Big data in Information Science literature

Luiz Carlos Felipe de Oliveira[1] and Beatriz Valadares Cendón[2]

[1] Master in Information Science. Graduate Program in Knowledge Management and Organization, School of Information Science, Federal University of Minas Gerais.
[2] Professor, Graduate Program in Knowledge Management and Organization, School of Information Science, Federal University of Minas Gerais.

Abstract: Big data, the massive amount of structured and unstructured data, too large or too complex to be analyzed by conventional database software, signals for new methods to record, organize, analyze and provide access to information. To help understand how information science (IS) has incorporated big data as a topic of study, this paper researched the scientific output of the area indexed in main referential databases. The paper reviews the emergence of large volumes of data, the definition, dimensions and components of big data, and the related concepts of e-science and data science. In this bibliometric study, data collection was done through searches in the main referential IS databases to identify the scientific literature on big data. Authors and their programs/ departments, institutions and countries were identified. Then, the works were categorized through content analysis of their titles, keywords and abstracts. The results present a map of how IS scientific output has approached big data, showing topics of research and a ranking of universities/organizations, graduate programs/ departments, authors and countries that produced the articles on big data indexed in the LISA and LISTA databases.

Key words: Big Data. Information Science. E-science. Data Science. Data scientist. Information scientist.

1. Introduction

Burke (1992) demonstrates that the recording of information is a constant in the evolutionary process of humanity. These initiatives can be seen from rock paintings, hieroglyphs and parchments to the present times of computers and of the web. Recently, with technological developments, there has been a growing volume of data available which comes from all societal contexts, generated by web applications, by governments, by scientific research, by publications or by organizations. Input could be, for example, from social networks chatters, web server logs, traffic flow sensors, satellite imagery, broadcast audio streams, banking transactions, MP3s of rock music, the content of web pages, scans of

Received: 8.7.2019 Accepted: 3.7.2020 ISSN 2241-1925

government documents, GPS trails, financial market data, among others. This massive amount of structured and unstructured data, in sets too large or too complex to be handled by conventional database software, has been called big data, a form of large-scale information recording.

Due to these new developments, we are, at the present moment, immersed in a new information explosion, comparable to the one that called for the development of information science (IS) in the post-war period. At that time, "The problem was (and basically still is) the massive task of making more accessible, a growing body of knowledge". (Saracenic, 1996, p. 42). Today we face a similar situation with the additional challenge that now information does not always have definite format and volumes are much greater.

The presence of Information Science (IS) in this new informational context is justified and necessary. It is important for IS to reaffirm itself within the informational scenario, in which big data is present, besides defining how the information scientist can act in this context where he can occupy new spaces, offer new services and seek new skills (Ribeiro, 2014; Souza, 2006). In the context of e-science, Hey and Hey (2006) point out the emergence of greater and new challenges for information professionals. As large amounts of data that originate from e-science become available, scientists will need the support of information professionals to manage, research, preserve and store these new data resources.

Within this context and the challenges presented by big data for the area, the current research sought to verify how the big data theme has been approached in the scientific literature of information science.

