Article

em questão
E-ISSN 1808-5245

Intersections between Information Science and Big Data

Joel Perozo-Vasquez [I]
https://orcid.org/0000-0002-1223-8927

Ana Depizzolatti [II]
https://orcid.org/0009-0003-8499-3686

Nathália Berger Werlang [I]
https://orcid.org/0000-0003-0172-6025

[I] Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil
[II] Universidade de São Paulo, São Paulo, SP, Brazil

Abstract: The aim of this work is to identify the interrelationships between Big Data and Information Science based on a literature review of scientific communications indexed in the Scopus and Web of Science databases. Methodologically, the generic search expression was ("big data" AND "information science"), and 233 records were retrieved. After careful filtering, the object of study was defined by a selection of 51 articles. This set was quantitatively characterized through bibliometric analysis to identify the most relevant elements for thematic classification, with ten main topics being identified. As qualitative results, methodological contributions of Information Science were verified in research involving Big Data across various fields, including Information Science methods and tools, medicine and public health, Information Science research, professional training, urban planning, and public policies. Additionally, ethical discussions, theoretical research, and practical applications were explored. It was concluded that Information Science has many interrelationships with Big Data and contributes significantly to the research and utilization of this type of analysis, while also strengthening the interdisciplinarity between Information Science and Data Science.

Keywords: information science; big data; data science; interdisciplinarity

1 Introduction

Over the past few decades, organizations have undergone significant transformations in how they make decisions and create value through the capture, storage, and analysis of data. With each passing year, they need to improve their method of analyzing and processing the large volume of data to which they have access to increase the value generation for the organization (Gobble, 2013; Alharthi; Krotov; Bowman, 2017).

Rapid technological advancement, which includes changes in equipment, software, and the Internet, has made data digital and brought about new forms of use and value to Big Data, which has been transforming various fields of knowledge (Ouyang; Wu; Huang, 2018).

The large volume of data generated (structured, semi-structured, unstructured, and with different complexity levels) requires agile computational analysis with the purpose of identifying patterns, trends, and associations related to various aspects of human, cybernetic, or machine agents. Big Data comprises specific computational approaches, strategies, and techniques for processing and managing such data (Dutra; Matias, 2014; NIH, 2022).

As the area of knowledge that "[…] investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability" (Borko, 1968, p. 3), Information Science (IS) shares with Data Science (DS) objectives, strategies, procedures, and analysis instruments of Big Data (Big Data Analytics), contemplating techniques and tools such as data mining, machine learning, and text analysis, among others (Dutra; Matias, 2014; De Mauro; Greco; Grimaldi, 2016; NIH, 2022).

Zhang, Wolfram and Ma (2023) pointed to the emergence of studies that approximate the use of Big Data in IS. To the authors, any empirical research is based on objectives, research methods, data, and their analysis. The method comprises scripts, techniques, and procedures used in a study to collect, process, and analyze data, produce results, and reach a conclusion that allows the researcher to address the research objectives. Most of the time, a large data set is

not self-explanatory and needs analysis and processing methodologies, especially when using diversified data.

The emerging trend and the use of Big Data have implications for empirical research based on data in different areas, especially in IS. In light of this, Zhang, Wolfram and Ma (2023) reiterated the need to analyze and expand the scope of research on Big Data in IS.

Given the above, this study aims to identify the relationships between Big Data and Information Science (IS) by conducting a literature review based on indexed scientific communications in Scopus and Web of Science (WoS) databases.

By analyzing 51 selected articles from Scopus and Web of Science databases, this study proposes a qualitative-quantitative analysis with the goal of identifying the contributions of Information Science (IS) to the field of Big Data. The analysis goes beyond merely identifying areas of overlap between the two fields, and instead explores conceptual and methodological advancements emerging from their interaction, highlighting the need for consolidation and expansion of the field.

Previous studies have investigated research trends on the theme in question: (a) relationships between BD and DS: Fuller (2015), Song and Zhu (2016), Cao (2017), Stockmann (2018), Daniel (2019); (b) relationships between DS and IS: Murtagh and Devlin (2018), Virkus and Garoufallou (2019, 2020), You, Joo and Katsurai (2022); and (c) relationships between IS and BD: Golub and Hansson (2017), Santos-D'Amorim *et al.* (2020), Garoufallou and Gaitanou (2021), Wang, Wang and Huang (2021), Lima (2022), Nascimento, Miguel and Costa (2022), Guimarães, Rocha and Mugnaini (2023), Zhang, Wolfram and Ma (2023). The present research differs from all these, first of all, due to the thematic scope, upon limiting the study exclusively to the discipline of IS relative to Big Data; second, due to the international coverage of the studied databases; third, due to the observation of publications until the end of 2022; and finally, due to the methodological procedures for data collection, processing, and analysis, explained below.

2 The approach between Big Data and Information Science

Soon after the coining of the concept of Big Data by Toffler (1990), different industries began to feel the need to use data to improve value delivery to their customers. The significant increase in the use of data and their processing exceeded the storage capacity of organizations (Wang *et al.*, 2019).

A 2008 publication in Nature of an album entitled "Big Data" discussed the challenges and opportunities of Big Data in different contexts, such as technology, economics, biomedicine, and others, which stimulated the research and use of Big Data in different sectors and rendered this topic urgent (Lynch, 2008).

Over the years, it became evident that Big Data directly impacts sectors such as medicine and public health, economics and business, and computer science (Boyd; Crawford, 2012; Einav; Levin, 2014; Khoury; Ioannidis, 2014).

Big Data may be defined as the processing and analysis of large amounts of data to support decision-making processes in organizational contexts (Allam; Dhunny, 2019; Gupta *et al.*, 2018; Davenport, 2014).

However, the Big Data phenomenon has been analyzed from two different contexts: the first focusing on the computational and technological infrastructure (Dong; Yang, 2018), and the other associated with the capacities to deal with the challenges of management and incorporation of Big Data into organizational processes (Gupta; George, 2016).

Studies have focused primarily on investigating the first context, which refers to infrastructure, due to the increase in the volume of available data created from different sources such as sensors, GPS devices, and the intensive use of technological resources, *e.g.*, devices such as smartphones connected to social networks, among others.

Seven data characteristics were defined to identify the main challenges in dealing with Big Data analysis (Mikalef *et al.*, 2018; Sivarajah *et al.*, 2017; Chen *et al.*, 2013; Barnaghi, Sheth; Henson, 2013).

The first characteristic is volume, which refers to the amount of data due to exponential growth, leading to data storage, acquisition, and processing

challenges. This induces the need for investments in technological equipment (George *et al.*, 2016; Barnaghi; Sheth; Henson, 2013).

The second attribute is variety, which refers to the significant heterogeneity of data – audio, video, text, and image – which generates a challenge in analyzing different forms of data (Constantiou; Kallinikos, 2015; Chen *et al.*, 2013).

The third characteristic is the speed at which the data flow is created, which in many cases requires real-time analysis and increases the ability of the data to become obsolete (Sivarajah *et al.*, 2017; George *et al.*, 2016).

The fourth characteristic is truthfulness, which concerns the quality of the data, *i.e.*, the accuracy and reliability of the data and their sources. The fifth attribute is visualization, which refers to the ability to present data in ways that render them meaningful (Seddon; Currie, 2017).

