



Gestão de dados: uma abordagem bibliométrica

André José Ribeiro Guimarães^I

<http://orcid.org/0000-0003-0874-7400>

Cícero Aparecido Bezerra^{II}

<http://orcid.org/0000-0001-9058-9469>

^I Universidade Federal do Paraná, Curitiba, PR, Brasil.
Doutorando em Ciência da Informação em Gestão da Informação.

^{II} Universidade Federal do Paraná, Curitiba, PR, Brasil
Professor Associado nível II.

<http://dx.doi.org/10.1590/1981-5344/4192>

Emprega técnicas bibliométricas para identificar as características primordiais da produção de artigos científicos sobre Gestão de Dados. De natureza aplicada, abordagem quantitativa e caráter exploratório, a pesquisa faz uso de um levantamento nas bases de dados Web of Science, SCOPUS, ScienceDirect e Emerald, recuperando um total de 3.683 artigos científicos publicados entre 1963 e 2018. Revela a tendência de crescimento no ritmo de publicações sobre o tema abordado, que obteve seu auge em 2017 com 210 artigos publicados. Pela amostra, coloca Yin Zang como autor mais produtivo com 18 documentos e, pela análise de rede, coloca H. Garcia-Molina como autor mais influente. Entre as palavras-chave mais recorrentes, destaca além dos conceitos relacionados à tecnologia da informação, o termo "human" como um dos mais influentes. A análise dos periódicos ressalta a superioridade das fontes voltadas a pesquisas sobre tecnologia, com predomínio do tema "bancos de dados", mas também destaca a presença de revistas da área médica e correlatas.

Palavras-chave: *Bibliometria. Gestão de dados. Bibliometria. Análise bibliométrica. Produção científica.*

Data managment: a bibliometric approach

This study uses bibliometric techniques to identify the primordial characteristics of the production of scientific articles on Data Managment. Classified as applied in nature, with a quantitative approach and exploratory character, the research retrieves from the Web of Science, SCOPUS, ScienceDirect and Emerald databases, a total of 3,683 scientific articles published between 1963 and 2018. As an outcome, the paper reveals the trend of growth in the rhythm of publications on the topic addressed, which reached its peak in 2017 with 210 articles published. For the sample, Yin Zang is the most productive author with 18 papers and, by network analysis, H. Garcia-Molina is the most influential author. Among the most recurrent keywords, it emphasizes the concepts non-related to information technology — the term "human" with one of the most influential. The analysis of the journals highlights the superiority of the sources of research on technology, with a predominance of the theme "databases", but also highlights the presence of journals in the medical field and correlates.

Keywords: *Bibliometrics. Data managment. Bibliometria. Bibliometricanalysis. Scientific production.*

Recebido em 12.11.2019 Aceito em 13.11.2019

1 Introdução

Conceitualmente, o conhecimento pode ser colocado no topo de uma estrutura hierárquica cuja base é o conceito de dado, seguido pelo elemento intermediário, a informação. Segundo o DAMA INTERNATIONAL (2012), enquanto dados podem ser compreendidos como representações de um fato, seja no formato de texto, número, gráfico, imagem, som ou vídeo, informações são dados colocados em um contexto, o que lhes agrega relevância. Quando o significado da informação é compreendido, tem-se a concepção do conhecimento, gerando valor à organização (DAMA INTERNATIONAL, 2012).

Neste cenário, os processos para a Gestão de Dados estão envolvidos em todo componente organizacional que faz uso da informação. Documentos técnicos, projetos, contratos, relatórios financeiros devem ser compartilhados entre os departamentos, respeitando os diferentes direitos de acesso, porém apresentando facilidade para recuperação, consulta e atualização do conteúdo (SAVULESCU; POLKOWSKI; DUTTA, 2016).

Porém, não basta à Gestão de Dados fornecer o acesso à informação. Segundo Chaffey e Wood (2005), o excesso, a escassez, bem como a qualidade questionável da informação impedem que um gestor tome a decisão mais adequada. A informação de qualidade e relevante deve estar disponível no lugar correto, e no momento certo para quem dela queira e possa fazer uso. Embora as organizações reconheçam o valor da informação e, por consequência, do conhecimento, muitas vezes não há uma estratégia que vise a qualidade da informação (CHAFFEY; WOOD, 2005).

Em contrapartida, o avanço da tecnologia da informação tem impactado a forma como vivemos, fazendo com que nossas coleções de dados cresçam cada vez mais rapidamente (CHEN *et al.*, 2013). Este crescimento nas mais diversas plataformas, por sua vez, parece superar o avanço das infraestruturas computacionais, tornando-as inadequadas para a quantidade de dados que necessitam ser gerenciados (CHEN *et al.*, 2013). Tal crescimento se deve, entre outros fatores, à utilização dos mais diversos tipos de dispositivos digitais que geram continuamente quantidades massivas de estruturas heterogêneas, constituídas por dados estruturados ou não, o que resulta no chamado *big data* (RUSSOM, 2011; KAMBATLA *et al.*, 2014).