2. Definition, dimensions and composition of big data

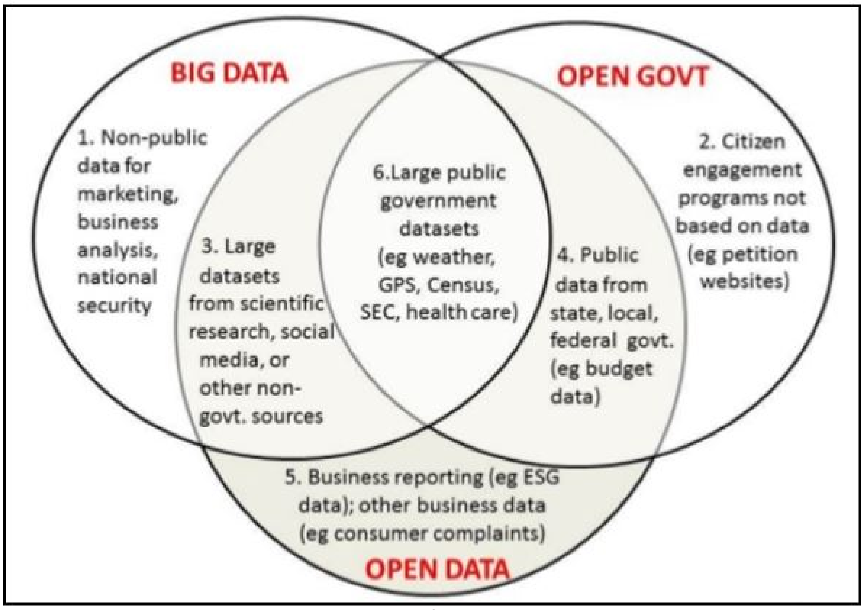
*2.1 B**ig Data***

By incorporating the big data phenomenon among its concerns, information science addresses a difficult issue. There is no definite authorship or single definition for big data. In a context in which both the corporate segment and the academy are involved (Azam, 2014; Taurion, 2012a), scholars try to define and analyze the big data phenomenon. Dumbill (2012, p.3) defines big data as:

> … data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of your database architectures. To gain value from this data, you must choose an alternative way to process it.

Taurion (2012b) and others argue that the big data phenomenon has 5 dimensions: Volume, Validity, Velocity, Veracity and Value. However, new dimensions and new attempts to define the phenomenon are constantly emerging (Kitchin and McArdle, 2016). In addition to the traditional big data V's approach, we highlight the view of Gurin (2013), which presents a model for understanding the composition of big data (Figure 1). The author points out that not all private data and not all open data make up the big data phenomenon, where open data can be defined as data that “can be freely used, modified, and shared by anyone for any purpose” (The open definition, 2020).

Figure 1 - Limits and classes of big data



Source: Adapted from Gurin (2013).

As shown in Figure 1, Gurin (2013) considers 3 limits and classes for big data which are detailed below:

- *Big data that are not open and non-governmental*: this category includes information of great relevance to business, such as consumption habits, information related to patients in hospitals, information related to credit card transactions, etc. The use of this information to obtain competitive advantages is a possibility;
- *Big data that are open and non-governmental*: data originated from scientific projects that share data, for example, large volumes of data from astronomical observations or from large biomedical projects such as the Human Genome project. Also included in this category are data maintained or analyzed by government agencies, which do not characterize them as governmental data as well as data obtained from social media;

- *Big data that are open and governmental:* Government-generated data sets in large volumes may be the most impact producing of all. Governments have the capacity and resources to gather large volumes of data. By making them open, one gets great economic benefits. Data such as census, social security, trade balance, public health, etc. are included here. This category becomes relevant because of the volume, robustness, and significance of its data.

2.2 ***Data Science, e-Science and its relationship with big data***

This section defines e-science and data science and relates these two concepts to big data. Science has become a great generator of data due to the use of digital equipment and accessories of high complexity and precision, often connected to the web, contributing to the big data phenomenon. Scientific research increasingly demands the analysis of large volumes of data, in addition to the scientific methods required for ongoing studies. Considering these characteristics, Bohle (2013) defines e-science as the application of computer technology to help modern scientific investigation, including data preparation, experimentations, collection, dissemination of results, storage and accessibility of material generated through scientific process.

On the other hand, data science is defined as:

> ...a set of fundamental principles that support and guide the principled extraction of information and knowledge from data. Possibly the most closely related concept to data science is data mining—the actual extraction of knowledge from data via technologies that incorporate these principles. (Provost e Fawcet, 2013, p. 2).

However, according to Provost and Fawcet (2013), data science is much more than just data mining. The field draws on statistics, causal analysis, data visualization methods, as well as on intuition, creativity and knowledge of the subject that is under analysis, to provide the data scientist a framework “to systematically treat problems of extracting useful knowledge from data” (Provost e Fawcet, 2013, p. 2).