The sixth attribute is the value of the data extracted from Big Data to an end user and their contribution to improving the outcome of organizations (Sivarajah *et al.*, 2017; Gandomi; Haider, 2015). Finally, the seventh characteristic is variability, which refers to the continuous and rapid variation in the meaning and interpretation of the data (Seddon; Currie, 2017; Sivarajah *et al.*, 2017).

Upon evaluating the main characteristics of the data, the challenges and opportunities for their use in IS are made evident (Boyd; Crawford, 2012). The existing concepts about IS are based on the stages of selection, organization, retrieval, dissemination, and use of information in organizations (Borko, 1968; Choo, 2003; Davenport; Prusak, 1998).

In recent years, the increasing access to information and interactions between information systems and information users have attracted many studies. All user accesses and searches are recorded in the browsing history, generating digital traces of the navigations in web portals. Moreover, interactions on social networks, media platforms, and others are archived in data reservoirs. Thus, Big Data, which comprises this growing number of user-generated data and the forms necessary for their analysis, is an important source of analysis for information science (Zhang; Wolfram; Ma, 2023).

3 Methodological procedures

As for its nature, this research is basic; regarding its objectives, it is exploratory and descriptive. Concerning its procedures, it is bibliographic research, as it is a literature scoping review. As for the research approach, it was mixed since a bibliometric study was carried out for the quantitative analysis, whereas thematic and content analyses of the selected articles were carried out through the qualitative approach. The steps listed and described below were performed to carry out this study. Where applicable, the simplifications of the 'Prisma 2020 Protocol checklist' and the '2009 article selection flow' were used (Moher *et al.*, 2009; Page *et al.*, 2021).

The Scopus (Elsevier) and Web of Science (WoS, Clarivate) databases were chosen for being the most comprehensive relative to internationality and the number of indexed journals and deposited articles. The selection of these databases is justified by their international scope and the large number of journals and articles indexed, which consolidate them as the main sources of scientific information for the field of Information Science and related areas. The choice of these databases was motivated by the need to access the most comprehensive possible corpus, ensuring the representativeness of research on the relationships between Information Science and Big Data. Additionally, the majority of relevant research in the areas of interest are indexed in these databases, which guarantees the robustness of the results without compromising quality due to the recollection.

The generic search expression was ("big data" AND "information science"), adapted for each database according to the syntactic peculiarities of each search system. The search was conducted using only English terms, as the language is widely recognized as a common tongue for scientific communication. In international databases such as Scopus and Web of Science, much of the scientific production, even in other languages, typically has title, abstract, and keywords in English.

The following inclusion criteria were used to filter the search results: Open Science: only publications with open access to the full text; Publication Type: complete research articles and reviews published in journals and peer-reviewed;

Time frame: articles published from any year, without restriction, until December 31, 2022.

The decision to analyze only Open-Access (OA) articles was motivated by the need to ensure accessibility and transparency in the research process. By prioritizing OA articles, we aimed to facilitate the dissemination of research findings, promote the sustainability of scientific research, and foster an open and collaborative environment. Furthermore, the growing trend of OA publishing suggests a promising future direction for future research.

The time frame for data collection, which includes publications from any year up to December 31, 2022, was established to analyze the development of the field over time without an initial specific temporal restriction. The goal is to chart the historical trajectory of the interactions between Information Science and Big Data from the earliest works to the most recent studies. The choice of the final date (December 31, 2022) is justified by the need to ensure that the latest available productions at the time of data collection were included. This data collection was performed in April 2023. The exclusion criteria applied to abstracts, expanded abstracts, critiques, editorials, and any published scientific communication that did not meet the inclusion criteria.

The results of the bibliographic records retrieved from each database were the following: 156 from Scopus and 77 from WoS, totaling 233 records. The results were merged, eliminating 55 duplicates, leaving 178 articles for the selection stage, in which 86 were excluded after screening titles and keywords, 40 after screening the abstracts, and one because it was retracted (Jiang, 2022, 2023), leaving a total of 51 articles for review and analysis, the results of which are described in the following section, with 2014 being the year of the first publication observed in the selected sample.

4 Results

In this section, the results of the qualitative analysis of the topics and content of the 51 selected articles are presented, initially discussing the quantitative characterization of said set of selected articles through the application of

bibliometric analysis techniques and subsequently presenting the results of the qualitative analysis of the content of said articles.

In the temporal dimension, the distribution of articles published annually was the following: 2014 (two articles), 2015 (four), 2016 (four), 2017 (four), 2018 (six), 2019 (eight), 2020 (ten), 2021 (eight), 2022 (eight), which denotes a linear trend of growth given by the mathematical expression [$f(x) = 0.767*x$; $R^2 = 0.774$], with a concentration of 61.54% of the articles published in the last four years, against 38.46% in the five years before

Analyzing the production by the authors, the sample only partially met Lotka's law due to the small sample size of 51 articles analyzed. Only Wang-Y participated in two articles, with the other 95 identified authors producing only one publication each, as observed in Table 1.

Table 1 - Lotka's Law: Number of Publications by Author

Publications	Lotka's Law	Lotka's Law (%)	Sample authors	Sample authors (%)
1	95	65.07%	95	98.96%
2	24	16.44%	1	1.04%
3	11	7.53%	0	0.00%
4	6	4.11%	0	0.00%
5	4	2.74%	0	0.00%
6	3	2.05%	0	0.00%
7	2	1.37%	0	0.00%
8	1	0.68%	0	0.00%

Source: Research data.

Lotka's Law describes the productivity distribution of authors in a given research area and predicts that most authors will publish only one article while a small number of authors will account for a large proportion of publications. For the sample analyzed, the expected values would be those presented in the columns "Lotka's Law" and "Lotka's Law (%)". However, the values obtained (presented in the columns "Sample authors" and "Sample authors (%)") show almost all authors participating in only one publication.

Bradford's Law analyzes the number of publications per journal. The 51 analyzed articles were published by 47 journals; the set of 17 articles that make up the first third of the total was published by a group of 12 journals (25.53% of the journals). Journal "Heliyon" published three articles, while "Perspectivas em

Ciência da Informação", "Journal of Information Science", and "Em Questão" published two articles each. In the other thirds of the sample, the average production was one article per journal.

As shown in Table 2, 25% of the journals concentrated 32% of the articles published. The first tertile presented a proportion of 17 articles for 12 journals, whereas the second and third tertiles presented a proportion of one article per journal.

Table 2 - Bradford's Law: Publications by Journal in tertiles

Group	Number of Articles	Perc. of Articles	Number of Journals	Perc. of Journals
1	17	32.69%	12	25.53%
2	17	32.69%	17	36.17%
3		34.62%	18	38.30%
Totals	52	100.00%	47	100.00%

Source: Research data.

By the applied lexicographic analysis, Zipf's Law was naturally met and represented by the mathematical expression [f(x) 226x^-0.694; $R^2 = 0.979$]. After eliminating from the count the terms "big", "data", "information", and "science", the terms most frequently identified in the titles and abstracts of the articles were the following: datum (309 occurrences), research (94), analysis (56), health (48), paper (42), development (42), field (41), approach (39), technology (34), knowledge (34), base (33), method (32), process (31), and system (30), thus being the most representative terms of the corpus (Table 3).

