visualizar os resultados (MINELI, CHAMBERS E DHIRAJ, 2013). Para isso, a

Figura 21 aponta três possíveis configurações em termos da utilização de disciplinas no processo de análise, que se apresentavam no passado e que ainda permanecem:

Figura 21 - Evolução da ciência de dados.



Fonte: Adaptado de Mineli, Chambers e Dhiraj (2013).

A primeira é a aplicação de negócios, juntamente com a matemática, resultando na utilização de heurísticas e criatividade para tornarem persuasivos os argumentos aos tomadores de decisão. Esta configuração reflete a origem do negócio de consultoria. Outro cenário é a aplicação da disciplina de negócios com tecnologia, decorrendo na automatização de tarefas algorítmicas, melhorando assim a produtividade e eficiência dos processos. Este processo reflete a origem do negócio da TI. O terceiro cenário é a aplicação da tecnologia, aliada à matemática, gerando inspiração para a criação de produtos de *software* que ajudariam a lidar com uma ampla gama de problemas de negócios e operar de forma proativa, com antecipação.

Com *big data*, o que está se apresentando no futuro é a aplicação da disciplina de negócios, com matemática, tecnologia e, ciências comportamentais. As ciências do comportamento irão permitir conectar os pontos entre as interações e desenvolver uma compreensão mais profunda do comportamento humano. Isto permite criar os incentivos adequados para conduzir comportamentos que se alinham aos objetivos de negócios.

Para Mineli, Chambers e Dhiraj (2013) a sustentação de uma cultura de ciências da decisão irá precisar de uma combinação das seguintes características: aprendizagem inferencial sobre conhecimento experimental, agilidade, escala e convergência, talento

multidisciplinar, inovação e efetividade nos custos. A Tabela 14 aprofunda na descrição destas características:

Tabela 14 - Tarefas e desafios no processo de análise de *big data*.

(continua)

Característica	Descrição
Aprendizagem inferencial sobre conhecimento experimental	Aprender e saber são duas coisas completamente diferentes. Aprendizado vem com a experiência e está associado com correr riscos. Conhecer é mais sobre o conhecimento prévio ou, talvez, o resultado de ser "dito" que um determinado fato é verdadeiro, sem questioná-lo. É o inverso de aprendizagem e muitas vezes não é baseado em fatos. A habilidade para aplicar princípios de abordagens e estruturas para resolver o problema, em vez de confiar excessivamente em <i>expertise</i> de domínio passado, é importante para melhorar a capacidade da organização em descobrir novos <i>insights</i> e reduzir a lacuna existente entre percepções, decisões e ações.
Agilidade	Organizações irão necessitar projetar suas estruturas para serem ágeis nas três perspectivas: pessoas, processos e tecnologias. Estas implicações adotam modelos ágeis que habilitam interações rápidas, pensar antecipadamente, hipóteses e visualização. Não se pode esperar durante semanas para entender o desempenho de uma campanha ou entender o comportamento do cliente. Em tal ambiente, para permitir à empresa tomar decisões baseadas em dados, as profissões de ciências da decisão precisam adotar agilidade como um valor fundamental e precisam lidar com a contínua transformação em seu pensamento e ações.
Escala e convergência	Ecossistema sinérgico de talento, capacidades, processos, clientes e parcerias que podem ser alavancados através de verticais, domínios e geografias. Com a indefinição de limites de valor da cadeia e do surgimento de novos modelos de negócios uma nova era de convergência na utilização de técnicas e estruturas vai entrar em jogo. A aplicação de novos modelos de negócio em empresas já existentes é acelerada pela necessidade de convergência e colaboração. Exemplos podem ser citados, como a entrada da Microsoft no segmento de varejo, a entrada de provedores de rede de telefonia celular na categoria de <i>netbooks</i> , a Dell se deslocando de um modelo de negócio com configurações personalizadas para ofertas pré-construídas.

(conclusão)

<b>Característica</b>	<b>Descrição</b>
Talento multidisciplinar	Habilidade de aplicar negócios + tecnologia + matemática + comportamento econômico + antropologia social. Com esta junção de talentos tornando-se uma prática formal nos procedimentos operacionais e estratégias corporativas poderão ser desenvolvidos reparos cognitivos, que podem garantir uma melhor aplicação das decisões. Ao longo dos próximos anos, as empresas irão se mover em direção a uma abordagem holística para entender seus clientes.
Inovação	Aumentar largura e profundidade da resolução de problemas pela constante pesquisa e posicionamento de técnicas emergentes, tecnologias e aplicações. Muitas empresas investem em estratégias de inovação incrementais, usando os investimentos que fizeram em seus ativos de dados. Ao mesmo tempo, os verdadeiros inovadores disruptivos, se concentram em criar a combinação certa de análises qualitativas e quantitativas, trazendo novas áreas como economia comportamental e antropologia social para criar e executar ideias inovadoras.
Efetividade nos custos	Assegurar sustentabilidade e institucionalização da resolução de problemas na organização. Qualquer ideia ou conceito que precisa de adoção em larga escala tem de ser rentável. Por exemplo, se o custo de um computador pessoal não descesse a um nível que fosse aceitável para uma grande parte da população, ele nunca teria se tornado um item doméstico. Assim, organizações e profissionais do setor de ciências da decisão precisam se esforçar para fazer disto algo comum.

FONTE: Adaptado de Mineli, Chambers e Dhiraj (2013).

Schönberger-Mayer e Cukier (2013) também comentam, porém sem evidências de estudos mais aprofundados, que as decisões tomadas com base em dados disputarão ou superarão o julgamento humano. Segundo os autores, *big data* no processo de análise anuncia o fim da influência dos especialistas em muitas áreas, bem como da causalidade<sup>8</sup>. Nesta visão o especialista iria ceder lugar ao estatístico ou analista de dados, que não utilizará mais as velhas práticas, mas sim deixará os dados se manifestarem, baseando-se em correlações<sup>9</sup>. Para embasar esta reflexão, o autor cita o exemplo de Jeff Bezos, fundador da Amazon, que abriu mão dos críticos internos de livros quando os dados mostraram que as recomendações algorítmicas geravam mais vendas.

Na visão dos autores isto não significa que os especialistas desaparecerão, mas sim que sua supremacia diminuirá. Eles defendem que a especialidade, assim como a causalidade

8 Causalidade é o conjunto de relações de causa e efeito (SCHÖNBERGER-MAYER E CUKIER, 2013).

9 Correlações são representadas por uma relação mútua entre dois termos, matematicamente é a medida padronizada da relação entre duas variáveis (SCHÖNBERGER-MAYER E CUKIER, 2013).

é adequada ao mundo de pequenos dados, onde não há informações o bastante ou a informação certa, e assim é preciso confiar na intuição ou na experiência para se orientar. Nesse mundo, a experiência exerce papel fundamental, já que é o acúmulo de conhecimento latente (que não se pode transmitir ou aprender em um livro, talvez nem mesmo ter consciência dele) que permite a melhor tomada de decisão.

### **2.2.3 Aspectos humanos no processo de análise com *big data***

Fleisher e Bensoussan (2007) corroboram os aspectos abordados nas seções anteriores ao declararem que fazer melhores escolhas e tomar melhores decisões, pela primeira vez, geram uma maior necessidade de análise eficaz e inteligência. Os gestores de hoje enfrentam uma grande quantidade de informações em seus contextos de tomada de decisão, e às vezes essa abundância de informação faz com eles sejam paralisados. Neste contexto, os analistas têm os meios para ajudar a reduzir o volume e a taxa deste fluxo de informação, enquanto simultaneamente asseguram a qualidade do produto a ser entregue. Eles podem melhorar substancialmente a capacidade real do executivo e a confiança na eficaz tomada de decisão.

Os autores esclarecem que as organizações que podem efetivamente gerar, captar, disseminar e aplicar o conhecimento de uma forma melhor e mais rápida do que seus concorrentes estão com uma maior probabilidade de alcançar uma performance de sucesso. Os analistas são uma parte crítica desse processo orientado ao conhecimento. Eles são o elo com os dados recolhidos e as decisões-chave da organização.

Esta importância se dá pelo aumento no reconhecimento do valor das boas habilidades de pensamento. A análise não pode ser conduzida na ausência de pensar. Sem isso, seria uma escolha aleatória e de sorte. Assim, o instinto poderá ser aplicado juntamente com os mais fidedignos e testados métodos de análise. Como o acesso a dados e informações tem aumentado, o valor mais elevado é agora colocado em não apenas obter os dados apropriados, mas, mais importante ainda, fazer com que eles façam sentido. Com isso, os profissionais em inteligência modernos são estimulados a usar seu conjunto único de competências, conhecimentos, habilidades e instintos (FLEISHER E BENSOUSSAN, 2007).

Fuld (1995) apresentou à época, as características de bons pesquisadores e analistas: ser um bom ouvinte, ter criatividade, ter persistência, utilizar estratégia, ter experiência e, acessar coleções de relatórios corporativos. Contudo, em tempos de *big data*, Cukier (2010) declara que os *Chief Information Officers* (CIOs) têm tomado uma posição de maior destaque



nas organizações e que neste cenário surge um novo tipo de profissional chamado de cientista de dados.

Este profissional combina as habilidades de programador de *software*, estatístico e contador de histórias ou artista para extrair valor dos grandes volumes de dados. Davenport e Patil (2012) definem o cientista de dados, como um profissional de alto escalão com qualificação e curiosidade para fazer descobertas no mundo de *big data*. Entretanto, a designação tem poucos anos de vida. Foi cunhada em 2008, por D.J Patil, e por Jeff Hammerbacher, que na época lideravam iniciativas de dados e analítica no LinkedIn e no Facebook, respectivamente.

Este novo papel vem rapidamente ganhando destaque nas organizações. Entretanto, ainda são escassos os cursos de nível superior para formar esse tipo de profissional. Além disso, há pouco consenso sobre o lugar e o cargo que este profissional deveria ocupar na organização. Assim, Davenport e Patil (2012) colocam que o primeiro passo para localizar cientistas e dados é entender qual a sua função na empresa. Em seguida, indagar que qualificação este profissional precisa ter e em que áreas seria mais fácil encontrar pessoas com o currículo desejado.

Nesta abordagem o cientista de dados é alguém capaz de achar respostas para grandes dúvidas da empresa em meio a rios de informação desestruturada. Ele fica à vontade no universo digital e é capaz de estruturar grandes volumes de dados amorfos, tornando possível a análise. Este profissional identifica fontes abundantes de dados, combina isso com a informação potencialmente incompleta de outras fontes e depura o conjunto resultante. Em um cenário competitivo, onde os desafios estão sempre mudando, o cientista de dados tem função de ajudar os tomadores de decisão a migrar da análise pontual para o diálogo permanente com os dados (DAVENPORT E PATIL, 2012).

De acordo com os autores, o cientista de dados tem perfil para superar as limitações técnicas, não permitindo que estas prejudiquem a busca de novas soluções. Assim, a medida em que vai fazendo as descobertas, este profissional irá informar sobre sua aprendizagem e sugerir implicações de suas descobertas para novos rumos na empresa. Posteriormente, as informações serão exibidas em uma forma visualmente criativa, apontando os padrões detectados de forma clara e convincente. Esse profissional também irá orientar executivos e gerentes de produtos sobre implicações dos dados para produtos, processos e decisões.

Assim, o traço dominante do cientista de dados é uma intensa curiosidade - o desejo de ir além da superfície de um problema, de chegar às interrogações em seu cerne e de sintetizar

isso tudo em uma série de hipóteses bem claras que possam ser testadas. Para isso, é necessário o raciocínio associativo que caracteriza os cientistas mais criativos em qualquer campo. Por isso empresas tem obtido sucesso buscando indivíduos com formação acadêmica ou experiência profissional em ciências físicas ou sociais, porém com um foco forte em dados e computação (DAVENPORT E PATIL, 2012).

Os autores complementam dizendo que um analista quantitativo pode ser ótimo na análise de dados, mas precisa aprimorar suas habilidades para tratar com dados não estruturados e deixá-los num formato que permita a análise. Assim, um especialista em gestão de dados pode saber como ninguém gerar e organizar dados em formato estruturado, mas não ser bom em transformar não estruturados em estruturados, e tampouco analisá-los. Outros sem fortes habilidades sociais para terem um bom desempenho no papel de cientistas de dados, precisarão desenvolver o traquejo para se tornarem eficazes.

Mineli, Chambers e Dhiraj (2013) retratam que nem todos os analistas de dados estão totalmente confortáveis com a ideia de serem chamados de cientistas de dados. Mas há um sentimento emergente e de comum acordo de que, para ser verdadeiramente eficaz, um analista de dados deve ter principalmente a forte capacidade de análise de dados. Drew Conway, um ex-membro da comunidade de inteligência, e PhD em ciência política pela Universidade de Nova York, é agora *scientist-in residence* no IA Ventures. Na formulação de Conway, as principais habilidades para o cientista de dados são: habilidades *hackers*, conhecimento matemática e estatística e conhecimentos substantivos, *expertise* real. Resumindo, em sua visão o cientista de dados é um "*hacker* falador".

O termo faz sentido, uma vez que a maioria dos dados, hoje existe como uma sequência de "uns" e "zeros" em um banco de dados. Assim, os cientistas de dados não têm de ser *experts* em computação, mas eles têm que saber o caminho em torno do cenário de TI porque é lá que vivem os dados. As habilidades *hackers* são importantes, pois os dados tendem a residir em vários locais, e em vários sistemas. Assim, encontrar e recuperar requer, por vezes as habilidades de um "ladrão", mesmo quando os dados são de domínio público, ou de propriedade da própria organização, ou outra que tenha autorizado o seu uso. Assim, talvez o termo cientista de dados reflita o desejo de ver a análise de dados de retornando às suas raízes científicas (MINELI, CHAMBERS E DHIRAJ, 2013).

Os autores ainda colocam que a criação de uma cultura de análise, tomada de decisão baseada em dados e em evidências, exige novos conhecimentos ciência de dados/decisão,

habilidades e valores. Assim, os traços necessários em cada membro da equipe para fazer a ciência da decisão sustentável estão expostos na Tabela 15:

Tabela 15 - Resumo dos atributos de qualidade da informação.

<b>Características necessárias</b>	<b>Descrição</b>
Aprendizagem ao longo do saber	Capacidade de aplicar os primeiros princípios e abordagens estruturadas para resolver problemas em vez de confiar excessivamente na <i>expertise</i> de domínio passado.
Agilidade	Capacidade de lidar com a transformação contínua.
Escala e convergência	Ecossistema sinérgico de talentos, capacidades, processos, clientes e parceiros que podem ser aproveitados em mercados verticais, domínios e geografias.
Talento multidisciplinar	Habilidade de aplicar negócios, matemática, tecnologia e ciências comportamentais.
Inovação	Aumentar a amplitude e a profundidade da resolução de problemas por pesquisa constante e implantação de técnicas emergentes, tecnologias e aplicações.
Relação custo-eficácia	Garantir a sustentabilidade e a institucionalização da resolução de problemas nas organizações.

Fonte: Mineli, Chambers e Dhiraj, 2013.

Em 2010, Drew Conway, desenvolveu um diagrama de Venn, mostrando as habilidades sobrepostas de um cientista de dados. Conway iniciou sua carreira como um cientista social computacional na comunidade de inteligência dos EUA e tornou-se um especialista na aplicação de métodos computacionais para os problemas sociais e comportamentais em grande escala (BARLOW, 2013). Essas habilidades podem ser visualizadas na Figura 22:

Figura 22 - Habilidades sobrepostas de um cientista de dados.



Fonte: Barlow, 2013.

Barlow (2013) expõe que na perspectiva de Conway, um cientista de dados deve dominar todas as três áreas, embora seja raro um indivíduo que possa encarnar todas as habilidades de um cientista de dados, para desempenhar um papel útil em um time de grandes análises. Assim, Kandel, Paepcke, Hellerstein, e Heer (2012) no *paper* intitulado "*Enterprise Data Analysis and Visualization: An Interview Study*", mostram resultados de 35 entrevistas com analistas de dados que trabalham em organizações comerciais na área da saúde, varejo, finanças e redes sociais. Com base nestas entrevistas, os pesquisadores identificaram três arquétipos básicos de analistas: (1) *hacker*; (2) roteirista; (3) usuário de aplicação.

O *hacker* é tipicamente um programador fluente e manipulador de dados. O roteirista executa a maioria de seu trabalho dentro de um pacote de *software* existente e trabalha principalmente em dados que foram obtidos a partir de *datawarehouse* pela equipe de TI. O usuário do aplicativo se baseia em planilhas ou aplicativos especializados, que funcionam normalmente em conjuntos de dados menores do que os *hackers* e roteiristas.

Assim, os autores esclarecem que é importante gerenciar e entender as diferenças entre estes tipos de analistas quando se está trabalhando em um time. *Hackers* são mais propensos a terem uma formação em ciência da computação, mas menos facilidade com estatística e alguns dos aspectos mais científicos da ciência de dados. Eles têm um conhecimento contextual das questões de domínio específico a serem exploradas nos dados. Por outro lado, roteiristas são mais propensos a serem treinados em estatística e os usuários de aplicativos são mais propensos a serem pessoas de negócios. Para ambos, os recursos de *software* podem

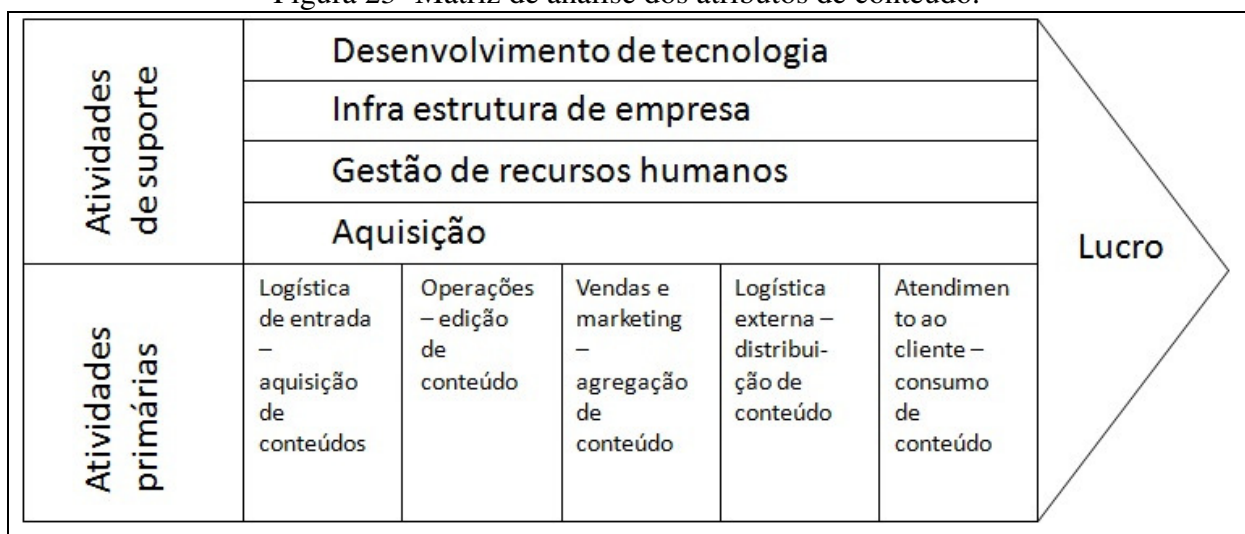
contribuir para tornar a preparação de dados menos penosa e diminuir as distinções entre os três tipos de analistas descritos.

Ao final das últimas seções, pode-se analisar em teoria, como o fenômeno *big data* está gerando mudanças no contexto da IE. Foram apresentados pontos de mudança em aspectos do processo de análise e também nas características profissional de inteligência. Estes elementos forneceram subsídios para o atingimento do objetivo geral deste estudo que é de examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica. Assim, o próximo passo a ser tratado refere-se ao potencial de agregação de valor destas mudanças, também em teoria. Ou seja, quais serão os prováveis benefícios de *big data*, se incluído no processo de análise para a atividade de inteligência estratégica. Posteriormente estas constatações serão confrontadas com a opinião dos profissionais entrevistados.

### 2.3. VALOR DE *BIG DATA* PARA ORGANIZAÇÃO

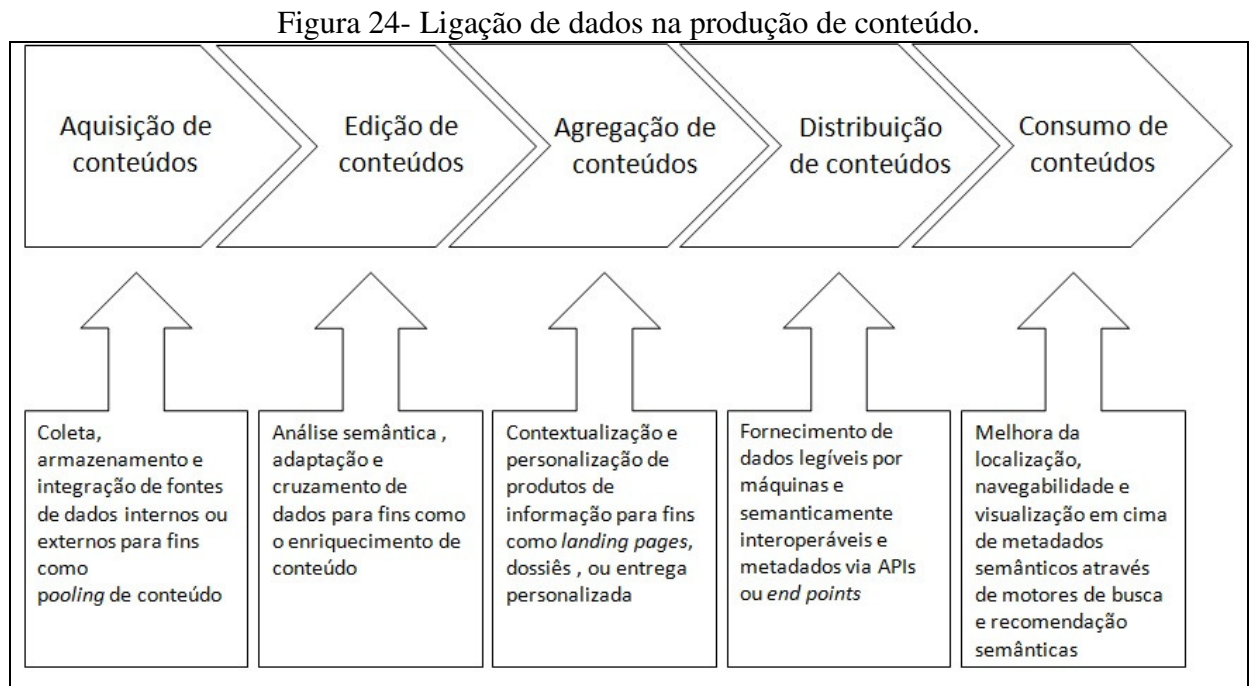
A análise dos atributos de conteúdo em cada uma das atividades dos processos organizacionais, tem sido utilizada para distinguir a lógica de produção de setores informacionais, como: impressão, transmissão, telecomunicações, software e mídia on-line entre si, e ilustrar as suas especificidades com relação a ativos de informação e modelos de receita (ZERDICK ET. AL., 1999). A Figura 23 ilustra as atividades dos processos organizacionais, e suas relações com a produção de conteúdo:

Figura 23- Matriz de análise dos atributos de conteúdo.



Fonte: AKERKAR, 2014 (Adaptado de PORTER, 1985).

Assim, as especificidades da criação de conteúdo podem ser descritas em cinco etapas: a) aquisição de conteúdo, b) edição de conteúdo, c) agregação de conteúdo, d) distribuição de conteúdo e, e) consumo de conteúdo. A Figura 24 expõe que a ligação dos dados pode contribuir para apoiar a função de produção de conteúdo associada:



Fonte: AKERKAR, 2014 (Adaptado de Pellegrini, 2012).

Na abordagem do autor, a aquisição de conteúdo é principalmente a preocupação com a coleta, armazenamento e integração de informações relevantes, necessárias para produzir um item de conteúdo. No decurso deste processo, a informação é reunida a partir de fontes externas ou internas para processamento adicional. O processo de edição envolve todas as medidas necessárias que lidam com a adaptação semântica, ligação, e enriquecimento de dados. Na adaptação, os dados adquiridos são apresentados de forma que possam ser reutilizados em processos editoriais. Interligação e enriquecimento podem ser realizados por meio de processos como anotação e/ou referência, quer por esclarecer conceitos existentes ou fornecer um conhecimento mais aprofundado.

Para Akerkar (2014), o processo de agregação está preocupado principalmente com a contextualização e personalização de produtos de informação. Ele pode ser usado para fornecer acesso personalizado a informações e serviços, utilizando metadados (dados sobre dados), para a entrega de conteúdo sensível ao dispositivo ou para compilar o material

tematicamente relevante em *landing pages* ou dossiês, melhorando assim a navegabilidade, localização, e reutilização de informações. Já o processo de distribuição de conteúdo, em um ambiente de dados ligados, lida principalmente com o fornecimento de dados legíveis por máquinas e semanticamente interoperáveis, através da *Application Programming Interfaces* (APIs) ou *SPARQL Endpoints* (ZIMMERMANN, 2011). Estes podem ser concebidos para servir à fins internos, em ambientes controlados, e para fins externos, com o compartilhamento de dados entre usuários anônimos.

O último passo da produção de conteúdo é lidar com o seu consumo. Isto implica em quaisquer meios que permitam ao usuário procurar e interagir com itens de conteúdo de uma forma agradável e útil. Assim, esta etapa trata principalmente de aplicativos para o usuário final, que fazem uso das ligações de dados para fornecer acesso a itens de conteúdo e gerar uma percepção mais profunda sobre eles.

Neste sentido, Schönberger-Mayer e Cukier, (2013) apresentam alguns atores que detêm o valor de *big data*: detentores de dados, especialistas em dados e empresas e pessoas com mentalidade de *big data*. Os detentores de dados são os atores que controlam o acesso à informação e a usam ou licenciam para que outros possam extrair valor dela. Algumas empresas se posicionaram no centro do fluxo da informação e por isto podem alcançar escala e captar o valor dos dados.

Um exemplo desta categoria é a indústria de cartões de crédito nos Estados Unidos. De acordo com os autores, durante muito tempo, o alto custo da batalha contra fraudes levou alguns bancos a evitarem lançar cartões de crédito. Estas operações foram encaminhadas para instituições financeiras maiores, que tinham o tamanho e a escala para investir na tecnologia apropriada. Esta terceirização acabou por privar os bancos dos dados sobre padrões de gastos, que lhes permitiriam saber mais sobre clientes, podendo agregar soluções customizadas. Com isso, empresas como MasterCard licenciaram dados para terceiros, criando outras divisões como MasterCard Advisors, que atualmente agrega e analisa 65 bilhões de transações dos 1,5 bilhão de clientes em 210 países, afim de prever tendências corporativas e de consumo e depois vender estas informações ao mercado (SCHÖNBERGER-MAYER E CUKIER, 2013).

A segunda categoria apontada pelos autores é a dos especialistas em dados. Empresas com o conhecimento ou a tecnologia para fazer análises complexas. Um exemplo que pode ser citado é o da consultoria Accenture, que trabalha com empresas de vários segmentos, empregando avançadas tecnologias e sensores sem fio para analisar os dados coletados. Esta empresa realizou um trabalho na cidade de St. Louis, no Missouri, onde instalou sensores sem

fio em vários ônibus públicos para monitorar os motores, prever quebras e/ou determinar a hora mais adequada para realizar a manutenção. Com isso, diminuiu os custos dos serviços em até 10%, com economias de mais de US\$ 1 mil por veículo. As técnicas utilizadas por estas empresas podem chegar a correlações surpreendentes, com descobertas realizadas por máquinas. Uma pessoa que incansavelmente estudasse os dados, talvez nunca fizesse a mesma descoberta.

A terceira categoria é composta por empresas e pessoas com mentalidade de *big data*, que veem as oportunidades antes dos outros, mesmo que não possuam dados ou habilidades para agir em relação às oportunidades. A noção de mentalidade de *big data* é o papel de um observador externo criativo, com uma ideia brilhante. Um exemplo que pode ser citado é o de Bradford Cross. Ele e mais alguns amigos, criaram a Flight Caster que realizava previsões sobre possíveis atrasos de voos nos Estados Unidos. Para isto, eles analisavam todos os voos dos 10 anos anteriores e os comparavam com dados antigos e atuais sobre o clima. As previsões da empresa eram tão precisas que até mesmo os funcionários das companhias aéreas as utilizavam, uma vez que estas divulgavam os atrasos apenas no último minuto (SCHÖNBERGER-MAYER E CUKIER, 2013).

Esta visão é compartilhada por Ross, Beath e Quaadgras (2013), que propõe três tipos de organizações que estão colhendo valor do processamento massivo de grandes volumes de dados: a) empresas com uma tradição de tomada de decisão baseada em fatos, b) empresas de pesquisa em engenharia e, c) empresas que dependem da análise de *big data* para tomar decisões operacionais críticas.

Como exemplos, podem ser citadas as empresas Procter & Gamble - que em 1920 tornou-se uma das primeiras a tomar decisões significativas sobre produtos e publicidade com base em dados, e a UPS - que em 1980, começou a monitorar os movimentos de seus veículos e encomendas. A ExxonMobil, que inventou a tecnologia sísmica em 3-D, revolucionando a forma como a indústria de petróleo e gás decide onde perfurar. Hoje, cientistas e engenheiros já usam análise 4-D para reduzir ainda mais os custos e os riscos de exploração. Empresas que se conectam com os clientes unicamente pela *internet* podem capturar uma enorme quantidade de dados sobre o comportamento do consumidor.

Akerkar (2014) coloca que os impactos de *big data* estão presentes nos negócios e na ciência. Nos negócios, a rápida análise de *big data* se traduz em melhor satisfação dos clientes, melhores serviços, e, pode também contribuir para a criação e manutenção de um negócio bem sucedido. Por exemplo, empresas coletam dados e fazem o monitoramento



contínuo de seus clientes para saberem o que eles dizem e como interagem com os sistemas, em um site de *e-commerce*, em uma rede social, ou em um sistema de navegação GPS, entre outros. Esses dados, uma vez coletados, precisam ser analisados rapidamente, a fim de descobrir as tendências interessantes. Desta forma, estas empresas podem obter vários *terabytes* de dados em uma base diária e os tipos de tendências observadas podem mudar de um dia para outro, de uma hora para outra.

Neste contexto, Dumbill (2012) refina o valor de *big data* para uma organização em termos de duas categorias: as de uso analítico, e permitindo a criação de novos produtos. Atualmente a análise de *big data* pode revelar percepções ocultas, pois antes os dados eram muito caros para o processo. Assim, ser capaz de processar cada item de dados em tempo razoável elimina a necessidade problemática pela amostragem e promove uma abordagem investigativa de dados, em contraste com a natureza um tanto estática de execução de relatórios pré-definidos.

*Startups* de sucesso da *web*, da última década, são exemplos de *big data*, utilizado como um facilitador de novos produtos e serviços. Por exemplo, através da combinação de um grande número de sinais a partir de ações de um usuário e de seus amigos, o Facebook tem sido capaz de criar uma experiência de usuário altamente personalizada e criar um novo tipo de negócio de publicidade. Não é coincidência que a maior parte das ideias e ferramentas que sustentam *big data* tenham surgido a partir do Google, Yahoo, Amazon e Facebook (DUMBILL, 2012).

*Big data* já atingiu todos os setores da economia mundial. Ele está transformando as oportunidades competitivas em todos os setores da indústria, incluindo bancos, saúde, seguros, manufatura, varejo, atacado, transporte, comunicações, construção, educação e serviços públicos. Ele também desempenha um papel chave nas operações comerciais, tais como *marketing*, operações, cadeia de suprimentos e novos modelos de negócios. Assim, está se tornando evidente que as empresas que não conseguirem usar os seus dados de forma eficiente estarão em desvantagem competitiva se comparadas àquelas que analisam e atuam sobre os dados (AKERKAR, 2014).

O impacto da abundância de dados se estende bem além dos negócios. Justin Grimmer, por exemplo, é um dos cientistas políticos mais importantes da nova geração de. Há vinte e oito anos como professor assistente na Universidade de Stanford, ele combinou a matemática com a ciência política na sua graduação e pós-graduação, vendo "uma oportunidade, porque a disciplina está se tornando cada vez mais intensiva de dados". Sua

pesquisa envolve a análise automatizada por computador de postagens de *blogs*, discursos do Congresso e comunicados de imprensa e notícias, em busca de *insights* sobre como disseminar ideias políticas (LOHR, 2012).

Para Akerkar (2014), na ciência, a rápida análise de *big data* pode fornecer contribuições positivas. Todas as ciências hoje lutam com o gerenciamento de dados, por exemplo, astronomia, biologia, entre outros. Entretanto, a expectativa é que nos próximos anos, a capacidade de coletar dados antecipadamente irá aumentar. Por exemplo, o Telescópio Large Synoptic Survey - projeto de especialistas dos EUA, possui atualmente uma coleção diária de 20 *terabytes* de dados, o Large Hadron Collider - CERN na Europa, já possui uma quantidade de dados ainda maior. Com vários *terabytes* em uma base diária, a exploração de dados torna-se essencial, a fim de permitir que os cientistas consigam focar rapidamente em partes onde há uma boa probabilidade de encontrar observações interessantes.

Em termos de pesquisa científica, Schönberger-Mayer e Cukier (2013) complementam que *big data* pode oferecer um novo olhar em relação ao método científico, pelo fato de atualmente expressarmos mais aspectos da realidade na forma de dados. Assim, embora todo o processo de coleta e análise se baseie em teorias, com a dataficação a compreensão de um fenômeno será cada vez mais motivada por dados que por hipóteses. Os autores colocam que no futuro, as correlações de *big data* serão rotineiramente usadas para discordar de nossa intuição causal, mostrando que geralmente há pouco ou nenhuma conexão estatística entre o efeito e o que se supõe ser a causa. Entretanto, no momento, a utilização mais avançada da dataficação é no mundo dos negócios.

