



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de Marília



CULTURA  
ACADÊMICA  
*Editora*

# O contexto da produção científica de Big Data:

análise cientométrica

Ely Franci na Tannuri de Oliveira

Leilah Santiago Bufrem

Rafael Gutierrez Castanha

Rene Faustino Gabriel Junior

**Como citar:** OLIVEIRA, E. F. T. *et al.* O contexto da produção científica de Big Data: análise cientométrica. *In:* MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (org.). **Informação, Conhecimento, Ação Autônoma e Big Data: Continuidade ou Revolução?** Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2019. p. 47-66.  
DOI: <https://doi.org/10.36311/2019.978-85-7249-055-9.p47-66>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

# CONTEXTO DA PRODUÇÃO CIENTÍFICA DE *BIG DATA*: ANÁLISE CIENTOMÉTRICA

*Ely F. Tannuri de Oliveira*  
*Rafael Castanha*  
*Rene F. Gabriel Junior*  
*Leilah Santiago Bufrem*

A diversidade, a geração e o processamento de grande quantidade de dados disponíveis digitalmente são características associadas ao fenômeno *Big Data*, que despontou, mais significativamente, a partir do início do século XXI, com o crescimento exponencial da quantidade de dispositivos e de usuários conectados. *Big Data* consiste, primariamente, de conjuntos de dados, caracterizados pelo volume, variedade e velocidade (3Vs), cuja organização e gerenciamento demandam arquiteturas escaláveis para que os processos de armazenamento, manipulação e análise sejam eficientes.

Com o movimento pela ciência aberta e as formas de democratizar os saberes produzidos, surgem os modelos de produção de conhecimento interativos, abertos, colaborativos e livres para que pesquisadores possam deles dispor, reutilizar

<https://doi.org/10.36311/2019.978-85-7249-055-9.p47-66>

e compartilhar. Renovam-se possibilidades graças à tecnologia e ao desenvolvimento de políticas que favorecem essa abertura para a construção científica, que tem sido chamada de “Revolução na sociedade”.

No contexto da produção sobre o tema, há uma surpreendente motivação para a busca por caminhos alternativos, não só na produção de modelos e estruturas propícias à interação entre pesquisadores e conhecimento, mas também nas formas de comunicação desses dados. Ilustra bem essa motivação o artigo de Roa-Martínez, Vidotti e Santana (2017), que propõe uma estrutura para o assim chamado artigo de dados, como uma modalidade de publicação científica. Se analisado o conceito de dado – atribuível a uma representação, uma informação, um evento ou uma coisa – decorre uma ampla possibilidade de definições. Daí a necessidade de sua especificação, ou seja, para o caso deste estudo, o foco é nos dados utilizados no processo de investigação científica, o que facilitará a discussão sobre o seu compartilhamento e publicação. A partir das definições correntes de dados, os autores procuram ir além, para especificar como caracterizar os dados de investigação ou científicos, aqueles de interesse para a comunidade de pesquisadores. Assim, indicam, para facilitar os estudos sobre o tema, a definição construída por Torres-Salinas e outros (2012):

[...] todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza [...] que debe provenir de una fuente única», sin tener en cuenta otros elementos como las notas de laboratorio, los análisis previos, bocetos, informes o conversaciones informales con otros colegas (ROA-MARTÍNEZ; VIDOTTI; SANTANA, 2017, p. 2).

Com sólido material de pesquisa, os autores propõem uma estrutura, com os elementos necessários para permitir o compartilhamento e publicação de dados para que possam ser reutilizados. Essa estrutura de referência pode facilitar o processo de revisão por pares especialistas e as validações técnicas necessárias a partir de um conjunto de elementos ou parâmetros definidos, o que, por sua vez, segundo os autores, posiciona esse tipo de comunicação da produção científica como uma publicação de caráter científico. Defendem a possibilidade de que tenham o mesmo rigor científico e reconhecimento do que têm artigos tradicionais ou convencionais (ROA-MARTÍNEZ; VIDOTTI; SANTANA, 2017). Essa estrutura compõe-se de metadados do conjunto de dados de pesquisa publicados, tendo implícito que os dados não serão apenas compartilhados, mas que deverão ser publicados obedecendo a normas e padrões pertinentes ao processo de publicações que os dados requerem.