Diante deste panorama, como forma de viabilizar o processamento de grandes volumes de dados, têm-se aliado os avanços da computação à utilização de métricas como a cientometria, a informetria e a bibliometria como ferramentas para o processamento destes grandes volumes de dados (WOLFRAM, 2017). Segundo Araújo (2007, p. 12), a bibliometria é a "aplicação de técnicas estatísticas e matemática para descrever aspectos da literatura e de outros meios da comunicação (análise quantitativa da informação)".

Andrés (2009) complementa a definição do conceito de bibliometria, afirmando que a produção científica de qualquer campo pode ser analisada a partir da utilização de técnicas bibliométricas; o único pré-requisito é que se possua um conjunto adequado de publicações. Pendlebury (2008), por sua vez, define bibliometria como aplicação de análise quantitativa e estatística a publicações científicas, oferecendo vantagens durante o processo de coleta da informação objetiva necessária para tomada de decisões. A análise quantitativa da pesquisa bibliométrica oferece uma perspectiva global, uma revisão *top-down* que situa trabalhos em um contexto; também fornece dados de todas as atividades realizadas em uma área, um resumo destes dados, e uma perspectiva abrangente

sobre cada atividade, bem como os resultados alcançados (PENDLEBURY, 2008).

Na bibliometria são utilizados diversos indicadores para estudar a produção científica, as relações e interações entre seus elementos (MARICATO, 2011, p. 2). Em linhas gerais, estes indicadores buscam medir a produtividade, estabelecer relações, identificar colaboração e coocorrência de elementos em uma amostra de documentos. O papel da área acadêmica, em especial os cursos de pós-graduação *Stricto Sensu*, é fundamental na produção e disseminação de conhecimento, visando "atender às emergentes demandas das organizações e da sociedade" (MACHADO JUNIOR et al., 2016, p. 112). Artigos científicos, dissertações e teses apresentam, além de levantamentos teóricos baseados em outros documentos científicos, padrões rígidos de metodologia e padrões de pesquisa pautados em normas técnicas, sendo consideradas fontes fidedignas para extração de informações a partir de metodologias bibliométricas (ibidem, p. 113).

Diante da conjuntura apresentada, pergunta-se qual seria o panorama da produção científica sobre Gestão de Dados, segundo parâmetros da bibliometria? De forma a responder a esta questão, a presente pesquisa tem como objetivo aplicar técnicas de bibliometria para identificar as principais características da produção em periódicos científicos sobre Gestão de Dados.

2 Procedimentos metodológicos

Em relação à sua natureza, a presente pesquisa pode ser classificada como aplicada. Também apresenta abordagem quantitativa e, sob o ponto de vista dos objetivos, é considerada exploratória. Por fim, sob a perspectiva dos procedimentos técnicos, constitui-se em uma pesquisa bibliográfica (SILVA; MENEZES, 2005). Foram seguidos os passos metodológicos propostos pela tese de Maricato (2011), na qual o autor discute desde os processos de coleta e tratamento dos dados à construção de indicadores bibliométricos.

Como etapa precedente à coleta de dados, definiram-se as bases a serem consultadas: Web of Science, Scopus, ScienceDirect e Emerald e a estratégia de busca: (TITLE("data managment") AND DOCTYPE(ar)). Como observado, o termo de busca "*data managment*" foi restrito ao título dos recursos e o tipo de documento foi definido como artigo. Uma vez que um dos objetivos do trabalho foi identificar um panorama sobre as publicações referentes à Gestão de Dados, não foi aplicado nenhum tipo de filtro referente à data de publicação dos artigos. A pesquisa às bases de dados foi realizada em 13 de agosto de 2018. Além das bases internacionais, foi realizada uma busca na base de dados Scielo.org que não retornou nenhum resultado. Esta busca atendeu aos mesmos critérios estabelecidos, todavia com o termo de busca em português ("*gestão de dados*").

Adicionalmente, foi realizada uma busca com o objetivo de identificar pesquisas preliminares sobre o tema deste trabalho. A

estratégia adotada foi ((TITLE("data management") AND TITLE("bibliometric*")) AND DOCTYPE(ar)). As bases Web of Science e Scopus retornaram um único documento: *Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research* escrito por Corral, Kennan e Afzal (2013). No entanto, o conteúdo do documento pouco se relaciona com o objetivo aqui estabelecido.

O detalhamento dos passos metodológicos é apresentado no Quadro 1:

Quadro 1 - Protocolo de pesquisa

Etapa	Objetivos	Materiais
1 Coleta de dados	Coletar os artigos com "data management" no título para compor o corpus de análise.	Bases de dados Web of Science, Scopus, ScienceDirect e Emerald
2 Tratamento de dados	Remover documentos duplicados (retornados por mais de uma base de dados).	EndNote® X6 (THOMSON REUTERS, 2015)
3 Indicadores descritivos	Relatar indicadores extraídos manualmente da amostra de documentos.	EndNote® X6 (THOMSON REUTERS, 2015)
4 Visualização de redes	Identificar as conexões formadas pelos componentes da amostra.	VOSviewer (VAN ECK; WALTMAN, 2018)
5 Análise dos resultados	Descrever as informações relevantes obtidas pelas etapas 3 e 4.	