Voltando a este foco, Manyika et al. (2011) expõe no relatório "*Big data: the next frontier for innovation, competition, and productivity*", números crescentes em termos de geração de dados, em função de que estes são criados como subprodutos que outras atividades, que hoje se incorporam ao mundo físico, em dispositivos como telefones celulares, medidores inteligentes de energia, automóveis e máquinas industriais. Outro aspecto que deve ser considerado, refere-se ao fato de que hoje é fisicamente impossível armazenar tudo. De acordo com Manyika et al. (2011) prestadores de cuidados de saúde, por exemplo, tem que descartar 90% dos dados que geram.

O mesmo relatório, por meio de entrevistas com especialistas da indústria, revisão da literatura atual e identificação de alavancas específicas, quantificou o potencial de geração de valor dos grandes volumes de dados e os habilitadores necessários ao seu uso (MANYIKA ET AL. 2011).

Assim, segundo Manyika et al. (2011), o fenômeno *big data* tem potencial de criação de valor por meio de cinco alavancas: criar transparência; experimentação, permitindo descobrir as necessidades, expor variabilidade, e melhorar o desempenho; segmentação de populações e ações personalizadas; substituição, apoiando a tomada de decisão humana com algoritmos automatizados; e, modelos inovadores de negócios, produtos e serviços. O potencial de geração de valor adicional dos grandes volumes de dados, foi organizado em cinco domínios, que juntos representaram cerca de 40% do PIB global em 2010, onde explorar o potencial dos grandes volumes de dados: a) cuidados de saúde nos Estados Unidos, b) administração do setor público na União Europeia, c) varejo nos Estados Unidos, d) manufatura (global) e, e) dados de localização pessoal (global).

O domínio saúde é um segmento grande e importante da economia dos EUA, responsável por mais de 17 por cento do PIB e que emprega em torno de 11 por cento dos trabalhadores do país. A histórica taxa de crescimento dos gastos com saúde vem aumentando anualmente, cerca de 5 por cento ao longo da última década. Os dados deste setor estão fragmentados em quatro áreas: dados clínicos, dados de sinistros e custos, dados de pesquisa e desenvolvimento para produtos farmacêuticos e dados de comportamentos e sentimentos dos pacientes. Se fossem removidas as barreiras estruturais dos sistemas de saúde e com os investimentos necessários, *big data* poderia ajudar a desbloquear mais de US\$ 300 bilhões por ano neste setor.

Manyika et al. (2011) buscou encontrar formas sobre como *big data* poderia auxiliar no aumento da produtividade do setor público com foco na Europa. A pesquisa mostrou que este setor poderia, potencialmente, reduzir os custos de atividades administrativas - 15 a 20 por cento, criando o equivalente a € 150 bilhões para € 300 bilhões (\$ 223.000.000.000 para 446.000.000 mil dólares). Esta estimativa inclui ganhos de eficiência e uma redução da lacuna entre a atual e a potencial cobrança de receitas fiscais. Estas aplicações poderiam acelerar a produtividade anual - crescimento em até 0,5 pontos percentuais - ao longo dos próximos dez anos. O relatório coloca que *big data* também poderia desempenhar um papel similar em outros países e regiões.

No setor de varejo dos EUA, o uso da tecnologia da informação e dados digitais já tem impulsionado a rentabilidade e a produtividade durante décadas. Nos próximos anos, a adoção contínua e o desenvolvimento de aplicações para grandes volumes de dados tem o potencial para aumentar a produtividade do setor em pelo menos 0,5 por cento ao ano até 2020. Entre as empresas individuais, estas aplicações poderiam aumentar as margens operacionais em mais

de 60 por cento para aqueles pioneiros que maximizam o uso de grandes volumes de dados. Tal impulso na rentabilidade seria especialmente significativo em um setor onde as margens são notoriamente justas. Esta também seria uma forma para examinar as interações entre varejistas e consumidores.

O setor manufatureiro foi um dos primeiros a utilizar-se de dados para orientar qualidade e eficiência, adotar tecnologia da informação e automação para projetar, construir e distribuir produtos. Em especial este setor enfrenta o desafio de gerar produtividade significativa e melhorias nas indústrias que já se tornaram relativamente eficientes. Os ganhos advindos de uma nova onda virão numa maior eficiência na concepção e produção, melhorias na qualidade do produto, e desenvolvimentos de acordo com as necessidades dos clientes, com mecanismos eficazes de promoção e distribuição. Por exemplo, grandes volumes de dados podem ajudar os fabricantes a reduzir o tempo de desenvolvimento de produto de 20 a 50 por cento e eliminar os defeitos anteriores à produção através de simulação e testes. Usando dados em tempo real, as empresas podem também gerir o planejamento de demanda em empresas ampliadas e cadeias globais de abastecimento, ao mesmo tempo reduzir os defeitos e retrabalho na produção plantas.

De acordo Manyika et al. (2011), pode-se assistir a uma explosão na quantidade de informação disponível sobre o local em que as pessoas estão no mundo. Tecnologias como GPS permitem localizar rapidamente um dispositivo pequeno como um telefone celular, dentro de algumas dezenas de metros. Estes dados estão sendo usados para criar uma onda de novos negócios e modelos de negócios inovadores que estão tocando a vida das pessoas ao redor do globo. Ao contrário de outros domínios, as aplicações para estes dados passam por muitas indústrias, incluindo telecomunicações, varejo e mídia. Este domínio oferece o potencial para criação de um novo valor ao longo dos próximos dez anos que foi estimado em mais de US \$ 100 bilhões em receitas para prestadores de serviços e até US\$ 700 bilhões em valor para consumidores e empresas usuários finais. Capturar esse valor exigirá investimentos em tecnologia, infraestrutura e pessoal, bem como ações apropriadas do governo.

A percepção do *big data* no contexto brasileiro não é generalizada e sim setorizada. Assim uma oportunidade relacionada ao uso de *big data* é o setor energético. Neste segmento, as companhias de petróleo estão utilizando sensores distribuídos, comunicações de alta velocidade e *data mining* para monitorar e aperfeiçoar as operações de perfuração em locais remotos. O objetivo é utilizar dados em tempo real para tomar melhores decisões e prever falhas (SIEGELE, 2013).

De acordo com Paul Siegele, presidente da Companhia de Tecnologia de Energia da Chevron, mesmo com o avanço das tecnologias renováveis, a demanda por petróleo ainda será de crescimento no ano de 2035. Entretanto, atualmente a extração tornou-se mais difícil e quase US\$ 20 trilhões de investimentos serão necessários para atender às necessidades futuras. Com utilização de dados em tempo real para tomar decisões de colaboração em operações de perfuração, ou gestão de poços e reservatórios de imagem para rendimentos mais elevados de produção, estima-se economizar R\$ 1 bilhão por ano (SIEGELE, 2013).

Para Akerkar (2014) as possibilidades para que *big data* continue a se desenvolver rapidamente é crescente e se mantém impulsionada pela inovação nas plataformas de tecnologias subjacentes e capacidades analíticas para a manipulação de dados, bem como a evolução do comportamento entre as pessoas, que cada vez mais vivem vidas digitais. Em contrapartida, Ross, Beath e Quaadgras (2013), relatam que, algumas empresas que estão investindo pesadamente em cientistas de dados, bancos de dados e *software*, não tem visto muitos resultados para seus esforços. A maior razão para isto é que as empresas não fazem um bom trabalho com as informações que já têm em mãos, não sabendo como controlá-las, analisá-las de forma a melhorar a sua compreensão e, logo, tomar ações em respostas às descobertas.

Schönberger-Mayer e Cukier (2013) colocam que neste caso, os executivos confiam em si mesmos por instinto e agem de acordo. Porém, isto está começando a mudar à medida que as decisões gerenciais são tomadas ou, pelo menos confirmadas por modelos de precisão e análise de *big data*. Assim, de acordo com Ross, Beath e Quaadgras (2013), a tomada de decisões com base em evidências e análise de dados fornece respostas, mas isso também requer uma grande mudança cultural. Os autores sugerem quatro alterações na forma de como as operações são gerenciadas: determinar uma única fonte de verdade, usar marcadores de desempenho, gerenciar as regras de negócios de forma explícita e, treinar para melhorar o desempenho.

### 3 MÉTODO DA PESQUISA

As pesquisas qualitativas se caracterizam por proporcionar ao pesquisador a observação do comportamento real de um fenômeno. Assim, se distinguem das quantitativas, no sentido em que possibilitam o foco na interpretação, subjetividade e flexibilidade de condução. Possuem ênfase no entendimento da situação pesquisada e não num objetivo predeterminado. Neste contexto, o pesquisador exerce influência sobre a situação de pesquisa e é influenciado por ela (MOREIRA, 2002). De acordo com Gil (2010), as pesquisas exploratórias proporcionam maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses.

Para Bardin (1977), a análise de conteúdo é um conjunto de técnicas de análise das comunicações que possui duas funções: 1) heurística - enriquece a tentativa exploratória e aumenta a propensão à descoberta e, 2) administração da prova - trabalha com hipóteses sob a forma de questões ou de afirmações provisórias servindo de diretrizes para serem verificadas no sentido de uma confirmação ou eliminação.

Moraes (1999) sugere a análise de conteúdo como sendo constituída por cinco etapas: 1) preparação das informações, 2) unitarização ou transformação do conteúdo em unidades, 3) categorização ou classificação das unidades em categorias, 4) descrição e, 5) interpretação. A etapa de interpretação pode ser facilitada com o uso de *softwares* como ATLAS/ti (GIL, 2010).

Esse estudo apresenta-se como qualitativo exploratório, por tratar de um tema novo como *big data*, em que há pouca teoria disponível para orientar as previsões. Neste caso, o estudo qualitativo é o mais adequado para que seja atingido o objetivo de examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica, pois permite a observação do comportamento real do fenômeno e sua influência na atividade de IE. A técnica de análise de conteúdo com apoio do *software* ATLAS/ti permitiu identificar o conteúdo e as características de informações presentes nos textos das entrevistas realizadas e, por meio disso, encontrar respostas para o problema de pesquisa definido e, demais objetivos deste estudo.

### 3.1 PARTICIPANTES DA PESQUISA

As concepções da pesquisa qualitativa utilizam estratégias graduais de amostragem, baseadas, em grande parte, na amostragem teórica (FLICK, 2009). De acordo com Glaser e Strauss (1967), amostragem teórica é o processo de coleta de dados, pelo qual o analista coleta, codifica e analisa conjuntamente seus dados, decidindo que dados coletar a seguir e onde encontrá-los. De acordo com Flick (2009), as decisões relativas à amostragem podem ser tomadas no nível dos grupos a serem comparados ou podem concentrar-se diretamente em pessoas específicas.

O autor explica que na amostragem teórica, indivíduos ou grupos são selecionados de acordo com critérios que digam respeito a seu conteúdo, em vez de aplicar critérios metodológicos abstratos, próprios da amostragem estatística. Assim, o prosseguimento da amostragem ocorre de acordo com a relevância dos casos, e não conforme a sua representatividade. Flick (2009) esclarece que este é um princípio característico da pesquisa qualitativa.

Assim, os participantes da pesquisa foram selecionados por meio de amostragem teórica e obtidos a partir da técnica *snowball sampling* (bola de neve). Esta técnica teve origem no trabalho de Goodman (1961), sendo utilizada em pesquisas sociais, onde os participantes iniciais (sementes) indicam novos participantes (filhos ou frutos), que por sua vez indicam novos participantes e assim sucessivamente, até que seja alcançado o objetivo proposto, ou ponto de saturação (esgotamento dos membros acessíveis).

Segundo Goodman (1961), em alguns estudos, os participantes contatados na aplicação da pesquisa recrutam o maior número de pessoas possível; em outros, os próprios pesquisadores podem efetuar esse recrutamento, por meio de agentes que atuam em um dado campo, com conhecimento aprofundado e trânsito em uma dada comunidade (são os *outreach workers*). Nesta pesquisa, os participantes foram selecionados pela autora por meio dos seguintes agentes: Associação Brasileira dos Analistas de Inteligência Competitiva (ABRAIC), IBM Brasil - Indústria, Máquinas e Serviços Ltda. - uma das subsidiárias da IBM *World Trade Corporation*, Sociedade Brasileira de Gestão do Conhecimento (SBGC) e Universidade de Caxias do Sul (UCS). Estas instituições foram designadas para a seleção dos participantes, pois possuem as características citadas por Goodman (1961) em termos de atuação, conhecimento e trânsito nos campos de inteligência estratégica e/ou *big data*.

A ABRAIC foi criada em 2000, por um grupo de profissionais de várias organizações brasileiras que realizaram cursos em nível de pós-graduação em inteligência competitiva no Brasil, na França e na Bélgica, e outros que já atuavam em áreas afins. O principal objetivo da associação é a disseminação da mentalidade de inteligência competitiva no país, visando o aumento da competitividade das organizações que atuam no Brasil, de forma a garantir espaço para essas organizações em um mundo globalizado. Sua sede é em Brasília - DF e atualmente a ABRAIC possui em torno de 2070 analistas e 50 empresas associadas em todo o território nacional.<sup>10</sup>

De acordo com Bertolucci (2013), a IBM atualmente configura entre os líderes em aquisições de empresas de análises de dados de *big data*. Ela tem se concentrado em empresas e produtos que possam auxiliar na entrega de "análises mais inteligentes", incluindo aumento da simplicidade e integração de vários componentes de *software*, que estejam alinhados com a sua estratégia de mercado. A IBM também vem trabalhando para abordar a crescente necessidade de cientistas de dados e aplicações de *big data*. Para isto, tem construído novas parcerias com o Estado de Michigan, a Universidade Northwestern, Yale e com a Universidade do Sul da Califórnia, todas nos Estados Unidos.

A SBGC desenvolve suas atividades desde 2001, com o objetivo estimular a gestão do conhecimento no Brasil. Para isto, reúne profissionais e organizações em um grande fórum de discussão sobre os temas como: inovação e aprendizagem organizacional, colaboração e redes de valor, inteligência competitiva e de negócios, gestão de capital intelectual, economia criativa e trabalho, dentre outros de relevância para a gestão do conhecimento. Atualmente a SBGC conta com uma unidade nacional em São Paulo e quatro regionais nos estados de Rio Grande do Sul, Rio de Janeiro, Paraná e no Distrito Federal. Possui 26.000 membros cadastrados, sendo mais de 5 mil empresas nacionais<sup>11</sup>.

A UCS é uma instituição de ensino superior, de caráter comunitário e regional, com atuação direta na região nordeste do estado do Rio Grande do Sul. A UCS atualmente possui núcleos de pesquisa voltados para inteligência e no ano de 2009 recebeu o 1º lugar no II Prêmio Excelência em Inteligência Competitiva, promovido pela ABRAIC, em parceria com o *Internacional Quality & Productivity Center (IQPC)*. O prêmio reconhece projetos que demonstram as melhores práticas das organizações em inteligência competitiva, uma vez que

---

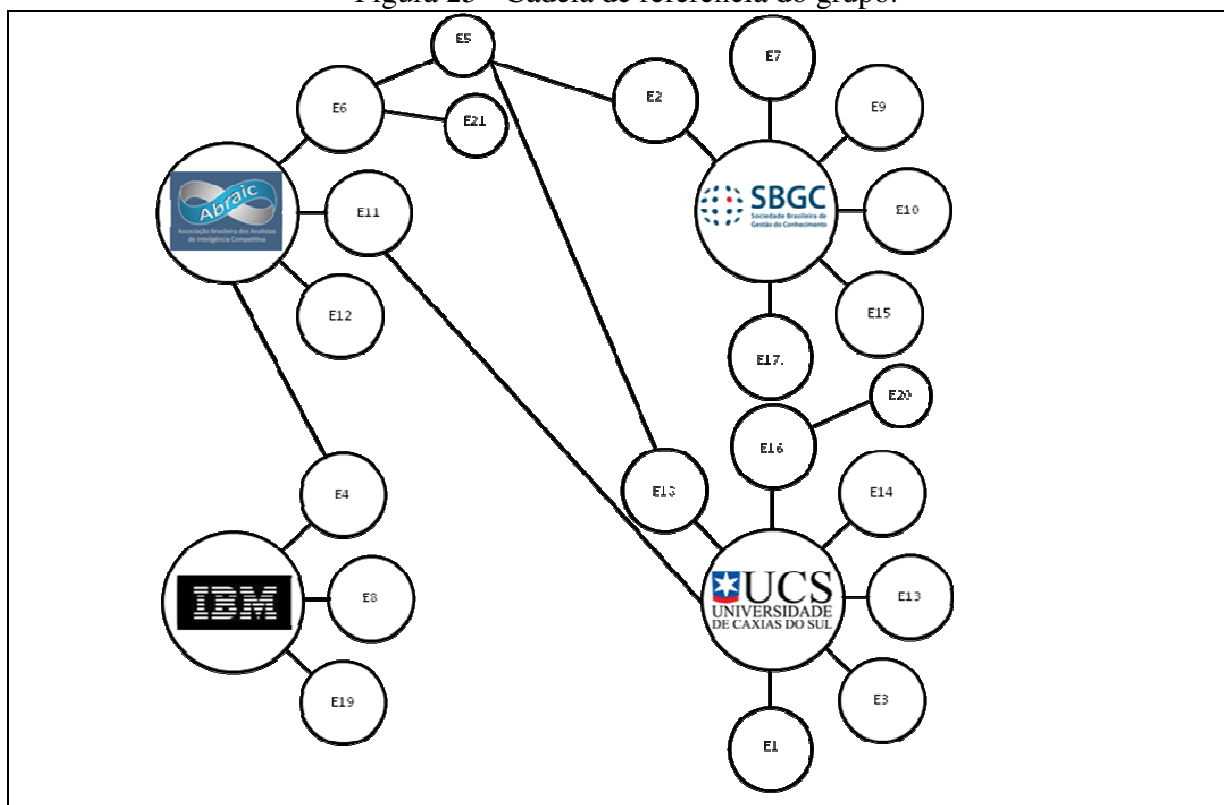
10 (<<http://www.abraic.org.br>>. Acessado em 21 de dezembro de 2013).

11 (<<http://www.sbgc.org.br>>. Acessado em 21 de dezembro de 2013).



estes auxiliam na consolidação da atividade no Brasil.<sup>12</sup> A Figura 25 ilustra a cadeia de referência do grupo selecionado:

Figura 25 - Cadeia de referência do grupo.



Fonte: elaborado pelo próprio autor.

Segundo Glaser e Strauss (1967) tão importante quanto a decisão sobre a amostragem é a decisão sobre quando interromper a integração de casos adicionais. Para isso, os autores sugerem a saturação teórica. A saturação significa que não estão sendo encontrados dados adicionais por meio dos quais o pesquisador possa desenvolver as propriedades da categoria. Assim, por meio dos agentes acima descritos foram identificados 21 respondentes, que geraram mais de 18 horas de diálogo sobre as influências do fenômeno *big data* na análise para inteligência estratégica. O detalhamento dos respondentes encontra-se no apêndice A.

### 3.2 TÉCNICA DE ANÁLISE

De acordo com a técnica de análise de conteúdo utilizada para o tratamento dos dados, foram determinadas algumas categorias *a priori* para este estudo. Para Moraes (1999), a

<sup>12</sup> (<<http://www.ucs.br>>. Acessado em 21 de dezembro de 2013).

categorização é um procedimento de agrupar dados considerando a parte comum existente entre eles. Classifica-se por semelhança ou analogia, segundo critérios previamente estabelecidos ou definidos no processo. As categorias devem ser válidas, exaustivas e homogêneas. Para o autor a validade da categoria se refere aos objetivos da análise, a natureza do material que está sendo analisado e às questões que se pretende responder por meio da pesquisa; a exaustividade possibilita a inclusão de todos os dados significativos definidos de acordo com os objetivos da análise; e, a homogeneidade fundamenta a organização dos dados com base em um único princípio ou critério de classificação.

Para Moraes (1999), categorias definidas *a priori* já devem atender aos critérios de classificação de antemão, isto é antes de proceder à classificação propriamente dita do conteúdo. Nesta pesquisa foram tomados como critérios de qualidade as categorias definidas *a priori*, com base nos pressupostos teóricos sobre os temas foco do estudo. De acordo com Gibbs (2009), a teoria fundamentada é uma abordagem para codificação, cujo foco central está em gerar de forma indutiva ideias teóricas novas ou hipóteses a partir dos dados, em vez de testar teorias especificadas de antemão. Assim, como as categorias advêm dos dados e são sustentadas por eles, são consideradas fundamentadas.

Contudo, a construção da validade e das outras características das categorias de uma análise de conteúdo ocorrem ao longo de todo o processo. Assim, as dimensões identificadas na literatura sobre os temas *big data* e atividade analítica do processo de análise para inteligência estratégica constituem-se em categorias definidas *a priori* que deverão ser confirmadas, eliminadas de acordo com Bardin (1977), ou ainda incrementadas, conforme proposto por Moraes (1999).

Bardin (1977) defende que fundamentar a vinculação das categorias com a teoria é um critério do método de análise de conteúdo. Assim, para inteligência estratégica, foram tomadas como categorias o conjunto de atividades analíticas de um profissional em inteligência definidas por Fachinelli, Rech e Bertolini (2011) a partir dos trabalhos de Heuer (1999) e Bruce e George (2008): 1) domínio da matéria; 2) compreensão de métodos de investigação; 3) imaginação e rigor científico; 4) compreensão dos métodos de coleta; 5) consciência das próprias potencialidades e outras influências cognitivas; 6) mentalidade aberta a opiniões contrárias ou modelos alternativos; e, 7) autoconfiança para admitir e aprender com os erros analíticos. Estas atividades foram confrontadas com os três Vs - volume, variedade e velocidade de *big data* propostas por Betser e Belanger (2013), para

examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica.

A análise e interpretação dos dados foram realizadas seguindo as etapas sugeridas por Moraes (1999): 1) preparação das informações, 2) unitarização ou transformação do conteúdo em unidades, 3) categorização ou classificação das unidades em categorias, 4) descrição e, 5) interpretação. A etapa de interpretação foi facilitada com o uso de *softwares* como ATLAS/ti (GIL, 2010).

### 3.3 COLETA E ORGANIZAÇÃO DOS DADOS

De acordo com Alves-Mazzotti e Gewandsznajder (1999), por sua natureza interativa, a entrevista pode ser a principal técnica de coleta de dados. Os autores colocam que as entrevistas qualitativas são pouco estruturadas, e o entrevistador geralmente está interessado em compreender o significado atribuído pelos sujeitos a eventos, situações, processos ou personagens que fazem parte do cotidiano. Desta forma, neste estudo foram realizadas entrevistas semiestruturadas presenciais ou via teleconferência, com 21 profissionais.

O roteiro que serviu de base para as entrevistas foi estruturado por meio de uma matriz que utilizou as dimensões identificadas na literatura sobre os temas *big data*, por Betser e Belanger (2013) e processo de análise para a atividade de inteligência estratégica, por (Fachinelli, Rech e Bertolini (2011), a partir dos trabalhos de Heuer (1999) e Bruce e George (2008). Este roteiro encontra-se no apêndice B.

No decorrer das primeiras entrevistas, foi possível observar que as respostas coletadas eram bem diferentes umas das outras. Ou seja, alguns entrevistados possuíam completo domínio do tema *big data*, confirmando e acrescentando reflexões de aspectos práticos aliados à base teórica considerada para este estudo. Outros não possuíam conhecimento sobre o tema e foi possível identificar que se tratavam de profissionais que estavam mais focados nas atividades de rotina de inteligência nas organizações. Assim, optou-se por relacionar estas constatações à referência teórica de Schönberger-Mayer e Cukier, (2013), quando estes apresentam que os atores que detém o valor de *big data* são: detentores de dados, especialistas em dados e empresas e pessoas com mentalidade de *big data*.

Assim, as entrevistas foram agrupadas seguindo o critério: se o entrevistado ou a organização de vínculo fossem enquadrados como detentores de dados, especialistas em dados e/ou entrevistado/organização com mentalidade de *big data*, então a entrevista era

alocada no grupo "Especialista". Se não fosse possível enquadrar o entrevistado/organização em algum destes aspectos, então a entrevista era alocada no grupo "Profissional". Desta forma, foi possível comparar as diferentes visões acerca de como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica, objetivo geral deste trabalho.

As entrevistas (E) foram codificadas com uma numeração sequencial de 1 a 21, correspondentemente como códigos resultantes foram gerados de E1 a E21. Na unitarização foram identificadas as "unidades de contexto", e a elas foram atribuídos os grupos: Especialista e Profissional - de acordo com o critério anteriormente descrito. As "unidades de registro" foram definidas de acordo com as categorias conceituais definidas *a priori*. Na categorização, onde os dados são agrupados considerando a parte comum existente, foram totalizadas 62 categorias resultantes, partindo daquelas definidas *a priori*. A Tabela 16 demonstra a organização dos dados:

Tabela 16 - Organização dos dados da entrevista.

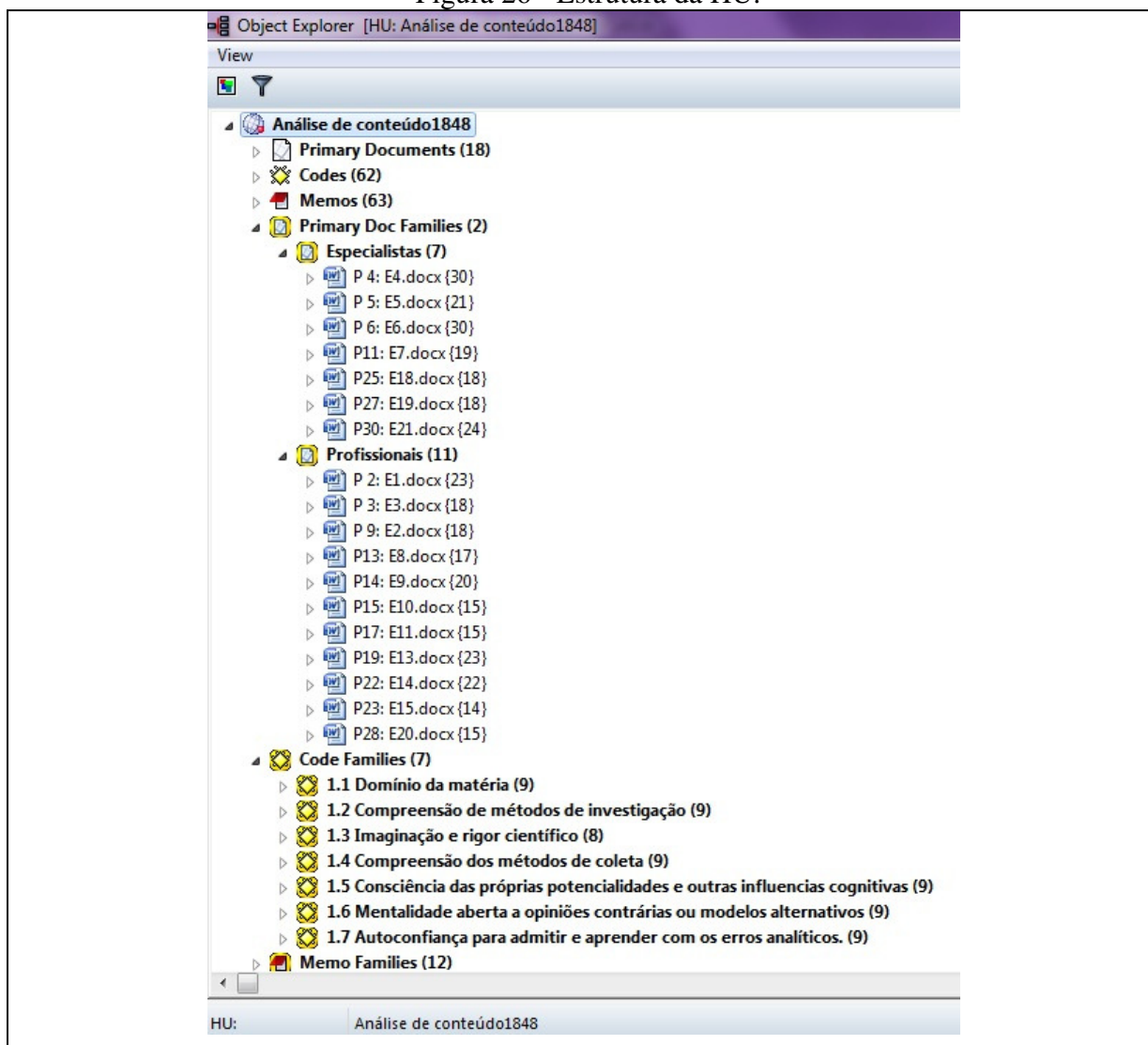
1. Preparação das entrevistas	2. Unitarização (unidades de contexto)	2. Unitarização (unidades de registro: categorias conceituais)	3. Categorização (categorias resultantes - hipóteses)
E1 - E21 <u>Exceto:</u> E8 - não gravada E16 - não gravada E17 - não considerada	Especialista Profissional	1.1 Domínio da matéria 1.2 Compreensão de métodos de investigação 1.3 Imaginação e rigor científico 1.4 Compreensão dos métodos de coleta 1.5 Consciência das próprias potencialidades e outras influencias cognitivas 1.6 Mentalidade aberta a opiniões contrárias ou modelos alternativos 1.7 Autoconfiança para admitir e aprender com os erros analíticos.	1 Volume contribuí 2 Volume prejudica 3 Volume não afeta 4 Velocidade contribuí 5 Velocidade prejudica 6 Velocidade não afeta 7 Variedade contribuí 8 Variedade prejudica 9 Variedade não afeta

Fonte: elaborado pelo próprio autor.

Foi utilizado *software* ATLAS.ti ® versão 7 para análise das entrevistas e organização dos dados. Para isso, as mesmas foram importadas para o *software*, onde simultaneamente foi

possível efetuar a transcrição e posteriormente a categorização conforme Bardin (1977), gerando a criação de um projeto único denominado pelo *software* de Unidade Hermenêutica (em inglês *Hermeneutic Unit* - HU). A Figura 26 demonstra como esta estrutura foi montada internamente no *software*:

Figura 26 - Estrutura da HU.



Fonte: elaborado pelo próprio autor.

O *software* ATLAS.ti ® versão 7, proporcionou algumas facilidades quando da geração da análise. Com exceção da preparação dos dados, que foi feita externamente ao *software*, todas as demais etapas e atividades foram realizadas com o seu auxílio. De acordo com a Figura 26, primeiramente os grupos Especialista e Profissional foram determinados como "famílias". Em um segundo momento, foram inseridas as entrevistas e seguindo a codificação definida, foi realizado o vínculo da entrevista à respectiva família. Na sequência

se deu a análise de cada uma das entrevistas, que foi sendo conduzida com a criação das categorias resultantes, conforme mostra a Figura 27:

Figura 27 - Categorias contextuais e resultantes.

The screenshot shows the Code Family Manager interface. At the top, there is a menu bar with 'Families', 'Edit', 'Miscellaneous', and 'View'. Below the menu is a toolbar with various icons and a search box labeled 'Search (Name)'. The main area is a table with the following columns: Name, Size, Author, Created, and Modified. The table lists seven categories under the family '1.1 Domínio da matéria'. Below the table, there are two panels: 'Codes in family (9):' and 'Codes not in family (53):'. The 'Codes in family' panel lists codes 1.1.1 through 1.1.9 with their respective counts in parentheses. The 'Codes not in family' panel lists codes 1.2.1 through 1.3.1 with their respective counts in parentheses. At the bottom of the window, there is a status bar showing '7 Families' and '1.1 Domínio da matéria'.

Name	Size	Author	Created	Modified
1.1 Domínio da matéria	9	Super	08/01/20...	17/02/20...
1.2 Compreensão de métodos de investigação	9	Super	08/01/20...	16/02/20...
1.3 Imaginação e rigor científico	8	Super	08/01/20...	17/02/20...
1.4 Compreensão dos métodos de coleta	9	Super	08/01/20...	16/02/20...
1.5 Consciência das próprias potencialidades e outras influencias cognitivas	9	Super	08/01/20...	16/02/20...
1.6 Mentalidade aberta a opiniões contrárias ou modelos alternativos	9	Super	08/01/20...	16/02/20...
1.7 Autoconfiança para admitir e aprender com os erros analíticos.	9	Super	08/01/20...	16/02/20...