Ao identificarem os principais nichos e vertentes de publicação sobre o *Big Data analytics*, Furlan e Laurindo (2017) admitem que a era do *Big Data* já é realidade para empresas e indivíduos, e a literatura acadêmica sobre o tema tem crescido rapidamente nos últimos anos. Realizando uma pesquisa bibliométrica na base de dados *ISI Web of Science*, utilizam-se do termo *Big Data* para focar as práticas de gestão e conseguem identificar cinco grupos distintos dentre os artigos encontrados: evolução; gestão, negócios e estratégia; comportamento humano e aspectos socioculturais; mineração dos dados (*data mining*) e geração de conhecimento; e Internet das Coisas. Concluíram que, embora o tema seja emergente e tenha crescido nos últimos anos, é pouco consolidado, apresentando grande variação nos termos empregados, o que influencia nas buscas bibliográficas. Como resultado complementar da pesquisa, foram identifica-

das as principais palavras-chave empregadas nas publicações sobre *Big Data analytics*, o que contribui para as pesquisas bibliográficas futuras.

A atual configuração da dinâmica relativa à produção e à comunicação científicas permite que se revele o protagonismo da Ciência Orientada a Dados, em concepção abrangente, representada principalmente por termos como *e-Science* e *Data Science*. Com o objetivo de apresentar a produção científica mundial relativa à Ciência Orientada a Dados a partir dos termos *e-Science* e *Data Science* nas bases de dados *Scopus* e *Web of Science*, entre 2006 e 2016, Bufrem *et al.* (2016) estruturaram uma pesquisa na qual buscam informações nas referidas bases para obtenção e representação dos registros bibliométricos, tendo como resultado o destaque, entre a produção científica analisada, para os termos *Distributed computer systems* (2006), *Grid computing* (2007 a 2013) e *Big Data* (2014 a 2016). Na área de Biblioteconomia e Ciência de Informação, a ênfase foi dada aos temas *Digital library* e *Open access*, evidenciando a centralidade do campo nas discussões sobre dispositivos para dar acesso à informação científica em meio digital. Concluem que, sob um olhar diacrônico, constata-se uma visível mudança de foco das temáticas voltadas às operações de compartilhamento de dados para a perspectiva analítica de busca de padrões em grandes volumes de dados.

Nesse contexto de transformações efetivas, Coneglian, Gonzalez e Santarem (2017) voltam-se aos profissionais da informação, desafiando-os a trabalhar com a rapidez imposta tanto pela geração de informação e de dados, quanto pela variedade de fontes em que se encontram. Assim, sua atuação apresenta-se como fundamental ao buscar a sinergia entre

as necessidades de gestores e a expertise dos profissionais da computação.

Ao procurar identificar e discutir a intersecção entre as funções executadas pelos administradores (gestores), cientistas da computação e os profissionais da informação em ambientes de *Big Data Analytics*, os autores partem dos estudos da Ciência da Informação, do Profissional da Informação e do *Big Data*, para estabelecer um elo entre tais conceitos. Descrevem a atuação do profissional da informação em todas as fases da análise dos dados, verificando uma intersecção entre a proposta de relacionamento entre os atores “Administrador”, “Profissional da Informação” e “Profissional da Computação” com o chamado cientista de dados, concluindo que, devido a diferentes competências exigidas do cientista, tanto o profissional da informação quanto o profissional da computação podem se especializar nesta área, atuando em campos diferentes, mas complementares, fornecendo informações em cenários de *Big Data*, auxiliando a tomada de decisões e agregando valores para as organizações.

Pode-se questionar se com esse processo transformador ocorreria, de fato, uma cultura participativa, superando a questão da obsolescência das mídias, além da solução no armazenamento em nuvem (*in cloud*). A literatura parece indicar que o *Big Data*, embora enseje a recuperação e uso de um imenso volume de dados, hoje configura novos desafios, sendo o mais premente a longevidade digital, como sugeriu Besser (2010), ao defender a necessidade de “começar a pensar em um modo de assegurar que a informação digital continue acessível durante longo período”. Ele descreve o problema do desaparecimento da informação digital e analisa os cinco fatores causadores de problemas na sua longevidade: o da visua-

lização, o do embaralhamento, o da inter-relação, o da custódia e o da tradução. Propõe ações para reduzir o problema de preservar os documentos digitais para o futuro e alerta sobre a importância da cooperação internacional para a preservação da informação digital.

