

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

LUIZ CARLOS FELIPE DE OLIVEIRA

**O BIG DATA NA PRODUÇÃO CIENTÍFICA DA CIÊNCIA DA INFORMAÇÃO**

Belo Horizonte

2017

LUIZ CARLOS FELIPE DE OLIVEIRA

## **O BIG DATA NA PRODUÇÃO CIENTÍFICA DA CIÊNCIA DA INFORMAÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Mestre em Gestão e Organização do Conhecimento.

Linha de Pesquisa: Gestão e Tecnologia – GET.

Orientadora: Prof<sup>a</sup>. Dra. Beatriz Valadares Cendón.

BELO HORIZONTE

2017

O48b

Oliveira, Luiz Carlos Felipe de.

O Big Data na produção científica da Ciência da Informação [manuscrito] /  
Luiz Carlos Felipe de Oliveira.– 2017.  
133 f., enc. : il. + 1 CD-Rom

Orientadora: Beatriz Valadares Cendón.

Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de  
Ciência da Informação.

Referências: f. 127-132.

Apêndice: em CD-Rom

1. Ciência da informação – Teses. 2. Big Data – Teses. 3. Cientistas da  
informação – Teses. I. Título. II. Cendón, Beatriz Valadares. III. Universidade  
Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 02:004.62



**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO**  
**CONHECIMENTO**



## FOLHA DE APROVAÇÃO

### O BIG DATA NA PRODUÇÃO CIENTÍFICA DA CIÊNCIA DA INFORMAÇÃO

**LUIZ CARLOS FELIPE DE OLIVEIRA**

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração REPRESENTAÇÃO DO CONHECIMENTO, linha de pesquisa Gestão e Tecnologia.

Aprovada em 26 de abril de 2017, pela banca constituída pelos membros:

Profa. Beatriz Valadares Cendon - Orientadora  
ECI/UFMG

Profa. Silvana Aparecida Borsetti Gregório Vidotti  
UNESP

Profa. Renata Maria Abrantes Baracho Porto  
ECI/UFMG

Belo Horizonte, 26 de abril de 2017



**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO  
CONHECIMENTO



**ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO**  
**LUIZ CARLOS FELIPE DE OLIVEIRA**

Realizou-se, no dia 26 de abril de 2017, às 14:00 horas, Sala 1000 - ECI/UFMG, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *O BIG DATA NA PRODUÇÃO CIENTÍFICA DA CIÊNCIA DA INFORMAÇÃO*, apresentada por LUIZ CARLOS FELIPE DE OLIVEIRA, número de registro 2015665832, graduado no curso de TECNOLOGIA EM PROCESSAMENTO DE DADOS, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Profa. Beatriz Valadares Cendon - ECI/UFMG (Orientadora), Profa. Silvana Aparecida Borsetti Gregório Vidotti - UNESP (por videoconferência), Profa. Renata Maria Abrantes Baracho Porto - ECI/UFMG.

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 26 de abril de 2017

Profa. Beatriz Valadares Cendon (orientadora)

Profa. Silvana Aparecida Borsetti Gregório Vidotti

Profa. Renata Maria Abrantes Baracho Porto

## DEDICATÓRIA

Dedico este trabalho a minha família.

A minha esposa Vânia das Graças Rocha Simões Oliveira, pelo incentivo, cooperação, encorajamento, inspiração e garra.

Aos meus filhos, Felipe Luiz Simões de Oliveira e Eduarda Luiza Simões de Oliveira, pela compreensão durante minhas ausências.

A minha mãe, Vera Lúcia Oliveira Araújo, por me fazer acreditar que tudo terminaria bem.

A minha sogra, Dona Albertina Rocha Simões pelo encorajamento e incentivo.

Ao meu sogro, Senhor Nilton Simões, pelo apoio.

A minha cunhada, Simone Simões pelo suporte e aconselhamento.

À memória de meu pai.

## **AGRADECIMENTOS**

Primeiramente à Deus, por ter me guiado no caminho certo, por ter me amparado e fortalecido nos momentos de desafio. Aos professores e funcionários do Programa de Pós-Graduação em Ciência da Informação (PPGCI), pelo acolhimento, aos professores e funcionários do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPGGOC) pela coragem, atenção, inspiração e ousadia.

A minha orientadora, Professora Beatriz Cendón, pela inspiração, paciência, direcionamento e amizade.

Um especial agradecimento aos senhores Guilherme Teixeira e Romualdo Roza, pela ajuda espontânea e sincera.

Também um especial e carinhoso agradecimento a todos os meus queridos e inspiradores ex-professores, que contribuíram com cada parcela de conhecimento que me capacitou para chegar a este momento.

“Não fique preso a dogmas - que é viver com o resultado do pensamento de outras pessoas.

Não deixe o barulho da opinião dos outros abafar a sua voz interior. E mais importante, tenha a coragem para seguir seu coração e sua intuição. Eles de alguma forma já sabem o que você realmente quer se tornar.”

Steve Jobs



## RESUMO

O grande volume de dados estruturados e não estruturados presente em nossos dias denominado *Big Data*, constituem demandas de atuação aos profissionais da informação e de pesquisa para a Ciência da Informação. Entretanto, observa-se que outras áreas têm se apropriado do estudo e uso do fenômeno com mais agilidade que a Ciência da Informação. Para auxiliar o entendimento de como a Ciência da Informação tem incorporado o *Big Data*, esta pesquisa tem o objetivo de estudar a abordagem que a Ciência da Informação atribui ao tema através da análise de sua produção científica. Na revisão de literatura são tratados o surgimento dos grandes volumes de dados, da chamada *e-science*, da *data science*, do *data scientist* e as consequências vivenciadas pela Ciência da Informação e pelo cientista da informação. A coleta de dados foi realizada através de pesquisa bibliográfica para identificar a produção científica sobre o tema *Big Data* na área da Ciência da Informação. Docentes, programas, universidades, países e abordagens foram identificados através da análise dos registros bibliográficos da produção científica sobre o tema *Big Data* na área da Ciência da Informação. Os dados foram analisados quantitativa e qualitativamente. Os resultados apresentam o mapeamento de como esta ciência tem abordado o tema *Big Data* em um *ranking* de universidades/entidades, programas de formação, autores e países.

**Palavras-chave:** *Big Data*. Ciência da Informação. *E-science*. *Datascience*. Cientista de dados. Cientista da Informação.

## ABSTRACT

The large amount of structured and unstructured data present in our days called Big Data constitute demands for information professionals and research for information science. However, it has been observed that other areas have been appropriated to the study and use of the phenomenon with more agility than the information science. To help understand how information science has incorporated Big Data, this research has the objective of studying the approach that information science attributes to the theme through the analysis of its scientific production. In the literature review are treated the emergence of large volumes of data, the called e-science, data science, data scientist and the consequences experienced by information science and information scientist. Data collection was done through a bibliographical research to identify the scientific production on the topic Big Data in the area of information science. Teachers, programs, universities, countries and approaches were identified through the analysis of the bibliographic records of the scientific production on the theme Big Data in the area of information science. The data were analyzed quantitatively and qualitatively. The results present the mapping of how this science has approached the theme Big Data in a ranking of universities/entities, training programs, authors and countries.

**Key-Words:** Big Data. Information Science. e-science. Data Scientist. Information Scientist.

## LISTA DE ILUSTRAÇÕES

FIGURA 1 – Volume de informação gerada a cada minuto na internet .....	21
FIGURA 2 – Limites e classes do <i>Big Data</i> .....	23
FIGURA 3 – Relacionamento entre <i>Big Data</i> , <i>e-science</i> e <i>data science</i> .....	28
FIGURA 4 – Fluxograma de desenvolvimento da pesquisa .....	38
GRÁFICO 1 – Mapa geral da classificação dos artigos.....	45
GRÁFICO 2 – Classificação dos artigos e países de origem .....	52
GRÁFICO 3 – Países com artigos classificados como Objeto de Pesquisa e sua qualificação .....	55
GRÁFICO 4 – Número de universidades/instituições com produção acadêmica sobre <i>Big Data</i> por ano .....	56
GRÁFICO 5 – Publicações por continente .....	97
GRÁFICO 6 – Mapa geral da classificação dos artigos: produção brasileira.....	114
GRÁFICO 7 – Número de universidades / instituições com produção acadêmica sobre <i>Big Data</i> por ano – bases BRAPCI e PERI.....	115
GRÁFICO 8 – Mapa da produção acadêmica anual sobre <i>Big Data</i> – universidades / instituições – Brasil .....	116
GRÁFICO 9 – Comparativo de tendências mundiais .....	120

## LISTA DE TABELAS

TABELA 1 – Países com artigos relacionados ao <i>Big Data</i> .....	43
TABELA 1 – Países com artigos relacionados ao <i>Big Data</i> .....	44
TABELA 2 – Número de universidades / instituições com publicação.....	57
TABELA 3 – Países e o número dos programas/departamentos com publicações .....	60
TABELA 4 – Países e o número de autores que publicaram.....	63
TABELA 4 – Países e o número de autores que publicaram.....	64
TABELA 5 – Programas com publicações – bases BRAPCI e PERI - 2013.....	117
TABELA 6 – Programas com publicações – bases BRAPCI e PERI – 2014 .....	117
TABELA 7 – Programas com publicações - bases BRAPCI e PERI - 2015.....	118

## LISTA DE QUADROS

QUADRO 1 –	Exemplos de instituições dedicadas à e-science .....	31
QUADRO 2 –	Recursos tecnológicos.....	40
QUADRO 3 –	Classificação das abordagens ao Big Data.....	41
QUADRO 4 –	Três exemplos de artigos classificados como Assunto Secundário .....	46
QUADRO 5 –	Três exemplos de artigos classificados como Reflexão .....	48
QUADRO 6 –	Exemplo de artigos classificados como Objeto de Pesquisa .....	49
QUADRO 7 –	Programas na University of California.....	58
QUADRO 8 –	Programas da University of Southampton .....	59
QUADRO 9 –	Programas da University of North Carolina.....	59
QUADRO 10 –	Publicações de autores sem vinculação – bases LISA e LISTA - 2010 a 2016 .....	67
QUADRO 11 –	Autores com três ou mais publicações .....	70
QUADRO 12 –	Universidades/Instituições com mais publicações – bases LISA e LISTA – 2010 a 2016 .....	71
QUADRO 13 –	Programas que mais publicaram – bases LISA e LISTA – 2010 a 2016 ...	73
QUADRO 14 –	Países que mais publicaram – bases LISA e LISTA – 2010 a 2016 .....	75
QUADRO 15 –	Classificação das publicações em cada país.....	75
QUADRO 16 –	Países com produção de artigos classificados como Objeto de Pesquisa.	76
QUADRO 17 –	Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Disseminação da Informação .....	76
QUADRO 18 –	Artigos classificados em Objeto de Pesquisa e qualificados em Disseminação da Informação .....	77
QUADRO 19 –	Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados em Infraestrutura.....	78
QUADRO 20 –	Artigos com classificação Objeto de Pesquisa e qualificados em Infraestrutura .....	79
QUADRO 21 –	Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Recuperação da Informação .....	81
QUADRO 22 –	Artigos classificados como Objeto de Pesquisa, qualificados como Recuperação da Informação – China e Alemanha.....	83
QUADRO 23 –	Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação. ....	87
QUADRO 24 –	Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha... ..	89
QUADRO 25 –	Comportamento das publicações por continente .....	98

QUADRO 26 – Empresas com publicações relacionadas ao Big Data – bases LISA e LISTA 2010 a 2016.....	98
QUADRO 27 – Empresas que mais publicaram e seus artigos.....	100
QUADRO 28 – Empresas que publicaram artigos classificados como Objeto de Pesquisa .. .....	103
QUADRO 29 – Ciências que publicaram artigos classificados como Objeto de Pesquisa – bases LISA e LISTA - 2010 a 2016.....	105
QUADRO 30 – Distribuição dos artigos classificados como Objeto de Pesquisa – por área de conhecimento .....	105
QUADRO 31 – Produção acadêmica no Brasil .....	108
QUADRO 32 – Autores com publicações relacionados ao Big Data – BRAPCI e PERI – 2013 a 2015 .....	118

## LISTA DE ABREVIATURAS

AS	Assunto Secundário
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CAPES	Comissão de Aperfeiçoamento de Pessoal do Nível Superior
CI	Ciência da Informação
CNC	Computação em Nuvem para Ciência
DeiC	Danish e-infrastructure Cooperation
DIS	Disseminação da Informação
FIOCRUZ	Fundação Oswaldo Cruz
IBICT	Instituto Brasileiro de Informação em Ciência da Informação
IBM	International Business Machines
INF	Infraestrutura
<i>IOT</i>	<i>Internet of Things</i>
<i>iSchool</i>	<i>Information School</i>
LISA	Library and Information Science Abstracts
LISTA	Library, Information Science & Technology Abstracts with Full Text
MSN	Microsot Network
NeSC	National e-Science Centre
NLeSC	Netherlands eScience Center
OP	Objeto de Pesquisa
PERI	Base de dados da Escola de Ciência da Informação da UFMG
PGCIN	Programa de Pós-Graduação em Ciência da Informação
PPGCI	Programa de Pós-Graduação em Ciência da Informação
PPGCINF	Programa de Pós-Graduação em Ciência da Informação
PPGOC	Programa de Pós-Graduação em Gestão e Organização do Conhecimento
RE	Reflexão
REC	Recuperação da Informação
RNP	Rede Nacional de Ensino e Pesquisa
SE-CNC	Serviço Experimental Computação em Nuvem para Ciência
SeRC	Swedish e-Science Research Centre
TRA	Tratamento da Informação
UFMG	Universidade Federal de Minas Gerais
UFPA	Universidade Federal do Pará
UFPB	Universidade Federal da Paraíba
UFRJ	Universidade Federal do Rio de Janeiro
UFSC	Universidade Federal de Santa Catarina
UNB	Universidade de Brasília
W3C	Consórcio World Wide Web

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>13</b>
<b>1.1</b>	<b>Problema de pesquisa</b> .....	<b>15</b>
<b>1.2</b>	<b>Objetivos</b> .....	<b>16</b>
1.2.1	Objetivo principal.....	16
1.2.2	Objetivos específicos .....	16
<b>1.3</b>	<b>Justificativa</b> .....	<b>16</b>
<b>2</b>	<b>REVISÃO DE LITERATURA</b> .....	<b>19</b>
<b>2.1</b>	<b>O fenômeno Big Data, sua origem e suas características</b> .....	<b>20</b>
<b>2.2</b>	<b>Novas habilidades para os profissionais da informação frente ao fenômeno Big Data</b> .....	<b>25</b>
<b>2.3</b>	<b>Data science (ciência de dados), e-science (e-ciência) e sua relação com o Big Data</b> .....	<b>27</b>
2.3.1	Data science .....	28
2.3.2	E-science .....	30
<b>2.4</b>	<b>O fenômeno Big Data e a sociedade</b> .....	<b>32</b>
<b>2.5</b>	<b>O fenômeno Big Data e a Ciência da Informação</b> .....	<b>33</b>
<b>2.6</b>	<b>O cientista da informação no contexto do Big Data</b> .....	<b>34</b>
<b>3</b>	<b>METODOLOGIA</b> .....	<b>36</b>
<b>3.1</b>	<b>Caracterização da pesquisa</b> .....	<b>36</b>
<b>3.2</b>	<b>Coleta de dados</b> .....	<b>37</b>
<b>3.3</b>	<b>Instrumentos de pesquisa</b> .....	<b>39</b>
3.3.1	Recursos tecnológicos .....	39
3.3.2	Métodos de obtenção dos dados .....	40
<b>3.4</b>	<b>Análise de dados</b> .....	<b>40</b>
<b>4</b>	<b>APRESENTAÇÃO DOS RESULTADOS</b> .....	<b>43</b>
<b>4.1</b>	<b>Análise da produção científica nas bases LISA e LISTA</b> .....	<b>43</b>
<b>4.2</b>	<b>Análise da produção científica nas bases BRAPCI e PERI</b> .....	<b>108</b>
<b>4.3</b>	<b>Comparativo de tendências mundiais</b> .....	<b>119</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>122</b>
<b>5.1</b>	<b>Conclusão</b> .....	<b>122</b>
<b>5.2</b>	<b>Contribuições</b> .....	<b>124</b>
<b>5.3</b>	<b>Limitações</b> .....	<b>125</b>
<b>5.4</b>	<b>Pesquisas futuras</b> .....	<b>126</b>
	<b>REFERÊNCIAS</b> .....	<b>127</b>
	<b>APÊNDICE A – Levantamento de dados da pesquisa</b> .....	<b>133</b>



## 1 INTRODUÇÃO

Reconhecida por Saracevic (1996) como ciência interdisciplinar, a Ciência da Informação (CI), empenha-se na solução de problemas relacionados ao dado, à informação e ao conhecimento, afirmando-se como ciência gestora do conhecimento. A revolução tecnológica e o conseqüente acúmulo de conhecimento sempre desafiaram a CI. Um dos recentes desafios para a área é o fenômeno *Big Data*, que se caracteriza pela grande escala dos dados que não permite seu processamento por sistemas ou mecanismos convencionais.

Burke (1992), demonstra que o ato de registrar eventos, seja para informar, ou com a intenção de posterior recuperação, esteve presente na evolução da humanidade. Tais iniciativas são percebidas na utilização de desenhos nas paredes das cavernas, nos hieróglifos egípcios em pedra, nos textos em pergaminhos, no papel, nos livros, na organização das bibliotecas, na imprensa, na máquina de perfurar cartões, até a era atual da informática e da web. Com a chegada da era digital, percebe-se o aumento no ritmo de geração, armazenamento, processamento e transmissão de registros de dados pela humanidade.

Em uma sociedade conectada, ideias, notícias e decisões assumem velocidade e volatilidade (JENKIN, 2010). O pensamento, o trabalho, a diversão e o relacionamento tornaram-se remotos e digitais. O mundo 'diminuiu de tamanho' e conforme preconizado por McLuhan (1969), surge a Aldeia Global. Passarelli, Junqueira e Angeluci (2014), comentam que geramos e consumimos grandes volumes de dados através do uso de aparatos tecnológicos, que, nos dias atuais podem ser dispositivos móveis, aplicativos, redes sociais, computação em nuvem e da comunicação máquina x máquina, a chamada *Internet of Things (IOT)*, que é definida por Guinard e Trifa (2016, p. 4, tradução nossa), como:

[...] simplesmente um mundo onde a internet torna-se mais do que a coleção de páginas multimídia que é hoje e estendendo-se para o mundo físico, em tempo real usando uma infinidade de dispositivos embarcados.

Então, Saracevic (1999, p. 1052, tradução nossa) destaca o relacionamento desta ciência com o volume informacional quando comenta:

O ritmo acelerado dos avanços científicos e técnicos que se acumulavam desde o início do século XX, produziu em meio século, uma revolução científica e técnica. A manifestação mais visível dessa revolução foi o fenômeno da "explosão informacional", referindo-se ao crescimento exponencial e ininterrupto de publicações científicas, técnicas e registros de informação de todos os tipos.

Sobre a dinâmica estabelecida entre a tecnologia e o volume informacional, Costa e Orrico (2007, não paginado) acrescentam:

Essa nova dinâmica informacional produz o desenvolvimento de conhecimentos que necessitam, por sua vez, da produção de novos conhecimentos que estabelecem novos paradigmas sociais, científicos, tecnológicos e, acrescentaríamos, culturais.

Mysore; Khupat e Jain (2014, *online*) observam que o grande volume de dados presente em nossos dias demandam novas formas de registrar, organizar, indexar, analisar e disponibilizar esses dados. Esse contexto apresenta novos desafios para a CI e particularmente na contribuição de quatro ciências que foram fundamentais desde os primórdios desta ciência: a biblioteconomia, a inteligência artificial, a ciência da computação e a tecnologia da informação, conforme pode-se verificar em Souza, Almeida e Baracho (2013) e Saracevic (1996).

Ciência moderna e aberta à interdisciplinaridade, a CI tem sua origem na resposta à necessidade de gestão, tratamento, organização e disponibilização da informação e do conhecimento. Reconhecido como um dos precursores da CI, Bush (1945) propôs o Memex – um dispositivo capaz de expandir a capacidade de armazenamento do cérebro humano –, publicando sobre isto no artigo *As we may Think*. A ideia, futurista à época, buscava meios de expandir a capacidade de guardar e recuperar informações, que àquela época, já extrapolavam as capacidades humanas e era motivo de preocupação, conforme descreve Saracevic (1996, p. 42-43):

O problema era (e, basicamente, ainda é) “a tarefa massiva de tornar mais acessível, um acervo crescente de conhecimento”; Bush (1945) identificou o problema da explosão informacional – o irreprimível crescimento exponencial da informação e de seus registros, particularmente em ciência e tecnologia. A solução por ele proposta era a de usar as incipientes tecnologias de informação para combater o problema. E foi mais longe, propôs uma máquina chamada MEMEX, incorporando (em suas palavras) a capacidade de associar ideias, que duplicaria “os processos mentais artificialmente”. É bastante evidente a antecipação do nascimento da CI e, até mesmo, da inteligência artificial.

A ideia de Bush (1945) foi percebida por pesquisadores e governos, que ao longo das décadas seguintes, dedicaram-se à busca de solução para a explosão informacional presente àquele momento, conforme Saracevic (1996, p. 43):

Cientistas e engenheiros de todo o mundo, e os mais importantes governos e agências de financiamento em muitos países ouviram e agiram. Nos Estados Unidos, o Congresso e outras agências governamentais aprovaram, durante os anos 50 e 60, inúmeros programas estratégicos que financiaram os esforços em larga escala para controlar a explosão informacional, primeiro na ciência e tecnologia, e depois em todos os outros campos. Empresas privadas uniram-se a eles. Eventualmente, esses programas e esforços foram responsáveis pelo desenvolvimento da moderna indústria da informação e das concepções que a direcionam.

Diante do fenômeno *Big Data*, é factível admitir que a CI esteja diante da mesma necessidade que determinou o seu surgimento: expandir a capacidade de guardar e recuperar informações que extrapolam as capacidades humanas de processamento, conforme Saracevic (1996, p. 42):

Um campo é definido pelos problemas que são propostos e a Ciência da Informação é definida como um campo englobando, tanto a pesquisa científica quanto à prática profissional, pelos problemas que propões e pelos métodos que escolheu, ao longo do tempo, para solucioná-los.

Ao incorporar o fenômeno *Big Data* entre suas preocupações, a CI aborda um tema de difícil definição. Não há uma origem, formato, limite ou autoria definida. Tanto o segmento corporativo quanto a academia estão envolvidos com o *Big Data* (AZAM, 2014; TAURION, 2012b). Percebem-se ações que partem do mercado da computação reivindicando o que denominam novo profissional da informação. Porém, observa-se que a academia iniciou-se na formação deste novo profissional o qual se denomina cientista de dados.

## 1.1 Problema de pesquisa

Ao envolver-se com temáticas relacionadas ao fenômeno *Big Data*, a CI poderá enfrentar a necessidade do aprendizado de novas ferramentas, tecnologias computacionais, semânticas, modelagem, e dedicar-se à criação de novos serviços de informação, além de ampliar seu próprio entendimento do que é o fenômeno *Big Data*, além de outras questões como *cloud computing*, redes sociais, web semântica, *IOT*, computação cognitiva, entre outros, conforme observam Souza, Almeida e Baracho (2013).

Para Ribeiro (2014), é importante para a CI, reafirmar-se dentro do cenário informacional, no qual o *Big Data* se faz presente, além de definir como o cientista da informação poderá atuar neste contexto. Igualmente importante, é avaliar se esse está capacitado para as exigências do *Big Data*, e possibilitar ao cientista da informação a escolha de uma linha de trabalho e pesquisa relacionada ao tema.

Souza, Almeida e Baracho (2013, p. 170-171) acrescentam:

Muitos programas de graduação e pós-graduação em Ciência da Informação têm buscado intercâmbios maiores com outras áreas. O contato com a ciência da computação aparece no sentido de se desenvolverem melhores concepções de sistemas de informação ou, juntamente com a engenharia, a estatística e a matemática aplicada, ajudando a prover modelos que permitam a evolução das tarefas de indexação, classificação/clustering e os pipelines analíticos necessários ao tratamento da informação.

O fenômeno *Big Data* provoca mudanças em formatos, suportes informacionais, modos de obtenção, produção e disseminação da informação, conforme relatam Mysore, Khupat e Jain (2014, *online*):

Os problemas com *Big Data* são, geralmente, difíceis de serem analisados e solucionados. O volume absoluto, a velocidade e a variedade dos dados dificultam a extração das informações e dos insights de negócios. Uma primeira etapa adequada é classificar o problema do *Big Data* de acordo com o formato dos dados a serem processados, o tipo de análise a ser aplicada, as técnicas de processamento usadas e as fontes de dados para os dados que o sistema alvo deve adquirir, carregar, processar, analisar e armazenar.

Conforme Souza, Almeida e Baracho (2013), o fenômeno *Big Data* relaciona-se com as ciências, organizações, governos e pessoas.

O desenvolvimento desta pesquisa visa responder à pergunta: como o tema *Big Data* tem sido abordado na produção científica da CI?

## 1.2 Objetivos

Os objetivos, principal e específicos, são apresentados abaixo.

### 1.2.1 Objetivo principal

Identificar como o tema *Big Data* tem sido abordado na produção científica da CI.

### 1.2.2 Objetivos específicos

Esta pesquisa se propõe especificamente a:

- a) levantar a produção científica que aborda o tema *Big Data*, nas bases internacionais Library and Information Science Abstracts (LISA) e Library, Information Science & Technology Abstracts with Full Text (LISTA) e nas bases nacionais Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) e Base de dados da Escola de Ciência da Informação (PERI) da Universidade Federal de Minas Gerais (UFMG);
- b) identificar as universidades/instituições, autores, programas e países responsáveis por estas produções científicas;
- c) classificar como estas produções científicas tem abordado o tema *Big Data*.

## 1.3 Justificativa

Souza, Almeida e Baracho (2013), reconhecem o impacto do *Big Data* para a ciência e em especial para a CI e comentam sobre a necessidade das publicações dedicadas ao tratamento, registro e utilização da informação, próprios da CI, além de

abordagens dedicadas às contribuições que esta ciência pode trazer neste novo contexto informacional. Diante da percepção destes autores, os mesmos comentam que “são necessários recortes explícitos e muitas vezes arbitrários, pois a quantidade de publicações e fontes disponíveis sobre assuntos específicos é frequentemente intratável.” (p. 168).

Souza, Almeida e Baracho (2013) e Mysore, Khupat e Jain (2014, *online*), destacam o reflexo que o *Big Data* causa através dos grandes volumes de dados, não somente nas universidades, mas em diversos segmentos da sociedade. Barbieri (2011), destaca a evolução no tratamento do dado, da informação e do conhecimento em que a tecnologia tem possibilitado a criação constante de novos serviços, evidenciando-se as altas capacidades de armazenamento e propagação de imensos volumes de dados, a conectividade, o poder de processamento em nuvem, a capilaridade e a instantaneidade.

Albino e Reinhard (2009, p. 4), destacam:

Novas tecnologias para comunicação com grande largura de banda, trabalho corporativo e remoto, objetos e multimídia, ampliaram o ambiente informacional e hoje já existem inúmeras ferramentas que facilitam ou suportam os atuais projetos de gestão do conhecimento.

Percebe-se que o fenômeno *Big Data* representa uma categoria de obtenção, geração e transmissão de conhecimento entre comunidades científicas, organizações e pessoas, através de serviços de *backoffice* globais, como Google, Microsoft Network (MSN), DropBox, Amazon, One Drive, etc., e nos ambientes informáticos acadêmicos e corporativos, hospedados em nuvem, conforme pode-se observar no website do projeto de Computação em Nuvem para Ciência (CNC) (COMPUTAÇÃO..., 2017). Destaca-se neste projeto, o envolvimento de instituições como a Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES), UFMG, Universidade Federal do Pará (UFPA), Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Rio Grande do Norte (UFRN), Rede Nacional de Ensino e Pesquisa (RNP) e Anolis – Tecnologia da Informação. No site deste projeto, comenta-se:

Nuvens de armazenamento de dados são ferramentas cada vez mais comuns no cotidiano de professores e pesquisadores. Estes usuários utilizam nuvens para armazenar e compartilhar dados, como notas de aula e resultados de pesquisas científicas. Para isto, utilizam-se de nuvens comerciais como o *DropBox* ou o *Microsoft One Drive*. Estas ferramentas são muitas vezes disponibilizadas gratuitamente e oferecem um grande número de benefícios a seus usuários. (COMPUTAÇÃO..., 2017, *online*)

Antes mesmo do surgimento do atual contexto informacional, Saracevic (1996, p. 2, tradução nossa) observou:

A Ciência da Informação está inexoravelmente ligada à tecnologia da informação. O imperativo tecnológico determina a Ciência da Informação, como ocorre também em outros campos. Em sentido amplo, o imperativo tecnológico está impondo a transformação da sociedade moderna em sociedade da informação, era da informação ou sociedade pós-industrial.

Conforme Ekbia (2015, p. 1524, tradução nossa), “*Big Data* é uma temática desafiadora e presente em diversos círculos de discussão”. Percebe-se que a informação se tornou uma espécie de ativo intangível de alto valor agregado, além de insumo indispensável para o cotidiano humano, criando assim diversas terminologias que conceituam o papel da informação nos dias atuais, conforme Araújo e Freire (1999, p. 5):

A informação, na verdade, é indispensável para toda e qualquer atividade humana, sendo cada vez mais vista como uma força importante e poderosa a ponto de dar origem a expressões como: sociedade da informação, explosão da informação, era da informação, indústria da informação, revolução da informação, sociedade, pós-sociedade da informação.

Essa forma de obter, manter e apresentar dados constitui um grande desafio à ciência e ao segmento corporativo (BIG DATA BUSINESS, 2015). Como exemplo, a International Business Machines (IBM) Developerworks (2013), cita que a população mundial cria diariamente 2,5 quintilhões de dados, concluindo que 90% dos dados existentes atualmente, foram gerados nos últimos dois anos.

Portanto, o estudo detalhado e o desenvolvimento desta pesquisa, justifica-se pelo fato de identificar como as produções científicas da CI sobre o *Big Data* têm sido abordadas pelas universidades e instituições, em função da inserção desta ciência no cenário informacional e a necessidade de entendimento de sua postura.

## 2 REVISÃO DE LITERATURA

Conforme Taurion (2012b), o ambiente informacional transcendeu o modelo de informações estruturadas, onde as demandas ocorriam em formatos e escalas previsíveis. Esse autor demonstra que atualmente tem-se um modelo não linear, sem formatos ou origens definidas e em grandes volumes, provocando mudanças no comportamento dos usuários, organizações, governos, centros de pesquisa e na academia.

Com a explosão informacional, as ciências ligadas à informação, dados e conhecimento atravessam grandes transformações, conforme comenta Bates(2015, *online*, tradução nossa):

Neste momento histórico particular, estamos participando de uma extraordinária mudança radical no modo como a Ciência da Informação, as bibliotecas, os arquivos e todas as disciplinas relacionadas à informação são vistos. Cada semana, ao que parece, aprendemos algo de um novo campo relacionado à informação.

Nesse contexto, percebe-se a possibilidade de o cientista da informação ocupar novos espaços, oferecer novos serviços e buscar novas competências, conforme Souza (2006). Hey e Hey (2006) apontam que a ciência irá demandar novos conhecimentos e habilidades dos profissionais da informação.

Segundo Souza, Almeida e Baracho (2013) pode-se buscar uma ordenação entre os interesses de estudo da Ciência da Informação, de tal forma que o acervo informacional pré-*Big Data*, com volume, suportes e formatos conhecidos, convivam harmonicamente com a nova face tecnológica, que se apresenta como opção ao cientista da informação, que pretende se dedicar ao *Big Data*, em um novo contexto, onde a informação apresenta volume, suporte e formatos desconhecidos.

Percebe-se que, a literatura não descreve limitações à atuação da CI quanto a questões relacionadas ao *Big Data*, o que nos permite entender que existe compatibilidade e aderência entre os estudos, ferramentas, o campo de atuação da CI e o fenômeno *Big Data*. Passarelli, Junqueira e Angeluci (2014, p. 160) afirmam:

[...] a reboque da sociedade contemporânea em rede, emergem novas lógicas, novas semânticas, novas literacias, novos modelos de negócios e novas práticas que ultrapassam as dualidades emissor receptor da comunicação de massa do século passado, relocando a atenção dos teóricos da comunicação, das instituições de ensino e pesquisa e das empresas da chamada “nova economia” para a reciprocidade das ações comunicacionais onde os usuários da modernidade agora, na contemporaneidade, são denominados *prosumers* (produtor + consumidor) com a conseqüente redefinição dos papéis destes atores em rede.

## 2.1 O fenômeno *Big Data*, sua origem e suas características

Com o propósito de compreender o fenômeno *Big Data*, Macedo (2013) afirma que se pode visualizar milhares de pessoas e empresas com seus dispositivos móveis, seus computadores, suas ações de compra, movimentações bancárias e fiscais, milhões de e-mails, milhões de opiniões em redes sociais e sites especializados, telefonemas e históricos de vidas.

Considerando Chang, Mishra e Lin (2015), podem-se visualizar sensores e equipamentos de controle, câmeras, automóveis e outros tantos dispositivos conectados, todos gerando dados que são transmitidos através da teia da web. Desta forma, muitas de nossas ações convertem-se em dados.

Suportado pelo entendimento do que diz o autor acima, cita-se como exemplo, que uma simples caminhada de um indivíduo em um parque permite aos olhares digitais com interesses específicos, detectar o comportamento de um potencial cliente através dos sinais transmitidos por seu celular e como uma alquimia cibernética, transformar estes dados em informação e conhecimento, que podem embasar ações comerciais, governamentais ou pesquisas científicas.

Conforme a Figura 1, a seguir representada, o site Bluesyemre (2017) destaca o volume de informação gerado a cada minuto na internet, de forma a contribuir para o entendimento do fenômeno *Big Data*. Observa-se nesta figura o volume de informações gerado a partir de redes sociais, mecanismos de buscas, sites de entretenimento, *e-mails*, sites de compra e de prestação de serviços. Vale ressaltar que não foram considerados pelo autor, a participação e contribuição do volume de informação gerado a partir do ambiente acadêmico e científico.



**FIGURA 1–Volume de informação gerada a cada minuto na internet**



Fonte: adaptado de Bluesyemree (2017, *online*).

Segundo Turkle (2016) o avanço tecnológico propiciou alteração nos hábitos das pessoas e acrescentou volatilidade e velocidade ao processo de comunicação e informação, de maneira global e instantânea, corroborando com a definição do cenário da "Aldeia Global", preconizada por McLuhan (1969).

Dumbill et al. (2012) contribuiu destacando que a velocidade e a volatilidade dos dados e dos formatos impulsionam novos suportes tecnológicos, demonstrando que as ondas de mudança são extremamente rápidas.

Ferrer-Sapena e Sánchez-Pérez (2013, p.150, tradução nossa), relacionam o fenômeno *Big Data* à necessidade do acesso além da informação, aos dados, em formato aberto e público:

Já não se trata somente de compartilhar informação e permitir livre acesso às publicações, agora queremos que estejam acessíveis também os dados, de maneira livre e gratuita. Os dados chegam de todos os contextos da sociedade, gerados pelos aplicativos da web 2.0, pelos governos, investigações científicas, publicações e organizações.

O site Open Definition (2017, *online*, tradução nossa), define dados abertos como “um dado que pode ser livremente utilizado, reutilizado e redistribuído por qualquer um”.

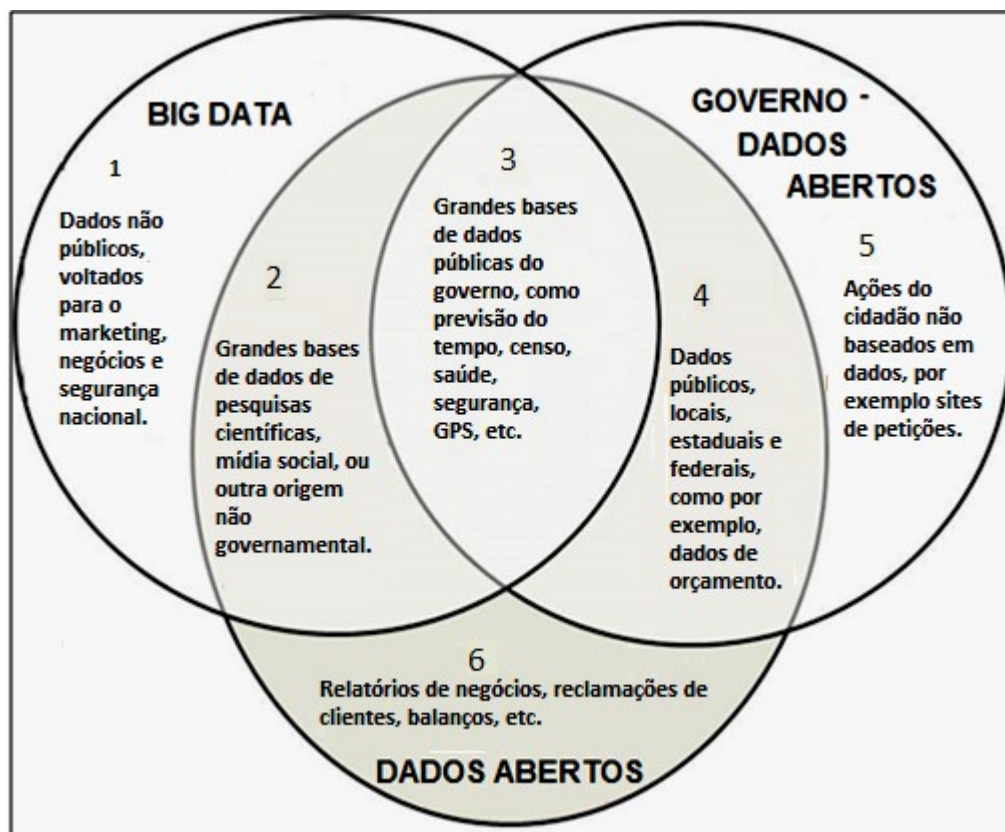
Neste contexto, o Consórcio World Wide Web (W3C) (2011?, p. 7), através de seu site acrescenta:

Dados abertos, especialmente os governamentais, são um ótimo recurso ainda muito pouco explorado. Muitos indivíduos e organizações coletam uma ampla gama de diferentes tipos de dados para executar suas tarefas. O governo é particularmente importante nesse contexto, tanto por causa da quantidade e da centralidade dos dados que coleta quanto pelo fato de que tais dados são públicos.