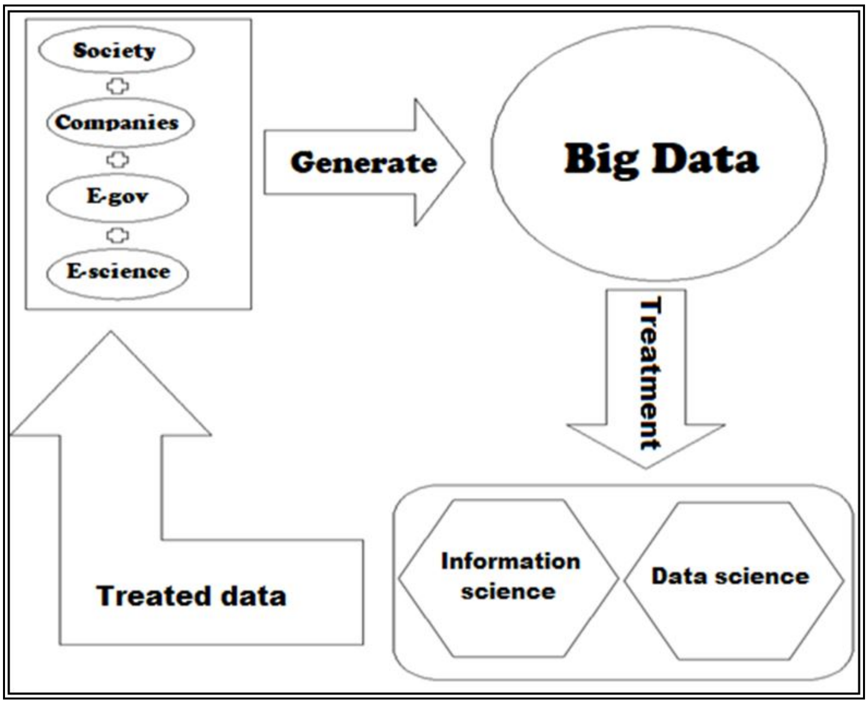
The terms "data science" and "data scientist" have aroused interest in the corporate segment and its activities, which involve the statistical treatment of large volumes of data. Ferreira (2015, our translation) complements:

> In practice, the data scientist is the professional responsible for the development of mathematical algorithms and the alignment of these to the business for better performance of the companies. With big data platforms, companies are able to predict or understand some points such as: market trends, repercussions of some fact/ marketing campaign and consumer behavior.

For the data scientist, techniques required include signal processing, probabilistic models, machine learning, data mining, pattern recognition and learning, data warehousing, data engineering, uncertainty models, data compression, artificial intelligence, and computing high performance, which comes from areas such as mathematics, statistics and computing (Noughabi et al., 2017).

Figure 2 identifies the relationship between big data, data science and e-science, mapped from authors who conceptualize big data, such as Taurion (2012a) and Dumbill et. al. (2012), and authors who point out the need for new skills to information professionals such as Stryker (2014) and Garritano and Carlson (2009).

Figure 2 - Relationship between big data, e-science and data science



Source: Elaborated by the authors (2017).

As shown in Figure 2, the open and private data originated from society, companies, governments and science, constitute the big data phenomenon. These large-volume data can be addressed by data science for extraction of information and knowledge as exemplified above or, in face of the need for classification, dissemination and treatment, can be addressed by information science. Subsequently, after these data is analyzed, classified, and disseminated it will be available for use.

3. Methodology

The data reported here were extracted from articles on big data in the Library and Information Science (LISA) and Library, Information Science and Technology (LISTA) databases. The choice of these two sources was due to their focus and to the representativeness of the publications they index to the area of IS as well as the observation of a great overlap in the content of specialized IS databases. The option of utilizing articles only was due to the fact that they should represent the most consolidated and highest quality research.

In the databases, the search for articles related to the big data phenomenon, in the databases used the following search strings, present in the title, abstract, and keywords fields of the articles: "big data" OR "e-science" OR "data science". The use of terms related to big data intended to broaden the search to ensure all works related to the topic would be retrieved.
Data was collected between July and August 2016.

The publications identified were dated from 2010 to 2016. In the metadata of each article, the study obtained the following information: first author, title, bibliographic reference, abstract, institution of affiliation, title of the journal and country. Subsequently, the authors' programs/ departments of origin were identified using the author and university fields as a search key at the universities' web sites, blogs or social networks.