Katz e Martin (1997) apontam a coautoria como indicador da atividade de colaboração científica e apresentam algumas de suas vantagens: constitui-se de dados objetivos, podendo ser ratificada por estudos de outros pesquisadores; representa uma metodologia acessível e amigável para quantificar a colaboração. As pesquisas indicam que a coautoria incrementa a produção científica e a visibilidade do país, e tem promovido várias iniciativas governamentais dirigidas ao comportamento colaborativo, no âmbito de pesquisadores, instituições ou países.

Muitos estudos voltam-se para as relações existentes entre produtividade científica e coautoria científica. Entre eles, destaque-se os estudos de Eaton *et al.* (1999), que apontam uma forte relação entre o número de coautores e a produtividade científica, sinalizando que grupos que produzem com maior número de coautores tendem a publicar mais artigos.

Neste contexto, a análise de coautoria reflete um rol possível de intercâmbios e trocas entre os pesquisadores e constitui um procedimento significativo, sendo medida pelo número de publicações em colaboração entre autores, instituições ou países e empregada para identificar e mapear a cooperação regional, nacional ou internacional. A coautoria identifica a circulação do conhecimento e das novas ideias, dentro, especialmente, da comunidade mundial. A coautoria em âmbito internacional, no tocante à internacionalização, adquire relevância especial.

A questão desta pesquisa assim se explicita: quais as áreas e subáreas e quais as questões sobre o *Big Data*, enquanto objeto científico, estão sendo mencionadas em artigos científicos, no período de janeiro de 2016 a dezembro de 2017 registrados nos periódicos na base *Scopus*, levando em consideração as diferentes áreas do conhecimento? Pergunta-se, ainda, quais veículos de comunicação científica estão sendo mais utilizados?

A pesquisa tem como objetivo geral identificar a ênfase dada a artigos sobre *Big Data*. De forma mais especial, objetiva explicitar em quais áreas do conhecimento se destacam, quais as temáticas mais candentes e analisar o perfil das publicações periódicas que mais produzem sobre o tema. Por fim, apresentar os países mais produtivos e as coautorias internacionais, por meio de procedimentos cientométricos.

Esta pesquisa justifica-se pela necessidade de analisar este fenômeno nas diversas áreas do conhecimento e como tem se comportado a dinâmica dos estudos científicos sobre a temática em uma perspectiva sincrônica, em outras áreas, além da Ciência da Computação. Ainda, verificar em qual intensidade este tema está presente em grandes áreas do conhecimento, tais como, Ciências da Vida, Ciências Exatas, Tecnológicas e Multidisciplinar e Humanidades (CAPES, 2014). Pretende-se que este estudo se preste a subsidiar políticas de pesquisa sobre o tema.

## 1. METODOLOGIA

Essa pesquisa, de caráter exploratório e analítico, está situada no segundo grupo alvo de Glänzel (2003). O autor destaca três grupos-alvo em Bibliometria, a saber, *Bibliometria*



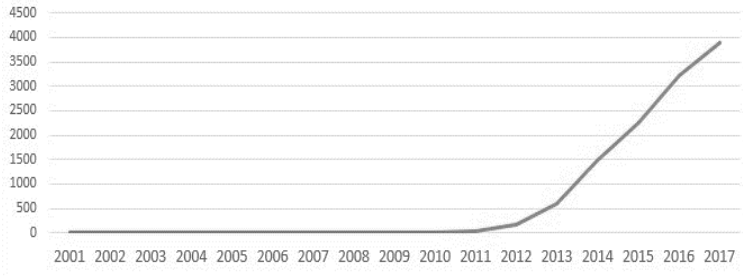
*para profissionais da bibliometria* ( $G_1$ ), que busca desenvolver e debater a bibliometria como metodologia, isto é, está preocupada com o seu próprio desenvolvimento conceitual-teórico-metodológico; *Bibliometria aplicada às disciplinas científicas* ( $G_2$ ): esse é grupo de pesquisa bibliométrica “aplicada” e se constitui o maior e mais diversificado grupo de interesse na bibliometria e ainda, *Bibliometria para a política científica e gestão* ( $G_3$ ): esse é o domínio da avaliação da pesquisa com fins de orientar políticas científicas.

