Ernane Rosa Martins
(Organizador)

A Abrangência da Ciência da **Computação na Atualidade**

Atena Editora
Ano 2019

Ernane Rosa Martins

(Organizador)

A Abrangência da Ciência da Computação na Atualidade

Atena Editora
2019

2019 *by Atena Editora*

Editora Executiva: Profª Drª Antonella Carvalho de Oliveira
Diagramação: Karine de Lima
Edição de Arte: Lorena Prestes
Revisão: Os Autores

Atena Editora
Ano 2019

Dados Internacionais de Catalogação na Publicação (CIP)
(eDOC BRASIL, Belo Horizonte/MG)

A161 A abrangência da ciência da computação na atualidade [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa, PR: Atena Editora, 2019.

Formato: PDF
Requisitos de sistema: Adobe Acrobat Reader
Modo de acesso: World Wide Web
Inclui bibliografia
ISBN 978-85-7247-488-7
DOI 10.22533/at.ed.887190908

1. Computação – Pesquisa – Brasil. I. Martins, Ernane Rosa.
CDD 004

Elaborado por Maurício Amormino Júnior – CRB6/2422

Atena Editora
Ponta Grossa – Paraná - Brasil
www.atenaeditora.com.br
contato@atenaeditora.com.br

Atena Editora
Ano 2019

APRESENTAÇÃO

A área da Ciência da Computação apresenta atualmente uma constante ascensão, seus profissionais estão sendo cada vez mais valorizados e requisitados pelas empresas, tornando-a mais importante, prestigiada e reconhecida. As empresas de todos os portes e setores necessitam de profissionais qualificados desta área, que apresentem potencial para promover inovação, desenvolvimento e eficiência.

A Ciência da Computação é uma área com amplas possibilidades de atuação, como por exemplo: a elaboração de programas e softwares, o gerenciamento de informações, a atuação acadêmica, a programação de aplicativos mobile ou ainda de forma autônoma. A abrangência da Ciência da Computação exige de seus profissionais conhecimentos diversos, tais como: novos idiomas, pensamento criativo, capacidade de comunicação e de negociação, além da necessidade de uma constante atualização de seus conhecimentos.

Dentro deste contexto, este livro aborda diversos assuntos importantes para os profissionais e estudantes desta área, tais como: API de localização da google, identificação de etiquetas RFID, ferramentas para recuperação de dados, ensino de computação, realidade virtual, interação humano computador, gestão do conhecimento, computação vestível, gerência de projetos, big data, mineração de dados, Internet das coisas, monitoramento do consumo de dados na Internet, pensamento computacional, análise de sentimentos, filtros ópticos, rede óptica elástica translúcida, algoritmo de roteamento, algoritmo de atribuição espectral, algoritmo de utilização de regeneradores e algoritmo genético.

Assim, certamente que os trabalhos apresentados nesta obra exemplificam um pouco a abrangência da área de Ciência da Computação na atualidade, permitindo aos leitores analisar e discutir os relevantes assuntos abordados. A cada autor, nossos agradecimentos por contribuir com esta obra, e aos leitores, desejo uma excelente leitura, repleta de boas reflexões.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 8

EXPLORING *BIG DATA* CONTENT AND INFORMATION METRICS: INTERSECTIONS AND ANALYSIS TO SUPPORT DECISION-MAKING

Rafael Barcellos Gomes
(ORCID: https://orcid.org/0000-0003-1179-2923)
Universidade Federal do Rio de Janeiro – UFRJ
Rio de Janeiro – RJ

Vânia Lisboa da Silveira Guedes
Universidade Federal do Rio de Janeiro – UFRJ
Rio de Janeiro – RJ

ABSTRACT: It proposes to expose theoretical and practical aspects in metric studies of scientific information associating them with the exploration of *Big Data* to delineate tendencies and scenarios on Public Health in Brazil. All the while dissecting the contemporary concept of information and its strategic importance to support decision-making in organizations inserted in the context of the Information and Knowledge Society, examining the importance of scientometric studies and network produced data analysis for the development of valuable information discovery processes. To do this, it investigates possible relationships between data from *Big Data*, official data published by the Brazilian Ministry of Health, and technical data extracted from scientific publications indexed in research databases. Patterns of convergence among the observations were recorded, noting that some occurrences of endemic outbreak peaks are preceded by peaks of informational demands through terms related to infectious agents involved in endemics. Finally, it highlights the relevance of analytical methodology - based on data visualization - for the development of studies that support decision-making processes and organizational strategies, outlining scenarios and trends to support Public Administration and private management.

KEYWORDS: *Big data*, data mining, scientometric analysis, strategic management, decision-making.

EXPLORANDO DADOS DE *BIG DATA* E MÉTRICAS DE INFORMAÇÃO: INTERSEÇÕES E ANÁLISES PARA SUPORTE A TOMADAS DE DECISÃO

RESUMO: Propõe expor aspectos teóricos e práticos em estudos métricos da informação científica, associando-os à exploração de dados de *Big Data* para deslindamento de tendências e cenários em Saúde Pública, no Brasil. Neste ínterim, disseca o conceito contemporâneo de informação e a sua importância estratégica para apoio a tomadas de decisão em organizações inseridas no contexto da Sociedade da Informação e do Conhecimento, examinando a importância dos estudos cienciométricos e da análise de dados produzidos em rede no desenvolvimento de processos de descoberta de informações de valor. Para tal, investiga

possíveis relações entre dados de *Big Data*, dados oficiais publicados pelo Ministério da Saúde do Brasil e órgãos correlatos e, dados técnicos extraídos de publicações científicas indexadas em bases de pesquisa. Registrou-se padrões de convergência entre as observações colhidas, constatando que alguns picos de ocorrência de surtos endêmicos são precedidos por picos de demandas informacionais por termos relacionados aos agentes infecciosos envolvidos nas endemias. Por fim, destaca a relevância da metodologia analítica – baseada na visualização de dados – para o desenvolvimento de estudos que apóiem processos decisórios e estratégias organizacionais em políticas de Saúde, delineando cenários e tendências para suporte à Administração pública e à gestão privada.
PALAVRAS-CHAVE: *Big data*, mineração de dados, análise cienciométrica, gestão estratégica, processo decisório.