Codes in family (9):

- 1.1.1 {5-0}
- 1.1.2 {9-0}
- 1.1.3 {2-0}
- 1.1.4 {1-0}
- 1.1.5 {5-0}
- 1.1.6 {3-0}
- 1.1.7 {2-0}
- 1.1.8 {5-0}
- 1.1.9 {2-0}

Codes not in family (53):

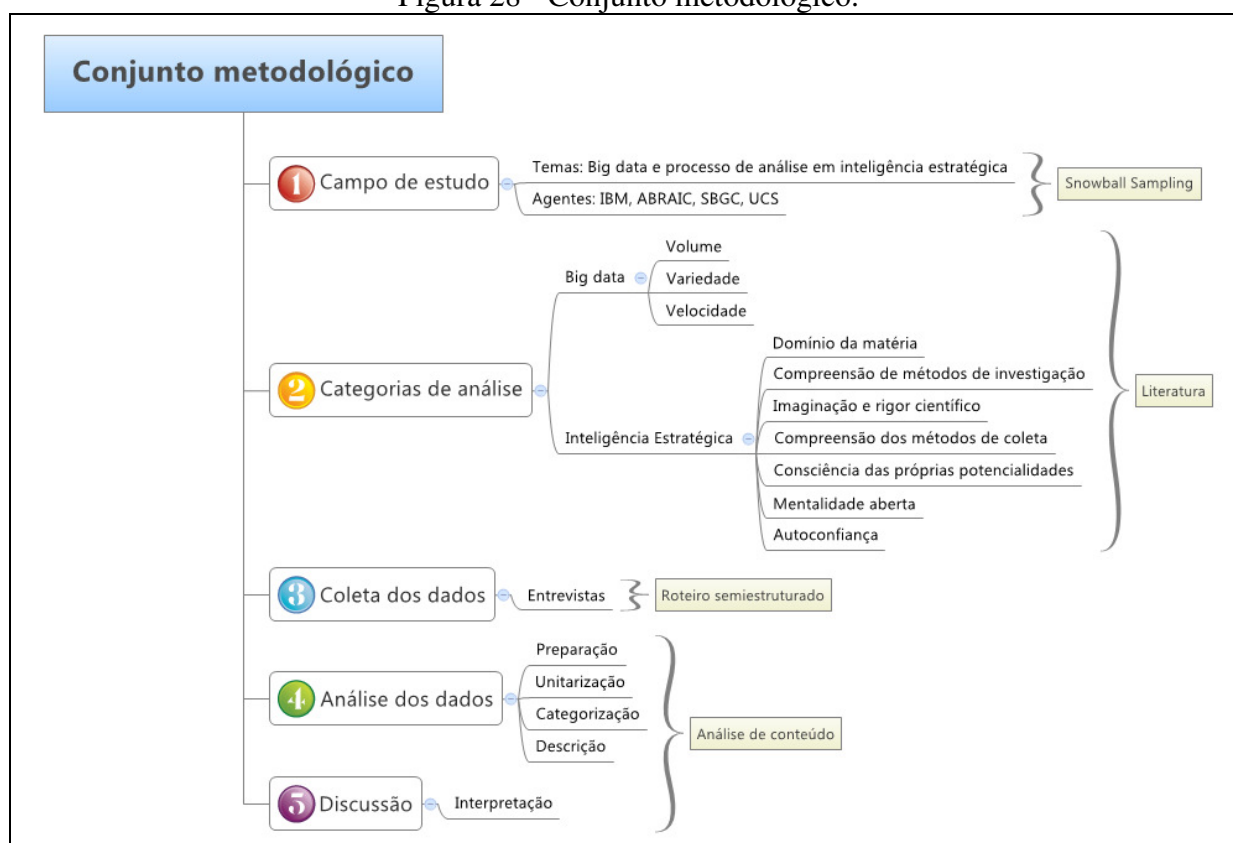
- 1.2.1 {8-0}
- 1.2.2 {9-0}
- 1.2.3 {7-0}
- 1.2.4 {2-0}
- 1.2.5 {1-0}
- 1.2.6 {4-0}
- 1.2.7 {5-0}
- 1.2.8 {5-0}
- 1.2.9 {5-0}
- 1.3.1 {5-0}

7 Families    1.1 Domínio da matéria

Fonte: elaborado pelo próprio autor.

Na Figura 27 pode-se observar que todas as categorias resultantes, onde houveram citações, foram alocadas às categorias contextuais. Assim, o conjunto metodológico desta pesquisa foi formado com base no conteúdo exposto neste capítulo e é apresentado de forma resumida na Figura 28:

Figura 28 - Conjunto metodológico.



FONTE: elaborado pelo próprio autor.

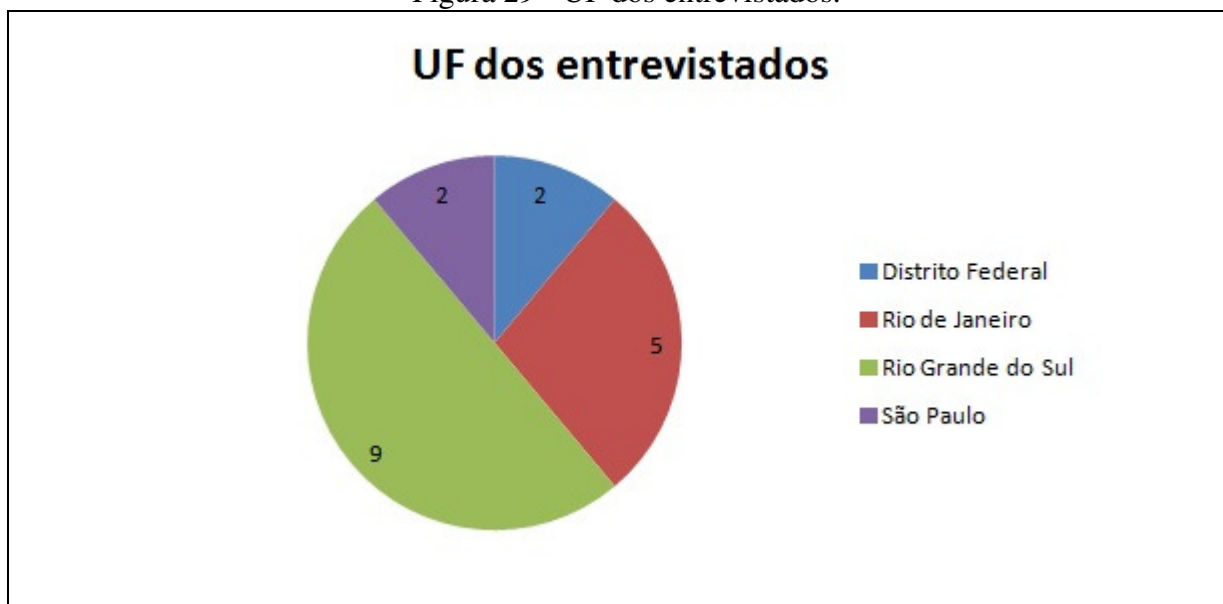
De acordo com a Figura 28, pode-se identificar os 5 passos do desenvolvimento deste conjunto metodológico: 1) campo de estudo: por meio da técnica *snowball sampling* foram identificados 4 agentes com conhecimento e trânsito em comunidades relacionadas aos temas pesquisados, e a partir destes foram identificados 21 entrevistados, 2) categorias de análise: de acordo com a literatura especializada, foram identificadas categorias para o fenômeno *big data* e atividade analítica, 3) coleta dos dados: por meio de entrevistas presenciais e remotas com base em roteiro semiestruturado, 4) análise dos dados: realizada com base na técnica de análise de conteúdo, e 5) discussão: com a interpretação dos resultados gerados.

## 4 RESULTADOS E ANÁLISE

No total, foram realizadas 21 entrevistas, sendo que duas delas não puderam ser gravadas, em função de problemas técnicos. Também uma das entrevistas não foi considerada para os efeitos da análise de conteúdo por não estar adequada aos objetivos da pesquisa. Assim, nas 18 entrevistas consideradas válidas, foi possível identificar diferentes percepções sobre como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica.

Primeiramente, foi realizada uma análise acerca do perfil dos entrevistados. Foi identificada a Unidade Federativa (UF) de residência do entrevistado, o nível de formação, o tempo de experiência na função e o setor de atuação. A Figura 29 apresenta a UF dos entrevistados:

Figura 29 - UF dos entrevistados.



Fonte: elaborado pelo próprio autor.

De acordo com a Figura 29, do total de entrevistas realizadas, a maioria foi composta por entrevistados residentes no Estado do Rio Grande do Sul, com 9 entrevistas. Sendo que as demais dividiram-se entre os Estados do Distrito Federal, Rio de Janeiro e São Paulo. A Figura 30 expõe dados referentes à formação dos entrevistados:



Figura 30- Nível de formação.



Fonte: elaborado pelo próprio autor.

Com relação ao nível de formação, os entrevistados dividiram-se em nível de pós-graduação (mestrado e doutorado) e especialização, com 8 entrevistas ambos. O próximo ponto a ser analisado foi o tempo de experiência do entrevistado na função. A Figura 31 expõe esta distribuição:

Figura 31 - Tempo de experiência.



Fonte: elaborado pelo próprio autor.

Para a realização deste estudo, a *expertise* dos entrevistados nos temas abordados foi considerada fator crítico. Assim a maioria dos entrevistados possui mais de 5 anos de atuação na academia e, em organizações, totalizando 14 entrevistados. Destes, metade possui mais de 10 anos de experiência. A Figura 32 apresenta os setores de atuação dos entrevistados:

Figura 32 - Setores de atuação.



Fonte: elaborado pelo próprio autor.

Nesse aspecto, pode-se observar que a maior parte dos entrevistados está vinculada ao setor de serviços, com 9 entrevistas. Estes entrevistados tem atuação em consultoria/pesquisa, instituições de ensino profissionalizante, e outros. As demais áreas ficaram divididas entre indústria, tecnologia e varejo.

Nas entrevistas, pode-se perceber que todas as categorias resultantes exceto uma foram identificadas, ou seja, houveram percepções pulverizadas acerca dos efeitos do *big data* no processo de análise. Esta é uma representação de como o conceito de *big data*, bem como seus efeitos ainda não são claros, ao menos entre os entrevistados deste estudo. A Figura 33 apresenta o *software* e a disposição do conteúdo de uma forma que facilita o processo de análise e interpretação:

Figura 33 - Disposição dos dados.

Name	Grounded	Families
1.6.9	12	1.6 Mentalidade aberta a opiniões contrárias ou modelos alternativos
1.6.6	12	1.6 Mentalidade aberta a opiniões contrárias ou modelos alternativos
1.6.3	11	1.6 Mentalidade aberta a opiniões contrárias ou modelos alternativos
1.7.3	9	1.7 Autoconfiança para admitir e aprender com os erros analíticos.
1.4.1	9	1.4 Compreensão dos métodos de coleta
1.1.2	9	1.1 Domínio da matéria
1.7.6	9	1.7 Autoconfiança para admitir e aprender com os erros analíticos.
1.2.2	9	1.2 Compreensão de métodos de investigação
1.4.8	8	1.4 Compreensão dos métodos de coleta
1.2.1	8	1.2 Compreensão de métodos de investigação
1.7.9	8	1.7 Autoconfiança para admitir e aprender com os erros analíticos.
1.5.2	7	1.5 Consciência das próprias potencialidades e outras influencias cognitivas
1.5.8	7	1.5 Consciência das próprias potencialidades e outras influencias cognitivas
1.4.7	7	1.4 Compreensão dos métodos de coleta
1.3.3	7	1.3 Imaginação e rigor científico
1.2.3	7	1.2 Compreensão de métodos de investigação
1.5.3	6	1.5 Consciência das próprias potencialidades e outras influencias cognitivas
1.4.3	6	1.4 Compreensão dos métodos de coleta
1.5.1	6	1.5 Consciência das próprias potencialidades e outras influencias cognitivas
1.3.6	6	1.3 Imaginação e rigor científico
1.3.1	5	1.3 Imaginação e rigor científico
1.2.9	5	1.2 Compreensão de métodos de investigação
1.1.5	5	1.1 Domínio da matéria
1.7.1	5	1.7 Autoconfiança para admitir e aprender com os erros analíticos.
1.1.8	5	1.1 Domínio da matéria
1.2.8	5	1.2 Compreensão de métodos de investigação
1.5.6	5	1.5 Consciência das próprias potencialidades e outras influencias cognitivas
1.1.1	5	1.1 Domínio da matéria
1.2.7	5	1.2 Compreensão de métodos de investigação
1.5.9	5	1.5 Consciência das próprias potencialidades e outras influencias cognitivas

Fonte: elaborado pelo próprio autor.

Como pode-se observar na Figura 33, as categorias resultantes com maior frequência de citações percebem o volume, a velocidade e a variedade, como não exercendo efeitos na dimensão "mentalidade aberta a opiniões contrárias e modelos alternativos" (1.6.3/6/9), também o volume foi colocado como não exercendo efeitos na dimensão "autoconfiança para admitir e aprender com erros analíticos" (1.7.1). Seguindo esta linha, o primeiro ponto onde é levantada a contribuição do *big data*, em termos de volume, é na dimensão "volume contribui para a compreensão dos métodos de coleta" (1.4.1). O efeito prejudicial do volume de *big data* é percebido primeiramente na dimensão "volume prejudica o domínio da matéria" (1.1.2).

Além disso, a ferramenta dispunha do recurso dos memorandos. Estes são objetos independentes que podem estar vinculados com outros objetos ou podem ser utilizados livremente, como memorandos independentes. A Figura 34 mostra a estrutura dos memorandos:

Figura 34 - Estrutura dos memorandos.

Name	Size	Author	Created	Modified
1.1 Domínio da matéria	1	Super	09/01/20...	16/02/20...
1.2 Compreensão de métodos de investigação	3	Super	09/01/20...	15/02/20...
1.3 Imaginação e rigor científico	4	Super	09/01/20...	15/02/20...
1.4 Compreensão dos métodos de coleta	6	Super	09/01/20...	15/02/20...
1.5 Consciência das próprias potencialidades e outras influencias cognitivas	1	Super	09/01/20...	15/02/20...
1.6 Mentalidade aberta a opiniões contrárias ou modelos alternativos	0	Super	09/01/20...	15/02/20...
1.7 Autoconfiança para admitir e aprender com os erros analíticos	1	Super	09/01/20...	15/02/20...
1.8 Outras capacidades	10	Super	09/01/20...	16/02/20...
1.9 Outras dimensões de BD	7	Super	15/02/20...	15/02/20...
Autores e artigos	1	Super	15/02/20...	15/02/20...
Comentários que remetem a teoria	27	Super	15/02/20...	15/02/20...
Processo de análise	1	Super	15/02/20...	15/02/20...

Memos in family (1):

Análises com base em intuição (3-Co) - Super

Memos not in family (62):

Agilidade (2-Co) - Super  
 Ainda não existem estudos com respaldo acadêmico sobre big data (1-Co) - Super  
 Algoritmo (3-Co) - Super  
 Alto volume de demandas compromete o desempenho do analista (5-Co) - Super  
 Alto volume de demandas leva a análises mais superficiais (3-Co) - Super  
 Associação de BD com inteligência (4-Co) - Super  
 Ausência de um conceito para BD (1-Co) - Super  
 Ausência de um novo modelo cognitivo (2-Co) - Super  
 Autores e artigos (13-Co) - Super  
 BD é para grandes empresas (1-Co) - Super

12 Families | 1.1 Domínio da matéria

Fonte: elaborado pelo próprio autor.

Este recurso foi utilizado para auxiliar na marcação de importantes trechos das entrevistas, que faziam referência aos objetivos específicos do estudo. Também foram criadas famílias de memorandos, seguindo as categorias conceituais e outras famílias para marcar: 1.8 Outros aspectos em IE; 1.9 Outras dimensões de *big data*.

Outros recursos do *software*, como as observações, também foram colocados em todos os níveis da HU, com a finalidade de facilitar o processo de entendimento acerca das marcações realizadas, para posterior interpretação. Como funcionalidades do *software* para explorar e consultar a informação foram utilizadas a recuperação simples, para uma única categoria resultante, ou recuperação complexa com a ferramenta "*Query Tool*", para códigos múltiplos. A funcionalidade "*Network View*" foi utilizada para a visualização das relações entre as categorias e famílias. A próxima seção irá apresentar o conteúdo gerado e interpretado com o auxílio da ferramenta.

#### 4.1 PROCESSO DE ANÁLISE DO ESTUDO

Este capítulo irá apresentar, por meio dos dados coletados nas entrevistas, como foi possível abordar os objetivos específicos deste estudo. Primeiramente serão apresentadas as

principais etapas das práticas de análise dos profissionais de IE, objetivo específico "a": identificar as principais etapas das práticas de análise dos profissionais de IE. Posteriormente serão exibidas as percepções da influência do *big data* na atividade de IE, objetivo específico "b": verificar a percepção da influência do *big data* na atividade de IE. Após serão reveladas as dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data*, objetivo específico "c": identificar as dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data*. A interface entre *big data* e atividade analítica será tratada na seção seguinte, referente ao objetivo específico "d": analisar as interfaces entre *big data* e atividade analítica no processo de interpretação de informações para a atividade de IE. Por fim, serão apresentados os efeitos do fenômeno *big data* no processo de análise na atividade de inteligência estratégica, objetivo geral deste estudo.

#### 4.1.1 Principais etapas das práticas de análise dos Profissionais de IE

Na realização das entrevistas, primeiramente os entrevistados foram convidados a expor as principais etapas das práticas de análise, como primeiro objetivo específico colocado para este estudo. Na análise das respostas foram identificados diferentes focos de atuação, métodos e ferramentas utilizadas com relação aos temas *big data* e inteligência estratégica. Isso proporcionou a criação de dois grupos: Especialistas e Profissionais. A Figura 35 demonstra esta distribuição:

Figura 35 - Grupos de entrevistas.



Fonte: elaborado pelo próprio autor.

O grupo Especialistas é composto pelos respondentes identificados como detentores de dados e especialistas em dados com melhor percepção de *big data*. São responsáveis por desenvolver pesquisas sobre assuntos relacionados à gestão das informações e inteligência estratégica e dar o acompanhamento às empresas nas áreas relacionadas. Estão vinculados a instituições de ensino e pesquisa e empresas de tecnologia ou consultoria. Atuam na elaboração de livros e artigos sobre os temas e ministram palestras.

Das 6 entrevistas realizadas, especificamente em uma o conteúdo diferenciou-se das demais. Trata-se da utilização de técnicas de análise de dados aderentes ao fenômeno *big data*, em um laboratório de pesquisa de empresa de tecnologia, no Brasil. Estas técnicas consistem em métricas de caráter preditivo com a utilização de estatística, baseada em grandes volumes de dados. Este processamento pode requerer inclusive, a utilização de tecnologia não existente no Brasil. Neste caso, o poder de *big data* apontado pelo entrevistado foi traduzido em termos da melhora na qualidade das pesquisas realizadas. Assim, quando são realizadas as análises não se parte de uma hipótese pré-determinada. A prática consiste em deixar emergir dos dados algum padrão que seja relevante ao tema que está sendo tratado (SCHÖNBERGER-MAYER E CUKIER, 2013).

O grupo Profissionais foi formado por 12 entrevistas. Os respondentes foram identificados como não sendo detentores de dados e especialistas em dados com melhor percepção de *big data*. Neste grupo estão os Profissionais em IE, pessoas que exercem suas atividades nas áreas de inteligência estratégica das organizações. Possuem focos de trabalho definidos e geralmente estão ligados à áreas de *marketing*, estratégia e inovação. Possuem menos tempo de contato com IE e *big data*, em comparação ao grupo Especialistas.

Neste modelo as etapas do processo de análise estão vinculadas a clara definição de metas, hipóteses definidas *a priori* e comumente, com o apoio de consultorias em inteligência estratégica. Os elementos marcantes para a realidade do grupo Profissionais é um cenário de forte pressão por parte do mercado e das empresas na entrega dos produtos de inteligência. Assim, o grupo Profissionais recebe um grande volume de demandas e relatara ter pouco tempo para análise, e pouca ou quase nula aplicação de grandes volumes de dados na geração de valor. Ao contrário, o que ocorre é a análise de pequenos volumes de dados, com fontes limitadas ao máximo, no intuito de mitigar o risco de erros, fontes dúbias e o desvio do foco da análise. Para este perfil a percepção da agregação de valor de *big data* não é um fator presente no dia-a-dia da inteligência.

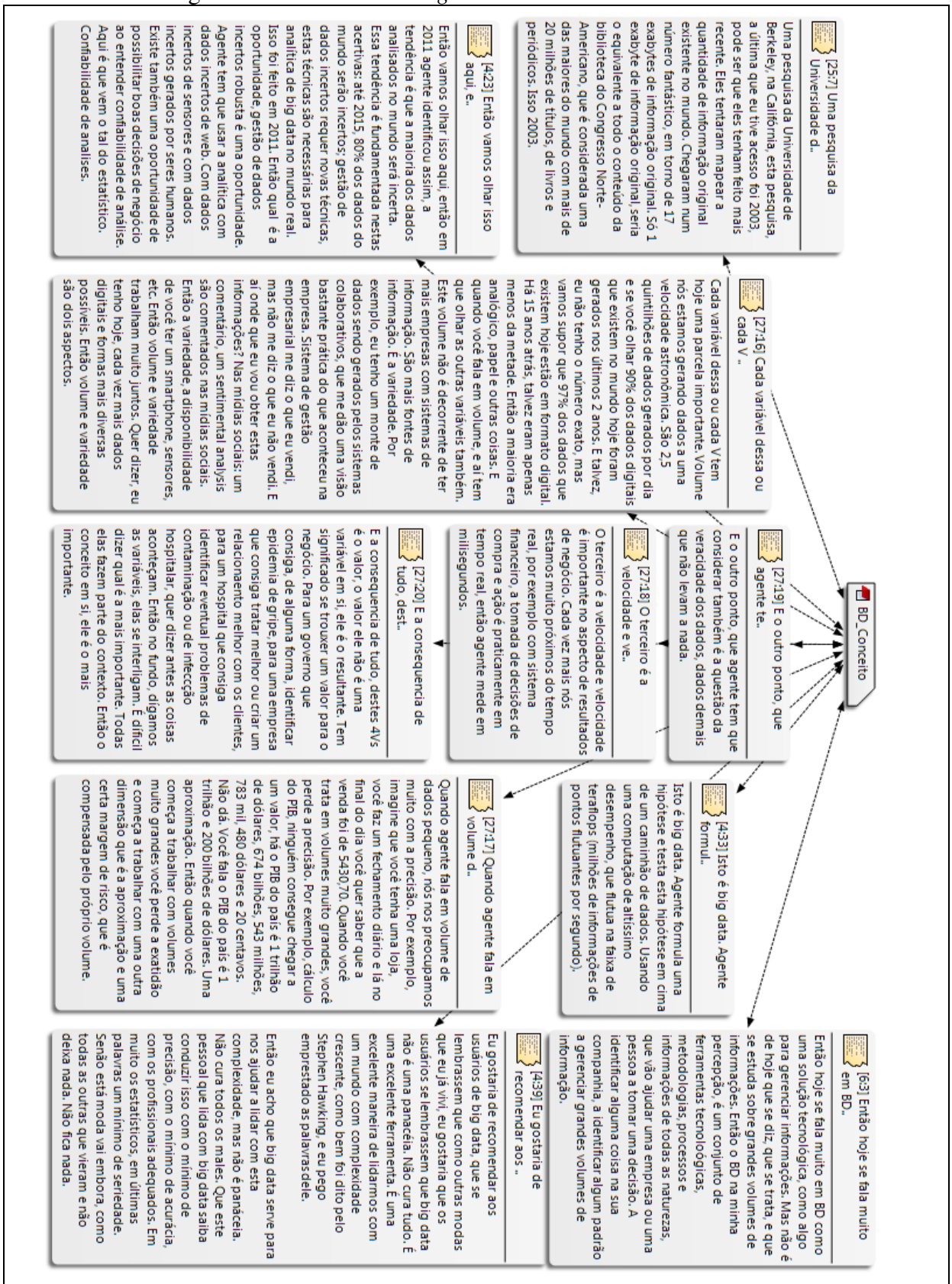
Uma prática utilizada pelos respondentes do grupo Profissionais é a troca de informações entre grupos ou equipes dentro da organização, com a formação de times multifuncionais. A função destas equipes vai desde o levantamento de dados, análise conjunta e validação das informações resultantes. Neste sentido, é grande a dependência por pessoas com experiência e domínio da matéria para a criação dos produtos de inteligência. Esse é um fator apontado por Fuld (1995), na literatura pesquisada em termos de IE.

Fuld (1995), coloca que análise é realmente a aplicação de bom senso e experiência para informação crua. Heuer (1999) expõe a questão da influencia de experiências do passado, educação, valores culturais, requisitos, e normas organizacionais e as especificidades da informação recebida, como fatores de influência no processo de análise. Entretanto, esta mentalidade demonstra certa fragilidade dos Profissionais frente a atual complexidade que se apresenta em meio do fenômeno *big data*.

Este grupo também traz em si, a análise de dados estruturados em processos direcionados por métodos definidos para a criação da inteligência. O monitoramento de dados externos à organização é realizado, porém de forma controlada e com uma finalidade específica. É baseado na utilização de ferramentas tradicionais na busca de informação, como Google Alertas, por exemplo. Em alguns casos o trabalho é feito manualmente pelo profissional, sem auxílio de ferramentas tecnológicas.

#### **4.1.2 Percepções da influência do *big data* na atividade de IE**

As percepções coletadas tanto do grupo Especialistas, quanto grupo Profissionais com relação a influência do *big data* na atividade de IE foram organizadas e confrontadas em relação à base teórica desenvolvida para este estudo. No resumo exposto na Figura 36 é possível observar as percepções dos entrevistados em relação ao conceito de *big data*:

Figura 36 - Influência do *big data* na atividade de IE: conceito.

Fonte: elaborado pelo próprio autor.



Nas citações expostas na Figura 36, pode-se verificar a percepção fundamentada do aumento do volume informacional, com ênfase em dados incertos, não estruturados. Outro ponto interessante mencionado é o fato de que com *big data* é diminuída a preocupação com a exatidão em detrimento da aproximação. A seguir são expostos trechos que embasam estas constatações:

*E18 (falando do aumento do volume informacional): Chegaram num número fantástico, em torno de 17 exabytes de informação original. Só um exabyte de informação original, seria o equivalente a todo o conteúdo da biblioteca do Congresso Norte-Americano.*

*E19 (falando da exatidão): Quando a gente fala em volume de dados pequeno, nós nos preocupamos muito com a precisão. Então quando você começa a trabalhar com volumes muito grandes, você perde a exatidão e começa a trabalhar com uma outra dimensão que é a aproximação.*

É apontada também a necessidade do desenvolvimento de técnicas apropriadas para a sua manipulação, e o papel da estatística no sentido de promover um melhor nível de confiabilidade das análises. Na sequência são apresentados trechos de entrevistas que apoiam esta constatação:

*E4 (falando de técnicas apropriadas): Gestão de dados incertos requer novas técnicas, estas técnicas são necessárias para analítica de big data no mundo real.*

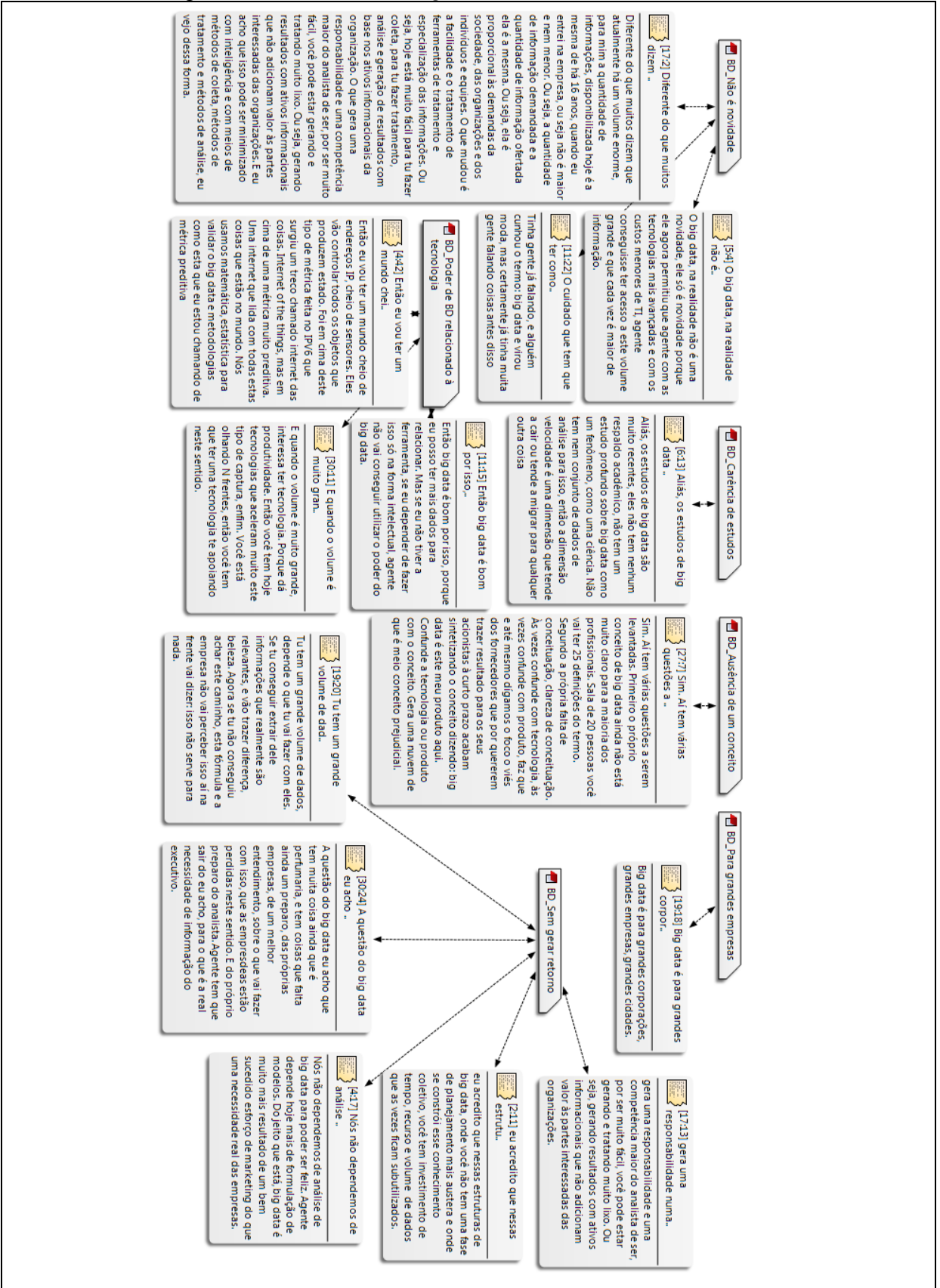
*E4 (falando da estatística): Que este pessoal que lida com big data saiba conduzir isso com o mínimo de precisão, com os profissionais adequados. Em muito os estatísticos.*

Foram descritos os 3 Vs de *big data*, complementados pela veracidade e geração de valor. A confusão de dados, trazida na base teórica por Schönberger-Mayer e Cukier, (2013), também foi mencionada, juntamente com a consciência da aplicação de *big data* em meio a um mundo de complexidade. A seguir é exposto trecho que fala da questão do valor de *big data*:

*E19 (falando das características de big data): E a consequência destes 4Vs é o valor, ele não é uma variável em si, ele é o resultante. Tem significado se trazer valor ao negócio.*

Estas reflexões tem sua origem nas entrevistas do grupo Especialistas. A Figura 37 apresenta outras percepções dos entrevistados em relação à aspectos diversos da influência do *big data* na atividade de IE:

Figura 37 - Influência do big data na atividade de IE: diversos.



Fonte: elaborado pelo próprio autor.

Neste resumo da Figura 37, pode-se observar uma citação que expõe a relação de oferta e demanda do volume de informações como sendo a mesma do passado. Nesta visão, o que se apresenta atualmente, é uma maior facilidade no tratamento das informações, com a indispensável utilização de ferramentas tecnológicas. A carência de estudos em *big data*, e a ausência de um conceito do fenômeno que seja claro para os Profissionais também foram pontos levantados e confirmados na base teórica do estudo. A seguir é dada ênfase a dois trechos que relatam esta situação:

*E06 (falando da carência de estudos em big data): Aliás, os estudos de big data são muito recentes, eles não tem nenhum respaldo acadêmico, não tem um estudo aprofundado sobre big data como um fenômeno, como uma ciência. Não tem nem conjunto de dados de análise para isso.*

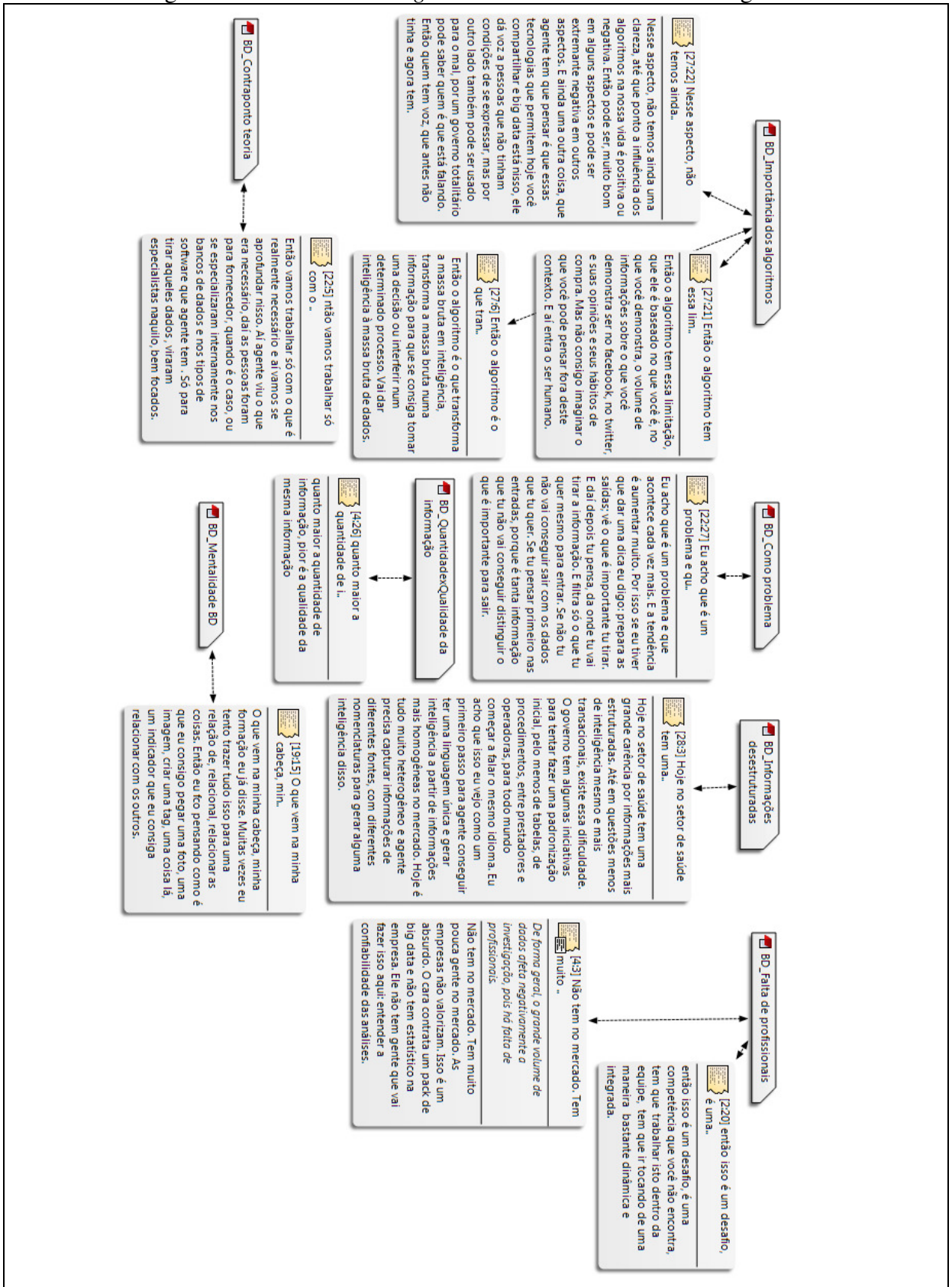
*E19 (falando da ausência de um conceito): Primeiro, o próprio conceito de big data ainda não está muito claro para a maioria dos profissionais. Segundo, a própria falta de conceituação, clareza de conceituação, às vezes confunde com tecnologia, às vezes confunde com produto.*

Outro ponto mencionado pelos entrevistados, refere-se ao fato de *big data* estar direcionado apenas à grandes corporações. Além disso, a percepção da não geração de valor de *big data*, também foi citada. Sobre este aspecto, as citações relatam que não basta que se tenham grandes volumes de dados, mas sim que haja um propósito para eles. Com isto surge a questão da formulação de métodos que possibilitem a utilização dos dados para a real necessidade de informação do executivo. Na sequência segue trecho que trata desta questão:

*E11 (falando da não geração de valor): Por ser muito fácil, você pode estar gerando e tratando muito lixo. Ou seja, gerando resultados com ativos informacionais que não adicionam valor às partes interessadas das organizações.*

Outros apontamentos feitos pelos entrevistados remetem ao viés negativo de *big data*. Foram mencionados aspectos referentes à falta de clareza sobre os atuais impactos da utilização da tecnologia e a limitação dos algoritmos neste processo. A Figura 38 resume as de entrevistas:

Figura 38 - Influência do big data na atividade de IE: viés negativo.