Assim, esta pesquisa situa-se em âmbito da cientometria e faz parte do grupo da *Bibliometria aplicada às disciplinas científicas*, considerando que serão utilizados procedimentos bibliométricos para análise dos objetivos propostos.

Como procedimento de pesquisa, em 25 de julho de 2018, consultou-se a base de dados *Scopus*, com o termo de busca “*Big Data*”, presente na tipologia artigos, em seus títulos, resumos e palavras-chave, publicados de 2016 a 2017. Justifica-se a análise desses dois anos por corresponderem a 60,9% de toda produção sobre o tema. Como resultado, foram encontrados 7.103 artigos publicados, sendo que 3.213 deles eram do ano de 2016 e 3.890 do ano de 2017, verificando-se um acréscimo de 21,07% de um ano para outro. Desse total de artigos, somente 436 (6,1%) estão em periódicos *open access*.

A produção sobre *Big Data* saltou de menos de 30 artigos publicados em 2011 para 3.890 em 2017, tendo um alto crescimento.

**Gráfico 1** - Representação Sincrônica da produção sobre o tema.



**Fonte:** Construção dos autores

Para verificação das temáticas mais usadas, utilizaram-se os dados da própria base, buscando-se as dez palavras-chave mais presentes, adicionando-se às incidências de *Human* com as de *Humans*.

Apresentou-se a tabela dos periódicos que mais publicam no tema, com seus respectivos países, Quartil ao qual pertencem, Índice h e áreas do conhecimento dos periódicos. Ainda, apresentaram-se os países mais produtivos nas temáticas relacionando-os com o desenvolvimento em Ciência, Tecnologia e Informação.

Como procedimento final, mapearam-se, por meio do software *Vosviewer*, as coautorias internacionais dos países envolvidos nas publicações relacionadas ao tema *Big Data*. Para o mapeamento, utilizou-se a totalidade dos dados recuperados na pesquisa e traçou-se a rede de coautorias internacionais, entre os países coautores, procedendo-se à análise.

## 2. APRESENTAÇÃO DOS DADOS

Ao recuperar os 7.103 artigos segundo os critérios descritos anteriormente, analisaram-se as principais temáticas relacionadas a *Big Data*, bem como as principais fontes de publicação periódicas, países que mais publicaram sobre o tema e como estes países relacionam-se em termos de coautorias internacionais.

Para a análise das principais temáticas, recuperaram-se as palavras-chave com maior incidência no conjunto de artigos obtidos pelo termo de busca *Big Data*, assim como apresentado no Quadro 1. Percebe-se que, somadas, estas palavras-chave estão presentes em 5.123 artigos.

O Quadro 1 apresenta nove palavras-chave como as principais relacionadas aos artigos. Dessa maneira, destaca-se que a temática mais incidente nesta coleta de dados é representada pela palavra-chave *Human(s)*. Entretanto, como já descrito anteriormente, as palavras-chave *Human* (860) e *Humans* (610) foram somadas, visto que apresentam apenas a variação relativa à flexão de número.

**Quadro 1** - Principais temas relacionados (palavras-chaves)

Keywords*	Palavras-Chave	Nº de Artigos
<i>Human(s)</i>	Humano(s)	1470
<i>Data mining</i>	Mineração de dados	893
<i>Data Handling</i>	Manipulação de dados	622
<i>Article</i>	Artigo	607
<i>Digital Storage</i>	Armazenamento Digital	452
<i>Cloud Computing</i>	Processamento em nuvens	405
<i>Information Management</i>	Gerenciamento de Informação	346
<i>Distributed Computer Systems</i>	Sistemas de Computação distribuídos	328
	<b>Total</b>	5.123

Nesse sentido, pode-se dizer que as temáticas mais relevantes desta pesquisa são *Data Mining* (Mineração de dados) presente em 893 artigos, seguidas por *Data Handling* (Manipulação de dados) em 622 artigos. Tanto *Data Mining* quanto *Data Handling* podem ser relacionados a áreas do conhecimento com forte aporte computacional e tecnológico, como Ciência da Computação e Engenharias, corroborando com o que expressa a literatura relacionada a *Big Data*.

Pode-se ainda destacar, neste mesmo segmento, as palavras-chave *Digital Storage* (Armazenamento Digital), *Cloud Computing* (Processamento em nuvens), *Information Management* (Gerenciamento de Informação) e *Distributed Computer Systems* (Sistemas de Computação Distribuída) com 452, 405, 346 e 328 respectivas incidências nos artigos, palavras também que remetem a pesquisas relacionadas à computação e tecnologia.