Table 3 - Most used terms in absolute and relative frequencies

Terms	Freq.	Perc.%	Terms	Freq.	Perc. %	Terms	Freq.	Perc. %
datum	309	13.08%	health	48	2.03%	knowledge	34	1.44%
big	146	6.18%	paper	42	1.78%	base	33	1.40%
information	141	5.97%	development	42	1.78%	method	32	1.35%
science	116	4.91%	field	41	1.74%	process	31	1.31%
research	94	3.98%	approach	39	1.65%	system	30	1.27%
analysis	56	2.37%	technology	34	1.44%			

Source: Research data.

The quantitative information resulting from the bibliometric analysis facilitates contextual profiling for classification. Ten main topics were identified, listed in descending order by their frequency: IS methods and tools (16 articles), medicine and public health (12), IS research (seven), professional training (four), urban planning and public policies (three), surveillance and privacy (three), organizational management (three), ethics (two), information security (one), and Natural Language Processing (NLP, one). Together, the first four topics represented 75% of the publications (30.77%, 23.08%, 13.46%, and 7.69%, respectively).

Table 4 - Classification of thematic areas of the publication

Topic	Publications	Perc. %
IS Methods and Tools	16	30.77%
Medicine and Public Health	12	23.08%
Scientific Research and Communication	7	13.46%
Education and Professional Training	4	7.69%
Surveillance and Privacy	3	5.77%
Urban Planning	3	5.77%
Organizational Management	3	5.77%
Ethics	2	3.85%
Information Security	1	1.92%
Natural Language Processing (Nlp)	1	1.92%

Source: Research data.

Based on the thematic classification carried out, the articles were analyzed qualitatively. The results are presented in the following section.

5 Qualitative analysis

Studies in different areas of knowledge, namely materials science (Ghiringhelli *et al.*, 2017), natural sciences (Gryk; Ludascher, 2017), sustainability science (Ahmad *et al.*, 2016), climate resilience (Harwell *et al.*, 2019), history (Volodin, 2019), journalism (Garcia Jimenez; Catalina Garcia, 2018; Opdahl; Tessem, 2021), sports (Wei, 2022), education, Internet of Things (IoT), industry,

innovation, and logistics, used IS methods and tools (Xu; Shi, 2015) to explore Big Data in their specific fields.

Liang (2022), Wang, Zhao and Li (2022), and Yin and Tsai (2021) researched the use of Big Data in education: outcome-based education, "educational Big Data", and virtual reality classrooms, respectively.

Sødring, Reinholdtsen and Massey (2020) suggested using the principles of Recordkeeping (accountability, transparency, integrity, protection, compliance, availability, retention, and disposition) to manage IoT data. IS methods and tools have also been used in innovation to solve problems in supply chains (Maghsoodi *et al.*, 2018), in new applications in digital manufacturing (Wang *et al.*, 2020), and in agricultural development systems and China's rural digital commerce sector (Zhan; Zhang; Wang, 2021).

Healthcare is a significant producer of Big Data, with data stemming from examinations, diagnostics, laboratory tests, genetic sequencing, and public health data, among others. The retrieved studies demonstrated the concern of the field with the correct processing of such data for their effective use, emphasizing the importance and growth of the integration of the IS professional in the processing of "medical Big Data" (Chen *et al.*, 2019; Du; Chen; Zhang, 2021; Sayed, 2020; Fu *et al.*, 2020; Liyanage *et al.*, 2014; Nakashima, 2015; Pearce *et al.*, 2019; Scott *et al.*, 2018; Simpao; Ahumada; Rehman, 2015; Sobral; Lima; Sobral, 2021; Waldman; Terzic, 2016; Zhou *et al.*, 2020).

Another field that produces a large amount of data is related to aspects of urban planning, in which meteorological, disaster, and socioeconomic data are collected to assist the decision-making process of public managers. In this field, studies show IS acting mainly in data integration and homogenization (Wu *et al.*, 2017; Zhao *et al.*, 2016) and visualization (Valls; Roca, 2021).

Although management is closely related to SI, the studies identified as organizational management only tangentially addressed IS, with Big Data applications for public policies, business and industrial innovation, and design (Arnaboldi; Azzone, 2020; Lozada; Arias-Perez; Perdomo-Charry, 2019; Nunes, 2017).

Freund *et al.* (2019) analyzed technological information security mechanisms that contribute to the veracity of data in Big Data environments (encryption, access control, hashing, backup, data replication, digital certificates, digital signatures, and time stamps), aiming to contribute to research within IS. Aspects of surveillance and privacy were addressed in the research by Aradau and Blanke (2015), who discussed security practices in the digital age from a critical IS perspective, and Galvão, Costa and Pimenta (2018), who discussed the relationship between power, surveillance, and competence in information. In a literature review, Lott and Cianconi (2018) found little academic production on the topics of surveillance and privacy related to the terms Big Data and personal data in academic IS research in Brazil.

Privacy issues are closely related to ethical issues, addressed by Deighton (2019), who discussed the concept of Big Data and its impact on society, especially on how we collect and use information, which, according to the author, represents a discontinuity in the practice of IS that plays an essential role in understanding the impact of Big Data on society and in developing strategies to manage its use ethically and responsibly. The ethical challenges for IS with the use of Big Data applications were also discussed by Morán-Reyes (2022).

To overcome the new ethical and technological challenges imposed on the field by Big Data, professional training in 'MUST' be in constant transformation. The studies by Aparac-Jelusic (2019), Pirela Morillo and Salazar Álvarez (2021), Oh (2016), and Ortiz-Repiso, Greenberg and Calzada-Prado (2018) analyzed IS curricula of Chinese, Colombian, and i-schools and concluded that most curricula already offer courses related to data processing, with much room for an increase in the offer of courses on the topic.

As for the research carried out by IS, literature reviews were retrieved. More generally, Guallar *et al.* (2020) investigated Spanish IS journals and found five main areas, among which were 'open data and Big Data'. Concerning more specifically the relationship between IS and Big Data, three reviews found the following: a limitation of studies to the perspective of the library and information without an analysis of the status of research and development trends (Jiang, 2022); predominance of a more applied, technological profile focused on processes

related to the collection, curation, and use of data (Meschini; Francelin, 2022, p. 35); and, conversely, an exploratory and predominantly theoretical nature of these studies to the detriment of experimentation and field research (Santos-D'Amorim, 2020, p. 1).

Iwami *et al.* (2020) proposed a bibliometric approach that involved a Big Data analysis in IS publications to evaluate how this area of knowledge coevolves with information technology. Shiri (2014) proposed a faceted analysis approach to understand Big Data better. Thor *et al.* (2021) used new script methods to identify the most influential publications from each field indexed in WoS, including IS.

Investigating how the areas of knowledge have been using the developments of NLP, Falcão, Lopes and Souza (2022) found little use of the NLP approach by IS in extracting information from Big Data.

Information Science (IS) is an area that enables the treatment of various forms of information, both in production and organization, as well as in use and circulation aspects. This approach impacts topics such as urban planning, public policies, information security, and privacy. The analysis of identified themes reveals that areas related to IS methods and tools, medicine and public health, research in IS, education, and professional training are interconnected. This demonstrates that Big Data affects various dimensions of knowledge production and use across different fields.

The study's findings pointed to a diversity of research on Big Data and Information Science (IS). Although the research was extensive but not exhaustive, it resulted in the classification into ten main themes: IS methods and tools, medicine and public health, scientific research and communication, education and professional training, urban planning and surveillance, organizational management, ethics, information security, and Natural Language Processing (NLP).