In the next step, through systematic reading of the title, abstract, and descriptors, the analysis of how the term big data was addressed in the text was performed. Based on this analysis, each article was classified by the first author of the paper according to categories that emerged from the data. These are described in Table 1 below.

Table 1- Approaches to big data

Category	**Code**	**Meaning**
Research Object	RO	The article reports on research having big data as object of research. The RO articles are further qualified as: IR: Research is on information retrieval; DIS: Research is on dissemination of information; IP: Research is on treatment of information;

		INF: Research is on infrastructure.
Secondary Subject	SS	The article reports on research or event, where big data is a secondary subject.
Reflection	RE	The author makes a reflection on the big data theme.

Source: Elaborated by the authors (2017).

The categories in Table 1 are further explained below:

- *Research Object (RO):* RO refers to the research article in which big data is the main object of research. This classification is subdivided into four subcategories, here called "qualifications":
  - *Information Retrieval (IR):* articles received this qualification when big data research is on information retrieval. For example, research on retrieving historical data, such as behavior of a population or cycle of a particular disease;
  - *Dissemination of Information (DIS):* articles received this qualification when big data research is on dissemination of information. This subcategory includes articles that portray the creation of applications, frameworks, or algorithms to make information available*;*
  - *Information Processing (IP):* articles received this code when big data research is on treatment of information. The creation of applications, algorithms, and frameworks for the purpose of information processing as well as actions for obtaining, processing, and transforming information are included here;
  - *Infrastructure (INF):* articles received this qualification, when big data research is on the infrastructure needed to process big data.
- *Secondary Subject (SS):* article reports on event or research not directly related to big data; however, the subject of big data is present in the article.
- *Reflection (RE):* This code includes articles which presented reflections on the impact of the big data phenomenon in society, in professions and professionals, for example.

In a second stage of the categorization of the works, a copy of the database without codes was submitted to another peer from the area of information science, who repeated the procedure of categorization described in the previous item. Next, a comparison was made between the researcher's and the peer's categorizations of the articles, in order to evaluate the degree of coherence between them. For the articles where the categories attributed by the two coders were different, final categorization was obtained through discussion and consensus. The articles classified as RO were still submitted to a third pair, the second author of this paper, and their codes underwent further adjustments.

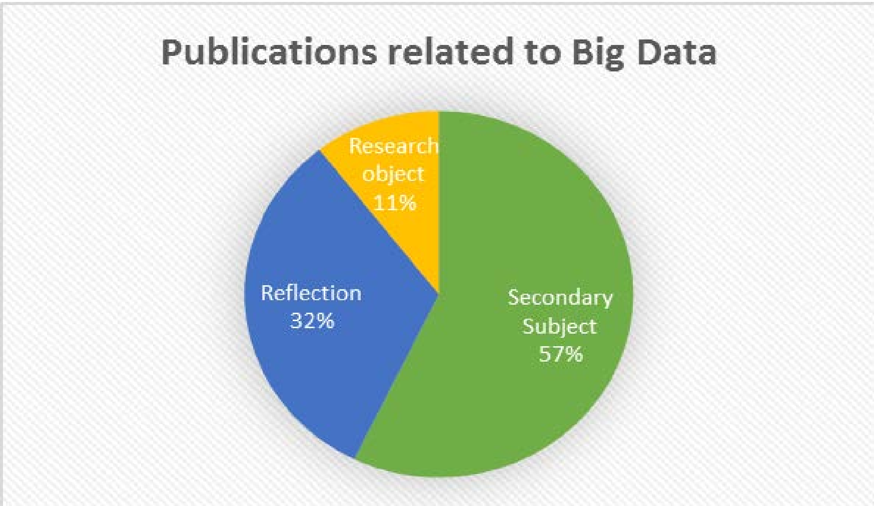
It should be pointed out that as the categorization of each article was based on subject metadata, that is, the title, the keywords, and the abstract, rather than on the full text, the resulting categorization could be only indicative.

4. Results

Below we present some of the research results, obtained from the analysis of the LISA and LISTA databases articles on big data. For complete results, the reader should access the full text of the dissertation research available at http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/ECIP-ARUJ8D.