Fonte: Elaborado pelos autores (2019).

A etapa de coleta de dados resultou em uma amostra de 4.913 artigos, assim distribuídos: Web of Science (1.648), Scopus (2.764), ScienceDirect (437) e Emerald (64). Todos os artigos foram exportados das bases com o maior número de informações possíveis, incluindo referências utilizadas no formato RIS (*Research Information Systems*) para serem importados no software de gerenciamento de referências, EndNote® X6. Na segunda etapa, tratamento de dados, por meio do EndNote® X6 foi possível identificar e remover os documentos duplicados. Por padrão, o EndNote® X6 considera como duplicado aquele documento com mesmo título, autores e ano. Porém, esta configuração inicial não se mostrou totalmente eficaz, uma vez os títulos apresentavam inconsistência entre as bases. Muitos foram os casos onde o título e subtítulo eram separados por hífen em uma base e por dois pontos (:) em outras. Em outros casos, alguns termos foram escritos de maneira diferente como "*data-base*" e "*database*". Nestes casos, o software não considerou o título como sendo o mesmo. Por isso, foi necessário alterar a configuração para que fossem considerados como duplicados artigos com os mesmos autores, anos, periódico, volume e número. Após o primeiro procedimento de identificação de duplicados, a amostra caiu para 3.945 documentos e, ao final, consolidou-se em 3.683 artigos, correspondendo ao *corpus* de análise.

O EndNote® X6 foi utilizado para exportar dados descritivos sobre a amostra de documentos. Por meio dele, foi possível identificar a quantidade de artigos por ano, quantidade de artigos por autor, palavras-chave mais recorrentes e periódicos com mais artigos publicados. Esta foi a terceira etapa de análise e seus resultados são apresentados na próxima seção.

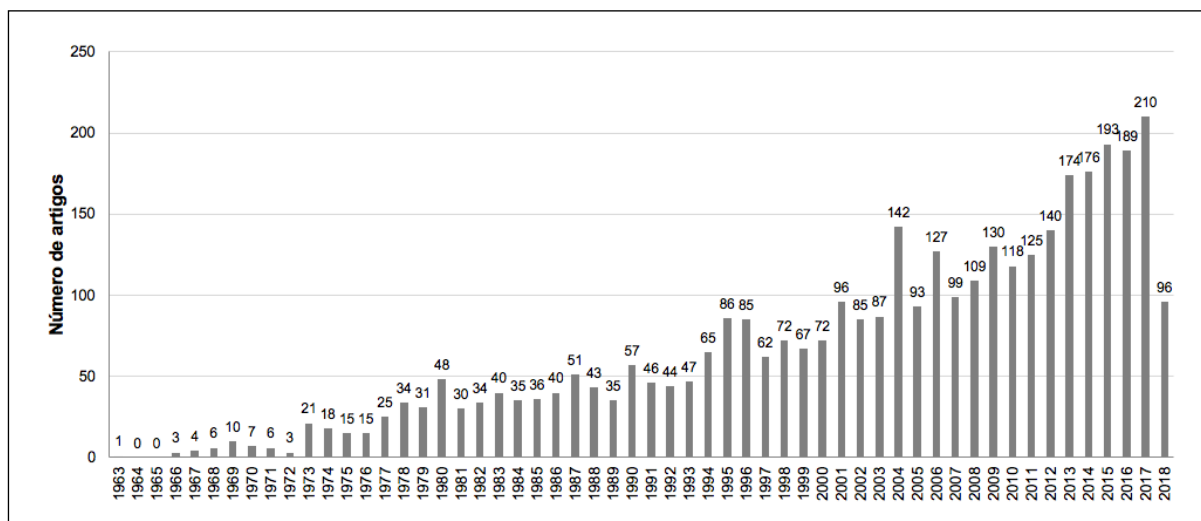
A quarta fase se caracterizou pela utilização da ferramenta VOSviewer, um software para criar, visualizar e explorar mapas bibliométricos baseados em dados de redes (VAN ECK; WALTMAN, 2018, p. 3). Uma vez que para explorar um maior número de possibilidades do VOSviewer é necessário trabalhar separadamente com os arquivos exportados de uma única base de dados, optou-se pela utilização dos 2.764 artigos recuperados pela SCOPUS. Ou seja, a análise de rede foi restrita a 75,04% do *corpus* inicialmente estabelecido. É possível importar os conteúdos de todas as bases no formato RIS, porém, algumas construções de rede, como cocitação, ficam desabilitadas.

Por fim, a etapa de análise dos resultados buscou encontrar informações e padrões relevantes diante dos métodos aplicados. A seguir, resultados são apresentados juntamente com a discussão dos achados da pesquisa.

3 Análise e discussão dos resultados

Primeiramente, a Figura 1 apresenta a distribuição dos artigos recuperados ao longo dos anos.

Figura 1 - Artigos sobre Gestão de Dados



Fonte: Elaborado pelos autores (2019).

