

BIG DATA EXPLORATION AND INFORMATION METRICS EXAMS: INTERSECTIONS AND ANALYSIS TO SUPPORT TO DECISION-MAKING

Rafael Barcellos Gomes
Master's degree student in Information Science at PPGCI – IBICT / UFRJ
(rafaelbarcellos@ufrj.br)

Vânia Lisboa da Silveira Guedes
Post-Doctorate in Linguistics and Post-Graduation in Information Science – UFRJ
(vanialisboa@facc.ufrj.br)

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. It dissects contemporary concept of information and its strategic importance to support to decision-making in inserted organizations 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. It was recorded patterns of convergence among the observations, noting that some occurrence of endemic outbreaks peaks went preceded by peaks of informational demands for terms related to infectious agents involved in endemics. It was also observed that scientific informational offers that satisfy the public demands for related information were late when compared to the maximum demands peaks. 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.

Resumo

Propõe expor aspectos teóricos e práticos em estudos métricos da informação científica, associando-os à exploração de *Big Data* para deslindamento de tendências e cenários em Saúde Pública, no Brasil. Neste íterim, 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. Observou-se, também, que ofertas informacionais científicas que satisfazem às demandas públicas por informação relacionada são, às vezes, tardias quando comparadas aos picos máximos de demandas. Por fim, destaca a relevância da metodologia analítica – baseada na

visualização de dados – para o desenvolvimento de estudos que apoiem 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.

Resumen

Propone exponer aspectos teóricos y prácticos en estudios métricos de la información científica, asociándolos a la explotación de *Big Data* para deslindamiento de tendencias y escenarios en Salud Pública, en Brasil. En este ínterin, diseña el concepto contemporáneo de información y su importancia estratégica para apoyar la toma de decisiones en organizaciones insertadas en el contexto de la Sociedad de la Información y del Conocimiento, examinando la importancia de los estudios cienciométricos y del análisis de datos producidos en red en el desarrollo de procesos de descubrimiento de información de valor. Para ello, investiga posibles relaciones entre datos de *Big Data*, datos oficiales publicados por el Ministerio de Salud de Brasil y órganos correlatos y, datos técnicos extraídos de publicaciones científicas indexadas en bases de investigación. Se registraron patrones de convergencia entre las observaciones tomadas, constatando que algunos picos de ocurrencia de brotes endémicos son precedidos por picos de demandas informacionales por términos relacionados a los agentes infecciosos involucrados en las endemias. Se observó, también, que ofertas informativas científicas que satisfacen a las demandas públicas por información relacionada son, a veces, tardías cuando comparadas a los picos máximos de demandas. Por último, destaca la relevancia de la metodología analítica - basada en la visualización de datos - para el desarrollo de estudios que apoyen procesos decisorios y estrategias organizativas en políticas de Salud, delineando escenarios y tendencias para soporte a las administraciones pública y privada.

Palabras clave: *Big data*, minería de datos, análisis cienciométrico, gestión estratégica, proceso de toma de decisiones.

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 Brazil, in early 1980s.

This moment was characterized, above all, by the migration of a characteristically industrial economic paradigm - based on manufacturing and industrial assets, to 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 aborded 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 spatio-temporal context in which they are acting, transposing the challenges and taking advantage of opportunities (NEELY; GREGORY; PLATTS, 1995). This discernment and perception by the organization are born of the generation of knowledge, product of this element that moves the globalized world: information.

1 Conceptualization and contextualization

In the context of the post-industrial society, information becomes more important in the process of capital generation, a fact by which it comes to be compared to a "[...] economic asset (commodity) [...]" and to a "[...] product/merchandise [...]" (TARAPANOFF, 2001, p.91) that are essential to the creation of knowledge and intelligence, whether for leisure, for work or for teaching, Research and Development. From this unusual socio-economic context emerged a new understanding of effective business management, whose decision-making processes are based on the search 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 given exists as a potential source of information and not as information itself).

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

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

It is observed, therefore, that in Information Management in the context of the New Economy it is possible to recognize the information under various aspects. 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.