Percebem-se diversas ações voltadas às questões dos dados abertos públicos e governamentais em diversos países e esferas públicas, dos quais, sem pretensão de exaustão do tema, podemos destacar:

- d) (<http://www.mindtrek.org/2009/node/127>), da Finlândia;
- e) (<http://wheredoesmymoneygo.org/> – para onde vai o meu dinheiro?), da Grã-Bretanha;
- f) (<http://folketsting.dk/>), da Dinamarca;
- g) no Brasil:
  - (<http://cmsp.topical.com.br/>);
  - (<http://alagamentos.topical.com.br/>);
  - (<http://legisdados.org/>);
  - (<http://dadosabertos.senado.leg.br/>);
  - (<http://dados.gov.br/>);
  - (<http://www.transparencia.mg.gov.br/dados-abertos>).

Gurin (2013) apresenta na Figura 2 um modelo de entendimento da composição do *Big Data*. Valendo-se desse modelo, baseado em limites e classes, o autor apresenta o *Big Data* sob a ótica dos dados abertos (*open data*) e privados (*private data*) e demonstra que nem todos os dados privados e nem todos os dados abertos, compõem o fenômeno *Big Data*.

FIGURA 2–Limites e classes do *Big Data*

Fonte: adaptado de Gurin (2013, *online*).

Os limites e classes do *Big Data* citados por Gurin (2013) são detalhados a seguir:

- Big Data* que não são dados abertos: nessa categoria inserem-se informações de grande relevância para negócios, como hábitos de consumo, informações relacionadas a pacientes em hospitais, informações relacionadas a transações com cartão de crédito, etc. Considera-se o uso dessas informações para obtenção de vantagens competitivas. Alguns dados governamentais se inserem nessa categoria;
- Big Data* que são dados abertos, porém não governamentais: são dados originados de projetos científicos que compartilham dados. Grandes volumes de dados de observações astronômicas, de grandes projetos biomédicos, como o projeto Genoma Humano. Adicionam-se também nesta categoria, dados mantidos ou analisados por agências governamentais, o que não os caracteriza como dados governamentais. Inserem-se ainda nesta categoria, dados obtidos das mídias sociais;
- dados abertos, não em grandes volumes, que não são originados do governo. Dados de companhias privadas compartilhados para diversas finalidades, dentre elas, relacionamento com investidores, comunidade e

- meio ambiente. Os dados de reclamação dos consumidores, também compõem esta categoria;
- d) dados do governo que são dados abertos: dados do governo produzidos através do engajamento (opinião) do cidadão;
  - e) dados abertos do governo que não são *Big Data*: dados governamentais mesmo em pequenos volumes possuem valor. Em alguns casos, conjuntos menores de dados de estados, cidades ou governo federal, podem ter um maior impacto quando tratados de forma distinta. Como exemplo, podem ser citados os dados de orçamento, onde os cidadãos são convidados a participar na escolha dos investimentos que deverão ser priorizados. Também são interessantes no direcionamento de investimentos públicos ou ações voltadas para a saúde pública e orientação ao cidadão no uso de serviços públicos;
  - f) dados originados pelo governo em grandes volumes: esses conjuntos de dados podem ser os mais impactantes de todos. Governos têm capacidade e recursos para reunir grandes volumes de dados. Ao torná-los abertos, consegue-se grandes benefícios econômicos. Dados como Censo, seguridade social, balança comercial, saúde pública, etc. Essa categoria torna-se relevante pelo volume, robustez e significância de seus dados.

Em um cenário moldado a tantas mãos e *bytes*, a vasta gama de formatos e padrões dos dados, que hoje podem ser em forma binária, textual, hipertextual, de imagem, de filmes, de sons, entre outros, em volumes imensos, compõem o fenômeno denominado *Big Data*. Dumbill et al. (2012, p. 3, tradução nossa) define o *Big Data* como:

Dado que excede a capacidade de processamento dos sistemas convencionais de gestão de dados. Os dados são tão grandes, movem-se tão rápido ou não cabem nas estruturas convencionais de dados. Para tirar proveito destes dados, você precisa escolher um modo alternativo de processá-lo.

Não há um consenso sobre a autoria da denominação *Big Data* atribuído a esse imenso volume de dados que transita entre nosso meio e transcende a capacidade de entendimento humano, com formatos, fontes, participantes e interesses variados. Taurion (2012b, *online*) comenta:

O termo *Big Data* está cada vez mais popular, embora ainda esteja mal compreendido. [...] O *Big Data* vem chamando atenção pela acelerada escala em que volumes cada vez maiores de dados são criados pela sociedade. Já falamos comumente em petabytes de dados gerados cada dia e *zetabytes* começam a ser uma escala real e não mais imaginária e futurista. O que era futuro há uma década, *terabytes*, hoje nós já temos nas nossas próprias casas.

Percebe-se que constantemente surgem novos atributos e novas tentativas de definir o o fenômeno *Big Data*. Russom et al. (2011, *online*), McAfee e Brynjolfsson (2012,

*online*), por exemplo, defendem que o *Big Data* é composto por três dimensões: Volume, Veracidade e Velocidade. Demchenko e Membrey (2014), Taurion (2012a) entre outros defendem que o fenômeno *Big Data* possui cinco dimensões, acrescentando Veracidade e Valor.

De acordo com os autores acima, as dimensões ou Vs do *Big Data*, podem ser assim definidos:

- a) volume: identifica grandes quantidades de dados que devido ao seu volume não é possível processá-los através das tecnologias e bancos de dados tradicionais. Taurion (2012a) comenta que a tecnologia que suporta o *Big Data* se apoia em computação distribuída e *softwares* apropriados a este tipo de cenário;
- b) velocidade: segundo Taurion (2012a), está relacionada à maneira como os dados são criados e propagados. Através da tecnologia de suporte ao *Big Data*, pode-se analisar estes dados instantaneamente, sem a necessidade de que façam parte de uma base de dados específica;
- c) variedade: para Taurion (2012a), refere-se à diversidade de tipos de dados estruturados ou não, que podem ser analisados pelas ferramentas tecnológicas relacionadas ao *Big Data*, e convertidas em padrões tradicionais e estruturados de processamento e análise;
- d) veracidade: Taurion (2012a) comenta que as ferramentas tecnológicas permitem analisar os dados, cuja origem possa ser pouco confiável. Porém, salienta-se que a avaliação da relação volume disponível x a qualidade do dado, deve ser realizada com o propósito de apurar se há uma relação positiva que justifique o processo;
- e) valor: conforme Taurion (2012a), esta possa ser a principal variável a ser avaliada. Se não é possível transformar estes dados em valor, os mesmos não terão utilidade.

## **2.2 Novas habilidades para os profissionais da informação frente ao fenômeno *Big Data***

Garritano e Carlson (2009) e Stryker (2014) chamam a atenção para questões relacionadas à formação do profissional da informação e às novas habilidades necessárias a esse profissional no atual ambiente informacional, por onde circula o fenômeno *Big Data*.

A importância dos objetos de estudo do cientista da informação contrasta com uma aparente dificuldade de organização entre as ciências que alicerçam a CI, conforme Souza, Almeida e Baracho (2013, p. 162):

Os cursos de graduação ainda buscam prover bibliotecários e arquivistas com formação instrumental, com pouca ênfase à pesquisa. E a pós-graduação têm (*sic*) recebido egressos de áreas diversificadas, mas que pouco contribuem para a transformação da área.

Esses mesmos autores, demonstram preocupação com tal perceptível perda sinérgica entre as ciências dedicadas ao dado, informação e conhecimento, sejam as ciências pré-*Big Data* ou as pós-*Big Data* e se nada for feito, talvez as futuras.

Para Souza, Almeida e Baracho (2013, p. 171), “essa essência interdisciplinar exorta o cientista da informação a navegar nos espaços teóricos, adaptar-se aos contextos tecnológicos e reinventar-se continuamente”. Bates (1999, p. 1046, tradução nossa), acrescenta:

Se você deseja trabalhar com organização e recuperação da informação, você deve ser um profissional da informação, não um especialista sem treinamento em informação. Todas as especializações do mundo não são suficientes, se você não possui habilidades e conhecimentos do trabalho com a informação.

Os grandes volumes de dados que se originam da *e-science*, provocam reflexos no meio científico, que se torna grande demandante da presença dos profissionais da informação. Conforme comentam Hey e Hey (2006) à medida que essas grandes quantidades de dados se tornam disponíveis, cientistas necessitarão do apoio dos profissionais da informação para gerenciar, pesquisar, preservar e armazenar esses novos recursos de dados. Cabendo então aos cientistas da informação que se encontrem preparados para este novo cenário.

Souza (2006, p. 26) destaca a possibilidade de o bibliotecário e o cientista da informação participarem ativamente da dimensão tecnológica da CI:

Em uma síntese dessas posições, pode-se dizer que as questões atuais relacionadas à formação do profissional bibliotecário e cientista da informação perpassam por sua capacitação para responder a problemas práticos e teóricos que se produzem em: Gestão e administração da informação, Tratamento da informação, Tecnologia da Informação, Atendimento e interação com o Usuário e as que afetam suas próprias Atitudes e qualidades pessoais. (SOUZA, 2006, p. 26).

Garritano e Carlson (2009) delinearão um conjunto de habilidades para bibliotecários que procuram apoiar as necessidades da *e-science*. Eles identificaram cinco categorias de habilidades que são esperadas ou devem ser desenvolvidas pelos novos bibliotecários para participar de projetos relacionados à *e-science*:

- a) conhecimento na área de biblioteconomia e CI;
- b) conhecimento no tema da pesquisa;
- c) parcerias e divulgação (internas e externas);

- d) participação em pesquisas;
- e) trabalho em equipe.

Stryker (2014) aponta para as habilidades necessárias a quem se lança a trabalhar com *Big Data*:

- a) habilidade em estruturar ideias;
- b) habilidade em encontrar *insights* nos grandes volumes de dados;
- c) habilidade de transformar esses *insights* em questões utilizáveis e úteis à sociedade;
- d) habilidade de raciocínio e ações rápidas.

### **2.3 Data science (ciência de dados), e-science (e-ciência) e sua relação com o Big Data**

Bryant, Katz e Lazowska (2008, tradução nossa), comentam que a ciência se tornou grande geradora de dados em virtude do emprego de equipamentos e acessórios de alta complexidade e precisão, digitais e muitas vezes conectados à web, contribuindo para o fenômeno *Big Data*:

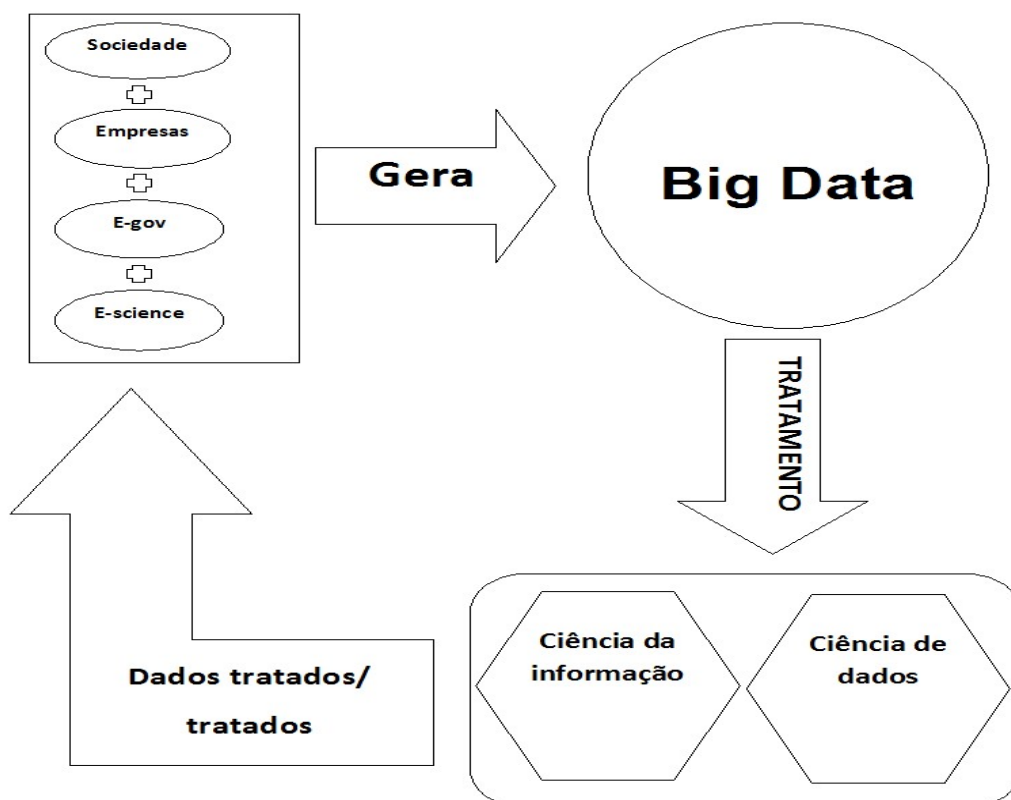
Muitas disciplinas científicas tornaram-se direcionadas a dados. Pesquisas como a *Large Synoptic Survey Telescope* (LSST) fará a varredura do céu de uma montanha no Chile, registrando 30 trilhões de *bytes* de imagens por dia, registrará o dobro do volume de dados gerado pelo satélite *Sloan Digital Sky* em suas Pesquisas. Astrônomos aplicarão um imenso poder de computação para analisar esses dados afim de sondar as origens do nosso universo. O *Large Hadron Collider* (LHC), um acelerador de partículas que vai revolucionar a nossa compreensão do funcionamento do universo, vai gerar 60 *terabytes* de dados por dia - 15 *petabytes* (15 milhões de *gigabytes*) por ano. Semelhantes projetos de *e-science* são propostas ou estão em andamento em uma ampla variedade de outras disciplinas, a partir da biologia, ciência ambiental e oceanografia. Esses projetos geram enormes conjuntos de dados onde a análise automatizada é necessária. Além disso, torna-se impraticável para replicar cópias nos locais de grupos de pesquisa individuais, para que os investigadores unifiquem os seus recursos para a construção de um grande centro de dados para que possam executar os programas de análise para todos os cientistas filiados. (BRYANT; KATZ; LAZOWSKA, 2008, não paginado, tradução nossa)

Azam (2014) acrescenta que as pesquisas demandam cada vez mais de análise de grandes volumes de dados, em adição aos métodos científicos necessários aos estudos em curso. Nem sempre estes cientistas possuem formação estatística ou computacional para tratarem o volume de dados crescente a cada pesquisa. Ao cientista de uma maneira geral, o fenômeno *Big Data* demanda conhecimentos em computação, estatística e matemática.

Na Figura 3, identifica-se o relacionamento existente entre o *Big Data*, *data science* e *e-science*, mapeado a partir da observação de autores que conceituam o *Big*

*Data*, como Taurion (2012b) e Dumbill et al. (2012), além de autores que apontam a necessidade de novas habilidades aos profissionais da área de informação, como Stryker (2014) e Garritano e Carlson (2009).

**FIGURA 3–** Relacionamento entre *Big Data*, *e-science* e *data science*



Fonte: elaborado pelo autor, 2017.

Os dados abertos e privados originados a partir da sociedade, empresas, governos e a ciência, constituem o fenômeno *Big Data*. Esses dados em grandes volumes podem ser diretamente tratados pela ciência de dados, ou diante da necessidade de classificação, disseminação e tratamento, podem ser tratados pela CI. Posteriormente, estes dados tornam-se analisados, classificados, tratados e disseminados. A partir deste ponto, esses dados estarão disponíveis para uso, possibilitando o reinício do processo.

A partir das relações observadas na Figura 3, torna-se necessário o entendimento do conceito de *data science* e *e-science*.

### 2.3.1 *Data science*

Provost e Fawcet (2013, p. 2) definem *data science* como:



Um conjunto de princípios fundamentais que suportam e orientam os princípios da extração de informações e conhecimento a partir dos dados. Possivelmente o conceito relacionado mais próximo com a ciência de dados seja a mineração de dados.

Entretanto, segundo Provost e Fawcet (2013) data science é muito mais que apenas mineração de dados. A estatística, métodos de visualização de dados, intuição, criatividade e conhecimento do assunto que está sob análise, constituem fatores fundamentais do pensamento analítico dos dados. Métodos utilizados para Big Data são de particular interesse à data science.

Josh Bloom, através de Azam (2014, não paginado, tradução nossa) acrescenta:

A primeira regra da ciência de dados é: "não pergunte como definir a ciência de dados". Assim diz Josh Bloom, professor da UC Berkeley da astronomia e um estudioso de destaque (PI) no Instituto Berkeley para a ciência de Dados (BIDS). Se esta abordagem parece problemática, isso é porque a data science é mais que uma filosofia interdisciplinar emergente, um amplo *modus operandi* que implica uma mudança cultural na comunidade acadêmica. O termo significa algo diferente para cada cientista de dados [...].

Os termos "data science" e "data scientist" têm despertado interesse ao segmento corporativo e suas atividades, que envolvam o tratamento estatístico de grandes volumes de dados, denominado *Big Data*, conforme destaca Soares (2013, online):

Com aumento do interesse das organizações pela implementação de projetos de *Big Data*, a profissão de cientista de dados ganhou importância e tornou-se uma carreira promissora, tendo sido considerada pela revista Harvard Business Review como o emprego mais sexy do século 21.

Ferreira (2015, online) complementa:

Na prática, o cientista de dados é o profissional responsável pelo desenvolvimento de algoritmos matemáticos e o alinhamento destes ao negócio para melhor desempenho das empresas. Com as plataformas de *Big Data*, as empresas são capazes de prever ou compreender alguns pontos como: tendências de mercado, repercussão de algum fato/campanha de marketing pelo mercado e comportamento do consumidor.

Ainda em Azam (2014), para o cientista de dados são requeridas técnicas que incluem processamento de sinais, modelos probabilísticos, aprendizagem de máquina, mineração de dados, reconhecimento e aprendizagem de padrões, *datawarehousing* (armazenamento de dados), engenharia de dados, modelos de incerteza, compressão de dados, inteligência artificial e computação de alta performance, que vem de áreas como a matemática, estatística e computação.

Neste contexto, Taurion (2012b, *online*) acrescenta:

O cientista de dados precisa de conhecimentos de estatística, matemática, entender do negócio e ter familiaridade com tecnologias e linguagens como *Hadoop* e *Pig*. Para os profissionais envolvidos com *Big Data* aparece um novo desafio que é a modelagem de dados não estruturados.

Lima (2015, *online*) comenta que o cientista de dados é um "profissional com conhecimento em áreas de exatas e tecnológicas, como estatística, computação e matemática".

Harbert (2013), corroborando com as definições de Azam (2014), Taurion (2013) e Lima (2015, *online*) para o cientista de dados, complementa ser necessária uma boa dose de curiosidade, devendo ser esta uma importante característica para este profissional.

### 2.3.2 *E-science*

Bohle (2013, *online*, tradução nossa) define *e-science* como:

A aplicação de tecnologia computacional para suportar a investigação científica moderna, incluindo preparação, experimentação, coleta de dados, disseminação dos resultados, armazenamento e acessibilidade dos materiais gerados através do processo científico.

Portanto a *e-science* utiliza de grandes quantidades de recursos computacionais para lidar com imensos conjuntos de dados, aplicando tecnologia de computação para preparação, experimentação, coleta de dados, divulgação de resultados, armazenamento e acessibilidade de todos os materiais gerados no processo da investigação. Visa apoiar cientistas com o desenvolvimento de métodos para análise de grandes quantidades de dados acessíveis através da internet, o chamado *Big Data*.

Alguns países possuem iniciativas voltadas a suportar as demandas específicas da *e-science*, como o Danish e-Infrastructure Cooperation da Dinamarca, que define:

*E-science* é um método de pesquisa que envolve a coleta, processamento e utilização de informações científicas na forma de dados. Métodos cada vez mais eficientes de coleta, geração, digitalização e subsequente armazenamento de dados estão abrindo oportunidades completamente novas para a realização de pesquisas em todas as áreas profissionais e acadêmicas. (DANISH E-INFRASTRUCTURE COOPERATION, 2017, *online*, tradução nossa)

No Quadro 1, são apresentadas, a título de exemplificação, algumas instituições que são dedicadas à *e-science* conforme o Danish e-infrastructure Cooperation (DeiC) e CNC:

QUADRO 1– Exemplos de instituições dedicadas à *e-science*

País	Instituição	Descrição
Brasil	O Serviço Experimental Computação em Nuvem para Ciência (SE-CNC)	O SE-CNC visa oferecer um serviço de armazenamento de dados em nuvem genuinamente nacional, principalmente para professores e pesquisadores. Para isso, o SE-CNC dispõe de servidores no Brasil e utiliza-se de tecnologias de código aberto. O serviço também prioriza o sigilo de dados, de forma que as chaves criptográficas são salvas em <i>hardware</i> seguro.
Dinamarca	Danish e-Infrastructure Cooperation (DeiC)	Tem o objetivo declarado de apoiar o desenvolvimento da <i>e-science</i> na Dinamarca, através da entrega de infraestruturas (informática, <i>data warehousing</i> , ligações de rede e serviços auxiliares), orientação e iniciativas a nível nacional. DeiC é uma unidade virtual subordinada ao Ministério Dinamarquês do Ensino Superior e da ciência. Sendo também o resultado de um acordo concluído entre as oito universidades dinamarquesas e a Agência Dinamarquesa para ciência, Tecnologia e Inovação.
Reino Unido	National e-Science Centre (NeSC)	O NeSC e o <i>e-Science Institute</i> da Universidade de Edimburgo tornaram-se importantes comunidades para discutir <i>e-science</i> .
Holanda	Netherlands eScience Center (NLeSC)	O NLeSC é encarregado de coordenar um programa científico colaborativo, trabalhando tanto com a academia como com a indústria.
Suécia	Swedish e-Science Research Centre (SeRC)	Este instituto é formado pelas universidades: Kungliga Tekniska Högskolan, Stockholms Universitet, Karolinska Institutet e Linköpings Universitet, com programas de investigação colaborativa e estratégica em <i>e-science</i> .
Estados Unidos	<i>E-science Institute</i>	Este instituto é composto por um grupo diversificado de indivíduos com formação em física, astronomia, bio-engenharia, bioinformática, técnicas de gestão de dados e informática, que atuam como grupos de pesquisas, ajudando aos pesquisadores a aplicar a tecnologia mais apropriada à sua pesquisa.

Fonte: adaptado de DANISH E-INFRASTRUCTURE COOPERATION (2017, *online*)<sup>1</sup>; COMPUTAÇÃO... (2017, *online*)<sup>2</sup> e E-SCIENCE INSTITUTE (2017, *online*)<sup>3</sup>.

Jankowski (2007, tradução nossa) cita alguns benefícios da *e-science* para a comunidade científica:

- a) colaboração internacional entre pesquisadores;
- b) o uso crescente de computadores interconectados de alta velocidade aplicando arquitetura distribuída;
- c) visualização de dados;
- d) desenvolvimento de ferramentas e procedimentos baseados na Internet;

1 Disponível em: <<https://vidensportal.deic.dk/eScience-centres?language=en:>>.

2 Disponível em: <<https://cnc.rnp.br/instituClonal/sobre.php>>.

3 Disponível em: <<http://esClence.washington.edu/>>.

- e) construção de estruturas organizacionais virtuais para a realização de pesquisas;
- f) distribuição eletrônica e publicação de pesquisas.

Segundo o mesmo autor, o termo *e-science* foi criado em 1999 por John Taylor, então diretor do Instituto de Ciência e Tecnologia do Reino Unido. Em 2001, foi criado o *National E-science Centre (NeSC)*, primeiro veículo de financiamento para projetos de *e-science* no Reino Unido, que destaca:

No futuro, *e-science* fará referência à ciência em grande escala que serão cada vez mais realizadas através de colaborações geradas a partir de diversas partes do mundo, realizadas pela Internet. Normalmente, uma característica de tais colaborações e empreendimentos científicos é que eles vão requerer acesso a coleções muito grandes de dados, recursos de computação em grande escala e visualização de alta performance para suportar os cientistas e usuários individuais. (NATIONAL E-SCIENCE CENTRE, 2001, *online*, tradução nossa)

Azam (2014) e Taurion (2013) vinculam o cientista de dados ao fenômeno *Big Data*. Para esses autores, o cientista de dados tem conhecimento em computação, estatística, matemática, bem como uma visão holística. Mediante observação dos autores citados, percebe-se que o cientista de dados pode atuar tanto na *data science* como na *e-science*, apoiado no fato em que as ferramentas a serem utilizadas são as mesmas, alterando-se apenas o contexto e o objetivo quanto ao desenvolvimento da atividade.

Patil e Davenport (2013, tradução nossa) citam que a formação dos cientistas de dados recebe destaque nas seguintes instituições americanas: instituições como Stanford, MIT, Berkeley, Harvard, Canegie Mellon e outras com currículo forte. De novo nos Estados Unidos, seriam Instituições como North Carolina State, UC Santa Cruz, University of Maryland, The University of Washington e UT Austin.

Portanto, observa-se que a relação existente entre *datascience*, *e-science* e o *Big Data* se constituem de forma que a primeira trata dos grandes volumes de dados, a segunda contribui com a geração de grandes volumes de dados, e por último, o *Big Data* é o somatório do grande volume de dados.

## 2.4 O fenômeno *Big Data* e a sociedade

Segundo Abreu (2015), o fenômeno *Big Data* envolve pessoas, empresas e instituições, de origens, culturas, legislações e valores diferentes. Algumas questões estão relacionadas a esse fenômeno, tais como:

- a) a privacidade dos dados, questionada por diversos segmentos;
- b) o valor monetário da informação e o benefício à pessoa que gerou tais dados;

- c) a quem deve ser atribuída a função de definir o que deverá ser salvo;
- d) métodos de visualização e tomadas de decisão sobre estes dados;
- e) concentração do “poder” de manipulação e o domínio dos mecanismos de operação do *Big Data* em um círculo reduzido de empresas e técnicos, o que pode fomentar diferenças econômicas e relações de submissão entre povos.

Lobo (2013) identifica que a sociedade se beneficia deste grande volume de dados através da disponibilidade da informação, e da qualidade dos serviços orientados por estas informações.

A colaboração de Ekbia et al. (2015, p. 1524, tradução nossa), é a de que essa sociedade se interessa pelo tema *Big Data*. Esses autores complementam que: “junto com o interesse acadêmico, profissionais em negócios, indústria e governo, descobriram que o *Big Data* oferece enormes oportunidades para o comércio, inovação e engenharia social”.

A partir desses autores, entende-se que a sociedade tem um papel importante quanto à geração e obtenção de informação, bem como apropriar-se do conhecimento tecnológico que suporta o *Big Data*.

## **2.5 O fenômeno *Big Data* e a Ciência da Informação**

Cavalcanti (2014, p. 34) comenta que “o *Big Data* não é uma tecnologia”. Corroborando com essa ideia, Saracevic (1996, 1999) e Taurion (2012b) destacam que pode ser traçado um paralelo entre a explosão informacional de onde se originou a CI e a que ocorre neste momento. Novamente os mecanismos existentes não são suficientes para armazenar e tratar o volume de dados gerados, obrigando a busca por novas ferramentas e metodologias.

A CI possui movimentos tecnológicos avançados, e interdisciplinaridade com as disciplinas ligadas à informação. Ela possui características que a diferencia das demais ciências, conforme mencionam Galvão e Borges (2000, p. 48):

[...] a Ciência da Informação não é uma ciência clássica, nem uma ciência básica. É uma ciência recursiva que se vale dos conhecimentos já existentes nas ciências (normais), dos avanços tecnológicos e suas possibilidades, bem como se define segundo os nichos de oportunidade (demandas sociais).

Souza, Almeida e Baracho (2013) comentam sobre o esvaziamento conceitual da CI, e ao mesmo tempo orientam acerca da necessidade de uma transformação conceitual a partir de sua interdisciplinaridade.

Ribeiro (2014, p. 96) destaca:

[...] é lícito supor que o profissional de informação deve refletir um pouco sobre como poderá se envolver nas discussões sobre o tema *Big Data*, pois afinal, o uso de dados e informação sempre foi objeto de estudo para a CI.

Os autores Souza, Almeida e Baracho (2013) e Ribeiro (2014), destacam a relação existente entre a CI e o fenômeno *Big Data*. Os mesmos ainda reconhecem a interdisciplinaridade dessa ciência e a necessidade de buscar novos conhecimentos e habilidades necessários a sua adequação ao estudo contemporâneo da informação.

Corroborando com os autores supracitados, Santos e Mannes (2009) esperam que o cientista da informação, seja um agente ativo no atendimento às demandas relacionadas à informação, moldando-se rapidamente diante das novas possibilidades que emergem a partir do avanço tecnológico e do estudo da informação.

Oard (2016) enfatiza que escolas de pós-graduação em CI perceberam a importância da abordagem tecnológica, a atenção aos novos suportes informacionais, aos grandes volumes de informação e aos novos atores na geração e consumo de informação. Ainda conclui, que o futuro da CI passa por um novo conceito interdisciplinar, o qual denomina-se *Information School (iSchool)*.

O que pode ser compreendido em Hey e Hey (2006), quando comentam sobre novos conhecimentos e habilidades demandadas pela ciência e também em Ribeiro (2014), conforme trecho supracitado. Oard (2016) reconhece a *iSchool* como evolução direta dos espaços de estudo da informação fortalecidos pela interdisciplinaridade, quando destaca a integração de “pessoas, informação e tecnologia”.

## **2.6 O cientista da informação no contexto do *Big Data***

Souza, Almeida e Baracho (2013, p. 171) comentam sobre a CI: “essa essência interdisciplinar exorta o cientista da informação a navegar nos espaços teóricos, adaptar-se aos contextos tecnológicos e reinventar-se continuamente”.

Ribeiro (2014, p. 102) destaca que:

A Ciência da Informação nasceu e foi gestada com o objetivo maior de apresentar solução para problemas ligados ao uso de dados e informação e como tal, tem um importante papel nos estudos que envolvem o tema *Big Data*.

Através de Oard (2016), reconhece-se que diante da tecnologia e volume informacional, a CI necessita mudar a forma de preparar o cientista da informação em um novo contexto. É reconhecido por Hey e Hey (2006) que o volume de dados científicos demandará novos conhecimentos e habilidades dos profissionais da informação.

Assim, é possível conceber que o fenômeno *Big Data* e suas novas definições demandarão do cientista da informação novos conhecimentos e conceitos, conforme preconiza Souza (2006).

O fenômeno *Big Data* é reconhecido pelos autores presente nesta pesquisa, ao mesmo tempo em que é compreendida a necessidade de “atualização” da CI, quando essa se apresenta de forma interdisciplinar frente às outras ciências. Observa-se, a importância em rever o perfil e as habilidades dos profissionais da informação diante das demandas desse fenômeno e da sociedade.

### 3 METODOLOGIA

Esta pesquisa tem caráter indutivo, conforme aponta Marconi e Lakatos (2007), diante da observação sistemática e a classificação dos fenômenos selecionados. Ela está fundamentada nas circunstâncias e frequências em que as publicações sobre esse tema ocorreram e a medição de suas diferentes intensidades.

#### 3.1 Caracterização da pesquisa

Do ponto de vista dos procedimentos técnicos e sob a perspectiva da coleta de dados, esta pesquisa assume características de pesquisa bibliográfica por ser elaborada a partir de material já publicado, como livros, artigos em periódicos impressos, teses, dissertações, revistas e periódicos digitais, conforme Ribas e Fonseca (2008, p. 6): "A pesquisa bibliográfica abrange toda teoria já tornada pública em relação ao tema de estudo, desde publicações avulsas, jornais, revistas, livros, pesquisas, monografias, teses, etc., isto é, envolve teorias que já receberam um tratamento científico".

O fenômeno *Big Data* e seus assuntos correlatos constituem tema recente na ciência, o que ocasiona escassez de artigos e material científico em grande profundidade, ou que trate exaustivamente o tema, principalmente sob a ótica da CI, denotando, portanto, a necessidade de ampla busca por fontes de conteúdo científico relacionado.

Deste modo, conforme Marconi e Lakatos (2001), este trabalho possui caráter de pesquisa bibliográfica, considerando a busca por informações que conduzissem ao mapeamento de abordagens voltadas ao fenômeno *Big Data*, através de artigos sobre assunto dentro da CI.

A população ou universo da pesquisa é constituído pelas publicações no idioma inglês, que abordam os termos *Big Data*, contidos nas Bases LISA e LISTA, e pelas publicações no idioma português, que abordam os termos *Big Data* em seus artigos contidos nas bases BRAPCI e PERI. Prodanov e Freitas (2013, p. 98), assim definem e relacionam universo da pesquisa e população "População (ou universo da pesquisa) é a totalidade de indivíduos que possuem as mesmas características definidas para um determinado estudo". Considerou-se também a busca pelo tema *Big Data* em publicações relacionadas a *e-science*, *escience* e *datascience*. A escolha dessas bases se deu pelo fato de as mesmas estarem orientadas para a área da CI, tema tratado na pesquisa.

A classificação desta pesquisa, segundo Gil (2002), quanto aos meios é empírica, pois não existe interferência do pesquisador, considerando que os resultados a serem obtidos, terão origem a partir das observações das produções científicas nas bases a serem pesquisadas.



Quanto aos fins, esta pesquisa se classifica como descritiva, pois tem como objetivo a descrição de ocorrência e frequência dos dados obtidos nas pesquisas realizadas nas bases. Para Gil (2002, p. 42), “as pesquisas descritivas têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis”. E complementa que uma das características da pesquisa descritiva “está na utilização de técnicas padronizadas de coleta de dados”. Mattar (1996) confirma quando diz que na pesquisa descritiva é aquela em o pesquisador possivelmente irá analisar seus dados de forma indutiva.

As técnicas de estatística descritiva aplicadas nesta pesquisa, objetivaram demonstrar e analisar os resultados obtidos a partir dos recursos do *software Microsoft Excel*.

### 3.2 Coleta de dados

Os dados coletados foram artigos sobre o tema *Big Data* nas bases de dados selecionadas: BRAPCI e PERI para identificar a literatura nacional e as bases LISA e LISTA para a literatura internacional. A escolha destas bases, deu-se pela sua especialidade em temas relacionados à CI, pela grande representatividade das mesmas e pela observação de uma grande sobreposição no conteúdo das bases de dados especializadas na CI. A opção pela filtragem de somente artigos, deu-se pelo fato de representarem a pesquisa mais consolidada e de maior qualidade.

O período de realização do levantamento de dados para o desenvolvimento desta pesquisa ocorreu entre julho e agosto de 2016.

A coleta e seleção dos dados ocorreram em seis etapas, a saber:

Na primeira etapa foram coletados os dados dos artigos com os termos *Big Data*, *e-science* e *datascience* existentes nos campos Título, *Abstract* e *Subject* das bases BRAPCI, PERI, LISA e LISTA. De posse desse conjunto de dados sobre as publicações, procedeu-se com a eliminação dos itens repetidos, dando origem ao banco de dados de suporte a esta pesquisa.

Com base nos primeiros resultados analisados, foi elaborado um código de classificação.

Na segunda etapa, o banco de dados obtido na etapa anterior foi submetido ao processo de complementação das informações, através da visita aos websites das universidades/instituições e programas aos quais os autores das produções bibliográficas identificada estavam afiliados. Destaca-se que cerca de 80% das informações sobre autores e suas instituições e programas de origem foram extraídos diretamente das bases de dados pesquisadas.

Na terceira etapa, o pesquisador realizou a leitura do abstract de todos os artigos constantes na base de dados e realizou-se a classificação e/ou qualificação das produções.

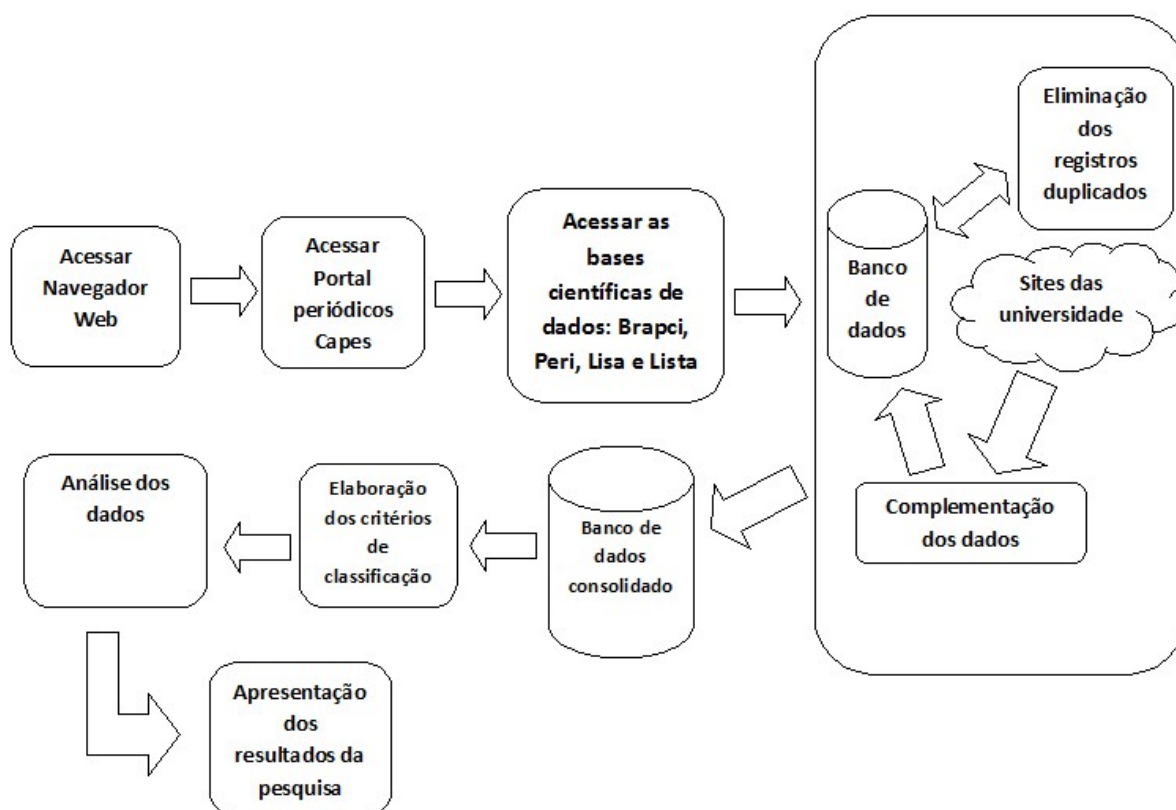
Na quarta etapa, uma cópia do banco de dados sem as classificações foi submetida a dois pares, para que os mesmos realizassem o procedimento classificatório descrito no item anterior, sem acesso ao resultado obtido acima. Somente um par retornou seus resultados obtidos para esta pesquisa.

Na quinta etapa, foi realizado o confronto entre as classificações do pesquisador e do par, afim de avaliar o grau de coerência entre as mesmas.

Na sexta etapa foi realizado um debate entre o pesquisador e o par em torno das divergências das classificações. Ao final obteve-se uma avaliação unânime entre os atores envolvidos.