O primeiro artigo recuperado foi escrito por Hall e publicado em 1963 com o título "*Modern Data Managment of Medical Board Reports*", no periódico suíço *Svenskaläkartidningen*, da área de medicina. Após dois anos sem artigos publicados, cujo título possuísse o termo "*data managment*", em 1966 três artigos foram recuperados. Desde então, a quantidade de artigos publicados anualmente demonstra uma forte

tendência de aumento, chegando ao ápice em 2017 com 210 registros. Salienta-se também, que os artigos publicados nos últimos cinco anos já completados (2013 a 2017) somam 942 registros e representam 25,58% das publicações contidas no *corpus* analisado.

3.1 EndNote – autores

A segunda informação retirada do EndNote® X6 foi a quantidade de artigos encontrados por autor. É importante salientar que o software separa os autores de um documento, ou seja, um mesmo artigo com três coautores é contabilizado para cada um dos três. A Tabela 1 apresenta os autores com maior número de artigos publicados.

Tabela 1 - Autores com mais artigos publicados

Posição	Autor	Artigos	Qtde.
1	Zhang, Y.	18	1
2	Wang, J.	12	1
3	Li, Y.	11	1
4	Ooi, B. C.; Wang, X.; Zhang, X.	9	3
5	Benson, M.; Chen, G.; Chen, L.; Kim, J.; Tan, K. L.; Zhang, J.	8	6
6	Castellanos, I.; Hempelmann, G.; Junger, A.; Kumar, N. Lee, J. S.; Prokosch, H. U.; Wang, Y.; Wu, S.; Yokoyama, H.; Zeadally, S.	7	10

Fonte: Elaborado pelos autores (2019).

O autor com maior número de artigos publicados foi Yin Zhang, cujo primeiro trabalho data de 1998 e o mais recente, de 2017. O segundo autor com mais publicações é J. Wang com 12 registros e o terceiro, Y. Li, com 11 documentos. Na sequência, três autores apresentam nove trabalhos; seis apresentam oito e, finalmente, sete artigos foram encontrados para sete autores. No extremo oposto, 1.514 autores publicaram dois documentos e 8.233, apenas um artigo. Isso indica que 81,94% dos 10.047 autores encontrados publicaram um único documento que atendia aos critérios da busca.

3.2 EndNote – palavras-chave

Do EndNote® X6, também foi possível exportar as palavras-chave mais recorrentes nos artigos. A Tabela 2 apresenta os 20 termos mais encontrados.

Tabela 2 - Palavras-chave mais recorrentes

Posição	Palavra-chave	Artigos
1	<i>data managment</i>	580
2	<i>Information managment</i>	518
3	<i>human</i>	360
4	<i>information processing</i>	323
5	<i>database</i>	288
6	<i>humans</i>	238
7	<i>database systems</i>	206
8	<i>data managment systems</i>	187
9	<i>software</i>	165
10	<i>computer program</i>	163
11	<i>data processing</i>	162
12	<i>information system</i>	132
13	<i>internet</i>	117
14	<i>quality control</i>	109
15	<i>organization and managment</i>	106
16	<i>data collection</i>	103
17	<i>information retrieval</i>	95
18	<i>managment information systems</i>	95
19	<i>computer analysis</i>	83
20	<i>big data</i>	77

Fonte: Elaborado pelos autores (2019).

Naturalmente, o termo que direcionou a coleta de dados, "*data managment*" foi o mais recorrente com 580 artigos. Nota-se que mesmo que o termo se encontre nos títulos de todos os documentos, nem todos os artigos utilizaram a expressão entre suas palavras-chave. O segundo termo da lista é "*information managment*" com 518 ocorrências. Os conceitos de Gestão da Informação e Gestão de Dados estão intimamente ligados, sendo muitas vezes utilizados como sinônimos (DAMA INTERNATIONAL, 2012).

Foi possível verificar que os termos "*human*", com 360 artigos, e "*humans*", com 238 artigos, referem-se a um mesmo conceito, mesmo que o EndNote® X6 tenha lhes tratados de maneira individualizada. Juntos eles somaram 598 ocorrências, tornando-se a palavra-chave mais recorrente.

Além disso, destacaram-se termos referentes ao processamento de dados e informação, como "*information processing*", "*data processing*", "*data collection*" e "*informationretrieval*". Por fim, os termos ligados à tecnologia da informação também se destacam: "*database*", "*database*

systems", *software*", *computerprogram*". Todavia, é importante ressaltar que estes termos muitas vezes são relacionados e possuem áreas de intersecção. A visualização de redes pode fornecer novas análises.

3.3 EndNote – periódicos

A última informação retirada do EndNote® X6 foi a quantidade de artigos encontrada por periódico, que é apresentada na Tabela 3:

Tabela 3 - Periódicos com mais artigos publicados

Posição	Periódicos	Artigos
1	SIGMOD Record	38
2	Computer Integrated Manufacturing Systems, CIMS	27
3	Proceedings of the VLDB Endowment	25
4	Therapeutic Innovation & Regulatory Science	22
5	Bioinformatics	21
	BMC Bioinformatics	
	IEEE Transactions on Knowledge and Data Engineering	
6	Procedia Computer Science	21
7	Fusion Engineering and Design	17
8	American Laboratory	16
	Drug Information Journal	
9	Computer Methods and Programs in Biomedicine	15
	Computers in Industry	
10	ACM SIGMOD Record	14

Fonte: Elaborado pelos autores (2019).