Graph 1 shows the distribution of the articles analyzed among the 3 categories. Of the articles, 144 (32%)were classified as RE, 255 (57%) as SS, and 47 (11%) as research on big data .

Graph 1 - Categorization of publications related to big data indexed in the LISA and LISTA Bases - 2010 to 2016



Source: Research data (2017).

Table 2 presents the number of publications by country in each category. Of note is the volume of publications by the United States (206 publications), by the United Kingdom (54 articles), and by China (37 articles), totaling 297 articles (67%). China was the country that most published articles categorized as RO, with ten articles published, followed by the United States with nine articles. Next, are Germany and Italy with the publication of six articles each.

Table 2 – Number of publications by country and category

Country	Category			Total
	SS	RO	RE	
United States	116	9	81	206
UK	33	3	18	54
China	21	10	6	37
Canada	9	1	11	21
South Corea	11	1	3	15
Germany	4	6	3	13
Italy	5	6	2	13
Spain	5	1	3	9
Netherlands	7	-	2	9
Australia	4	1	3	8
Denmark	2	1	5	8
France	5	1	-	6
New Zealand	5	1	-	6
India	5	-	-	5
Malta	-	-	3	3
Poland	3	-	-	3
Switzerland	2	1	-	3
Algeria		1	1	2
Austria	1	1	-	2
Greece	1	-	1	2
Hungary	2	-	-	2
Iran	1	1	-	2
Ireland	2	-	-	2
Israel	2	-	-	2
Malaysia	1	-	1	2
Norway	2	-	-	2
Chile	-	1	-	1
Finland	1	-	-	1

Japan	-	1	-	1
Latvia	1	-	-	1
Morocco	-	-	1	1
Palestine	1	-	-	1
Portugal	1	-	-	1
Sweden	1	-	-	1
Turkey	1	-	-	1
Total	**255**	**47**	**144**	**446**

Source: Research data.

A total of 384 programs/ departments distributed in 35 countries published the 446 articles on big data. Table 3 identifies the number of programs/ departments per country. In the top three positions in number of programs/ departments, are the United States, United Kingdom and China. It is of notice the high dispersion of publications among the programs/ departments with an average of 1,16 publications per program/ department.

Table 3 - Number of programs/ departments and publications by country

Country	Total		Country	Total	
	Prog.	N. Art.		Prog.	N. Art.
United States	189	206	Algeria	2	2
UK	48	54	Austria	2	2
China	36	37	Greece	2	2
Canada	21	21	Hungary	2	2
South Korea	15	15	Iran	2	2
Germany	13	13	Ireland	2	2
Italy	13	13	Israel	2	2
Spain	9	9	Malaysia	2	2
Netherlands	9	9	Norway	2	2
Australia	8	8	Chile	1	1
Denmark	8	8	Finland	1	1
France	6	6	Japan	1	1
New Zealand	5	6	Latvia	1	1
India	5	5	Morocco	1	1

Malta	1	3	Palestine	1	1
Poland	3	3	Portugal	1	1
Switzerland	3	3	Sweden	1	1
			Turkey	1	1
Total				**384**	**446**

Source: Research data.

The publications emanated from 307 universities/ institutions. Table 4 shows the universities/ institutions with the largest number of publications, where independent writer indicates authors without affiliation.

Table 4- Universities / Institutions with most publications on big data

University/Institution	Country	N. Art
University of California	United States	17
Independent writer	United States (8) United Kingdom (1)	9
University of Southampton	United Kingdom	8
University of North Carolina	United States	7
George Mason University	United States	6
CILIP: The Library and Information Association	United Kingdom	5
Johns Hopkins University	United States	4
Initiatives.ORG	New Zealand	4
Editorial Services at Bank of Philadelphia	United States	4
University of Minnesota	United States	4
Drexel University	United States	4
Information Today	United States	4
University of Southern California	United States	4
Indiana University	United States	4
University of Wisconsin	United States	4

Source: Research data.

As shown in Table 4, the three most prominent universities in number of publications, totaling 32 articles, are:

- University of California, USA, with 17 articles;
- University of Southampton, UK, with eight articles;
- University of North Carolina, USA, with seven articles.