Fonte: elaborado pelo próprio autor.

De acordo as citações expostas na Figura 38, ainda pode-se perceber uma preocupação com a limitação das fontes consultadas, e uma super especialização de alguns entrevistados do grupo Profissionais em temas específicos. A seguir é apresentado um trecho de entrevista que trata desta questão:

*E14 (falando da especialização): Vamos trabalhar só com o que é realmente necessário e aí vamos nos aprofundar nisso. As pessoas foram para fornecedor, ou se especializaram internamente nos bancos de dados e nos tipos de software que a gente tem. Só para tirar aqueles dados, viraram especialistas naquilo, bem focados.*

Estes são pontos que na base teórica utilizada, foram colocados como não apropriados, pois em tempos de *big data* a ação deveria estar direcionada em fazer com que novos padrões emergissem dos grandes volumes de dados e que profissionais de perfil multidisciplinar, conduzissem este processo. Também é apontada a falta de profissionais com perfil específico para o trato com grandes volumes de dados, de acordo com trecho a seguir:

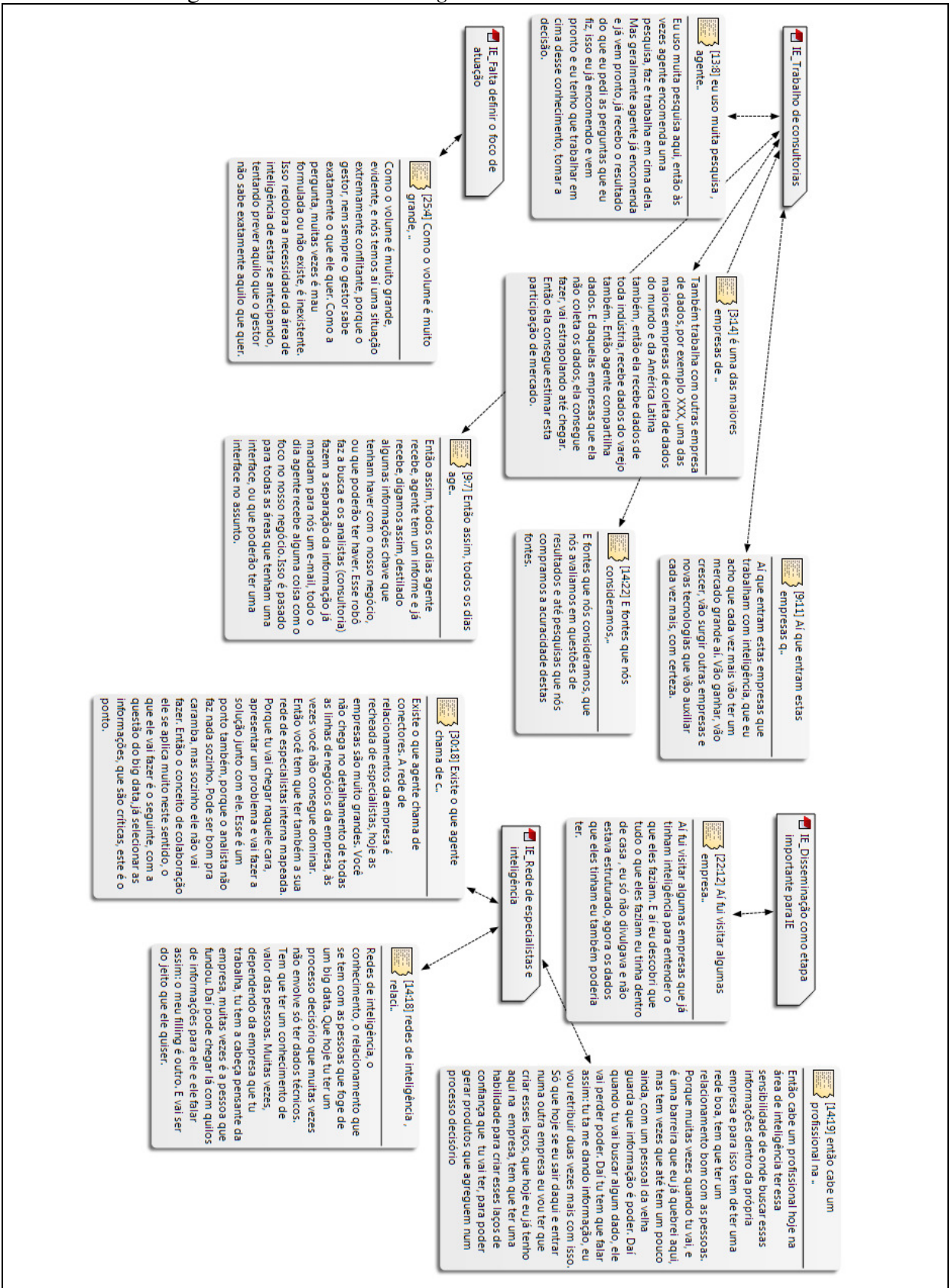
*E4 (falando da falta de profissionais): Não tem no mercado. Tem pouca gente no mercado. Isso é um absurdo. O cara contrata um pack de big data e não tem estatístico na empresa. Ele não tem gente que vai fazer isso aqui: entender a confiabilidade das análises.*

Outros pontos abordados nas entrevistas tratam da questão do *big data*, como um problema para os profissionais nas organizações, principalmente em função do grande volume de informações. A seguir são apresentados dois trechos referentes a esta percepção:

*E14 (falando do big data como problema): Eu acho que é um problema e que acontece cada vez mais. E a tendência é aumentar muito. Por isso eu digo: prepara as saídas. Se não tu não vai conseguir sair com os dados que tu quer. Se tu pensar primeiro nas entradas, é tanta informação que tu não vai conseguir distinguir o que é importante para sair.*

A Figura 39 relata mais algumas percepções diversas acerca do fenômeno, no que tange ao trabalho de consultorias, foco de atuação, processo de disseminação na IE e rede de especialistas:

Figura 39 - Influência do big data na atividade de IE: diversos.



Fonte: elaborado pelo próprio autor.

Nas citações da Figura 39, chama atenção a compra de análises prontas de empresas de consultoria, por profissionais em IE. Nesta visão, a prestação de serviços em inteligência tende a crescer no mercado. As consultorias são atores que detêm o valor de *big data*, são empresas com o conhecimento ou a tecnologia para fazer análises complexas, segundo Schönberger-Mayer V.; Cukier K, (2013). A seguir é exposto trecho que fundamenta esta questão:

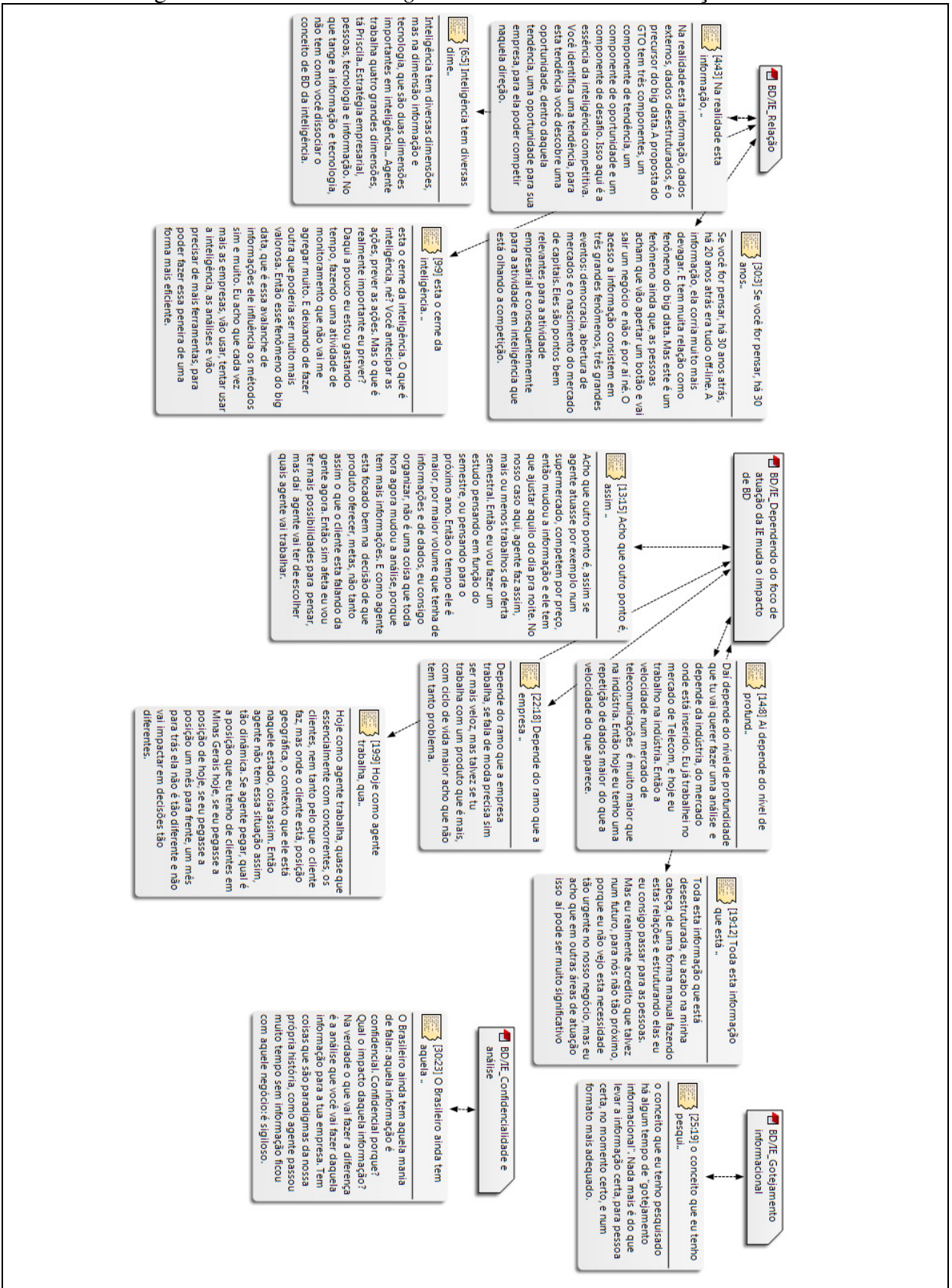
*E8(falando do trabalho de consultorias): Eu uso muita pesquisa aqui, então às vezes, a gente encomenda uma pesquisa, faz e trabalha em cima dela. Mas geralmente a gente já encomenda e já vem pronto, já recebo o resultado do que eu pedi.*

Outro apontamento refere-se a utilização das redes de especialistas. Este ponto é colocado como importante pelo grupo Profissionais em função da falta de domínio das linhas de negócio da empresa, fato que demanda a colaboração dos membros da rede na utilização dos dados internos da organização. A seguir é apresentado trecho que confirma esta constatação:

*E14(falando das redes): Então cabe ao profissional hoje na área de inteligência, ter essa sensibilidade de onde buscar essas informações dentro da própria empresa e para isso tem de ter uma rede boa, tem que ter um relacionamento bom com as pessoas.*

A consciência da utilização do *big data* traz evidências da influência do fenômeno em IE. As questões relativas ao método e às ferramentas tecnológicas que devem ser utilizadas para trabalhar com *big data* é o ponto de mutação para IE, com o surgimento do fenômeno. A influência de algumas dimensões, como a velocidade, principalmente, tem diferentes efeitos no processo de análise, em função do setor de atuação da indústria. Ou seja, varejo e indústria possuem ritmos diferentes de negócios, em função do mercado. Isso reflete na geração de informações e na sua velocidade de atualização. A Figura 40 traz citações que falam sobre estas questões:

Figura 40 - Influência do big data na atividade de IE: relação com IE.



Fonte: elaborado pelo próprio autor.



Na Figura 40, outro ponto de reflexão dos entrevistados é a questão da utilização de *big data* no contexto de IE. Neste sentido, não é possível dissociar o conceito de *big data* da inteligência em se tratando das dimensões tecnologia e informação. No trecho a seguir é considerada a utilização de tendências, oportunidades e desafios para o direcionamento da organização. Isto é apontado como a essência a inteligência competitiva, que toma como base dados externos e não estruturados, ou seja, *big data* para a sua concepção:

*E4(falando da relação big data e IE): Dados externos, dados desestruturados é o precursor de big data. Isso aqui é a essência da inteligência competitiva. Você identifica uma tendência, você descobre uma oportunidade dentro daquela tendência, uma oportunidade para sua empresa, para ela poder competir naquela direção.*

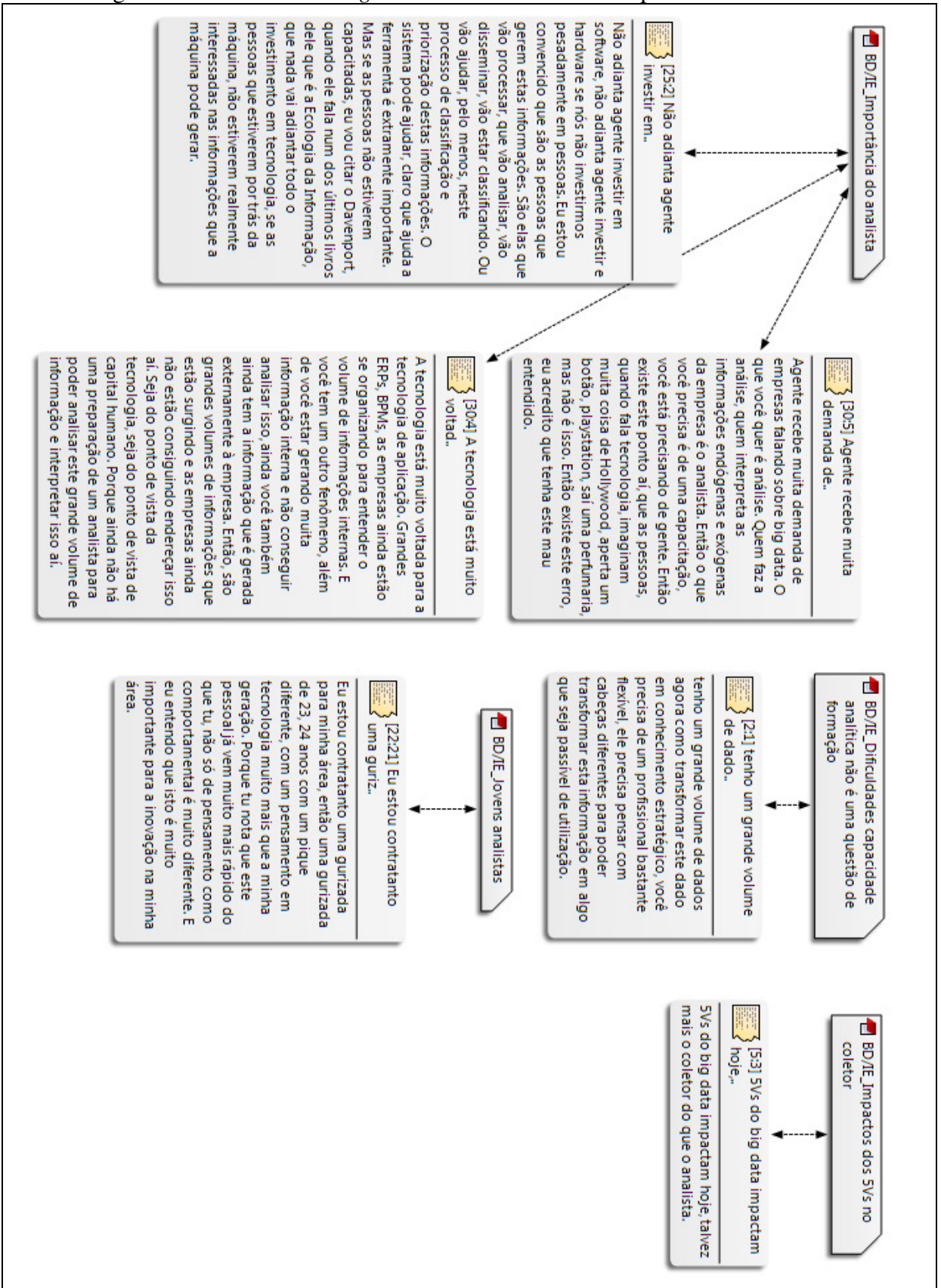
Outra observação advinda das citações foi a utilização mais ou menos intensa de *big data*, dependendo do foco de atuação da organização. Nestes casos foi mencionado que em segmentos como telecomunicações e varejo a velocidade na geração e atualização de informação é mais intensa. Em contrapartida, segmentos relacionados à indústria teriam um ritmo mais lento para a geração e atualização de informações. A seguir é apresentado trecho que descreve esta diferenciação:

*E8(falando do foco de atuação): Um supermercado compete por preço, então mudou a informação, ele tem que ajustar aquilo do dia para noite. No nosso caso aqui a gente faz assim, mais ou menos trabalhamos com oferta semestral. Por maior volume que tenha de informações e de dados, eu consigo organizar.*

A fragilidade da confidencialidade da informação também é mencionada. Este é um paradigma para IE, mas em tempos de *big data*, passa a perder um pouco o sentido que vinha tendo em outras épocas, onde a informação era protegida e dita como sigilosa. Atualmente e talvez mais do que antes, o valor está na descoberta de um sentido para estas informações, por meio do processo de análise e não apenas no seu acesso. Com *big data* este processo irá ocorrer por meio do profissional em IE, que irá utilizar dados não estruturados, em grandes volumes, para a descoberta de padrões nunca antes pensados. Neste cenário, a tecnologia também terá papel fundamental, no sentido de facilitar o trabalho do profissional.

Entretanto, as ferramentas tecnológicas por si só, não farão com que *big data* possa proporcionar valor às organizações, alguns entrevistados apontaram para a questão humana no processo de análise. A Figura 41 traz evidências destas constatações:

Figura 41 - Influência do *big data* na atividade de IE: importância do analista.



Fonte: elaborado pelo próprio autor.

Os profissionais que estão envolvidos neste processo possuem diversas denominações - analistas, cientistas de dados e outros, dependendo da área de atuação e setor. Independente disso, a importância deste tipo de profissional em análise é justificada pelos entrevistados, pelo fato que a gestão das informações, a classificação, a priorização e a análise, pode ser facilitada com a ajuda de sistemas, mas se as pessoas não estiverem interessadas ou capacitadas para os resultados gerados isto não será revertido em valor para as organizações. A seguir é apresentado o trecho que relata esta questão:

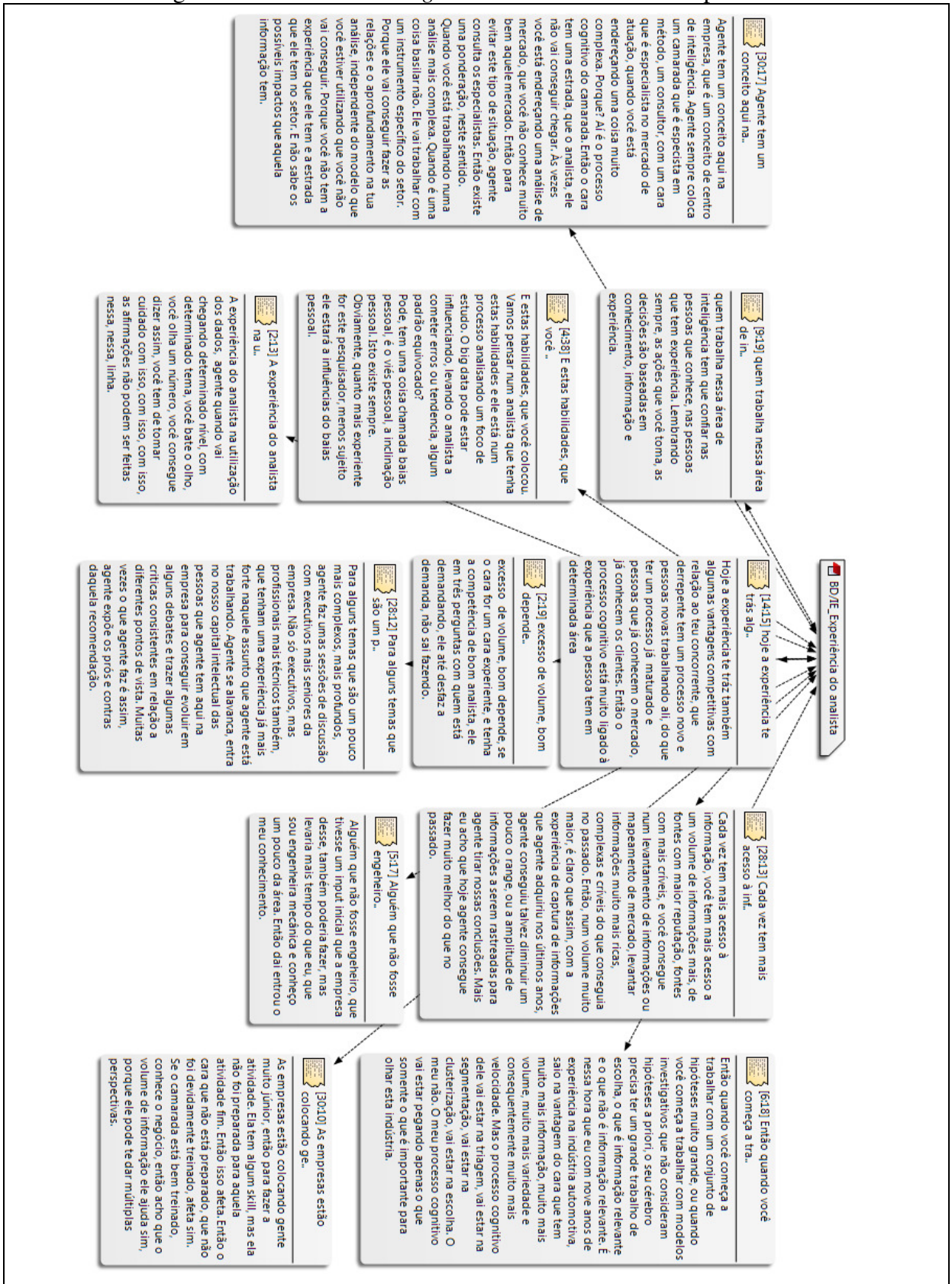
*E18(falando da importância do profissional): São as pessoas que gerem estas informações. São elas que vão processar, que vão analisar, vão disseminar, vão estar classificando. Nada vai adiantar todo o investimento em tecnologia, se as pessoas que estiverem por trás da máquina, não estiverem realmente interessadas nas informações.*

As citações também mencionam que as organizações já encontram um grande desafio no entendimento do grande volume de informações internas. Então o advento do *big data* potencializa o *gap* da deficiência em análise, do ponto de vista tecnológico e humano. Neste trecho é dada ênfase na falta de preparação do analista para analisar e interpretar a informação:

*E21(falando da deficiência em análise): As empresas ainda estão se organizando para entender o volume de informações internas. Então, são grandes volumes de informações que estão surgindo e as empresas ainda não estão conseguindo endereçar isto aí. Porque ainda não há uma preparação de um analista para poder analisar este grande volume de informação e interpretar.*

Com o advento do *big data*, as suas funções e habilidades do profissional em análise estão se modificando. Um elemento que ressalta neste contexto é a questão do domínio das ferramentas tecnológicas, que implicam diretamente nos métodos utilizados para coletar, investigar, organizar e analisar os dados e na habilidade de profissionais mais jovens em tratar com estas questões. Em contraponto, a Figura 42 traz citações que falam da questão da experiência do analista:

Figura 42 - Influência do *big data* na atividade de IE: experiência.



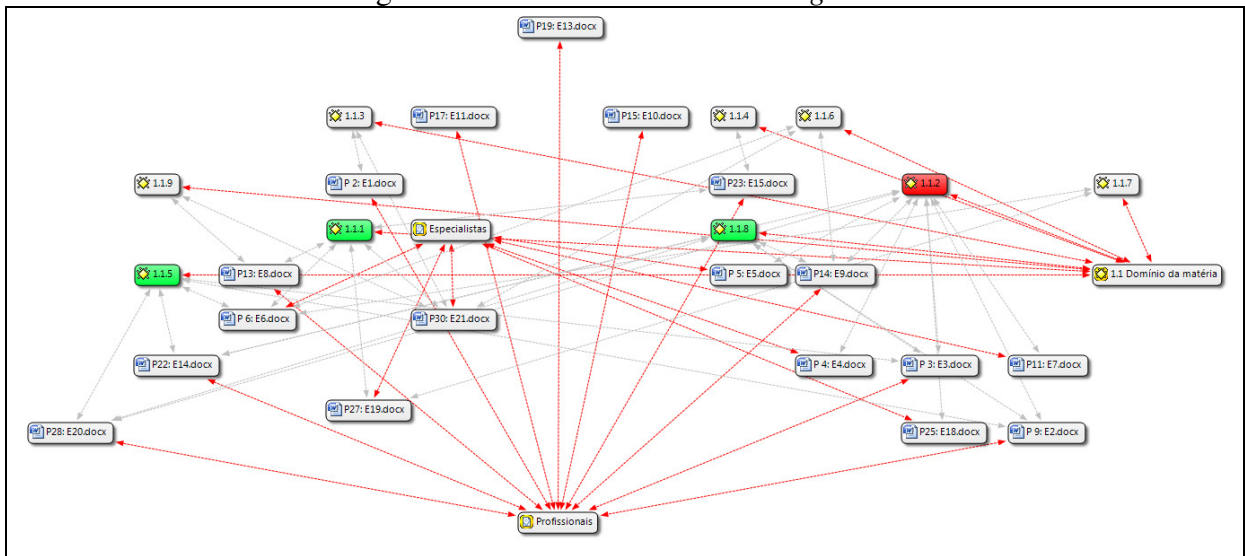
Fonte: elaborado pelo próprio autor.

Neste momento se apresentam dois extremos, pois de um lado se tem um profissional experiente que possui domínio da matéria, conteúdo e bom trânsito em determinada comunidade, do outro se tem um jovem profissional que já nasce em um ambiente tecnologicamente adequado ao *big data* e com isso tem mais facilidade em tratar com as ferramentas de manipulação para descobrir padrões em grandes volumes de dados. São habilidades diferentes e complementares. Como *big data* talvez não seja ainda uma realidade para as organizações, predominam as citações que valorizam a experiência do profissional em detrimento da utilização da tecnologia com elemento facilitador do processo de análise.

#### **4.1.3 Dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data***

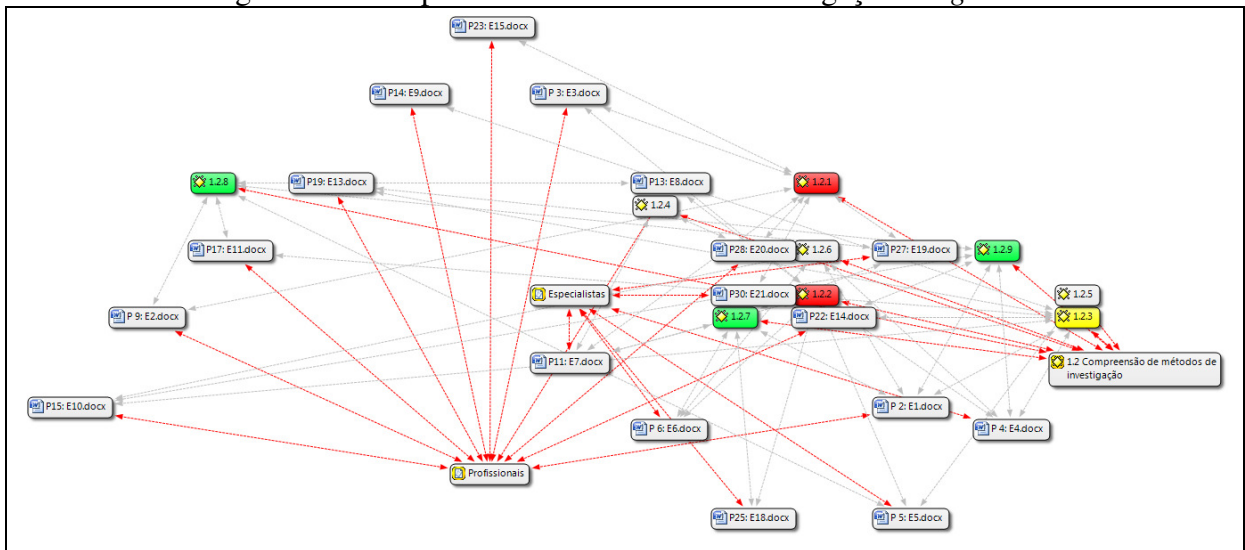
Nesta seção serão identificadas as dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data*. Este processo foi realizado com base nas percepções dos Especialistas e Profissionais, coletadas durante as entrevistas realizadas pela autora. Na ocasião os entrevistados foram questionados com relação aos efeitos dos 3 Vs de *big data* para cada uma das atividades do profissional em IE, utilizadas neste estudo como categorias. As categorias resultantes com maior número de citações foram identificadas nas figuras a seguir pelas cores vermelha, que representa mais de 8 citações, amarela, que representa entre 6 e 7 citações, e verde com 5 citações. Esta foi a distribuição média adotada em função de que as categorias resultantes mais citadas tiveram no máximo 12 citações e no mínimo 1 citação. Nesta análise, as categorias com menos de 5 citações não foram consideradas.

Em se analisando especificamente a primeira categoria conceitual "domínio da matéria", na Figura 43 é possível identificar as resultantes com maior número de citações:

Figura 43 - Domínio da matéria e *big data*.

Fonte: elaborado pelo próprio autor.

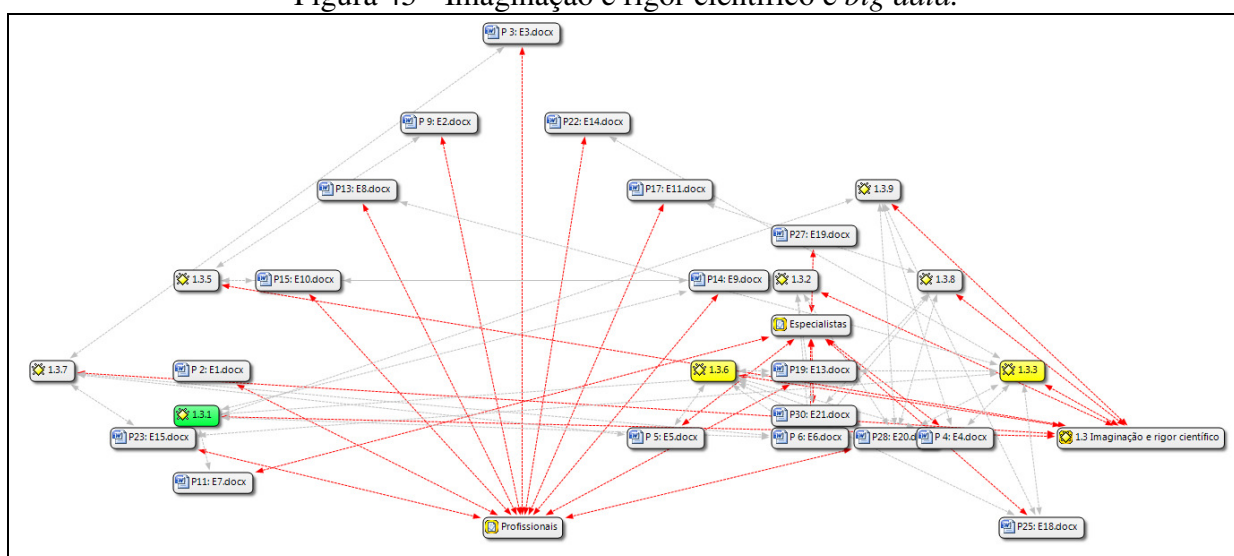
Assim, 1.1.2 - volume prejudica o domínio da matéria; seguida por, 1.1.1 - volume contribui para o domínio da matéria; 1.1.5 velocidade prejudica o domínio da matéria; 1.1.8 - variedade prejudica o domínio da matéria, foram as categorias resultantes mais citadas, demonstradas na Figura 43 em cor vermelha e verde. Em geral, o maior número de citações desta categoria, caracteriza o fenômeno *big data* como prejudicial. A Figura 44 apresenta a segunda categoria conceitual "compreensão de métodos de investigação":

Figura 44 - Compreensão de métodos de investigação e *big data*.