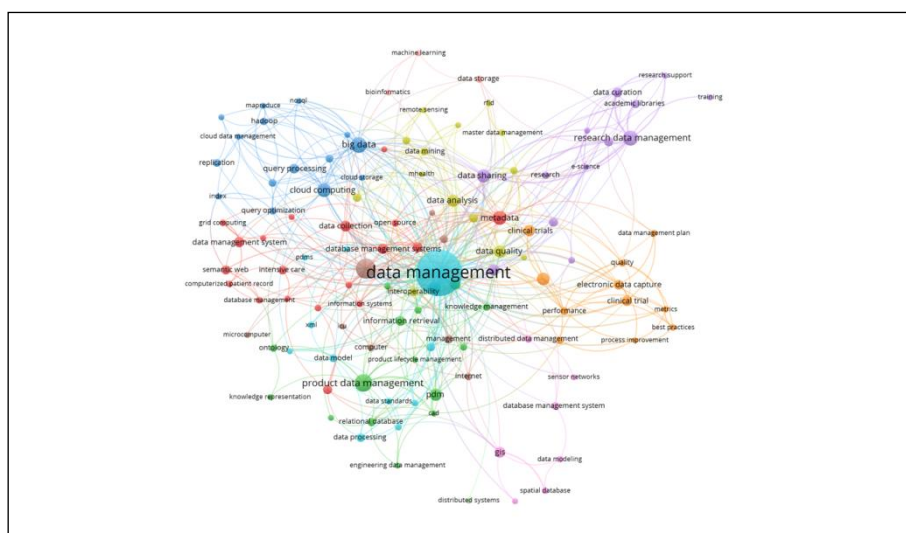
A lista contém um ranking com 10 posições, mas apresenta 14 periódicos, uma vez que três deles apresentaram 21 artigos (ocupando a quinta posição), dois apresentaram 16 artigos (oitava posição) e outros dois, 15 artigos (nona posição). O periódico com maior número de artigos foi SIGMOD Record com 38 publicações. Esta revista científica possui publicações trimestrais e é dedicada ao estudo, desenvolvimento e aplicações em banco de dados e tecnologia da informação (SPECIAL INTEREST GROUP ON MANAGEMENT OF DATA, 2018).

Além dos periódicos da área da tecnologia da informação, destaca-se a presença de publicações de áreas relacionadas à saúde como *"Therapeutic Innovation & Regulatory Science"*, *"Bioinformatics"*, *"BMC Bioinformatics"*, *"American Laboratory"* (química), *"Drug Information Journal"* e *"Computer Methods and Programs in Biomedicine"*. Esta maior incidência pode estar relacionada ao fato de que as áreas correlatas à medicina e à biologia apresentam enorme demanda na gestão de dados. O volume de dados gerados tem se tornado um desafio para a rápida recuperação e precisão de informações necessárias à tomada de decisão, como um diagnóstico médico (NG *et al.*, 2015, p. 1). O fato do primeiro artigo publicado no *corpus* analisado, apresentado no início desta seção, ser da área de medicina reforça que esta necessidade não é recente.

3.4 VOSviewer – palavras-chave dos autores

Para o primeiro mapa de rede gerado pelo VOSviewer por meio dos registros obtidos na base de dados SCOPUS, optou-se pela utilização das palavras-chaves utilizadas pelos autores, uma vez que o mesmo conteúdo foi analisado com o uso do EndNote® X6. A Figura 2 apresenta este primeiro mapa.

Figura 2 - Mapa das Palavras-chave dos autores



Fonte: Elaborado pelos autores (2019).