Em uma análise combinada das palavras-chave com os artigos, apresentada na Figura 2, pode-se identificar duas grandes temáticas de estudo sobre *Big Data*. A primeira envolvendo a Ciência da Computação principalmente com os temas relacionados a metodologias, descrição de metadados, modelos reducionistas, redes neurais, manipulação de informação, computação na nuvem entre outros, formando um grupo que se relaciona aos estudos sobre o próprio *Big Data*.

A outra grande temática relaciona o *Big Data* a suas aplicações, principalmente nos estudos relativos aos seres humanos, envolvendo principalmente a genética, genoma, doenças, com destaque para o Alzheimer, gripe H1N1 e o diabetes, entre outras, caracterizando-se como a pesquisa que busca identificação de padrões.



ambos situados no Quartil Q4 e com Índice h, igual a 3 e 6, respectivamente. Representando as áreas de Engenharia Civil, Ciências de Matérias e Ciências Planetárias e da Terra, essas revistas podem ser caracterizadas como *outliers*.

**Quadro 2** - Principais fontes de publicação

Fontes	País	Nº de Artigos	SJR Rankings		Áreas
			Quartil	Índice h	
<i>Boletín Técnico</i>	Venezuela	156	Q4	3	Engenharia Civil; Ciências de Materiais
<i>Revista de La Facultad de Ingeniería</i>	Venezuela	120	Q4	6	Ciências Planetárias e da Terra
<i>IEEE Access</i>	EUA	102	Q1	36	Ciências da Computação; Engenharias
<i>Future Generation Computer Systems</i>	Holanda	65	Q1	85	Ciência da Computação
<i>Agro Food Industry Hi Tech</i>	Itália	60	Q3	16	Engenharia
<i>Neurocomputing</i>	Holanda	54	Q1/Q2	100	Ciência da Computação; Neurociências
<i>Lectures Notes In Computer Science</i>	Alemanha	53	Q2	296	Ciência da Computação; Matemática
<i>Cluster Computing</i>	Holanda	51	Q2	31	Ciência da Computação
<i>International Journal of Control Theory and Applications</i>	Índia	50	Q4	13	Ciência da Computação
<i>Jisuanji Yanjiu Yu Fazhan Computer Research and Development</i>	China	47	Q3	30	Ciência da Computação
<b>Total</b>	-	758	-	-	-

**Fonte:** Construção dos autores

O fato dessas publicações serem oriundas da Venezuela pode suscitar discussões com relação ao papel do país no âmbito do desenvolvimento da temática e possível incentivo a pesquisas relacionadas a Ciência, Tecnologia e Informação, visto que, historicamente, a Venezuela, fazendo parte da América Latina, não está situada entre as grandes produtoras de ciência no mundo como os EUA, a China e a Alemanha. Esta questão apontada carece de maior aprofundamento de estudos.

Em contraponto, os periódicos com Quartil de maior relevância (Q1) são oriundos dos EUA e Holanda: *IEEE Access* (EUA), *Future Generation Computer Systems* (Holanda) e *Neurocomputing* (Holanda). Tais periódicos foram responsáveis pela produção de 102, 65 e 54 artigos respectivamente. Pode-se também apontar a notoriedade destas publicações ao observar seu Índice h, visto que apresentam respectivamente os valores de 36, 85 e 100 como índice. Com relação às áreas do conhecimento às quais estão associadas estas publicações, nota-se a Ciência da Computação, as Engenharias e as Neurociências.

Ademais, deve-se destacar que o periódico com o maior Índice-H é o alemão *Lectures Notes In Computer Science*, com 296, e é relacionado à Matemática e à Ciência da Computação com Quartil igual a Q2. Dessa maneira, pode-se observar, que mesmo que os periódicos mais produtivos sejam oriundos da Venezuela, os de maior relevância segundo o Quartil e Índice h são provenientes dos EUA, Holanda e Alemanha, países com destaque internacional na produção científica mundial. Além disso, nota-se que todos os periódicos são relacionados a áreas do conhecimento das Ciências Exatas. Nesse contexto, analisaram-se ainda os países mais produtivos com relação ao tema *Big Data*. Os países são descritos na Quadro 3, podendo-se destacar China e EUA como os mais produtivos.