1 I INTRODUCTION

Over the last four decades public and private administrations have been strongly impacted by new dynamics in the political, social, economic and technological fields, as a result of the accelerated changes that have taken place since the emergence of the so-called Information and Knowledge Society - in the early 1980s Brazil.

This moment was characterized above all by the migration of a characteristically industrial economic paradigm - based on manufacturing and industrial assets into a post-industrial one - characterized by an essentially tertiary economy, in which social, economic, and power relations settle based on information mediation, made possible by the development of new information and communication technologies.

The problem here approached is derived from the understanding that - in this new and challenging conjuncture - the survival and success of contemporary organizations depend on their adaptive capacity in face of the new configurations of demand and market, as well as the ability to distinguish the spatiotemporal context in which they are acting, transposing the challenges and taking advantage of opportunities (NEELY; GREGORY; PLATTS, 1995). These discernment and perception by the organization are born from the generation of knowledge, product of this element that moves the globalized world: information.

2 I CONCEPTUALIZATION AND CONTEXTUALIZATION

In the context of post-industrial society, information becomes more important in the process of capital generation, a fact by which it comes to be compared to an "[...] economic asset (commodity) [...]" and to a "[...] product/merchandise [...]" (TARAPANOFF, 2001, p.91), which are essential to the creation of knowledge and intelligence, whether it is for leisure, for work or for teaching, research and development. From this unusual socio-economic context there emerged a new understanding of effective business management, whose decision-making processes are based on the search for and use

of information resources. This was called Information Management (IM).

From a philosophical-epistemological point of view, Rascão (2006, p.33) understands "information" under four fundamental perspectives: information as a thing, as a process, as social construction, and information as "potentiality" (in the sense that a piece of data exists as a potential source of information and not as a piece of information itself).

Rascão then proposes to interweave the fundamental concepts of information to the concept of Information Management, noting that in the context of Information Society this management seeks "[...] to provide the right information (as a thing) to the right person at the right moment and in the right place [...]", but also to preserve the information generated from the individual cognitive process, as well as information generated with "[...] the language, thought and dialogue [...]" (RASCÃO, 2006, p.51-52) of social groups.

From this understanding, it is possible to comprehend that the whole process of signification, valorization and the use of analyzed information will occur under cultural influence, which will also interfere in the profile of the strategies constructed for decision-making - both individual and organizational.

It is observed, therefore, that in Information Management in the context of the New Economy it is possible to recognize information under various aspects (KIRK, 1999, p.12). In this perspective, the concept of potential information - discussed by Rascão (2006), Shannon and Weaver (1948), among others - emerges as a particularly interesting aspect, especially with regard to the practice of Information Management in the last three decades.

Valentim (2010) calls this period New Economy, an expression that reinforces the perception of the concept of information as "commodity" (TARAPANOFF, 2001, p.91). In this model, social and economic links are each time less structured on tangible products (SCHWAB, 2016) and, increasingly, in the provision of services: online, on-time, full-time.

3 I DATA AND INFORMATION

Becker (2015) analyzes a series of statisticians and information theorists and evaluates that "data" correspond to the starting point in which to base an observation, allocating them at a primary level; the information in turn corresponds to the product of contextualized and interpreted data and is therefore classified at a secondary level - thus understanding that data when decontextualized have no meaning at all. Turning to the concepts proposed by Becker, and considering Wiener's definitions, it is possible to trace important relations with the proposal of Rascão.

Wiener (1948) believed that data processed from an old piece of information could give rise to new information, which in turn, would have new values. This understanding that processed data can (have the potential to) give rise to precious information, is in

line with the idea initially exposed by Rascão when defining “information as potentiality” (2006, p.33), notable in the scope of the New Economy.

In the meantime, Ginman (1987) contributes to this observing that often a precious data is not valued for not being correctly and widely identified (visualized); this has prevented users and organizations from fully exploiting accessible and potentially valuable informational content. Some of these contents are generated from initiatives of digital users who, proceeding in search of information that meets their demand, leave informational “traces” often not perceived. This individual’s way of being, of acting or reacting - in the search for information - receives the name of informational behavior (SILVA; MOREIRA; SILVA, 2014).

Fialho and Andrade (2007) understand that informational behavior is a complex process, affected by external factors - environmental, demographic, economic, social, etc. It is observed, therefore, that data generated from actions of information users bring - implicitly - I get informational contents that reflect the context in which they were created. It is exactly this “potential information” load (RASCÃO, 2006, p.33) that gives the diluted data the probability/potential to generate new information.

4 I EXPLORING *BIG DATA* CONTENT

From the foregoing exposition, it is possible to understand, with greater clarity, the Market’s interest in unprecedented and timely information, treating it as a socio-economic element of powerful strategic value. In this context, many of the contents used have been obtained from the analysis of information in its grossest state: that of “data”.

However, the largest source of raw data for strategic information generation is non-standardized; a mass of virtual data currently estimated at more than 3 Zettabytes (3 billion Terabytes), an exponentially increasing volume - something in the order of 2.2 million Terabytes per day, worldwide (GARTNER, 2013). This immense informational content, available for exploration and analysis in digital networks, is called *Big Data*.

Big Data is the largest source of raw material for the generation of knowledge and innovation of the present time, formed by the profusion of structured and unstructured data available via electronic means and fed by the productions and publications of companies, governments and individuals, in the form of texts, photos, music, videos and social media contents.

According to Taurion (2014), the need for faster and faster data analysis tools is increasing, as the speed with which new data is created each day causes some data to become obsolete - for organizations - in a short time. Therefore, organizations have been keen to solve constant problems with storage space and data processing capacity; after all, for contemporary managers to acquire timely information is to acquire power (ROGERS, 2010).