A Figura 4 apresenta a sequência de desenvolvimento desta pesquisa.

**FIGURA 4– Fluxograma de desenvolvimento da pesquisa**



Fonte: elaborado pelo autor, 2017.

A partir do fluxograma de desenvolvimento da pesquisa, foi consolidado um banco de dados no *software* Microsoft Excel, com as informações sobre a produção científica da CI, abordando a temática central desta pesquisa que é o *Big Data*. Os dados consolidados nesse banco de dados são:

- a) base de dados LISA e LISTA:
  - título;
  - autor;
  - abstract;
  - palavras chave;
  - ano da publicação;
  - universidade de afiliação;
  - *Big Data* Como (Classificação do artigo);
  - Objeto de Pesquisa como;
  - programa;
  - país de origem;
- b) base de dados BRAPCI e PERI:
  - título;
  - autor;
  - abstract;
  - palavras chave;
  - ano da publicação;
  - universidade de afiliação;
  - *Big Data* Como (Classificação do artigo);
  - Objeto de Pesquisa como;
  - país de origem.

### **3.3 Instrumentos de pesquisa**

Os instrumentos de pesquisa necessários foram baseados no uso dos recursos tecnológicos e no método de obtenção dos dados, conforme descritos a seguir.

#### **3.3.1 Recursos tecnológicos**

Os recursos tecnológicos necessários para o desenvolvimento desta pesquisa, composição da amostragem e coleta de dados estão apresentados no Quadro 2.

QUADRO 2 – Recursos tecnológicos

Recurso	Objetivo
Microcomputador, Notebook ou Tablet, equipado com navegador web.	Executar os aplicativos e armazenar as informações obtidas na pesquisa.
Acesso à internet.	Acessar o ambiente que contém as bases objeto da pesquisa.
Usuário e senha para acesso à Base Portal Periódicos Capes.	Acessar as bases de dados do Portal.
Acesso à base BRAPCI (produção no idioma em português).	Extrair informações sobre <i>Big Data</i> , e-Science e e-science, de acordo com o filtro pré-estabelecido.

Fonte: elaborado pelo autor, 2017.

### 3.3.2 Métodos de obtenção dos dados

O método utilizado para a obtenção dos dados foi o bibliométrico, que conforme Araújo (2007), consiste em uma técnica de distribuição que informa o número de autores, publicações, universidades, países, dentre outras variáveis, que existem em cada categoria de busca, como utilidade, produção, frequência e o que mais for necessário saber. Para Vanti (2002) os dados foram coletados quantitativamente, pois tratam da identificação do conhecimento, a frequência desses nas bases pesquisadas e suas análises, relacionando-os na pesquisa às categorias: universidades/instituições, programas, países, autores e anos das publicações.

### 3.4 Análise de dados

Com a identificação das publicações relacionadas ao fenômeno *Big Data*, presentes nas bases pesquisadas, constatou-se a ocorrência dessas publicações no período compreendido entre 2010 e 2016. Posteriormente, a partir do banco de dados consolidado, foram realizadas as seguintes etapas, para a análise do mesmo:

Identificação dos Programas de origem dos autores: utilizando como chave de pesquisa os campos autor e universidade, foi pesquisado nos sites das universidades, *blogs* ou redes sociais, a qual programa o autor da publicação científica está afiliado. O dado foi registrado manualmente no campo "Programa" do banco de dados, no registro referente ao artigo em questão.

Análise de conteúdo: o conteúdo das publicações foi classificado através da análise de conteúdo. Foi realizada a leitura sistemática dos campos chave da pesquisa, título, *abstract* e *subject*, tendo por finalidade a análise de como o termo *Big Data* se apresenta no texto. Diante da análise, o artigo foi classificado conforme Quadro 3.

QUADRO 3 – Classificação das abordagens ao *Big Data*

Classificação	Sigla	Significado
Objeto de Pesquisa	OP	O artigo relata pesquisa tendo como objeto o <i>Big Data</i> Caso seja OP, deve ser qualificado: REC - A pesquisa é relacionada à recuperação da informação DIS - A pesquisa é relacionada à disseminação da informação; TRA - A pesquisa é relacionada ao tratamento da informação; INF - A pesquisa é relacionada à Infraestrutura.
Assunto Secundário	AS	O artigo relata pesquisa ou evento, onde o <i>Big Data</i> é assunto secundário.
Reflexão	RE	O autor faz uma reflexão sobre o tema <i>Big Data</i> , sem se tratar de uma pesquisa.

Fonte: elaborado pelo autor, 2017.

- a) Objeto de Pesquisa (OP): para os fins de classificação desta pesquisa, OP diz respeito ao artigo em que o *Big Data* é o tema principal do artigo. Essa classificação é subdividida em quatro subcategorias, aqui denominadas de “qualificação”:
- b) Recuperação da Informação (REC): nessa qualificação, inserem-se artigos em que o *Big Data* é OP e diz respeito à recuperação da informação. Por exemplo, em situações em que informações são recuperados a partir de dados históricos, como comportamento de uma população ou ciclo de determinada doença;
- c) Disseminação da Informação (DIS): nessa qualificação, são introduzidos artigos em que o *Big Data* é OP e corresponde à disseminação da informação. Artigos que retratem a criação de aplicativos, *frameworks* ou algoritmos que tornem a informação disponível, acrescenta-se aqui;
- d) Tratamento da Informação (TRA): nessa qualificação, são inseridos artigos em que o *Big Data* é OP e está relacionado ao tratamento da informação. A criação de aplicativos, algoritmos e *frameworks* com a finalidade de tratamento da informação, ações de obtenção, processamento e transformação da informação também se inserem aqui;
- e) Infraestrutura (INF): nesta qualificação, adicionam-se artigos em que o *Big Data* é OP e refere-se à infraestrutura necessária ao processamento de *Big Data*. Artigos que retratem situações e estudos relacionados a bancos de dados, balanceamento de carga de processamento, ambientes computacionais, computação em nuvem, redes de dados e tecnologias relacionadas ao *Big Data* são introduzidas aqui.

- f) Assunto Secundário (AS): nessa classificação, inserem-se artigos em que o *Big Data* é item secundário, a pesquisa não está relacionada diretamente a ele. Também são inseridos aqui, artigos em que o *Big Data* é um entre vários temas abordados, artigos que retratem eventos, ou que o termo *Big Data* conste em seu título ou *abstract*, palavras-chave ou *subject* e não façam parte de seu *abstract*.
- g) Reflexão (RE): nessa classificação, inserem-se artigos em que são realizados reflexões e comentários sobre o *Big Data*, ou seus impactos, consequências, ou a manifestação de opinião sobre o fenômeno, sem que sejam realizados estudos específicos sobre o *Big Data*.

Para a análise conclusiva da pesquisa, foram utilizadas as parametrizações abaixo apresentadas e desenvolvidas pelo pesquisador, tendo como objetivo elaborar as classificações dos campos como:

O campo **Big Data Como** responde como a CI tem abordado o tema *Big Data* em seus programas. Caso ocorra a maioria dos artigos na categoria OP, isso reporta que a CI tem realizado pesquisas sobre *Big Data* como tema central. A categoria AS, demonstra o *Big Data* é citado no artigo sem que seja o tema central. A categoria RE, significa que o artigo é apenas uma reflexão sobre o tema *Big Data*.

O campo **Objeto de Pesquisa Como** é preenchido sempre que o campo anterior receber a classificação de OP. Esse funciona como uma subcategorização que indica se o objeto da pesquisa está relacionado a REC, DIS, TRA ou INF.

O Campo **Universidade** responde em qual universidade ou instituição o autor publicou ou está afiliado.

O Campo **Programa** responde em qual Programa o autor publicou.

O Campo **País** responde em qual o país o autor publicou.

O Campo **Autor** demonstra quais são os autores que publicaram nesta temática.

Nesta pesquisa, será considerado como autor, sempre o primeiro autor ou o autor principal. Em artigos com mais de um autor, a informação dos demais será descartada.

O Campo **Ano** demonstra a produtividade da CI ao longo dos tempos.

## 4 APRESENTAÇÃO DOS RESULTADOS

Para a análise observou-se que os dados obtidos a partir das bases nacionais, PERI e BRAPCI, deveriam ser analisados de maneira distinta, de modo a evitar distorções diante do comportamento dos artigos em função do idioma de sua publicação.

Os resultados e análises das pesquisas realizadas nas bases LISA, LISTA, PERI e BRAPCI, acerca do tema *Big Data* compreendidos entre os anos de 2010 a 2016, são apresentados abaixo.

### 4.1 Análise da produção científica nas bases LISA e LISTA

TABELA1– Países com artigos relacionados ao *Big Data*

(Continua)

País	Total AS	Total REC	Total OP	DIS	INF	REC	TRA	Total	Total acumulado	% sobre publicações	% acumulado
Estados Unidos	116	81	9	2	-	-	7	206	206	46,23	46,23
Reino Unido	33	18	3	1	-	1	1	54	260	12,11	58,34
China	21	6	10	-	1	6	3	37	297	8,3	66,64
Canadá	9	11	1	-	1	-	-	21	318	4,72	71,36
Coréia do Sul	11	3	1	-	-	1	-	15	333	3,36	74,72
Alemanha	4	3	6	-	1	2	3	13	346	2,91	77,63
Itália	5	2	6	-	2	1	3	13	359	2,91	80,54
Espanha	5	3	1	-	-	-	1	9	368	2,02	82,56
Holanda	7	2	-	-	-	-	-	9	377	2,02	84,58
Austrália	4	3	1	-	-	1	-	8	385	1,79	86,37
Dinamarca	2	5	1	-	-	-	1	8	393	1,79	88,16
França	5	-	1	-	1	-	-	6	399	1,36	89,52
Nova Zelândia	5	-	1	-	-	-	1	6	405	1,35	90,87
Índia	5	-	-	-	-	-	-	5	410	1,12	91,99
Malta	-	3	-	-	-	-	-	3	413	0,66	92,65
Polônia	3	-	-	-	-	-	-	3	416	0,66	93,31
Suíça	2	-	1	-	-	1	-	3	419	0,66	93,97

TABELA 1 – Países com artigos relacionados ao *Big Data*

(Conclusão)

País	Total AS	Total REC	Total OP	DIS	INF	REC	TRA	Total	Total acumulado	% sobre publicações	% acumulado
Hungria	2	-	-	-	-	-	-	2	421	0,45	94,42
Argélia	-	1	1	-	-	-	1	2	423	0,45	94,87
Áustria	1	-	1	-	-	-	1	2	425	0,45	95,32
Grécia	1	1	-	-	-	-	-	2	427	0,45	95,77
Irã	1	-	1	-	1	-	-	2	429	0,45	96,22
Irlanda	2	-	-	-	-	-	-	2	431	0,45	96,67
Israel	2	-	-	-	-	-	-	2	433	0,45	97,12
Malásia	1	1	-	-	-	-	-	2	435	0,45	97,57
Noruega	2	-	-	-	-	-	-	2	437	0,45	98,02
Chile	-	-	1	-	-	-	1	1	438	0,22	98,24
Finlândia	1	-	-	-	-	-	-	1	439	0,22	98,46
Japão	-	-	1	-	1	-	-	1	440	0,22	98,68
Letônia	1	-	-	-	-	-	-	1	441	0,22	98,9
Marrocos	-	1	-	-	-	-	-	1	442	0,22	99,12
Palestina	1	-	-	-	-	-	-	1	443	0,22	99,34
Portugal	1	-	-	-	-	-	-	1	444	0,22	99,56
Suécia	1	-	-	-	-	-	-	1	445	0,22	99,78
Turquia	1	-	-	-	-	-	-	1	446	0,22	100
	<b>255</b>	<b>144</b>	<b>47</b>	<b>3</b>	<b>8</b>	<b>13</b>	<b>23</b>	<b>446</b>		<b>100</b>	

Fonte: elaborado pelo autor, 2017.

De acordo com a Tabela 1, identificaram-se 446 produções científicas com origem em 35 países. Constata-se que sete países são responsáveis por 80,54% ou 359 publicações, demonstrando polarização na produção científica analisada no universo desta pesquisa. Destacam-se neste contexto, Estados Unidos, Reino Unido e China, que conjuntamente são responsáveis por 66,64% ou 297 publicações.

Verifica-se que 28 países, o equivalente a 80,00% dos países do universo da pesquisa, publicaram de um a dez artigos. Esses países publicaram ao todo 87 artigos, o equivalente a 19,40% dos artigos do universo da pesquisa: Espanha e Holanda publicaram nove artigos cada; Austrália e Dinamarca publicaram oito artigos cada; França e Nova



Zelândia publicaram seis artigos cada; a Índia publicou cinco artigos; Malta, Polônia e Suíça publicaram três artigos cada; Argélia, Áustria, Grécia, Hungria, Irã, Irlanda, Israel, Malásia e Noruega, publicaram dois artigos cada; Chile, Finlândia, Japão, Letônia, Marrocos, Palestina, Portugal, Suécia e Turquia publicaram um artigo cada.

Coréia do Sul, o equivalente a 2,85% dos países que publicaram artigos sobre *Big Data*, publicou quinze artigos, o equivalente a 3,36% do total dos artigos desta pesquisa;

Alemanha, o equivalente a 2,85% dos países que publicaram artigos sobre *Big Data*, produziu treze artigos, o equivalente a 2,91% do total dos artigos desta pesquisa;

A Itália, que equivale a 2,85% dos países que publicaram artigos sobre *Big Data*, produziu treze artigos, o equivalente a 2,91% do total dos artigos do universo da pesquisa

Canadá, o equivalente a 2,85% dos países que publicaram artigos sobre *Big Data*, publicou 21 artigos, o equivalente a 4,70% do total dos artigos desta pesquisa;

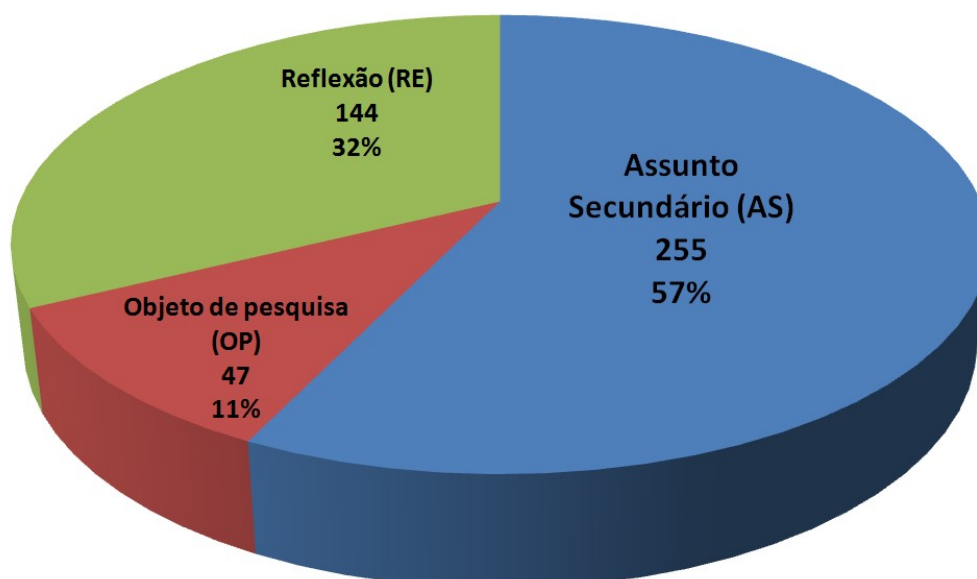
China, o equivalente a 2,85% dos países que publicaram artigos sobre *Big Data*, publicou 37 artigos, o equivalente a 8,30% do total dos artigos desta pesquisa;

Reino Unido, o equivalente a 2,85% dos países que publicaram artigos sobre *Big Data*, publicou 55 artigos, 12,30% do universo de artigos da pesquisa.

Estados Unidos, o equivalente a 2,85% % dos países que publicaram artigos sobre *Big Data*, publicaram 206 artigos, 46% do universo de artigos da pesquisa.

O GRÁFICO 1 apresenta o comportamento dos artigos quanto à classificação:

**GRÁFICO 1 – Mapa geral da classificação dos artigos**  
Publicações relacionadas ao Big data - bases Lisa e Lista - 2010 a 2016



Fonte: elaborado pelo autor, 2017.

Conforme o GRÁFICO 1, percebe-se que o volume de publicações classificadas como AS e RE indicam que a CI pode não estar abordando o *Big Data* como tema de pesquisa em suas publicações. Verifica-se que 255 artigos foram classificados na categoria *Big Data* como AS, o equivalente a 57% do universo de artigos da pesquisa. Essa categoria indica que o *Big Data* é citado no artigo como tema secundário.

144 artigos foram classificados na categoria *Big Data* como RE, o equivalente a 32,% dos artigos do universo de artigos da pesquisa. Essa categoria indica que no artigo são realizadas reflexões sobre o *Big Data*.

47 artigos foram classificados na categoria *Big Data* como OP, o equivalente a 11% do universo dos artigos da pesquisa. Essa categoria indica que nestes artigos o *Big Data* é o tema central da pesquisa.

O Quadro 4, exemplifica artigos classificados como AS:

QUADRO 4 – Três exemplos de artigos classificados como Assunto Secundário

(Continua)

Título	Autor	Ano	Referência
Acting on Big Data: A Data Scientist Role for Info Pros	Affelt, Amy	2014	AFFELT, Amy. Acting on Big Data: A Data Scientist Role for Info Pros. <b>Online Searcher</b> , v. 38, n. 5, p. 10-14, 2014.
<b>Abstract</b>			
How can information professionals become 'accidental' data scientists? How can we learn the skills that hiring managers are looking for, in addition to parlaying our existing strengths into the arena of Big Data - or, just 'data,' which librarians have been expert at working with since the dawn of librarianship? Although 'Big Data' is a new term, it is not a new concept for information professionals. Data has always been 'big' in our organizations-many of us used to crash Lotus 1-2-3 back in the 1980s when we ran out of cells to fit in all of the numbers in our analysis. it is important to know what people mean when they talk about Big Data. Gaining a basic understanding of the skills required and the processes used makes info pros start speaking the language of the Big Data project teams in our organizations. This article discusses training in data science, user friendly programming, using a checklist, storytelling with data, and tools for data analysis. Adapted from the source document.			
Título	Autor	Ano	Referência
Data Analytics: The Next Big Thing in Information	Crowe, June	2013	ARNOLD, Stephen E. Analytics for Pain-Free Number-Crunching. <b>Information Today</b> , v. 30, n. 1, p. 18-19, 2013.

## QUADRO 4 – Três exemplos de artigos classificados como Assunto Secundário

(Conclusão)

<b>Abstract</b>			
<p>Information is now available in an overabundance, so much so, that distinguishing the noise from the signal has become very problematic. In the past, the collection and storage of information was the primary issue. Currently, there are massive amounts of data both structured and unstructured, that need to be analyzed in an iterative, as well as in a time sensitive manner. In response to this need, data analytical tools and services have emerged as a means to solve this problem. Grey literature repositories, libraries, and information centers are well positioned to take advantage of these new tools and services. The current trend is to make grey literature more easily discoverable, accessible, and with the new data analytical tools and services, more easily analyzed. The intent of our survey of the Grey Literature community was to provide a snapshot of the Community's use, planned use, and knowledge of data analytical tools/services for big data as it affects grey literature. The survey summary that follows indicates where the Community currently stands in regards to the use of data analytical tools and services. Adapted from the source document.</p>			
Título	Autor	Ano	Referência
A review of large-scale 'How much information?' inventories: variations, achievements and challenges.	Hilbert, Martin	2015	HILBERT, Martin. A Review of Large-Scale "How Much Information?" Inventories: Variations, Achievements and Challenges. <b>Information Research: An International Electronic Journal</b> , v. 20, n. 4, <i>online</i> , 2015.
<b>Abstract</b>			
<p>Introduction. Pressed by the increasing social importance of digital information, including the current attention given to the 'big data paradigm', several research projects have taken up the challenge to quantify the amount of technologically mediated information. Method. This meta-study reviews the eight most important inventories in a descriptive and comparative manner, focusing on methodological differences and challenges. Results. It shows that approaches differ in terms of scope and research focus. This leads to different answers to the question of 'how much information?'. Differences include how the information realm is conceptualised (e.g., in terms of stocks or flows, or in terms of creation or consumption, etc.); differences in the unit of measurement (words, bits, minutes, etc.); varying geographic and temporal scopes; and diverse additional attributes that highlight complementary aspects of the amount of information (e.g., the kind of technology, the sort of content, the type of user sector, etc.). Conclusion. The study reveals how different answers to the 'How much information?' question hinges upon the particular question on the researchers' mind and on the subsequent methodological choices. Differences in findings stem from different research interests. The review ends with a discussion of the remaining theoretical and practical challenges. [ABSTRACT FROM AUTHOR]</p>			

Fonte: elaborado pelo autor, 2017.

No Quadro 5, exemplos de artigos classificados na categoria *Big Data* como RE.

QUADRO 5 – Três exemplos de artigos classificados como Reflexão

Título	Autor	Ano	Referência
What Big Data means to me.	Bourne, Philip E	2013	BARNAGHI, Payam; SHETH, Amit; HENSON, Cory. From data to actionable knowledge: Big data challenges in the web of things [Guest Editors' Introduction]. <b>IEEE Intelligent Systems</b> , v. 28, n. 6, p. 6-11, 2013.
<b>Abstract</b>			
The author reflects on Big Data. He mentions its capability in changing business and draw stakeholders through conferences, workshops and fundraising events. He also notes its definition which involves the integration of large datasets as well as points out its representation in digital business growth.			
Título	Autor	Ano	Referência
Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship	Bruns, Axel	2013	BRUNS, Axel. Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship. <b>First Monday</b> , v. 18, n. 10, 2013.
<b>Abstract</b>			
The promise of 'big data' has generated a significant deal of interest in the development of new approaches to research in the humanities and social sciences, as well as a range of important critical interventions which warn of an unquestioned rush to 'big data'. Drawing on the experiences made in developing innovative 'big data' approaches to social media research, this paper examines some of the repercussions for the scholarly research and publication practices of those researchers who do pursue the path of 'big data' -- centric investigation in their work. As researchers import the tools and methods of highly quantitative, statistical analysis from the 'hard' sciences into computational, digital humanities research, must they also subscribe to the language and assumptions underlying such 'scientificity'? If so, how does this affect the choices made in gathering, processing, analysing, and disseminating the outcomes of digital humanities research? In particular, is there a need to rethink the forms and formats of publishing scholarly work in order to enable the rigorous scrutiny and replicability of research outcomes? Adapted from the source document.			
Título	Autor	Ano	Referência
Exploring big data's strategic consequences	Bhimani, Alnoor	2015	BHIMANI, Alnoor. Exploring big data's strategic consequences. <b>Journal of Information Technology</b> , v. 30, n. 1, p. 66-69, Mar.2015.
<b>Abstract</b>			
The article reports that big data is used by an increasing number of organizations in combination with established data sources and affects their managerial decisions and actions. For many business enterprises, it creates asset value through the storage of information whose potential grows by virtue of its volume and from results of its examination			

Fonte: elaborado pelo autor, 2017.

No Quadro 6 exemplifica-se artigos classificados na categoria *Big Data* como OP.

QUADRO 6 – Exemplo de artigos classificados como Objeto de Pesquisa

(Continua)

Qualificação	Título	Autor	Ano	Referência
TRA	Knowledge discovery from social media using big data-provided sentiment analysis (SoMABiT).	Bohlouli, Mahdi	2015	BOHLOULI, Mahdi et al. Knowledge discovery from social media using big data-provided sentiment analysis (SoMABiT). <b>Journal of Information Science</b> , v. 41, n. 6, p. 779-798, 2015.

#### Abstract

In today's competitive business world, being aware of customer needs and market-oriented production is a key success factor for industries. To this aim, the use of efficient analytical algorithms ensures better understanding of customer feedback and improves the next generation of products. Accordingly, the dramatic increase in the use of social media in daily life provides beneficial sources for market analytics. Yet how traditional analytic algorithms and methods can be scaled up for such disparate and multistructured data sources is a major challenge. This paper presents and discusses the technological and scientific focus of SoMABiT as a social media analysis platform using big data technology. Sentiment analysis has been employed in order to discover knowledge from social media. The use of MapReduce and the development of a distributed algorithm towards an integrated platform that can scale for any data volume and provide social media-driven knowledge is the main novelty of the proposed concept in comparison to the state-of-the-art technologies. [ABSTRACT FROM AUTHOR]

Qualificação	Título	Autor	Ano	Referência
DIS	The center for causal discovery of biomedical knowledge from big data.	Cooper, Gregory F.	2015	COOPER, Gregory F. et al. The center for causal discovery of biomedical knowledge from big data. <b>Journal of the American Medical Informatics Association</b> , v. 22, n. 6, p. 1132-1136, Nov. 2015.

#### Abstract

The Big Data to Knowledge (BD2K) Center for Causal Discovery is developing and disseminating an integrated set of open source tools that support causal modeling and discovery of biomedical knowledge from large and complex biomedical datasets. The Center integrates teams of biomedical and data scientists focused on the refinement of existing and the development of new constraint-based and Bayesian algorithms based on causal Bayesian networks, the optimization of software for efficient operation in a supercomputing environment, and the testing of algorithms and software developed using real data from 3 representative driving biomedical projects: cancer driver mutations, lung disease, and the functional connectome of the human brain. Associated training activities provide both biomedical and data scientists with the knowledge and skills needed to apply and extend these tools. Collaborative activities with the BD2K Consortium further advance causal discovery tools and integrate tools and resources developed by other centers. [ABSTRACT FROM AUTHOR]

QUADRO 6 – Exemplo de artigos classificados como Objeto de Pesquisa

(Conclusão)

Qualificação	Título	Autor	Ano	Referência
INF	Recent Developments on Security and Reliability in Large-Scale Data Processing with MapReduce	Esposito, Christian	2016	ESPOSITO, Christian; FICCO, Massimo. Recent Developments on Security and Reliability in Large-Scale Data Processing with MapReduce. <b>International Journal of Data Warehousing and Mining (IJDWM)</b> , v. 12, n. 1, p. 49-68, 2016.

#### Abstract

The demand to access to a large volume of data, distributed across hundreds or thousands of machines, has opened new opportunities in commerce, science, and computing applications. MapReduce is a paradigm that offers a programming model and an associated implementation for processing massive datasets in a parallel fashion, by using non-dedicated distributed computing hardware. It has been successfully adopted in several academic and industrial projects for Big Data Analytics. However, since such analytics is increasingly demanded within the context of mission-critical applications, security and reliability in MapReduce frameworks are strongly required in order to manage sensible information, and to obtain the right answer at the right time. In this paper, the authors present the main implementation of the MapReduce programming paradigm, provided by Apache with the name of Hadoop. They illustrate the security and reliability concerns in the context of a large-scale data processing infrastructure. They review the available solutions, and their limitations to support security and reliability within the context MapReduce frameworks. The authors conclude by describing the undergoing evolution of such solutions, and the possible issues for improvements, which could be challenging research opportunities for academic researchers.

Qualificação	Título	Autor	Ano	Referência
INF	Single-pass and linear-time k-means clustering based on MapReduce.	Shahrivari, Saeed	2016	SHAHRIVARI, Saeed; JALILI, Saeed. Single-pass and linear-time k-means clustering based on MapReduce. <b>Information Systems</b> , v. 60, p. 1-12, 2016.

#### Abstract

In recent years, k-means has been fitted into the MapReduce framework and hence it has become a very effective solution for clustering very large datasets. However, k-means is not inherently suitable for execution in MapReduce. The iterative nature of k-means cannot be modeled in MapReduce and hence for each iteration of k-means an independent MapReduce job must be executed and this results in high I/O overhead because in each iteration the whole dataset must be read and written to slow disks. We have proposed a single-pass solution based on MapReduce called mrk-means which uses the reclustering technique. In contrast to available MapReduce-based k-means implementations, mrk-means just reads the dataset once and hence it is several times faster. The time complexity of mrk-means is linear which is lower than the iterative k-means. Due to usage of k-means++ seeding algorithm, mrk-means results in clusters with higher quality, too. Theoretically, the results of mrk-means are  $O(\log 2 k)$  - competitive to optimal clustering in the worst case, considering k as the number of clusters. During our experiments which were done on a cluster of 40 machines running the Hadoop framework, mrk-means showed both faster execution times, and higher quality of clustering results compared to available MapReduce-based and stream-based k-means variants. [ABSTRACT FROM AUTHOR]

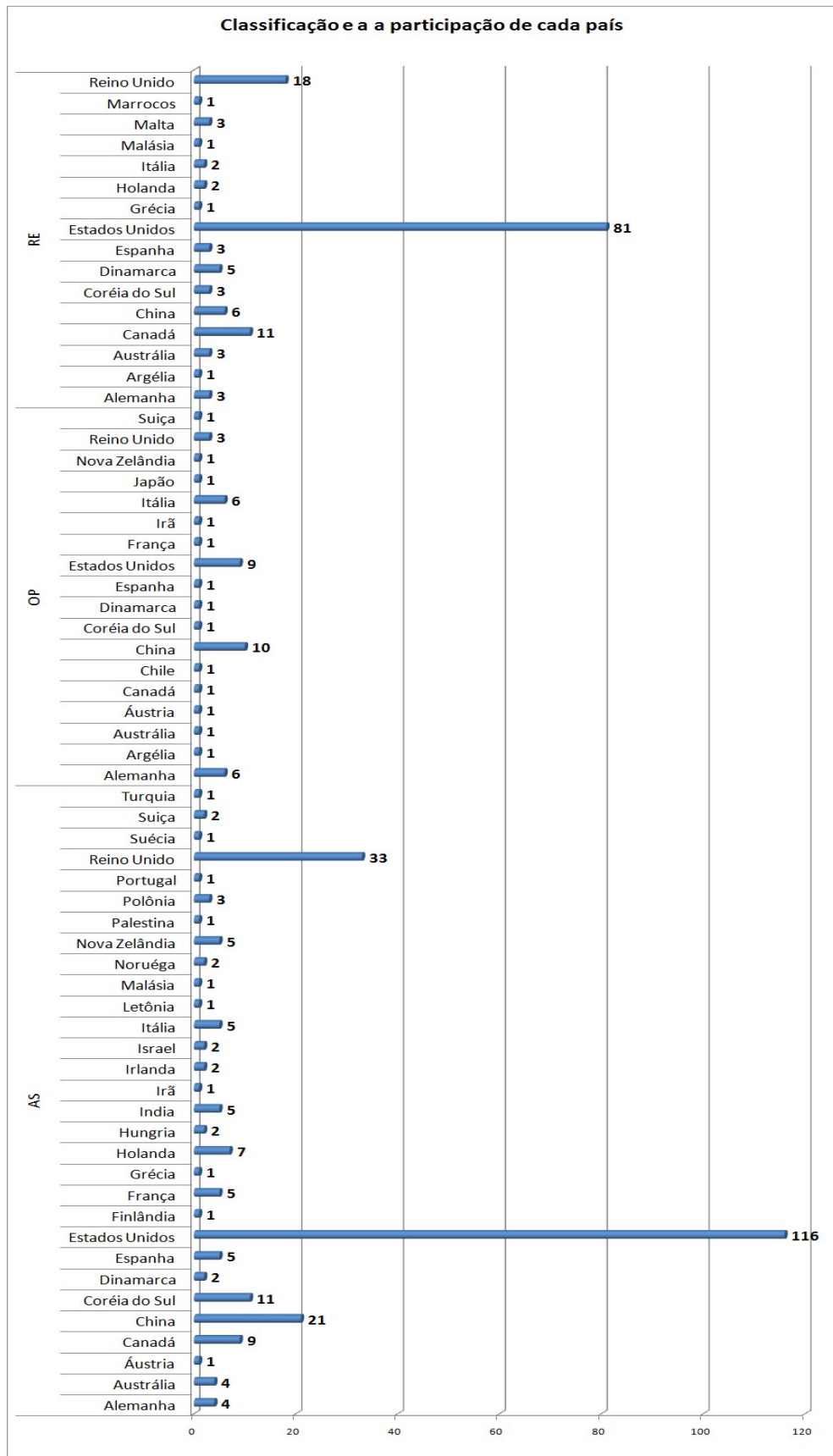
Fonte: elaborado pelo autor, 2017.

No Quadro 6 demonstra-se um exemplo para cada qualificação dada aos artigos que receberam a classificação OP: TRA, DIS, REC e INF. É possível perceber que os

artigos são provenientes da medicina e da computação, constituindo exemplos da presença de outras ciências publicando em espaços especializados da CI.

No GRÁFICO 2 são demonstradas as classificações dos artigos e os países que os publicaram, de forma consolidada:

GRÁFICO 2 – Classificação dos artigos e países de origem



Fonte: elaborado pelo autor, 2017.



O GRÁFICO 2 demonstra que os Estados Unidos são o país que mais publicou artigos em números absolutos, mantendo esta posição dentre os artigos classificados como RE, contribuindo com 81 dos 144 artigos. Os Estados Unidos também aparecem como o país que mais publicou artigos dentre os artigos classificados com AS. No entanto, a China é o país que mais publicou artigos como objetos de pesquisa, contribuindo com dez dos 47 artigos classificados nesta categoria.

No GRÁFICO 2 percebe-se que a produção científica analisada e classificada na categoria *Big Data* como AS, é composta por 255 artigos distribuídos por 30 países. Nesses artigos, o tema *Big Data* está presente, porém não é o objeto principal da pesquisa:

27 países, o equivalente a 90% dos países que possuem artigos classificados como AS publicaram 84 artigos, o equivalente a 33,20% deste total de artigos:

Publicou-se dez artigos na Coreia do Sul, oito no Canadá, sete na Holanda, seis na Itália, cinco na Nova Zelândia, cinco na França, cinco na Espanha e cinco na Índia. Publicou-se quatro artigos na Alemanha, quatro na Austrália, três na Polônia, dois na Hungria, dois na Noruega, dois na Irlanda, dois na Dinamarca, dois em Israel e dois na Suíça. Publicou-se um artigo Grécia, Portugal, Suécia, Irã, Áustria, Finlândia, Palestina, Turquia, Letônia e Malásia;

A China, o equivalente a 3,33% dos países que possuem artigos classificados como AS, publicou 21 artigos, o equivalente a 8,30% desse total de artigos;

Reino Unido, o equivalente a 3,33% dos países que possuem artigos classificados como AS publicou 33 artigos, o equivalente a 13,04% deste total de artigos;

Estados Unidos, o equivalente a 3,33% dos países que possuem artigos classificados como AS, publicou 115 artigos, o equivalente a 45,45% deste total de artigos.

Ainda de acordo com o GRÁFICO 2, percebe-se que a produção científica analisada e classificada na categoria *Big Data* como RE, é composta por 144 artigos distribuídos em 16 países. Nesses artigos, são realizadas reflexões sobre o tema *Big Data*, sem que haja aprofundamento ou abordagem direta ao mesmo.

Treze países, o equivalente a 81,25% dos países que possuem artigos classificados como RE, publicou de um a dez artigos, totalizando 34 artigos, o equivalente a 23,61% do total dos artigos classificados como RE. Publicou-se seis artigos na China, cinco na Dinamarca, três na Alemanha, três na Austrália, três na Coreia do Sul, três na Espanha, três em Malta, dois na Holanda, dois na Itália e um na Grécia, Malásia e Marrocos.

Canadá, o equivalente a 6,66% dos países que possuem artigos classificados como RE, publicou onze artigos, o equivalente a 7,63% do total dos artigos com classificação como RE;

Reino Unido, o equivalente a 6,66% dos países que possuem artigos classificados como RE, publicou dezoito artigos, o equivalente a 12,50% do total desses arquivos;

Estados Unidos, o equivalente a 6,66% do total dos países que possuem artigos classificados como RE, publicaram 81 artigos, o equivalente a 56,25% do total desses artigos:

No GRÁFICO 2 percebe-se que a produção científica analisada e classificada na categoria *Big Data* como OP é composta por 47 artigos distribuídos em 18 países. Nesses artigos o *Big Data* é o tema principal da pesquisa:

Quatorze países, o equivalente a 29,78% dos países que possuem artigos classificados como OP, publicaram menos que seis artigos. Totalizaram dezesseis artigos, o equivalente a 34,04% do total dos artigos com classificação OP. Publicou-se três artigos no Reino Unido e um artigo no Irã, Canadá, Japão, Argélia, França, Coreia Do Sul, Áustria, Suíça, Chile, Nova Zelândia, Espanha, Austrália e Dinamarca.

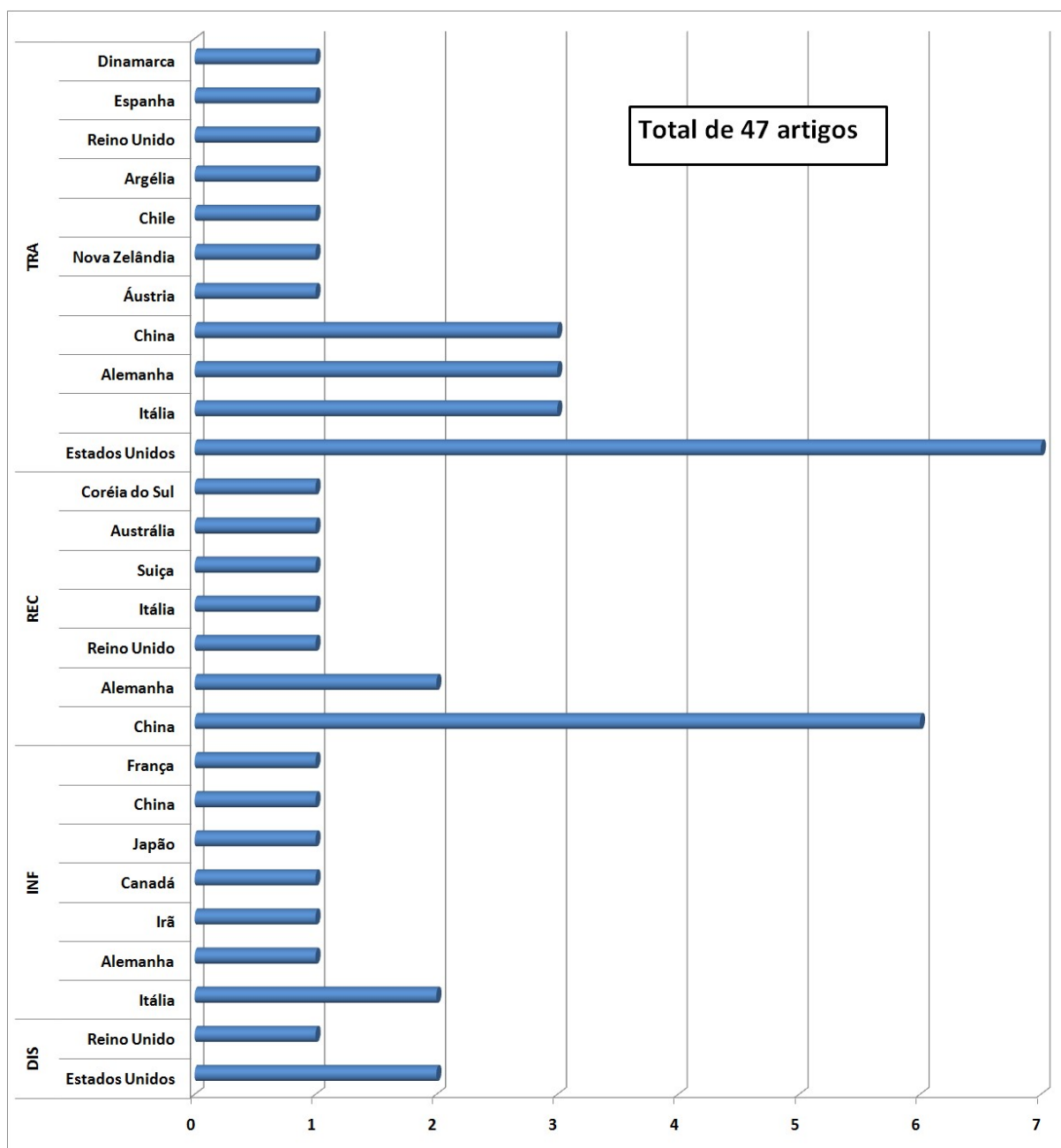
Alemanha, o equivalente a 2,12% dos países que possuem artigos classificados como OP, publicou seis artigos, o equivalente a 12,76% do total dos artigos com classificação como OP.