The interconnections between areas, particularly with regards to methodological aspects, demonstrate the centrality of IS in data analysis. For example, the healthcare field benefits from IS methods for handling the large amounts of data generated by medical exams and records, but lacks studies that

delve into the relationship between health data and public health data, despite the "medicine and public health" theme existing. The same situation occurs in the other areas, which points to the need for a more in-depth examination, especially in terms of methodological aspects.

The study also revealed a strong presence of theoretical studies and reviews, while empirical studies were less common, indicating a research gap that needs to be filled in future investigations. Additionally, some methodologies present, such as bibliometric analysis and network analysis, have been repeated in the field, pointing to the need for a more nuanced look at developing new methodological approaches, especially those proposing to investigate the practical use of Big Data in Information Science. The discussion on ethics remains theoretical and general, but lacks specific ethical aspects that occur in the daily practice of information professionals.

The analysis of the interconnections between IS and Big Data highlights the importance of Information Science not only in treating and organizing data, but also in critically debating its ethical and social impacts. Ethics discussions, for instance, tend to focus on theoretical and general issues, rather than addressing the specific ethical challenges encountered by information professionals on a daily basis.

The rapid expansion of big data has brought forth challenges that require attention from both researchers and information professionals. The issue of privacy, for instance, manifests in various forms through the use of big data, from unauthorized collection of personal data to the abusive use of information for commercial and political purposes.

In this context, IS must take a stance in the discussion to develop solutions that enable the ethical use of technology and protect citizens' rights. Furthermore, the information inequality problem, characterized by unequal access to information and technology, is a structural issue that requires addressing through public policies and initiatives aimed at democratizing access to information and technology.

Algorithms also pose a challenge, as they often replicate societal biases, resulting in discrimination and social injustice. Although these challenges are

common in Big Data research, there is still a need for further in-depth study within Information Science to understand them and address them effectively.

The identification of primary themes revealed the diversity of Big Data applications in CI, but also highlighted the lack of empirical studies that demonstrate its practical application by information professionals.

6 Conclusion

A literature review was carried out to describe, quantitatively and qualitatively, how IS and Big Data are related, based on bibliometric and thematic analyses of scientific communications indexed in Scopus and Web of Science. We analyzed 51 articles from the 233 records retrieved. We found the use of IS methods and tools for Big Data applications in fields as diverse as IS methods and tools, medicine and public health, scientific research and communication, education and professional training, surveillance and privacy urban planning, organizational management, ethics, information security, and Natural Language Processing (NLP).

The importance of multidisciplinary teams for data recovery, processing, and treatment is highlighted. Researchers revealed the importance of considering the Big Data topic in the formulation and execution of curricula in IS programs. Ethical and information security challenges were topics addressed in the analyzed research. This confirms the interdisciplinary dynamics between DS and IS, as pointed out by most of the research previously identified in the state of the art.

Studies indicated the importance of considering the need for Big Data in IS research methodologies given the breadth and diversity of data and sources of information for analysis and decision-making.

We conclude that the intersections between Big Data and Information Science (IS) go beyond the technical and methodological, requiring a critical analysis of the ethical and social impacts of technology. IS is essential for ensuring the responsible use of Big Data, which demands a deeper exploration of empirical research and the development of practical solutions that balance technological advancement with the protection of citizens' rights.

Professional training must be improved to prepare future information professionals to tackle the ethical challenges arising from the use of Big Data. In this sense, this study highlights the importance of considering not only technical aspects but also ethical and social implications in future research, emphasizing the need for professionals to be equipped to handle everyday data treatment issues.

Lack of exploration of qualitative methodologies and limited discussion about the impacts of Big Data on socially vulnerable or historically oppressed groups must be addressed in future investigations. The analysis revealed a limited use of IS methods and tools to explore Big Data in areas such as culture and society, with little emphasis on topics like network analysis and qualitative approaches, which are fundamental to the field.

For future research, it is recommended to deepen discussions about the intersection between IS and Big Data in the following areas: development of more sophisticated data analysis methods, empirical studies to map the impacts of Big Data on different sectors of society, exploration of the use of Big Data in less investigated areas, such as human sciences and social sciences, and research into new qualitative methodologies for data analysis, which are still little explored in the context of Big Data.

References

AHMAD, Nasir *et al.* Using Fisher information to track stability in multivariate systems. **Royal Society Open Science**, United Kingdom, v. 3, n. 11, p. 160582, 2016. Available at: https://doi.org/10.1098/rsos.160582 . Access: Apr. 27, 2023.

ALHARTHI, Ahmed; KROTOV, Vladimir; BOWMAN, Michael. Addressing barriers to big data. **Business Horizons**, Amsterdam, v. 60, n. 3, p. 285-292, 2017. Available at: https://doi.org/10.1016/j.bushor.2017.01.002 . Access: Apr. 02, 2024.

ALLAM, Zeeshan; DHUNNY, Zainab Ali. On big data, artificial intelligence and smart cities. **Cities**, Amsterdam, v. 89, p. 80-91, 2019. Available at: https://doi.org/10.1016/j.cities.2018.10.003 . Access: Apr. 05, 2024.

APARAC-JELUŠIĆ, Tatjana. Sveučilišno obrazovanje informacijskih stručnjaka: izazovi i mogući iskoraci. **Bosniaca**, Saravejo, n. 24, p. 17-25, 2019. Available at: https://doi.org/10.37083/bosn.2019.24.17 . Access: Apr. 27, 2023.

ARADAU, Claudia; BLANKE, Tobias. The (Big) Data-security assemblage: knowledge and critique. **Big Data & Society**, Thousand Oaks, v. 2, n. 2, p. 1-12, 2015. Available at: https://doi.org/10.1177/2053951715609066 . Access: Apr. 27, 2023.

ARNABOLDI, Michela; AZZONE, Giovanni. Data science in the design of public policies: dispelling the obscurity in matching policy demand and data offer. **Heliyon**, Amsterdam, v. 6, n. 6, p. 1-13, 2020. Available at: https://doi.org/10.1016/j.heliyon.2020.e04300 . Access: Apr. 27, 2023.

BARNAGHI, Payam; SHETH, Amit; HENSON, Cory. From data to actionable knowledge: big data challenges in the web of things. **IEEE Intelligent Systems**, New York, v. 28, n. 6, p. 6-11, 2013. Available at: https://doi.org/10.1109/MIS.2013.142 . Access: Apr. 27, 2023.

BORKO, Harold. Information Science: what is it? **American Documentation**, New Jersey, v. 19, n. 1, p. 3-5, 1968. Available at: https://doi.org/10.1002/asi.5090190103 . Access: May 17, 2023.

BOYD, Danah; CRAWFORD, Kate. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Information. **Communication & Society**, London, v. 15, n. 5, p. 662-679, 2012. Available at: https://doi.org/10.1080/1369118X.2012.678878 . Access: Apr. 17, 2024.

CAO, Longbing. Data science: challenges and directions. **Communications of the ACM**, New York, v. 60, n. 8, p. 59-68, 2017. Available at: https://doi.org/10.1145/3015456 . Access: Apr. 13, 2023.

CHEN, Jie *et al*. Big data challenge: a data management perspective. **Frontiers of Computer Science**, New York, v. 7, n. 2, p. 157-164, 2013. Available at: https://doi.org/10.1007/s11704-013-3903-7 . Access: Apr. 13, 2024.