Jewell et al (2014) classify the *Big Data* data into three fundamental types:

structured, semi-structured and unstructured. Currently, it is estimated that ninety percent of all data produced in the world is in unstructured form.

Big Data can also be defined through the so-called "V's" - attributes that describe and qualify this virtual mega-mass. Initially, it was considered that the concept of *Big Data* rested on the traditional 3 V's: *Volume*, *Velocity* and *Variety* (DUMBILL, 2012). However, two other attributes were suggested to integrate with the original ones: *Veracity* and *Value* (SAPORITO, 2013). *Validity* (KHAN, UDDIN, GUPTA, 2014, ELDER, 2014), *Visualization* (SEDDON; CURRIE, 2016), *Variability* and *Volatility* (OWAIS; HUSSEIN, 2016) appeared and were incorporated at the same time in which new observations emerged. Semantic studies (DUAN, 2015; SMITS, 2018) around data took into account the existence of another attribute little explored: *Vocabulary*. Finally, two other criteria were evaluated - *Venue* (BALAANAND et al 2017) and *Vulnerability* (ALLODI; MASSACCI, 2017; AYDINOGLU; BOVKIR, 2018) - concepts that, later, were integrated to the existing ones.

In a scenario where data is treated as the new natural resource of efficient organizations, to develop the ability to collect, store, classify, visualize and apply intelligence and innovation through analysis and interpretation is seen as a major organizational differential. This would be a reasonably simple task if the volume in *Big Data* were constant, uniform, clear, and debugged - which is not always the case.

Nowadays, the great difficulty for public and private organizations to adapt to this volatile and contingent scenario is precisely to deal with the characteristic attributes of Big Data. Paradoxically, these attributes - which express opportunities for organizations - also express major challenges.

For a better understanding of this concept (KHAN; UDDIN; GUPTA, 2014; OWAIS; HUSSEIN, 2016; BALAANAND et al., 2017; ALLODI; MASSACCI, 2017; AYDINOGLU; BOVKIR, 2018) each of the twelve attributes mentioned above will be listed and defined:

a) "Volume" attribute: it refers to the dimensional characteristic that *Big Data* occupies: a virtual mass of about 1 billion Terabytes composing the world Data platform. The challenge of contemporary organizations here is to store large volumes in limited virtual spaces;

b) "Velocity" attribute: it concerns both the speed with which data travels on this network and the speed with which new data are generated. The challenge of contemporary organizations here is to follow continuous and rapid production;

c) "Variety" attribute: should be appreciated as synonymous with heterogeneity, diversity: a whole formed by parts (data) of different natures, sizes, extensions and formats. The challenge of contemporary organizations here is to adapt to data processing with totally different natures;

d) "Veracity" attribute: refers to the importance of filtering the contents of *Big Data*, which are not always worthy of credibility. The challenge of contemporary

organizations here is to discern between what is true and what is false;

e) "Value" attribute: is related to the usefulness, the potential importance that *Big Data* data possess to add value to other contents. This attribute is directly related to the information concept conceptualized by Wiener (1948), which proposed that data processed from old information could be useful to give rise to new information of value. The challenge of contemporary organizations here is to evaluate between what can be useful and what can not;

f) "Visualization" attribute: also called *"Visibility"*, refers to the capacity that data - when properly processed - have to explain naturally hidden facts. This idea corroborates the epistemological concept of "information as potentiality" (RASCÃO, 2006, p.33), since it assumes that in *Big Data*, data - even if initially meaningless - constitute potential information. The challenge for contemporary organizations here is to have the right visualization tools for exploring raw content and thus to discover the hidden potential of the data;

g) "Volatility" attribute: unlike what used to happen two or three decades ago, today an organization can no longer keep the same data indefinitely; because they occupy a huge space, *Big Data* data need to be re-evaluated and discarded from time to time. Even if they are within "validity", data tend to be discarded after they have been thoroughly analyzed and applied. The challenge of the organizations here is to assertively select the data without applicability to be discarded;

h) "Vocabulary" attribute: considered by some as "Verbality", this attribute refers to the semantic-descriptive content stored in Big Data that, paradoxically, comes up against the difficulty of representation through metadata, indexing, ontologies and taxonomies. This thematic wealth of Big Data alternatively presents itself as an opportunity, and as a problem: an opportunity as it may consist of an immeasurable source of new ideas; a problem because there are enormous difficulties in defining and classifying unstructured data, which is why the concept of *"Vocabulary"* is sometimes referred to as *"Vagueness"*. The challenge for contemporary organizations here is to adequately label these hard-to-define content so they can store and retrieve them in the future;

i) "Venue" attribute: it talks about the situational nature of data in Big Data, which leads to reflect about the context of production, collection, analysis and application of these data. The challenge for contemporary organizations here is to understand the relationship between the scope of application of particular sets of data and the context from which they were drawn;

j) "Variability" attribute: also described as *"Viscosity"*, it stems from *"Velocity"*. It refers to the dynamic, fluid and unstable behavior of data; in Big Data, no perception is one hundred percent right, they are just odds. The challenge of contemporary organizations here is to manage contingencies and variations;

k) "Validity" attribute: is one of the most striking features in Big Data; it points to

the fact that the use of a given data has *timing*, that is, an optimal moment and a validity period. The challenge of contemporary organizations here is to seize opportunities in a timely manner;

l) "Vulnerability" attribute: while a culture based on data forecasting can be seen as an opportunity to add more security to decision making, it can be highly worrisome for managers to work with data that may be the gateway to contamination or breach of security protocols. The challenge of contemporary organizations here is to deal with all kinds of data without forgetting security.

It should be noted, therefore, that the complexity involving both the concept and the nature and applicability of *Big Data* often make it difficult to explore and apply. *Big Data* can be an excellent raw material for brainstorming, forecasting and intelligence generation, both in corporate environments and government agencies.