Itália, o equivalente a 2,12% dos países que possuem artigos classificados como OP, publicou seis artigos, o equivalente a 12,76% do total dos artigos com como classificação OP.

Estados Unidos, o equivalente a 2,12% dos países que possuem artigos classificados como OP, publicou nove artigos, o equivalente a 19,14% do total dos artigos com classificação OP.

China, o equivalente a 2,12% dos países que possuem artigos classificados como OP, publicou dez artigos, 21,27% do total dos artigos com classificação OP.

GRÁFICO 3 – Países com artigos classificados como Objeto de Pesquisa e sua qualificação



Fonte: elaborado pelo autor, 2017.

De acordo com o GRÁFICO 3 observam-se os países cuja produção científica analisada foi classificada na categoria *Big Data* como OP, após ser estratificada e combinada com o campo “Objeto de Pesquisa como”, apresenta o seguinte comportamento:

TRA: 23 artigos, o equivalente a 48,93% dos artigos classificados como OP receberam esta qualificação. Foram publicados por onze países, o equivalente a 61,11% dos países que publicaram artigos classificados como OP: publicou-se sete artigos nos Estados Unidos, três na Itália, Alemanha, China, Áustria, Alemanha e China. Publicou-se um artigo nos seguintes países: Áustria, Nova Zelândia, Chile, Argélia, Reino Unido, Espanha e Dinamarca.

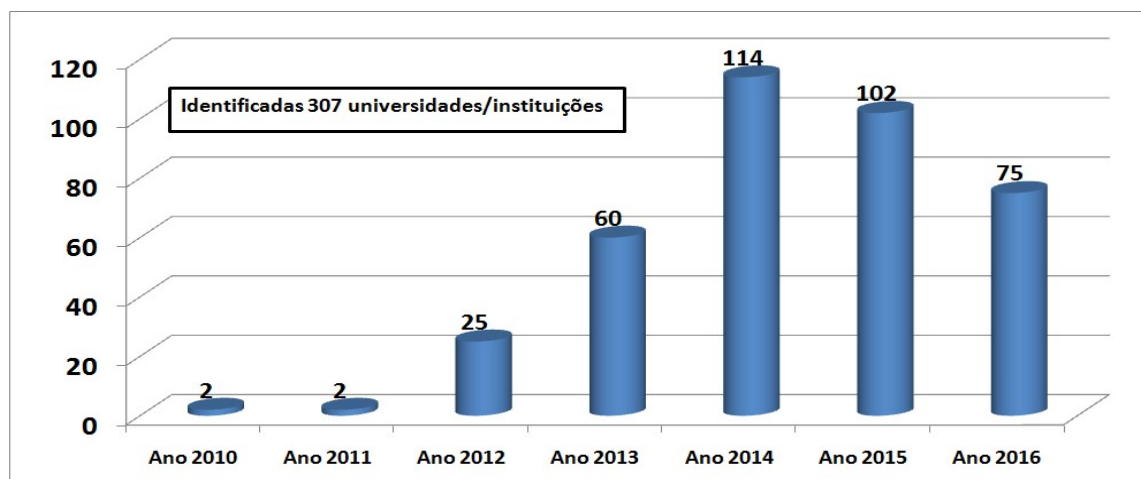
REC: treze artigos, o equivalente a 27,65% dos artigos classificados como OP, receberam esta qualificação. Foram publicados por sete países, o equivalente a 38,88% dos países que publicaram artigos classificados como OP. Publicou-se seis artigos na China, dois artigos na Alemanha. Publicou-se um artigo no Reino Unido, Itália, Suíça, Austrália e Coréia do Sul.

INF: oito artigos, o equivalente a 17,02% dos artigos classificados como OP receberam esta qualificação. Foram publicados por sete países, o equivalente a 38,88% dos países que publicaram artigos classificados como OP. Publicou-se dois artigos na Itália, um artigo na Alemanha, Irã, Canadá, Japão, China e França.

DIS: três artigos, o equivalente a 6,38% dos artigos classificados como OP receberam esta qualificação. Foram publicados por dois países, o equivalente a 11,11% dos países que publicaram artigos classificados como OP. Publicou-se dois artigos nos Estados Unidos e um artigo no Reino Unido.

Informações detalhadas sobre os países presentes nos levantamentos desta pesquisa, podem ser visualizadas no Apêndice A. O GRÁFICO 4 demonstra o número de universidades que publicaram artigos relacionados ao *Big Data* a cada ano:

**GRÁFICO 4 – Número de universidades/instituições com produção acadêmica sobre *Big Data* por ano**



Fonte: elaborado pelo autor, 2017.

O GRÁFICO 4 demonstra que o universo da pesquisa é composto por 307 universidades/instituições que publicaram artigos envolvendo a temática *Big Data*. Observou-se que em 2010 e 2011, duas universidades publicaram artigos relacionados ao *Big Data* a cada ano. Em 2012, 25 universidades/instituições publicaram artigos sobre *Big Data*. Em 2013, 60 universidades/instituições publicaram artigos; em 2014, 114 universidades publicaram artigos; em 2015, 102 universidades publicaram artigos e em 2016, 75 universidades/instituições publicaram artigos relacionados ao *Big Data*.

A Tabela 2 demonstra a quantidade de universidades que publicaram artigos em cada país, ano a ano.

**TABELA 2 – Número de universidades/instituições com publicação**

País	2010/11		2012		2013		2014		2015		2016		Total	
	Univ.	Artigos	Univ.	Artigos	Univ.	Artigos	Univ.	Artigos	Univ.	Artigos	Univ.	Artigos	Univ	Artigos
Estados Unidos	4	4	13	16	31	43	52	69	35	45	26	29	161	206
Reino Unido	-	-	9	10	9	12	10	13	9	10	8	9	45	54
China	-	-	-	-	4	4	9	10	14	15	7	8	34	37
Canadá	-	-	1	1	2	2	8	8	3	4	5	6	19	21
Coréia do Sul	-	-	-	-	1	1	5	5	4	4	5	5	15	15
Alemanha	-	-	-	-	2	2	5	5	5	5	1	1	13	13
Itália	-	-	1	1	-	-	1	1	4	4	6	7	12	13
Espanha	-	-	-	-	1	1	2	2	4	4	2	2	9	9
Holanda	-	-	-	-	-	-	1	1	4	4	3	4	8	9
Austrália	-	-	-	-	2	2	4	4	2	2	-	-	8	8
Dinamarca	-	-	-	-	2	3	1	1	2	2	2	2	7	8
França	-	-	-	-	2	2	-	-	1	1	3	3	6	6
Nova Zelândia	-	-	1	1	1	1	2	3	-	-	1	1	5	6
Índia	-	-	-	-	-	-	3	3	1	1	1	1	5	5
Malta	-	-	-	-	1	3	-	-	-	-	-	-	1	3
Polônia	-	-	-	-	-	-	1	1	1	1	1	1	3	3
Suíça	-	-	-	-	-	-	2	2	1	1	-	-	3	3
Argélia	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Áustria	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Grécia	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Hungria	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Irã	-	-	-	-	-	-	-	-	1	1	1	1	2	2
Irlanda	-	-	-	-	-	-	1	1	-	-	1	1	2	2
Israel	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Malásia	-	-	-	-	-	-	-	-	2	2	-	-	2	2
Noruega	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Chile	-	-	-	-	-	-	1	1	-	-	-	-	1	1
Finlândia	-	-	-	-	-	-	-	-	-	-	1	1	1	1
Japão	-	-	-	-	1	1	-	-	-	-	-	-	1	1
Letônia	-	-	-	-	1	1	-	-	-	-	-	-	1	1
Marrocos	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Palestina	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Portugal	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Suécia	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Turquia	-	-	-	-	-	-	-	-	-	-	1	1	1	1
<b>Total</b>	<b>4</b>	<b>4</b>	<b>25</b>	<b>29</b>	<b>60</b>	<b>78</b>	<b>114</b>	<b>136</b>	<b>103</b>	<b>116</b>	<b>75</b>	<b>83</b>	<b>-</b>	<b>446</b>

Fonte: elaborado pelo autor, 2017.

Na análise da Tabela 2 foi observado que 307 universidades, publicaram 446 artigos relacionados ao *Big Data* entre 2010 e 2016, demonstrando que algumas universidades/instituições publicaram em mais de uma oportunidade. Nesse contexto, destacam-se as três universidades que mais publicaram:

a) *University of California*, Estados Unidos, com 17 artigos:

QUADRO 7 – Programas na University of California

Programa	Ciência	Artigos
Institute for Research on Labor and Employment, University of California-Berkeley	Multidisciplinar	7
Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California Irvine	Computação	1
UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, USA	Biologia	1
Department of Communication, University of California, Davis,	Comunicação	1
Innovation and Industry Alliances, University of California San Diego	Inovação	1
Department of Anthropology University of California	Antropologia	1
Department of Family and Preventive Medicine University of California San Diego	Medicina	1
Department of Communication, University of California,	Comunicação	1
Santa Barbara, Anthropology (Archaeology) Department University of California,	Antropologia	1
Research Data Curation Program, University of California San Diego Library	Biblioteconomia	1
School of Information, University of California, Berkeley, CA, USA	<i>iSchool</i>	1
	<b>Total</b>	<b>17</b>

Fonte: elaborado pelo autor, 2017.

Ao analisar o Quadro 7, percebe-se que a produção acadêmica dessa universidade é composta por 11 programas/departamentos, destacando-se o *Institute for Research on Labor and Employment*, com sete artigos. Esse é um Instituto multidisciplinar, voltado a temáticas de trabalho e emprego. Em seguida, percebe-se uma dispersão entre dez programas/departamentos, que publicaram um artigo cada. Podemos visualizar produções de ciências voltadas aos estudos dos dados, como a computação, a biblioteconomia e a *iSchool*. ciências relacionadas aos estudos antropológicos, comunicação, biologia e medicina.

b) *University of Southampton*, Reino Unido, com oito artigos:

QUADRO 8 – Programas da University of Southampton

Programa	Ciência	Artigos
University of Southampton, UK	Computação	7
Electronics and Computer Science, University of Southampton, Southampton, UK	Computação	1
<b>Total</b>		<b>8</b>

Fonte: elaborado pelo autor, 2017.

Ao analisar o Quadro 8, percebe-se que a *University of Southampton* publicou oito artigos através de três programas, todos relacionados ao segmento da computação. Destaca-se o programa “*Computiv and Visiting Professor, University of Southampton, UK*”, destinado a professores visitantes na universidade.

c) *University of North Carolina*, Estados Unidos, com sete artigos:

QUADRO 9 – Programas da University of North Carolina

Programa	Ciência	Artigos
Bryan School of Business and Economics, The University of North Carolina at Greensboro	Administração	1
Bryan School of Business and Economics, The University of North Carolina at Greensboro	Administração	2
College of Arts & Sciences; CONNECTIONS (Mathematics); MATHEMATICIANS	Computação	1
Digital Innovation Lab, University of North Carolina at Chapel Hill,	Inovação Digital	1
School of Information and Library Science, University of North Carolina, Chapel Hill	Ciência da Informação	2
<b>Total geral</b>		<b>7</b>

Fonte: elaborado pelo autor, 2017.

Percebe-se que os programas das universidades/instituições americanas destacadas acima, *University of California* e *University of North Carolina*, apresentam maior diversidade entre as ciências que publicaram, enquanto os programas da universidade de *South Hampton* apresentam programas exclusivamente relacionados à computação e web.

Todas as informações obtidas nos levantamentos desta pesquisa, que dizem respeito às universidades que publicaram artigos relacionados ao *Big Data* podem ser acessadas no Apêndice A.

TABELA 3 – Países e o número dos programas/departamentos com publicações

País	2010/11		2012		2013		2014		2015		2016		Total	
	Prog.	Art.	Prog.	Art.	Prog.	Art.	Prog.	Art.	Prog.	Art.	Prog.	Art.	Prog.	Art.
Estados Unidos	4	4	14	16	37	43	61	69	44	45	29	29	189	206
Reino Unido	-	-	9	10	10	12	10	13	10	10	9	9	48	54
China	-	-	-	-	4	4	9	10	15	15	8	8	36	37
Canadá	-	-	1	1	2	2	8	8	4	4	6	6	21	21
Coréia do Sul	-	-	-	-	1	1	5	5	4	4	5	5	15	15
Alemanha	-	-	-	-	2	2	5	5	5	5	1	1	13	13
Itália	-	-	1	1	-	-	1	1	4	4	7	7	13	13
Espanha	-	-	-	-	1	1	2	2	4	4	2	2	9	9
Holanda	-	-	-	-	-	-	1	1	4	4	4	4	9	9
Austrália	-	-	-	-	2	2	4	4	2	2	-	-	8	8
Dinamarca	-	-	-	-	3	3	1	1	2	2	2	2	8	8
França	-	-	-	-	2	2	-	-	1	1	3	3	6	6
Nova Zelândia	-	-	1	1	1	1	2	3	-	-	1	1	5	6
Índia	-	-	-	-	-	-	3	3	1	1	1	1	5	5
Malta	-	-	-	-	1	3	-	-	-	-	-	-	1	3
Polônia	-	-	-	-	-	-	1	1	1	1	1	1	3	3
Suíça	-	-	-	-	-	-	2	2	1	1	-	-	3	3
Argélia	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Áustria	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Grécia	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Hungria	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Irã	-	-	-	-	-	-	-	-	1	1	1	1	2	2
Irlanda	-	-	-	-	-	-	1	1	-	-	1	1	2	2
Israel	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Malásia	-	-	-	-	-	-	-	-	2	2	-	-	2	2
Noruega	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Chile	-	-	-	-	-	-	1	1	-	-	-	-	1	1
Finlândia	-	-	-	-	-	-	-	-	-	-	1	1	1	1
Japão	-	-	-	-	1	1	-	-	-	-	-	-	1	1
Letônia	-	-	-	-	1	1	-	-	-	-	-	-	1	1
Marrocos	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Palestina	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Portugal	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Suécia	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Turquia	-	-	-	-	-	-	-	-	-	-	1	1	1	1
<b>Total</b>	<b>4</b>	<b>4</b>	<b>26</b>	<b>29</b>	<b>68</b>	<b>78</b>	<b>123</b>	<b>136</b>	<b>115</b>	<b>116</b>	<b>83</b>	<b>83</b>	<b>384</b>	<b>446</b>

Fonte: elaborado pelo autor, 2017.



Na Tabela 3 observa-se que 384 programas publicaram 446 artigos, distribuídos em 35 países, dos quais, pelo volume de publicações, destacam-se Estados Unidos, Reino Unido e China.

Nos anos de 2010 e 2011 foram identificados quatro programas de origem norte-americana que publicaram o total de quatro artigos.

Estados Unidos: 189 programas publicaram o total de 206 artigos entre os anos de 2010 e 2016. Nos anos de 2010 e 2011 conjuntamente, quatro programas publicaram o total de quatro artigos. Em 2012, quatorze programas publicaram o total de dezesseis artigos. Em 2013, 37 programas publicaram o total de 43 artigos. Em 2014, 61 programas publicaram o total de 69 artigos. Em 2015, 44 programas publicaram o total de 45 artigos e em 2016, 29 programas publicaram o total de 29 artigos.

Reino Unido: 48 programas publicaram o total de 54 artigos entre os anos de 2012 e 2016. Nove programas publicaram dez artigos em 2012. Em 2013, dez programas publicaram o total de doze artigos. Em 2014, dez programas publicaram o total de treze artigos. Em 2015, dez programas publicaram o total de dez artigos e em 2016 nove programas publicaram nove artigos.

China: 36 programas publicaram o total de 37 artigos entre os anos de 2013 e 2016. Em 2013, quatro programas publicaram o total de quatro artigos. Em 2014, nove programas publicaram o total de dez artigos. Em 2015, quinze programas publicaram o total de quinze artigos e em 2016, oito programas publicaram o total de oito artigos.

Canadá: 21 programas publicaram o total de 21 artigos entre os anos de 2012 e 2016. Um programa publicou um artigo em 2012. Em 2013, dois programas publicaram o total de dois artigos. Em 2014, oito programas publicaram o total de oito artigos. Em 2015, quatro programas publicaram o total de quatro artigos e em 2016, seis programas publicaram o total de seis artigos.

Coréia do Sul: quinze programas publicaram o total de quinze artigos entre os anos de 2013 e 2016. Em 2013 um programa publicou um artigo. Em 2014, cinco programas publicaram o total de cinco artigos. Em 2015, quatro programas publicaram o total de quatro artigos e em 2016, cinco programas publicaram o total de cinco artigos.

Alemanha: treze programas publicaram o total de treze artigos entre os anos de 2013 e 2016. Dois programas publicaram o total de dois artigos em 2013, cinco programas publicaram o total de cinco artigos em 2014. Cinco programas publicaram o total de cinco artigos em 2015 e em 2016 um programa publicou um artigo.

Itália: treze programas publicaram o total de treze artigos nos anos de 2012, 2014, 2015 e 2016. Um programa publicou um artigo em 2012. Em 2014 um programa publicou um artigo. Em 2015 quatro programas publicaram o total de quatro artigos e em 2016, sete programas publicaram o total de sete artigos.

Espanha: nove programas publicaram o total de nove artigos entre os anos de 2013 e 2016. Em 2013 um programa publicou um artigo. Em 2014, dois programas publicaram dois artigos. Em 2015, quatro programas publicaram o total de quatro artigos e em 2016, dois programas publicaram o total de dois artigos.

Holanda: nove programas publicaram o total de nove artigos entre os anos de 2014 e 2016. Em 2014 um programa publicou um artigo. Em 2015 quatro programas publicaram o total de quatro artigos e em 2016, quatro programas publicaram o total de quatro artigos.

Austrália: oito programas publicaram o total de oito artigos entre os anos de 2013 e 2016. Dois programas publicaram o total de dois artigos em 2013. Em 2014, quatro programas publicaram o total de quatro artigos e em 2015, dois programas publicaram o total de dois artigos.

Dinamarca: oito programas publicaram o total de oito artigos entre os anos de 2013 e 2016. Em 2013, três programas publicaram o total de três artigos. Em 2014, um programa publicou um artigo. Em 2015, dois programas publicaram o total de dois artigos e em 2016, dois programas publicaram o total de dois artigos.

França: seis programas publicaram o total de seis artigos nos anos de 2013, 2015 e 2016. Em 2013 dois programas publicaram o total de dois artigos. Em 2015 um programa publicou um artigo e em 2016, três programas publicaram o total de três artigos.

Nova Zelândia: cinco programas publicaram o total de seis artigos entre os anos de 2012, 2013, 2014 e 2016. Um programa publicou um artigo em 2012. Em 2013 um programa publicou um artigo. Em 2014 dois programas publicaram o total de três artigos e em 2016 um programa publicou um artigo.

Índia: cinco programas publicaram o total de cinco artigos entre os anos de 2014 e 2016. Em 2014 três programas publicaram o total de três artigos. Em 2015 um programa publicou um artigo e em 2016, um programa publicou um artigo.

Malta: um programa publicou três artigos no ano de 2013.

Polônia: três programas publicaram o total de três artigos entre os anos de 2014 e 2016. Um programa publicou um artigo em 2014. Em 2015, um programa publicou um artigo e 2016 um programa publicou um artigo.

Suíça: três programas publicaram o total de três artigos entre os anos de 2014 e 2015. Dois programas publicaram o total de dois artigos em 2014. Em 2015 um programa publicou um artigo.

Argélia: dois programas publicaram o total de dois artigos nos anos de 2014 e 2015.

Áustria: dois programas publicaram o total de dois artigos entre os anos de 2014 e 2015.

Grécia: dois programas publicaram o total de dois artigos nos anos de 2014 e 2015. Em 2014, um programa publicou um artigo e em 2015 um programa publicou um artigo.

Hungria: dois programas publicaram o total de dois artigos nos anos de 2014 e 2015. Em 2014, um programa publicou um artigo e em 2015 um programa publicou um artigo.

Irã: dois programas publicaram o total de dois artigos nos anos de 2015 e 2016.

Irlanda: dois programas publicaram dois artigos nos anos de 2014 e 2016.

Israel: dois programas publicaram dois artigos nos anos de 2014 e 2015.

Malásia: dois programas publicaram o total de dois artigos em 2015.

Noruega: publicou-se o total de dois artigos.

Chile: um programa publicou um artigo em 2014.

Finlândia: um programa publicou um artigo em 2016.

Japão: um programa publicou um artigo em 2013.

Letônia: um programa publicou um artigo em 2013.

Marrocos: um programa publicou um artigo em 2015.

Palestina: um programa publicou um artigo em 2015.

Portugal: um programa publicou um artigo em 2015.

Suécia: um programa publicou um artigo em 2015.

Turquia: um programa publicou um artigo em 2016.

Todas as informações obtidas nos levantamentos desta pesquisa, relacionadas aos programas que publicaram artigos relacionados ao *Big Data*, podem ser acessadas no Apêndice A.

**TABELA 4 – Países e o número de autores que publicaram**

(Continua)

País	2010/11		2012		2013		2014		2015		2016		TOTAL	
	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.
Estados Unidos	4	4	15	16	38	43	63	69	43	45	28	29	191	206
Reino Unido	-	-	9	10	10	12	11	13	10	10	9	9	49	54
China	-	-	-	-	4	4	9	10	15	15	8	8	36	37
Canadá	-	-	1	1	2	2	8	8	4	4	6	6	21	21
Coréia do Sul	-	-	-	-	1	1	5	5	4	4	5	5	15	15
Alemanha	-	-	-	-	2	2	5	5	5	5	1	1	13	13
Itália	-	-	1	1	-	-	1	1	4	4	7	7	12	13
Espanha	-	-	-	-	1	1	2	2	4	4	2	2	9	9
Holanda	-	-	-	-	-	-	1	1	4	4	4	4	9	9
Austrália	-	-	-	-	2	2	4	4	2	2	-	-	8	8

TABELA 4 – Países e o número de autores que publicaram

(Conclusão)

País	2010/11		2012		2013		2014		2015		2016		TOTAL	
	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.	Autores	Art.
Dinamarca	-	-	-	-	2	3	1	1	2	2	2	2	7	8
França	-	-	-	-	2	2	-	-	1	1	3	3	6	6
Índia	-	-	-	-	-	-	3	3	1	1	1	1	5	5
Nova Zelândia	-	-	1	1	1	1	2	3	-	-	1	1	4	6
Polônia	-	-	-	-	-	-	1	1	1	1	1	1	3	3
Argélia	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Áustria	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Grécia	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Hungria	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Irã	-	-	-	-	-	-	-	-	1	1	1	1	2	2
Irlanda	-	-	-	-	-	-	1	1	-	-	1	1	2	2
Israel	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Malásia	-	-	-	-	-	-	-	-	2	2	-	-	2	2
Noruega	-	-	-	-	-	-	1	1	1	1	-	-	2	2
Chile	-	-	-	-	-	-	1	1	-	-	-	-	1	1
Finlândia	-	-	-	-	-	-	-	-	-	-	1	1	1	1
Japão	-	-	-	-	1	1	-	-	-	-	-	-	1	1
Letônia	-	-	-	-	1	1	-	-	-	-	-	-	1	1
Malta	-	-	-	-	1	3	-	-	-	-	-	-	1	3
Marrocos	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Palestina	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Portugal	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Suécia	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Suíça	-	-	-	-	-	-	2	2	1	1	-	-	1	3
Turquia	-	-	-	-	-	-	-	-	-	-	1	1	1	1

Fonte: elaborado pelo autor, 2017.

Na Tabela 4 observa-se que 396 autores publicaram o total de 446 artigos, demonstrando que alguns publicaram em mais de uma oportunidade. Nos anos de 2010 e 2011 quatro autores americanos publicaram o total de quatro artigos. Destacam-se Estados Unidos, Reino Unido e China, demonstrando a polarização na produção científica da CI entre estes três países.

Estados Unidos: 191 autores publicaram o total de 206 artigos. Nos anos de 2010 e 2011, conjuntamente, quatro autores publicaram o total de quatro artigos. Em 2012, quinze autores publicaram o total de dezesseis artigos. Em 2013, 38 autores publicaram o total de 43 artigos. Em 2014, 63 autores publicaram o total de 69 artigos. Em 2015, 43 autores publicaram o total de 45 artigos e em 2016, 28 autores publicaram o total de 29 artigos.

Reino Unido: 49 autores publicaram o total de 54 artigos entre os anos de 2012 a 2016. Nove autores publicaram dez artigos em 2012. Em 2013, dez autores publicaram o total de doze artigos. Em 2014, onze autores publicaram o total de treze artigos. Em 2015 dez autores publicaram o total de dez artigos e em 2016 nove programas publicaram nove artigos.

China: 36 autores publicaram o total de 37 artigos entre os anos de 2013 e 2016. Em 2013, quatro autores publicaram o total de quatro artigos. Em 2014, nove autores publicaram o total de dez artigos. Em 2015, quinze autores publicaram o total de quinze artigos e em 2016, oito autores publicaram o total de oito artigos.

Canadá: 21 autores publicaram o total de 21 artigos. Um autor publicou um artigo em 2012. Em 2013, dois autores publicaram o total de dois artigos. Em 2014, oito autores publicaram o total de oito artigos. Em 2015, quatro autores publicaram o total de quatro artigos e em 2016, seis autores publicaram o total de seis artigos.

Coréia do Sul: quinze autores publicaram o total de quinze artigos, no período de 2013 a 2016. Em 2013 um autor publicou um artigo. Em 2014, cinco autores publicaram o total de cinco artigos. Em 2015, quatro autores publicaram o total de quatro artigos e em 2016, cinco autores publicaram o total de cinco artigos.

Alemanha: treze autores publicaram o total de treze artigos entre os anos de 2013 a 2016. Dois autores publicaram o total de dois artigos em 2013, cinco autores publicaram o total de cinco artigos em 2014. Cinco autores publicaram o total de cinco artigos em 2015 e em 2016 um autor publicou um artigo.

Itália: treze autores publicaram o total de treze artigos nos anos de 2012, 2014, 2015 e 2016. Um autor publicou um artigo em 2012. Em 2014 um autor publicou um artigo. Em 2015 quatro autores publicaram o total de quatro artigos e em 2016, sete autores publicaram o total de sete artigos.

Espanha: nove autores publicaram o total de nove artigos entre 2013 e 2016. Em 2013 um autor publicou um artigo. Em 2014, dois autores publicaram dois artigos. Em 2015, quatro autores publicaram o total de quatro artigos e em 2016, dois autores publicaram o total de dois artigos.

Holanda: Nove autores publicaram o total de nove artigos entre os anos de 2014 e 2016. Em 2014, um autor publicou um artigo. Em 2015, quatro autores publicaram o total de quatro artigos e em 2016, quatro autores publicaram o total de quatro artigos.

Austrália: oito autores publicaram o total de oito artigos entre os anos de 2014 e 2015. Dois autores publicaram o total de dois artigos em 2013. Em 2014, quatro autores publicaram o total de quatro artigos e em 2015, dois autores publicaram o total de dois artigos.

Dinamarca: sete autores publicaram o total de oito artigos, entre os anos de 2013 a 2016. Em 2013, dois autores publicaram o total de três artigos. Em 2014, um autor publicou um artigo. Em 2015, dois autores publicaram o total de dois artigos e em 2016, dois autores publicaram o total de dois artigos.

França: seis autores publicaram o total de seis artigos nos anos de 2013, 2015 e 2016. Em 2013, dois autores publicaram o total de dois artigos. Em 2015 um autor publicou um artigo e em 2016, três autores publicaram o total de três artigos.

Índia: cinco autores publicaram o total de cinco artigos entre os anos de 2014 e 2015. Em 2014, três autores publicaram o total de três artigos. Em 2015, um autor publicou um artigo e em 2016, um autor publicou um artigo.

Nova Zelândia: quatro autores publicaram o total de seis artigos nos anos de 2012, 2013, 2014 e 2016. Um autor publicou um artigo em 2012. Em 2013 um autor publicou um artigo. Em 2014 dois autores publicaram o total de três artigos e em 2016 um autor publicou um artigo.

Polônia: três autores publicaram o total de três artigos entre os anos de 2014 e 2016. Um autor publicou um artigo em 2014. Em 2015, um autor publicou um artigo e 2016 um autor publicou um artigo.

Argélia: Dois autores publicaram o total de dois artigos.

Áustria: dois autores publicaram o total de dois artigos entre os anos de 2014 e 2015.

Grécia: dois autores publicaram o total de dois artigos nos anos de 2014 e 2015. Em 2014, um autor publicou um artigo e em 2015 um autor publicou um artigo.

Hungria: dois autores publicaram o total de dois artigos nos anos de 2014 e 2015. Em 2014, um autor publicou um artigo e em 2015 um autor publicou um artigo.

Irã: dois autores publicaram o total de dois artigos entre os anos de 2015 e 2016.

Irlanda: dois autores publicaram o total de dois artigos nos anos de 2014 e 2015.

Israel: dois autores publicaram o total de dois artigos entre os anos de 2014 e 2015.

Malásia: dois autores publicaram o total de dois artigos, no ano de 2015.

Noruega: dois autores publicaram o total de dois artigos entre os anos de 2014 e 2015.

Chile: um autor publicou um artigo em 2014.

Finlândia: um autor publicou um artigo em 2016.

Japão: um autor publicou um artigo em 2013.

Letônia: um autor publicou um artigo em 2013.

Malta: um autor publicou 3 artigos em 2013.

Marrocos: um autor publicou um artigo em 2015.

Palestina: um autor publicou um artigo em 2015.

Portugal: um autor publicou um artigo em 2015.

Suécia: um autor publicou um artigo em 2015.

Suíça: três autores publicaram o total de três artigos nos anos de 2014 e 2015.

Dois autores publicaram o total de dois artigos em 2014. Em 2015, um autor publicou um artigo.

Turquia: um autor publicou um artigo em 2016.

O Quadro 10, apresenta os artigos cuja vinculação dos autores não foi identificada:

QUADRO 10 – Publicações de autores sem vinculação – bases LISA e LISTA - 2010 a 2016

(Continua)

Classificação	Ciência	Periódico	Autor	Título	Artigos	
AS	Biblioteconomia	<i>Online Searcher: Information Discovery, Technology, Strategies</i>	Ojala, Marydee	Making Information a Business: The Voice Behind the Curtain	1	
				<i>Online Information's Extreme Makeover</i>	1	
		Information Today	Padgett, Lauree	Big ...	1	
				Mind Games	1	
		Journal of the Leadership & Management Section.	Burke, Timothy	President's and Editor's Message.	1	
		Records Management Journal.	McDonald, John	Whither the retention schedule in the era of big data and open data?	1	
		<b>BIBLIOTECONOMIA TOTAL</b>				<b>6</b>
	Computação	Communications Law: Journal of Computer, Media & Telecommunications Law.	Davies, Clive	Data, data everywhere and not a (proper) law in sight.	1	
		<b>COMPUTAÇÃO TOTAL</b>				<b>1</b>

QUADRO 10 – Publicações de autores sem vinculação – bases LISA e LISTA - 2010 a 2016

(Conclusão)

Classificação	Ciência	Periódico	Autor	Título	Artigos
	Governo Finanças	Editorial Services at the Federal Reserve Bank of Philadelphia	Brynko, Barbara	DataContent: Where Data Meets Community, Markets	1
				R&D Workflows: What's in the Pipeline	1
<b>GOVERNO FINANÇAS TOTAL</b>					<b>2</b>
	Medicina	21st Century Publishing Group	Qiulin, Zhang	Children's Reading in Big Data Time How to Think and What to Do.	1
				Journal of the American Medical Informatics Association	Ohno-Machado, Lucila
			Richesson, Rachel L.	Health information technology data standards get down to business: maturation within domains and the emergence of interoperability.	1
<b>MEDICINA TOTAL</b>					<b>3</b>
<b>TOTAL AS</b>					<b>12</b>
RE	Biblioteconomia	Information Today	Kaser, Dick	Is It Search or Discovery?	1
<b>BIBLIOTECONOMIA TOTAL</b>					<b>1</b>

Fonte: elaborado pelo autor, 2017.

Ao analisar o Quadro 10, percebe-se que doze autores, cuja vinculação não foi identificada, foram responsáveis pela publicação de dezesseis artigos, o equivalente a 3,58% dos artigos e a 3,03% dos autores do universo desta pesquisa. Nesse contexto, destacam-se publicações realizadas através de sete periódicos especializados em temas relacionados à informação, que juntos totalizam nove artigos, distribuídos pelas seguintes ciências: biblioteconomia, ciência da computação e CI. Nota-se a presença do governo americano, com a publicação de quatro artigos e em seguida dois periódicos especializados em medicina, com três artigos publicados. Quanto à classificação, verifica-se a publicação de doze artigos classificados como AS e quatro artigos classificados como RE.

Todas as informações obtidas nos levantamentos desta pesquisa, que dizem respeito aos autores que publicaram artigos relacionados ao *Big Data* podem ser acessadas no Apêndice A.

Portanto, ao identificar-se 307 universidades/instituições incluindo empresas, governos e organizações, produzindo artigos relacionados ao *Big Data*, em bases



especializadas da CI, percebe-se a importância do tema *Big Data* para a academia, em especial para essa ciência, para sociedade e governos.

Ribeiro (2014) aponta a licitude do envolvimento do profissional da informação com as questões relacionadas à informação e à importância dos dados para a CI. Nesse contexto, autores como Saracevic (1996, 1999) e Taurion (2012b) reconhecem a explosão informacional que atravessa a sociedade, gerando e consumindo grandes volumes de dados. Reconhece-se a face tecnológica da CI através de Saracevic (1996) e sua recursividade em Galvão e Borges (2000).

Azam (2014) aponta que as universidades participam do atual cenário informacional, produzindo conhecimentos, realizando estudos e também formando os novos profissionais da informação.

Souza (2006) aponta a necessidade do cientista da informação e do bibliotecário buscarem conhecimentos para manterem-se atualizados quanto aos progressos tecnológicos e teóricos que cercam a informação. Nesse contexto, Hey e Hey (2006), corroboram com essa idéia, quando apontam que a ciência demandará novas habilidades do profissional da informação e ainda Bates (1999) aponta a disseminação da informação como um dos princípios desta ciência. Portanto ao identificar-se 396 autores publicando artigos relacionados ao *Big Data* em bases de dados dedicadas à CI, LISA e LISTA, percebe-se a disseminação do conhecimento relativo ao fenômeno *Big Data*, entre os profissionais da informação, a busca por novos conhecimentos e a geração de conhecimentos a partir de achados científicos que geram novos conhecimentos, conforme Costa e Orrico (2007).

Saracevic (1996) e Souza, Almeida e Baracho (2013) reconhecem a perspectiva interdisciplinar da CI.

Esta pesquisa identificou 385 programas que publicaram artigos relacionados ao *Big Data*, a partir de 35 países de diversas culturas e de diversas ciências. O achado desta pesquisa, deu-se em bases de dados reconhecidamente dedicadas à CI, onde se percebem artigos gerados a partir de outras ciências como medicina, computação e da própria CI, constituindo exemplo da interdisciplinaridade desta ciência.

Souza, Almeida e Baracho (2013) refletem sobre as consequências que esta característica interdisciplinar pode trazer para esta ciência. Porém, autores como Galvão e Borges (2000) e Bates (1999) apontam que a proximidade com as demais ciências, além de fonte de conhecimentos é uma das ferramentas da CI em seu papel de ciência gestora da informação e do conhecimento.

Mysore, Khupat e Jain (2014, *online*) comentam que o padrão informacional atual demanda novas formas de registro, organização e análise destes dados, e observam

que o volume de dados presente em nossos dias, demandam novas formas de registrar, organizar, indexar, analisar e disponibilizar estes dados.

Observa-se que a prática de intercâmbios e busca de melhores soluções através de parcerias com outras ciências, constituem conforme apontam Souza, Almeida e Baracho (2013), prática presente nesta ciência, podendo ser a diversidade na origem dos programas identificados nesta pesquisa, uma demonstração da interdisciplinaridade dessa ciência.

Os levantamentos realizados nesta pesquisa demonstram alguns destaques conforme se demonstra abaixo.

Dos 396 autores identificados nesta pesquisa, 28 publicaram mais que um artigo.

Destes, quatorze autores publicaram dois artigos, dez publicaram três artigos, dois autores publicaram quatro artigos, um autor publicou cinco artigos e um autor publicou sete artigos.

O Quadro 11 destaca os autores com três ou mais publicações:

QUADRO 11 – Autores com três ou mais publicações

(Continua)

Autor	País	Origem do autor	Ano	Classificação	Artigos
Huwe, Terence K	Estados Unidos	University of California	2012	AS	1
				RE	1
			2013	AS	1
			2014	AS	1
			2015	RE	2
	RE	1			
<b>Huwe, Terence K Total</b>			<b>Total</b>	<b>7</b>	
Boeri, Robert	Estados Unidos	IEEE.org	2012	RE	1
			2013	AS	1
		Sociotechnological dataScientist	2013	RE	1
			2013	AS	1
<b>Boeri, Robert Total</b>			<b>Total</b>	<b>4</b>	
Brynko, Barbara	Estados Unidos	Editorial Services at the Federal Reserve Bank of Philadelphia	2012	AS	1
			2013	AS	1
				RE	2
<b>Brynko, Barbara Total</b>			<b>Total</b>	<b>4</b>	
Arnold, Stephen E	Estados Unidos	Tandon School of Engineering	2012	AS	1
			2013	AS	1
			2014	AS	1
<b>Arnold, Stephen E Total</b>			<b>Total</b>	<b>3</b>	
Springer, Robert	Estados Unidos	Independent writer	2013	AS	1
			2014	AS	1
				RE	1
<b>Springer, Robert Total</b>			<b>Total</b>	<b>3</b>	

QUADRO 11 – Autores com três ou mais publicações

(Conclusão)

Autor	País	Origem do autor	Ano	Classificação	Artigos
Ojala, Marydee	Estados Unidos	Online Searcher: Information Discovery,	2013	AS	1
		Technology, Strategies	2014	AS	1
		WebSearch University	2012	RE	1
<b>Ojala, Marydee Total</b>				<b>Total</b>	<b>3</b>
<b>Total geral</b>					<b>24</b>

Fonte: elaborado pelo autor, 2017.