Fonte: elaborado pelo próprio autor.

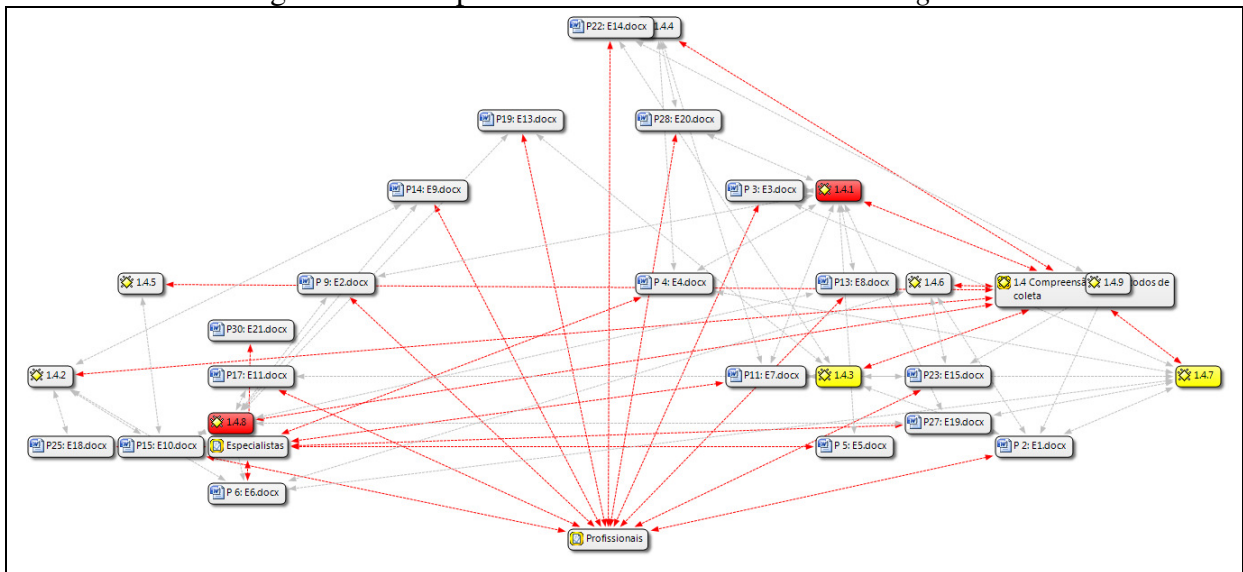
Nesta categoria, de acordo com a Figura 44, as resultantes com maior número de citações foram: 1.2.1 - volume contribui para a compreensão de métodos de investigação; e, 1.2.2 - volume prejudica a compreensão de métodos de investigação, representadas em vermelho. Em amarelo está a categoria 1.2.3 - volume não afeta a compreensão de métodos de investigação. As categorias em cor verde foram: 1.2.7 - variedade contribui para a compreensão de métodos de investigação; 1.2.8 - variedade prejudica a compreensão de métodos de investigação; 1.2.9 - variedade não afeta a compreensão dos métodos de investigação. Nesta categoria há citações controversas sobre os efeitos do fenômeno. A Figura 45 expõe a terceira categoria conceitual "imaginação e rigor científico":

Figura 45 - Imaginação e rigor científico e *big data*.



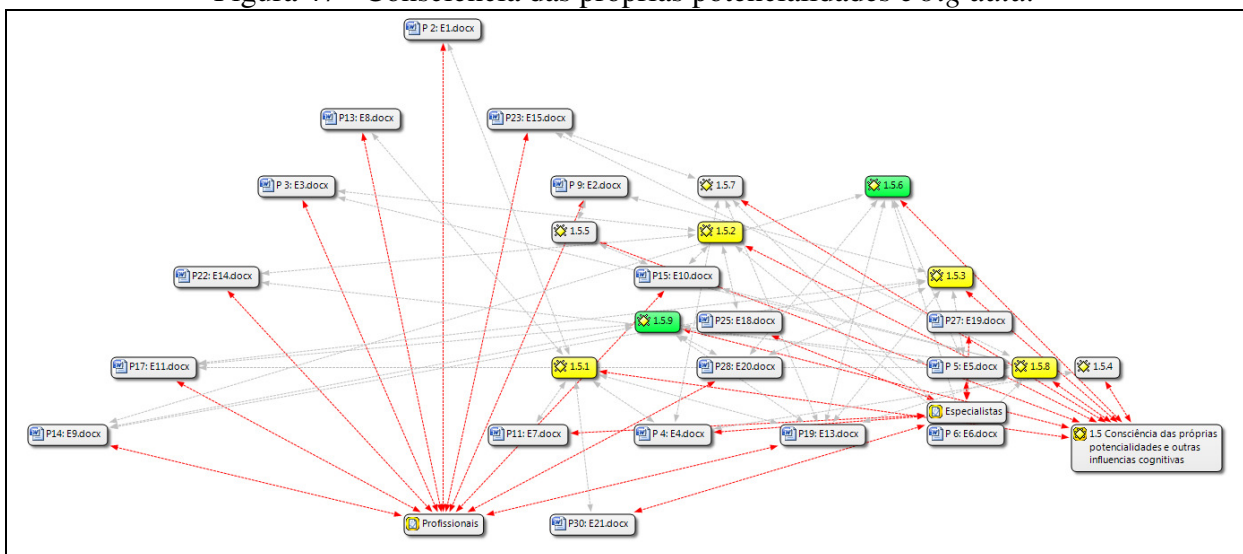
Fonte: elaborado pelo próprio autor.

Nesta categoria, as resultantes mais citadas foram: 1.3.3 e 1.3.6 - volume e velocidade não afetam a imaginação e rigor científico, em amarelo; e, 1.3.1 - volume contribui para a imaginação e rigor científico, em verde. A Figura 46, apresenta as citações da quarta categoria "compreensão dos métodos de coleta":

Figura 46 - Compreensão dos métodos de coleta e *big data*.

Fonte: elaborado pelo próprio autor.

Nesta categoria as resultantes com mais citações foram: 1.4.1 - volume contribui para a compreensão dos métodos de coleta; e, 1.4.8 - variedade prejudica a compreensão dos métodos de coleta, em vermelho. As amarelas foram: 1.4.3 - volume não afeta a compreensão dos métodos de coleta; e, 1.4.7 - variedade contribui para a compreensão dos métodos de coleta. A Figura 47, apresenta a quinta categoria conceitual "consciência das próprias potencialidades e outras influencias cognitivas":

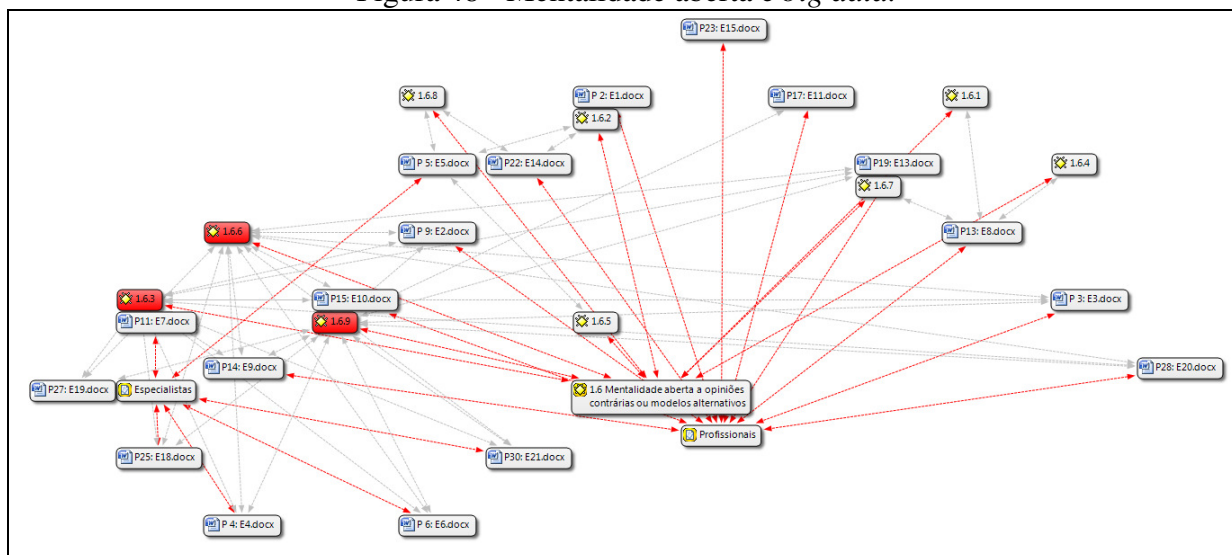
Figura 47 - Consciência das próprias potencialidades e *big data*.

Fonte: elaborado pelo próprio autor.



De acordo com a Figura 47, nesta categoria as resultantes mais citadas foram: 1.5.1 - volume contribui para a consciência das próprias potencialidades e outras influências cognitivas; 1.5.2 - velocidade prejudica a consciência das próprias potencialidades e outras influências cognitivas; 1.5.3 - volume não afeta a consciência das próprias potencialidades e outras influências cognitivas. Outra categoria que aparece em amarelo é: 1.5.8 - variedade prejudica a consciência das próprias potencialidades e outras influências cognitivas, em amarelo. As resultantes em verde foram: 1.5.6 e 1.5.9 - velocidade e variedade não afetam a consciência das próprias potencialidades e outras influências cognitivas. A Figura 48 apresenta a sexta categoria "mentalidade aberta a opiniões contrárias ou modelos alternativos":

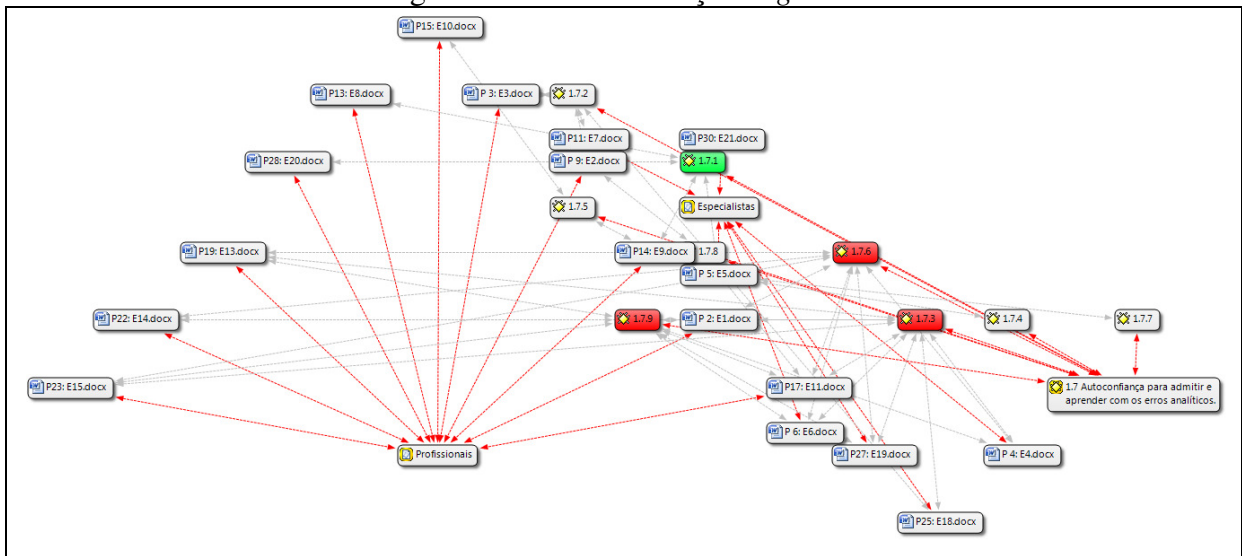
Figura 48 - Mentalidade aberta e *big data*.



Fonte: elaborado pelo próprio autor.

De acordo com a Figura 48, as categorias resultantes em vermelho foram: 1.6.3, 1.6.6 e 1.6.9 - volume/velocidade/variedade não afetam a mentalidade aberta a opiniões contrárias ou modelos alternativos. A Figura 49 expõe a sétima categoria conceitual "autoconfiança para admitir e aprender com erros analíticos":

Figura 49 - Autoconfiança e big data.

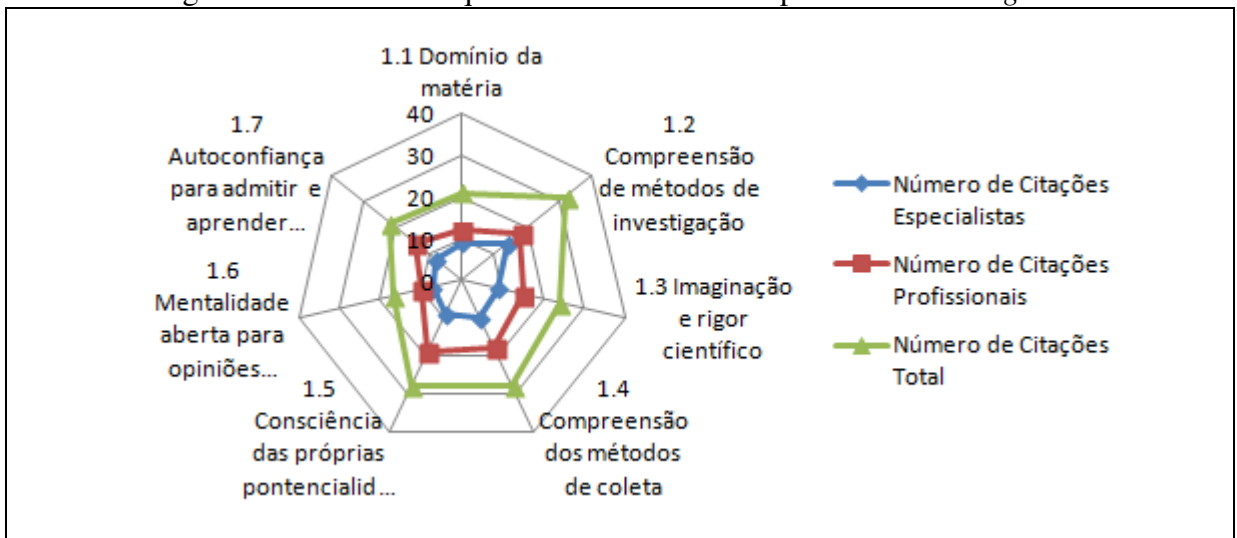


Fonte: elaborado pelo próprio autor.

De acordo com a Figura 49, as resultantes mais citadas foram: 1.7.3, 1.7.6, 1.7.9 - volume/velocidade/variedade não afetam a autoconfiança para admitir e aprender com erros analíticos, em vermelho. Em verde, também recebeu destaque a categoria resultante: 1.7.1 - volume contribui para a autoconfiança para admitir e aprender com erros analíticos.

Assim, atendendo ao objetivo específico "c) identificar as dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data*", foram apresentadas de acordo com cada categoria conceitual as resultantes com maior número de citações. A Figura 50 apresenta uma síntese dos resultados encontrados:

Figura 50 - Dimensões que estão sendo afetadas pelo fenômeno *big data*.



Fonte: elaborado pelo próprio autor.

Na Figura 50 foram agrupadas as citações das categorias resultantes totalizando as citações da categoria conceitual, que foram analisadas na perspectiva do grupo Especialistas e Profissionais. O número total de citações analisadas serviu de base para a construção da escala apresentada no gráfico. Desta forma, a categoria conceitual com maior número de citações para o grupo Especialistas, e número total de citações foi "compreensão de métodos de investigação", para os Profissionais foi "consciência das próprias potencialidades e outras influências cognitivas". Tanto para Especialistas, quanto para Profissionais, a categoria conceitual com menor número de citações foi "mentalidade aberta para opiniões contrárias ou modelos alternativos".

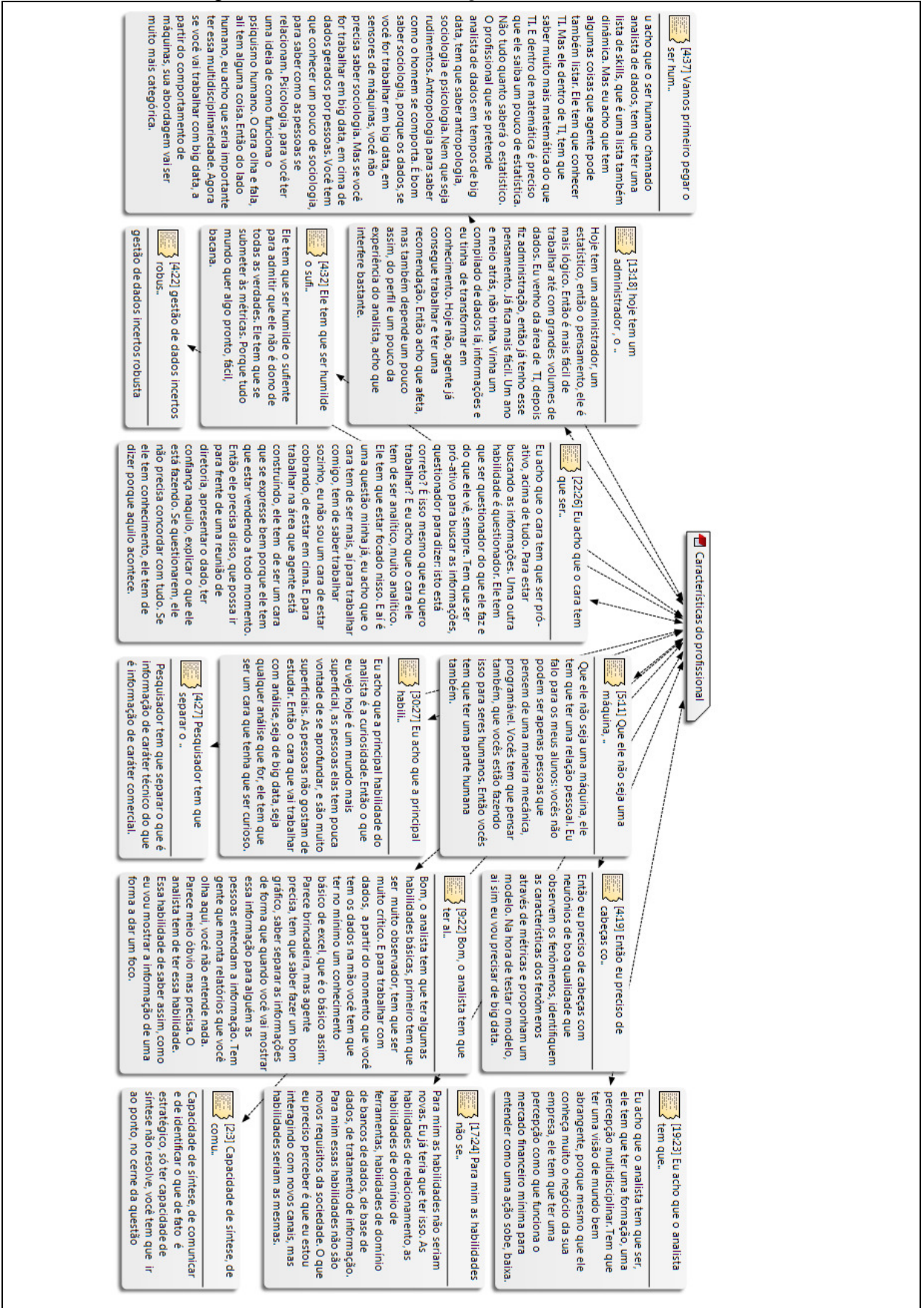
No sentido de buscar uma interpretação para estes dados, é válido mencionar que entre as características do grupo Especialistas está o fato da consciência e aceitação do fenômeno *big data*. Assim, o desafio que é apresentado a este grupo está na adequação de métodos e ferramentas para que seja possível extrair valor dos grandes volumes de dados. Este desafio passa pela construção de modelos de gestão da informação e da atividade de IE, e persiste com a obtenção e entendimento acerca das ferramentas tecnológicas atualmente disponíveis, tanto no nível de *hardware* quanto *software* para que seja possível armazenar, recuperar e processar a informação que serve como matéria-prima para o processo de análise.

É característica do grupo Profissionais o fato de que o fenômeno *big data* ainda é algo desconhecido e ameaçador. Desta forma há ainda uma necessidade de um melhor entendimento acerca das próprias potencialidades do profissional, bem como de que forma ele irá lidar com as influências cognitivas a que estará sujeito devido a incidência de um maior volume, variedade e velocidade de informações. Por fim, mentalidade aberta para opiniões contrárias ou modelos alternativos foi menos citada, em função de que de forma geral os entrevistados entendem que está é uma característica inerente à atividade de IE. A próxima seção irá tratar mais detalhadamente desta interface identificada que diz respeito às características do profissional em IE.

#### **4.1.4 Interfaces entre *big data* e atividade analítica no processo de interpretação de informações para a atividade de IE**

Durante a realização das entrevistas, as características do profissional em IE foram interpretadas como interface entre *big data* e atividade analítica. Estas características foram identificadas por meio de 13 citações, apresentadas na Figura 51:

Figura 51 - Interface entre big data e atividade analítica.



Fonte: elaborado pelo próprio autor.

Dos trechos de entrevistas marcados com questões referentes às características do profissional, pode-se ressaltar a questão das habilidades multifuncionais. Um dos entrevistados coloca a importância de três domínios, antropologia, sociologia e psicologia, além de TI com ênfase em matemática. Isto para conseguir o entendimento sobre como as pessoas se comportam, como se relacionam e como funciona o psiquismo humano, uma vez que este profissional está lidando com dados gerados por pessoas.

A capacidade de síntese de conteúdos e o saber comunicar este conhecimento, dentro da organização também foram mencionados como importantes para o profissional, juntamente com o domínio de ferramentas tecnológicas para conseguir fazer a gestão de dados incertos robusta. Características da literatura que fala de IE, também foram citadas como "curioso", "pró-ativo" e "questionador". Assim, percebe-se neste ponto uma interface entre os temas, que está centrada na figura do profissional.

#### **4.1.5 Efeitos fenômeno *big data* no processo de análise**

Esta seção apresentará os efeitos do fenômeno *big data* no processo de análise na atividade de inteligência estratégica, objetivo geral deste estudo. A análise das respostas foi organizada seguindo uma lógica de interface das categorias estabelecidas.

##### **4.1.5.1 Volume/Velocidade/Variabilidade contribuem para o domínio da matéria**

De acordo com os Especialistas entrevistados, o volume contribui para o domínio da matéria - domínio de conteúdos e temas, no processo de análise, em função de trazer múltiplas perspectivas ao processo. Entretanto para isto, o profissional deve estar preparado para a atividade fim, ele deve ser capaz de analisar este volume de informações, pois quanto menos experiência ele tiver, mais sujeito ele estará às influências cognitivas do processo.

A possibilidade da utilização de diversas dimensões de conteúdo acerca de um tema está também relacionada à variedade. No ponto de vista de outro entrevistado, esta variedade de fontes, inclusive, pode fazer com que uma pessoa sem experiência, consiga internalizar muito mais conhecimento do que um profissional com experiência, num curto espaço de tempo. Não foram encontradas evidências de contribuições da velocidade na capacidade de domínio da matéria, na visão dos Especialistas. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E21 (falando de volume): Então o cara que não está preparado, que não foi devidamente treinado, afeta sim. Se o camarada está bem treinado, conhece o negócio, então acho que o volume de informação ele ajuda sim, porque ele pode te dar múltiplas perspectivas.*

*E6 (falando sobre volume/variedade): Aprendi muito sobre indústria automotiva, com um time meu que estava trabalhando há um ano. Porque? Porque estes caras foram atrás de grandes volumes de informação, estes caras tiveram uma capacidade ímpar de dominar um conjunto maior de informações do que eu, quer dizer, eles pegaram diversas dimensões do conteúdo da indústria automotiva - eu associaria isto à variedade; e eles tiveram uma velocidade de pesquisa, triagem e processamento muito maior do que eu que estava na indústria.*

Na visão dos Profissionais, o volume de dados e informações contribui no sentido de qualificar melhor a informação e o conhecimento a ser utilizado na tomada de decisão, em forma de inteligência. O volume e a velocidade na geração da informação também foram colocados como pontos que contribuem ao processo de inteligência. Não foram identificadas citações sobre contribuições da variedade nesta categoria. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E8 (falando do volume): Então eu não avalio que a quantidade de informação vá me atrapalhar, bem pelo contrário, quanto mais informação e quanto mais eu conseguir qualificar ela e transformar em conhecimento, eu vou tomar a decisão mais assertiva.*

*E15 (falando do volume/velocidade): Hoje nós temos mais informação. Há um tempo atrás eles nos davam informação de 2004, 2006, hoje se tu olhar o eletrodoméstico, já consigo quase que anual a informação da produção brasileira. Automotivo tu consegue até o emplacamento mensal de cada carro.*

#### 4.1.5.2 Volume/Velocidade/Variedade prejudicam o domínio da matéria

Com relação aos efeitos negativos do volume no domínio da matéria, na visão dos Especialistas, foram expostos pontos referentes à questão da qualidade das informações e da ausência de ferramentas para tratar com grandes volumes de dados. Em função disto, uma crítica deve ser feita sobre qual informação está sendo buscada, pois diante do volume, a qualidade pode ficar prejudicada. Também, o tratamento destas informações torna-se fundamental para que não hajam sobrecargas, que possam vir a prejudicar o domínio de conteúdo do profissional no processo de análise.

Ainda na visão dos Especialistas, o ponto negativo da velocidade no domínio da matéria está relacionado à questão da obsolescência do conhecimento. De acordo com os Especialistas, em se partindo do princípio que, para se ter domínio sobre algum tema o

profissional deverá ou ter experiência, ou estar envolvido em seu contexto, o conhecimento internalizado ou recém adquirido rapidamente perderá o valor em função da velocidade da geração de conteúdos e informação. Não foram encontradas evidências para efeitos negativos da variedade na capacidade de domínio da matéria, na visão dos Especialistas. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E4 (falando do volume): Isso acontece, já acontece e eu lamento te informar que vai acontecer cada vez mais. O que agente propõe para eles, é que eles façam uma crítica com relação a quantidade de informação que eles estão buscando, porque agente já sabe que quanto maior a quantidade de informação, pior é a qualidade da mesma informação.*

*E7 (falando do volume): Eu acho que sim, eu acho que quando torna mais complexo, torna mais difícil. A questão toda é ter ferramentas. A quantidade, todo mundo vai dizer que é bom, só que se você não tiver como tratar a quantidade, vai ocorrer o que se chama sobrecarga, vai ficar perdido nas informações.*

Na visão dos Profissionais o efeito negativo do volume de informações está no fato da restrição. Assim, o foco na informação relevante é uma prática adotada pelos Profissionais, para que estes possam se proteger do grande volume de dados e informações. Entretanto, esta situação foi interpretada como negativa pela autora, pois de acordo com a base teórica de *big data*, muitas vezes existem padrões desconhecidos que somente serão possíveis de se identificar a partir da utilização de grandes volumes de dados. Esta restrição do foco é mencionada também em outros trechos de entrevistas. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E14 (falando de volume): Sempre primando pelo pouco, pouca informação, mas que fosse realmente relevante. Então, se a informação não é relevante, não saí da área. Esse foi o primeiro critério, justamente por ter muito dados. Se eu trabalhar com muita coisa, primeiro ninguém entende nada e não muda nada.*

*E2 (falando de volume / velocidade / variedade): Atrapalha , tu acaba tendo um volume de informação muito grande para ter que estar monitorando, olhando, lendo. Assim, quanto mais informação mais tempo você tem que ter para digerir isso e separar o que presta e o que não presta o que é sinal forte, do que é sinal fraco, ou nada. Agente acaba, às vezes, eu às vezes deixo até de olhar algumas coisas, quando eu vejo que aquilo pode me tirar um pouco do foco.*

#### 4.1.5.3 Volume/Velocidade/Variedade não afetam o domínio da matéria

Na visão dos Especialistas *big data* não afeta o domínio da matéria quando há um processo de governança para IE instituído na organização. Assim, os efeitos positivos e negativos do volume e variedade não serão sentidos pelo profissional, se ele estiver amparado

por um método. Quando o questionamento foi sobre a categoria velocidade, o ponto de vista colocado por outro entrevistado foi de que esta característica não seria uma dimensão a ser considerada para *big data*, uma vez que a informação útil em formato original é apenas uma. Neste contexto, o que ocorre é a disseminação em massa de uma mesma informação por vários canais. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E21 (falando do volume/variedade): Vai depender do quanto o analista vai estar preparado. Agente tem uma atividade, quando agente vai fazer um trabalho que é o mapa de governança. O mapa de governança trata sobre as fontes que vão ser trabalhadas, vai tratar sobre escopo central, que é o debriefing, qual é o bloco de conteúdo que você vai trabalhar, se você vai precisar acessar analista ou não, se você vai trabalhar com base de dados, então o volume de informação, se o analista está preparado, se ele tem método para isso eu acho que não é um ponto negativo. Porque ele consegue separar.*

*E6 (falando da velocidade): Eu não sei como velocidade impacta nisso, eu tenho uma visão bem particular sobre velocidade, que velocidade não tem nada a ver com big data. Nada a ver, o mundo está produzindo o mesmo volume de informações que produzia, só que como o crescimento de produtores foi exponencial, a velocidade de produção é exponencial. Agora, você está falando de informação ou de informação útil. Existe uma diferença tremenda nisso.*

Para os Profissionais, as dificuldades relativas ao domínio da matéria, estão relacionadas à falta de profissionais com conhecimento dos dados disponíveis. Outro entrevistado, Profissional, concorda com a visão colocada pelo entrevistado (E6), Especialista, nesta categoria resultante, de que hoje se tem uma repetição de dados muito maior do que, de fato, velocidade. Assim, o que passa a influenciar a questão do domínio da matéria é mais o setor de atuação do profissional em inteligência. Isto porque dependendo da indústria, a velocidade na propagação de dados é maior do que outras, exemplo: moda e metal mecânico. Outra visão da não influência do fenômeno *big data*, está quando os Profissionais não utilizam *big data* para a análise em inteligência estratégica. Utilizam mais dados estruturados, internos à organização e desta forma, não percebem seus efeitos. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E1 (falando de volume): O que agente se depara é justamente com uma dificuldade de identificar, de chegar, de ter nas nossas equipes profissionais que tenham um conhecimento um pouco mais profundo e conceitual dos dados que estão sendo disponibilizados, para que se possa de fato utilizar, para transformar isso em informação.*

*E9 (falando da velocidade): Daí depende do nível de profundidade que tu vai querer fazer uma análise e depende da indústria, do mercado onde está inserido. Eu já trabalhei no mercado de Telecom, e hoje eu trabalho na indústria. Então a velocidade*



*num mercado de telecomunicações é muito maior que na indústria. Então hoje eu tenho uma repetição de dados maior do que a velocidade do que aparece.*

*E8 (falando da variedade): Na verdade agente aqui, hoje trabalha com dados estruturados, então boa parte das informações eu tenho dentro da casa. Eu tenho dados já de série histórica, dados que eu já tenho e dados dos meus clientes porque agente tem um volume muito grande. Eu não chego a coletar informações de redes sociais.*

#### 4.1.5.4 Volume/Velocidade/Variabilidade contribuem para a compreensão de métodos de investigação

Na visão dos Especialistas, volume, velocidade e variedade contribuem para esta categoria, no sentido do aperfeiçoamento e evolução, adequada à atual complexidade, de métodos de investigação para organizar e avaliar dados. Em se falando especificamente do volume, a forte influência de estatística robusta utilizando grandes massas de dados, diminui o erro amostral e leva ao refinamento nas visões. Para os Especialistas, a variedade também exerce um efeito positivo no sentido de proporcionar outras dimensões informacionais, que podem agregar ao desenvolvimento de um modelo preditivo, complementando assim a dimensão volume. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E21 (falando do volume): Ele fazia muita análise estatística, então o volume afetava muito o desempenho dele. Porque, como ele não conseguiu pegar uma amostragem muito grande, então o erro estatístico dele era muito grande. A tecnologia neste sentido, deu um ganho muito grande para ele. Então imagina só, numa modelagem estatística você passa de uma amostra de 1000, ela pula para 80000, você vai ter diferentes visões para trabalhar e muitas análises para serem feitas. Então neste sentido que eu estou te colocando, a tecnologia, o big data, ele traz um refinamento estatístico que possibilita visões mais refinadas.*