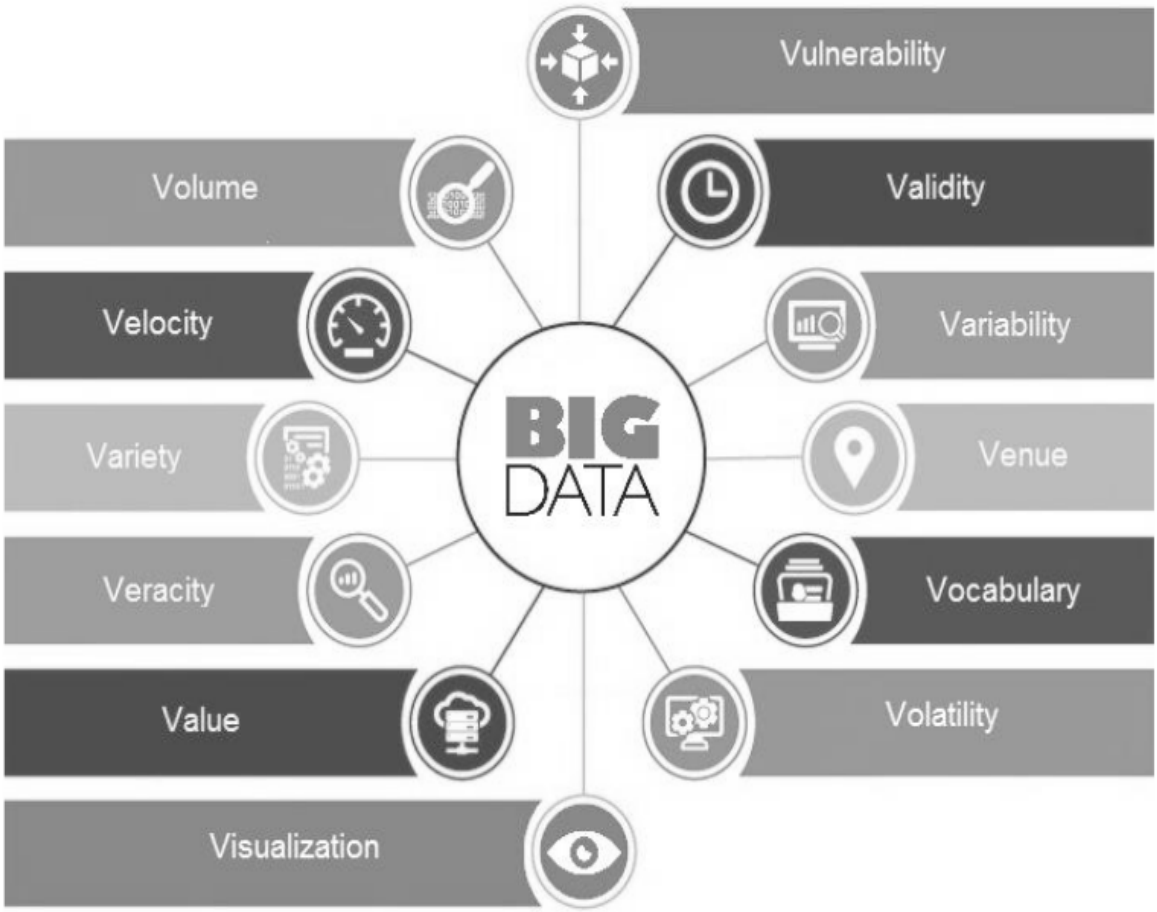

Figure 1 - The 12 V's in *Big Data*
Font: Adapted by author, 2018

However, for a successful application and exploitation of opportunities it is necessary – prior to giving any step related to the data - to have a clear idea about the questions that one wishes to answer. The visualization and analysis of data are intended to answer one or more problem questions, which must be well defined before the beginning of each process (TAURION, 2014).

In this way, there can be said to be four types of basic approaches *Big Data* can be worked with (MAFRA, 2017). They are: Descriptive, Diagnostic, Predictive and Prescriptive. Barga, Fontama and Hyong (2015) and Mafra (2017) argue that the Descriptive and Diagnostic approaches dissect the past, while the Predictive and Prescriptive ones project their gaze to the future.

Once the question (or questions) to be answered has been defined, the manager proceeds with the six steps for data analysis, to which Fayyad, Piatetsky-Shapiro and Smyth (1996) report: Selection and Extraction, Processing, Transformation, Data Mining (DM), Interpretation, Evaluation, and Application.

Therefore, it is observed that not only the origin of the data or the software technologies employed are important in the process of *Big Data* analysis, but also the methodology of study and the work of the researcher.

Finally, in order to obtain successful analyzes and responses from data, it is indispensable to define well what one wishes to respond to, to know the attributes of *Big Data* well, to understand the focuses of approach and to correctly apply the stages of evaluation and action.

5 I EXAMINING INFORMATION

For decades, there has been growing interest in content related to scientific and technological production - by researchers and those responsible for studying and creating public policies on security, development, health and education. Professionals related to research and development, both in the private and public sectors have sought in Information Science methods, tools and techniques to help them identify, select, process and extract scientific information for policies, academic assistance and decision support.

Thus universities, research centers and companies have focused on the study of productivity metrics, impact factor, collaboration studies and other investigations related to authors and contents in Science, using for this criteria and quali-quantitative processes explored in Scientometry.

It is called Scientometry the infometrics branch dedicated to the investigation of the various aspects related to the quantification and measurement of information in scientific production. Due to his contribution, Derek Solla Price was identified as the father of Scientometry. In his book *Little Science, Big Science*, published in 1963, Price had not yet coined the term *"Scientometry"*, however, that work was of enormous importance for its development (GARFIELD, 2007). It is observed that much of the methodology and fundamentals adopted by Price, and later by Garfield, for the development of the scientometric field derived from bibliometric theoretical bases - notably older ones. For this reason, many authors today recognize Scientometry as a branch of Bibliometry, field of library science (HOOD, WINSON, 2001).