Observa-se ainda os autores de destaque na publicação de artigos relacionados ao *Big Data*, a identificação do país, universidade e programa de origem, além da classificação de seus artigos e o ano da publicação.

Destacam-se: Huwe, Terence K., dos Estados Unidos, Universidade de Berkeley com sete artigos publicados, seguido por James, Roger do Reino Unido, IEEE.org com cinco artigos publicados. Em terceiro lugar aparece Boeri Robert, dos Estados Unidos, Editorial Services at the Federal Reserve Bank of Philadelphia, com quatro artigos publicados. Percebe-se as contribuições que surgem a partir de setores acadêmicos, organizações profissionais e governo.

QUADRO 12– Universidades/Instituições com mais publicações–bases LISA e LISTA –2010 a 2016<sup>4</sup>

(Continua)

Universidade/Instituição	País	Artigos
University of California	Estados Unidos	17
Independent writer	Estados Unidos (8)	9
	Reino Unido (1)	
University of Southampton	Reino Unido	8
University of North Carolina	Estados Unidos	7
George Mason University	Estados Unidos	6
CILIP	Reino Unido	5
Johns Hopkins University	Estados Unidos	4
Initiatives.ORG	Nova Zelândia	4
Editorial Services at the Federal Reserve Bank of Philadelphia	Estados Unidos	4
University of Minesota	Estados Unidos	4
Drexel University	Estados Unidos	4
Information Today	Estados Unidos	4
University of Southern California	Estados Unidos	4

4 Autores independentes agrupados como Universidade/Intituição.

QUADRO 12 – Universidades/Instituições com mais publicações – bases LISA e LISTA – 2010 a 2016<sup>5</sup>

(Conclusão)

Universidade/Instituição	País	Artigos
Indiana University	Estados Unidos	4
University of Wisconsin	Estados Unidos	4
Arizona State University	Estados Unidos	3
University of Copenhagen	Dinamarca	3
University of Alberta	Canadá	3
Delft University of Technology	Holanda	3
University of Malta	Malta	3
Microsoft Research	Estados Unidos	3
Tandon School of Engineering	Estados Unidos	3
MIT	Estados Unidos	3
Aarhus University	Dinamarca	3
Montana State University	Estados Unidos	3
University of Illinois	Estados Unidos	3
University of Western Ontario	Canadá	3
University of Pittsburgh	Estados Unidos	3
Purdue University	Estados Unidos	3
Sociotechnological dataScientist	Estados Unidos	3
IBM	Canadá	(1)
	Estados Unidos	(2)
University of Geneva	Suíça	2
NATIONAL Institutes of Health (U.S.)	Estados Unidos	2
Montclair State University	Estados Unidos	2
YeungNam University	Coréia do Sul	2

Fonte: elaborado pelo autor, 2017.

Observa-se no Quadro 12 as universidades que publicaram mais que um artigo relacionado ao *Big Data* e o número de publicações. Dentro dessas, a University of California, destaca-se com 17 artigos publicados. Em seguida, percebe-se que os autores independentes apresentaram volume expressivo de produção, com oito artigos publicados nos Estados Unidos e um artigo publicado no Reino Unido. Nota-se que do total de 309 universidades identificadas, 68 (22%) publicaram mais que um artigo.

5 Autores independentes agrupados como Universidade/Intituição.

QUADRO13– Programas que mais publicaram – bases LISA e LISTA – 2010 a 2016<sup>6</sup>

(Continua)

Programa	Universidade/Instituição	País	Artigos
Independent writer	Independent writer	Estados Unidos	8
		Reino Unido	1
Institute for Research on Labor and Employment	University of California CILIP	Estados Unidos	7
		Reino Unido	5
Computiv and Visiting Professor	University of Southampton	Reino Unido	5
Editorial Services	Editorial Services at the Federal Reserve Bank of Philadelphia	Estados Unidos	4
Initiatives.ORG	Initiatives.ORG	Nova Zelândia	4
Polytechnical Institute	Tandon School of Engeneering	Estados Unidos	3
Sociotechnological dataScientist	Sociotechnological dataScientist	Estados Unidos	3
School of Public Policy	George Mason University	Estados Unidos	3
Social Sciences & Professional Programs	University of Minesota	Estados Unidos	3
College of Computing and Informatics	Drexel University	Estados Unidos	3
Library Information & Archive Sciences	University of Malta Information Today	Malta	3
		Estados Unidos	3
Editor-in-Chief at <i>Online Searcher: Information Discovery, Technology, Strategies</i>	<i>Online Searcher: Information Discovery, Technology, Strategies</i>	Estados Unidos	2
Department of Software Design and Management	Gachon University	Coréia do Sul	2
Oak Ridge National Laboratory	Institute of Preventive Medical Information, Hubei Center for Disease Control and Prevention Oak Ridge National Laboratory Information International Associates AMIA - Informatics Professionals Leading the Way	Estados Unidos	1
		Estados Unidos	1
		Estados Unidos	2
		Estados Unidos	2
University of Southampton	University of Southampton	Reino Unido	2
Faculty of Information and Media Studies	University of Western Ontario Cambridge University	Canadá	2
		Reino Unido	2
Oxford Internet Institute	University of Oxford Delft University of Technology	Reino Unido	2
		Holanda	2
School of Public Health	University of Illinois intranetfocus.com	Estados Unidos	2
		Reino Unido	2
School of Software	Beijing Institute of Technology, Journal of the American Medical Informatics Association	China	2
		Estados Unidos	2

---

6 Autores independentes agrupados como Programa.

QUADRO 13 – Programas que mais publicaram – bases LISA e LISTA – 2010 a 2016<sup>7</sup>

(Conclusão)

Programa	Universidade/Instituição	País	Artigos
Department of Archaeology and Anthropology	University of Bristol	Reino Unido	2
Department of Media and Communication, YeungNam University	YeungNam University	Coréia do Sul	2
Bryan School of Business and Economics	University of North Carolina	Estados Unidos	2
Microsoft Research	Microsoft Research	Estados Unidos	2
Institute of High Performance Computing and Networking (ICAR) University of Salerno	University of Salerno	Itália	2

Fonte: elaborado pelo autor, 2017.

Observa-se no Quadro 13 os programas e instituições que publicaram mais de um artigo relacionado ao *Big Data* e o número de publicações. Destaca-se a presença dos autores independentes com volume de publicação superior ao programa formal com maior número de publicações. Observa-se que o programa Institute for Research on Labor and Employment, University of California-Berkeley publicou sete artigos, sendo o programa com o maior número de artigos publicados. Destaca-se a produção realizada pela organização CILIP com a publicação de cinco artigos.

Identifica-se no Quadro 14 os países que publicaram mais de um artigo relacionado ao *Big Data* e o número de publicações. Destaca-se o volume de publicações dos países: Estados Unidos, com 206 publicações, Reino Unido com 54 artigos e China, com 37 artigos publicados.

---

7 Autores independentes agrupados como Programa.

QUADRO 14 – Países que mais publicaram – bases LISA e LISTA – 2010 a 2016

País	Artigos	País	Artigos
Estados Unidos	206	Índia	5
Reino Unido	54	Suíça	3
China	37	Malta	3
Canadá	21	Polônia	3
Coréia do Sul	15	Irlanda	2
Itália	13	Hungria	2
Alemanha	13	Noruega	2
Holanda	9	Áustria	2
Espanha	9	Malásia	2
Dinamarca	8	Israel	2
Austrália	8	Argélia	2
Nova Zelândia	6	Grécia	2
França	6	Irã	2

Fonte: elaborado pelo autor, 2017.

Identifica-se no Quadro 15 os países e a quantidade de cada classificação atribuída aos países que publicaram artigos sobre *Big Data*, nas bases LISA e LISTA de 2010 a 2016.

QUADRO 15 – Classificação das publicações em cada país

País	Classificação			Total Geral	País	Classificação			Total Geral
	AS	OP	RE			AS	OP	RE	
Estados Unidos	116	9	81	206	Áustria	1	1	-	2
Reino Unido	33	3	18	54	Grécia	1	-	1	2
China	21	10	6	37	Hungria	2	-	-	2
Canadá	9	1	11	21	Irã	1	1	-	2
Coréia do Sul	11	1	3	15	Irlanda	2	-	-	2
Alemanha	4	6	3	13	Israel	2	-	-	2
Itália	5	6	2	13	Malásia	1	-	1	2
Espanha	5	1	3	9	Noruega	2	-	-	2
Holanda	7	-	2	9	Chile	-	1	-	1
Austrália	4	1	3	8	Finlândia	1	-	-	1
Dinamarca	2	1	5	8	Japão	-	1	-	1
França	5	1	-	6	Letônia	1	-	-	1
Nova Zelândia	5	1	-	6	Marrocos	-	-	1	1
Índia	5	-	-	5	Palestina	1	-	-	1
Malta	-	-	3	3	Portugal	1	-	-	1
Polônia	3	-	-	3	Suécia	1	-	-	1
Suíça	2	1	-	3	Turquia	1	-	-	1
Argélia		1	1	2					
					<b>Total Geral</b>	<b>255</b>	<b>47</b>	<b>144</b>	<b>446</b>

Fonte: elaborado pelo autor, 2017.

No Quadro 16, é destacado o comportamento dos países que publicaram artigos relacionados ao *Big Data* nas bases LISA e LISTA entre os anos de 2010 e 2016 e que receberam a classificação como OP:

QUADRO 16 – Países com produção de artigos classificados como Objeto de Pesquisa

Países	Qualificação				Total Geral
	DIS	INF	REC	TRA	
China	-	1	6	3	10
Estados Unidos	2	-	-	7	9
Alemanha	-	1	2	3	6
Itália	-	2	1	3	6
Reino Unido	1	-	1	1	3
Argélia	-	-	-	1	1
Austrália	-	-	1	-	1
Áustria	-	-	-	1	1
Canadá	-	1	-	-	1
Chile	-	-	-	1	1
Coréia do Sul	-	-	1	-	1
Dinamarca	-	-	-	1	1
Espanha	-	-	-	1	1
França	-	1	-	-	1
Irã	-	1	-	-	1
Japão	-	1	-	-	1
Nova Zelândia	-	-	-	1	1
Suíça	-	-	1	-	1
<b>Total geral</b>	<b>3</b>	<b>8</b>	<b>13</b>	<b>23</b>	<b>47</b>

Fonte: elaborado pelo autor, 2017.

No Quadro 16 verifica-se que a China foi o país que mais publicou artigos que receberam a classificação como OP, com dez artigos publicados, seguida pelos Estados Unidos com a publicação de nove artigos. Em seguida aparecem Alemanha e Itália com a publicação de seis artigos cada.

QUADRO 17 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Disseminação da Informação

País	Universidade	Programa/departamento	Artigos
Estados Unidos	Stanford University	Department of Bioengineering	1
	University of Pittsburgh	Department of Biomedical Informatics	1
<b>Estados Unidos Total</b>			<b>2</b>
Reino Unido	Cambridge University	Cambridge University	1
<b>Reino Unido Total</b>			<b>1</b>
<b>Total Geral</b>			<b>3</b>

Fonte: elaborado pelo autor, 2017.



No Quadro 17 verifica-se que os Estado Unidos foram o país que mais publicou artigos que receberam a classificação como OP e a qualificação como DIS, este publicou dois artigos. O Reino Unido publicou um artigo. Esses artigos são apresentados no Quadro 18:

QUADRO 18 – Artigos classificados em Objeto de Pesquisa e qualificados em Disseminação da Informação

(Continua)

País	Título	Autor	Referência
Estados Unidos	The center for causal discovery of biomedical knowledge from big data.	Cooper, Gregory F.	COOPER, Gregory F. et al. The center for causal discovery of biomedical knowledge from big data. <b>Journal of the American Medical Informatics Association</b> , v. 22, n. 6, p. 1132-1136, 2015.
<b>Abstract</b>			
<p>The Big Data to Knowledge (BD2K) Center for Causal Discovery is developing and disseminating an integrated set of open source tools that support causal modeling and discovery of biomedical knowledge from large and complex biomedical datasets. The Center integrates teams of biomedical and data scientists focused on the refinement of existing and the development of new constraint-based and Bayesian algorithms based on causal Bayesian networks, the optimization of software for efficient operation in a supercomputing environment, and the testing of algorithms and software developed using real data from 3 representative driving biomedical projects: cancer driver mutations, lung disease, and the functional connectome of the human brain. Associated training activities provide both biomedical and data scientists with the knowledge and skills needed to apply and extend these tools. Collaborative activities with the BD2K Consortium further advance causal discovery tools and integrate tools and resources developed by other centers. [ABSTRACT FROM AUTHOR]</p>			
País	Título	Autor	Referência
Estados Unidos	The mobilize center: an NIH big data to knowledge center to advance human movement research and improve mobility.	Ku, Joy P.	KU, Joy P. et al. The mobilize center: an NIH big data to knowledge center to advance human movement research and improve mobility. <b>Journal of the American Medical Informatics Association</b> , v. 22, n. 6, p. 1120-1125, 2015.
<b>Abstract</b>			
<p>Regular physical activity helps prevent heart disease, stroke, diabetes, and other chronic diseases, yet a broad range of conditions impair mobility at great personal and societal cost. Vast amounts of data characterizing human movement are available from research labs, clinics, and millions of smartphones and wearable sensors, but integration and analysis of this large quantity of mobility data are extremely challenging. The authors have established the Mobilize Center (<a href="http://mobilize.stanford.edu">http://mobilize.stanford.edu</a>) to harness these data to improve human mobility and help lay the foundation for using data science methods in biomedicine. The Center is organized around 4 data science research cores: biomechanical modeling, statistical learning, behavioral and social modeling, and integrative modeling. Important biomedical applications, such as osteoarthritis and weight management, will focus the development of new data science methods. By developing these new approaches, sharing data and validated software tools, and training thousands of researchers, the Mobilize Center will transform human movement research. [ABSTRACT FROM AUTHOR]</p>			
<b>Estados Unidos Total</b>			<b>2</b>

QUADRO 18 – Artigos classificados em Objeto de Pesquisa e qualificados em Disseminação da Informação

(Conclusão)

País	Título	Autor	Referência
Reino Unido	Using Data to Understand How the Statute Book Works	Sheridan, John	SHERIDAN, John. Using Data to Understand How the Statute Book Works. <b>Legal Information Management</b> , v. 14, n. 4, p. 244-248, 2014.
<b>Abstract</b>			
<p>The statute book is a large, complex system; a vast corpus of texts dating back to the thirteenth century, now evolving at a rate of around 100,000 words a month. The volume and pace of change combine with the constraints of current generation of digital tools to present a real barrier to researchers, limiting the type of research that is currently possible. The statute book is simply too big, and changes too rapidly, for any one person to easily comprehend. This situation is transformed if you view legislation as data, and then apply big data technologies and new data analysis techniques to that data. The aim of the Big Data for Law research project is to do just that; applying the latest analytical techniques to legislation, making it possible to research, interrogate and understand the statute book as a whole system. An important part of the initiative is to make available the raw data for conducting this type of research, alongside new tools and methods for working with the content. In this article, John Sheridan, Head of Legislation Services at The National Archives, sets out some of the ideas that underpin the project and describes the new service that researchers can use from Spring 2015. Adapted from the source document.</p>			
<b>Reino Unido Total</b>			<b>1</b>
<b>Total Geral</b>			<b>3</b>

Fonte: elaborado pelo autor, 2017.

O Quadro 19 apresenta os países que publicaram artigos classificados como OP e qualificados como INF:

QUADRO 19 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados em Infraestrutura

(Continua)

País	Universidade / Instituição	Programa	Artigos
Alemanha	Humboldt-Universität zu Berlin	Institute of Informations Systems, Spandauer Str. 1, 10178 Berlin, Germany	1
<b>Alemanha Total</b>			<b>1</b>
Canadá	IBM	Canada Research and Development Center, Markham, Toronto, ON, Canada	1
<b>Canadá Total</b>			<b>1</b>
China	Zhejiang University	College of Computer Science and Technology	1
<b>China Total</b>			<b>1</b>
França	INRIA Montpellier, France	INRIA Montpellier	1
<b>França Total</b>			<b>1</b>
Irã	Tarbiat Modares Universit	Computer Engineering Department	1

QUADRO 19 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados em Infraestrutura

(Conclusão)

País	Universidade / Instituição	Programa	Artigos
<b>Irã Total</b>			<b>1</b>
Itália	Politecnico di Milano	Politecnico di Milano	1
	University of Salerno	Institute of High Performance Computing and Networking (ICAR)	1
<b>Itália Total</b>			<b>2</b>
Japão	The Institute of Statistical Mathematics (10-3 Midori-cho Tachikawashi	The Research Organization of Information and Systems	1
	<b>Japão Total</b>		<b>1</b>
<b>Total Geral</b>			<b>8</b>

Fonte: elaborado pelo autor, 2017.

No Quadro 19 verifica-se que a Itália publicou dois artigos e os seguintes países publicaram um artigo cada: Alemanha, Canadá, China, França, Irã e Japão. Esses artigos são apresentados no Quadro 20:

QUADRO 20 – Artigos com classificação Objeto de Pesquisa e qualificados em Infraestrutura

(Continua)

País	Título	Autor	Referência
Alemanha	Collaborative and secure sharing of healthcare data in multi-clouds.	Fabian, Benjamin	FABIAN, Benjamin; ERMAKOVA, Tatiana; JUNGHANNS, Philipp. Collaborative and secure sharing of healthcare data in multi-clouds. <b>Information Systems</b> , v. 48, p. 132-150, 2015.
<b>Abstract</b>			

In healthcare, inter-organizational sharing and collaborative use of big data become increasingly important. The cloud-computing paradigm is expected to provide an environment perfectly matching the needs of collaborating healthcare workers. However, there are still many security and privacy challenges impeding the wide adoption of cloud computing in this domain. In this paper, we present a novel architecture and its implementation for inter-organizational data sharing, which provides a high level of security and privacy for patient data in semi-trusted cloud computing environments. This architecture features attribute-based encryption for selective access authorization and cryptographic secret sharing in order to disperse data across multiple clouds, reducing the adversarial capabilities of curious cloud providers. An implementation and evaluation by several experiments demonstrate the practical feasibility and good performance of our approach. [ABSTRACT FROM AUTHOR].

## QUADRO 20 – Artigos com classificação Objeto de Pesquisa e qualificados em Infraestrutura

(Continuação)

País	Título	Autor	Referência
Canadá	Real-Time and Retrospective Health-Analytics-as-a-Service: A Novel Framework.	Khazaei, Hamzeh	KHAZAEI, Hamzeh et al. Real-time and retrospective health-analytics-as-a-service: a novel framework. <b>JMIR medical informatics</b> , v. 3, n. 4, não paginado, 2015.
<b>Abstract</b>			
<p>Background: Analytics-as-a-service (AaaS) is one of the latest provisions emerging from the cloud services family. Utilizing this paradigm of computing in health informatics will benefit patients, care providers, and governments significantly. This work is a novel approach to realize health analytics as services in critical care units in particular. Objective: To design, implement, evaluate, and deploy an extendable big-data compatible framework for health-analytics-as-a-service that offers both real-time and retrospective analysis. Methods: We present a novel framework that can realize health data analytics-as-a-service. The framework is flexible and configurable for different scenarios by utilizing the latest technologies and best practices for data acquisition, transformation, storage, analytics, knowledge extraction, and visualization. We have instantiated the proposed method, through the Artemis project, that is, a customization of the framework for live monitoring and retrospective research on premature babies and ill term infants in neonatal intensive care units (NICUs). Results: We demonstrated the proposed framework in this paper for monitoring NICUs and refer to it as the Artemis-In-Cloud (Artemis-IC) project. A pilot of Artemis has been deployed in the SickKids hospital NICU. By infusing the output of this pilot set up to an analytical model, we predict important performance measures for the final deployment of Artemis-IC. This process can be carried out for other hospitals following the same steps with minimal effort. SickKids' NICU has 36 beds and can classify the patients generally into 5 different types including surgical and premature babies. The arrival rate is estimated as 4.5 patients per day, and the average length of stay was calculated as 16 days. Mean number of medical monitoring algorithms per patient is 9, which renders 311 live algorithms for the whole NICU running on the framework. The memory and computation power required for Artemis-IC to handle the SickKids NICU will be 32 GB and 16 CPU cores, respectively. The required amount of storage was estimated as 8.6 TB per year. There will always be 34.9 patients in SickKids NICU on average. Currently, 46% of patients cannot get admitted to SickKids NICU due to lack of resources. By increasing the capacity to 90 beds, all patients can be accommodated. For such a provisioning, Artemis-IC will need 16 TB of storage per year, 55 GB of memory, and 28 CPU cores. Conclusions: Our contributions in this work relate to a cloud architecture for the analysis of physiological data for clinical decisions support for tertiary care use. We demonstrate how to size the equipment needed in the cloud for that architecture based on a very realistic assessment of the patient characteristics and the associated clinical decision support algorithms that would be required to run for those patients. We show the principle of how this could be performed and furthermore that it can be replicated for any critical care setting within a tertiary institution.</p>			
País	Título	Autor	Referência
China	Deploying Data-Intensive Service Composition with a Negative Selection Algorithm	Deng, Shuiguang	DENG, Shuiguang et al. Deploying data-intensive service composition with a negative selection algorithm. <b>International Journal of Web Services Research (IJWSR)</b> , v. 11, n. 1, p. 76-93, 2014.

## QUADRO 20 – Artigos com classificação Objeto de Pesquisa e qualificados em Infraestrutura

(Conclusão)

Abstract
<p>With the development of information technology, data on the Internet is growing even faster than Moore's Law. At the age of big data, more and more services are created to deal with big data, which are called data-intensive services. In most cases, multiple data-intensive services are assembled into a service composition to meet complicated requirements. Since the big-data transmission, which is occurred among component services as well as between a service and a data center, has great influence on the overall performance of a composition, deploying those services cannot be considered independently. This paper proposes an optimal deployment method based on a negative selection algorithm for a data-intensive service composition to reduce the cost of the data transmission. When making a deployment schedule, it considers not only the cost of data transmission among component services, but also the load balance of data centers where component services are deployed. It models the deployment problem as a combination optimization problem and extends a negative selection algorithm to get an optimal deployment plan. A series of experiments are carried out to evaluate the performance of the proposed method using different settings as well as to compare with other methods. The results show that the method outperforms others for the problem of data-intensive service composition deployment. Adapted from the source document.</p>

Fonte: elaborado pelo autor, 2017.

O Quadro 21, apresenta os países que publicaram artigos classificados como OP e qualificados como REC:

QUADRO 21 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Recuperação da Informação

(Continua)

País	Autor	Universidade / Instituição	Programa/Departamento	Artigos
Alemanha	Huang, Ying	Karlsruhe Institute of Technology	Department of Economics and Management	1
	Rill, Sven	University of Applied Sciences Hof	Institute of Information Systems	1
<b>Alemanha Total</b>				<b>2</b>
Austrália	Nguyen, Chinh	Swinburne University of Technology	Swinburne University of Technology	1
<b>Austrália Total</b>				<b>1</b>
China	Hang Yang	University of Macau	Department of Computer and Information Science	1
	Jieyu Li	University Zhenjiang	Researcher in the Library of Jiangsu University, Zhenjiang, China	1
	Liu, Zheli	Nankai University	College of Computer and Control Engineering	1
	Ma, Baojun	Beijing University	School of Economics and Management	1

QUADRO 21 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Recuperação da Informação

(Conclusão)

País	Autor	Universidade / Instituição	Programa/Departamento	Artigos
	Wang, Shuliang	Beijing Institute of Technology,	School of Software	1
	Yu, Chung-Yen	National Taiwan Normal University	Graduate Institute of Library and Information Studies	1
	<b>China Total</b>			<b>6</b>
Coréia do Sul	Kim, Younghoon	Seoul National University	School of Electrical and Computer Engineering	1
	<b>Coréia do Sul Total</b>			<b>1</b>
Itália	Miele, Antonio	Politecnico di Milano	Dipartimento di Elettronica e Informazione	1
	<b>Itália Total</b>			<b>1</b>
Reino Unido	Trovati, Marcello	University of Derby	Department of Computing and Mathematics	1
	<b>Reino Unido Total</b>			<b>1</b>
Suíça	Mohamed, Hisham	University of Geneva	Computer Science Department, University of Geneva	1
	<b>Suíça Total</b>			<b>1</b>
<b>Total Geral</b>				<b>13</b>

Fonte: elaborado pelo autor, 2017.

Ao analisar-se o Quadro 21, observa-se que sete países publicaram o total de treze artigos que receberam a classificação como OP e a qualificação como REC. Com seis artigos publicados, a China foi o país que mais publicou artigos que receberam esta qualificação, seguido pela Alemanha com dois artigos. Austrália, Coréia do Sul, Itália, Reino Unido e Suíça, publicaram um artigo cada.

O Quadro 22 apresenta os artigos dos dois países que mais publicaram artigos qualificados como REC.

QUADRO 22 – Artigos classificados como Objeto de Pesquisa, qualificados como Recuperação da Informação – China e Alemanha

(Continua)

País	Titulo	Autor	Referência
China	Balancing efficiency and effectiveness for fusion-based search engines in the 'big data' environment.	Jieyu Li	LI, Jieyu et al. Balancing efficiency and effectiveness for fusion-based search engines in the 'big data' environment. <b>Information Research</b> , v. 21, n. 2, 2016.
<b>Abstract</b>			
<p>Introduction. In the big data age, we have to deal with a tremendous amount of information, which can be collected from various types of sources. For information search systems such as Web search engines or <i>online</i> digital libraries, the collection of documents becomes larger and larger. For some queries, an information search system needs to retrieve a large number of documents. On the other hand, very often people are only willing to visit no more than a few top-ranked documents. Therefore, how to develop an information search system with desirable efficiency and effectiveness is a research problem. Method. In this paper, we focus on the data fusion approach to information search, in which each component search model contributes a result and all the results are combined by a fusion algorithm. Through empirical study, we are able to find a feasible combination method that balances effectiveness and efficiency in the context of data fusion. Analysis. It is a multi-optimisation problem that aims to balance effectiveness and efficiency. To support this, we need to understand how these two factors affect each other and to what extent. Results. Using some groups of historical runs from TREC to carry out the experiment, we find that using much less information (e.g., less than 10% of the documents in the experiment), good efficiency is achievable with marginal loss on effectiveness. Conclusions. We consider that the findings from our experiment are informative and this can be used as a guideline for providing more efficient search service in the big data environment. [ABSTRACT FROM AUTHOR]</p>			
País	Titulo	Autor	Referência
China	HASTA: A Hierarchical-Grid Clustering Algorithm with Data Field	Wang, Shuliang	WANG, Shuliang; CHEN, Yasen. HASTA: a hierarchical-grid clustering algorithm with data field. <b>International Journal of Data Warehousing and Mining (IJDWM)</b> , v. 10, n. 2, p. 39-54, 2014.
<b>Abstract</b>			
<p>In this paper, a novel clustering algorithm, HASTA (HierArchical-grid cluStering based on daTA field), is proposed to model the dataset as a data field by assigning all the data objects into quantized grids. Clustering centers of HASTA are defined to locate where the maximum value of local potential is. Edges of cluster in HASTA are identified by analyzing the first-order partial derivative of potential value, thus the full size of arbitrary shaped clusters can be detected. The experimented case demonstrates that HASTA performs effectively upon different datasets and can find out clusters of arbitrary shapes in noisy circumstance. Besides those, HASTA does not force users to preset the exact amount of clusters inside dataset. Furthermore, HASTA is insensitive to the order of data input. The time complexity of HASTA achieves <math>O(n)</math>. Those advantages will potentially benefit the mining of big data. Adapted from the source document.</p>			
País	Titulo	Autor	Referência
China	Solving Problems of Imperfect Data Streams by Incremental Decision Trees.	Hang Yang	YANG, Hang. Solving problems of imperfect data streams by incremental decision trees. <b>Journal of Emerging Technologies in Web Intelligence</b> , v. 5, n. 3, p. 322-331, 2013.

QUADRO 22 – Artigos classificados como Objeto de Pesquisa, qualificados como Recuperação da Informação – China e Alemanha

(Continuação)

Abstract			
<p>Big data is a popular topic that attracts highly attentions of researchers from all over the world. How to mine valuable information from such huge volumes of data remains an open problem. Although fast development of hardware is capable of handling much larger volume of data than ever before, in the author's opinion, a well-designed algorithm is crucial in solving the problems associated with big data. Data stream mining methodologies propose onepass algorithms that discover knowledge hidden behind massive and continuously moving data. These provide a good solution for such big data problems, even for potentially infinite volumes of data. In this paper, we investigate these problems and propose an algorithm of incremental decision tree as the solution. [ABSTRACT]</p>			
País	Título	Autor	Referência
China	The Study of Analytical Model of Library Electronic Resources Usage: A Case of Medical Electronic Resources	Yu, Chung-Yen	YU, Chung-Yen; SHIEH, Jiann-Cherng. The Study of Analytical Model of Library Electronic Resources Usage: A Case of Medical Electronic Resources. <b>Journal of Educational Media &amp; Library Sciences</b> , v. 51, 2014.
Abstract			
<p>With the advents of internet, the importance of electronic resources is growing. Due to the increasing expensiveness of electronic resources, university libraries normally received budgets from parent institutions annually. They necessarily applied effective and systematic methods for decision making in electronic resources purchase or re-subscription. However, there are some difficulties in practices: First of all, libraries are unable to receive user records; second, the COUNTER statistics does not include details about users and their affiliation. As a result, one cannot conduct advanced user analysis based on the usage of users, institutions, and departments. To overcome the difficulties, this study presents a feasible model to analyze electronic resource usage effectively and flexibly. We set up a proxy server to collect actual usage raw data. By analyzing items in internet browsing records, associated with original library automatic system, this study aims at exploring how to use effective ways to analyze big data of website log data. We also propose the process of how original data to be transformed, cleared, integrated, and demonstrated. This study adopted a medical university library and its subscription of medical electronic resources as a case. Our data analysis includes (1) year of subscription, (2) title of journal, (3) affiliation, (4) subjects, and (5) specific journal requirements, etc. The findings of the study are contributed to obtain further understanding in policy making and user behavior analysis. The integrated data provides multiple applications in informatics research, information behavior, bibliominig, presenting diverse views and extended issues for further discussion.</p>			
País	Título	Autor	Referência
China	SQL-Based Fuzzy Query Mechanism Over Encrypted Database	Yu, Chung-Yen	LIU, Zheli et al. Sql-based fuzzy query mechanism over encrypted database. <b>International Journal of Data Warehousing and Mining (IJDWM)</b> , v. 10, n. 4, p. 71-87, 2014.



QUADRO 22 – Artigos classificados como Objeto de Pesquisa, qualificados como Recuperação da Informação – China e Alemanha

(Continuação)

Abstract			
<p>With the development of cloud computing and big data, data privacy protection has become an urgent problem to solve. Data encryption is the most effective way to protect privacy however, it will change the data format and result in: 1. database structure and application software will be changed; 2. structured query language (SQL) operations cannot work properly, especially in SQL-based fuzzy query. As a result, it is necessary to provide an SQL-based fuzzy query mechanism over encrypted databases, including traditional databases and cloud outsourced databases. This paper establishes a secure database system using format-preserving encryption (FPE) as the underlying primitive to protect the data privacy while not change the database structure. It further proposes a new SQL-based fuzzy query mechanism supporting directly query over encrypted data, which is constructed by FPE and universal hash function (UHF). The security of the proposed mechanism is analyzed as well. In the end, it makes extensive experiments on the system to demonstrate its practical performance. Adapted from the source document.</p>			
País	Título	Autor	Referência
China	Semantic search for public opinions on urban affairs: A probabilistic topic modeling-based approach	Ma, Baojun	MA, Baojun et al. Semantic search for public opinions on urban affairs: A probabilistic topic modeling-based approach. <b>Information Processing &amp; Management</b> , v. 52, n. 3, p. 430-445, 2016.
Abstract			
<p>The explosion of <i>online</i> user-generated content (UGC) and the development of big data analysis provide a new opportunity and challenge to understand and respond to public opinions in the G2C e-government context. To better understand semantic searching of public comments on an <i>online</i> platform for citizens opinions about urban affairs issues, this paper proposed an approach based on the latent Dirichlet allocation (LDA), a probabilistic topic modeling method, and designed a practical system to provide users -- municipal administrators of B-city -- with satisfying searching results and the longitudinal changing curves of related topics. The system is developed to respond to actual demand from B-city's local government, and the user evaluation experiment results show that a system based on the LDA method could provide information that is more helpful to relevant staff members. Municipal administrators could better understand citizen's; <i>online</i> comments based on the proposed semantic search approach and could improve their decision-making process by considering public opinions.</p>			
China Total			6
País	Título	Autor	Referência
Alemanha	A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'.	Huang, Ying	HUANG, Ying et al. A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'. <b>Scientometrics</b> , v. 105, n. 3, p. 2005-2022, 2015.

QUADRO 22 – Artigos classificados como Objeto de Pesquisa, qualificados como Recuperação da Informação – China e Alemanha

(Conclusão)

Abstract			
<p>Bibliometric and 'tech mining' studies depend on a crucial foundation-the search strategy used to retrieve relevant research publication records. Database searches for emerging technologies can be problematic in many respects, for example the rapid evolution of terminology, the use of common phraseology, or the extent of 'legacy technology' terminology. Searching on such legacy terms may or may not pick up R&amp;D pertaining to the emerging technology of interest. A challenge is to assess the relevance of legacy terminology in building an effective search model. Common-usage phraseology additionally confounds certain domains in which broader managerial, public interest, or other considerations are prominent. In contrast, searching for highly technical topics is relatively straightforward. In setting forth to analyze 'Big Data,' we confront all three challenges-emerging terminology, common usage phrasing, and intersecting legacy technologies. In response, we have devised a systematic methodology to help identify research relating to Big Data. This methodology uses complementary search approaches, starting with a Boolean search model and subsequently employs contingency term sets to further refine the selection. The four search approaches considered are: (1) core lexical query, (2) expanded lexical query, (3) specialized journal search, and (4) cited reference analysis. Of special note here is the use of a 'Hit-Ratio' that helps distinguish Big Data elements from less relevant legacy technology terms. We believe that such a systematic search development positions us to do meaningful analyses of Big Data research patterns, connections, and trajectories. Moreover, we suggest that such a systematic search approach can help formulate more replicable searches with high recall and satisfactory precision for other emerging technology studies. [ABSTRACT FROM AUTHOR]</p>			
País	Título	Autor	Referência
Alemanha	PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis	Rill, Sven	RILL, Sven et al. Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. <b>Knowledge-Based Systems</b> , v. 69, p. 24-33, 2014.
Abstract			
<p>In this work, we present a system called PoliTwi, which was designed to detect emerging political topics (Top Topics) in Twitter sooner than other standard information channels. The recognized Top Topics are shared via different channels with the wider public. For the analysis, we have collected about 4,000,000 tweets before and during the parliamentary election 2013 in Germany, from April until September 2013. It is shown, that new topics appearing in Twitter can be detected right after their occurrence. Moreover, we have compared our results to Google Trends. We observed that the topics emerged earlier in Twitter than in Google Trends. Finally, we show how these topics can be used to extend existing knowledge bases (web ontologies or semantic networks) which are required for concept-level sentiment analysis. For this, we utilized special Twitter hashtags, called sentiment hashtags, used by the German community during the parliamentary election.</p>			
<b>Alemanha Total</b>			<b>2</b>
<b>Total Geral</b>			<b>13</b>

Fonte: elaborado pelo autor, 2017.

No Quadro 23 demonstra-se que onze países publicaram 23 artigos que receberam a classificação como OP e a qualificação como TRA:

QUADRO 23 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação

(Continua)

País	Universidade / Instituição	Programa	Artigos
Estados Unidos	Ohio State University	Department of Biomedical Informatics,	1
	University of North Carolina of Chapel Hill	Population Informatics Research Group, Department of Computer Science	1
	Purdue University	Computational Science Information Specialist at Purdue University Libraries	1
	Sam Houston State University	Sam Houston State University	1
	University of Pittsburgh	The RODS Laboratory, Department of Biomedical Informatics	1
	University of Southern California	Division of Interactive Media and Games, Division of Media Arts + Practice in the School of Cinematic Arts	1
	University of Wisconsin	College of Letters & Science English	1
	<b>Estados Unidos Total</b>		<b>7</b>
China	Chinese Academy of Sciences	Institute of Remote Sensing and Digital Earth	1
	National Kaohsiung University of Applied Sciences	Department of Electrical Engineering,	1
	Sichuan Health Information Center	Sichuan Provincial Health Information Center	1
		<b>China Total</b>	<b>3</b>
Itália	European Commission - JRC (IPSC)	European Commission - JRC (IPSC)	1
	Second University of Naples	Second University of Naples	1
	University of Salerno	Institute of High Performance Computing and Networking (ICAR)	1
	<b>Itália Total</b>	<b>3</b>	
Reino Unido	Fujitsu Laboratories of Europe Ltd	Intelligent Society Platform Research Division	1
	<b>Reino Unido Total</b>		<b>1</b>
Canadá	IBM	Canada Research and Development Center	1
	University of Alberta	Health Law Institute, Faculty of Law	1
		<b>Canadá Total</b>	<b>2</b>

QUADRO 23 – Países que publicaram artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação

(Conclusão)

País	Universidade / Instituição	Programa	Artigos
Alemanha	High Performance Computing Center Stuttgart (HLRS)	High Performance Computing Center Stuttgart (HLRS)	1
	University of Siegen	University of Siegen	1
	<b>Alemanha Total</b>		<b>2</b>
Áustria	University of Vienna	University of Vienna	1
	<b>Áustria Total</b>		<b>1</b>
Nova Zelândia	Utrecht University	Department of Information and Computing Sciences	1
	<b>Nova Zelândia Total</b>		<b>1</b>
Argélia	University of Saïda	Department of Computer Sciences	1
	<b>Argélia Total</b>		<b>1</b>
Chile	University of Waikato	Departamento de Cinecia da Computacao	1
	<b>Chile Total</b>		<b>1</b>
Espanha	University of Jaén	Computer Science Department	1
	<b>Espanha Total</b>		<b>1</b>
Dinamarca	University of Copenhagen	Communication and IT, Department of Media, Cognition and Communication	1
	<b>Dinamarca Total</b>		<b>1</b>
<b>Total Geral</b>			<b>23</b>

Fonte: elaborado pelo autor, 2017.