CHEN, Donghua *et al.* A bibliometric analysis of the development of ICD-11 in medical informatics. **Journal of Healthcare Engineering**, New Jersey, v. 2019, p. 1-12, 2019. Available at: https://doi.org/10.1155/2019/1649363 . Access: Apr. 27, 2023.

CHOO, Chun Wei. **Criação do conhecimento na empresa**: como as empresas japonesas geram a dinâmica da inovação. 11. ed. Rio de Janeiro: Campus, 2003.

CONSTANTIOU, Ioannis D.; KALLINIKOS, Jannis. New games, new rules: big data and the changing context of strategy. **Journal of Information Technology**, Thousand Oaks, v. 30, n. 1, p. 44-57, 2015. Available at: https://doi.org/10.1057/jit.2014.17 . Access: Apr. 13, 2024.

DANIEL, Ben Kei. Big Data and data science: A critical review of issues for educational research. **British Journal of Educational Technology**, New Jersey,

v. 50, n. 1, p. 101-113, 2019. Available at: https://doi.org/10.1111/bjet.12595 . Access: Apr. 13, 2023.

DAVENPORT, Thomas H. How strategists use “big data” to support internal business decisions, discovery and production. **Strategic Leadership**, Leeds, v. 42, n. 4, p. 45-50, 2014. Available at: https://doi.org/10.1108/SL-05-2014-0034 . Access: Apr. 13, 2024.

DAVENPORT, Thomas; PRUSAK, Laurence. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.

DE MAURO, Andrea; GRECO, Marco; GRIMALDI, Michele. A formal definition of Big Data based on its essential features. **Library Review**, Leeds, v. 65, n. 3, p. 122-135, 2016. Available at: https://doi.org/10.1108/LR-06-2015-0061 . Access: May 17, 2023.

DEIGHTON, John. Big data. **Consumption Markets & Culture**, London, v. 22, n. 1, p. 68-73, 2019. Available at: https://doi.org/10.1080/10253866.2017.1422902 . Access: Apr. 27, 2023.

DONG, J.Q., YANG, C.H. Business value of big data analytics: a systems-theoretic approach and empirical test. **Information & Management**, Amsterdam, v. 57, n. 1, p. 1-9, 2018. Available: https://doi.org/10.1016/j.im.2018.11.001 . Access: Apr. 27, 2023.

DU, Jian; CHEN, Ting; ZHANG, Luxia. Measuring the interactions between health demand, informatics supply, and technological applications in digital medical innovation for China: content mapping and analysis. **JMIR Medical Informatics**, Toronto, v. 9, n. 7, p. 1-18, 2021. Available at: https://doi.org/10.2196/26393 . Access: Apr. 27, 2023.

DUTRA, Moises Lima; MATIAS, Marcio. Visualização de correspondências semânticas no universo Big Data. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15, Belo Horizonte, 2014. **Anais** […].Belo Horizonte: ECI/UFMG, 2014. p. 4250-4257.

EINAV, Liran; LEVIN, Jonathan. Economics in the age of big data. **Science**, Washington, v. 346, n. 6210, p. 1243089, 2014. Available at: https://www.science.org/doi/10.1126/science.1243089 . Access: Mar 17, 2024.

FALCÃO, Luander Cipriano de Jesus; LOPES, Brenner; SOUZA, Renato Rocha. Absorção das tarefas de processamento de Linguagem Natural (NLP) pela Ciência da Informação (CI): uma revisão da literatura para tangibilização do uso de NLP pela CI. **Em Questão**, Porto Alegre, v. 28, n. 1, p. 13-34, 2022. Available at: https://doi.org/10.19132/1808-5245281.13-34 . Access: Apr. 27, 2023.

FREUND, Gislaine Parra *et al.* Mecanismos tecnológicos de segurança da informação no tratamento da veracidade dos dados em ambientes Big Data. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 2, p. 124-142, 2019. Available at: https://doi.org/10.1590/1981-5344/3348 . Access: Apr. 27, 2023.

FU, Mei R. *et al.* Precision health: a nursing perspective. **International Journal of Nursing Sciences**, Amsterdam, v. 7, n. 1, p. 5-12, 2020. Available at: https://doi.org/10.1016/j.ijnss.2019.12.008 . Access: Apr. 27, 2023.

FULLER, Michael. Big data: new science, new challenges, new dialogical opportunities. **Zygon**, New Jersey, v. 50, n. 3, p. 569-582, 2015. Available at: https://doi.org/10.1111/zygo.12187 . Access: Apr. 13, 2023.

GALVÃO, Fernanda do Valle; COSTA, Diego da Silva; PIMENTA, Ricardo Medeiros. Tecnovigilância e desafios à competência em informação. **Estudos em Comunicação**, Covilhã, v. 1, n. 26, p. 101-118, 2018. . Available at: https://doi.org/10.20287/ec.n26.v1.a07 . Access: Apr. 27, 2023

GANDOMI, Amirhossein; HAIDER, Murtaza. Beyond the hype: big data concepts, methods, and analytics. **International Journal of Information Management**, Amsterdam, v. 35, n. 2, p. 137-144, 2015. Available at: https://doi.org/10.1016/j.ijinfomgt.2014.10.007 . Access: Jan. 27, 2024.

GARCIA JIMÉNEZ, Antonio; CATALINA GARCÍA, Beatriz. Una perspectiva documental y bibliotecológica sobre el Big Data y el periodismo de datos. **Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información**, Ciudad de México, v. 32, n. 74, p. 77, 2018. Available at: https://doi.org/10.22201/iibi.24488321xe.2018.74.57910 . Access: Apr. 27, 2023.

GAROUFALLOU, Emmanouel; GAITANOU, Panorea. Big Data: opportunities and challenges in libraries, a systematic literature review. **College & Research Libraries**, Chicago, v. 82, n. 3, p. 410, 2021. Available at: https://doi.org/10.5860/crl.82.3.410 . Access: Apr. 13, 2023.

GEORGE, Gerard; OSINGA, Ernst C.; LAVIE, Dovev; SCOTT, Brent A. Big data and data science methods for management research [Editorial]. **Academy of Management Journal**, New York, v. 59, n. 5, p. 1493-1507, 2016. Available at: https://doi.org/10.5465/amj.2016.4005 . Access: Feb. 13, 2024.

GHIRINGHELLI, Luca M. *et al.* Learning physical descriptors for materials science by compressed sensing. **New Journal of Physics**, United Kingdom, v. 19, n. 2, p. 023017, 2017. Available at: https://doi.org/10.1088/1367-2630/aa57bf . Access: Apr. 27, 2023.

GOBBLE, Mary M. Big data: the next big thing in innovation. **Research-Technology Management**, London, v. 56, n. 1, p. 64-67, 2013. Available at: https://doi.org/10.5437/08956308X5601005 . Access: Mar. 27, 2024.

GOLUB, Koraljka; HANSSON, Joacim. (Big) Data in library and information science: a brief overview of some important problem areas. **JUCS - Journal of Universal Computer Science**, Austria, v. 23, n. 11, p. 1098-1108, 2017. Available at: https://doi.org/10.3217/jucs-023-11-1098 . Access: Apr. 13, 2023.