Scientometry, therefore - as a bibliometric branch - is part of the Social Sciences

group dedicated to supporting the development of scientific, education, technology and innovation policies (BUFREM and PRATES, 2005). Among its applications, scientometric studies focus: the occurrence analysis, frequency and impact of publications of scientific content, the mapping of productivity by author and by institution, and the exploration of complex networks involving terms and relations of collaboration in Sciences.

Surveys such as these enable development agencies, education and research institutions, and Public Administration itself to improve its policies and strategies for hiring staff, priorities establishment and resources allocation.

6 I METHODOLOGICAL PROCEDURES

Both the study of *Big Data* (related to Technological Sciences), and Scientometry (belonging to the Applied Social Sciences) work with data analysis and statistical methods in its methodology. For this reason Bibliometry - from which the foundations for the development of Scientometry originated - was initially called Statistical Bibliography (GUEDES; BORSCHIVER, 2005). Thus, it can be said that both procedures - with data - apply to the exercise of techniques for the visualization and analysis of information, in order to originate quali-quantitative studies of descriptive nature.

Therefore, this research sought to demonstrate how data produced from informational behavior of digital users in searches can be related to scientific productions data and - at the same time – to the interests of the Brazilian Ministry of Health. For this purpose, it was developed a scientometric study of terms in the Health field using the *Web of Science* and *Scopus* databases. The timeline cut was limited between May 2013 and June 2018 in Brazil.

The highlighted terms in the scientometrics study were then analyzed under the user's perspective in their electronic demands for information. For this, the *Google Trends*© - a tool for visualization of search trends in *Google*© which operates through the data mining engine using available data in *Big Data* - was selected. Online software *Google Trends*© operates by ranking the most used terms in searches made by users in *Google*©, the most used metasearcher in Brazil. According to studies, 94.31% of the online researches in Brasil are made through the *Google Brasil*© website, followed by 2.05% through *Google*'s international website (SERASA EXPERIAN, 2015).

For experimentation and methodological demonstration, an exploratory-descriptive research was developed, using a quali-quantitative approach.

Finally, we will analyze the patterns of information evidenced among the three types of data sets - the first one, whose data were indexed from search attitudes in a popular search engine - the second, extracted from the two scientific databases, and the third, obtained from information sources of the Brazilian Government.

7 I RESULTS

Analyzing the graph (**Figure 2**) that reflects the data research indexed in the *Scopus* base, the expression *"zika virus"* has a direct and strong relationship with the word *"microcephaly"*, and also with other similar ones such as: *"neonatal"*, *"malformation"*, *"fetal death"*, *"pregnancy"*, *"zika infection"* and *"pernambuco"*.

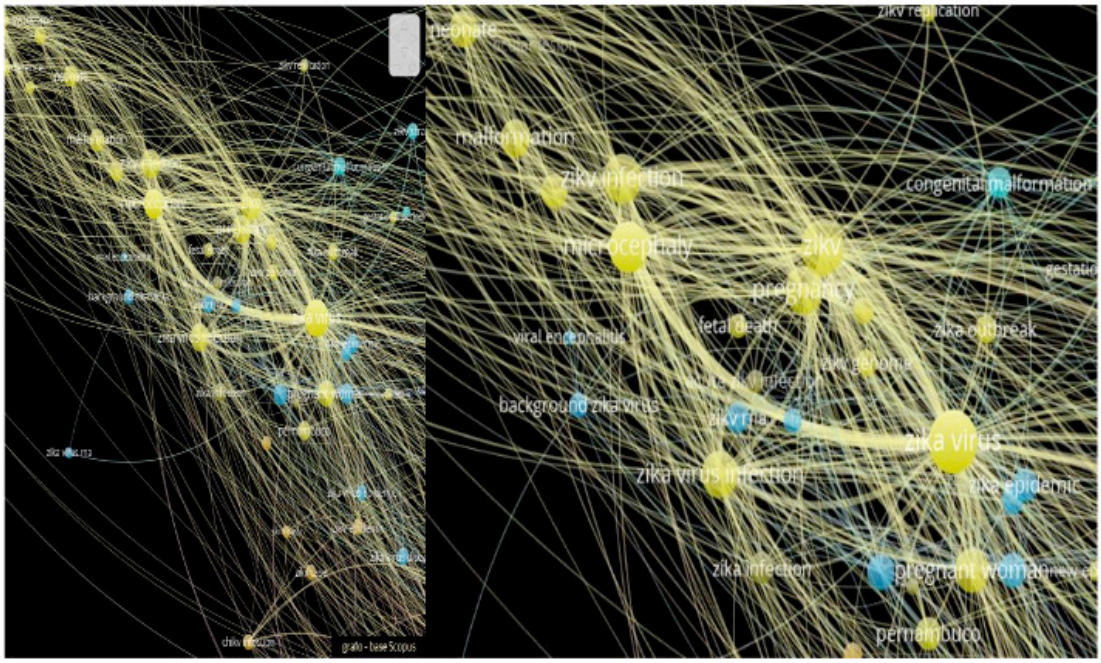

Figure 2 - VosVIEWER© software screenshot with *Scopus* base data graph.

Font: Research data, 2018

Analyzing the graph obtained by extracting data from the Web of Science database (**Figure 3**), the same perception is obtained: the node with the word *"congenital zika syndrome"* links directly and forcefully to the terms *"Pernambuco estate"*, *"Gestation"* and *"malformation"*. Attention is drawn to the presence of other weight terms, in the same image, such as *"infected mother"*, *"fetus"* and *"fetal death"*.