To this period Valentim (2010) gives the name of 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 fewer structured on tangible products (SCHWAB, 2016) and, increasingly, in the provision of services: online, ontime, fulltime.

2 Data and information

Becker (2015) analyzes a series of statisticians and information theorists and evaluates that "data" correspond to the starting point 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 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 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 exploit accessible and potentially valuable informational content.

Some of these contents are generated from initiatives of digital users that, starting in search of information that meets their demand, leave informational "traces" often not perceived. To this way of being, to act or to react of the individual - in the search for information - gives 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.

3 Exploring *Big Data*

From the foregoing, 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*¹.

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 the unstructured form.

Big Data can also be defined through 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 the efficient organizations, to develop the ability to collect, store, classify, visualize and apply intelligence

¹ *Big Data* can be understood as a set of high-volume, high-volume, large-scale digital information assets that require cost-effective and innovative forms of information processing to support the improvement of ideas and decision-making in organizations. (GARTNER, 2013).

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.

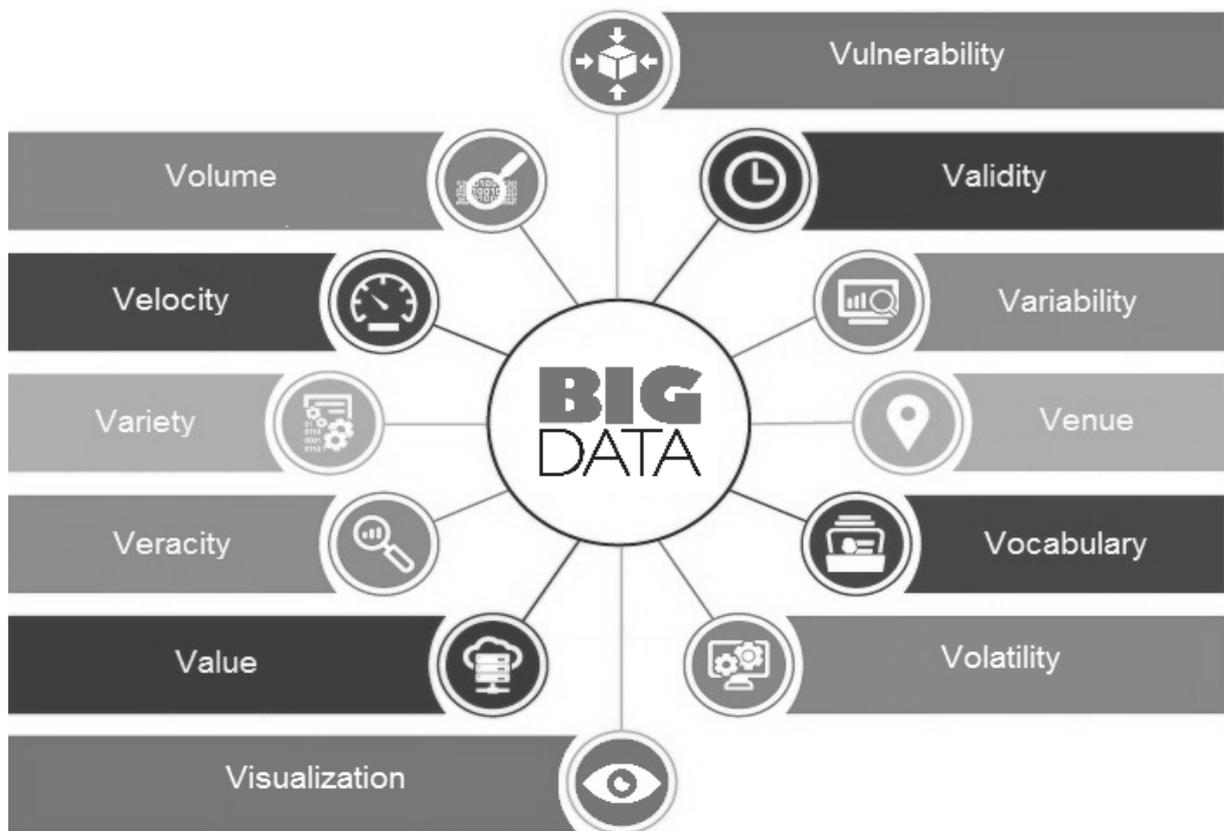


Figure 1 - The 12 V's in *Big Data*
Font: Adapted by the author, 2018

Nowadays, the great challenge of organizations, public and private to adapt to a 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. Opportunities because the immeasurable supply, diversity and richness of *Big Data* data can be seen as an inexhaustible resource for adding value to new management models. Challenge because this feature, although rich, is generally found in the raw, latent and heterogeneous state.