Table 5 shows the number of articles published by each of the programs/ departments at University of California, as well as their areas. For the definition of the area of knowledge, the title of the program or department of origin of the article was used. A total of 11 programs/ departments at the University of California published on big data, of which the Institute for Research on Labor and Employment, stands out with seven articles. This is a multidisciplinary institute, focused on work and employment issues. Other than that, there is a dispersion of the publications among ten programs/ departments, with one article each. The programs/ departments belong to a variety of areas such as computer science, library science, information science, anthropology, communication, biology, and medicine.

Table 5 - Programs and publications at the University of California

Program/Department	Area	N. of Art.
Institute for Research on Labor and Employment, University of California, Berkeley	Multidisciplinary	7
Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California Irvine	Computer Science	1
UC Santa Cruz Genomics Institute, University of California, Santa Cruz	Biology	1
Department of Communication, University of California, Davis	Communication	1
Innovation and Industry Alliances, University of California San Diego	Medicine	1
Department of Anthropology University of California	Anthropology	1
Department of Family and Preventive Medicine University of California San Diego	Medicine	1
Department of Communication, University of California	Communication	1
Anthropology (Archaeology) Department, University of	Anthrop	1

California	ology	
Research Data Curation Program, University of California San Diego Library	Library Science	1
School of Information, University of California, Berkeley, CA, USA	Information Science	1
	Total	**17**

Source: Elaborated by the authors.

Table 6 shows the number of articles published by each of the programs/ departments at University of Southampton, UK, as well as their areas. At this university, eight articles were published by three programs, all related computer science. The "Computational Visiting Professor, University of Southampton, UK" a program for visiting professors at the university, stands out with 7 articles.

Table 6 - Programs and publications of the University of Southampton

Program/Department	Area	N. of Articles
University of Southampton, UK Computational Visiting Professor	Computer Science	7
Electronics and Computer Science, University of Southampton, UK	Computer Science	1
Total		**8**

Source: Research data.

Table 7 shows the number of articles published by each of the programs/ departments at University of North Carolina, United States, as well as their areas:

Table 7 - Programs and publications of the University of North Carolina

Program	Area	N. of Art.
Bryan School of Business and Economics, The University of North Carolina at Greensboro	Administration	1

Bryan School of Business and Economics, The University of North Carolina at Greensboro	Administration	2
College of Arts & Sciences; Connections (Mathematics); Mathematicians	Mathematics	1
Digital Innovation Lab, University of North Carolina at Chapel Hill,	Computer Science	1
School of Information and Library Science, University of North Carolina, Chapel Hill	Information science	2
Total		**7**

Source: Research data.

As seen on Tables 5, 6 and 7, the University of California and the University of North Carolina show greater diversity among the areas of the programs which published on big data, while programs at the University of Southampton show only programs related to computing. Of special interest for the current research is the fact that although the articles analyzed were indexed in IS databases, only a minority of them (3 out of 32 articles on big data published by the three most productive universities) emanated from programs in the area of library and/or information science.

Table 8 presents the distribution of articles classified as RO, grouped by area of knowledge. For the definition of the area, the title of the program or department of origin of the article was used. The 47 articles classified as RO, belonged to fourteen areas of knowledge. Only 4 of the 47 research articles emanated from an IS program/ department.

Table 8 - Areas that published articles classified as Research Object

Area	N. of Art.	Area	N. of Art.
Computer Science	26	Language	1
Medicine	4	Statistics	1
Library Science	3	Mathematic	1
Economics	2	Bioengineering	1
Engineering	2	Communication	1
Information Science	1	Arts	1
Multidisciplinary	2	Law	1

		Total	**47**

Source: Research data.

Table 9 shows that 396 (first) authors published the 446 articles, showing that 28 of them published more than one article. The average number of publications per (first) author is 1,13.

Table 9 –Number of first authors and number of articles by country
Source: Elaborated by the author.