GRYK, Michael R.; LUDÄSCHER, Bertram. Workflows and provenance: toward information science solutions for the natural sciences. **Library Trends**, Baltimore, v. 65, n. 4, p. 555-562, 2017. Available at: https://dx.doi.org/10.1353/lib.2017.0018 . Access: Apr. 27, 2023.

GUALLAR, Javier *et al.* Revistas españolas de Documentación en Web of Science: análisis bibliométrico y evolución temática de 2015 a 2019. **El profesional de la información**, Spain, v. 29, n.6, p. 1-27, 2020. Available at: https://doi.org/10.3145/epi.2020.nov.06 . Access: Apr. 27, 2023.

GUIMARÃES, Suzane Albino da Silva; ROCHA, Ednéia Silva Santos; MUGNAINI, Rogério. Estudo cientométrico da atividade acadêmica sobre as temáticas de humanidades digitais e big data nas universidades estaduais paulistas. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 28, p. 1-34, 2023. Available at: https://doi.org/10.5007/1518-2924.2023.e90566 . Access: May 17, 2023.

GUPTA, Sanjay; KAR, Ajoy Kumar; BAABDULLAH, Asif; AL-KHOWAITER, Waleed Abdullah Al. Big data with cognitive computing: a review for the future. **International Journal of Information Managemen**t, Amsterdam, v. 42, p. 78-89, 2018. Available at: https://doi.org/10.1016/j.ijinfomgt.2018.06.005 . Access: Mar. 17, 2024.

GUPTA, Manish; GEORGE, Joey F. Toward the development of a big data analytics capability. **Information & Management**, Amsterdam, v. 53, n. 8, p. 1049-1064, 2016. Available at: https://doi.org/10.1016/j.im.2016.07.004 Access: Mar. 17, 2024.

HARWELL, Linda C. *et al.* Scientific data management in the age of big data: an approach supporting a resilience index development effort. **Frontiers in Environmental Science**, Lausanne, v. 7, p. 72, 2019. Available at: https://doi.org/10.3389/fenvs.2019.00072 . Access: Apr. 27, 2023.

IWAMI, Shino *et al.* A bibliometric approach to finding fields that co-evolved with information technology. **Scientometrics**, New York, v. 122, n. 1, p. 3-21, 2020. Available at: https://doi.org/10.1007/s11192-019-03284-9 . Access: Apr. 27, 2023.

JIANG, Shujie. Hotspot mining in the field of library and information science under the environment of Big Data. **Journal of Environmental and Public Health**, New Jersey, v. 2022, p. 1-9, 2022. Available at: https://doi.org/10.1155/2022/2802835 . Access: Apr. 27, 2023.

JIANG, Shujie. [Retracted]: hotspot mining in the field of library and information science under the environment of Big Data. **Journal of Environmental and Public Health**, New Jersey v. 2023, p. 1-9, 2023. Available at: https://doi.org/10.1155/2023/9821819 . Access: July 30, 2023.

KHOURY, Muin J.; IOANNIDIS, John P. A. Big data meets public health. **Science**, Washington, v. 346, p. 1054-1055, 2014. Available at: https://doi.org/10.1126/science.aaa2709 . Access: Apr. 27, 2024.

LIANG, Jifeng. "OBE" Concept for new training mode of electronic information science and technology professionals under Big Data analysis. **Computational Intelligence and Neuroscience**, New Jersey, v. 2022, p. 1-10, 2022. Available at: https://doi.org/10.1155/2022/8075708 . Access: Apr. 27, 2023.

LIMA, Carina Munhoz. **O big data e a ciência de dados na produção bibliográfica brasileira da biblioteconomia e da ciência da informação**. 2022. Dissertação (Mestrado) - Curso de Ciência da Informação, Universidade Federal Fluminense, Niterói, 2022.

LIYANAGE, Harshana *et al.* Big Data usage patterns in the health care domain: a use case driven approach applied to the assessment of vaccination benefits and risks: contribution of the IMIA Primary Healthcare Working Group. **Yearbook of Medical Informatics**, Germany, v. 23, n. 1, p. 27-35, 2014. Available at: https://doi.org/10.15265/IY-2014-0016 . Access: Apr. 27, 2023.

LOTT, Yuri Monnerat; CIANCONI, Regina de Barros. Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 23, n. 4, p. 117-132, 2018. Available at: https://doi.org/10.1590/1981-5344/3313 . Access: Apr. 27, 2023.

LOZADA, Nelson; ARIAS-PÉREZ, Jose; PERDOMO-CHARRY, Geovanny. Big data analytics capability and co-innovation: an empirical study. **Heliyon**, Amsterdam, v. 5, n. 10, p. 1-7, 2019. Available at: https://doi.org/10.1016/j.heliyon.2019.e02541 . Access: Apr. 27, 2023.

LYNCH, Clifford. How do your data grow? **Nature**, London, v. 455, p. 28-29, 2008. Available at: https://doi.org/10.1038/455028a. Access: Apr. 27, 2024.

MAGHSOODI, Abteen Ijadi *et al.* CLUS-MCDA: a novel framework based on cluster analysis and multiple criteria decision theory in a supplier selection problem. **Computers & Industrial Engineering**, Amsterdam, v. 118, p. 409-

422, 2018. Available at: https://doi.org/10.1016/j.cie.2018.03.011 . Access: Apr. 27, 2023.

MESCHINI, Fabio Orsi; FRANCELIN, Marivalde Moacir. Big data e organização do conhecimento: reflexões iniciais a partir de uma proposta classificatória da produção científica. **Em Questão**, Porto Alegre, v. 28, n.1, p. 35-63, 2022. Available at: https://doi.org/10.19132/1808-5245281.35-63 . Access: Apr. 27, 2023.

MIKALEF, Patrick; PAPPAS, Ilias O.; KROGSTIE, John; GIANNAKOS, Michail. Big data analytics capabilities: a systematic literature review and research agenda. **Information Systems and e-Business Managemen**t, New York, v. 16, n. 3, p. 547-578, 2018. Available at: https://doi.org/10.1007/s10257-017-0362-y . Access: May 05, 2023.

MOHER, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **PLoS medicine**, California, v. 6, n. 7, p. e1000097, 2009. Available at: https://doi.org/10.1371/journal.pmed.1000097 . Access: July 05, 2022.

MORÁN-REYES, Ariel Antonio. Towards an ethical framework about Big Data era: metaethical, normative ethical and hermeneutical approaches. **Heliyon**, Amsterdam, v. 8, n. 2, p. 1-8, 2022. Available at: https://doi.org/10.1016/j.heliyon.2022.e08926 . Access: Apr. 27, 2023.

MURTAGH, Fionn; DEVLIN, Keith. The development of data science: implications for education, employment, research, and the data revolution for sustainable development. **Big Data and Cognitive Computing**, Basel, v. 2, n. 2, p. 1-16, 2018. Available at: https://doi.org/10.3390/bdcc2020014 . Access: July 05, 2022.

NAKASHIMA, Naoki. What Big Data means. **Japanese Journal of Neurosurgery**, Japan, v. 24, n. 10, p. 664-671, 2015. Available at: https://doi.org/10.7887/jcns.24.664 . Access: Apr. 27, 2023.

NASCIMENTO, Lucileide Andrade de Lima; MIGUEL, Marcelo Calderari; COSTA, Rosa da Penha Ferreira. Big Data na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci): um enfoque situacional. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 12, n. 2, p. 126-150, 2021. Available at: https://doi.org/10.11606/issn.2178-2075.v12i2p126-150 . Access: May 17, 2023.