Com sete artigos publicados, os Estados Unidos foram o país que mais publicou artigos classificados como OP e qualificados como TRA. Com três artigos publicados em cada um, seguem: China, Itália e Alemanha. Os seguintes países publicaram um artigo: Áustria, Nova Zelândia, Chile, Argélia, Reino Unido, Espanha e Dinamarca.

Os artigos publicados nos Estados Unidos, China, Itália e Alemanha, que foram os países que mais publicaram artigos nesta qualificação, serão destacados no Quadro 24:

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continua)

País	Título	Autor	Referência
Estados Unidos	An analytics appliance for identifying (near) optimal over-the-counter medicine products as health indicators for influenza surveillance.	Rexit, Ruhsary	REXIT, Ruhsary et al. An analytics appliance for identifying (near) optimal over-the-counter medicine products as health indicators for influenza surveillance. <b>Information Systems</b> , v. 48, p. 151-163, 2015.

**Abstract**

In the era of "Big Data", a challenge is how to optimize our use of huge volumes of data. In this paper, we address this challenge in the context of a public health surveillance system which identifies disease outbreaks using individual and population health indicators. Our goal is to automate and improve the accuracy of the selection process of the health indicators, a process which is data-intensive and computationally expensive. The health indicators selection process traditionally has been carried out manually by public health experts in collaboration with health data providers. In particular, we present an approach for identifying sets of over-the-counter (OTC) medicine products whose aggregate sales correlate optimally with aggregate counts of emergency department (ED) visits. Towards this goal, we propose an OTC Analytics Appliance which utilizes a distributed search engine to efficiently generate time series of time-stamped records and supports "plug-and-play" search and correlation functionalities. Using the OTC Analytics Appliance with the Pearson correlation coefficient function, we evaluate Brute-force search, Greedy search, and Knapsack search for their ability to select the optimal or suboptimal set of OTC products automatically. Our results show that greedy search is the most preferable, producing a set of OTC products whose sales that correlate optimally or near optimally to ED visits, while achieving acceptable search times with large datasets. Also, our evaluations show that our approach using the greedy search can be potentially used to efficiently identify different optimal OTC medicine products for detection of different types of disease outbreaks. [ABSTRACT FROM AUTHOR]

País	Título	Autor	Referência
Estados Unidos	Data Sharing and Re-Use Policies for Webcam Video Feeds from International Sources.	Pouchard, Line C.	POUCHARD, Line C.; NELSON, Megan Sapp; LU, Yung-Hsiang. Data Sharing and Re-Use Policies for Webcam Video Feeds from International Sources. <b>IASSIST QUARTERLY</b> , p. 15, 2015.

**Abstract**

The Continuous Analysis of Many Cameras (CAM2)<sup>2</sup> project is a research project at Purdue University for Big Data and visual analytics. CAM2 has the ability to collect over 60,000 publicly accessible video feeds from many regions around the world. These video feeds were originally collected for improving the scalability of image processing algorithms and are now becoming of interest to ecologists, city planners, and environmentalists. With CAM2's ability to acquire millions of images or many hours of videos per day, collecting these large quantities of data raises questions about data management. The data sources have heterogeneous policies for data use, and some sources have no policies. Separate agreements had to be negotiated between each video stream source and the data collector. In this paper, we propose to compare data use policies that are attached to the video streams and study their implications for open access. The need for common points of legal guidance for webcam stream users and publishers is demonstrated through this analysis of usage agreements. [ABSTRACT FROM AUTHOR]

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continuação)

País	Título	Autor	Referência
Estados Unidos	Detecting near-duplicate text documents with a hybrid approach.	Varol, Cihan	VAROL, Cihan; HARI, Sairam. Detecting near-duplicate text documents with a hybrid approach. <b>Journal of Information Science</b> , v. 41, n. 4, p. 405-414, 2015.

**Abstract**

Near duplicate data not only increase the cost of information processing in big data, but also increase decision time. Therefore, detecting and eliminating nearly identical information is vital to enhance overall business decisions. To identify near-duplicates in large-scale text data, the shingling algorithm has been widely used. This algorithm is based on occurrences of contiguous subsequences of tokens in two or more sets of information, such as in documents. In other words, if there is a slight variation among documents, the overall performance of the algorithm decreases. Therefore, to increase the efficiency and accuracy performances of the shingling algorithm, we propose a hybrid approach that embeds Jaro distance and statistical results of word usage frequency for fixing the ill-defined data. In a real text dataset, the proposed hybrid approach improved the shingling algorithm's accuracy performance by 27% on average and achieved above 90% common shingles. [ABSTRACT FROM AUTHOR]

País	Título	Autor	Referência
Estados Unidos	Privacy preserving interactive record linkage (PPIRL).	Hye-Chung Kum	KUM, Hye-Chung et al. Privacy preserving interactive record linkage (PPIRL). <b>Journal of the American Medical Informatics Association</b> , v. 21, n. 2, p. 212-220, 2014.

**Abstract**

Objective Record linkage to integrate uncoordinated databases is critical in biomedical research using Big Data. Balancing privacy protection against the need for high quality record linkage requires a human-machine hybrid system to safely manage uncertainty in the ever changing streams of chaotic Big Data. Methods In the computer science literature, private record linkage is the most published area. It investigates how to apply a known linkage function safely when linking two tables. However, in practice, the linkage function is rarely known. Thus, there are many data linkage centers whose main role is to be the trusted third party to determine the linkage function manually and link data for research via a master population list for a designated region. Recently, a more flexible computerized third-party linkage platform, Secure Decoupled Linkage (SDLink), has been proposed based on: (1) decoupling data via encryption, (2) obfuscation via chaffing (adding fake data) and universe manipulation; and (3) minimum information disclosure via recoding. Results We synthesize this literature to formalize a new framework for privacy preserving interactive record linkage (PPIRL) with tractable privacy and utility properties and then analyze the literature using this framework. Conclusions Human-based third-party linkage centers for privacy preserving record linkage are the accepted norm internationally. We find that a computer-based third-party platform that can precisely control the information disclosed at the micro level and allow frequent human interaction during the linkage process, an effective human-machine hybrid system that significantly improves on the linkage center model both in terms of privacy and utility. [ABSTRACT FROM AUTHOR]

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continuação)

País	Título	Autor	Referência
Estados Unidos	Statistical Genre Analysis: Toward Big Data Methodologies in Technical Communication	Graham, S Scott	GRAHAM, S. Scott et al. Statistical genre analysis: Toward big data methodologies in technical communication. <b>Technical Communication Quarterly</b> , v. 24, n. 1, p. 70-104, 2015.

**Abstract**

This article pilots a study in statistical genre analysis, a mixed-method approach for (a) identifying conventional responses as a statistical distribution within a big data set and (b) assessing which deviations from the conventional might be more effective for changes in audience, purpose, or context. The study assesses pharmaceutical sponsor presentations at the Food and Drug Administration (FDA) drug advisory committee meetings. Preliminary findings indicate the need for changes to FDA conflict-of-interest policies. Adapted from the source document.

País	Título	Autor	Referência
Estados Unidos	The TOKEn project: knowledge synthesis for in silico science.	Payne, Philip R. O.	PAYNE, Philip R. O. et al. The TOKEn project: knowledge synthesis for in silico science. <b>Journal of the American Medical Informatics Association</b> , v. 18, n. Supplement 1, p. i125-i131, 2011.

**Abstract**

**Objective** The conduct of investigational studies that involve large-scale data sets presents significant challenges related to the discovery and testing of novel hypotheses capable of supporting in silico discovery science. The use of what are known as Conceptual Knowledge Discovery in Databases (CKDD) methods provides a potential means of scaling hypothesis discovery and testing approaches for large data sets. Such methods enable the high-throughput generation and evaluation of knowledge-anchored relationships between complexes of variables found in targeted data sets. **Methods** The authors have conducted a multipart model formulation and validation process, focusing on the development of a methodological and technical approach to using CKDD to support hypothesis discovery for in silico science. The model the authors have developed is known as the Translational Ontology-anchored Knowledge Discovery Engine (TOKEn). This model utilizes a specific CKDD approach known as Constructive Induction to identify and prioritize potential hypotheses related to the meaningful semantic relationships between variables found in large-scale and heterogeneous biomedical data sets. **Results** The authors have verified and validated TOKEn in the context of a translational research data repository maintained by the NCI-funded Chronic Lymphocytic Leukemia Research Consortium. Such studies have shown that TOKEn is: (1) computationally tractable; and (2) able to generate valid and potentially useful hypotheses concerning relationships between phenotypic and biomolecular variables in that data collection. **Conclusions** The TOKEn model represents a potentially useful and systematic approach to knowledge synthesis for in silico discovery science in the context of large-scale and multidimensional research data sets.

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continuação)

País	Título	Autor	Referência
Estados Unidos	Coping with the big data dump: Towards a framework for enhanced information representation.	Kratky, Andreas	KRATKY, Andreas; KUHN, Virginia; EIKENES, Jon Olav. Coping with the big data dump: Towards a framework for enhanced information representation. <b>First Monday</b> , v. 20, n. 6, 2015.

**Abstract**

Easy and efficient access to large amounts of data has become an essential aspect of our everyday life. In this paper we investigate possibilities of supporting information representation through the combined use of multiple modalities of perceptions such as sight, touch and kinesthetics. We present a theoretical framework to analyze these approaches and exemplify our findings with case studies of three emergent projects. The results are a contribution to a larger discussion of multimodal information representation at the intersection of theory and practice. [ABSTRACT FROM AUTHOR]

**Estados Unidos Total**

**7**

País	Título	Autor	Referência
China	Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking	Lee, Chung-Hong	LEE, Chung-Hong; CHIEN, Tzan-Feng. Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. <b>Journal of Information Science</b> , v. 39, n. 4, p. 523-543, 2013.

**Abstract**

Although diverse groups argue about the potential and true value benefits from social-media big data, there is no doubt that the era of big data exploitation has begun, driving the development of novel data-centric applications. Big data is notable not only because of its size, but also because of the complexity caused by its relationality to other data. In the past, owing to the limited possibilities of accessing big data, few data sources were available to allow researchers to develop advanced data-driven applications, such as monitoring of emerging real-world events. In fact, social media is greatly impacting the growth of big data; and big data is providing enterprises with the data to help them understand how to better detect marketing demands. Microblogging is a social network service capable of aggregating messages to explore facts and unknown knowledge. Nowadays, people often attempt to search for trending news and hot topics in real time from microblogging messages to satisfy their information needs. Under such a circumstance, a real demand is to find a way to allow users to organize a large number of microblogging messages into understandable events. In this work, we attempt to tackle such challenges by developing an *online* text-stream clustering approach using a modified density-based clustering model with collected microblogging big data. The system kernel combines three technical components, including a dynamic term weighting scheme, a neighbourhood generation algorithm and an *online* density-based clustering technique. After acquiring detected event topics by the system, our system provides functions for recommending top-priority event information to assist people to effectively organize emerging event data through the developed topic ranking algorithm. [Reprinted by permission of Sage Publications, Ltd., copyright holder.]



QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continuação)

País	Título	Autor	Referência
China	Particle Swarm Optimization based dictionary learning for remote sensing big data	Wang, Lizhe	LEE, Chung-Hong; CHIEN, Tzan-Feng. Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. <b>Journal of Information Science</b> , v. 39, n. 4, p. 523-543, 2013.

**Abstract**

Dictionary learning, which is based on sparse coding, has been frequently applied to many tasks related to remote sensing processes. Recently, many new non-analytic dictionary-learning algorithms have been proposed. Some are based on *online* learning. In *online* learning, data can be sequentially incorporated into the computation process. Therefore, these algorithms can train dictionaries using large-scale remote sensing images. However, their accuracy is decreased for two reasons. On one hand, it is a strategy of updating all atoms at once; on the other, the direction of optimization, such as the gradient, is not well estimated because of the complexity of the data and the model. In this paper, we propose a method of improved *online* dictionary learning based on Particle Swarm Optimization (PSO). In our iterations, we reasonably selected special atoms within the dictionary and then introduced the PSO into the atom-updating stage of the dictionary-learning model. Furthermore, to guide the direction of the optimization, the prior reference data were introduced into the PSO model. As a result, the movement dimension of the particles is reasonably limited and the accuracy and effectiveness of the dictionary are promoted, but without heavy computational burdens. Experiments confirm that our proposed algorithm improves the performance of the algorithm for large-scale remote sensing images, and our method also has a better effect on noise suppression.

País	Título	Autor	Referência
China	R language-based analysis of big data about drugs prescribed in grass root clinics	Wang, Shuai	LEE, Chung-Hong; CHIEN, Tzan-Feng. Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. <b>Journal of Information Science</b> , v. 39, n. 4, p. 523-543, 2013.

**Abstract**

The objective of this paper is to provide the evidence for health management decision-making and rational use of drugs grass root clinics by studying their drug prescription rules. The prescribed drugs in clinics of 5 township health centers from September 2012 to September 2014 were retrieved from The Management Information System of Sichuan Grass Root Medical Institutions. Their big data were analyzed using R language. The commonly prescribed drugs in clinics were vitamin B6, vitamin C and cefixime tablets, which were usually used in combination. Health administrative organizations can strengthen their supervision and management of prescribed drugs and promote their rational use in grass root clinics using unified management information system of grass root medical institutions in combination with information technology.

**China Total**

**3**

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continuação)

País	Título	Autor	Referência
Itália	A Distributed and Scalable Solution for Applying Semantic Techniques to Big Data	Amato, Alba	AMATO, Alba; VENTICINQUE, Salvatore; DI MARTINO, Beniamino. A Distributed and Scalable Solution for Applying Semantic Techniques to Big Data. In: INFORMATION RESOURCES MANAGEMENT ASSOCIATION (Ed.). <b>Big Data: Concepts, Methodologies, Tools, and Applications</b> . [S.l.]: IGI Global, 2016. p. 1091-1109.

**Abstract**

The digital revolution changes the way culture and places could be lived. It allows users to interact with the environment creating an immense availability of data, which can be used to better understand the behavior of visitors, as well as to learn about their thoughts on what the visit creates excitement or disappointment. In this context, Big Data becomes immensely important, making possible to turn this amount of data in information, knowledge, and, ultimately, wisdom. This paper aims at modeling and designing a scalable solution that integrates semantic techniques with Cloud and Big Data technologies to deliver context aware services in the application domain of the cultural heritage. The authors started from a baseline framework that originally was not conceived to scale when huge workloads, related to big data, must be processed. They provide an original formulation of the problem and an original software architecture that fulfills both functional and not-functional requirements. The authors present the technological stack and the implementation of a proof of concept.

País	Título	Autor	Referência
Itália	A knowledge-based platform for Big Data analytics based on publish/subscribe services and stream processing	Esposito, Christian	ESPOSITO, Christian et al. A knowledge-based platform for Big Data analytics based on publish/subscribe services and stream processing. <b>Knowledge-Based Systems</b> , v. 79, p. 3-17, 2015.

**Abstract**

Big Data analytics is considered an imperative aspect to be further improved in order to increase the operating margin of both public and private enterprises, and represents the next frontier for their innovation, competition, and productivity. Big Data are typically produced in different sectors of the above organizations, often geographically distributed throughout the world, and are characterized by a large size and variety. Therefore, there is a strong need for platforms handling larger and larger amounts of data in contexts characterized by complex event processing systems and multiple heterogeneous sources, dealing with the various issues related to efficiently disseminating, collecting and analyzing them in a fully distributed way. In such a scenario, this work proposes a way to overcome two fundamental issues: data heterogeneity and advanced processing capabilities. We present a knowledge-based solution for Big Data analytics, which consists in applying automatic schema mapping to face with data heterogeneity, as well as ontology extraction and semantic inference to support innovative processing. Such a solution, based on the publish/subscribe paradigm, has been evaluated within the context of a simple experimental proof-of-concept in order to determine its performance and effectiveness.

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Continuação)

País	Título	Autor	Referência
Itália	An intelligent Web agent that autonomously learns how to translate.	Turchi, Marco	TURCHI, Marco; BIE, Tijl; CRISTIANINI, Nello. An intelligent Web agent that autonomously learns how to translate. <b>Web Intelligence and Agent Systems: An International Journal</b> , v. 10, n. 2, p. 165-178, 2012.
<b>Abstract</b>			
<p>We describe the design of an autonomous agent that can teach itself how to translate from a foreign language, by first assembling its own training set, then using it to improve its vocabulary and language model. The key idea is that a Statistical Machine Translation package can be used for the Cross-Language Retrieval Task of assembling a training set from a vast amount of available text (e.g. a large multilingual corpus, or the Web) and then train on that data, repeating the process several times. The stability issues related to such a feedback loop are addressed by a mathematical model, connecting statistical and control-theoretic aspects of the system. We test it on controlled environment and real-world tasks, showing that indeed this agent can improve its translation performance autonomously and in a stable fashion, when seeded with a very small initial training set. We develop a multiprocessor version of the agent that directly accesses the Web using a Web search engine and taking advantage of the big amount of data available there. The modelling approach we develop for this agent is general, and we believe that it will be useful for an entire class of self-learning autonomous agents working</p>			
<b>Itália Total</b>			<b>3</b>
País	Título	Autor	Referência
Alemanha	A Keyquery-Based Classification System for CORE	Michael Völske	VÖLSKE, Michael et al. A Keyquery-Based Classification System for CORE. <b>D-Lib Magazine</b> , v. 20, n. 11/12, 2014.
<b>Abstract</b>			

We apply keyquery-based taxonomy composition to compute a classification system for the CORE dataset, a shared crawl of about 850,000 scientific papers. Keyquery-based taxonomy composition can be understood as a two-phase hierarchical document clustering technique that utilizes search queries as cluster labels: In a first phase, the document collection is indexed by a reference search engine, and the documents are tagged with the search queries they are relevant -- for their so-called keyqueries. In a second phase, a hierarchical clustering is formed from the keyqueries within an iterative process. We use the explicit topic model ESA as document retrieval model in order to index the CORE dataset in the reference search engine. Under the ESA retrieval model, documents are represented as vectors of similarities to Wikipedia articles; a methodology proven to be advantageous for text categorization tasks. Our paper presents the generated taxonomy and reports on quantitative properties such as document coverage and processing requirements.

QUADRO 24 – Artigos classificados como Objeto de Pesquisa e qualificados como Tratamento da Informação – Estados Unidos, China, Itália e Alemanha

(Conclusão)

País	Título	Autor	Referência
Alemanha	Knowledge discovery from social media using big data-provided sentiment analysis (SoMABiT).	Bohlouli, Mahdi	BOHLOULI, Mahdi et al. Knowledge discovery from social media using big data-provided sentiment analysis (SoMABiT). <b>Journal of Information Science</b> , v. 41, n. 6, p. 779-798, 2015.

**Abstract**

In today's competitive business world, being aware of customer needs and market-oriented production is a key success factor for industries. To this aim, the use of efficient analytical algorithms ensures better understanding of customer feedback and improves the next generation of products. Accordingly, the dramatic increase in the use of social media in daily life provides beneficial sources for market analytics. Yet how traditional analytic algorithms and methods can be scaled up for such disparate and multistructured data sources is a major challenge. This paper presents and discusses the technological and scientific focus of SoMABiT as a social media analysis platform using big data technology. Sentiment analysis has been employed in order to discover knowledge from social media. The use of MapReduce and the development of a distributed algorithm towards an integrated platform that can scale for any data volume and provide social media-driven knowledge is the main novelty of the proposed concept in comparison to the state-of-the-art technologies. [ABSTRACT FROM AUTHOR]

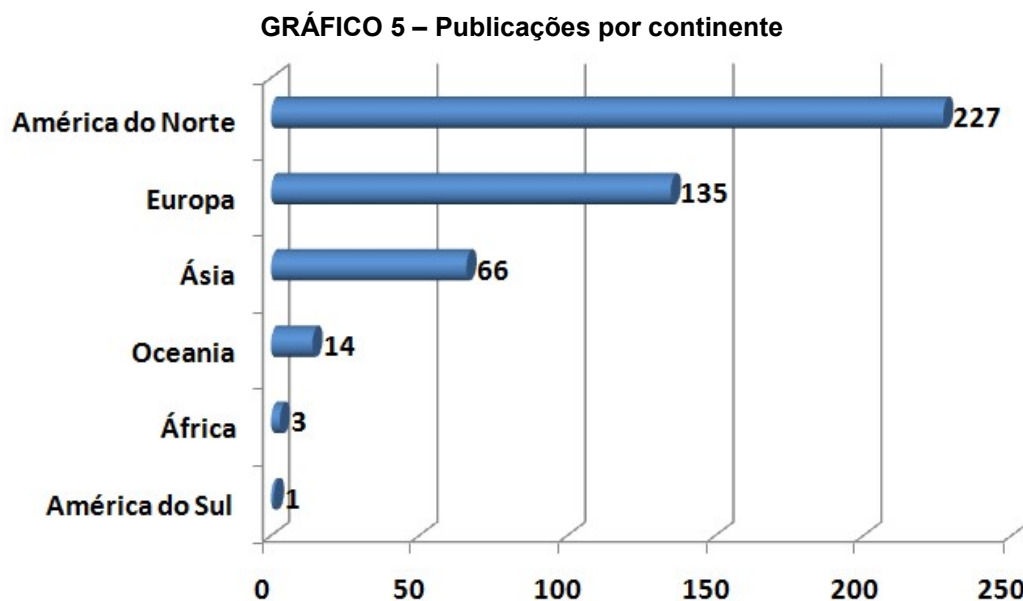
País	Título	Autor	Referência
Alemanha	Service-Oriented Development of Workflow-Based Semantic Reasoning Applications	Cheptsov, Alexey	CHEPTSOV, Alexey; WESNER, Stefan; KOLLER, Bastian. Service-Oriented Development of Workflow-Based Semantic Reasoning Applications. <b>International Journal of Distributed Systems and Technologies (IJ DST)</b> , v. 5, n. 1, p. 40-53, 2014.

**Abstract**

The modern Semantic Web scenarios require reasoning algorithms to be flexible, modular, and highly-configurable. A solid approach, followed in the design of the most currently existing reasoners, is not sufficient when dealing with today's challenges of data analysis across multiple sources of heterogeneous data or when the data amount grows to the 'Big Data' sizes. The 'reasoning as a workflow' concept has attracted a lot of attention in the design of new-generation Semantic Web applications, offering a lot of opportunities to improve both flexibility and scalability of the reasoning process. Considering a single workflow component as a service offers a lot of opportunities for a reasoning algorithm to target a much wider range of potentially enabled Semantic Web use cases by taking benefits of a service-oriented and component-based implementation. We introduce a technique for developing service-oriented Semantic Reasoning applications based on the workflow concept. We also present the Large Knowledge Collider -- a software platform for developing workflow-based Semantic Web applications, taking advantages of on-demand high performance computing and cloud infrastructures. Adapted from the source document.

Fonte: elaborado pelo autor, 2017.

O GRÁFICO 5 demonstra a distribuição das publicações através dos continentes:



Fonte: elaborado pelo autor, 2017.

No GRÁFICO 5 percebe-se que a América do Norte foi o continente que mais publicou artigos relacionados ao *Big Data*. Observa-se o seguinte comportamento das publicações:

América do Norte: publicou 227 artigos, o que equivale a 51% do volume total dos artigos do universo da pesquisa.

Europa: publicou 135 artigos o que equivalente a 30% dos artigos do universo da pesquisa.

Ásia: publicou 66 artigos, o que equivale a 15% do universo da pesquisa.

Oceania: publicou quatorze artigos, o que equivale a 3% do universo da pesquisa.

África: publicou três artigos, o que equivale a 1% do universo da pesquisa.

América do Sul: publicou um artigo.

O Quadro 25 demonstra o comportamento das publicações em cada continente:

QUADRO 25 – Comportamento das publicações por continente

Continente	Artigos	Continente	Artigos
América do Norte	227	RE	10
AS	125	Oceania	14
OP	10	AS	9
RE	92	OP	2
Europa	135	RE	3
AS	78	África	3
OP	20	OP	1
RE	37	RE	2
Ásia	66	América do Sul	1
AS	43	OP	1
OP	13	<b>TOTAL</b>	<b>446</b>

Fonte: elaborado pelo autor, 2017.

Com 227 artigos publicados, a América do Norte é o continente que mais publicou artigos relacionados ao *Big Data*, seguidos pela Europa e Àsia com respectivamente 135 e 66 artigos publicados.

Mesmo publicando um número menor de artigos, a Ásia destaca-se como maior produtor de artigos classificados como OP.

Percebe-se a presença do segmento corporativo na elaboração de artigos relacionados ao *Big Data*. Juntas, as companhias de diversos países publicaram 20 artigos conforme mostra o Quadro 26:

QUADRO 26 – Empresas com publicações relacionadas ao *Big Data* – bases LISA e LISTA 2010 a 2016

(Continua)

País	Classificação	Empresa	Artigos
Estados Unidos	AS	Cycorp Incorporation	1
		HP Labs	1
		IBM	1
		Intel Corporation, Hillsboro, OR, USA, Intel Corporation, Santa Clara, CA, USA	1
		Larson & Associates	1
		PricewaterhouseCoopers	1
		<b>AS Total</b>	<b>6</b>
	RE	IBM	1
		Microsoft Research	3
		Oak Ridge National Laboratory	1
		VALVE Software (Company)	1
	<b>RE Total</b>	<b>6</b>	
	<b>Estados Unidos Total</b>		

QUADRO 26 – Empresas com publicações relacionadas ao *Big Data* – bases LISA e LISTA 2010 a 2016

			(conclusão)
Reino Unido	AS	Beaworthy Consulting	1
		intranetfocus.com	1
	AS Total		2
	OP	Fujitsu Laboratories of Europe Ltd	1
	OP Total		1
	AS	intranetfocus.com	1
	RE Total		1
<b>Reino Unido Total</b>			<b>4</b>
Letônia	AS	RIX Technologies	1
	AS Total		1
	<b>Letônia Total</b>		<b>1</b>
China	AS	CoMed Technology & Consulting Co., Ltd., Hong Kong, China	1
	AS Total		1
	<b>China Total</b>		<b>1</b>
Canadá	OP	IBM	1
	OP Total		1
	<b>Canadá Total</b>		<b>1</b>
Coréia do Sul	AS	Samsung Electronics	1
	AS Total		1
	<b>Coréia do Sul Total</b>		<b>1</b>
<b>Total Geral</b>			<b>20</b>

Fonte: elaborado pelo autor, 2017.

No Quadro 26, percebe-se que quinze companhias, distribuídas por seis países publicaram vinte artigos, o que equivale a 6,51% das universidades/instituições e 4,47% dos artigos do universo da pesquisa. *Microsoft* foi a empresa que mais publicou, com três artigos. Percebe-se o seguinte comportamento destas produções:

Estados Unidos: nove empresas publicaram o total de doze artigos. Seis empresas publicaram o total de seis artigos, que receberam a classificação como AS. Quatro empresas publicaram seis artigos que foram classificados como RE.

Reino Unido: quatro empresas publicaram o total de quatro artigos. Dois artigos receberam a classificação como AS, um artigo recebeu a classificação como OP e uma empresa publicou um artigo que recebeu a classificação como RE.

Letônia: uma empresa publicou um artigo classificado como AS.

China: uma empresa publicou um artigo classificado como AS.

Canadá: uma empresa publicou um artigo classificado como OP.

Coréia do Sul: uma empresa publicou um artigo classificado como AS.

O Quadro 27 destaca os artigos publicados pelas empresas que mais publicaram artigos: Microsoft e IBM:

QUADRO 27 – Empresas que mais publicaram e seus artigos

(Continua)

Empresa	Autor	Titulo	Referência
IBM	Khazaei, Hamzeh	Real-Time and Retrospective Health-Analytics-as-a-Service: A Novel Framework.	KHAZAEI, Hamzeh et al. Real-time and retrospective health-analytics-as-a-service: a novel framework. <b>JMIR medical informatics</b> , v. 3, n. 4, <i>online</i> , 2015.

#### Abstract

Background: Analytics-as-a-service (AaaS) is one of the latest provisions emerging from the cloud services family. Utilizing this paradigm of computing in health informatics will benefit patients, care providers, and governments significantly. This work is a novel approach to realize health analytics as services in critical care units in particular. Objective: To design, implement, evaluate, and deploy an extendable big-data compatible framework for health analytics as a service that offers both real-time and retrospective analysis. Methods: We present a novel framework that can realize health data analytics-as-a-service. The framework is flexible and configurable for different scenarios by utilizing the latest technologies and best practices for data acquisition, transformation, storage, analytics, knowledge extraction, and visualization. We have instantiated the proposed method, through the Artemis project, that is, a customization of the framework for live monitoring and retrospective research on premature babies and ill term infants in neonatal intensive care units (NICUs). Results: We demonstrated the proposed framework in this paper for monitoring NICUs and refer to it as the Artemis-In-Cloud (Artemis-IC) project. A pilot of Artemis has been deployed in the SickKids hospital NICU. By infusing the output of this pilot set up to an analytical model, we predict important performance measures for the final deployment of Artemis-IC. This process can be carried out for other hospitals following the same steps with minimal effort. SickKids'NICU has 36 beds and can classify the patients generally into 5 different types including surgical and premature babies. The arrival rate is estimated as 4.5 patients per day, and the average length of stay was calculated as 16 days. Mean number of medical monitoring algorithms per patient is 9, which renders 311 live algorithms for the whole NICU running on the framework. The memory and computation power required for Artemis-IC to handle the SickKids NICU will be 32 GB and 16 CPU cores, respectively. The required amount of storage was estimated as 8.6 TB per year. There will always be 34.9 patients in SickKids NICU on average. Currently, 46% of patients cannot get admitted to SickKids NICU due to lack of resources. By increasing the capacity to 90 beds, all patients can be accommodated. For such a provisioning, Artemis-IC will need 16 TB of storage per year, 55 GB of memory, and 28 CPU cores. Conclusions: Our contributions in this work relate to a cloud architecture for the analysis of physiological data for clinical decisions support for tertiary care use. We demonstrate how to size the equipment needed in the cloud for that architecture based on a very realistic assessment of the patient characteristics and the associated clinical decision support algorithms that would be required to run for those patients. We show the principle of how this could be performed and furthermore that it can be replicated for any critical care setting within a tertiary institution.

Empresa	Autor	Titulo	Referência
IBM	Miller, Steven	Preparing the next generation for the cognitive era.	MILLER, Steven. Preparing the next generation for the cognitive era. <b>Information Services &amp; Use</b> , v. 36, n. 1/2, p. 23-25, 2016.



## QUADRO 27 – Empresas que mais publicaram e seus artigos

(Continuação)

Abstract			
<p>After decades of data scarcity, we are finally arriving in the era of data abundance. Cognitive systems such as IBM's Watson, a cloud-based cognition service platform, can help us utilize all the data available to make better decisions in business, society, and our personal lives. But even with the help of cognitive assistants, data literacy will be a key new skill for students and professionals, as data transforms industries and occupations, and creates enormous entrepreneurial and startup opportunities. This brief paper highlights some current data-driven innovations. [ABSTRACT FROM AUTHOR]</p>			
Empresa	Autor	Titulo	Referência
IBM	Spohrer, Jim	Cognition as a Service: An Industry Perspective	SPOHRER, Jim; BANAVAR, Guruduth. Cognition as a service: an industry perspective. <b>AI Magazine</b> , v. 36, n. 4, p. 71-87, 2015.
Abstract			
<p>Recent advances in cognitive computing componentry combined with other factors are leading to commercially viable cognitive systems. From chips to smart phones to public and private clouds, industrial strength cognition as a service is beginning to appear at all scales in business and society. Furthermore, in the age of zettabytes on the way to yottabytes, the designers, engineers, and managers of future smart systems will depend on cognition as a service. Cognition as a service can help unlock the mysteries of big data and ultimately boost the creativity and productivity of professionals and their teams, the productive output of industries and organizations, as well as the gross domestic product of regions and nations. In this and the next decade, cognition as a service will allow us to reimage work practices, augmenting and scaling expertise to transform professions, industries, and regions.</p>			
Empresa	Autor	Titulo	Referência
IBM	Spohrer, Jim	Cognition as a Service: An Industry Perspective	SPOHRER, Jim; BANAVAR, Guruduth. Cognition as a service: an industry perspective. <b>AI Magazine</b> , v. 36, n. 4, p. 71-87, 2015.
Abstract			
<p>Recent advances in cognitive computing componentry combined with other factors are leading to commercially viable cognitive systems. From chips to smart phones to public and private clouds, industrial strength cognition as a service is beginning to appear at all scales in business and society. Furthermore, in the age of zettabytes on the way to yottabytes, the designers, engineers, and managers of future smart systems will depend on cognition as a service. Cognition as a service can help unlock the mysteries of big data and ultimately boost the creativity and productivity of professionals and their teams, the productive output of industries and organizations, as well as the gross domestic product of regions and nations. In this and the next decade, cognition as a service will allow us to reimage work practices, augmenting and scaling expertise to transform professions, industries, and regions.</p>			
IBM Total			3
Empresa	Autor	Titulo	Referência
Microsoft Research	Baym, Nancy K	Data not seen: The uses and shortcomings of social media metrics	BAYM, Nancy K. Data not seen: The uses and shortcomings of social media metrics. <b>First Monday</b> , v. 18, n. 10, 2013.

## QUADRO 27 – Empresas que mais publicaram e seus artigos

(Conclusão)

Abstract			
<p>This paper looks at the use of <i>online</i> data, especially social media metrics, to assess media audiences, with particular focus on musicians. It shows how audiences are defined by different sets of people, and grounds the use of social media to understand audiences in the history of mass media audience measurement. The second half of the paper focuses on visible social media metrics -- likes, followers, and other such counts -- outlining their appeal as measures and highlighting their fallibility and ambiguity. Throughout, the paper argues that different people construct information systems to collect, store, analyze, and interpret data, and that these are shaped by value systems. Metric and big data analysis generally serves economic values, while other approaches to data may be more appropriate for assessing social and personal values. Adapted from the source document.</p>			
Empresa	Autor	Titulo	Referência
Microsoft Research	Boyd, Danah	CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon	BOYD, Danah; CRAWFORD, Kate. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. <b>Information, communication &amp; society</b> , v. 15, n. 5, p. 662-679, 2012.
Abstract			
<p>The era of Big Data has begun. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions. Diverse groups argue about the potential benefits and costs of analyzing genetic sequences, social media interactions, health records, phone logs, government records, and other digital traces left by people. Significant questions emerge. Will large-scale search data help us create better tools, services, and public goods? Or will it usher in a new wave of privacy incursions and invasive marketing? Will data analytics help us understand <i>online</i> communities and political movements? Or will it be used to track protesters and suppress speech? Will it transform how we study human communication and culture, or narrow the palette of research options and alter what 'research' means? Given the rise of Big Data as a socio-technical phenomenon, we argue that it is necessary to critically interrogate its assumptions and biases. In this article, we offer six provocations to spark conversations about the issues of Big Data: a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric. Adapted from the source document.</p>			
Empresa	Autor	Titulo	Referência
Microsoft Research	Marshall, Cathy	Big Data, the crowd and me.	MARSHALL, Cathy. Big Data, the Crowd, and Me. <b>Information Services and Use</b> , v. 32, n. 3/4, p. 215-226, 2012.
Abstract			
<p>The article presents the author's views on the implications of analysis and manipulation of big data. She states that researchers from different disciplines are seduced to the availability of big data despite its limitations. She offers information on big data which refers to digital data from various sources including the Internet, video, and social network communications. She discusses how she conducted a study which involved the microblogging site Twitter to analyze big data.</p>			
<b>Microsoft Research Total</b>			<b>3</b>
<b>Total Geral</b>			<b>6</b>

Fonte: elaborado pelo autor, 2017.

O Quadro 28 apresenta as empresas que publicaram artigos classificados como OP:

QUADRO 28 – Empresas que publicaram artigos classificados como Objeto de Pesquisa

(Continua)

Empresa	Título	Qualificação	Autor	Referência
Fujitsu Laboratories of Europe Ltd Reino Unido	Towards Big Linked Data: A Large-Scale, Distributed Semantic Data Storage	TRA	Hu, Bo	HU, Bo et al. Towards big linked data: a large-scale, distributed semantic data storage. <b>International Journal of Data Warehousing and Mining</b> , v. 9, n. 4, p. 19-43, Oct. 2013.
<b>Abstract</b>				
<p>In light of the challenges of effectively managing Big Data, the authors are witnessing a gradual shift towards the increasingly popular Linked Open Data (LOD) paradigm. LOD aims to impose a machine-readable semantic layer over structured as well as unstructured data and hence automate some data analysis tasks that are not designed for computers. The convergence of Big Data and LOD is, however, not straightforward: the semantic layer of LOD and the Big Data large scale storage do not get along easily. Meanwhile, the sheer data size envisioned by Big Data denies certain computationally expensive semantic technologies, rendering the latter much less efficient than their performance on relatively small data sets. In this paper, the authors propose a mechanism allowing LOD to take advantage of existing large-scale data stores while sustaining its 'semantic' nature. The authors demonstrate how RDF-based semantic models can be distributed across multiple storage servers and the authors examine how a fundamental semantic operation can be tuned to meet the requirements on distributed and parallel data processing. The authors' future work will focus on stress test of the platform in the magnitude of tens of billions of triples, as well as comparative studies in usability and performance against similar offerings. Adapted from the source document.</p>				
Empresa	Título	Qualificação	Autor	Referência
IBM Canadá	Real-Time and Retrospective Health-Analytics-as-a-Service: A Novel Framework.	TRA	Khazaei, Hamzeh	KHAZAEI, Hamzeh et al. Real-time and retrospective health-analytics-as-a-service: a novel framework. <b>JMIR medical informatics</b> , v. 3, n. 4, p. e36, 2015.