Country	N. Authors	N. Art.	Country	N. Authors	N. Art.
United States	**191**	**206**	**Hungary**	**2**	**2**
United Kingdom	**49**	**54**	**Iran**	**2**	**2**
China	**36**	**37**	**Ireland**	**2**	**2**
Canada	**21**	**21**	**Israel**	**2**	**2**
South Korea	**15**	**15**	**Malaysia**	**2**	**2**
Germany	**13**	**13**	**Norway**	**2**	**2**
Italy	**12**	**13**	**Chile**	**1**	**1**
Spain	**9**	**9**	**Finland**	**1**	**1**
Netherlands	**9**	**9**	**Japan**	**1**	**1**
Australia	**8**	**8**	**Latvia**	**1**	**1**
Denmark	**7**	**8**	**Malta**	**1**	**3**
France	**6**	**6**	**Morocco**	**1**	**1**
India	**5**	**5**	**Palestine**	**1**	**1**
New Zealand	**4**	**6**	**Portugal**	**1**	**1**
Poland	**3**	**3**	**Sweden**	**1**	**1**
Algeria	**2**	**2**	**Switzerland**	**1**	**3**
Austria	**2**	**2**	**Turkey**	**1**	**1**
Greece	**2**	**2**			
			Total	**396**	**446**

Source: Research data.

The data evidences the dispersion of the scientific output among the authors, and the lack of connection of the majority of the most prolific authors with programs in information science. Fourteen authors published two articles, ten published three articles, two authors published four articles, and one author published seven articles. It is important to highlight that the most prolific author, with 7 articles (Terence K Huwe, Director of Library and Information Resources, Institute for Research on Labor and Employment, UCLA) is connected to the field of IS although his program/ department is not. Other most productive authors are: Robert Boeri, IEEE.org (4 articles), Barbara Brynko, Editorial Services of the Federal Reserve Bank of Philadelphia (4 articles); Robert Springer, no affiliation (3 articles); Marydee Ojala, Websearch University (3 articles); and Stephen Arnold, Tandem School of Engineering (3 articles). Another point that should be underlined is that the works by the authors mentioned above as being the most productive in the topic of big data are not research articles and were categorized as RE or SS.

The data also showed that fifteen companies, distributed in six countries published 20 articles, corresponding to 6.51% of the institutions and 4.47% of articles published. Microsoft was the company that most published, with three articles.

5. Final considerations

Recently, with technological developments, there has been a growing volume of data available. The large amounts of structured and unstructured data sets, that are too large or too complex to be dealt with by traditional data processing software, have been called big data. Challenges posed by big data include capturing, storing, analyzing, searching, sharing, transferring, visualizing, querying, updating, and information privacy. There is a call for a role for IS and the IS professionals in addressing the problems brought by big data. In this context, it is relevant to understand how IS has positioned itself regarding this phenomenon. To shed light on this issue, this article surveyed and analyzed the IS literature about big data to understand how the big data theme has been approached in the scientific articles indexed in the two major IS databases, LISA and LISTA.The results of this research show that 446 articles related to the big data phenomenon were published in the LISA and LISTA databases between 2010 and 2016, by 35 countries, 307 universities/ institutions, 396 first authors, and 384 programs/Departments. In number of publications, the United States, the United Kingdom and China stand out with 206, 54 and 37 publications, respectively. The three most productive universities, totaling 32 publications, are in the US and UK. Of the 446 articles, 255 (52%) were classified as Secondary Subject (articles were big data is not the main concern), 144 (32%) as Reflections about the theme, and 47 (11%) as Research on big data. These 47 articles belonged to fourteen areas of knowledge.

It is of notice that little amount of research apparently has been conducted in the area of IS as the 47 research articles represent only 11% of the publications. It is also relevant to the objectives of the current study that only 4 (8,5%) of these research articles originated in Library and Information science programs/ departments. The remaining 43 research works (91,5%) originated in programs belonging to fields such as computer science, medicine, economics, engineering, language, statistics, mathematic, bioengineering, communication and arts.