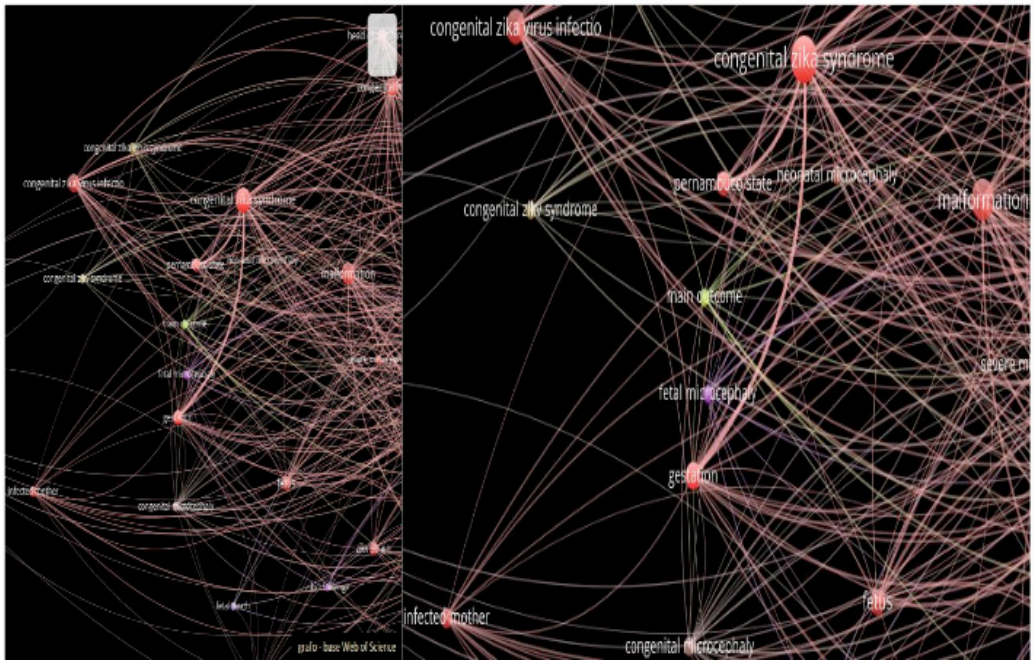

Figure 3 - VosVIEWER© software screenshot with *Web of Science* base data graph.

Font: Research data, 2018

It was observed that in the interstice from May 25 2013 to June 24 2018, the demands for information related to the terms *"Microcephaly"* (blue chart) and *"Zika"* (red chart) obtained maximum peaks of interest in exactly the same period (**Figure 4**): in the second week of December 2015 and in the second week of February 2016 in *Google Trends*©.

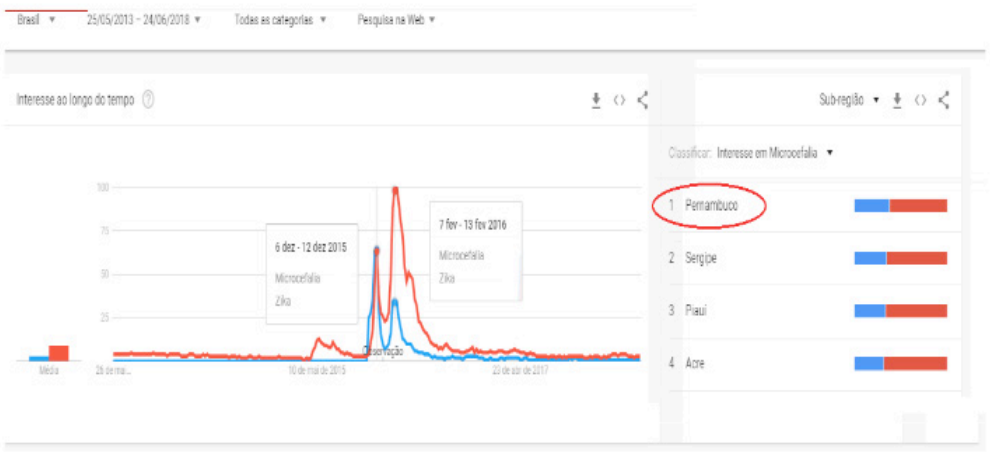

Figure 4 - Screenshot of the *Google Trends*© tool with analysis graphics for the *Big Data* content on the terms *"Zika"* and *"Microcephaly"* in the period from 05/25/2013 to 06/24/2018 in Brazil.

Font: Research data, 2018

It was verified that, among the sub-regions with the highest demand for search terms, the state of Pernambuco stands out leading the number of popular searches (**Figure 4**). It was also observed that, in the years 2015 and 2016, there were two peaks of searches related to the themes *"Microcephaly"* (blue line) and *"Zika"* (red line).

Associating this data with official data on cases of microcephaly, made available by the Ministry of Health in 2015 (**Table 1**), it was noted that Pernambuco stood out in this period as the state with the highest number of cases of microcephaly resulting from infection of newborns by Zika virus in the gestation period: 150 cases reported and 1,031 cases suspected only in this state. These numbers put Pernambuco with more than 37% of all cases registered in the country, and with more than twice as many cases as those reported in the state that appears in second place.

Distribuição dos casos suspeitos de microcefalia notificados à SVS/MS até a semana epidemiológica 50, por número de municípios e Unidade da Federação de residência, Brasil, 2015

Unidade da Federação	Total de municípios com casos notificados	Casos suspeitos de microcefalia relacionada ao vírus Zika		Óbitos suspeitos (n)
		n	%	
Centro-Oeste				
Distrito Federal	1	11	0,4	1
Goiás	12	40	1,44	0
Mato Grosso	10	78	2,8	0
Mato Grosso do Sul	2	3	0,11	0
Nordeste				
Alagoas	44	114	4,1	0
Bahia	64	271	9,74	10
Ceará	30	127	4,57	1
Maranhão	30	88	3,16	1
Paraíba	69	429	15,42	5
Pernambuco	150	1.031	37,06	3
Piauí	21	51	1,83	1
Rio Grande do Norte	42	154	5,54	10
Sergipe	40	136	4,89	5
Norte				
Pará	8	32	1,15	0
Tocantins	27	58	2,08	0
Sul				
Rio Grande do Sul	1	1	0,04	0
Sudeste				
Espírito Santo	10	18	0,65	0
São Paulo	6	6	0,22	0
Minas Gerais	33	52	1,87	1
Rio de Janeiro	18	82	2,95	2
Brasil	618	2.782	100	40

Table 1 - Reported and suspected cases of microcephaly in Brazil, in 2015.