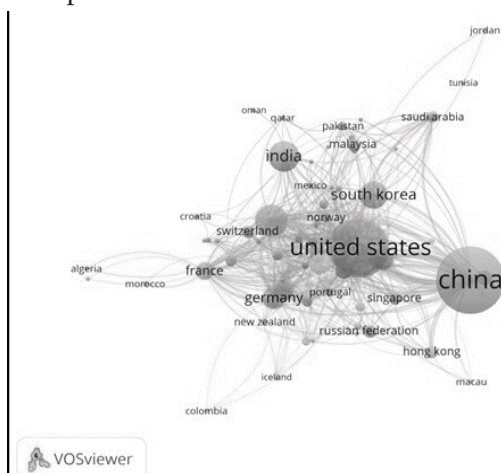
### Quadro 3 – Países mais produtivos

Países	Artigos
China	2.288
EUA	1.759
UK	550
Índia	495
Coreia do Sul	402
Alemanha	289
Áustria	289
Canadá	242
Itália	231
Espanha	214

Fonte: Construção dos autores

Ao se destacarem os dez países mais produtivos, a China ganha notoriedade por estar presente em 2.288 artigos relacionados ao tema, seguida dos EUA, com 1.759 artigos. Entretanto, para que se possa ter uma maior dimensão dos países mais produtivos, mapearam-se as coautorias entre estes países, assim como apresentado na Figura 2.

Figura 2 - Mapeamento das Coautorias Internacionais.



Fonte: Elaborada pelos autores por meio do software Vosviewer no formato CSV (Excel).



Os fenômenos relacionados ao *Big Data* estão ligados diretamente à supercomputação e de acordo com o site TOP500list, que divulga duas vezes ao ano o ranking dos computadores mais velozes do mundo, a China aparece em primeiro lugar, tendo 206 dos 500 supercomputadores listados, Estados Unidos com 124, Japão com 36, Alemanha e França com 21 e 18 supercomputadores respectivamente. Essa hegemonia tecnológica dos Estados Unidos e da China reflete-se diretamente na produção científica, como é ilustrado na Figura 2, com a indicação dos maiores círculos para os países mais produtivos.

Entretanto, observa-se que existe uma colaboração em coautorias entre os pesquisadores de diferentes países, destacando as coautorias entre Coreia do Sul, Suíça e Estados Unidos com a China, e da França com os Estados Unidos. Pode-se identificar também o domínio Chinês, principalmente com os países asiáticos como Coreia do Sul, Singapura, Macau e Hong Kong. Dos países mais produtivos que apresentam pouca coautoria internacional destacam-se o Reino Unido, Alemanha, Austrália, Canadá, Itália e Espanha.

É possível destacar que o conhecimento dessa produção científica relacionada aos temas que compõem a configuração aqui analisada permite a identificação das características da discussão científica que ocorre no contexto teórico empírico em que é produzida.

## CONSIDERAÇÕES FINAIS

Com a análise da produção de artigos na base *Scopus*, relativa aos anos de 2016 e 2017, foi possível identificar que os temas mais pesquisados envolvem os seres humanos, seja em

técnicas de coleta de dados, seja na identificação de padrões de comportamento ou genético.

Examinando o tema sob a dimensão metodológica, observa-se que para realizar essas análises são necessários o desenvolvimento de técnicas de mineração e manipulação de dados, gerenciamento de informações, principalmente na otimização do armazenamento de grandes massas de dados. Conta-se, para isso, com estudos relacionados a sistemas de computação distribuída e processamento na nuvem. Esse resultado reflete as palavras-chave mais utilizadas na análise dos artigos deste estudo.

Na análise das revistas mais produtivas, descartado o desvio das revistas venezuelanas, sugestivo do caráter geopolítico pelo qual pode ser analisada a América do Sul e seu processo de integração em curso, pode-se observar que a concentração dos estudos é nas áreas de Ciência da Computação e Engenharia. Salienta-se, entretanto, que há uma intensificação no processo de integração regional que tem desencadeado movimentos de natureza político-estratégica em termos de produção científica. Observou-se que os países mais produtivos são a China e os Estados Unidos, refletindo diretamente o investimento realizado por sua política no desenvolvimento de supercomputadores, ou o agrupamento (*cluster*) de computadores, sendo esses países os detentores do maior número de supercomputadores instalados.