*E7 (falando do volume / velocidade / variedade): Eu acho que sim, as pesquisas estão gerando novos métodos, novas metodologias de coleta de análise, isto talvez até impulsionado mesmo pelos 3 Vs. Eu acho que o volume está crescendo, a variedade está crescendo, a complexidade está crescendo, então a complexidade dos métodos, das ferramentas também tem que aumentar na mesma medida.*

*E6 (falando da variedade): Eles não conseguem ter um modelo preditivo, um modelo que diga para eles qual vai ser o tamanho do mercado e qual vai ser a projeção de venda deles. Eles não tem isso. Porque eles não estão levando em consideração uma variedade maior de informação do que eles têm hoje. Eles estão pensando que este problema é volume, então 3 anos de série histórica está ruim, deixa eu pegar mais informação. Vamos pegar 5 anos, agora já está em 6 anos, eles não viram que o problema não é volume, é a variedade. Existem outras dimensões informacionais que melhorariam muito o modelo de previsão, mas eles não estão enxergando isso.*

Para os Profissionais, o volume proporciona facilidades para o cruzamento de dados. A questão do melhor acesso à dados e informações também foi citada, como benefício relacionado ao volume e a variedade para investigar dados. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E3 (falando do volume): O trabalho ficou mais fácil. Quanto mais informação eu tenho para mim mais fácil, pois daí tu vai cruzando os dados.*

*E20 (falando do volume/variedade): Em relação a dados externos isso também é verdadeiro, porque o volume de informações foi ficando tão mais robusto com a internet, com uma facilidade de acesso à artigos, à conhecimento e à formação de opinião, além de movimentos em outros países. Agente consegue monitorar isso de forma mais fácil.*

#### 4.1.5.5 Volume/Velocidade/Variedade prejudicam a compreensão de métodos de investigação

Os efeitos prejudiciais de *big data* na compreensão dos métodos de investigação para organizar e avaliar dados foi uma das categorias resultantes mais mencionadas pelos Especialistas. Dos 7 entrevistados, 6 confirmaram estes efeitos na categoria. Eles estão relacionados a falta de profissionais preparados para lidar com os métodos adequados ao *big data*. Assim, a falta de profissionais com conhecimento estatístico, a falta das ferramentas tecnológicas adequadas e a utilização dos métodos investigativos tradicionais desenvolvidos a partir de hipóteses, estão entre os pontos destacados nos trechos a seguir:

*E7 (falando do volume): Mas quem está no mercado normalmente só repete coisas. Não consegue definir novos métodos, novas metodologias, é só o que as empresas estão impondo. É só o que tem no software. É só o que o software permite fazer. Então muitas vezes os analistas não combinam sua capacidade de análise com a capacidade de análise do software, eles deixam o software fazer tudo. O pessoal não está fazendo, não está explorando todo poder do big data. Então, na verdade o big data hoje é um entrave. Não é uma vantagem para esse pessoal que está no mercado.*

*E6 (falando do volume): eu acredito que quando você trabalha método de investigação a gente ainda não sabe trabalhar com big data. Agente ainda está muito baseado em métodos investigativos a partir de hipóteses. Agente não está construindo estas hipóteses a partir do grande volume de dados.*

*E5 (falando do volume / velocidade / variedade): Eu acho que não estão não. Pelo que eu tenho visto, eu acho que eles ainda estão nos velhos e bons métodos antigos de análise: SWOTs, modelos de excel e macros.*

*E19 (falando da variedade): numa corporação, você tem vários sistemas, muitas vezes isolados, vencidos. Você tem o sistema de gestão empresarial que tem em muitas empresas, e tem vários sistemas departamentais gerando arquivos diferentes com formatos e bancos de dados diferentes. Daí você teria que ter uma tecnologia, já existe, mas às vezes a empresa não tem quem colete e consiga acessar estes dados e dar nexos e*

*correlacionar isso. E tem que ter uma ferramenta de correlacionamento de dados. Não adianta você ter os dados, espalhados se eu nem sei onde eles estão, não consigo tratá-los, analisá-los em tempo hábil para conseguir tomar uma decisão de negócios.*

Para os Profissionais, o volume faz com que seja necessário filtrar mais a informação e isto demanda um tempo maior por parte do analista, no processo de análise. Estes filtros também prejudicam ao passo que para que se faça um monitoramento de qualidade, devem ser utilizados sempre os mesmos critérios e isto, por vezes, pode ser esquecido pelos profissionais que estão manipulando a informação. Os Profissionais também apontaram o fato de que o volume lhes proporciona ir no detalhe da informação, mas sem a certeza de que ao fim da investigação, serão encontradas informações úteis para o processo. Na visão de outro entrevistado a variedade gera uma necessidade de mudança nos relacionamentos com os canais de geração e tratamento de informação na investigação, para organizar e avaliar dados. Esta mudança é interpretada como prejudicial, pois vai requer uma adequação nas habilidades técnicas, relacionais e emocionais do profissional. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E3 (falando do volume): A grande quantidade de informações que vem, acaba prejudicando um pouco. Porque a pessoa também tem que estar se dedicando a filtrar e em determinar bem, qual é o foco. No passado, agente tinha mais tempo para parar, analisar, planejar e a velocidade, a pressão por metas, a pressão por produtividade é muito grande hoje nas empresas.*

*E3 (falando do volume): Há uma grande dificuldade de cruzamento desses dados, especialmente de separar o que é a qualidade desses dados. O que é relevante e o que é não relevante. Hoje muitas informações, muitos cruzamentos e muitos relatórios que se geram, a gente se depara com uma série de informações, muitas vezes agente tem de estar elegendando: essas aqui são a prioridade, essas aqui deixa de lado. Porque pega um assunto, se quiser estratificar, segmentar ou clusterizar, realmente consegue ir no detalhe de cada item.*

*E14 (falando do volume): Por causa do volume de dados e da variedade de filtros que eu tenho para chegar no dado que eu quero, pode impactar diretamente. Este é o maior problema que agente tem. Usar sempre o mesmo critério e como tem muito critério, isso pode gerar um monte de erro.*

*E11 (falando da variedade): Cada indivíduo hoje é uma fonte geradora de conteúdo, no mundo. Então eu acho que a variedade impacta sim, nos relacionamentos com os canais de geração e tratamento de informação. Isso vai requerer habilidades técnicas, relacionais e emocionais do analista de ser para lidar com isso.*

#### 4.1.5.6 Volume/Velocidade/Variedade não afetam a compreensão de métodos de investigação

Na visão de um dos Especialistas, a não influência do volume, velocidade e variedade nesta categoria, se deve ao fato de que mesmo antes do fenômeno *big data* já se utilizavam técnicas robustas na investigação, para organizar e avaliar dados, conforme trecho a seguir:

*E4 (falando do volume / velocidade / variedade): Então muito antes do big data ser a moda que é hoje, agente já lida com grandes quantidades de dados. A gente inclusive trata estas quantidades massivas de dados com modelos de informática que a gente chama de HPC (high performance computing), neste cenário, sendo muito honesto com você, não é a moda do big data que está melhorando a capacidade que a gente tem de realizar medições, de realizar pesquisas e algumas inferências. Não é.*

Na visão dos Profissionais, a adoção de metodologias e a restrição do foco da análise, faz com que os efeitos de *big data* não sejam percebidos nesta atividade em IE. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E9 (falando do volume): Informação sempre teve muita, era uma questão de necessidade de estruturar realmente uma metodologia de trabalho para a pesquisa, para a própria empresa. Mas não em função da quantidade de informação, falamos hoje de quatro anos atrás, então de quatro anos atrás para agora, o nível de informação ele não mudou.*

*E14 (falando do volume/variedade): Acho que volume hoje, por agente adotar isso de somente o que é preciso. A gente não tem mais problema com isso. A gente focou e disse: eu preciso olhar isso. Então o volume de dados que eu tenho e que eu procuro são suficientes para fazer o que eu quero agora.*

#### 4.1.5.7 Volume/Velocidade/Variedade contribuem para a imaginação e rigor científico

A categoria imaginação e rigor científico para chegar a hipóteses, na visão dos Especialistas, é beneficiada pelo fenômeno *big data* quando o profissional não parte de premissas *a priori*, para depois testá-las, de acordo com o método tradicional. Nesta visão, para que seja possível extrair todo o valor de *big data*, a condução do processo deve ser feita no sentido de identificar grandes dimensões ou taxonomias, e por meio dos grandes volumes de dados, oriundos de fontes diversificadas, entender fenômenos que antes não seriam pensados. A seguir é apresentado trechos de entrevista que confirma a análise realizada:

*E6 (falando de volume/variedade): Quando você não parte de premissas, o fenômeno big data, te ajuda a identificar grandes dimensões, te ajuda a identificar uma taxonomia para aquilo que você quer estudar. A partir das taxonomias você pode ir entendendo os fenômenos. Estes fenômenos podem ser fenômenos de consumo, fenômenos de comportamento de clientes, fenômenos de comportamento de vendas de produtos, fenômenos de comportamento de perda e ganho de clientes, fenômenos de comportamento de preços, fenômenos de comportamento de matéria-prima, tem N inferências que você pode tirar através do estudo destas grandes dimensões. Agora, se*

*you have little variety of information and if you have little volume of information, you will not be able to do it. Your inferences will be poor.*

Na visão dos Profissionais, o volume contribui para esta categoria conceitual, pois antes, eram inferidas ideias em cima de conjuntos limitados de dados. Assim, mesmo tendo a percepção de que as informações não eram as mais adequadas, os profissionais tomavam ações em cima daqueles dados que não eram precisos. Outro Profissional aponta como contribuição da variedade, o fato do fácil acesso à informação e possibilidade de cruzamento de dados de boas fontes para a atividade. Ele lembra quando não haviam dados disponíveis e a informação era vendida a altos preços, para que se pudessem gerar análises. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E13 (falando de volume): Você percebia que existia um risco nesta conclusão. Inferir ideias em cima de um conjunto limitado de dados e pressupor aquilo, em determinado momento, como verdade sem ter realmente um conjunto mais consistente de dados que validassem aquelas hipóteses. Eram situações que se tinha essa percepção de que estas informações, os dados iniciais não eram consistentes o suficiente para levar àquela conclusão.*

*E3 (falando de variedade): Ele é mais fácil, porque hoje você tem uma variedade muito grande de informações, então ele acaba sendo mais fácil. Você trabalha com um potencial de consumo, tu consegue fácil chegar à dados. A dificuldade de acessar a informação era muito maior e as informações eram vendidas. Hoje você coloca na internet, faz alguns cruzamentos ali, você consegue chegar à dados de boas fontes e de uma forma segura. Então agente consegue de uma forma científica ter decisões muito mais assertivas.*

Não houveram citações relacionadas às contribuições da velocidade para a imaginação e rigor científico, nem da parte de Especialistas, nem Profissionais.

#### 4.1.5.8 Volume/Velocidade/Variedade prejudicam a imaginação e rigor científico

Na visão dos Especialistas, o volume e a variedade prejudicam a atividade de imaginação e rigor científico para construção de hipóteses, em função da veracidade dos dados e informações. Assim, a utilização de dados de fontes duvidosas pode acarretar a construção de hipóteses fundamentadas em erros, conforme trecho a seguir:

*E21 (falando de volume/variedade): Tem muita coisa que está disponível que não foi checada. Então se você está lendo, acessando um trabalho de mestrado, um estudo, um artigo científico, você está acessando algum conteúdo não estruturado, tem muita informação que pode ter erro, então se você não checar, não fizer uma checagem disso aí, a tua hipótese já pode começar com erro. É sempre mais fácil quando você está trabalhando com uma informação que de alguma maneira ela tem confiabilidade.*

*Então você também tem maneira de dar um peso para aquele tipo de informação. Então tem método para isso, para você ver a confiabilidade daquela informação, mas o volume pode te induzir ao erro.*

Na visão dos Profissionais, o volume e a velocidade fazem com que, em alguns setores, a decisão tomada com base em informações oriundas de atividades de IE, tenha que ser revistas antes mesmo de serem implementadas, pois mudam muito rapidamente. A questão da variedade também foi citada, se referindo ao grande volume de dados não estruturados que circulam e causam dificuldades de análise, em função da falta de capacidade dos profissionais em tratar com eles. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E9 (falando do volume/velocidade) Todo final do mês, tem uma reunião de gestão, daí foi decidido um produto final para aquela reunião, um dia depois muda o cenário. Tem que voltar tudo, daí tu faz de novo uma seleção de cenário, passa 4 ou 5 dias muda de novo. Eu estou falando por exemplo, de volume de peças que tem de ser produzidas, então tem vezes que a quantidade de informações que tu tem, de vez em quando atrapalha.*

*E20 (falando de volume/variedade) Hoje principalmente falando de setor de saúde acho que o número de informações que trafegam entre prestadores, e prestadores e operadoras de saúde é uma quantidade de dados imensa. E essas informações, elas não estão organizadas. Então agente tem uma dificuldade de transformar estes dados em inteligência e trazer como resultado prático uma coisa que agregue valor para a empresa tomar suas decisões ou mesmo para construir ações, que ajudem a criar valor.*

#### 4.1.5.9 Volume/Velocidade/Variedade não afetam a imaginação e rigor científico

Na visão do *big data* não influenciando a categoria imaginação e rigor científico, outros elementos passam a figurar nos trechos das entrevistas coletadas com o grupo Especialistas e Profissionais. Para o grupo Especialistas, a questão da observação do fenômeno a partir pelo olhar do profissional é apontada como elemento fundamental para a construção de hipóteses. E isto, sem a influência de *big data*, de acordo com trecho a seguir:

*E4 (falando de volume / velocidade / veracidade): A existência de uma massa grande de dados serve de balcão de ensaio para validar ou refutar a hipótese. Mas não é a massa de dados que me ajuda a formular a hipótese. A formulação da hipótese, ela é muito mais dependente de uma boa observação do fenômeno do que de uma massa de dados. Eu confio muito mais no neurônio do pesquisador que está observando o fenômeno e a partir da observação do fenômeno, ele vai me dar uma hipótese inacreditável. Aí nós vamos atrás da grande massa de dados para confirmar a hipótese.*

A definição prévia de uma questão bem formulada, foi mencionada tanto pelo grupo Especialistas quanto pelo grupo Profissionais. Na visão de ambos, a partir dela podem ser desenvolvidas hipóteses estruturadas. O fato é que esta definição depende do estabelecimento de políticas de informação que digam o que realmente é informação relevante. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E18 (falando de volume / velocidade / veracidade): Eu acredito que essa construção dos produtos de inteligência tem que partir de uma pergunta bem formulada. Eu tenho muita dúvida se estas perguntas estão sendo bem formuladas. Eu tenho muita dúvida se existe nas organizações uma política de informação. Tenho mais dúvida ainda, se as organizações se preocupam em realmente determinar o que é informação relevante para elas. Essas três questões que eu estou te falando, impactam até mais que o volume a flexibilidade, disponibilidade, tempo. Porque se eu não tenho uma boa pergunta, como que eu vou buscar as respostas?*

*E13 (falando de volume): Tu passa a ter um certo procedimento de onde buscar a informação, quais são os mecanismos, as chaves de busca. Tu começa a estabelecer um processo, mesmo que ele seja do dia-a-dia, uma coisa empírica, não formal. Tu passa a ter mais facilidade e a partir das relações que tu fez no passado, tu passa a replicar isso e ter um aprendizado. Nas análises que a gente faz hoje, a gente não tem essa questão do volume.*

*E20 (volume/velocidade/variedade): Mais fácil pelo fato de agente ter conseguido estruturar a nossa ambição e os pontos mais críticos para alguma ruptura de inovação no setor. Então eu acho que a partir do momento que agente conseguiu fazer esta definição ficou mais fácil. E agente tem uma série de variáveis, dentro de cada tema de possível ruptura no setor, e isso ajuda a gente a ganhar tempo e velocidade nas análises, estruturação de inteligência com recomendações para nossa diretoria.*

Para o grupo Profissionais a restrição de tempo faz com que recomendações sejam tomadas com um número limitado de dados. Assim, o tempo afeta a categoria, não *big data*. Outro ponto novamente mencionado é a questão do método, do processo. Como a atuação da inteligência geralmente é restrita a um ambiente de *small data*, em função de processos fechados de análise e inteligência, os efeitos de *big data* não são percebidos. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E8 (falando de volume): Poderia ter mais hipóteses à testar e a pensar. Agora na prática a gente tem um tempo para fazer. Então eu tenho um tempo, o que eu consigo fazer neste tempo, com as informações que eu tenho é isso que eu vou recomendar.*

#### 4.1.5.10 Volume/Velocidade/Variedade contribuem para a compreensão dos métodos de coleta

Na categoria compreensão dos métodos de coleta, de acordo com a visão dos Especialistas, *big data* tem suas contribuições se houver a utilização de ferramentas

tecnológicas adequadas à captura de grandes volumes de dados - volume, a partir de uma diversidade de fontes - exemplo *Twitter*, com o processamento das informações na mesma velocidade em que elas são geradas - velocidade. Neste sentido, apenas a existência de grandes volumes de dados ou a utilização de relações de causa e efeito para a sua interpretação, não fazem com que o valor seja extraído, de acordo com o trecho a seguir:

*E4 (falando de volume / velocidade / variedade): Na verdade o que agente fez aqui foi desenvolver uma ferramenta muito robusta de processamento semântico aplicada numa massa enorme de dados capturada no Twitter. Então isso aqui é uma aplicação de big data. Mas o que eu quero mostrar para você é que se nós não tivéssemos desenvolvido a ferramenta os dados do Twitter estariam lá.*

Na visão do grupo Profissionais, o grande volume de dados e informações fez com que a coleta se qualificasse mais, com a utilização de dados primários e secundários. A utilização de tecnologia com robôs de busca, e a compra de pesquisas prontas com empresas de consultoria, foram apontadas como mudanças nos métodos de coleta, em tempos de *big data*. A possibilidade de cruzamento e melhorias na qualidade das informações, advindas da variedade, também foram citadas. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E3 (falando do volume): A gente coleta dados primários, secundários, a gente trabalha, a gente tem um banco interno aqui, um BI bem grande de informações.*

*E2 (falando do volume): Com certeza (os métodos são diferentes), hoje a gente precisa de um robzinho de busca auxiliando a fazer uma peneira, porque que a quantidade de informação é grande.*

*E8 (falando do volume): Faz com que a gente precise se organizar melhor, e em algumas vezes, a gente preferiu fazer um grupo focal. Em alguns casos usa mesmo pesquisas, encomenda e faz em função também dessa grande disposição de informações.*

*E3 (falando de variedade): São tantas as informações que tu tem hoje a possibilidade de cruzar elas. Assim como você vê numa rede social que rastreia teus dados, que cruza com a rede, o CRM faz muito isso. Então a gente vem realmente procurando melhorar a qualidade das informações.*

#### 4.1.5.11 Volume/Velocidade/Variedade prejudicam a compreensão dos métodos de coleta

Os métodos de coleta são prejudicados pelo *big data*, na visão do grupo Especialistas, quando o monitoramento de diversas fontes gera muito mais informação do que as pessoas - clientes da inteligência, estão preparadas para receber, ou que as ferramentas de extração estão preparadas para capturar. Outro ponto mencionado foi com relação à busca da



informação específica e pontual, ou seja, aquela informação que irá ser matéria-prima para a inovação na organização e está em meio ao grande volume de dados. A influência da velocidade não foi citada, neste contexto, pelo grupo Especialistas. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E18 (falando de volume) Na inovação eu vou estar trabalhando com uma informação muito mais específica. E é aquela informação que pode gerar uma melhoria de processo ou até uma ruptura. E dentro deste processo de inovação, a informação é crucial: matéria-prima. Não tem como não lidar com ela. Como eu estou falando em volumes cada vez maiores, normalmente a minha dificuldade de encontrar aquela informação específica e pontual ela é realmente muito mais complexa.*

*E6 (falando de volume/variedade): A gente está monitorando 26 mil fontes de informação, agente viu que destas só 2100 fontes são usadas com frequência pelas pessoas. Então o que está acontecendo: o volume de informações que eu estou capturando e trazendo para os meus clientes não está ajudando eles em absolutamente nada. Eu poderia ter 2100 fontes de informação que eles teriam a mesma ajuda do que as 26 mil fontes estão dando. Porque a pessoa não tem capacidade de processar aquilo, não tem tempo suficiente, não tem tecnologia para extração de conhecimento destes grandes volumes de informação.*

Para o grupo Profissionais, novamente foi citada a questão da seleção e restrição das fontes utilizadas para análise, a compra de pesquisas em função da baixa acurácia dos dados coletados internamente pela empresa e a dificuldade em manter os produtos da inteligência atualizados, em função da velocidade da mudança nos dados coletados. Também foi levantado pelo grupo Profissionais, o fato de que alguns tipos de informação (não estruturadas), não são possíveis de serem coletadas por robôs. O efeito prejudicial disto seria traduzido no tempo que esta coleta demanda do profissional, tornando o processo mais demorado. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E9 (falando do volume/variedade): Para a gente gerar os produtos de inteligência que tem aqui dentro, foram selecionadas algumas fontes. E fontes que nós consideramos, que nós avaliamos em questão de resultados e até pesquisas que nós compramos. Então foi uma questão de que se tem muita informação hoje, mas a gente fez um filtro.*

*E10 (falando do volume/variedade): Exige um trabalho muito grande do profissional na hora da escolha das fontes, para que facilite depois a busca. O universo é enorme de informações. Então, tu tem que tomar o cuidado para não se perder.*

*E10 (falando da velocidade): Então tu tem que ter essa noção de montar um relatório, mas pelo menos um dia antes, ou algumas horas antes da reunião, tu tem que dar uma olhada novamente nos conteúdos e atualizar. Não existe pronto, não existe um software que faça isso pronto.*

*E8 (falando da variedade): Posso conseguir um robô para coletar informações, mas no nosso caso ainda teria que ter uma ou duas pessoas coletando informações por telefone. Então o nosso mercado está se estruturando, e a gente consegue ter informações até mais estruturadas. Por outro lado, tem mais concorrentes, então eu levo mais tempo, porque tenho mais fontes para procurar e para buscar essas informações.*

#### 4.1.5.12 Volume/Velocidade/Variabilidade não afetam a compreensão dos métodos de coleta

Para o grupo Especialistas, a única citação com relação a esta categoria vem da entrevista onde é dito que a velocidade não é uma dimensão para *big data*, em função de que a informação útil em formato original é apenas uma. O que ocorre neste caso é a disseminação em massa de uma mesma informação original. Esta marcação já foi citada na categoria "volume/velocidade/variedade não afetam o domínio da matéria" e outras onde a categoria resultante é "não afeta", com relação à velocidade.

Para os Profissionais, o fator que influencia a compreensão dos métodos de coleta é o método em si. Se o analista tiver um método que garanta a efetividade do processo, então o volume não irá influenciar. Outro ponto colocado, foi no sentido de saber exatamente o que se quer como produto de inteligência e definir uma estrutura de informações voltada para isto. A ausência da utilização de grandes volumes de dados, também é um motivo pelo qual o grupo Profissionais não percebe a influência do fenômeno. A seguir, são apresentados trechos de entrevistas que confirmam a análise realizada:

*E11 (falando do volume): Pode ter o montante de informações que você quiser, se tu tiver um método que garanta o resultado final da inteligência que tu quer gerar, eu acho que facilita o trabalho do analista de ser. Eu quero chamar atenção, que a questão não está na quantidade de informações, não é a quantidade que vai definir a qualidade do trabalho do analista de ser, é a presença ou ausência de métodos para fazer isto.*

*E1 (falando do volume / velocidade / variedade): As ferramentas de coleta, metodologias de coleta, isso é uma coisa que não é entrave hoje, o que é entrave é você saber exatamente o que você quer e definir uma estrutura de informações de dados, que você vai construir para poder responder as perguntas que você almeja.*

*E15 (falando do volume / velocidade / variedade): A gente sempre procura as associações. Da mesma forma, com as mesmas ferramentas. A gente analisou outros mecanismos, eu fiz uma comparação de uma ferramenta das mais renomadas e fiz um comparativo com o google alertas. O google alertas me traz muito mais informação.*

#### 4.1.5.13 Volume/Velocidade/Variabilidade contribuem para a consciência das próprias potencialidades e outras influências cognitivas

Para o grupo Especialistas, *big data* tem contribuído para a consciência das próprias potencialidades e outras influências cognitivas, no ponto em que as ferramentas tecnológicas disponíveis podem estar sendo utilizadas para automatizar o processo de coleta de um maior número de dados provenientes de fontes diversificadas, na mesma velocidade em que eles são gerados. Então o desempenho do profissional melhora neste sentido, refletindo o aprimoramento de seu potencial. Isto também faz com que a análise não seja apenas uma questão de percepção, o que poderia levar a produzir falsas representações de mundo, com a ocorrência de tendenciosidades. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E4 (falando do volume): O fenômeno big data favoreceu muito. Principalmente a definição da amostra. Antigamente, antes deste tratamento de dados acumulados, antes do big data, como é que você fazia uma prova estatística? Você tinha que contratar pesquisadores, coletores de dados. Instruí-los. Dar uma pranchetinha para cada um e colocar eles na rua.*

*E7 (falando de volume / velocidade / variedade): O que a gente hoje está tentando fazer é construir modelos que não sejam só tácitos. Que se consiga explicitar isto numa planilha, num gráfico, ou até numa ferramenta de software. Então eu acho que essa é a grande diferença hoje, a intuição sempre existiu, sempre deu resultado, mas hoje talvez a gente consiga ter ferramentas que nos ajudem a entender como é que agente está tomando as decisões baseadas em intuição.*

*E6 (falando de variedade): Quanto mais informação eu te dou sobre um determinado espectro, maior viscosidade esta informação tem, logo ela é mais relevante. Então a viscosidade a capilaridade, quando eu digo capilaridade é a variedade de informação, isso eu acho que ajuda muito o analista. Poder usar muitas fontes de informação, poder usar muitas origens, para o mesmo espectro, de informação e poder fazer seus próprios cruzamentos.*

De acordo com o grupo Profissionais, as grandes estruturas de dados atuam nesta categoria facilitando a associação e interpretação do comportamento de variáveis. A questão volume tem seus efeitos positivos também na construção de uma mentalidade mais lógica no tratamento dos dados, com o desenvolvimento de um profissional com características multidisciplinares. A apresentação das análises baseadas num número maior de dados, de fontes variadas, contribui para categoria ao proporcionar para os clientes da inteligência diferentes visões acerca de um determinado negócio, com base em dados não em percepções do profissional envolvido. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E1 (falando do volume): Estas grandes estruturas de dados facilitam este processo cognitivo e esta associação na busca do entendimento e interpretação do comportamento das variáveis.*

*E8 (falando do volume): Hoje tem um administrador, um estatístico, então o pensamento é mais lógico. Então é mais fácil de trabalhar até com grandes volumes de dados. Eu venho da área de TI, depois fiz administração, então já tenho esse pensamento. Já fica mais fácil. Um ano e meio atrás, não tinha. Vinha um compilado de dados, informações e eu tinha de transformar em conhecimento. Hoje não, agente já consegue trabalhar e ter uma recomendação.*

*E13 (falando do volume/variedade): Acho que até tem ajudado a ser mais assertivo. Eu tenho meios de buscar as informações, buscar dados, isso mesmo que não se estruture de uma forma tão sistemática. No momento em que tu tem a capacidade de buscar um volume maior de dados e apresentar isso para as pessoas, vai afetar a percepção que elas tem do negócio. Para o nosso caso a variedade e volume.*

#### 4.1.5.14 Volume/Velocidade/Variedade prejudicam a consciência das próprias potencialidades e outras influências cognitivas

A categoria "consciência das próprias potencialidades e outras influências cognitivas" é colocada pelo grupo Especialistas como um dos principais desafios para o profissional em inteligência estratégica. Isto porque o profissional pode, muitas vezes, ficar com sua atenção voltada para a questão da triagem da informação, em detrimento do processo de análise em si. Os entrevistados ressaltam os limites da cognição humana, no que tange à priorização do que será considerado para análise, partindo-se do pressuposto de que a responsabilidade do processo não pode ser atribuída à máquinas. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E18 (falando de volume): Há uma necessidade muito grande de recuperação da informação em volumes cada vez maiores. Apesar de todas as ferramentas tecnológicas que nós temos robôs, softwares, data mining, datawarehouse, enfim, todos os "datas" possíveis e imaginários, a nossa capacidade cognitiva é limitada. Então não adianta eu ter um volume de informação cada vez maior do que a capacidade de análise. Esta análise não adianta transferir para máquina, porque na última instância, nós é que vamos estar fazendo esta priorização, deveríamos estar fazendo esta priorização, ou ajuste fino neste volume de informação. Eu acho que sim, isso afeta bastante o trabalho do analista, acho que esse é um dos grandes desafios de como você vai lidar com este excesso de informação num mundo hiperinformacional.*

*E6 (falando de volume/variedade): O que acontece, para o processo cognitivo, a maneira como uma mente tem capacidade de interpretar e fazer inferências, o volume de informações e a variedade trouxe um impacto muito perverso nestas pessoas. Elas estão tendo que usar o seu processo de cognição, o seu processo de inferência e o seu processo de construção de relações para fazer triagem do que é relevante, e o que não é relevante. Do que ela considera informação, e do que ela deve desconsiderar como informação. Do que ela contempla em suas análises, e do que ela vai ter que*

*obrigatoriamente abandonar. Então, grande parte do pensamento analítico tem ser focado em triagem informacional e não o processo analítico verdadeiramente.*

Quanto o grupo Profissionais, os efeitos negativos do volume são interpretados pela autora como a super especialização dos envolvidos no processo de análise, em dados ou *softwares* específicos de utilização da empresa. Com isso os profissionais se fecham em seus métodos e estruturas e isto faz com que os efeitos positivos de *big data* não sejam explorados. Também foi citado que o profissional perde a tranquilidade para executar a construção de hipóteses na análise, em função dos grandes volumes de informações que mudam rapidamente. O profissional também não tem segurança para construir sozinho sua visão com base em dados e eventualmente divide esta responsabilidade com outros membros da equipe. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E14 (falando do volume): Iria demorar muito tempo para entender do todo. Então vamos trabalhar só com o que é realmente necessário e aí vamos nos aprofundar nisso. Aí agente viu o que era necessário, as pessoas foram para fornecedor, quando é o caso, ou se especializaram internamente nos bancos de dados e nos tipos de software que agente tem. Só para tirar aqueles dados, viraram especialistas naquilo, bem focados.*

*E10 (falando do volume / velocidade / variedade): Aquela segurança que tu tinha, talvez por ter menos informações e tu conseguir montar teu quebra-cabeças com mais tranquilidade, um tempo maior de análise, agora tu tem um tempo mais curto, mais informações. Tu precisa ter uma velocidade muito maior de análise e de concentração para poder traçar um cenário, traçar uma hipótese assertiva, então com certeza os 3Vs afetaram nosso trabalho.*

#### 4.1.5.15 Volume/Velocidade/Variedade não afetam a consciência das próprias potencialidades e outras influências cognitivas

De acordo com o grupo Especialistas, *big data* não afeta a categoria consciência das próprias potencialidades e outras influências cognitivas, no ponto em que questões de maturidade e confiança são adquiridas ao longo do tempo e habilitam o profissional a trabalhar com bom desempenho, independe de um cenário *small data* ou *big data*. Neste contexto, a questão experiência é o ponto chave, tirando o foco de *big data* e colocando o profissional no centro do processo de análise, independente do ambiente. O trecho a seguir confirma a análise realizada:

*E5 (falando de volume / velocidade / variedade): Não é devido ao big data. É devido a maturidade, a confiança que você passa a ter em você mesmo. Eu contrato hoje um analista júnior e treino, daqui um ano eu posso colocar ele dentro de uma outra empresa, de uma empresa para trabalhar e ele vai ter um desempenho muito bom,*

*porque ele adquiriu confiança, para usar com um mundo big data. Não é só o mundo big data, que se não existisse faria com que ele fizesse melhor ou pior. Vai fazer sim, com que ele seja mais rápido, vai fazer com que ele seja mais antenado é isso que o big data fez. Mas hoje se eu pensar em mim. Eu sou melhor por uma questão de experiência também.*

Na visão do grupo Profissionais, o volume de dados não interfere nesta categoria, pois não é percebido por eles. A questão da experiência também é mencionada como fator que permite a influência do *big data* no processo cognitivo. Assim, se o profissional for experiente não sentirá os efeitos do volume, velocidade e variedade, e poderá se diferenciar de seus concorrentes, com análises mais assertivas. Outros entrevistados do mesmo grupo não percebem a influência nem a importância de *big data* no seu setor de atuação. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E11 (falando do volume): Eu não parto da premissa que tem um grande volume de dados é essa a discussão. Para mim o volume de dados é proporcional ao tamanho do planeta. Eu acho que o volume de dados que você está colocando, isto não vai interferir nos aspectos cognitivos dos analistas.*

*E9 (falando do volume / velocidade / variedade): Hoje a experiência te traz também algumas vantagens competitivas com relação ao teu concorrente, que de repente tem um processo novo e pessoas novas trabalhando ali, do que ter um processo já maturado e pessoas que já conhecem o mercado, já conhecem os clientes. Então o processo cognitivo está muito ligado à experiência que a pessoa tem em determinada área.*

*E13 (volume/velocidade/variedade): Mas eu realmente acredito que talvez num futuro, para nós não tão próximo, porque eu não vejo esta necessidade tão urgente no nosso negócio, mas eu acho que em outras áreas de atuação isso aí pode ser muito significativo.*

#### 4.1.5.16 Volume/Velocidade/Variedade contribuem para a mentalidade aberta a opiniões contrárias ou modelos alternativos

Não foram encontradas evidências desta categoria para o grupo Especialistas. Para o grupo Profissionais, a contribuição do fenômeno para a mentalidade aberta a opiniões contrárias ou modelos alternativos, se dá com a facilidade na realização de cursos, treinamentos on-line, que podem contribuir para o desenvolvimento do profissional. O trecho a seguir confirma a análise realizada:

*E8 (falando do volume / velocidade / variedade): Hoje tu tem vários locais, tem acesso a informações, métodos. A gente tem muito coisa gratuita, como os TEDs, então tem muita informação. Principalmente o que sai de tendências na área de tecnologia, na área educacional, tem que estar sempre acompanhando.*

#### 4.1.5.17 Volume/Velocidade/Varietade prejudicam a mentalidade aberta a opiniões contrárias ou modelos alternativos

Esta categoria foi gerada a partir da interpretação do trecho de entrevista, no qual é colocado, que os analistas continuam utilizando os mesmos antigos métodos de análise, como por exemplo, SWOTs, modelos de excel e macros. Ou seja, não se abriram para a utilização das novas tecnologias no trato de grandes volumes de dados, continuando com a mesma mentalidade de *small data*. A colocação também serviu de embasamento para a categoria "volume/velocidade/variedade prejudicam a compreensão de métodos de investigação".