## QUADRO 28 – Empresas que publicaram artigos classificados como Objeto de Pesquisa

(Conclusão)

<b>Abstract</b>	
<p>Background: Analytics-as-a-service (AaaS) is one of the latest provisions emerging from the cloud services family. Utilizing this paradigm of computing in health informatics will benefit patients, care providers, and governments significantly. This work is a novel approach to realize health analytics as services in critical care units in particular. Objective: To design, implement, evaluate, and deploy an extendable big-data compatible framework for health-analytics-as-a-service that offers both real-time and retrospective analysis. Methods: We present a novel framework that can realize health data analytics-as-a-service. The framework is flexible and configurable for different scenarios by utilizing the latest technologies and best practices for data acquisition, transformation, storage, analytics, knowledge extraction, and visualization. We have instantiated the proposed method, through the Artemis project, that is, a customization of the framework for live monitoring and retrospective research on premature babies and ill terminants in neonatal intensive care units (NICUs). Results: We demonstrated the proposed framework in this paper for monitoring NICUs and refer to it as the Artemis-In-Cloud (Artemis-IC) project. A pilot of Artemis has been deployed in the SickKids hospital NICU. By infusing the output of this pilot set up to an analytical model, we predict important performance measures for the final deployment of Artemis-IC. This process can be carried out for other hospitals following the same steps with minimal effort. SickKids' NICU has 36 beds and can classify the patients generally into 5 different types including surgical and premature babies. The arrival rate is estimated as 4.5 patients per day, and the average length of stay was calculated as 16 days. Mean number of medical monitoring algorithms per patient is 9, which renders 311 live algorithms for the whole NICU running on the framework. The memory and computation power required for Artemis-IC to handle the SickKids NICU will be 32 GB and 16 CPU cores, respectively. The required amount of storage was estimated as 8.6 TB per year. There will always be 34.9 patients in SickKids NICU on average. Currently, 46% of patients cannot get admitted to SickKids NICU due to lack of resources. By increasing the capacity to 90 beds, all patients can be accommodated. For such a provisioning, Artemis-IC will need 16 TB of storage per year, 55 GB of memory, and 28 CPU cores. Conclusions: Our contributions in this work relate to a cloud architecture for the analysis of physiological data for clinical decision support for tertiary care use. We demonstrate how to size the equipment needed in the cloud for that architecture based on a very realistic assessment of the patient characteristics and the associated clinical decision support algorithms that would be required to run for those patients. We show the principle of how this could be performed and furthermore that it can be replicated for any critical care setting within a tertiary institution.</p>	
<b>Total Geral</b>	<b>2</b>

Fonte: elaborado pelo autor, 2017.

No Quadro 28 percebe-se que duas companhias, distribuídas por dois países publicaram dois artigos, classificados como OP, o que equivale a 10% da publicação das empresas presentes neste levantamento e 3,70% das publicações que receberam a classificação como OP. No Reino Unido, a *Fujitsu Laboratories of Europe Ltd*, publicou um artigo que recebeu a qualificação como TRA, e no Canadá, a IBM publicou um artigo que recebeu a qualificação como TRA.

Detectou-se nesta pesquisa artigos procedentes de diversas outras ciências. Lança-se um olhar especial sobre os artigos classificados como OP, pois referem-se a pesquisas relacionadas ao fenômeno *Big Data* publicadas em espaços especializados à CI.

No Quadro 29 demonstra-se quais ciências publicaram artigos classificados como OP. Para a confecção deste quadro, adotou-se a declaração do título do programa ou departamento de origem do artigo.

QUADRO 29 – Ciências que publicaram artigos classificados como Objeto de Pesquisa – bases LISA e LISTA - 2010 a 2016

<b>Ciência</b>	<b>Artigos</b>	<b>Ciência</b>	<b>Artigos</b>
Computação	26	Letras	1
Medicina	4	Estatística	1
Biblioteconomia	3	Matemática	1
Economia	2	Bioengenharia	1
Engenharia	2	Comunicação	1
Ciência da Informação	1	Artes	1
Multidisciplinar	2	Direito	1
		<b>Total Geral</b>	<b>47</b>

Fonte: elaborado pelo autor, 2017.

O Quadro 30 apresenta a distribuição dos artigos classificados como OP, agrupados pela concentração da ciência que publicou:

QUADRO 30 – Distribuição dos artigos classificados como Objeto de Pesquisa – por área de conhecimento

<b>Área de conhecimento</b>	<b>Artigos</b>	<b>Área de conhecimento</b>	<b>Artigos</b>
Exatas e da terra	31	Engenharia Biomédica	1
Ciências da saúde	4	Letras	1
Ciência da Informação	3	Comunicação	1
Economia	2	Artes	1
Multidisciplinar	2	Direito	1
		<b>Total Geral</b>	<b>47</b>

Fonte: elaborado pelo autor, 2017.

A classificação do Quadro 30 foi realizada com base nas orientações contidas no Portal Capes, quanto às áreas de conhecimento (COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2017).

Foram publicados 47 artigos classificados como OP, com artigos provenientes de quatorze áreas de conhecimento, dentre as quais uma é nomeada como multidisciplinar, por agregar tantas ciências quantas forem necessárias à conclusão de suas pesquisas, composta pelas contribuições das seguintes instituições:

- a) European Commission - JRC (IPSC) - Itália; e
- b) Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences – China.

Ciências exatas e da terra foi a concentração que mais publicou, com 31 artigos, o equivalente a 65,95% do volume publicado. Ciências da saúde publicaram quatro artigos, o equivalente a 8,51% do volume publicado, seguido pela CI com três artigos publicados, o

equivalente a 6,38%. Seguem-se Economia e Multidisciplinar com dois artigos publicados, cada. Engenharia Biomédica, Letras, Comunicação, Artes e Direito publicaram um artigo cada, o equivalente a 2,12%.

Verifica-se portanto que as ciências exatas e da terra, em especial a computação, são as que mais publicam artigos relacionados ao *Big Data* e classificados como OP.

As três pesquisas que receberam a qualificação como DIS foram publicadas por programas como áreas da medicina e direito, e descrevem mecanismos e ferramentas para concentrar e compartilhar conhecimento técnico, através da criação de ferramentas tecnológicas digitais para a disseminação de grandes volumes de informação.

Quanto às pesquisas que receberam a qualificação como INF, praticamente todas são emanadas de programas da área de ciências exatas, como a computação, matemática e estatística, buscando mecanismos que proporcionem melhora na performance dos ambientes computacionais. Muitos artigos retratam ferramentas voltadas à otimização de processamento através de processamento computacional distribuído, ferramentas que otimizem o tráfego de dados, ou ferramentas que explorem de maneira otimizada as tecnologias que suportam os ambientes computacionais de *Big Data*. Destacadamente, percebe-se uma preocupação com a capacidade, a velocidade de processamento do volume informacional crescente e com a robustez do parque computacional relacionado ao *Big Data*.

As pesquisas qualificadas como REC abordam ferramentas e algoritmos voltados à recuperação de dados em grandes bases de dados, inclusive governamentais. A maioria dos artigos tem origem em escolas de computação, excetuando-se dois de origem na biblioteconomia e um de origem em um departamento de estudos em economia. Nota-se a preocupação com a assertividade e com o volume dos resultados das recuperações, através do emprego de tecnologias computacionais, como a clusterização, computação em nuvem e técnicas de programação, como a utilização de lógica *fuzzy*. Percebe-se também a preocupação com o usuário, levando em consideração seu contexto de pesquisa, e a relevância das respostas ao mesmo.

As pesquisas que receberam a qualificação como TRA, abordam pesquisas relacionadas a ferramentas ou *frameworks* destinados ao tratamento da informação, muitas vezes relacionada ao *data analytics*. Esse último assim definido por Runler (2012, p. 2, tradução nossa), "*Data analytics* é definido como a aplicação ou sistema computacional para a análise de grandes bases de dados para o suporte a decisões". Essas pesquisas são em sua maioria originadas em programas ou departamentos de computação ou engenharia. Inserem-se aqui os artigos que abordaram pesquisas relacionadas ao aprendizado de máquina, tratamento de dados relacionados à medicina e uso de medicamentos. Verifica-se

algumas pesquisas com desenvolvimento de ferramentas e/ou algoritmos para análise de sentimentos.

Autores de outras ciências publicando pesquisas em espaços especializados da CI pode ser uma demonstração da interdisciplinaridade desta ciência, conforme reconhecido por Saracevic (1999). Entende-se que o relacionamento próximo com outras ciências não constitui adversidade ou novidade para essa ciência e sim, a característica que permite à mesma adaptar-se continuamente e buscar novos conceitos diante da dinâmica do contexto informacional contemporâneo, contextualizado por Silva e Ribeiro (2004), com sua frequente criação de novos suportes informacionais, tecnologias, interesses, mudança no comportamento de usuários. Tal cenário exigirá dos profissionais da informação novos conhecimentos e novas habilidades, conforme Hey (2010) e Ohira e Prado (2004).

A presença de autores de outras ciências também pode indicar a quebra de um paradigma em relação ao novo perfil de quem pratica essa ciência. Khun (1996, p. 13) assim define paradigma: “[...] são as realizações científicas universalmente reconhecidas que, durante algum tempo, fornece problemas e soluções modelares para a comunidade de praticantes de uma ciência”. A presença interdisciplinar e busca por solução às questões relacionadas à informação pode estar nos conduzindo naturalmente ao novo conceito do que é o cientista da informação. Kuhn (2006, p. 77) acrescenta: “Quanto maiores forem a precisão e o alcance de um paradigma, tanto mais sensível este será como indicador de anomalias e, conseqüentemente de uma ocasião para a mudança de paradigma”.

Saracevic (1999, p. 1052) destaca que além de ser uma ciência interdisciplinar, próxima à tecnologia da informação, a CI é “parte ativa na evolução da sociedade da informação”. Nesse contexto, Dias (2002) demonstra preocupação com a interdisciplinaridade que pode ofuscar os limites do que é a CI e suas atribuições. Corroborando esta ideia, Souza, Almeida e Baracho (2013) comentam sobre a existência de um “esvaziamento conceitual” em função da interdisciplinaridade e hospitalidade dessa ciência.

Considerando-se a face interdisciplinar da CI, além do reconhecimento da exigência constante de novos conhecimentos ao profissional da informação, pode-se estar diante de uma dicotomia em assumir que a interdisciplinaridade pode estar “esvaziando” esta ciência ou assumir que essa característica é na verdade a grande raiz que se lança em busca dos conhecimentos e recursos (inclusive humanos) necessários a essa ciência e a mantém em sintonia com as questões referentes ao dado, à informação e ao conhecimento na contemporaneidade.

No universo desta pesquisa percebem-se estudos referentes a ações de recuperação da informação, tratamento, disseminação e alguns artigos dizem respeito à tecnologia necessária para suportar as questões anteriores (infraestrutura). Borko (1968)

legítima a atuação desta ciência com as temáticas tecnológicas atuais ao comentar que esta ciência:

Preocupa-se com o corpo de conhecimento relacionada à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão e utilização da informação. Incluem investigação da representação da informação em sistemas naturais e artificiais, o uso de códigos para transmissão eficiente de mensagens, estudos de dispositivos de processamento de dados e técnicas como computadores e seus sistemas de programação. É uma ciência derivada e relacionada a campos como matemática, lógica, linguística, psicologia, tecnologia computacional, pesquisa operacional, artes gráficas, comunicação, biblioteconomia, gestão e campos similares. (BORKO, 1968, p. 3, tradução nossa)

Portanto, apoiados em Borko (1968), Kuhn (2006) e Saracevic (1999), pode-se entender que os artigos identificados no universo desta pesquisa, inserem-se de forma legítima nos domínios da CI, considerando o conjunto de conhecimentos necessários, próprios da contemporaneidade, sugerindo o surgimento quase instintivo do que denomina-se *iSchool*.

## 4.2 Análise da produção científica nas bases BRAPCI e PERI

O Quadro 31, apresenta a produção científica brasileira relacionada ao *Big Data*, nas Bases BRAPCI e PERI, entre 2010 e 2015:

QUADRO 31 – Produção acadêmica no Brasil

(Continua)

Classificação AS				
Ano	Programa	Universidade/ Instituição	Autor	Título
2013	PPGCI – Instituto Brasileiro de Informação em Ciência da Informação (IBICT) – UFRJ	Universidade Federal do Rio de Janeiro (UFRJ)	Pimenta, R. M.	Big data e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do Estado

### Abstract

Este artigo é produto da pesquisa, em andamento, “dimensões tecnopolíticas do esquecimento: usos e abusos da memória na era digital” cujo objetivo é compreender as práticas e formas de memória no ambiente digital da World Wide Web como ação política e cultural da sociedade contemporânea. Dessa forma, neste artigo comunicar-se-ão alguns resultados preliminares no âmbito da revisão de literatura pertinente ao que é o Big Data e sua correlação possível com os estudos da CI em perspectiva interdisciplinar. Ao pensar sobre a possível “saturação” da memória através dos canais informacionais e suas respectivas tecnologias, também serão apresentados alguns dados da pesquisa de campo realizada sobre o Big Data na internet, enquanto lugar de produção e circulação de discursos e significações, capaz de gerar informações constituidoras de uma grande memória digitalizada da sociedade global, em muitos aspectos, criada à revelia de homens e mulheres na atualidade. Nesta primeira etapa de levantamento de dados e bibliografia, alguns resultados apontam para uma clara necessidade de se discutir o Big Data para além de sua dimensão econômica, refletindo sobre políticas públicas de acesso à informação e proteção à privacidade.



## QUADRO 31 – Produção acadêmica no Brasil

(Continuação)

Ano	Programa	Universidade/ Instituição	Autor	Título
2014	PPGCI	UFMG	BARACHO, Renata Maria Abrantes	O caminhar da ciência da informação e o XV encontro nacional de pesquisa em ciência da informação
<b>Abstract</b>				
<p>Este artigo descreve algumas características da Ciência da Informação como área do conhecimento e apresenta o Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) como o evento seminal da área. Busca-se criar a memória do ENANCIB, promovido pela Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), desde a sua criação em 1994, ao longo de suas quinze edições. O ENANCIB congrega pesquisadores e programas de pós-graduação nacionais, além de propiciar um fórum para discussão dos temas de pesquisa da área. Em 2014, o XV ENANCIB tem lugar em Belo Horizonte Minas Gerais e é organizado pelo Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais de Minas Gerais (UFMG). No âmbito do PPGCI, nasceu a percepção de que o evento deveria propor a discussão sobre como as mais recentes tecnologias da informação e da comunicação têm alterado a sociedade e a forma de interagir das pessoas, bem como o papel relevante da Ciência da Informação nesse contexto. Os dispositivos móveis, nuvens, big data, linked data, dentre outras formas de interagir com a informação têm causado transformações em vários âmbitos e exigido novas abordagens para os estudos em CI. O tema do XV ENANCIB 'Além das `nuvens`: expandindo as fronteiras da Ciência da Informação?', proporciona a oportunidade para refletir sobre essas mudanças que impactam na interação humana com a informação, bem como sobre suas implicações para o futuro da CI.</p>				
Ano	Programa	Universidade/ Instituição	Autor	Título
2013	PPGCINF	Universidade Federal de Brasília (UNB)	Gasque, Kelley	Competência em Informação: conceitos, características e desafios
<b>Abstract</b>				
<p>A Profa. Dra. Kelley Cristine Gonçalves Dias Gasque, especialista em competência informacional, aborda e esclarece distintos conceitos desta temática ainda em construção, apresentando um breve histórico e os principais modelos aceitos para o desenvolvimento da competência informacional em ambientes os mais diversos; e suas consequências para os indivíduos envolvidos. A Dra. Kelley Cristine complementa suas considerações explorando os efeitos do big data e do linked data nos processos de aprendizagem contínua e fornece interessantes insights para a temática nos próximos 5-10 anos.</p> <p>Palavras-chave: Letramento Informacional. Alfabetização Informacional. Competência Informacional. Habilidade Informacional. Aprendizagem Cognitiva.</p>				
Ano	Programa	Universidade/ Instituição	Autor	Título
2014	PPGCI/IBICT	UFRJ	Barreto, Aldo de Albuquerque	A aventura de perceber significados

## QUADRO 31 – Produção acadêmica no Brasil

(Continuação)

<b>Abstract</b>				
<p>Uma explosão de conteúdos aconteceu quando o volume de informação impressa disponibilizada no pós-guerra de 1945 mudou o regime de informação existente para uma nova configuração de: recursos humanos, acervamento, processamento e recuperação dos documentos estocados para atender a um novo fluxo da oferta e demanda. O fluxo da informação entre os estoques e os receptores permeiam critérios da tecnologia que almeja possibilitar o maior e melhor acesso ao acervo disponível e o critério do repasse intencional intervém para intentar uma interiorização adequada. Hoje com a condição <i>online</i> os estoques e os fluxos de informação, renomeados, de acordo com o gosto do momento, para "Big Data" são multidirecionados e levam condições virtuais em seu desatamento, quando o tempo se aproxima de zero, a velocidade se acerca do infinito e os espaços são de vivência pela não presença. A crescente produção de informação precisa ser reunida e armazenada de forma eficiente, e eficaz na distribuição. A chegada de uma sociedade eletrônica de informação modificou a delimitação de tempo e espaço dos conteúdos em relação aos receptores.</p>				
<b>Ano</b>	<b>Programa</b>	<b>Universidade /Instituição</b>	<b>Autor</b>	<b>Título</b>
2015	PPGCINF	UNB	Costa, Maira Murrieta.	A literatura internacional sobre e-science nas bases de dados LISA e LISTA.
<b>Abstract</b>				
<p>Apresenta brevemente o conceito de e-Science. Trata-se de um estudo quantitativo que analisa as publicações sobre o tema e-Science na área da CI. O universo da pesquisa é formado por 202 registros de documentos das bases de dados LISA e LISTA. A análise dos dados revela um crescimento do volume de publicações sobre o tema e demonstra que dentre os 202 registros analisados, 172 são de artigos de periódicos. Ainda, confirma o fenômeno de que poucos autores publicam muito. A Inglaterra foi o país que mais se destacou na análise dos dados, visto que demonstrou pioneirismo em programas de governo para e-Science. Dentre as associações profissionais da CI, destaca-se a atuação da Association of Research Libraries. Ressalta-se que foram encontrados 3 artigos de autores brasileiros indexados sobre o tema.</p>				
<b>Classificação OP</b>				
<b>Ano</b>	<b>Programa</b>	<b>Universidade /Instituição</b>	<b>Autor</b>	<b>Título</b>
2014	Departamento de CI » Grupos de Pesquisa	UFSC	Dutra, Moisés Lima	Big data aplicado a sistemas ciber-físicos da logística: Proposta de Modelo Conceitual
<b>Abstract</b>				
<p>Os Sistemas ciber-físicos combinam aspectos cibernéticos da computação e comunicação com a dinâmica dos sistemas físicos operados no mundo real. Juntamente com a grande relevância do gerenciamento e análise das informações com vistas ao planejamento e controle dos sistemas logísticos, considera-se que utilização de técnicas e tecnologias de Big Data e Business Analytics possa impulsionar a tomada de decisão dentro destes sistemas. Neste sentido, o processo de tomada de decisão é estratégico para se lidar com enormes quantidades de dados, analisar a informação e determinar como se trabalhar com tal volume e variedade de informações na velocidade adequada. Este trabalho é uma pesquisa aplicada, que propõe um modelo conceitual para combinar práticas de gestão com a aplicação de técnicas inovadoras para aquisição e análise de dados nos Sistemas ciber-físicos. Ao final, conclui-se que as organizações não irão colher os benefícios de uma transição para o uso de Big Data, a não ser que sejam capazes de gerenciar efetivamente este processo de mudança. Palavras-chave: Big Data. Business Analytics. Sistemas ciber-físicos. Logística.</p>				

## QUADRO 31 – Produção acadêmica no Brasil

(Continuação)

Ano	Programa	Universidade/ Instituição	Autor	Título
	Recherche en Sciences de l'Information et de la Communication pela Aix Marseille Université, France. Pesquisador da Fundação Oswaldo Cruz (FIOCRUZ)	FIOCRUZ	Magalhães, Jorge L	Extração e tratamento de dados na base lattes para identificação de core competencies em dengue

**Abstract**

A conjuntura de competitividade global requer das organizações práticas cada vez mais ousadas de Inteligência Competitiva, a fim de obter, analisar, tratar e disseminar as informações para auxiliar na tomada de decisão. No ambiente de Big Data na Web, é premente cautela para o resgate e análise de dados a fim de transformá-los em informações essenciais para o gestor. Objetivo: Identificar e extrair a produção científica, produtos tecnológicos, instituições, redes dos cientistas que trabalham com a doença Dengue. Metodologia: Usa-se a bibliométrica com técnicas para mensurar a produção e disseminação do conhecimento científico. Analisar os 2,5 milhões de currículos da base Lattes do CNPq, extrair e tratar os que possuem o termo "dengue" com o software ScriptLattes do conceito Web 2.0. Resultados: A identificação de 15.465 currículos específicos com o tema Dengue. Extraiu-se 424 cientistas renomados na área, bem como mais 971 colaborações nacionais e internacionais relacionados com o termo dengue. O método possibilitou extrair dos especialistas a geolocalização, publicações, orientações, dentre outros. Conclusões: A análise dos resultados mostrou a relação multidisciplinar dos cientistas em Dengue. O método pode ser replicado para qualquer área da ciência. A bibliometria como fonte de auxílio no tratamento de Big Data mostrou-se eficaz através da ferramenta ScriptLattes.

Ano	Programa	Universidade /Instituição	Autor	Título
	PPGCI	UFPB	Santana, Júnior	Microgisborne: uma ferramenta para recuperação das informações contidas em blogs baseadas em microformatos

**Abstract**

O objetivo deste trabalho é apresentar a ferramenta Microgisbone que realiza a recuperação da informação de blogs na Internet utilizando como referência o padrão de microformato rel-tag. A ferramenta proposta tem como princípios (i) a adequação aos padrões de Big Data, (ii) utilizar uma arquitetura de serviços escalável para que seja possível utilizar computação em nuvem a partir de Infraestrutura como Serviços (IaaS) e (iii) que esteja disponível como um serviço de informação a comu+A10nidade. Para validar o funcionamento da ferramenta e escolher a arquitetura de dados foi realizado um experimento utilizando um cenário real, blogs hospedados pelo wordpress.com, onde as marcações (tags) foram coletadas por um período de 3 dias. Ao término deste experimento, onde foram coletadas cerca de 6,6 milhões de tags, foram desenvolvidos alguns serviços de informação, baseados nas marcações coletada. Foi observado que, de fato, um volume relativamente grande de informação foi recuperado de uma quantidade pequena de blogs e de um tipo de informação que é pequeno (marcações). Foi observado também padrão rel-tag dos microformatos tornam mais simples a identificação e recuperação das marcações nos blogs por máquinas se comparados com os mecanismos formais de web semântica

## QUADRO 31 – Produção acadêmica no Brasil

(Continuação)

Ano	Programa	Universidade /Instituição	Autor	Título
	Programa de pós-Graduação em Ciência da Informação (PGCIN)	UFSC	Dutra, Moisés Lima	Visualização de correspondências semânticas no universo Big Data
<b>Abstract</b>				
<p>O Big Data consiste de uma vasta quantidade de dados disponíveis em diferentes níveis de complexidade ? criados por humanos ou por máquinas, em diferentes ritmos ? e que apresentam grandes níveis de ambiguidade, de forma que não podem ser processados computacionalmente por meio da utilização de tecnologias, dispositivos de comunicação, métodos de processamento e algoritmos tradicionais, além de qualquer outra solução similar. Desta forma, pode-se dizer que a recuperação da informação nessas grandes aglomerações de dados necessita de novas tecnologias e técnicas de apoio, que deverão modificar os processos tradicionais de coleta e análise da informação. Este trabalho utiliza abordagem quali-quantitativa. Trata-se de uma pesquisa exploratória e aplicada, que envolve pesquisa bibliográfica e faz parte de dois projetos de pesquisa em andamento. O objetivo deste trabalho é apresentar uma proposta preliminar para o desenvolvimento de visualizador de correspondências semânticas no contexto Big Data, de maneira a facilitar a interação humano-computador. Por fim, concluiu-se que a possibilidade de visualizar correspondências semânticas dentro do universo Big Data, com interfaces adequadas quanto a sua usabilidade, pode melhorar sensivelmente o processo humano de tomada de decisão, pois possibilita descobrir e trazer à tona informações importantes, que de outra forma permaneceriam ocultas.</p>				
<b>Classificação RE</b>				
Ano	Programa	Universidade /Instituição	Autor	Título
2013	PPGCI	UFMG	Souza, R. R	Ciência da Informação em transformação: Big Data, Nuvens, Redes Sociais e Web Semântica
<b>Abstract</b>				
<p>Discute, sob uma perspectiva crítica, o panorama atual da CI e algumas das perspectivas futuras, em face dos fenômenos informacionais contemporâneos. Apresenta considerações sobre processo de esvaziamento temático através de análise de objetos de pesquisa atualmente desenvolvidos na CI e aponta o processo de absorção de outros destes objetos em diversas áreas do conhecimento. Questiona a primazia e a quantidade de pesquisas interdisciplinares na área e tece comentários sobre o que pode redundar desse processo para o futuro da CI. Em segundo momento, analisa o impacto para a área da explosão informacional, da consolidação das redes sociais como espaços de interação, da computação em nuvem, bem como de elementos constituintes da Web Semântica. Palavras-chave: Ciência da Informação; tecnologia; Ciência da Informação; representação do conhecimento; interdisciplinaridade; big data; Cloud Computing; social networks; Big data; Ciência da Informação; Cloud Computing; interdisciplinaridade; representação do conhecimento; social networks; Tecnologia; web semântica</p>				
Ano	Programa	Universidade /Instituição	Autor	Título
	PPGCI – Escola de Comunicação e Artes Ciências da Comunicação	USP	Dias, Guilherme	Big data: questões éticas e legais emergentes

## QUADRO 31 – Produção acadêmica no Brasil

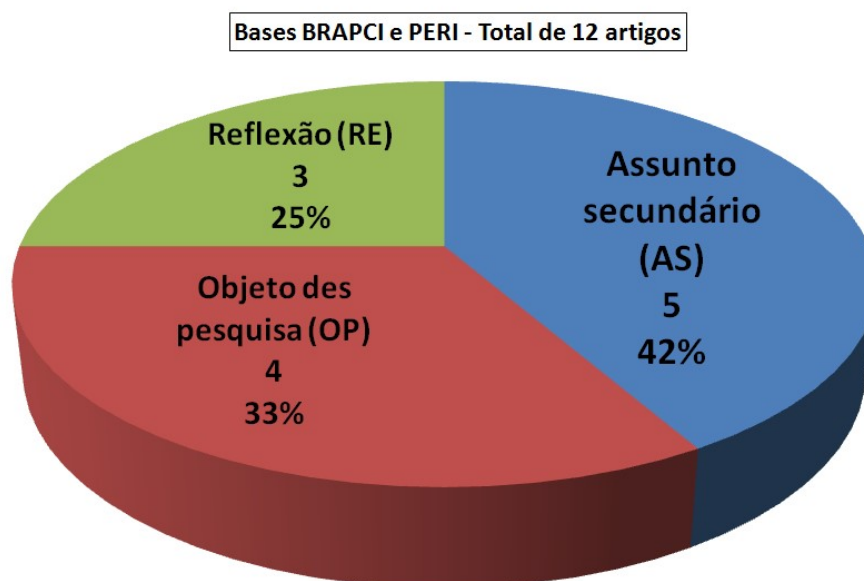
(Conclusão)

<b>Abstract</b>				
<p>Este artigo apresenta e discute questões de cunho ético e legais associadas ao fenômeno do Big Data, contextualizando esse conceito. Introduce questões éticas pertinentes ao Big Data conforme discutido pelo autor Kord Davis na obra intitulada Ethics of Big Data. A metodologia da pesquisa caracteriza-se por ser bibliográfica, exploratória e de cunho qualitativo. Com relação às questões de ordem ética relacionadas ao Big Data, entende-se que podem ser abordadas através do alinhamento das necessidades dos usuários em conjunto com as das organizações. No que tange à discussão jurídica, apresenta os principais desafios para os governos e sociedades e, em particular, coteja esses desafios com o ordenamento jurídico brasileiro. Ao final, conclui-se que o ordenamento jurídico brasileiro e suas instituições devem se adequar melhor aos vários desafios impostos pela era da informação.</p>				
<b>Ano</b>	<b>Programa</b>	<b>Universidade/Instituição</b>	<b>Autor</b>	<b>Título</b>
2014	PPGCI/IBICT	UFRJ	Ribeiro, C. J. S	Big Data: uma investigação com uso de dados abertos sobre acidentes de trabalho
<b>Abstract</b>				
<p>O tema Big Data tem despertado interesse nos profissionais que trabalham com a Gestão da Informação, pois trata de insumo essencial no processo de criação do conhecimento. Este relato apresenta o tema e explora os fundamentos que auxiliam no entendimento da abordagem de Big Data. Discute a explosão informacional e a avalanche de dados, chegando aos elementos que compõem o tema. Aborda os 4 Vs do Big Data e as fases de Discovery, Data Preparation, Model Planning e Analytics. O processo de pesquisa se baseia em um estudo exploratório desenvolvido por meio de pesquisa bibliográfica com análise documental. O marco empírico para análise dos elementos do domínio foi baseado no site de Dados Abertos da Dataprev, bem como com outras informações disponíveis em redes sociais e blogs especializados. Apresenta os resultados preliminares sobre estudos nas fases iniciais de projetos de Big Data, de forma a viabilizar alternativas para representação e disseminação dos grandes volumes de dados presentes na Internet. Ao final, apresenta alguns aspectos ligados ao perfil do profissional que está participando destes projetos.</p>				

Fonte: elaborado pelo autor, 2017.

No Quadro 31 demonstra-se que a produção científica brasileira publicada nas bases BRAPCI e PERI é composta por 12 artigos, publicados entre o anos de 2013 e 2015. Verifica-se no comportamento desta produção:

- a) artigos classificados como assunto secundário (AS): total de cinco artigos: Publicou-se dois artigos em 2013, dois artigos em 2014 e um artigo em 2015;
- b) artigos classificados como RE: total de três artigos: Publicou-se dois artigos em 2013 e um artigo em 2014.
- c) artigos classificados como OP: total de quatro artigos, todos publicados em 2014. Qualificaram-se dois como REC e dois como TRA.

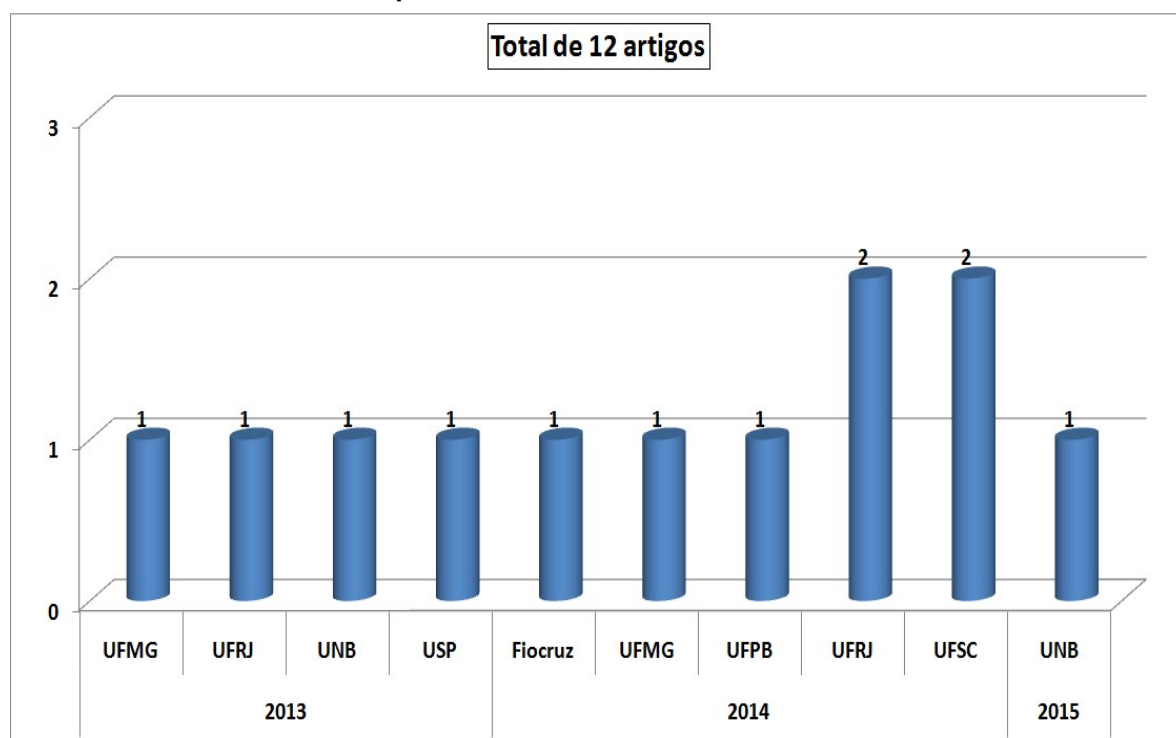
**GRÁFICO 6 – Mapa geral da classificação dos artigos: produção brasileira**

Fonte: elaborado pelo autor, 2017.

No GRÁFICO 6 percebe-se que a produção científica analisada é composta por 12 artigos, onde se verifica:

- classificou-se na categoria assunto secundário (AS), cinco artigos, o equivalente a 42% do universo da pesquisa, de 12 artigos. Essa categoria indica que o Big Data apenas é citado, não sendo esse o tema central do artigo;
- classificou-se três artigos, o equivalente a 25% do universo da pesquisa como reflexão - RE). Essa categoria indica que são realizadas reflexões sobre o Big Data, não sendo esse o tema central do artigo;
- classificou-se quatro artigos, o equivalente a 33% do universo da pesquisa como objeto de pesquisa (OP). Essa categoria indica que o Big Data é o tema central do artigo.

**GRÁFICO 7 – Número de universidades/instituições com produção acadêmica sobre *Big Data* por ano – bases BRAPCI e PERI**

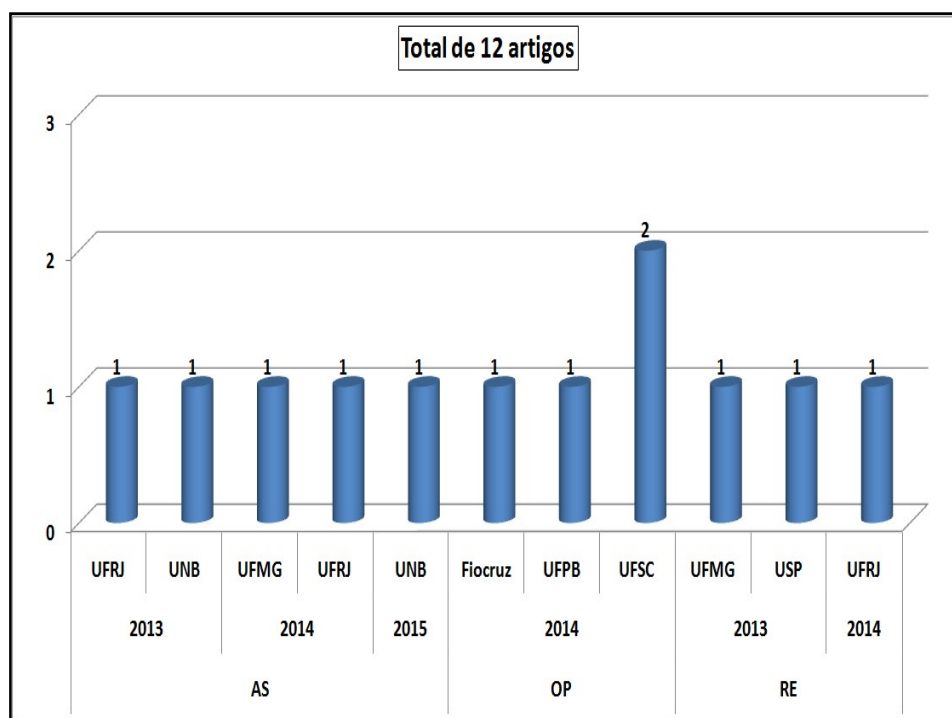


Fonte: elaborado pelo autor, 2017.

De acordo com os GRÁFICOS 6 e 7, conclui-se que sete universidades/instituições publicaram doze artigos envolvendo a temática *Big Data*.

O GRÁFICO 8 demonstra a classificação dos artigos publicados por cada universidade/instituição.

**GRÁFICO 8 – Mapa da produção acadêmica anual sobre *Big Data*–universidades / instituições – Brasil**



Fonte: elaborado pelo autor, 2017.

No GRÁFICO 8 verifica-se que cinco artigos receberam a classificação como (AS). A UFRJ e a UNB publicaram o total de dois artigos em 2013, em 2014 a UFMG e a UFRJ, publicaram o total de dois artigos, em 2015, a UNB publicou um artigo.

Quatro artigos receberam a classificação como (OP), todos publicados em 2014. A Fiocruz e a UFPB publicaram um artigo cada e a UFSC publicou dois artigos.

Três artigos receberam a classificação como reflexão (RE). Em 2013 a UFMG e a USP publicaram um artigo cada. E 2014 a UFRJ publicou um artigo.

A Tabela 5, apresenta quais programas publicaram artigos nas bases BRAPCI e PERI em 2013:



**TABELA 5 – Programas com publicações – bases BRAPCI e PERI - 2013**

<b>Programa</b>	<b>Universidade/ instituição</b>	<b>Total artigos do programa</b>	<b>%sobre produção 2013</b>	<b>%sobre produção Total</b>
PPGCINF – Ciência da Informação da Universidade de Brasília	UNB	1	25	8,33
PPGCI – Escola de Ciência da Informação da Universidade Federal de Minas Gerais	UFMG	1	25	8,33
PPGCI-ECA –Escola de Comunicação e Artes Ciências da Comunicação – Pós-Graduação em Ciência da Informação (USP)	USP	1	25	8,33
PPGCI-IBICT-UFRJ – Programa de Pós-Graduação em Ciência da Informação.	UFRJ	1	25	8,33
<b>Total</b>		<b>4</b>	<b>100</b>	<b>32,32</b>

Fonte: elaborado pelo autor, 2017.

Verifica-se na Tabela 5, que no ano de 2013 quatro programas publicaram um artigo cada, o equivalente a 36,5% da produção total do universo da pesquisa referente às bases BRAPCI e PERI, com o seguinte comportamento: publicou-se um artigo nos seguintes programas: PPGCINF/CI/UNB, PPGCI/ECI/UFMG, PPGCI/ECA/USP, PPGCI/IBICT/UFRJ.

A Tabela 6 apresenta quais programas publicaram artigos nas bases BRAPCI e PERI em 2014:

**TABELA 6 – Programas com publicações – bases BRAPCI e PERI – 2014**

<b>Programa/Departamento</b>	<b>Universidade/ instituição</b>	<b>Total artigos do programa</b>	<b>%sobre produção 2014</b>	<b>%sobre produção Total</b>
Recherche en Sciences de l'Information et de la Communication pela Aix Marseille Université, France. Pesquisador da FIOCRUZ	FIOCRUZ	1	14,28	8,33
PPGCI-IBICT-UFRJ Programa de Pós-Graduação em Ciência da Informação	UFRJ	2	28,56	16,66
PGCIN- Pós-graduação em Ciência da Informação.	UFSC	2	28,56	16,66
PPGCI- Programa de Pós-Graduação em Ciência da Informação.	UFPB	1	14,28	8,33
PPGCI- Escola de Ciência da Informação da Universidade Federal de Minas Gerais	UFMG	1	14,28	8,33
<b>Total</b>		<b>7</b>	<b>100</b>	<b>58,31</b>

Fonte: elaborado pelo autor, 2017.