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, forecast and intelligence generation, both in corporate environments and government agencies.

However, for a successful application and exploitation of opportunities it is necessary - before beginning any step related to the data - to have clear 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, it can be said that there are four types of basic approaches with which *Big Data* can be worked (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 approaches 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 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.

4 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 tools and techniques - to help them identify, select, process and extract scientific information for policies, academic assistance and decision support - in Information Science methods.

Thus, universities, research centers, companies, and information units have focused on the study of productivity metrics, impact factor, calculation of average life, collaboration studies, representativeness, use of information and other investigations related 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, a field of study pertaining to 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.

5 Methodological procedures

Both the *Big Data* study (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 -

"malformation". Attention is drawn to the presence of other weight terms, in the same image, such as "infected mother", "fetus" and "fetal death".

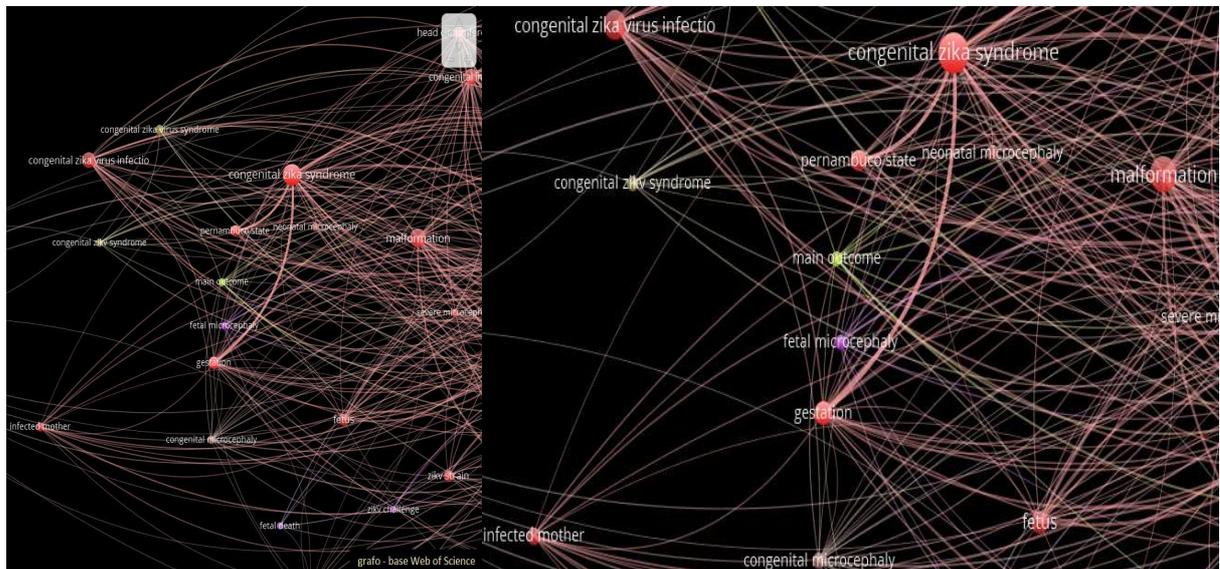


Figure 3 - VosVIEWER© software screen with *Web of Science* base data graph.
Font: Research data, 2018

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*©.