Para o grupo Profissionais, a citação que foi interpretada como relacionada a esta categoria foi aquela que se refere a super especialização dos profissionais, já citada anteriormente em "volume/velocidade/variedade prejudicam a consciência das próprias potencialidades e outras influencias cognitivas". Outro ponto citado por pelo grupo Profissionais, está na desconfiança a respeito da veracidade das informações, que em função da variedade, acaba fazendo com que profissionais restrinjam seu escopo de análise. O trecho a seguir confirma a análise realizada:

*E14 (falando da variedade): Tem o escopo fechado só do que a gente deveria realmente entregar. O fornecedor esta me dizendo que o mercado está desaquecido. Eu não acredito nisso, porque eu estou vendendo muito bem. Eu vou atrás de outra informação para descobrir se isso é certo. Eu vou procurar outro fornecedor, vou procurar em site, vou procurar outra informação.*

#### 4.1.5.18 Volume/Velocidade/Varietade não afetam a mentalidade aberta a opiniões contrárias ou modelos alternativos

Na visão do grupo Especialistas, esta categoria leva em consideração mais questões de comportamento do próprio profissional, aliadas a abertura de mercados, focos de atuação e a disponibilidade da tecnologia de informação e comunicações para definir o quão aberta será a mentalidade no processo de análise. Também é mencionada a questão da colaboração, com a formação de grupos para análise, como prática utilizada atualmente em algumas organizações. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E19 (falando de volume / velocidade / variedade): Também é muito característica das pessoas. Você tem o sujeito que tem um caráter, ou um lado psicológico mais autoritário... Então o profissional de inteligência ele tem que ter essa visão hoje, ele jamais pode colocar na cabeça que ele é o detentor de tudo. Porque as informações*

*estão espalhadas por aí, o que ele tem os outros também tem de acesso. O que ele tem de diferente, talvez seja o tempo que é dedicado a isso e os outros não.*

**E21** *(falando de volume / velocidade / variedade): Tem mercados que são muito complicados. Vai depender muito do mercado de atuação. Você tem mercados que são mais conservadores, aí tem a própria maneira de como o mercado foi concebido, então o analista ele tende a ser mais conservador. E tem mercados que são mais abertos. Mais abertos para a colaboração.*

**E6** *(falando de volume / velocidade / variedade): Acho que o big data não tem nenhuma influência nisso. Primeiro é a disponibilidade de tecnologia de informação e comunicação, que preconizam esta facilidade de contato. Segundo são os processos de abertura das próprias empresas. Antigamente se você falasse assim "pesquisa e desenvolvimento", isso era um castelo dentro da empresa. As portas eram até fechadas, hoje você até escuta falar em open innovation. Até os processos de inovação e pesquisa são abertos. Terceiro, com o mecanismo e o advento da evolução da sociedade da informação e do conhecimento a colaboração não é mais um problema. Certamente tinham a noção de que informação era poder.*

Em outra entrevista é mencionada a questão de que mentalidade aberta não seria uma atividade relevante para inteligência, em específico. Ela é importante de modo geral. Neste sentido, o diferencial estaria concentrado no oposto, precisão e rigor que o profissional irá assumir frente ao processo de análise. Este último trecho se apresenta como um contraponto à base teórica considerada neste estudo:

**E4** *(falando de volume / velocidade / variedade): Ele tem que ter uma mentalidade mais precisa. Aberta sempre ajuda, mas ele tem que ter mais precisão, mais rigor. O analista precisa ter uma mente mais precisa.*

As entrevistas do grupo Profissionais compartilha das visões do grupo Especialistas com relação a esta categoria, quando mencionam a questão da colaboração, da utilização de diferentes pessoas, com diferentes habilidades na organização para melhorar o produto da inteligência. Também foi mencionado que o próprio ciclo de inteligência carrega em si aspectos relacionados à melhoria dos modelos utilizados. Assim, o alinhamento dos produtos de inteligência às necessidades dos clientes representa uma oportunidade para a organização e requer uma visão aberta para a utilização de novos métodos. Isto reflete a imparcialidade do profissional frente às variadas formas de como gerar inteligência. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

**E3** *(falando do volume / velocidade / variedade): Eu valorizo bastante a questão do contraponto. Tem reuniões que a gente senta para discutir ações de um mês todo e que realmente é baseado em dados, em fatos, em tendências, em comportamento, em sazonalidade. E aí senta gerente de compras, gerente de comunicação, pessoal do planejamento para definir uma oferta. Há discussões e aí a gente vai até chegar ao consenso, até se tomar uma decisão corporativa.*



**E10** (falando do volume / velocidade / variedade): *Eu acho que o processo de inteligência, o fluxo mesmo, ele te pede fazer para sempre fazer uma revisão. Esse cuidado de como eles estão recebendo, de ter este feedback. Não simplesmente informar, mas pedir se aquelas informações que agente está passando nas análises estão sendo suficientes, se são importantes, eu acho que o processo mesmo exige isso.*

**E19** (falando do volume / velocidade / variedade): *Estou cada vez mais aberto. Estas coisas que eu desenvolvi ao longo do tempo, hoje elas atendem à demanda básica que a empresa e as pessoas precisam. Eu particularmente tenho buscado formas alternativas. Tenho me dedicado a estudar bastante, isso se baseia muito em se colocar na situação do cliente, se colocar como o cliente. Maneiras bem diversas e alternativas de fazer a coleta dos dados e a análise dos dados. (Big data tem a ver com essa tua posição?) Não. Neste momento não. Parte de uma percepção minha de que o negócio da nossa empresa tem que estar alinhado com a percepção, o desejo o sentimento do cliente.*

#### 4.1.5.19 Volume/Velocidade/Variedade contribuem na autoconfiança para admitir e aprender com os erros analíticos

Na visão do grupo Especialistas, volume, velocidade e variedade proporcionam um maior embasamento para as análises e isto reflete numa maior autoconfiança por parte do profissional. Esta visão é compartilhada pelo grupo Profissionais, que complementam que o volume possibilita uma maior precisão na realização dos testes de hipóteses. Eles colocam que no passado, este processo era comprometido em função da escassez de dados. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

**E5** (falando de volume / velocidade / variedade): *Então equívoco, tem milhões de motivos para isso. Motivo de que eu usei uma outra informação que eu tive, ou eu usei a informação do departamento financeiro. Eu acho que esta questão do big data vai ajudar para que as análises sejam feitas com mais embasamento. E daí talvez as pessoas fiquem com mais confiança.*

**E9** (falando do volume): *A única coisa que a gente tem certeza, lá na frente, é que a gente vai acertar ou errar. Então quanto mais dados tu tiver para embasar a tua decisão, é melhor.*

**E20** (falando do volume): *Uma coisa que dificultou muito, isso ocorreu com 6, 7 anos atrás, é pelo baixo volume de informações disponíveis. Isso limitou bastante, acho que a assertividade, as informações necessárias para fazer os testes de hipóteses necessários. Então tu conseguiria, com um volume de informações maior, mais confiáveis, fazer talvez recomendações diferentes do que a gente fez no passado e com mais precisão.*

#### 4.1.5.20 Volume/Velocidade/Variedade prejudicam a autoconfiança para admitir e aprender com os erros analíticos

A dimensão volume prejudica a autoconfiança para admitir e aprender com os erros analíticos, pois de acordo com o grupo Especialistas, as ferramentas tecnológicas disponíveis atualmente não suprem as necessidades do processo de análise, comprometendo o trabalho dos profissionais em IE e diminuindo a sua autoconfiança. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E7 (falando de volume): Eu acho que está mais perdido mesmo. Acho que as ferramentas que hoje são muito comuns, elas não ajudam muito no big data.*

*E18 (falando de volume): Então o resultado deste trabalho do analista é muito pior do que era anteriormente, quando tinha um volume muito menor. Então a possibilidade num volume menor de você recuperar informação era maior. Mesmo se as tecnologias não fossem tão avançadas.*

Para o grupo Profissionais, o grande volume aumenta a propensão a erros de análise. O volume e a variedade comprometem a questão da veracidade dos dados e informações e o próprio desempenho do profissional. Os efeitos prejudiciais da dimensão velocidade também são percebidos em alguns setores específicos, onde o profissional não consegue acompanhar a velocidade das mudanças no foco da análise. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E3 (falando do volume): Evidente que o volume, ele te deixa mais suscetível. Ele acaba sendo uma armadilha, o volume de informações, de dados.*

*E2 (falando do volume/variedade): Acho que o volume de dados atrapalha, também às vezes você confia em alguns dados que talvez não sejam tão robustos, qualidade dos dados. Às vezes, a gente tem que cuidar um pouquinho com a informação que está disponível na internet, nos meios de comunicação, dentro do que está ali está a verdade.*

*E10 (falando da velocidade): Algumas situações, a gente não consegue prever, muda tudo muito rápido. Sei que algumas empresas têm até agrônomos que trabalham na análise, então algumas coisas a gente não vai conseguir monitorar 100% e essas condições climáticas, previsões, são difíceis de acertar. A gente não sabe se vai dar uma seca, se não vai, e para nós a parte agrícola afeta muito.*

#### 4.1.4.21 Volume/Velocidade/Variedade não afetam a autoconfiança para admitir e aprender com os erros analíticos

Com relação à categoria autoconfiança para admitir e aprender com os erros analíticos, a falta de uma diretriz do tomador de decisão sobre o que deve ser pesquisado é um ponto que afeta os profissionais, na visão do grupo dos Especialistas. Segundo eles, também as questões relacionadas ao tempo disponível para realizar a atividade de IE estariam comprometendo a

autoconfiança. Outro ponto relevante mencionado pelo grupo Especialistas, refere-se à ausência de uma nova modelagem cognitiva e evolução do processo de tomada de decisão. Este ponto faz com que os produtos gerados pela análise - conhecimentos, não sejam efetivados em inteligência para as organizações. E por fim, a autoconfiança é colocada como uma característica vinculada a própria função do profissional, quanto maior for a sua responsabilidade, mais preocupado com o nível de acertos no processo este profissional estará e será cobrado por isso. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E18 (falando de volume): Eu me sinto mais inseguro, em função a insegurança do tomador de decisão, não pelo volume. Mas pelo tempo sim, porque cada vez ele é menor, o tempo necessário para você estar recuperando e disseminar esta informação, isso impacta numa pressão cada vez maior com relação aos analistas de ser. Mas principalmente a definição do que é necessário. Então como eu não tenho a clareza desta definição, naturalmente toda a pesquisa fica prejudicada.*

*E6 (falando de volume / velocidade / variedade): Aí eu volto lá no início da nossa conversa, os modelos analíticos, os modelos cognitivos e o processo de tomada de decisão não sofreram alteração no último quinquênio, não adianta, nós estamos há 50 anos com os mesmos processos decisórios. Nós estamos há 60 anos sem a publicação de uma nova modelagem cognitiva, para o nosso cérebro. Nós estamos há 210 anos estudando a diferença entre o hemisfério direito e esquerdo. Então, não é a questão informacional que vai endereçar isso, tem outras variáveis que se isoladas podem te levar a uma conclusão equivocada. 11 de Setembro: alguém tinha dúvida de que aconteceria? Nenhum americano hoje tinha dúvida, mas a CIA já não tinha dúvida. O que ela fez com esta informação? Nada. Então não adianta maior velocidade, maior volume, maior variedade se não há ação humana, da empresa ou da nação.*

*E19 (falando de volume / velocidade / variedade): Isso é uma característica muito pessoal. A medida que você vai subindo no conceito da organização, vai alcançando um determinado patamar você sente mais vulnerável a erros e falhas. As próprias cobranças começam a aumentar. Ele está na posição X, então tudo o que fala tem que estar correto. Então você começa a se cobrar muito e os erros começam a ser apontados com mais contundência.*

As entrevistas do grupo Profissionais compartilham com o grupo Especialistas a questão da falta de uma diretriz sobre o que deve ser pesquisado. Outros de ambos os grupos atribuem a questão da autoconfiança ao próprio profissional e não conseguem ver os efeitos do volume, velocidade e variedade nesta atividade. Outros ainda atribuem a autoconfiança ao contexto da organização, ou seja, se a empresa está em uma boa condição financeira, será mais tolerante a erros. Se a empresa está em uma condição financeira ruim, o profissional não poderá errar, e isto irá abalar a sua autoconfiança. Um ponto colocado que contribui para esta atividade, novamente, é a questão da tomada de decisão em conjunto. A seguir são apresentados trechos das entrevistas que confirmam a análise realizada:

*E1 (falando do volume / velocidade / variedade): Na questão da prospectiva, a gente tem de tomar cuidado, porque às vezes você está construindo um futuro desejável e isto pode ser frustrante. Você constrói um futuro desejável, e aí você pensa ações que viabilizem aquele futuro. No ponto de vista de utilização de dados, invariável acontece, na equipe mesmo, e na maioria das vezes pela falta de compreensão do que está sendo pedido.*

*E11 (falando do volume / velocidade / variedade): Eu não consigo ver uma mudança no ambiente em quantidade e em qualidade de informações que vai afetar diretamente, somente esse tipo de profissional. Não consigo fazer esta relação. É do indivíduo e não do cargo, da função.*

*E15 (falando do volume / velocidade / variedade): Abala (autoconfiança), mas amenizou porque foi uma ação tomada em conjunto. Como eu te disse, eu sou geradora de informação, mas a decisão não era minha.*

#### 4.1.5.22 Outras dimensões de *big data*

Durante as entrevistas, outras dimensões para *big data* também foram citadas. Entre elas está a questão da veracidade, com um maior número de citações, seguida de complexidade, viralidade e viscosidade. As duas primeiras citadas, tem fundamentação teórica mencionada neste estudo. A Figura 52 apresenta estes dados:

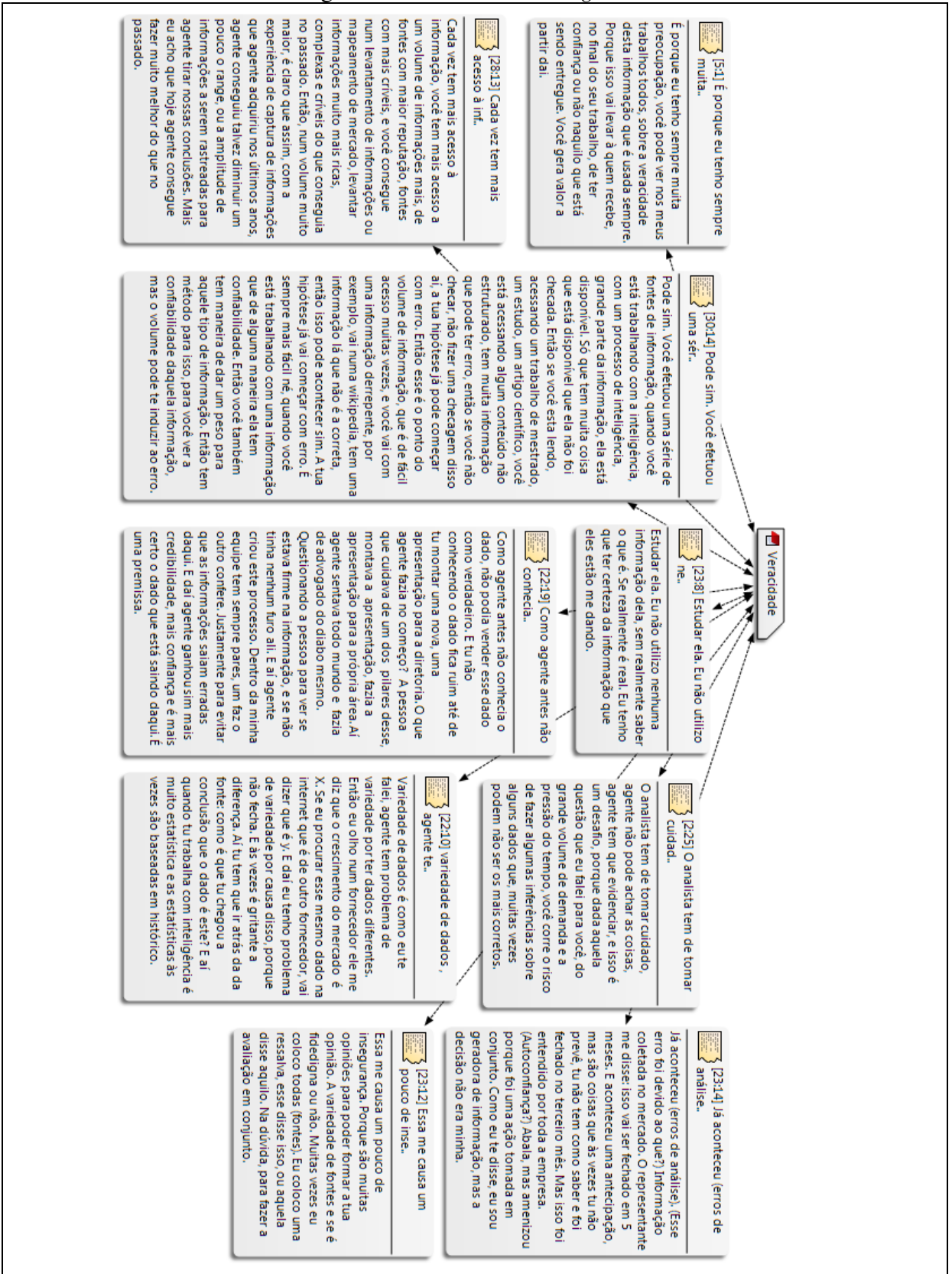
Figura 52 - Outras dimensões de *big data*.

Name	Type	Groun...	De...	Size	Author	Created	Modified	PDs	Families
Veracidade	Commentary	9	0	111	Super	12/01/20...	22/02/20...	-	1.9 Outras dimensões de BD
Complexidade	Commentary	3	0	0	Super	08/02/20...	23/02/20...	-	1.9 Outras dimensões de BD
Viralidade	Commentary	2	0	0	Super	08/02/20...	15/02/20...	-	1.9 Outras dimensões de BD
Viscosidade	Commentary	1	0	0	Super	08/02/2...	24/02/2...	-	1.9 Outras dimensões de BD

Fonte: elaborado pelo próprio autor.

A ênfase na questão da veracidade foi citada por alguns dos entrevistados e aparece como sendo um efeito do volume, variedade e velocidade. Isto é percebido em trechos que relatam que atualmente existe um maior volume de fontes mais críveis - volume/veracidade, ou que é importante verificar as fontes de informações, pois não podem haver erros na construção de hipóteses - variedade/veracidade. A Figura 53 apresenta trechos destas entrevistas:

Figura 53 - Veracidade de *big data*.



Fonte: elaborado pelo próprio autor.

A complexidade foi outra dimensão citada pelos entrevistados. Esta questão foi apresentada como sendo um aspecto problemático do século XXI, e neste contexto os métodos tradicionais para a resolução de problemas não teriam êxito, sendo necessária a utilização de metodologia científica. A seguir é apresentado trecho desta entrevista:

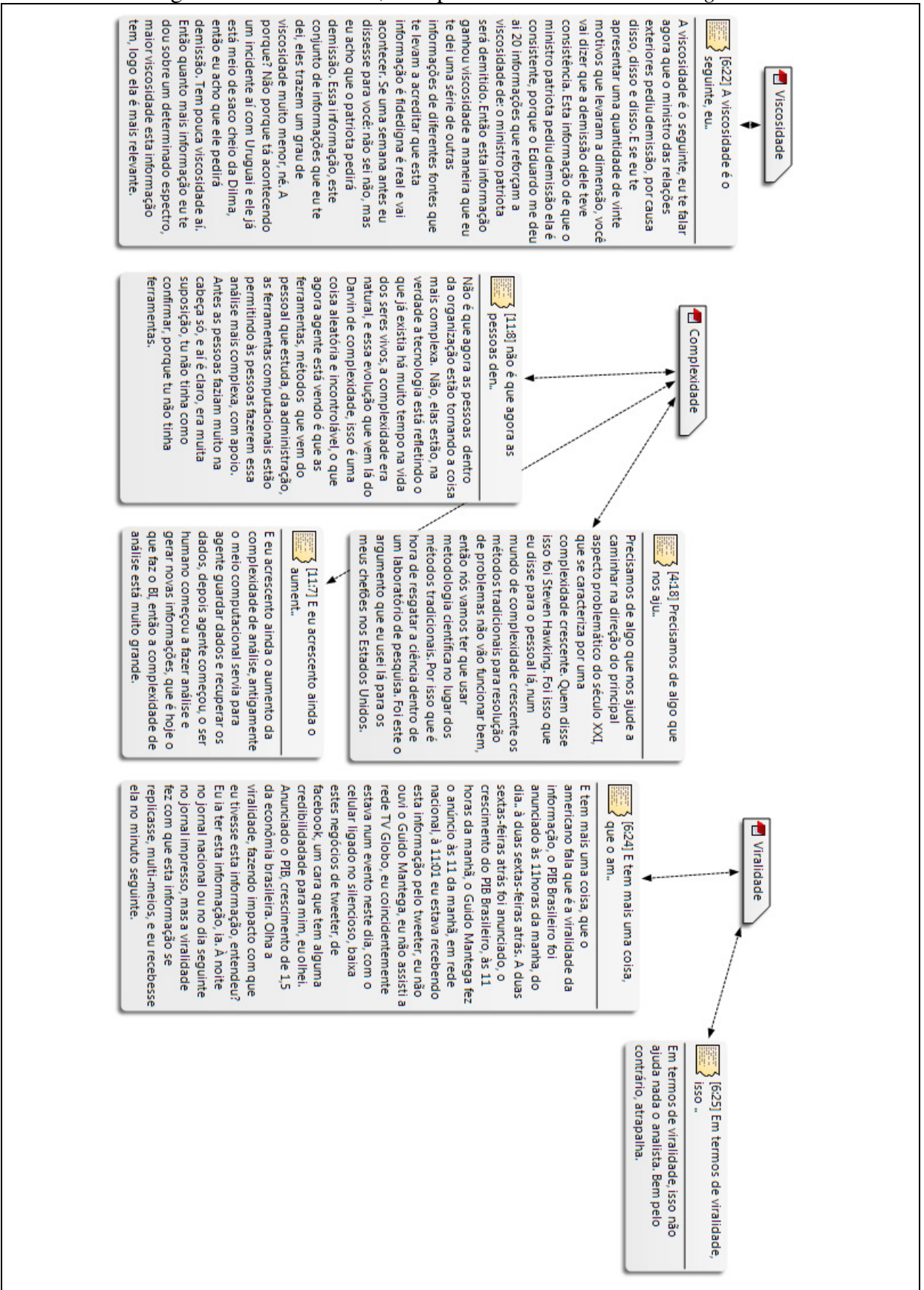
*E4 (falando da complexidade): Num mundo de complexidade crescente os métodos tradicionais para resolução de problemas não vão funcionar bem, então nós vamos ter que usar metodologia científica no lugar dos métodos tradicionais. Por isso que é hora de resgatar a ciência dentro de um laboratório de pesquisa.*

Relacionado com este trecho, outro aspecto citado remete a ausência de um novo modelo cognitivo que possa estar adequado a atual complexidade. Na visão do entrevistado, *big data* ainda não é uma realidade para as empresas porque os métodos de investigação e tomada de decisão não evoluíram. Os modelos analíticos e cognitivos não sofreram alterações no último quinquênio, esta é a dificuldade em estar utilizando o *big data* como geração de valor para as organizações.

*E6 (falando do método): Você ainda tem um conjunto de executivos, de pesquisadores, de consultores que constroem suas hipóteses, formulam suas hipóteses, definem suas estratégias, e depois eles vão em analistas de informação, analistas de mercado, estrategistas e informam aquelas hipóteses para estes caras testarem. Estes caras pegam grandes volumes de dados, massageiam estes grandes volumes de dados, tratam estes grandes volumes de dados e negam ou afirmam a hipótese do executivo, a hipótese do gestor, a hipótese do diretor, a hipótese de um líder de uma nação. Então o que acontece, você tem um poderio imenso, que é o poderio da extração de conhecimento de grandes volumes de dados e informação que não é utilizado.*

Outros aspectos como complexidade viscosidade e viralidade também foram citados. A viscosidade está relacionada à dimensão variedade e a viralidade à dimensão velocidade. De acordo com o entrevistado, viscosidade é colocada como o conjunto de informações complementares, que reforçam uma informação inicial. A viralidade está relacionada ao número de repetidores, empresas ou pessoas que replicam uma mesma informação original e fazem com que esta informação circule, chegando a um número maior de pontos. A Figura 54 apresenta estas dimensões:

Figura 54 - Viscosidade, Complexidade e Viralidade de big data.



Fonte: elaborado pelo próprio autor.

As dimensões complementares de *big data*, identificadas nas entrevistas não foram examinadas com relação às implicações na análise para IE, por se tratarem apenas de percepções dos entrevistados. Embora veracidade e complexidade tenham sido citadas na fundamentação teórica, não são estas as dimensões que mais se repetem na literatura.

Desta forma, conforme colocado na delimitação do tema, apenas as dimensões volume, velocidade e variedade serão tratadas neste estudo. Assim, o resumo proposto para este capítulo, no que tange aos efeitos do fenômeno *big data* no processo de análise para inteligência estratégica está apresentado na Figura 55:

Figura 55 - Efeitos do fenômeno *big data*.



Fonte: elaborado pelo próprio autor.

A Figura 55 mostra que os entrevistados, por meio de setenta e uma citações sintetizadas, percebem as contribuições do fenômeno *big data* para atividade analítica pela possibilidade de utilização de múltiplas perspectivas acerca de um tema e pela possibilidade de cruzamento de dados. Isto desde que, se tenham as ferramentas tecnológicas adequadas e profissionais com mentalidade mais lógica para o tratamento dos dados.

De acordo com os entrevistados, em noventa citações que foram sintetizadas na Figura 55, o fenômeno *big data* é prejudicial para a atividade analítica em função da baixa qualidade das informações e também em função da velocidade com que as informações são geradas, o que faz com que as informações coletadas tenham que ser revisadas, mesmo antes da



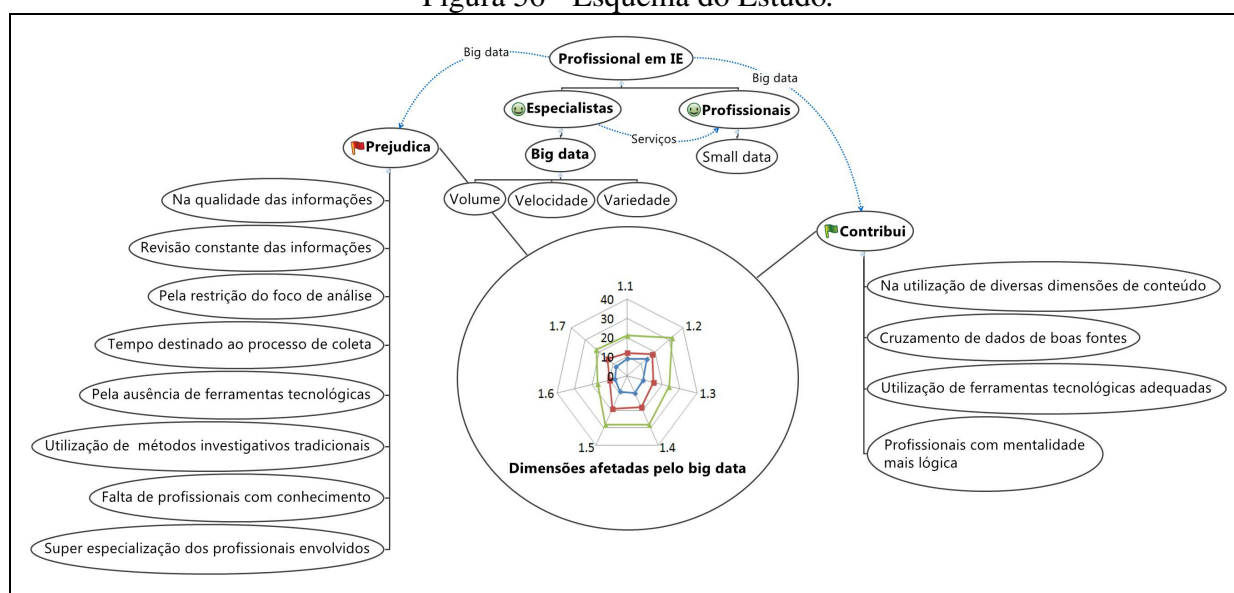
implementação de ações. Outro efeito decorrente de *big data* esta na restrição do foco de análise, inclusive para que o tempo que o profissional leva para coletar as informações não seja maior do que o tempo que ele leva para realizar a análise. Neste ponto, a ausência de ferramentas tecnológicas apropriadas e a utilização de métodos tradicionais de análise também são elementos fazem com que os 3 Vs de *big data* prejudiquem o processo de análise. Com relação aos aspectos humanos, foi citada a questão da falta de profissionais capacitados para lidar com o grande volume de dados e a super especialização de profissionais, como efeitos do fenômeno.

De acordo com os entrevistados, em cento e vinte e nove citações que foram sintetizadas na Figura 55, *big data* não afeta a atividade analítica quando há um processo de governança na organização e/ou quando existe uma questão e uma estrutura de informações já pré-determinada, ou seja, mecanismos que restringem a livre atuação do profissional. Neste caso o profissional não utiliza grandes volumes de dados para gerar suas análises, atuando muitas vezes, apenas com informações internas da organização. Outro ponto levantado foi que não é *big data* que tem suas implicações no processo de análise e sim o próprio profissional que está conduzindo o processo e suas capacidades.

## 4.2 DISCUSSÃO

A questão de pesquisa deste trabalho foi apresentada no sentido de examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica. As 18 entrevistas válidas, consideradas na análise de conteúdo trouxeram como primeira constatação, o fato de que o fenômeno *big data* ainda é algo distante das organizações tradicionais. Portanto pode-se dizer que com dados originados da pesquisa e de acordo com as categorias analisadas, em duzentas e dezenove citações *big data* não afeta o processo de análise, ou a prejudica. Assim, setenta e uma citações atestam que *big data* contribui para o processo de IE. A Figura 56 apresenta o Esquema do Estudo:

Figura 56 - Esquema do Estudo.



Fonte: elaborado pelo próprio autor.

Com base nas citações dos entrevistados, *big data* não afeta, não porque seja algo que não exista ou que não traga benefícios, mas porque é muito novo e ainda não está sendo utilizado e nem visto como necessário. Na pesquisa às bases de dados foi possível identificar que nenhum dos artigos publicados em *big data* e análise leva o contexto de inteligência estratégica em seu escopo. Este é um paradoxo que se apresenta neste estudo, uma vez que a maioria dos entrevistados do grupo Especialistas possui nível de formação de pós-graduação (mestrado e doutorado), desenvolve pesquisas sobre assuntos relacionados à gestão das informações e inteligência estratégica e está vinculado a instituições de ensino e pesquisa.

Os entrevistados deste grupo são atores detém o valor de *big data*. Conforme proposto por Schönberger-Mayer e Cukier (2013), estes atores são: detentores de dados, especialistas em dados e empresas e pessoas com mentalidade de *big data*. Detentores de dados são os atores que controlam o acesso à informação e a usam ou licenciam para que outros possam extrair valor dela; especialistas em dados são as empresas com o conhecimento ou a tecnologia para fazer análises complexas; e, empresas e pessoas com mentalidade de *big data* são aquelas que veem as oportunidades antes dos outros, mesmo que não possuam dados ou habilidades para agir. Neste sentido, estes atores estão colhendo os benefícios da utilização de *big data*, em suas empresas de pesquisa e consultoria, prestando serviços aos demais profissionais das áreas.

Fora do contexto de *big data*, os produtos de inteligência gerados por analistas competem fortemente com a análise baseada em intuição, fazendo com que o ciclo de

inteligência não seja completado (FULD, 1995). Para o autor as decisões de negócios são baseadas em uma combinação de experiência, bom senso e inteligência, sendo a última a mais importante para o processo. Schönberger-Mayer e Cukier (2013) esclarecem que ao longo de boa parte da história coletavam-se poucos dados porque os instrumentos para coletar, organizar, armazenar e analisar eram ruins. As informações eram peneiradas ao mínimo, para que fosse possível analisar com mais facilidade. De acordo com as entrevistas realizadas, atualmente os entrevistados do grupo Profissionais utilizam estes mesmos recursos, de reduzir as fontes e volume de dados coletados, como estratégia para realizar as análises de forma mais efetiva.