Considerando-se que os Estados Unidos e a China são os maiores produtores sobre a temática, esperava-se que a intensidade da colaboração internacional entre pesquisadores destes países fosse mais destacada.

Conclui-se que o *Big Data* é um tema novo na ciência, seu reflexo pode ser observado com a produção de trinta arti-

gos sobre o tema em 2011, para 3.890 artigos produzidos em 2017, e um crescimento de 20,07% em relação a 2016. Por estar concentrada em países mais ricos, a produção em acesso aberto representou apenas 6,1% de toda produção analisada.

Salienta-se, desse modo, o aprofundamento sobre a informação e os artefatos tecnológicos como suporte aos grandes sistemas de armazenamento e tratamento da informação especializada, em decorrência da forma de fazer ciência e suas transformações ao longo dos anos. É perceptível o círculo de inter-relações que se efetiva entre áreas do conhecimento, autores, instituições, países e configurações temáticas e isso tem sido um desafio aos especialistas em estudos métricos, especialmente pelo potencial de visibilidade que podem oferecer e a consequente complexidade para sua análise e interpretação.

Esse constante e dinâmico processo de influência mútua entre a ciência e as tecnologias utilizadas na produção e comunicação do conhecimento científico tem sido potencializado, por um lado pelos avanços das tecnologias de informação e comunicação, em particular das redes computacionais e, por outro, da possibilidade de compartilhamento dos dados por parte de intermediários que atuam a partir das possibilidades concretas de compartilhar, utilizar e reutilizar dados com os seus pares.

## REFERÊNCIAS

BESSER, H. Longevidade digital. *Acervo*, Rio de Janeiro, v. 23, no 2, p. 57-70, jul/dez 2010 –Request PDF. Disponível em: [https://www.researchgate.net/publication/260301901\\_Longevidade\\_Digital](https://www.researchgate.net/publication/260301901_Longevidade_Digital). Acesso em: 21/08/2018.

BUFREM, L. S. *et al.* Produção internacional sobre ciência orientada a dados: análise dos termos Data Science e E-science na Scopus e na Web of Science. *Informação & Informação*, v. 21, n. 2, p. 40-67, 2016. DOI: 10.5433/1981-8920.2016v21n2p40.

CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SEGUNDO, J. E. S. O profissional da informação na era do Big Data. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 22, n. 50, 2017.

CONEGLIAN, C. S.; SEGUNDO, J. E. S.; SANT'ANA, R. C. S. G. A. Big Data: fatores potencialmente discriminatórios em análise de dados. *Em Questão*, v. 23, n. 1, 2017.

COORDENAÇÃO APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). *Tabelas de Áreas do Conhecimento*. Brasília: Ministério da Educação, 2014, atualizado em 2018. Disponível em: <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>. Acesso em: 21/08/2018.

EATON, J. P. *et al.* Structural analysis of co-author relationships and author productivity in selected outlets for consumer behavior research. *Journal of Consumer Psychology*, v. 8, n. 1, p. 39-59, 1999.

FURLAN, P. K.; LAURINDO, F. J. B. Agrupamentos epistemológicos de artigos publicados sobre Big Data analytics. *Transinformação*, v. 29, n. 1, p. 91-100, 2017. DOI: 10.1590/2318-08892017000100009.

GLÄNZEL, W. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Bélgica: [s.n.], 2003. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.5311&rep=rep1&type=pdf>. Acesso em: 21/08/2018.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? *Research Policy*, Amsterdam, v. 26, p. 1-18, 1997.

ROA-MARTÍNEZ, S.M.; VIDOTTI, S.A.B.; SANTANA, R. C. Estructura propuesta del artículo de datos como publicación científica. *Revista Española de Documentación Científica*, n. 40, v.1, p. e167, 2017. DOI: 10.3989/redc.2017.1.1375.

TORRES-SALINAS, D. Compartir datos (data sharing) en ciencia: contexto de una oportunidad. *Anuario ThinkEPI*, v. 1, pp. 258-261, 2009. Disponível em: <http://www.thinkepi.net/compartir-datos-data-sharing-en-ciencia-el-contexto-de-una-oportunidad>. Acesso em: 21/08/2018.