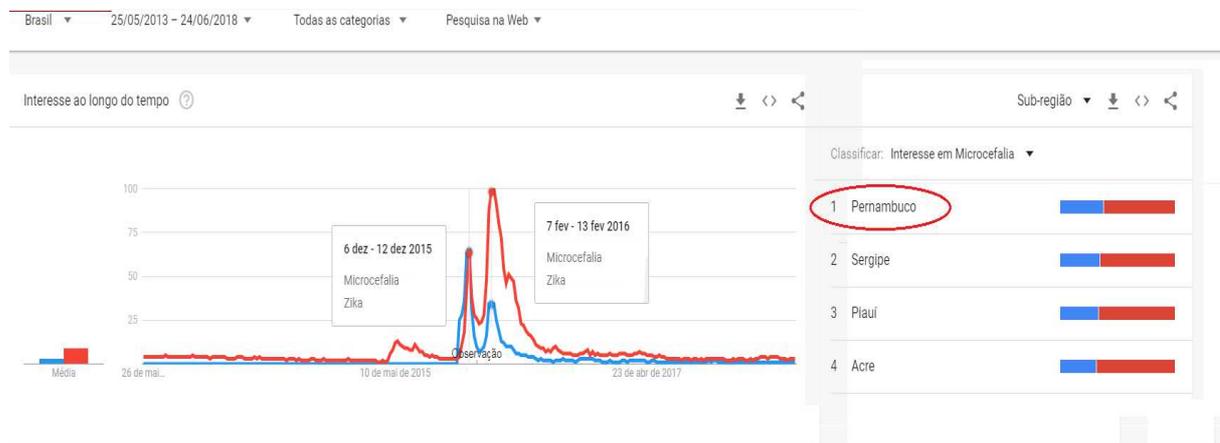


Figure 4 - Screen of the *Google Trends*© tool with *Big Data* content analysis graphics to the terms "Zika" and "Microcephaly" in the period from 05/25/2013 to 06/24/2018 in Brazil.
Font: Research data, 2018

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

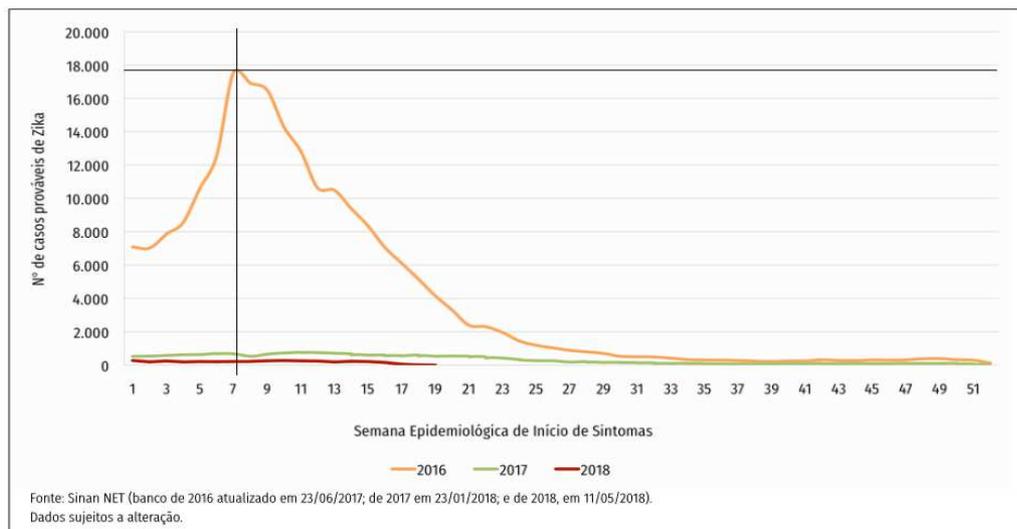
suspected only in this state. These numbers put Pernambuco with more than 37% of all cases registered in the country, and with more than twice as many cases as those reported in the state that appears in second place.

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 - Cases 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.



Casos prováveis de doença aguda pelo vírus Zika, por semana epidemiológica de início de sintomas, Brasil, 2017 e 2018

Figure 5 - Cases reported and suspected cases of acute illness by Zika virus, in Brazil, between 2016 and 2018.
Font: Sinan NET, 2018

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**).

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.

7 Considerations

From the discussions developed, it is inferred that the existing synchronism between 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 performed by 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 not always be 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.

BALANAND, 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 Brasil, 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. 83 p.

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; SILVA, J. **Comportamento informacional versus 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.