Percebe-se na Tabela 6 que no ano de 2014 cinco programas publicaram o total de sete artigos, correspondendo a 100% da produção acadêmica relacionada ao *Big Data* no ano de 2014 e 58,33% da produção total do universo da pesquisa referente às bases BRAPCI e Peri. Publicou-se 2 artigos nos seguintes programas: Programa de Pós-Graduação em Ciência da Informação (PPGCI-IBICT-UFRJ) e Programa de Pós-Graduação em Ciência da Informação na UFSC (PGCIN).

Publicou-se 1 artigo nos seguintes programas: Recherche en Sciences de l'Information et de la Communication pela Aix Marseille Université, France, Programa de Pós-Graduação em Ciência da Informação na UFPB (PPGCI) PPGCI e Escola de Ciência da Informação da Universidade Federal de Minas Gerais (UFMG).

**TABELA 7 – Programas com publicações - bases BRAPCI e PERI - 2015**

Programa	Universidade/ Instituição	Total artigos do programa	% sobre produção 2015	% sobre produção Total
PPGCINF - Ciência da Informação da UNB.	UNB	1	100	8,33
<b>Total</b>		<b>1</b>	<b>100</b>	<b>8,33</b>

Fonte: elaborado pelo autor, 2017.

Percebe-se na Tabela 7, que no ano de 2015 um programa publicou um artigo, correspondendo a 100% da produção acadêmica relacionada ao *Big Data* no ano de 2015 e 8,33% da produção total do universo da pesquisa referente às bases BRAPCI e Peri. Publicou-se 1 artigo no Programa de Pós-Graduação em Ciência da Informação da UNB (PPGCINF).

O Quadro 32, demonstra os autores com publicações de artigos relacionados ao *Big data* nas bases BRAPCI e PERI, entre 2013 e 2015:

QUADRO 32 – Autores com publicações relacionados ao *Big Data* – BRAPCI e PERI – 2013 a 2015  
(Continua)

Autor	Universidade /Instituição	Programa	Artigos
Baracho, Renata Maria Abrantes	UFMG	PPGCI/ECI/UFMG	1
Barreto, Aldo de Albuquerque	UFRJ	PPGCI/IBICT/UFRJ	1
Costa, Maira Murrieta,	UNB	PPGCINF - Ciência da Informação da UNB.	1
Dias, Guilherme	USP	PPGCI/ECA/USP	1
Dutra, Moisés Lima	UFSC	PGCIN	2
Gasque, Kelley	UNB	PPGCINF - Ciência da Informação da UNB.	1

QUADRO 32 – Autores com publicações relacionados ao *Big Data* – BRAPCI e PERI – 2013 a 2015  
(Conclusão)

<b>Autor</b>	<b>Universidade /Instituição</b>	<b>Programa</b>	<b>Artigos</b>
Magalhães, Jorge L	FIOCRUZ	Recherche en Sciences de l'Information et de la Communication pela Aix Marseille Université, France. Pesquisador da FIOCRUZ	1
Pimenta, R. M	UFRJ	PPGCI/IBICT/UFRJ	1
Ribeiro, C. J. S	UFRJ	PPGCI/IBICT/UFRJ	1
Santana Júnior	UFPB	PPGCI	1
Souza, R. R.;	UFMG	PPGCI/ECI/UFMG	1
<b>Total Geral</b>			<b>12</b>

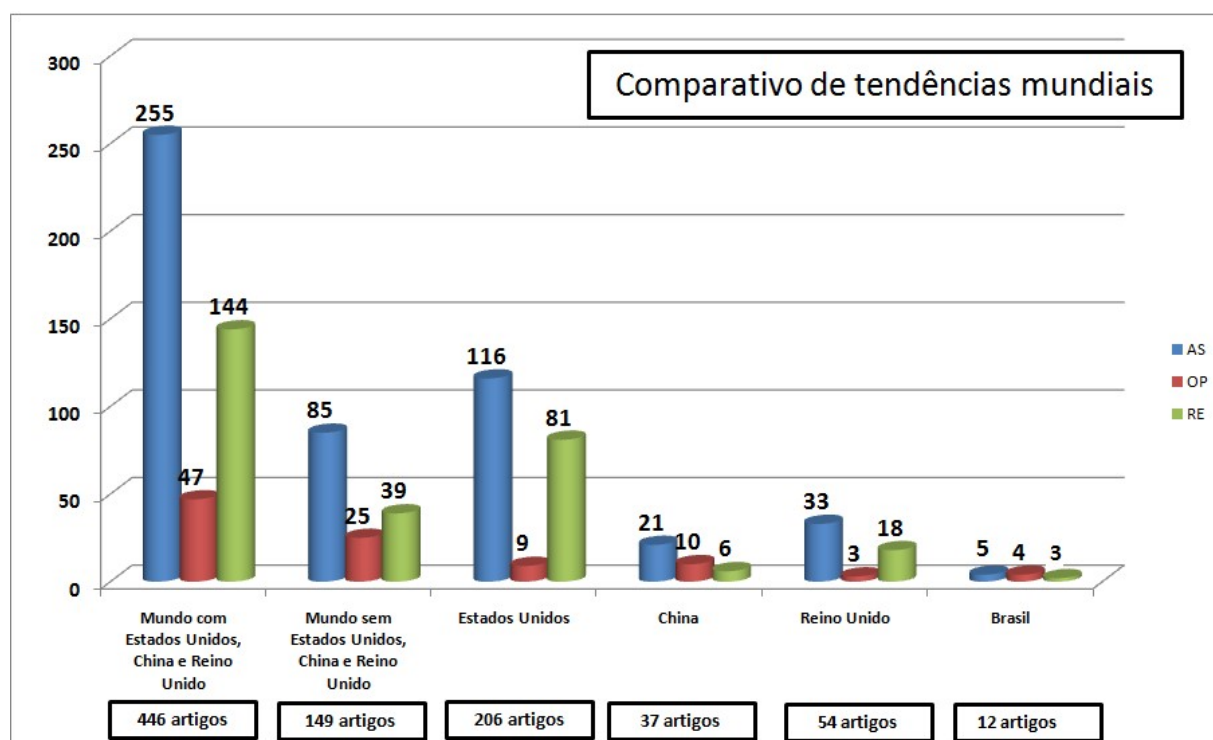
Fonte: elaborado pelo autor, 2017.

### 4.3 Comparativo de tendências mundiais

Foi perceptível o impacto da produção acadêmica dos Estados Unidos, Reino Unido e China ao longo desta pesquisa.

Com o objetivo de avaliar a influência destes países, realizou-se simulações de cenários eliminando-os do universo desta pesquisa, onde restariam 149 artigos e 32 países. Nota-se que a maioria dos artigos restantes, 85 ou 57,05% foi classificado como AS, seguindo a mesma tendência observada no cenário anterior com todos os países. Em seguida temos os arquivos classificados como RE, com 39 artigos ou 26,17% dos artigos desse cenário. Selando a conformidade com o primeiro cenário, os artigos classificados como OP, totalizam 25 ou 16,77% dos artigos deste cenário.

GRÁFICO 9 – Comparativo de tendências mundiais



Fonte: elaborado pelo autor, 2017.

No GRÁFICO 9 constata-se que os Estados Unidos de forma isolada, publicou 206 artigos, o equivalente a 46,19% dos artigos do universo da pesquisa. Na publicação desse país, 116 artigos ou 56,31% foram classificadas como AS. Em seguida, observa-se que os artigos classificados como RE totalizam 81 artigos ou 39,32% do total dos artigos produzidos por esse país. Os nove artigos restantes equivalem a 4,36% do total receberam a classificação OP. Portanto, proporcionalmente, os Estados Unidos publicam mais artigos classificados como RE em detrimento dos artigos classificados como OP. Mesmo sendo o maior produtor de artigos na categoria em números absolutos.

China produziu 37 artigos, o equivalente a 8,29% dos artigos do universo da pesquisa. Da publicação desse país, 21 artigos receberam a classificação como AS, o que equivale a 56,76% do total publicado por esse país, mantendo-se a tendência observada até o momento, no que tange à classificação como AS. Diferenciando-se da tendência observada até aqui, em segundo lugar, a China publica artigos classificados como OP. Foram 10 artigos, o que equivale a 27% do volume publicado por este país, classificando-o proporcionalmente como o país que mais publicou artigos classificados como OP, entre os países com grande volume de publicações. Em seguida, tem-se os artigos classificados como reflexão, totalizando seis artigos, equivalente a 16,21% da produção deste país. Tornando-o portanto, o país com alto volume de publicação que menos publicou artigos classificados como RE.

Reino Unido publicou 54 artigos, equivalente a 12,10% do universo da pesquisa. Dos artigos publicados por esse país, o total de 33 artigos foram classificados como AS, o que equivale a 61,11% de seu volume de publicações, tornando-o proporcionalmente entre os países que produzem grandes volumes de artigos, o que mais produz artigos classificados como AS. Portanto, mantem-se a tendência de publicar o maior número de artigos classificados como AS. Em seguida, observa-se que aparecem os artigos classificados como reflexão, em um total de 18, o que equivale a 33,33% do total do volume publicado por este país. Em seguida observa-se que o Reino Unido publicou quatro artigos que foram classificados como OP, que equivalem a 7,40% do volume produzido por este país.

A produção científica brasileira demonstrada na Figura 16, anteriormente, é a única de origem exclusiva das bases BRAPCI e PERI. Percebe-se que o universo da pesquisa é de doze artigos. Cinco desses artigos, o equivalente a 42% da produção acadêmica deste país recebeu a classificação como AS. Observa-se que quatro artigos, equivalente a 33% da produção acadêmica brasileira, receberam a classificação como OP. Os três artigos restantes, o equivalente a 25% da produção acadêmica brasileira, receberam a classificação como RE. Portanto percebe-se que o Brasil também apresenta a tendência em escrever mais artigos classificados na categoria aborda Big data como AS. Em seguida observam-se os artigos classificados como OP e na sequência os artigos classificados como RE.

## 5 CONSIDERAÇÕES FINAIS

### 5.1 Conclusão

O fenômeno *Big Data* diz respeito ao maior volume de informações já produzido pela humanidade. Nesse contexto, é relevante entender como a ciência e em especial a CI tem pesquisado e se posicionado diante desse fenômeno. Torna-se relevante entender as soluções, pesquisas e tendências em torno desse tema.

A CI gestou-se a partir de uma explosão informacional ocorrida após a Segunda Guerra Mundial. Naquele momento, nascia essa ciência interdisciplinar com a missão de encontrar soluções ao volume informacional então presente.

Vemo-nos imersos em uma nova explosão informacional, onde a informação nem sempre possui formato, autoria, suporte ou volume definidos. A tecnologia acessível propicia a inclusão de diversos novos geradores e consumidores de informação.

Nesta pesquisa, apresentou-se a origem da CI, sua interdisciplinaridade e seu desafio no atual cenário informacional, além de analisar-se o fato de outras ciências lançarem-se na busca por soluções relacionadas à informação.

Conclui-se que a CI busca esses conhecimentos e habilidades também em outras áreas de estudo ao mesmo tempo em que cede seus próprios conhecimentos às outras ciências, valendo-se de sua característica de meta-ciência e sua interdisciplinaridade, que apesar de ser apontada por autores supracitados, como canal de esvaziamento conceitual dessa ciência, é compreendido por esta pesquisa como o canal por aonde chegam os conhecimentos adicionais necessários às respostas aos desafios que se apresentam. Nesse ponto percebe-se que a CI se reinventa continuamente, em um ritmo distinto do que ocorre com o cientista da informação e seu processo de formação acadêmica, mais lento que o contínuo avanço tecnológico, o constante aumento do volume informacional e a mudança no perfil e no interesse dos usuários. Adiciona-se a esse contexto, a coexistência entre os antigos e os novos suportes informacionais, interesses culturais e comportamentais, corporativos, acadêmicos, governamentais e individuais.

Desta forma, conclui-se que a interdisciplinaridade da CI constitui seu passaporte para manter-se como ciência gestora da informação e do conhecimento, ao mesmo tempo que exige a busca de novos conhecimentos e habilidades necessários ao tratamento do dado, informação e do conhecimento, considerando os volumes informacionais atuais e futuros.

Constata-se que diversas áreas publicam artigos em espaços especializados da CI, o que além de ser demonstração de sua interdisciplinaridade, indica:

- a) a busca dos usuários (ciência, academia, empresas e a sociedade) pelos conceitos de organização, disseminação, tratamento e criação de serviços, próprios da CI. Esses conceitos são embutidos em aplicativos, ferramentas e *frameworks* destinados a resolver questões próprias a cada demanda específica. ciências como a computação e a medicina, destacam-se neste sentido;
- b) a presença de programas e pesquisadores de outras ciências pode indicar que os “nativos” desta ciência não estão instrumentalizados para o atual contexto informacional. Isto explica em parte a questão de a CI e conseqüentemente o cientista da informação não estarem se dedicando ao estudo, compreensão e análise do fenômeno *Big Data*, o que explica o reduzido volume de publicações produzidos por esta ciência presentes no universo desta pesquisa;
- c) a grande incidência de artigos classificados como AS pode ser atribuída ao fato de que pessoas de outras ciências estão publicando e os estudos referentes à informação é objeto próprio da CI.

Diante do espaço não ocupado pela CI, as demais ciências avançam no sentido de buscar suas soluções informacionais, dando origem a novos personagens na cena informacional, como o cientista de dados.

Neste contexto, a CI responde com a criação das *iSchools*, que pretendem através da interdisciplinaridade, instrumentalizar os nativos desta ciência, através de um espaço de formação que possa realizar uma aproximação com as demais ciências, somando pessoas, tecnologia e conhecimento em busca do que seria um novo perfil para esta ciência e conseqüentemente para o cientista da informação. Esta talvez seja a grande resposta que a CI tenha para a sociedade, a comunidade científica e sobretudo para si mesma, e assim como em suas origens, manter-se como ciência gestora da informação e do conhecimento.

Como resultado desta pesquisa, constata-se que os 446 artigos relacionados ao fenômeno *Big Data*, publicados nas bases LISA e LISTA entre 2010 e 2016, foram publicados por:

- a) 35 países;
- b) 307 universidades/instituições;
- c) 397 autores;
- d) 385 programas, que além da CI, têm origem e diversas outras ciências, como administração, antropologia, biologia, medicina, comunicação, computação, contabilidade, educação, marketing e sociologia, etc.

Quanto às bases brasileiras BRAPCI e PERI, identificaram-se:

- a) oito universidades/instituições;
- b) oito programas;
- c) nove autores;
- d) 10 artigos publicados.

Destaca-se que apenas a FIOCRUZ não é especializada em CI, demonstrando que no Brasil as publicações, majoritariamente, partem das escolas de Pós-Graduação em Ciência da Informação, todas públicas, sendo seis federais e uma estadual, a USP.

A despeito da origem dos programas presentes no universo da pesquisa, constatou-se que nas bases internacionais, 57% dos artigos relacionados ao *Big Data* publicados nas bases LISA e LISTA receberam a classificação como AS, 32% dos artigos recebeu a classificação RE e 11% recebeu a classificação OP, demonstrando que a CI, na maioria dos estudos relacionados ao *Big Data*, posiciona-o como AS.

No contexto brasileiro, percebe-se a tendência em dedicar-se igualmente a pesquisas onde o *Big Data* é OP 40% e a reflexões 40% e menos a pesquisas onde o *Big Data* é AS (30%).

Conclui-se que a CI estuda o fenômeno *Big Data*, entende a necessidade e o desafio de sua inserção frente a este fenômeno e o compartilhamento e busca de novos conhecimentos em outras ciências. Essa se reinventa com cautela, pois aos novos cientistas da informação serão exigidos novos conhecimentos, habilidades e atitudes, além do domínio dos conceitos basais dessa ciência.

## 5.2 Contribuições

Esta pesquisa contribui para a CI ao abordar uma temática contemporânea de grande interesse para esta ciência e ao demonstrar e analisar como a mesma tem se posicionado diante do fenômeno *Big Data*, e quais os atores e atitudes estão presentes nesse contexto. Identificou-se quais universidades, programas e países e autores publicam artigos relacionados ao *Big Data* e ainda analisou-se como abordam o tema *Big Data*.

Também contribui para que a CI possa entender o contexto científico em torno do *Big Data*, ao propor uma metodologia de identificação e classificação dos artigos e pesquisas publicados na área, o que permite a essa ciência entender qual é o seu papel, qual tem sido seu posicionamento frente ao fenômeno *Big Data* e quais as consequências.

Outra contribuição é tornar conhecidas quais outras ciências têm realizado pesquisas e buscam alternativas para solução de suas questões informacionais relacionadas ao *Big Data*, demonstrando quem realiza pesquisas voltadas à disseminação da informação, à infra-estrutura necessária para o processamento dos atuais grandes volumes de dados, ao tratamento da informação e à recuperação da informação.



O trabalho contribui para esclarecer o que é *e-science* (e-ciência), seu relacionamento com o fenômeno *Big Data*, e com esta ciência. Discutiu-se o surgimento da *data science* (ciência de dados) e do *data scientist* (cientista de dados) e demonstrou-se a possibilidades de atuação em conjunto das duas ciências frente ao fenômeno *Big Data*.

Elucidou-se o fluxo da informação que dá origem ao *Big Data*, além de auxiliar na compreensão de sua composição entre dados abertos e fechados, públicos e privados.

### 5.3 Limitações

Para a realização desta pesquisa, realizou-se ações de coleta de dados, durante as quais deparamo-nos com dificuldades relacionadas aos ambientes tecnológicos das bases BRAPCI e PERI e LISA e LISTA:

- a) o prazo reduzido diante do volume de dados identificados impediu o envolvimento de um número maior de pares durante a análise e classificação dos dados;
- b) fatores constantes que contribuíram para o consumo de tempo e redução da produtividade;
- c) dados incompletos, obrigando-se à busca dos dados complementares em sites de universidades, LinkedIn, Plataforma Lattes;
- d) muitos artigos inacessíveis através dos links oferecidos pelos portais das bases selecionadas;
- e) indisponibilidade técnica dos portais;
- f) a dificuldade de extração de dados a partir das bases pesquisadas, constituiu-se o maior entrave ao processo de pesquisa, obrigando a coleta de dados a partir do processo copiar e colar. Percebe-se que não há padronização na forma de dispor os dados, mecanismos de exportação e também ocorreram dificuldades com a incompletude dos dados. Encontrou-se diversos artigos sem *abstract*, autor, afiliação do autor, ano de publicação, etc.;
- g) algumas universidades, em geral asiáticas, no momento de coleta dos dados, não possuíam interface em inglês e/ou às vezes trabalhavam com um conjunto de caracteres que carregavam símbolos sem significado (lixo) para a base de dados, induzindo a erros de pesquisa e indexação do banco de dados;
- h) muitos autores não identificam sua afiliação no corpo dos artigos.

## 5.4 Pesquisas futuras

Lançam-se bases para pesquisas que considerem não apenas a quantidade de artigos produzidos pelos programas, universidades/instituições, autores e países. Trabalhos futuros podem inserir o peso das publicações, acrescentando a análise qualitativa a este contexto, como número de citações, índice h, fator de impacto, etc. Pode-se eventualmente propor um mecanismo ou espaço de avaliação de universidades, programas e autores com publicações voltadas à cena informacional ou exclusivamente ao fenômeno *Big Data*, a fim de detectar tendências e demandas.

Trabalhos futuros poderiam propor um selo de rastreabilidade ou verificabilidade dos metadados fornecidos pelo autor quando da publicação do artigo.

Recomenda-se também considerar a formação dos autores dos artigos que identificam pesquisas sobre o fenômeno *Big Data*.

Trabalhos futuros poderiam conceituar e comparar os perfis da CI no Brasil e em outros países, além de adicionar teses e dissertações ao universo da pesquisa.

## REFERÊNCIAS

- ABREU, G. O. L. **A Soberania dos dados versus a autonomia do usuário: big data, internet das coisas e as estratégias afirmativas do anonimato**. 2015. 121 f. Dissertação (Mestrado em Comunicação)– Centro de ciências, Letras e Artes, Universidade Federal da Paraíba, João Pessoa, 2015. Disponível em: <<http://tede.biblioteca.ufpb.br/bitstream/tede/7897/2/arquivototal.pdf>>. Acesso em: 23 jan. 2017.
- ALBINO, J. P.; REINHARD, N. KMAuditBr: uma ferramenta para diagnóstico e avaliação de sistemas de gestão do conhecimento. **Revista Gestão Industrial (Online)**, v. 5, p. 61-79, 2009. Disponível em: <<https://periodicos.utfpr.edu.br/revistagi/article/view/345>>. Acesso em: 29 jun. 2016.
- ARAÚJO, C. A. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, 2007.
- ARAÚJO, V. R. H.; FREIRE, I. M. Conhecimento para o desenvolvimento: reflexões para o profissional da informação. **Revista Informação & Sociedade: Estudos**, João Pessoa, v. 9, n. 1, 1999. Disponível em: <[www.isafreire.pro.br/ARAUJOeFREIREConhe.pdf](http://www.isafreire.pro.br/ARAUJOeFREIREConhe.pdf)>. Acesso em: 27 jun. 2016.
- AZAM, A. The first rule of data science. **Berkeley Science Review**. 27 Apr. 2014. Disponível em: <<http://berkeleysciencereview.com/article/first-rule-data-science/>>. Acesso em: 21 abr. 2016.
- BARBIERI, C. **BI2: business intelligence, modelagem e qualidade**. Rio de Janeiro: Elsevier, 2011.
- BATES, M. J. The invisible substrate of information science. **Journal of the American Society for Information Science**, v. 50, n. 12, 1999. Disponível em: <<http://www.gseis.ucla.edu/faculty/bates/substrate.html>>. Acesso em: 08 ago. 2016.
- BATES, M. The information professions: knowledge, memory, heritage. **Information Research: an International Electronic Journal**, v. 20, n. 1, 2015. Disponível em: <[http://www.informationr.net/ir/20-1/paper655.html#WY\\_Jn1GGPIU](http://www.informationr.net/ir/20-1/paper655.html#WY_Jn1GGPIU)>. Acesso em: 14 ago. 2017.
- BIG DATA BUSINESS. **Big Data pode prejudicar a privacidade das pessoas?** 10 ago. 2015. Disponível em: <<http://www.bigdatabusiness.com.br/o-big-data-pode-mesmo-prejudicar-a-privacidade-das-pessoas-2/>>. Acesso em: 22 jan. 2017.
- BLUESYEMRE. This is what happens in an #Internet minute 2017 (#infographic). 2017. Disponível em: <<https://bluesyemre.com/2017/03/21/this-is-what-happens-in-an-internet-minute-infographic/>>. Acesso em: 14 ago. 2017.
- BOHLE, S. What is e-science and how should it be managed?. **Scilogs: scientific and medical libraries**, 12 jun. 2013. Disponível em: <[http://www.scilogs.com/scientific\\_and\\_medical\\_libraries/what-is-e-science-and-how-should-it-be-managed/](http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed/)>. Acesso em: 26 mar. 2017.
- BORKO, H. Information Science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, Jan. 1968.

BRYANT, R. E.; KATZ, R. H.; LAZOWSKA, E. D. Big-data computing: creating revolutionary breakthroughs in commerce, science, and society. **Community Computing Consortium**, 22 Dec. 2008. Disponível em: <<http://www.datascienceassn.org/sites/default/files/Big%20Data%20Computing%202008%20Paper.pdf>>. Acesso em: 02 abr. 2016.

BURKE, P. (Org.). **A escrita da história: novas perspectivas**. São Paulo: Unesp, 1992.

BUSH, V. et al. As we may think. **The atlantic monthly**, Boston, v. 176, n. 1, p. 101-108, 1945.

CAVALCANTI, M. Big data não é tecnologia. **O Globo**, Rio de Janeiro, 23 mar. 2014. Disponível em:

<<http://acervo.oglobo.globo.com/busca/?tipoConteudo=pagina&pagina=&ordenacaoData=relevancia&allwords=BIG+DATA&anyword=&noword=&exactword=&decadaSelecionada=2010&anoSelecionado=&mesSelecionado=&diaSelecionado=>>>. Acesso em: 01 fev. 2017.

CHANG, H.-T.; MISHRA, N.; LIN, C.-C. lot big-data centred knowledge granule analytic and cluster framework for BI applications: a case base analysis. **PLoS One**, v. 10, n. 11, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4657997/>>. Acesso em: 21 jan. 2017.

COMPUTAÇÃO em nuvem para ciência. Disponível em:

<<https://cnc.rnp.br/institucional/sobre.php>>. Acesso em: 21 jan. 2017.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. Portal de periódicos CAPES/MEC. Disponível em: <<http://www.periodicos.capes.gov.br/>>. Acesso em: 30 jan. 2017.

COSTA, R. S.; ORRICO, E. G. D. Informação e memória quadranística: o caso watchmen. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais eletrônicos...** Salvador: PPGCI-UFBA, 2007. Não paginado. Disponível em: <<http://enancib.ibict.br/index.php/enancib/viiienancib/paper/viewFile/2854/1982>>. Acesso em: 13 jan. 2017.

DANISH E-INFRASTRUCTURE COOPERATION. eScience centers. Disponível em: <<https://vidensportal.deic.dk/eScience-centres?language=en>>. Acesso em: 22 jan. 2017.

DEMCHENKO, Y.; LAAT, C.; MEMBREY, P. Defining architecture components of the Big Data Ecosystem. In: INTERNATIONAL CONFERENCE ON COLLABORATION TECHNOLOGIES AND SYSTEMS, 2014, Minneapolis, MN. **Anais eletrônicos...** IEEE, 2014. p. 104-112.

DIAS, E. J. W. Ensino e pesquisa em Ciência da Informação. **DataGramZero**, v. 3, n. 5, p. A02-0, 2002. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/1267>>. Acesso em: 26 mar. 2017.

DUMBILL, E. et al. **Big data now current perspectives**. Beijing: O'reilly, 2012. Disponível em: <<http://www.oreilly.com/data/free/files/big-data-now-2012.pdf>>. Acesso em: 25 maio 2017.

EKBIA, H. et al. Big data, bigger dilemmas: a critical Review. **Journal of the Association for Information Science and Technology**. v. 66, n. 8, p. 1523-1545, Aug. 2015. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23294/abstract;jsessionid=657F99A44A42DD>>

EFBC1DA6A22CDF96B6.f01t02?userIsAuthenticated=false&deniedAccessCustomisedMessage=>. Acesso em: 03 jul. 2016.

E-SCIENCE INSTITUTE. Advancing data-intensive discovery in all fields. Disponível em: <<http://escience.washington.edu/>>. Acesso em: 21 jan. 2017.

FERREIRA, M. As 11 profissões que estarão em alta no Brasil em 2015. **Época Negócios**, 07 jan. 2015. Disponível em: <<http://epocanegocios.globo.com/Carreira/noticia/2014/12/10-profissoes-que-estarao-em-alta-no-brasil-em-2015.html>>. Acesso em: 30 jun. 2016.

FERRER-SAPENA, A.; SÁNCHEZ-PÉREZ, E. Open data, big data: ¿ hacia dónde nos dirigimos?. **Anuario ThinkEPI 2013**, v. 7, p. 150-156, 2013. Disponível em: <<http://eprints.rclis.org/21006/1/150-156-Ferrer-Sanchez-Open-data-Big-data.pdf>>. Acesso em: 27 maio. 2017.

GALVÃO, M. C. B.; BORGES, P. C. Ciência da Informação: ciência recursiva no contexto da sociedade da informação. **Ciência da Informação**, Brasília, DF, v. 29, n. 3, p. 40-49, set./dez. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n3/a05v29n3.pdf>>. Acesso em: 30 jun. 2016.

GARRITANO, J.; CARLSON, J. R. A subject librarian's guide to collaborating on e-science projects. 2009. **Purdue University**. Disponível em: <[http://docs.lib.purdue.edu/lib\\_research/140/](http://docs.lib.purdue.edu/lib_research/140/)>. Acesso em: 31 mar. 2016.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GUINARD, D.; TRIFA, V. M. **Building the web of thing**: with examples in Node. js and Raspberry Pi. [S.l.]: Manning Publications, 2016. 344 p. Disponível em: <<https://www.manning.com/books/building-the-web-of-things>>. Acesso em: 27 jun. 2016.

GURIN, J. Big data vs open data: mapping it out. **Open Data Now**, out. 2013. Disponível em: <[http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/#.V\\_1d-Y-cF9A](http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/#.V_1d-Y-cF9A)>. Acesso em: 26 mar. 2017.

HARBERT, T. Big data, big jobs? **Computerworld**, v. 47, n. 1, p. 23-27, 2013. Disponível em: <<http://www.computerworld.com/article/2492676/big-data/big-data--big-jobs-.html>>. Acesso em: 23 mar. 2017.

HEY, T. The next scientific revolution. **Harvard Business Review**, v. 88, n. 11, p. 56-63, 2010. Disponível em: <<http://i2ge.com/wp-content/uploads/2012/01/Next-Scientific-Revolution.pdf>>. Acesso em: 23 mar. 2017.

HEY, T.; HEY, J. E-Science and its implications for the library community. **Library Hi Tech**, v. 24, n. 4, p. 515-28, 2006. Disponível em: <<https://www.deepdyve.com/lp/emerald-publishing/e-science-and-its-implications-for-the-library-community-jcYvTiSio8>>. Acesso em: 05 jul. 2016.

INTERNATIONAL BUSINESS MACHINES. Developerworks. **Curso big data analytics na FGV Management Rio de Janeiro**. 2013. Disponível em: <[https://www.ibm.com/developerworks/community/blogs/bigdata/entry/curso\\_big\\_data\\_analytics\\_na\\_fgv\\_management\\_rio\\_de\\_janeiro?lang=en](https://www.ibm.com/developerworks/community/blogs/bigdata/entry/curso_big_data_analytics_na_fgv_management_rio_de_janeiro?lang=en)>. Acesso em: 30 jan. 2017.

JANKOWSKI, N. W. Exploring e-science: an introduction. journal of computer-mediated communication. **Journal of Computer-Mediated Communication**, v. 12, n. 2, p. 549-562, Jan. 2007. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00337.x/epdf>>. Acesso em: 03 jul. 2016.

JENKINS, H. TED x NYED. **Youtube**, 03 june 2010. Vídeo. (17:46). Disponível em: <<https://www.youtube.com/watch?v=AFCLKa0XRlw>>. Acesso em: 13 jan. 2017.

KUHN, T. S. **A estrutura das revoluções científicas**. 9. ed. Tradução Beatriz Vianna Boeira e Nelson Boeira. São Paulo: Perspectiva, 2006. 115 p.

LIMA, D. Big data é nova opção de carreira na ciência da computação; veja perfil. **Folha de São Paulo**. São Paulo, 27 set. 2015. Caderno Guia das Profissões. Disponível em: <<http://www1.folha.uol.com.br/educacao/2015/09/1686544-big-data-e-nova-opcao-de-carreira-na-ciencia-da-computacao-veja-perfil.shtml>>. Acesso em: 14 mar. 2016.

LOBO, A. P. Governo de Minas contrata big data da Microsoft. **Convergência Digital**, 27 ago. 2013. Disponível em: <<http://convergenciadigital.uol.com.br/cgi/cgilua.exe/sys/start.htm?infoid=34681&sid=16#.VD SRVMstCUk>>. Acesso em: 07 out. 2014.

MACEDO, J. Big data: transformando a indústria de telefonia móvel. **Revista Canaltech**, 03 mar. 2013. Disponível em: <<https://corporate.canaltech.com.br/noticia/big-data/Big-Data-transformando-a-industria-de-telefonia-movel/>>. Acesso em: 21 jan. 2017.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 4. ed. São Paulo: Atlas, 2001. 288 p.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2007.

MATTAR, F. N. **Pesquisa de marketing**. São Paulo: Atlas, 1996.

MCAFEE, A.; BRYNJALSSON, E. Big Data: The management Revolution. **Harvard Business Review**, Oct. 2012. Disponível em: <<http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf>>. Acesso em: 27 maio 2017.

MCLUHAN, M. **Os meios de comunicação como extensões do homem**. São Paulo: Cultris, 1969.

MYSORE, D.; KHUPAT, S.; JAIN, S. Arquitetura e padrões de big data, Parte 1: introdução à classificação e à arquitetura de big data. **Revista Developwork**, 25 abr. 2014. Disponível em: <<http://www.ibm.com/developerworks/br/library/bd-archpatterns1/>>. Acesso em: 14 jan. 2017.

NATIONAL E-SCIENCE CENTRE. Disponível em: <<http://www.nesc.ac.uk/>>. Acesso em: 12 jan. 2017.

OARD, D. Greetings from Maryland's iSchool. **Youtube**, 10 Jul. 2016. (5:37). Vídeo. Disponível em: <<https://www.youtube.com/watch?v=kOIMUHs2J8Y&feature=youtu.be>>. Acesso em: 24 jan. 2017.

OHIRA, M. L. B.; PRADO, N. S. Profissional da informação no limiar do século XXI: enfoque nos periódicos brasileiros em biblioteconomia e Ciência da Informação (1995/ 2002). **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 17, 1. sem. 2004. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/viewFile/1518-2924.2004v9n17p34/5269>>. Acesso em: 08 mar. 2017.

OPEN Definition. The Open Definioition. 2017. Disponível em: <<http://opendefinition.org/>>. Acesso em: 14 ago. 2017.

PASSARELLI, B.; JUNQUEIRA, A. H.; ANGELUCI, A. C. B. Os nativos digitais no Brasil e seus comportamentos diante das telas. **Matrizes**, v. 8, n. 1, p. 159-178, 2014. Disponível em: <<http://www.matrizes.usp.br/index.php/matrizes/article/viewFile/404/pdf>>. Acesso em: 05 jul. 2016.

PATIL, D. J., DAVENPORT, T. Cientista de dados o profissional mais cobiçado do século 21. **Harvard Business Review**, jul. 2013. Disponível em: <<http://hbrbr.com.br/cientista-de-dados-o-profissional-mais-cobicado-do-seculo-21/>>. Acesso em: 30 jan. 2016.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Universidade FEEVALE, 2013.

PROVOST, F.; FAWCETT, T. Data Science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51-59, Feb. 2013. Disponível em: <<http://online.liebertpub.com/doi/abs/10.1089/big.2013.1508>>. Acesso em: 27 jun. 2016.

RIBAS, C. C. C.; FONSECA, R. C. V. **Manual de metodologia OPET**. Curitiba. 2008. Disponível em: <[http://opet.com.br/biblioteca/PDF's/MANUAL\\_DE\\_MET\\_Jun\\_2011.pdf](http://opet.com.br/biblioteca/PDF's/MANUAL_DE_MET_Jun_2011.pdf)>. Acesso em: 27 mar. 2017.

RIBEIRO, C. J. S. Big data: os novos desafios para o profissional da informação. **Revista Informação & Tecnologia (ITEC)**, v. 1, n. 1, p. 96-105, jan./jun. 2014. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/itec/article/view/19380>>. Acesso em: 26 abr. 2016.

RUSSOM, P. et al. Big data analytics. **TDWI best practices report, fourth quarter**, v. 19, p. 40, 2011. Disponível em: <<https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf>>. Acesso em: 14 ago. 2017.

SANTOS, M. V. M.; MANNES, C. J. O estereótipo e o cientista da informação: quebras ou fortalecimentos. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 7, n. 1, p. 128-137, jul./dez. 2009. Disponível em: <<http://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1979>>. Acesso: 23 jan. 2017.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 26 jun. 2016.

SARACEVIC, T. Information science. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1051-1063, Oct. 1999.

SILVA, A. M.; RIBEIRO, F. Formação, perfil e competências do profissional da informação. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8., 2004, Estoril. **Anais eletrônicos...** Lisboa: BAD, 2004. Não paginado. Disponível em: <<http://ler.letras.up.pt/uploads/ficheiros/artigo4161.PDF>>. Acesso em: 08 mar. 2017.

SOARES, E. **Cientista de dados é profissão do futuro**. Computerworld, 22 mar. 2013. Disponível em: <<http://computerworld.com.br/carreira/2013/03/14/cientista-de-dados-e-profissao-do-futuro>>. Acesso em: 16 mar. 2016.

SOUZA, F. C. A Formação acadêmica de bibliotecários e cientistas da informação e sua visibilidade, identidade e reconhecimento social no Brasil. **Informação & Sociedade:**

Estudos, João Pessoa, v. 16, n. 1, p. 23-34, jan./jun. 2006. Disponível em: <<http://www.biblionline.ufpb.br/ojs/index.php/ies/article/viewFile/439/1491>>. Acesso em: 19 mar. 2016.

SOUZA, R. R.; ALMEIDA, M. B.; BARACHO, R. M. A. Ciência da Informação em transformação: big data, nuvens, redes sociais e web semântica. **Ciência da Informação**, Brasília, DF, v. 42 n. 2, p. 159-173, maio/ago. 2013. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1379>>. Acesso em: 14 jan. 2017.

STRYKER, C. Big data will impact every part of your life. **Youtube**, 22 set. 2014. (9:10). Vídeo. Disponível em: <<https://www.youtube.com/watch?v=0Q3sRSUYmys>>. Acesso em: 02 fev. 2016.

TAURION, C. **Cientista de dados**: o profissional de big data. 21 out. 2013. Disponível em: <[https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/cientista\\_de\\_dados\\_o\\_profissional\\_de\\_big\\_data?lang=en](https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/cientista_de_dados_o_profissional_de_big_data?lang=en)>. Acesso em: 30 jan. 2016.

TAURION, C. O caos conceitual e os 5V's do big data. **Revista CIO**, 11 maio 2012a. Disponível em: <<http://cio.uol.com.br/opiniaio/2012/05/11/o-caos-conceitual-e-os-5vs-do-big-data/>>. Acesso em: 13 jan. 2016.

TAURION, C. **Você realmente sabe o que é big data?** 30 abr. 2012b. Disponível em: <[https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce\\_realmente\\_sabe\\_o\\_que\\_e\\_big\\_data?lang=en](https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce_realmente_sabe_o_que_e_big_data?lang=en)>. Acesso em: 05 jul. 2016.

TURKLE, S. Os efeitos do mundo digital no nosso comportamento. **Globosat Play**. (22:07). Vídeo. Exibido em: 20 fev. 2016. Disponível em: <<http://globosatplay.globo.com/globonews/v/4850106/>>. Acesso em: 21 jan. 2017.

VANTI, N. A. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, Brasília, DF, v. 31, n. 2, p. 152-162, maio/ago. 2002.

WORLD WIDE WEB. Manual de dados abertos: governo. Traduzido e adaptado de opendatamanual.org. [S.l.]: Laboratório brasileiro de Cultura Digital, [2011?]. 54 p.



## **APÊNDICE A – Levantamento de dados da pesquisa**

NOTA - O Apêndice A encontra-se disponibilizado em CD anexado a esta dissertação.