Entretanto, isso é a evidência de que mesmo não havendo mais um limite artificial tecnológico, ainda não se tem a mentalidade voltada para trabalhar com *big data* (SCHÖNBERGER-MAYER E CUKIER, 2013). Assim, dotados da mesma mentalidade e dos mesmos métodos, os profissionais em inteligência ainda não conseguiram concretizar todo o potencial anunciado pelo fenômeno. Em uma das entrevistas realizadas foi dito que os métodos de investigação e tomada de decisão, são métodos que não evoluíram, pois as hipóteses são construídas, formuladas, estratégias são definidas e depois os analistas de informação, analistas de mercado, estrategistas as testam. Assim, grandes volumes de dados são manipulados no sentido de negar ou afirmar a hipótese previamente construída. Então o imenso poderio da extração de conhecimento de grandes volumes de dados e informação não é utilizado.

De acordo com esta mesma entrevista, em um ambiente de *big data*, este processo partiria de um problema de pesquisa, onde as informações envolvidas seriam manipuladas, sem nenhuma premissa *a priori*. Então, a empresa começaria a identificar alguns padrões e entender ou conhecer fenômenos até então desconhecidos. Com isso, quando se lida com grandes volumes de dados, com a utilização de métodos preditivos e prescritivos, com dados incertos e correlação é necessário que se tenham profissionais habilitados à forte utilização de matemática, de estatística, com um viés multidisciplinar apoiado em negócios, tecnologia, comportamento econômico e antropologia social, para que seja possível interpretar dados (MINELI, CHAMBERS E DHIRAJ, 2013).

Os autores colocam que com *big data*, o que está se apresentando no futuro é a aplicação da disciplina de negócios, combinada com matemática, tecnologia e, ciências comportamentais. As ciências do comportamento irão permitir conectar os pontos entre as interações e desenvolver uma compreensão mais profunda do comportamento humano. Isto

permite criar os incentivos adequados para conduzir comportamentos que se alinham aos objetivos de negócios. Mas isto é algo que ainda se restringe à teoria ou a poucas empresas especializadas, que conforme as entrevistas, citaram utilizar técnicas de análise de sentimento, por exemplo. Desta forma, o contato com empresas que fornecem pesquisas prontas é um ponto importante levantado na pesquisa, que momentaneamente atenderia às demandas fora da rotina de inteligência nas organizações, mas que talvez no longo prazo poderia vir a criar dependência.

Conforme abordado no referencial teórico deste estudo, os aspectos tecnológicos de *big data* são colocados como a parte fácil do processo. A parte mais difícil é colocada como descobrir para onde ir com a análise de dados e certificar-se de que a empresa tem as pessoas e os processos prontos antes de se comprometer a comprar a tecnologia (MINELI, CHAMBERS E DHIRAJ, 2013). Fleisher e Bensoussan (2007) compartilham deste pensamento ao declararem que fazer melhores escolhas e tomar melhores decisões pela primeira vez, gera uma maior necessidade de análise eficaz e inteligência. Os gestores de hoje enfrentam uma grande quantidade de informações em seus contextos de tomada de decisão, e às vezes essa abundância de informação faz com eles sejam paralisados.

Neste contexto está a importância das dimensões cognitivas propostas por Fachinelli, Rech e Bertolini (2011), a partir dos trabalhos de Heuer (1999) e Bruce e George (2008) abordadas neste estudo e pela primeira vez testadas em campo: 1) domínio da matéria; 2) compreensão de métodos de investigação, 3) imaginação e rigor científico; 4) compreensão dos métodos de coleta; 5) consciência das próprias potencialidades e outras influências cognitivas; 6) mentalidade aberta a opiniões contrárias ou modelos alternativos e, 7) autoconfiança para admitir e aprender com os erros analíticos.

Assim, a descoberta das características do profissional interpretadas como interface entre *big data* e atividade analítica, tomam a frente da discussão, refletindo o cenário de escassez de talento analítico e gerencial para transformar o conhecimento em melhoria do desempenho organizacional. De acordo com Manyika et al. (2011), os Estados Unidos sozinhos deverão enfrentar até o ano de 2018, uma escassez de 140.000 a 190.000 pessoas com habilidades analíticas. Isso representa um *gap* de 50% a 60% na demanda. Assim, a falta de 1,5 milhões de gerentes e analistas para analisar os grandes volumes tende a se manter uma vez que os profissionais/empresas ainda não perceberam o desenvolvimento destas habilidades como um fator essencial em IE. Seria o profissional em inteligência levantado por

Fuld (1995) o futuro cientista de dados? Ou seria esta uma outra função dentro da organização?

De acordo com Mineli, Chambers e Dhiraj (2013), nem todos os analistas de dados estão totalmente confortáveis com a ideia de serem chamados de cientistas de dados. Entretanto, há um sentimento emergente e de comum acordo de que, para ser verdadeiramente eficaz, um analista de dados deve ter principalmente a forte capacidade de análise. Então, se forem consideradas as duas hipóteses identificadas neste estudo que confirmam a influência de *big data*: 1) volume contribui para a compreensão dos métodos de coleta e, 2) volume prejudica o domínio da matéria, o caminho a ser traçado por estes profissionais deverá ser de revitalizar mentalidade e métodos para análise na atividade de IE. Neste sentido, Schönberger-Mayer e Cukier (2013) defendem que a supremacia dos especialistas diminuirá, e assim a dependência da intuição, ou experiência para orientar as decisões será substituída por métodos mais científicos e eficazes de análise.

## CONSIDERAÇÕES FINAIS

Esta dissertação apresentou um estudo sobre as implicações do fenômeno *big data* na análise para inteligência estratégica. O método utilizado foi a pesquisa qualitativa, que empregou um estudo exploratório, com técnica de análise de conteúdo. O conjunto metodológico foi desenvolvido a partir de 5 passos: 1) campo de estudo: onde por meio da técnica *snowball sampling* foram identificados 4 agentes com conhecimento e trânsito em comunidades relacionadas aos temas pesquisados, e a partir destes foram identificados 21 entrevistados, 2) categorias de análise: de acordo com a literatura especializada, foram identificadas categorias para o fenômeno *big data* e atividade analítica, 3) coleta dos dados: por meio de entrevistas presenciais e remotas com base em roteiro semiestruturado, 4) análise dos dados: com base na técnica de análise de conteúdo, considerando 18 entrevistas válidas, 5) discussão: com a interpretação dos resultados gerados.

A base teórica utilizada para o estudo, considerou o fenômeno *big data*, suas definições e características, bem como os aspectos tecnológicos que estão sendo desenvolvidos e empregados para se tratar com grandes volumes de dados. Este tema se apresentou dinâmico ao longo do processo de pesquisa, com muito conteúdo sendo gerado a partir de uma série de organismos, principalmente internacionais. Conforme dados apresentados ao longo do trabalho, pode-se perceber uma grande concentração de estudos de cunho comercial, com enfoque em tecnologia. A carência de estudos acadêmicos que tratassem do tema, bem como a sua aplicação nas ciências sociais justificaram o desenvolvimento deste trabalho.

Em meio a uma série de definições que estão sendo geradas para o fenômeno as dimensões consideradas neste estudo foram o volume, a velocidade e a variedade na geração de dados e conteúdo. Dentre estas dimensões, o volume é o mais perceptível em *big data*, pois pode ser observado por qualquer pessoa que tenha em suas mãos um *smartphone*, ou que acesse as redes sociais, ou utilize dispositivos GPS em seus automóveis. O volume de informações cresce exponencialmente, e sua origem está em cada operação que é realizada e armazenada em bancos de dados. Por isto, o desafio tecnológico intrínseco a esta dimensão é a questão da escalabilidade.

A velocidade consiste na rapidez com que os dados são criados, acumulados, internalizados e processados. Dar respostas em tempo real com paradigma de processamento de fluxo paralelo (*StreamProcessing*) é o desafio tecnológico vinculado a esta dimensão. A variedade trata da utilização de dados não estruturados, de diferentes tipos de fontes, como vídeos, imagens, e outros. Para esta dimensão, o desafio tecnológico que se apresenta é ter mecanismos que façam a ligação destas diversas classes diferentes em uma mesma estrutura externa.

Entretanto, os desafios que se apresentam para tratar com grandes volumes de dados não são apenas tecnológicos, mas também cognitivos e por meio da atividade analítica podem ser observados. Em função disto, a segunda parte da base teórica utilizada fala da inteligência estratégica, dos aspectos metodológicos e cognitivos que compõe o processo de análise e da interface entre os dois temas que é o profissional que trabalha com o processo análise de dados e informações para a atividade de inteligência estratégica.

A atividade de IE é antiga, entretanto a utilização dos processos de IE nos negócios é recente e suas implicações estão voltadas para a utilização do produto da inteligência para subsídio do processo de tomada de decisão nas organizações. Este produto consiste na informação analisada, interpretada e infundida com ações. Neste processo, a análise consiste na etapa onde o maior valor é agregado. Ela se divide em duas etapas, a fase física, onde o conhecimento é gerado e a fase mental, onde a inteligência é gerada.

Na fase física, aspectos metodológicos como análise estratégica, FAROUT e outras, auxiliam profissionais para a construção do conhecimento. Na fase mental, aspectos cognitivos passam a figurar e interferir na atividade analítica. Na literatura, estes aspectos foram traduzidos nas dimensões: 1) domínio da matéria; 2) compreensão de métodos de investigação, 3) imaginação e rigor científico; 4) compreensão dos métodos de coleta; 5) consciência das próprias potencialidades e outras influências cognitivas; 6) mentalidade aberta a opiniões contrárias ou modelos alternativos e, 7) autoconfiança para admitir e aprender com os erros analíticos. Estas dimensões foram tomadas como categorias conceituais para o presente estudo.

A terceira parte da base teórica deste estudo tratou das contribuições que *big data* pode trazer para a geração de valor, tanto no âmbito dos negócios, quanto no âmbito científico. Nos negócios, *big data* enriquece o processo de análise, proporcionando a utilização de técnicas de análises preditivas e prescritivas, permitindo a revelação de percepções antes ocultas e permitindo a criação de novos produtos. Na ciência *big data* oferece um novo olhar à pesquisa

científica pela possibilidade da utilização de dados de toda uma população, ao invés da utilização de uma amostra que eventualmente pode induzir a erros. Entretanto para que se possa obter o valor de *big data* é necessário que haja uma mudança de mentalidade por parte do profissional, que passará a trabalhar com grandes volumes de dados incertos, utilizando outras técnicas, como correlação.

Estes pressupostos teóricos foram importantes ao subsidiar o objetivo geral deste estudo que foi o de examinar como o fenômeno *big data* afeta o processo de análise na atividade de inteligência estratégica. Os resultados encontrados, por meio dos dados coletados nas entrevistas, trouxeram entre outras reflexões, que o fenômeno *big data* ainda não é uma realidade para os profissionais em IE de organizações tradicionais. Neste cenário, os atores que estão colhendo valor dos grandes volumes de dados são aqueles apontados Schönberger-Mayer e Cukier (2013): especialistas em dados, empresas com conhecimento ou a tecnologia para fazer análises complexas.

O primeiro objetivo específico deste estudo consistiu em identificar as principais etapas das práticas de análise dos profissionais de IE. Da análise das respostas foram originados os grupos de entrevistas: "Profissionais" e "Especialistas", que foram utilizados no decorrer da organização e análise dos dados. Posteriormente foram apresentadas as percepções da influência do *big data* na atividade de IE, segundo objetivo específico do estudo. Como resultados foram encontrados e descritos aproximadamente dez pontos de influência que vão desde a fragilidade na confidencialidade das informações, até uma maior facilidade no seu tratamento, entre outros. Estas reflexões foram confirmadas ao buscar o terceiro objetivo específico do estudo que foi o de identificar as dimensões relacionadas à atividade analítica que estão sendo afetadas pelo fenômeno *big data*.

Por fim, foram analisadas as interfaces entre *big data* e atividade analítica, que resultaram na hipótese de que o profissional em IE e suas características constituem esta área de interação. Este fato pode ser associado à constatação de Manyika et al. (2011) com relação a falta de 1,5 milhões de gerentes e analistas para analisar os grandes volumes de dados. Se considerarmos que o fenômeno *big data* ainda não é uma realidade para os profissionais em IE de organizações tradicionais participantes da presente pesquisa é possível supor que esta deficiência tende a aumentar já que o volume a variedade e a velocidade de dados são uma característica do *big data*. Assim, em função da complexidade do ambiente, profissionais em IE das organizações irão ficar dependentes das poucas empresas que possuem cientistas de dados habilitados a tratar com *big data*.



Estes objetivos deram origem às hipóteses construídas para este estudo exploratório, de acordo com as recomendações de Gil (2010): as pesquisas exploratórias proporcionam maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses. Assim, as categorias resultantes com maior frequência de citações percebem o volume, a velocidade e a variedade, como não exercendo efeitos na dimensão "mentalidade aberta a opiniões contrárias e modelos alternativos" (1.6.3/6/9), também o volume foi colocado como não exercendo efeitos na dimensão "autoconfiança para admitir e aprender com erros analíticos" (1.7.1). Seguindo esta linha, o primeiro ponto onde é levantada a contribuição do *big data*, em termos do volume é na dimensão "volume contribui para a compreensão dos métodos de coleta" (1.4.1). O efeito prejudicial do volume de *big data* é percebido primeiramente na dimensão "volume prejudica o domínio da matéria" (1.1.2).

Outras descobertas revelam que para alguns entrevistados *big data* já integra as práticas profissionais na realização de análises mais elaboradas e no desenvolvimento de projetos específicos. Entretanto para outros, *big data* ainda não é uma realidade, não sendo percebida a necessidade de utilizar grandes volumes de dados nas análises. Neste contexto, a principal contribuição identificada neste estudo foi a de que, apesar de o ambiente no qual o profissional em IE atua tenha se modificado ao longo dos anos, e este processo se intensificou com o advento do *big data*, nem todos o têm aplicado à IE. Este fato também denota um paradoxo entre a caracterização da produção de conhecimento no campo do *big data* e o seu uso no campo profissional de inteligência estratégica. Por um lado o maior volume de trabalhos sobre o *big data* está no campo profissional das organizações produtivas e não na academia e por outro os profissionais em inteligência ainda não percebem o valor do fenômeno para sua atuação profissional.

O fato é de que aspectos fundamentais que antes eram perseguidos por analistas como o acesso a informação rara e ainda não divulgada, a confidencialidade da informação e a própria experiência do analista no foco da observação já não representam as únicas formas de se gerar os produtos de inteligência. Isto porque no ambiente digital com *big data*, o acesso à informações primárias já é compartilhado entre a concorrência, e o conhecimento de domínio do analista que é decorrente de sua experiência, se não implementado ou atualizado em um curto espaço de tempo tende a se tornar obsoleto. Atualmente, com suporte da análise e ferramentas tecnológicas adequadas, os profissionais tem a oportunidade de identificar padrões desconhecidos em meio à grandes volumes de dados. Então já não se trata apenas de identificar a informação rara antes da concorrência e sim, a partir das informações

disponíveis, tentar identificar padrões relevantes ainda não percebidos pelos demais. É uma mudança de paradigma. Assim, das atividades analíticas tomadas como base para este estudo, pode-se identificar a carência daquelas que tratem sobre a questão dos aspectos tecnológicos e da multidisciplinaridade entre os profissionais em IE. Trata-se de um complemento acerca das atividades analíticas, adequadas ao *big data* e não da substituição daquelas já existentes.

Neste sentido, o aumento no número de publicações científicas sobre *big data*, no sentido de auxiliar na sua disseminação e aplicação em IE é um fator importante a ser considerado, visto que, em muitos momentos no decorrer das entrevistas pode-se perceber limitações conceituais dos entrevistados sobre o tema. Com isso, profissionais teriam contato com técnicas e utilização de grandes volumes de dados e poderiam chegar à conclusões para questões cotidianas ou novos projetos de IE dentro das organizações. Este é o primeiro passo no sentido do aumento no número de profissionais habilitados para realização de análise de grandes volumes de dados e, as organizações poderiam melhor avaliar o potencial de geração de valor de *big data* em diferentes aplicações. Esta é uma sugestão para estudos futuros.

Podem ser apontadas como limitações deste estudo: o fato de que o tema pesquisado é ainda novo do ponto de vista conceitual e pouco utilizado em termos práticos, o que dificultou o acesso à informações e bibliografia academicamente especializada; o aspecto da interpretação do pesquisador, por tratar-se de uma pesquisa qualitativa; o fato dos próprios profissionais em inteligência não serem facilmente identificados nas organizações, pois é raro que se tenha uma área específica da empresa dedicada a esta função.

Em se tratando do último caso, pode-se avaliar que o processo de análise atualmente perpassa a estrutura organizacional e atinge as mais variadas funções na organização. Outra área começa a se destacar neste contexto, embora somente observada em teoria dentro deste estudo: a ciência de dados. Com esta visão pessoas ou áreas dentro da organização passariam a desempenhar a função de criação e execução das políticas necessárias no tratamento dos dados e informações e sua disseminação. Estudos futuros não precisariam estar restritos a IE, mas a percepção de analistas das mais variadas áreas. Academicamente, por exemplo, poderia se entender a capacidade do pesquisador no acompanhamento das publicações.

Por fim, foi possível observar que em meio a atual complexidade de *big data*, os tradicionais métodos de análise e tratamento da informação já não são mais adequados e requerem uma revitalização. Para isto, como estudo futuro também pode ser proposta a construção de uma nova abordagem, que permita ao analista explorar melhor a questão cognitiva para a análise e tratamento de grandes volumes de dados.

## REFERÊNCIAS

- ALVES-MAZZOTTI, A. J.; GEWANDSZNAJDER, F. *O método nas ciências naturais e sociais: pesquisa quantitativa e qualitativa*. 2.ed. São Paulo: Pioneira, 1999.
- AKERKAR, R. *Big data computing*. Boca Raton: Taylor & Francis Group, LLC, 2014.
- BARDIN, L. *Análise de conteúdo*. São Paulo: Edições 70, 2011.
- BARLOW, M. *The culture of big data*. O'Reilly Media, Inc. 2013.
- BETSER, J.; BELANGER, D. *Architecting the enterprise via big data analytics*. In: LIEBOWITZ, Jay. *Big data and business analytics*. Boca Raton: Taylor & Francis Group, LLC, 2013.
- BEYER, M. A.; ANNE, L.; NICHOLAS, G.; DONALD, F.; VALENTIN, S. T. *Big Data is only the beginning of extreme information management*. Gartner, 2011.
- BRUCE, J. B.; George, R. Z. *Analyzing intelligence: origins, obstacles, and innovations*. Washington, D. C.: Georgetown University Press, 2008.
- CUKIER, K. *Data, data, everywhere: a special report on managing information*. *The Economist*, 394 (8671), Fevereiro 2010.
- DAVENPORT, T. H.; HARRIS, J. G. **Competição analítica: vencendo através da nova ciência**. Rio de Janeiro: Elsevier, 2007.
- \_\_\_\_\_ ; BARTH, P.; BEAN, R. *How "big data" is different*. *Opinion & Analysis*, July, 2012.
- DUMBILL, E. *What is big data? An introduction to the big data landscape*. O'Reilly Media, Inc., 2012. Disponível em: <<http://strata.oreilly.com/2012/01/what-is-big-data.html>>. Acesso em 05/03/2014.
- FACHINELLI, A. C.; RECH, J.; BERTOLINI, A. L. *La dinámica cognitiva de los procesos de análisis en la actividad de inteligencia estratégica*. In: SÁNCHEZ, América M.; ESTRADA, Martha C. *Administración de conocimiento y desarrollo basado en conocimiento*. Cidade do México: Cengage Learning, 2011.
- FLICK, U. **Introdução à pesquisa qualitativa**. 3.ed. Porto Alegre: Artmed, 2009.
- FLEISHER, C. S.; BENSOUSSAN, B. **Strategic and competitive analysis: methods and techniques for analyzing competition**. New Jersey: Prentice Hall. 2003.
- \_\_\_\_\_. *Business and competitive analysis: effective application of new and classic methods*. New Jersey: Pearson Education, Inc., 2007.
- FULD, L. M. *The new competitor intelligence: the complete resource for finding, analyzing, and using information about your competitors*. John Wiley & Sons, Inc., 1995.

- GIBBS, Graham R. **Análise de dados qualitativos**. Porto Alegre: Bookman, 2009.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 5.ed. São Paulo: Atlas, 2010.
- GLASER, B. G.; STRAUSS, A. L. *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine, 1967.
- GOMES, E.; BRAGA, F. Construção de um sistema de inteligência competitiva. In: STAREC, C.; GOMES, E. B. P.; CHAVES, J. B. L. **Gestão estratégica da informação e inteligência competitiva**. São Paulo: Saraiva, 2006.
- GOODMAN L. *Snowball sampling*. *Annals of Mathematical Statistics*. 1961.
- HALLIMAN, C. *A look at how a number of business and competitive intelligence professionals view or define the analysis process*. April 2003.
- HEUER, R. J. *Psychology of intelligence analysis*. *Center for the study of intelligence*, 1999.
- HIRSHBERG, P. O *big data* não faz milagre (Entrevista concedida a Bruno Ferrari). *Revista Exame*, São Paulo, Abril, v. 1037, n. 5, p. 100-101, Março 2013.
- IALEA. *Starting an analytic unit for intelligence led policing*. Lawrenceville, NJ: International Association of Law Enforcement Intelligence Analysts Inc. (IALEA), 2001.
- JOHNSTON, R. *Analytic culture in the U.S. intelligence community*. Washington, DC: The Center of the Study of Intelligence, 2005.
- KANDEL, S.; PAEPCKE, A.; HELLERSTEIN, J. M.; HEER, J. *Enterprise data analysis and visualization: an interview study*. *IEEE Transactions on visualization and computer graphics*, vol. 18, no. 12, December 2012.
- LETOUZÉ, E. *Big data for development: challenges & opportunities*. *UN Global Pulse*, May 2012.
- LOHR, S. *The age of big data*. *The New York Times*, 2012. Disponível em: <[http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all&\\_r=0](http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all&_r=0)>. Acesso em 05/03/2014.
- MAJUMDAR, A. K.; SOWA, J. F. *Big data: Structured and unstructured*. In: LIEBOWITZ, Jay. **Big data and business analytics**. Boca Raton: Taylor & Francis Group, LLC, 2013.
- MANYIKA, J.; et al. *Big data: the next frontier for innovation, competition, and productivity*. *McKinsey Global Institute*, June 2011.
- MATHAMS, R. H. *The intelligence analyst's notebook*. *Strategic intelligence: Theory and Application*, 2d. ed., Washington, DC: Joint Military Intelligence Training Center, pp. 77-96, 1995.
- MCAFEE, A. ; BRYNJOLFSSON, E. Big data: a revolução da gestão. *Harvard Business Review Brasil*, v. 90, n. 10, p. 39-45, Outubro 2012.

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

MORAES, R. Análise de conteúdo. **Revista Educação**. Porto Alegre, v.22, n. 37, p. 7-32, 1999.

MOREIRA, D. A. **O método fenomenológico na pesquisa**. São Paulo: Pioneira Thomson, 2002.

O' BRIEN, J. A. **Sistemas de informação e as decisões gerenciais na era da internet**. 3. ed. São Paulo: Saraiva, 2010.

PALOP, F.; VICENTE, J. M. *Vigilancia tecnológica e inteligencia competitiva: su potencial para la empresa española*. **Fundacion COTEC**. 1999. Disponível em: <[http://www.navactiva.com/web/es/descargas/pdf/avtec/potencial\\_vtec.pdf](http://www.navactiva.com/web/es/descargas/pdf/avtec/potencial_vtec.pdf)> Acesso em: 05/03/2014.

PAQUET, R. *Technology trends you can't afford to ignore*. **Gartner, Inc.**, 2010.

PETRY, A. O berço do *big data*. **Revista Veja**, São Paulo, Abril, v. 2321, n. 20, p. 71-76, Maio 2013.

PROCYSHYN, T. W. *The coming intelligence failure*. **Departament of National Defense (DND)**. Canada: 1998.

RADFAHRER, L. **Enciclopédia da Nuvem: 100 oportunidades e 550 ferramentas online**. Rio de Janeiro: Campus, 2012.

RAPPORTEUR, D. B. *The promise and peril of big data*. **The Aspen Institute**, January 2010.

ROSS, J. W.; BEATH, C. M.; QUAADGRA, A. *You may not need Big Data after all*. **Harvard Business Review**, December, 2013.

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

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

SIEGELE, P. *Big Oil Goes Mining for Big Data* (Entrevista concedida a Jessica Leber). **Mit Technology Review**. 2013. Disponível em: <<http://www.technologyreview.com/news/427876/big-oil-goes-mining-for-big-data>> Acesso em: 04 mar. 2013.

SFEZ, L. *Dictionnaire critique de la communication*. Paris: Presses Universitaires de France. 1993.

TAURION, C. **Big data**. Rio de Janeiro: Brasport, 2013. ePUB.

*World Economic Forum. **Big data, Big Impact: New Possibilities for International Development.** Switzerland, 2012.*

ZERDICK, A.; PICOT, A.; SCHRAPE, K.; ARTOPE, A.; GOLHAMMER, K.; LANGE, U.; VIERKANT, E.; LOPEZ-ESCOBAR, E.; SILVERSTONE, R. ***E-Conomics. Strategies for the Digital Market Place.*** New York: Springer, 1999.

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

ZIMMERMANN, A. *Leveraging the linked data principles for electronic communications.* In: ***IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.*** France, Lyon: 2011.

## APÊNDICE A - Detalhamento dos respondentes

(continua)

Entrevista	Formação	Função / Segmento onde atua	Tempo de atuação na função	Residência	Unidade de contexto
E1	Mestrado em Economia e Especialização em Gestão Financeira, Universidade Católica de Brasília, Brasil.	Gerente de Pesquisa e Unidade Prospectiva da Confederação Nacional da Indústria - CNI.	9 anos	Distrito Federal	Profissional
E2	Especialização em Gestão Empresarial, Universidade de Caxias do Sul, Brasil.	Área de Gestão Competitiva Marcopolo S/A.	4 anos	Rio Grande do Sul	Profissional
E3	Bacharel em Administração de empresas pela Universidade Norte do Paraná, Brasil.	Gerente de Planejamento de vendas na Lojas Colombo S/A.	3 anos	Rio Grande do Sul	Profissional
E4	PhD em Computação, Universidade Stanford, EUA; Especialização em Cirurgia Pediátrica, Toronto Sick Children, Canadá.	Cientista-Chefe da IBM Brasil; Coordenador da área de pesquisa na filial brasileira da companhia.	20 anos	São Paulo	Especialista
E5	Doutorado em Engenharia de Produção, Universidade Federal do Rio de Janeiro, Brasil; Mestrado em Engenharia Nuclear, Universidade Federal do Rio de Janeiro, Brasil.	Diretora de Consultoria e Educação na Plugar Informações Estratégicas, Professora Pesquisadora.	14 anos	Rio de Janeiro	Especialista
E6	Mestrado em administração, Grupo IBMEC, Brasil.	CEO da Plugar Informações Estratégicas; Pesquisador, Professor e Membro fundador da SBGC.	14 anos	Rio de Janeiro	Especialista
E7	Doutorado e Mestrado em Computação Universidade Federal do Rio Grande do Sul, Brasil.	Professor da Universidade Luterana do Brasil - ULBRA e da Faculdade de Tecnologia SENAC, além de pesquisador e consultor.	12 anos	Rio Grande do Sul	Especialista
E8 - não gravada	Especialização em Gestão de Empresas, Fundação Armando Alvares Penteado, Brasil; Extensão Universitária em Information Technology, Universidade da Califórnia at Berkeley, EUA.	Gerente de Prestação de Serviços em Empresa de Tecnologia da Informação para Indústria Automotiva.	15 anos	São Paulo	Especialista
E9	Especialização em Gestão Empresarial, Universidade Federal do Rio Grande, Brasil.	Gerente de Inteligência de Mercado SENAC-RS.	5 anos	Rio Grande do Sul	Profissional

(continuação)

<b>Entrevista</b>	<b>Formação</b>	<b>Função / Segmento onde atua</b>	<b>Tempo de atuação na função</b>	<b>Residência</b>	<b>Unidade de contexto</b>
E10	Especialização em Gestão Estratégica e Inteligência Competitiva, pela PUC-RS	Especialista de Informações de Mercado na DHB Componentes Automotivos S.A.	6 anos	Rio Grande do Sul	Profissional
E11	Especialização em Marketing, Fundação Getúlio Vargas, Brasil.	Analista de Marketing em Empresa do Setor Metal Mecânico.	8 anos	Rio Grande do Sul	Profissional
E12	Doutorado e Mestrado em Administração, Universidade de Brasília, Brasil.	Pesquisador da Área de Inovação Empresarial da Empresa Brasileira de Correios e Telégrafos - CORREIOS.	5 anos	Distrito Federal	Profissional
E13	Especialização em Gestão Estratégica da Tecnologia da Informação, Fundação Getúlio Vargas, Brasil.	Executivo de Inteligência de Mercado na Promob Software Solutions.	2 anos	Rio Grande do Sul	Profissional
E14	Especialização em Gestão Estratégica da Produção pela Ftec Faculdades.	Supervisor de Inteligência de Mercado em Empresa do Setor Calçadista.	2 anos	Rio Grande do Sul	Profissional
E15	Bacharel em Administração de empresas pela Universidade de Caxias do Sul, Brasil.	Consultora Comercial com Foco em Inteligência Estratégica, na Planef Consultoria e Gestão Empresarial.	5 anos	Rio Grande do Sul	Profissional
E16 - não gravada	Doutorado em Information Studies, University of Tampere, Finlândia; Mestrado em Ciências da Comunicação, Universidade de São Paulo, Brasil.	Professora e pesquisadora com foco na criação e comunicação de conhecimento tácito, especialmente em organizações em setores intensivos de conhecimento.	20 anos	São Paulo	Especialista
E17 - não considerada	Mestrado em Engenharia - Ciências dos Materiais, Universidade Federal do Rio Grande do Sul, Brasil.	Consultora do Ramo da Aviação na Área de Educação a Distância, Desenvolvimento e Treinamento Corporativo.	5 anos	Rio Grande do Sul	Especialista
E18	Doutorado em Ciência da Informação, Universidade Federal Fluminense, Brasil. Mestrado em Ciência da Informação, Universidade Federal do Rio de Janeiro, Brasil.	Professor dos programas de pós-graduação <i>Lato Sensu</i> do IBMEC, da ESPM, da Fundação Getúlio Vargas e da HSM.	14 anos	Rio de Janeiro	Especialista



(conclusão)

<b>Entrevista</b>	<b>Formação</b>	<b>Função / Segmento onde atua</b>	<b>Tempo de atuação na função</b>	<b>Residência</b>	<b>Unidade de contexto</b>
E19	Mestre em ciência da computação e MBA em marketing de serviços.	Gerente de Novas Tecnologias Aplicadas da IBM Brasil. Autor do primeiro livro brasileiro sobre <i>big data</i> .	30 anos	Rio de Janeiro	Especialista
E20	Global Executive MBA - University of North Carolina, United States; Tecnológico de Monterrey, Monterrey, Mexico; Fundação Getúlio Vargas, São Paulo, Brasil; RSM Erasmus University, Rotterdam, Netherlands; The Chinese University of Hong Kong, Hong Kong, China; Especialização em Gestão Estratégica, Fundação Getúlio Vargas, Brasil.	Responsável pelas Áreas de Estratégia, Inteligência Estratégica e Gestão Estratégica do Grupo Fleury.	7 anos	São Paulo	Profissional
E21	Especialização em Pesquisa de Mercado, Universidade do Estado do Rio de Janeiro, Brasil.	Gerente de Inteligência Competitiva na Plugar Informações Estratégicas.	15 anos	Rio de Janeiro	Especialista

FONTE: elaborado pelo próprio autor.

## APÊNDICE B - Roteiro de base para as entrevistas

Dimensões: <i>big data</i>  Atividade analítica	Volume	Velocidade	Variedade	Outro
Domínio da matéria: domínio de conteúdos e temas	<ul style="list-style-type: none"> <li>• Você acha que o volume, velocidade e variedade de informações pode estar afetando esta especialidade do analista?</li> <li>• Porque você acha que afeta ou não afeta?</li> </ul>			
Compreensão de métodos de investigação para organizar e avaliar dados	<ul style="list-style-type: none"> <li>• Você acha que o volume, velocidade e variedade de informações pode estar afetando a questão dos métodos de investigação, que é a busca pela informação relevante?</li> <li>• E sobre os métodos de coleta?</li> <li>• O que você pode me dizer dos métodos que você vinha usando para levantar a informação?</li> </ul>			
Compreensão dos métodos de coleta				
Imaginação e rigor científico para chegar a hipóteses	<ul style="list-style-type: none"> <li>• Na análise das informações com vistas a construção de hipóteses, como volume, variedade e velocidade podem estar afetando?</li> </ul>			
Consciência das próprias potencialidades e outras influencias cognitivas	<ul style="list-style-type: none"> <li>• Você sente a influência de VVV em aspectos cognitivos do processo de análise?</li> <li>• Você acredita que o seu desempenho enquanto analista pode estar sendo afetado por VVV?</li> </ul>			
Mentalidade aberta a opiniões contrárias ou modelos alternativos	<ul style="list-style-type: none"> <li>• Você acredita que VVV influencia a forma como você lida com diferentes tipos de opinião quanto à análise de informações?</li> <li>• E com relação a diferentes modelos de análise?</li> </ul>			
Autoconfiança para admitir e aprender com os erros analíticos.	<ul style="list-style-type: none"> <li>• Você já viveu uma situação em que cometeu equívocos de análise? Como ficou sua autoconfiança com esta situação? Você acredita que VVV influenciaram este processo?</li> </ul>			
Outro	<ul style="list-style-type: none"> <li>• Você gostaria de acrescentar alguma outra dimensão de <i>big data</i> ou analítica que eu não tenha citado?</li> </ul>			

FONTE: elaborado pelo próprio autor.