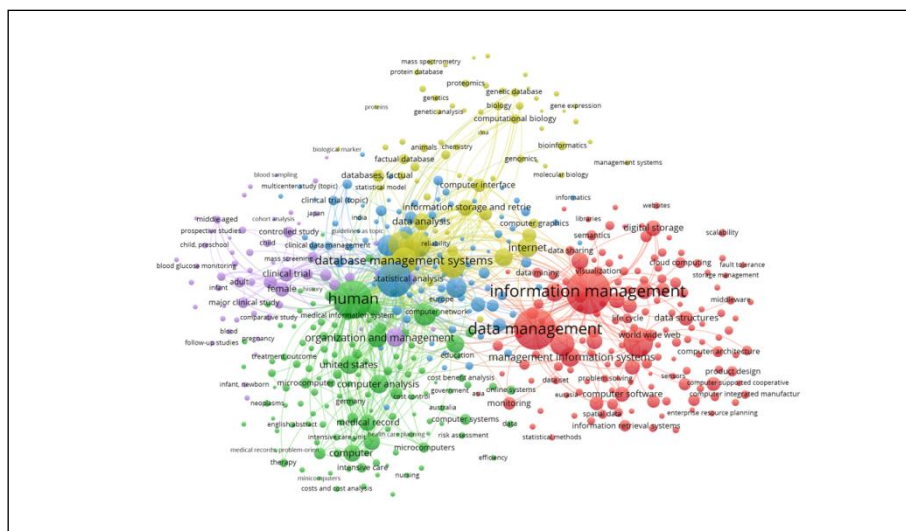
O mapa destaca o termo *"data management"* como sendo o conceito central de todas as relações estabelecidas. Dentre os outros conceitos que possuem fortes ligações com os demais e podem ser destacados encontram-se *"database"*, *"research data management"*, *"data sharing"*, *"big data"*, *"metadata"* e *"cloud computing"*. O VOSviewer organizou os elementos do mapa em 11 clusters. O maior deles contém 19 itens e é representado pela cor vermelha no mapa. Este grupo engloba termos relacionados a sistemas de informação e sistemas de gestão de banco de dados, mas também abriga itens como *"python"*, *"quality control"*, *"web semantic"* e *"open source"*. O segundo maior cluster, representado em verde no mapa, contém itens *"knowledge management"*, *"data"*, *"information retrieval"*, *"product data management"*, *"ontology"* e *"engineering data management"*. Destaca-se o cluster referente aos conceitos mais voltados à tecnologia, em especial, a computação na nuvem, com 15 itens, dentre eles: *"big data"*, *"cloud computing"*, *"cloud storage"*, *"query processing"*, *"distributed computing"*, *"hadoop"*, *"query optimization"*. Por fim, outro grupo que merece destaque se refere aos conceitos relacionados à Gestão de Dados na ciência com termos como: *"research data management"*, *"academic libraries"*, *"data sharing"* e *"e-science"*.

3.5 VOSviewer – todas as palavras-chave

Além das palavras-chave utilizadas pelo autor, o VOSviewer oferece a possibilidade de visualização do mapa formado por todas as palavras dos documentos importados, como título e resumo. O mapa resultado desta aplicação é mostrado na

Figura 3.

Figura 3 - Mapa de todas as Palavras-chave



Fonte: Elaborado pelos autores (2019).

Neste método de análise, os pesos dos conceitos, isto é, a força das ligações que apresentadas na rede, foram alterados. O termo *"human"* foi considerado o mais forte, seguido por *"information processing"* e *"information management"*. Gestão de Dados, com seu termo *"data management"* ocupou apenas a quinta posição nesta ordenação. Isso indica que embora o termo estivesse no título de todos os documentos, não era, necessariamente, o tema central dos artigos.

Em relação à formação de clusters, o mapa apresentou cinco agrupamentos. O maior deles, representado em vermelho no mapa, contém além de dois encontrados entre os mais fortes, *"information management"* e *"data management"*, outros conceitos relacionados à área da tecnologia da informação, como *"database systems"*, *"data processing"*, *"data structures"*, *"cloud computing"*, *"big data"*. O segundo maior cluster, representado pela cor verde no mapa, contém além do principal termo do mapa, *"humans"*, conceitos relacionados a sistemas de informação (*"information systems"*, *"organization and management"*, *"online systems"*) e conceitos relacionados à área médica (*"medical records"*, *"hospital information systems"*, *"medical documentation"*, *"patient care"*). O terceiro maior cluster, representado em roxo, também apresenta termos relacionados à saúde, mas indica ligações a estudos e

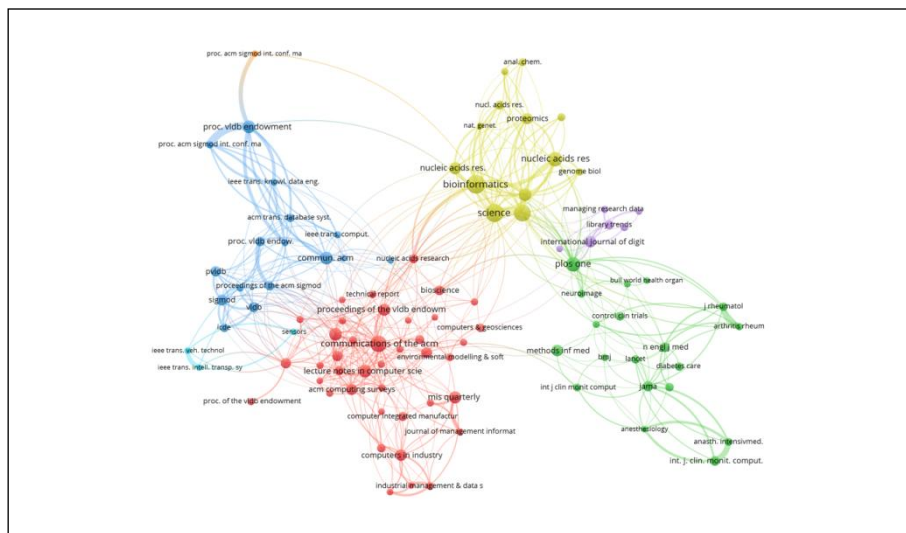
pesquisas na área (“*clinical trial*”, “*major clinical study*”, “*male*”, “*female*”, “*controlled study*”).

3.6 VOSviewer – cocitação dos periódicos

Ao analisar os artigos importados e as referências utilizadas, o VOSviewer consegue criar um mapa de cocitação dos periódicos mais relevantes para a temática estudada. O resultado desta análise é mostrado na

Figura 4.