NIH. MeSH descriptor data 2022. *In*: **Medical Subject Headings - MeSH**. Bethesda, MD: National Library of Medicine, 2022.

NUNES, Miguel Baptista. Understanding Big Data for industrial innovation and design: the missing information systems perspective. **Journal of Data and**

Information Science, Beijing, v. 2, n. 4, p. 1-6, 2017. Available at: https://doi.org/10.1515/jdis-2017-0017 . Access: Apr. 27, 2023.

OH, Sanghee *et al.* Results of a digital library curriculum field test. **International Journal on Digital Libraries**, New York, v. 17, n. 4, p. 273-286, 2016. Available at: https://doi.org/10.1007/s00799-015-0151-5 . Access: Apr. 27, 2023.

OPDAHL, Andreas L.; TESSEM, Bjørnar. Ontologies for finding journalistic angles. **Software and Systems Modeling**, New York, v. 20, n. 1, p. 71-87, 2021. Available at: https://doi.org/10.1007/s10270-020-00801-w . Access: Apr. 27, 2023.

ORTIZ-REPISO, Virginia; GREENBERG, Jane; CALZADA-PRADO, Javier. A cross-institutional analysis of data-related curricula in information science programmes: a focused look at the iSchools. **Journal of Information Science**, Thousand Oaks, v. 44, n. 6, p. 768-784, 2018. Available at: https://doi.org/10.1177/0165551517748149 . Access: Apr. 27, 2023.

OUYANG, Qiang; WU, Changxin; HUANG, Lei. Methodologies, principles and prospects of applying big data in safety science research. **Safety Science**, Amsterdam, v. 101, p. 60-71, 2018. Available at: https://doi.org/10.1016/j.ssci.2017.08.012 . Access: Mar. 27, 2023.

PAGE, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, London, v. 372 n. 71, p. 1-9, 2021. Available at: https://doi.org/10.1136/bmj.n71 . Access: July 05, 2022.

PEARCE, Christopher *et al.* Coding and classifying GP data: the POLAR project. **BMJ Health & Care Informatics**, Bethesda, v. 26, n. 1, p. e100009, 2019. Available at: https://doi.org/10.1136/bmjhci-2019-100009 . Access: Apr. 27, 2023.

PIRELA MORILLO, Johann Enrique; SALAZAR ÁLVAREZ, Leidy Marisol. Perfiles de los profesionales de la información: entre lo tradicional y lo emergente. **Revista Interamericana de Bibliotecología**, Medellín, v. 44, n. 3, p. 1-11, 2021. Available at: https://doi.org/10.17533/udea.rib.v44n3e344766 . Access: Apr. 27, 2023.

SANTOS-D'AMORIM, Karen Isabelle *et al*. Dos dados ao conhecimento: tendências da produção científica sobre Big Data na Ciência da Informação no Brasil. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 25, p. 1-23, 2020. Available at: https://doi.org/10.5007/1518-2924.2020.e70527 . Access: May 17, 2023.

SAYED, Esmat Mohamed Abdel Moniem el. How AI, Data science and technology is used to fight the pandemic covid-19: case study in Saudi Arabia

environment. **Research in World Economy**, Ontario, v. 11, n. 5, p. 409, 2020. Available at: https://doi.org/10.5430/rwe.v11n5p409 . Access: Apr. 27, 2023.

SCOTT, Philip J. *et al.* Learning health systems need to bridge the 'two cultures' of clinical informatics and data science. **BMJ Health & Care Informatics**, Bethesda, v. 25, n. 2, p. 126-131, 2018. Available at: https://doi.org/10.14236/jhi.v25i2.1062 . Access: Apr. 27, 2023.

SHIRI, Ali. Making sense of Big Data: a facet analysis approach. **Knowledge Organization**, Baden-Baden, v. 41, n. 5, p. 357-368, 2014. Available at: https://doi.org/10.5771/0943-7444-2014-5-357. Access: Apr. 27, 2023.

SEDDON, John J. J. M.; CURRIE, Wendy L. A model for unpacking big data analytics in high frequency trading. **Journal of Business Research**, Amsterdam, v. 70, p. 300-307, 2017. Available at: https://doi.org/10.1016/j.jbusres.2016.08.003. Access: Jan. 27, 2024.

SIMPAO, Allan F.; AHUMADA, Luis M.; REHMAN, Mohamed A. Big data and visual analytics in anaesthesia and health care. **British Journal of Anaesthesia**, Amsterdam, v. 115, n. 3, p. 350-356, 2015. Available at: https://doi.org/10.1093/bja/aeu552 .Access: Apr. 27, 2023.

SIVARAJAH, Uthayasankar *et al*. Critical analysis of Big Data challenges and analytical methods. **Journal of Business Research**, Amsterdam, v. 70, p. 263-286, 2017. Available at: https://doi.org/10.1016/j.jbusres.2016.08.001 . Access: Apr. 27, 2023.

SOBRAL, Natanael Vitor; LIMA, Gillian Leandro de Queiroga; SOBRAL, Ana Sara Pereira de Melo. Produção científica sobre hospitais no contexto da ciência de dados: um estudo a partir da Web of Science. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 26, n. esp. p. 1-16, 2021. Available at: https://doi.org/10.5007/1518-2924.2021.78824 . Access: Apr. 27, 2023.

SØDRING, Thomas; REINHOLDTSEN, Petter; MASSEY, David. A record-keeping approach to managing IoT-data for government agencies. **Records Management Journal**, Leeds, v. 30, n. 2, p. 221-239, 2020. Available at: https://doi.org/10.1108/RMJ-09-2019-0050 . Access: Apr. 27, 2023.

SONG, Il-Yeol; ZHU, Yongjun. Big data and data science: what should we teach? **Expert Systems**, New Jersey, v. 33, n. 4, p. 364-373, 2016. Available at: https://doi.org/10.1111/exsy.12130 . Access: Jan. 12, 2023.

STOCKMANN, Daniela. Toward area-smart data science: critical questions for working with Big Data from China. **Policy & Internet**, New Jersey, v. 10, n. 4, p. 393-414, 2018. Available at: https://doi.org/10.1002/poi3.192 . Access: Apr. 13, 2023.

THOR, Andreas *et al.* Which are the influential publications in the Web of Science subject categories over a long period of time? CRExplorer software used for big-data analyses in bibliometrics. **Journal of Information Science**, Thousand Oaks, v. 47, n. 3, p. 419-428, 2021. Available at: https://doi.org/10.1177/0165551520913817 . Access: Apr. 27, 2023.

TOFFLER, Alvin. **The third wave**. New York: Bantam Books, 1990.

VALLS, Francesc; ROCA, Josep. Visualizing digital traces for sustainable urban management: mapping tourism activity on the virtual public space. **Sustainability**, Basel, v. 13, n. 6, p. 3159, 2021. Available at: https://doi.org/10.3390/su13063159 . Access: Apr. 27, 2023.

VIRKUS, Sirje; GAROUFALLOU, Emmanouel. Data science from a library and information science perspective. **Data Technologies and Applications**, Leeds, v. 53, n. 4, p. 422-441, 2019. Available at: https://doi.org/10.1108/DTA-05-2019-0076 . Access: Apr. 13, 2023.