Font: Brazil, 2015

It was also observed that, in the data mining research generated by searches for the term *"Microcephaly"*, in Brazil (**Figure 4**), the greater national demand for related information came from the population of Pernambuco.

Relating these cases of microcephaly to the official data of the Ministry of Health on notifications of contamination by the Zika virus between 2016 and 2018 (**Figure 5**), it is observed that the peak of cases in the three years occurs in the 7th Week of 2016, that is, around the third week of February 2016.

Attention is drawn to the fact that the second highest peak in information demand related to *"Zika"* and *"Microcephaly"* - recorded in the second week of February 2016 (**Figure 4**) - occurred about ten days before an outbreak of Zika, recorded in the middle of the third week of February 2016 (**Figure 5**).

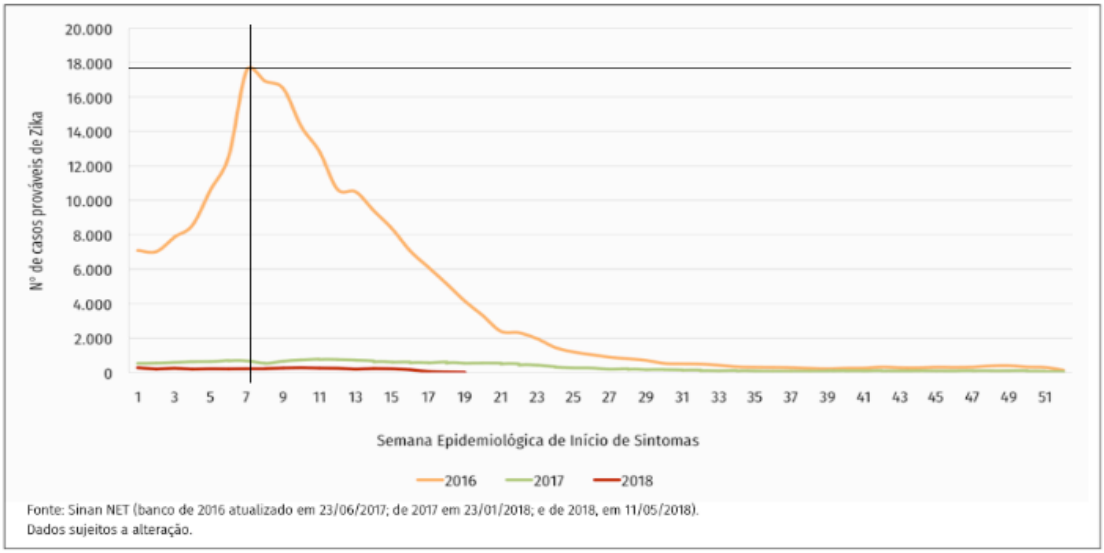

Figure 5 - Reported and suspected cases of acute illness by Zika virus, in Brazil, between 2016-2018.

Font: Sinan NET, 2018

It was observed, therefore, that there is a relation between peaks of volume of demand for information and peaks of endemic outbreaks. It was noted that the maximum peaks of information demand precede, in a short time, the maximum peaks of endemic crisis.

8 I CONSIDERATIONS

From the discussions developed, it is inferred that the existing synchronism among the peaks of data volumes in the *Google Trends*© chart, for the different terms, may be reflecting possible thematic associations established by digital users to define search criteria for information.

There was a strong relationship between the geographical location of the applicant user and the main geographic area from which the endemic outbreaks started, as well as a strong relationship of chronological proximity between the endemic outbreaks and the searches made for terms that represented them. These spatiotemporal relationships, which align informational demand behaviors (analyzed from data produced in *Big Data* environment) with official endemics data (published by the Brazilian Ministry of Health), suggest that the digital user acts - or reacts - driven by local and timely phenomena. In this context, aspects investigated here - such as locality (addressed under the term *"Venue"*) and timing (referred to as *"Validity"*), inherent in the processes of extraction and analysis of value data (referring to *"Value"*) for the discovery of useful information from *Big Data* - are remarkably instigating, revealing and relevant.

It was also evaluated that information provided by discursive genres related to the theme (scientific information offerings - dissected through scientometric study) derives from these perceptions of informational need; however, such studies are not always

timely in meeting population demands.

It was also observed that the peaks in demand for information occurred shortly before the endemic peaks recorded by the Ministry of Health and Secretariats. Thus, the peaks of searches for the term *"Zika"* with national demands headed by the state of Pernambuco (in the second week of December 2015 and in the second week of February 2016) were followed by peaks of Zika outbreak exactly in Pernambuco (in the third week of February 2016).

It was therefore assessed that informational behaviors of local users, manifested through informational attitudes with extraordinary volumes, seem to precede (and herald) critical peaks of impending social disorder.

It was found that data generated from informational "lanes" can reveal implicit information about the state, the location, and the political-economic-social moment in which the information user is. This occurs insofar as data generated, from an informational behavior, manifest the temporary and contextualized need of a specific group of regional users.

In the same way, it was possible to explore - methodologically - the intersections between scientometric research and data analysis from *Big Data* obtaining from them important contributions to the practice of extracting valuable information to assist organizations in the definition of scenarios and trends.

So, it is believed that judicious analyzes developed from metric studies - with data and information - can contribute to the construction of indicators that support research, policies and decision-making processes in organizations - public or private - inserted in the context of the Information and Knowledge Society.

REFERENCES

ALLODI, L; MASSACCI, F. Security events and vulnerability data for cybersecurity risk estimation. **Journal of Society for Risk Analysis**, n. 37, v. 8, Aug. 2017.