Another point that stands out in the research results is the high degree of dispersion in the number of publications by area of knowledge, by program/ departments, by countries and by (first) authors. This could indicate few departments or authors of the articles indexed in the LISA and LISTA databases are concentrating efforts on the subject. The data from the current study seem to indicate the need of more research on the topic of big data by Information Science programs. The predominance of articles emanating from programs and researchers from other sciences also raises questions. While this can be interpreted as a demonstration of information science's interdisciplinarity, it also may be indicative that the "natives" of information science are not instrumented to face the current informational context. If this is true, information science programs should adapt their curriculum to address the demand for a new profile of the information scientist by the new information context. In face of the space not occupied by information science, other sciences advance in seeking for solutions, giving rise to new players in the informational scene.

References

Azam, A., (2014). The first rule of data science. *Berkeley Science Review*. 27 Apr. Available at: <http://berkeleysciencereview.com/article/first-rule-data-science/>. Access in: 21 Apr. 2016.

Bohle, Shannon, (2013). What is e-science and how should it be managed? *Scientific and Medical Libraries,* June 12. Available at: http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed/. Access in:10 Apr. 2020.

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

Dumbill, Edd, (2012). What is dig data. In: O'Reilly Radar Team. *Big Data Now: Current Perspectives* from O'Reilly Media. Sebastopol: O'Reilly Media, Inc. Available at: <http://www.oreilly.com/data/free/files/big-data-now-2012.pdf>. Access in: 25 May 2017.

Ferreira, M., (2015). As 11 profissões que estarão em alta no Brasil em 2015. *Época Negócios*, 07 Jan. Available at: <https://epocanegocios.globo.com/Carreira/noticia/2014/12/10-profissoes-que-estarao-em-alta-no-brasil-em-2015.html>. Access in: 30 jun. 2016.

Garritano, J.; Carlson, J. R., (2009). *A Subject Librarian's Guide to Collaborating on E-science Projects.* Purdue University. Available at: <http://docs.lib.purdue.edu/lib_research/140/>. Access in: 31 Mar. 2016.

Gurin, J., (2013). Big data vs open data: mapping it out. *Open Data Now*, Oct. 2013. Available at: <http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/#.V_1d-Y-cF9A>. Access in: 26 Mar. 2017.

Provost, Foster, Fawcett, T., (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* v.1, n.1. Available at: http://doi.org/10.1089/big.2013.1508.

Hey, T.; Hey, J., (2006). E-Science and its implications for the library community. *Library Hi Tech*, v. 24, n. 4, p. 515-28. Available at: <https://www.deepdyve.com/lp/emerald-publishing/e-science-and-its-implications-for-the-library-community-jcYvTlSio8>. Access in: 05 Jul. 2016.Kitchin, Rob; McArdle, Gavin (5 January 2016). "What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets". *Big Data & Society*. 3 (1): 205395171663113. doi:10.1177/2053951716631130. ISSN 2053-9517.

Noughabi, Elham Akhond Zadeh, Raahemi, Bijan; Albadvi, Amir; Far, Behrouz H. (Eds) (2017). Handbook of Research on Data Science for Effective Healthcare Practice and Administration. Hershey, PA: IGI Global.

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

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

Souza, F. C. A., (2006). Formação acadêmica de bibliotecários e cientistas da informação e sua visibilidade, identidade e reconhecimento social no Brasil. *Informação & Sociedade: Estudos*, v. 16, n. 1, p. 23-34. Available at: <http://www.biblionline.ufpb.br/ojs/index.php/ies/article/viewFile/439/1491>. Access in: 19 Mar. 2016.

Stryker, C., (2014). *Big data will impact every part of your life*. Video. Available at: <https://www.youtube.com/watch?v=0Q3sRSUYmys>. Access in: 02 Feb. 2016.

Taurion, C., (2012a). O caos conceitual e os 5V´s do big data. *Revista CIO*, 11 May. Avaliable at: <http://cio.uol.com.br/opiniao/2012/05/11/o-caos-conceitual-e-os-5vs-do-big-data/.>. Access in: 13 Jan. 2016.

Taurion, C., (2012b). *Você realmente sabe o que é big data*? Available at: <https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce_realmente_sabe_o_que_e_big_data?lang=en>. Access in: 05 Jul. 2016.

The open definition. Avaliable at: https://opendefinition.org. Access in: 10 Apr. 2020.