VIRKUS, Sirje; GAROUFALLOU, Emmanouel. Data science and its relationship to library and information science: a content analysis. **Data Technologies and Applications**, Leeds, v. 54, n. 5, p. 643-663, 2020. Available at: https://doi.org/10.1108/DTA-07-2020-0167 . Access: Apr. 13, 2023.

VOLODIN, Andrey Y. Between 'data' and 'capta': the problem of datafication in historical research. **Вестник Пермского университета - История**, Russian, n. 3, p. 137-145, 2019. Available at: 10.17072/2219-3111-2019-3-137-145. Access: Apr. 27, 2023.

WALDMAN, Samuel; TERZIC, Andre. Big Data transforms discovery-utilization therapeutics continuum. **Clinical Pharmacology & Therapeutics**, New Jersey, v. 99, n. 3, p. 250-254, 2016. Available at: https://doi.org/10.1002/cpt.322 . Access: Apr. 27, 2023.

WANG, Junping *et al.* Constructing and inferring event logic cognitive graph in the field of big data. **Scientia Sinica Informationis**, Beijing, v. 50, n. 7, p. 988-1002, 2020. Available at: https://doi.org/10.1360/SSI-2019-0273 . Access: Apr. 27, 2023.

WANG, Ping; ZHAO, Pengfei; LI, Yingji. Design of education information platform on education Big Data visualization. **Wireless Communications and Mobile Computing**, New Jersey, v. 2022, p. 1-13, 2022. Available at: https://doi.org/10.1155/2022/6779105 . Access: Apr. 27, 2023.

WANG, Xiaoguang; WANG, Hongyu; HUANG, Han. Evolutionary exploration and comparative analysis of the research topic networks in information disciplines. **Scientometrics**, New York, v. 126, n. 6, p. 4991-5017, 2021. Available at: https://doi.org/10.1007/s11192-021-03963-6 . Access: Apr. 13, 2023.

WANG, Bo; WU, Chao; HUANG, Li; KANG, Liang. Using data-driven safety decision-making to realize smart safety management in the era of big data: a theoretical perspective on basic questions and their answers. **Journal of Cleaner Production**, Amsterdam, v. 210, p. 1595-1604, 2019. Available at: https://doi.org/10.1016/j.jclepro.2018.11.181 . Access: Apr. 13, 2023.

WEI, Xiaoni. Video sequence analysis for on-table tennis player ranking and analysis. **International Journal of Mobile Computing and Multimedia Communications**, New York, v. 13, n. 2, p. 1-9, 2022. Available at: https://doi.org/10.4018/IJMCMC.293750. Access: Apr. 27, 2023.

WU, XianHua *et al.* Review and prospect of the emergency management of urban rainstorm waterlogging based on big data fusion. **Chinese Science Bulletin**, Beijing, v. 62, n. 9, p. 920-927, 2017. Available at: https://doi.org/10.1360/N972016-01080 . Access: Apr. 27, 2023.

XU, Zongben; SHI, Yong. Exploring Big Data analysis: fundamental scientific problems. **Annals of Data Science**, New York, v. 2, n. 4, p. 363-372, 2015. Available at: https://doi.org/10.1007/s40745-015-0063-7 . Access: Apr. 27, 2023.

YIN, Ziyan; TSAI, Sang-Bing. Research on virtual reality interactive teaching under the environment of Big Data. **Mathematical Problems in Engineering**, New Jersey, v. 2021, p. 1-11, 2021. Available at: https://doi.org/10.1155/2021/7980383 . Access: Apr. 27, 2023.

YOU, Sukjin; JOO, Soohyung; KATSURAI, Marie. Data mining topics in the discipline of library and information science: analysis of influential terms and Dirichlet multinomial regression topic model. **Aslib Journal of Information Management**, Leeds, v. 76, n. 1, p. 65-85, 2022. Available at: https://doi.org/10.1108/AJIM-05-2022-0260 . Access: Apr. 13, 2023.

ZHAN, Hui; ZHANG, Xin; WANG, Haiwen. Influencing factor modeled examination on internet rural logistics talent innovation mechanism based on fuzzy comprehensive evaluation method. **Plos One**, California, v. 16, n. 3, p. 1-20, 2021. Available at: https://doi.org/10.1371/journal.pone.0246599 . Access: Apr. 27, 2023.

ZHANG, Jin; WOLFRAM, Dietmar; MA, Feicheng. The impact of big data on research methods in information science. **Data and Information Management**, Amsterdam, v. 7, n. 2, p. 1-7, 2023. Available at: https://doi.org/10.1016/j.dim.2023.100038 . Access: May 17, 2023.

ZHAO, Jing *et al.* Simulation data warehouse for integration and analysis of disaster information. **Journal of Disaster Research**, Japan, v. 11, n. 2, p. 255-264, 2016. Available at: https://doi.org/10.20965/jdr.2016.p0255 . Access: Apr. 27, 2023.

ZHOU, Yadi *et al.* Artificial intelligence in covid-19 drug repurposing. **The Lancet Digital Health**, Amsterdam, v. 2, n. 12, p. 1-10, 2020. Available at: https://doi.org/10.1016/S2589-7500(20)30192-8 . Access: Apr. 27, 2023.

Interseções entre Ciência da Informação e Big Data

Resumo: O objetivo deste trabalho é identificar as inter-relações entre Big Data e Ciência da Informação, com base em uma revisão sistemática da literatura de comunicações científicas indexadas nas bases de dados Scopus e Web of Science. Metodologicamente, a expressão genérica de busca foi ("big data" AND "information science"), e foram recuperados 233 registros. Após filtragens criteriosas, o objeto de estudo foi definido por uma seleção de 52 artigos. Mediante análise temática, foram identificados os elementos mais relevantes a serem classificados, identificando-se dez temas principais. Como resultados qualitativos, foram verificados aportes metodológicos da Ciência da Informação em pesquisas com Big Data em diversos campos, tais como: métodos e ferramentas da Ciência da Informação, medicina e saúde pública, pesquisa na Ciência da Informação, formação profissional, planejamento urbano e políticas públicas, além de discussões éticas e pesquisas teóricas e aplicações práticas. Conclui-se que a Ciência da Informação possui muitas interrelações com Big Data e tem muito a contribuir na pesquisa e uso desse tipo de análise, além de fortalecer a interdisciplinaridade entre a Ciência da Informação e a Ciência de Dados.

Palavras-chave: ciência da informação; big data; ciência de dados; interdisciplinaridade

Authorship and responsibility statement

Conception and design of the study: Joel Perozo-Vasquez, Ana Depizzolatti
Data collection: Joel Perozo-Vasquez
Data analysis and interpretation: Joel Perozo-Vasquez, Ana Depizzolatti
Writing: Joel Perozo-Vasquez, Ana Depizzolatti, Nathália Berger Werlang
Critical review of the manuscript: Ana Depizzolatti, Nathália Berger Werlang

Corresponding author
Joel Perozo-Vasquez
ufscjoel@gmail.com

How to cite
PEROZO-VASQUEZ, Joel; DEPIZZOLATTI, Ana; WERLANG, Nathália Berger. Intersections between Information Science and Big Data. **Em Questão**, Porto Alegre, v. 31, e-140613, 2025. https://doi.org/10.1590/1808-5245.31.140613

Received: 06/08/2024
Accepted: 01/31/2025