Figura 4 - Mapa de cocitação dos periódicos



Fonte: Elaborados pelo autores (2019).

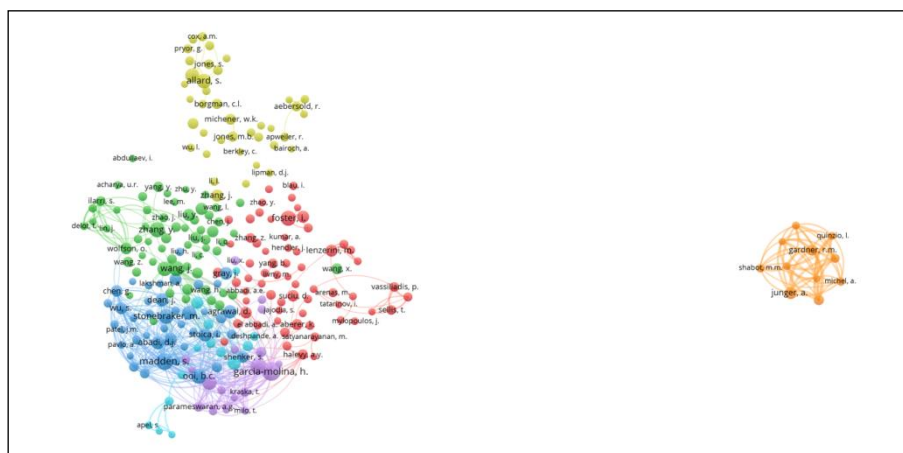
O periódico mais cocitado é o “*bioinformatics*” (158 citações), seguido pela “*science*” (142 citações) e pela “*nature*” (129 citações). Estas três revistas fazem parte do quarto maior número *cluster* do mapa, representando em amarelo. O maior *cluster*, formatado em vermelho, contém os periódicos mais diretamente ligados à área de Gestão de Dados (“*communications of the acm*”, “*acm transactions on data bases*”, “*acm computer surveys*”). Destacam-se também o *cluster* verde referente a periódicos relacionados à área médica e biológica (“*plus one*”, “*neuro image*”, “*j rheumatol*”) e o *cluster* azul que contém alguém periódicos que publicam sobre pesquisas em grandes bancos de dados (“*proc. vldbendowment*”, “*communications ofacm*”, “*iee trans. comput.*”).

3.7 VOSviewer – cocitação dos autores

Por fim, a última análise que utilizou o VOSviewer buscou identificar as redes de cocitação entre autores relacionados ao tema da Gestão de Dados, ou seja, quais autores são mais citados simultaneamente. A

Figura 5 apresenta o mapa resultante desta análise.

Figura 5– Mapa de cocitação dos autores



Fonte: Elaborado pelos autores (2019).

A análise apontou H. Garcia-Molina como autor mais cocitado, com 129 cocitações, e com as ligações mais fortes na rede, centralizado as relações no *cluster* roxo. Ele atua como professor no departamento de Ciências da Computação e Engenharia Elétrica da Universidade de Stanford e também é membro da Association for Computing Machinery (ACM) e da American Academy of Arts and Sciences, tendo recebido o

prêmio ACM SIGMOD Innovations de 1999¹. O segundo autor mais cocitado é Samuel Madden, 103 citações, e está localizado no *cluster* colorido de azul. Yin Zhang, autor que teve mais artigos encontrados nesta pesquisa, também ocupa posição destacada no mapa, no centro do *cluster* verde.

Por fim, destaca-se o *cluster* formatado em laranja, que não apresentou ligações com os demais autores. Em suma, os artigos que citam os autores deste *cluster* não citam autores situados em outros *clusters*. Um dos motivos pode ser a data na qual os artigos destes autores foram publicados. Dentro do *corpus*, R. M. Gartner, o autor mais relevante deste aglomerado, possui dois artigos, publicados em 1985 e 1990, respectivamente. O tempo transcorrido desde estas publicações fez com que seu conteúdo se tornasse obsoleto.

4 Considerações finais

Neste estudo, a aplicação de métodos bibliométricos em uma amostra de 3.683 de artigos cujo título continha o termo “*data managment*” buscou responder à pergunta: qual o panorama da produção científica sobre gestão de dados segundo parâmetros da bibliometria? A análise da distribuição dos artigos em anos permitiu identificar quando foi publicado o primeiro documento que atendia aos critérios de busca. Outrossim, relevou-se que há uma tendência de crescimento nas publicações sobre Gestão de Dados, sendo o ano de 2017 aquele com maior número de registros e os últimos cinco anos completos (2014 a 2018) correspondem a 25,58% de toda a amostra.

As análises referentes aos autores permitiram a listagem dos nomes mais produtivos, bem como dos mais relevantes para a temática de pesquisa. Em relação aos periódicos, foi possível estabelecer os grupos e instituições mais relevantes para a área. Por fim, analisando as palavras-chave ressaltaram-se os principais conceitos que estão relacionados ao tema central, Gestão de Dados. Destaca-se ainda, que o tema pesquisado se mostrou relevante não só para as áreas de tecnologia da informação e gestão, mas também para áreas correlatada à biologia, em especial, a medicina.