AYDINOGLU, A; BOVKIR, R. **Management of big geographic data for smart city applications**. FIG Congress, Istambul, 2018.

BALAANAND, M et al. A survey on *Big Data* with various v's on comparison of apache hadoop and apache spark. **Advances in Natural and Applied Sciences**, v.11, p. 362, 2017.

BARGA, R; FONTAMA, V; HYONG, T. **Predictive analytics with microsoft azure machine learning**. 2. ed. New York: Springer Sci + BM, New York, 2015.

BECKER, J. **Estatística básica:** transformando dados. Porto Alegre: Bookman, 2015.

BRASIL. **Sinan NET** – Casos notificados e casos suspeitos de doença aguda por Zika vírus, no Barsil, entre 2016 e 2018. Brasília, DF, 2018.

BRASIL, Secretaria de Vigilância em Saúde - Ministério da Saúde. **Boletim Epidemiológico**, Brasília, DF, v. 46, n° 6, 2015. ISSN 2358-9450.

BUFREM, L; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 9-25, maio/ago. 2005.

DUAN, H et al. Distributed in-memory vocabulary tree for real-time retrieval of *Big Data* images. **Ad Hoc Networks**, v.35, p.137-148, dez. 2015.

DUMBILL, E. **Planning for *Big Data***. Massachusetts: O'Reilly Media Inc., 2012.

ELDER, J. **Evaluate the validity of your discovery with target shuffling**. [S.l.]: ER, 2014.

FAYYAD, U, PIATETSKY-SHAPIRO, G; SMYTH. P. **From data mining to knowledge discovery in databases**. v.17, n.3, Providence, 1996.

FIALHO, J; ANDRADE, M. Comportamento informacional de crianças e adolescentes: uma revisão da literatura estrangeira. **Ciência da Informação**, Brasília, v. 36, n. 1, p. 20-34, 2007.

GARFIELD, E. From the science of science to scientometrics visualizing the history of science with HistCite Software. In: **ISSI**, 11., 2007, Madrid. Anais... Madrid, 25 jun. 2007.

GARTNER, Inc. **IT glossary**. New York: Gartner, [2013].

GINMAN, M. Information Culture and Business performance. **IATUL Quarterly**, v. 2, 1987.

GUEDES, V; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: **ENANCIB**, 6., 2005, Salvador: UFBA, 2005.

HOOD, W; WILSON, C. The literature of bibliometrics, scientometrics, and informetrics. **Scientometrics**, v. 52, n. 5. 2001.

JEWELL, D. et al. **Performance and capacity implications for *Big Data***. [S.l.: s.n.], 2014.

KHAN, Ali-ud-din; UDDIN, Muhammad Fahim; GUPTA, Navarun, 2014. **Seven v's of *Big Data*: understanding *Big Data* to extract value**. Conference of the American Society for Engineering Education, I, Bridgeport: IEEE, 2014.

MAFRA, F. **Business intelligence com sql server 2012 na prática**. [S.l.]: Alpha, 2017.

NEELY, A; GREGORY, M; PLATTS, K. Performance measurement system design: a literature review and research agenda. **IJO & PM**, [S.l.], v. 15, n. 4, p. 80–116. 1995.

OWAIS, S; HUSSEIN, N. Extract five categories CPIVW from the 9v's characteristics of the *Big Data*. **IJACSA**, v.7, n.3, 2016.

RASCÃO, J. **Da gestão estratégica à gestão estratégica da informação**: como aumentar o tempo disponível para a tomada de decisão estratégica. Rio de Janeiro: E-papers, 2006.

ROGERS, S. Information is power. **The Guardian**, London, 2010.

SAPORITO, P. The 5 v's of *Big Data*: value and veracity join three more crucial attributes that carriers should consider when developing a *Big Data* vision. **Best's Review**, v. 114, 2013.

SCHWAB, K. **A Quarta Revolução Industrial**. São Paulo: Edipro, 2016.

SEDDON, J; CURRIE, W. A model for unpacking *Big Data* analytics in high-frequency trading. **Journal of Business Research**, v. 70, p.300-307, 2016.

SERASA EXPERIAN. **Google Brasil tem 94,31% de participação nas buscas em dezembro,**

segundo Hitwise. São Carlos: Serasa Experian, 2015.

SHANNON, C; WAEVER, W. The mathematical theory of communication. **Bell System Technical Journal**. 1948. p.379-423.

SILVA, C; MOREIRA, J. **Comportamento informacional vs comunicação**: aplicação de modelos em contextos multidisciplinares. [S.l.: s.n.], 2014.

SMITS, G. et al. A soft computing approach to *Big Data* summarization. **FSets & S**, 2018.

TARAPANOFF, K. (Org.). **Inteligência organizacional e competitiva**. Brasília: UNB, 2001.

TAURION, C. Entre os v's do *Big Data*, velocidade cresce em importância. **CIO**, [S.l.], 2014.

VALENTIM, M. Informação e conhecimento no contexto de ambientes organizacionais. In: ____. **Gestão, mediação e uso da informação**. São Paulo: Cultura Acadêmica, 2010.

WIENER, N. **Cybernetics:** or the control and communication in the animal and the machine. Massachusetts: Institute of Technology, 1948.

SOBRE O ORGANIZADOR

Ernane Rosa Martins - Doutorado em andamento em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa, em Porto/Portugal. Mestre em Engenharia de Produção e Sistemas, possui Pós-Graduação em Tecnologia em Gestão da Informação, Graduação em Ciência da Computação e Graduação em Sistemas de Informação. Professor de Informática no Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia) ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador do Núcleo de Inovação, Tecnologia e Educação (NITE), certificado pelo IFG no CNPq. ORCID: https://orcid.org/0000-0002-1543-1108

ÍNDICE REMISSIVO

Agência Brasileira do ISBN
ISBN 978-85-7247-488-7