Diante do exposto, julga-se que a aplicação de técnicas de bibliometria se mostrou adequada para identificar as características primordiais da produção em periódicos científicos sobre Gestão de Dados. Todavia, a pesquisa apresenta limitações. Em princípio, é necessário reafirmar que as análises foram aplicadas a uma amostra obtida por quatro bases de dados, não podendo, assim, serem generalizadas. Ademais, conforme Wolfram (2017, p. 98), embora as análises de ligações baseadas em “citações (citação direta, cocitação, acoplamento bibliográfico) ou coautoria” ofereçam uma rotina para resumir e visualizar grandes massas de dados, elas também apresentam desvantagens. Uma

¹ (<http://infolab.stanford.edu/people/hector.html>).

delas é não permitir a identificação de similaridades entre documentos com interesses em comum se os mesmos não apresentarem relação entre si.

Dentre as sugestões para trabalhos futuros, ressalta-se a submissão dos dados às três leis fundadoras da bibliometria (Lei de Lotka, Lei de Bradford e Lei de Zipt). Outra recomendação é a exploração de outros tipos de análise disponíveis no VOSviewer. Por fim, uma abordagem dos conteúdos por décadas poderia ressaltar os conceitos predominantes em cada época, revelando o surgimento e desaparecimento de conceitos, sejam eles metodológicos ou tecnológicos.

Referências

ANDRÉS, A. *Measuring academic research: How to undertake a bibliometric study*. Elsevier, 2009.

ARAÚJO, C. A. A. Bibliometria: evolução histórica e questões atuais. *Em Questão*, v. 12, n. 1, 2007.

CHAFFEY, D.; WOOD, S. *Business information management: improving performance using information systems*. London: Financial Times, 2005.

CHEN, J.; CHEN, Y.; DU, X.; LI, C.; LU, J.; ZHAO, S.; ZHOU, X. Big data challenge: a data management perspective. *Frontiers of Computer Science*, v. 7, n. 2, p. 157–164, 6 abr. 2013. Disponível em: <http://link.springer.com/10.1007/s11704-013-3903-7>. Acesso em:

DAMA INTERNATIONAL. O Guia da DAMA para o Corpo de Conhecimento em Gestão de Dados (DAMA-DMBOK). TechnicsPublications, 2012.

KAMBATLA, K.; KOLLIAS, G.; KUMAR, V.; GRAMA, A. *Trends in big data analytics*. *Journal of Parallel and Distributed Computing*, v. 74, n. 7, p. 2561–2573, 2014. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0743731514000057>.

MACHADO JUNIOR, C.; SOUZA, M. T. S. de; PARISOTTO, I. R. dos S.; PALMISANO, A. As Leis da Bibliometria em Diferentes Bases de Dados Científicos. *Revista de Ciências da Administração*, v. 18, n. 44, p. 111–123, 2016. Disponível em: <https://periodicos.ufsc.br/index.php/adm/article/view/2175-8077.2016v18n44p111>. Acesso em: 30 out. 2019.

MARICATO, J. de M. Procedimentos metodológicos em estudos bibliométricos e cientométricos: opções e reflexões no contexto dos processos de recuperação e organização da informação. In: COSTA, R. L. M. (Ed.). *Estudos contemporâneos em comunicações e artes: melhores teses e dissertações da ECA/USP*. São Paulo: ECA/USP, 2011. p. 1–19.

NG, K. H.; FAUST, O.; SUDARSHAN, V.; CHATTOPADHYAY, S. Data Overloading in Medical Imaging: Emerging Issues, Challenges and Opportunities in Efficient Data Management. *Journal of Medical Imaging and Health Informatics*, v. 5, n. 4, p. 755–764, ago. 2015.

PENDLEBURY, D. A. White paper: Using bibliometrics in evaluating research. 2008.

RUSSOM, P. Big Data Analytics. TDWI Best Practices Report, v. 19, n. 4, p. 1–34, 2011.

SAVULESCU, C.; POLKOWSKI, Z.; DUTTA, N. *Collaborative data management for business: A review of collaborative techniques*. 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), p. 1–4, 2016. Disponível em: <http://ieeexplore.ieee.org/document/7861144/>. Acesso em: 30 out. 2019.

SILVA, E. L. da; MENEZES, E. M. *Metodologia da pesquisa e elaboração de dissertação*. 4. ed. Florianópolis: UFSC, 2005.

SPECIAL INTEREST GROUP ON MANAGEMENT OF DATA. SIGMOD Record. Disponível em: <https://sigmodrecord.org/>. Acesso em: 15 ago. 2018.

THOMSON REUTERS. *EndNote*. Disponível em: <https://endnote.com/>. Acesso em: 20 jan. 2015.

WOLFRAM, D. A pesquisa bibliométrica na era do big data: Desafios e oportunidades. In: MUGNAINI, R.; FUJINO, A.; KOBASHI, N. Y. (Ed.). *Bibliometria e cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na era do Big Data*. São Paulo: ECA/USP, 2017. p. 91–100.