**Sciences du Nord Nature Science and Technology**
**Vol. 2, No. 01, June 2025: pp. 21-29**
**E-ISSN: 3046-6504**
**P-ISSN: 3047-4922**
**DOI: 10.71238/snnst.v2i01.67**


# Bibliometric Study on the Role of Big Data in the Science Revolution

**Loso Judijanto [1], Diky Wardhani [2]**
IPOSS Jakarta, losojudijantobumn@gmail.com
Universitas Siber Indonesia, diky.wardhani@cyber-univ.ac.id
Coresponding Author: losojudijantobumn@gmail.com

**ARTICLE HISTORY**

**Received April, 2025** **Revised May, 2025** **Accepted June, 2025**

**ABSTRACT**

The rapid proliferation of big data has ushered in a transformative era in scientific research, often referred to as the data-driven or fourth paradigm of science. This study employs a bibliometric approach using VOSviewer to analyze the intellectual structure, thematic evolution, and interdisciplinary impact of big data in the context of the modern science revolution. Drawing from the Scopus database, we examined publications from 2000 to 2024, focusing on co-occurrence of keywords, citation networks, and temporal patterns. The findings reveal that big data serves as a central node connecting diverse domains such as artificial intelligence, data science, genomics, industrial automation, and social sciences. Temporal overlay analysis indicates a shift from early emphasis on data management and social computing toward recent trends in AI, deep learning, and industry 4.0. Density visualizations highlight concentrated scholarly attention on computational and biomedical applications, while ethical and social implications remain underrepresented. This study affirms that big data is not merely a technological tool but a transformative epistemological force reshaping the landscape of contemporary scientific inquiry.

***Keywords:*** *Big Data; Science Revolution; Bibliometric Analysis; Artificial Intelligence; Data Science.*

## INTRODUCTION

Over the past two decades, the advent of big data has significantly reshaped the contours of scientific inquiry and technological development. As data generation accelerates exponentially across various sectors—from genomics to astronomy, from climate modeling to social science—researchers have been confronted with new challenges and opportunities [1]. Big data, characterized by its high volume, velocity, and variety, has emerged not only as a technical paradigm but also as a transformative agent in the production of scientific knowledge. It allows for previously unattainable patterns, correlations, and predictions to be unearthed, thereby revolutionizing the methodology and epistemology of science [2], [3].

Historically, scientific revolutions have been marked by paradigm shifts, as famously described by Thomas Kuhn. These shifts often stem from the introduction of new technologies or concepts that redefine the boundaries of inquiry. The rise of big data mirrors this historical pattern. Just as the telescope redefined astronomy and the microscope transformed biology, big data now redefines scientific discovery. It enables hypothesis-free explorations, supports real-time experimentation, and augments decision-making across disciplines. This has led many scholars to argue that big data represents a new mode of science—often termed "data-intensive science" or the "fourth paradigm" [4], [5], [6].

The implications of this data-driven shift are far-reaching. In the biomedical sciences, for example, the integration of big data analytics with clinical records and genomic data has facilitated advances in precision medicine and epidemiological modelling [7]. In environmental science, satellite data combined with machine learning algorithms have enabled better climate forecasting and disaster management. Even in the social sciences, big data sourced from social media, mobile devices, and online behavior is offering unprecedented insights into human behavior, opinion dynamics, and societal trends [8], [9]. These developments UNDERSCORE the pivotal role that big data now plays in the epistemic infrastructure of modern science.

However, the integration of big data into the scientific process has also raised philosophical, ethical, and methodological debates. Critics caution that overreliance on correlation-driven models may marginalize theory-driven research. Concerns about data quality, reproducibility, algorithmic bias, and interpretability have sparked ongoing discussions about the responsible use of data in research [10]. Furthermore, disparities in data access and computational resources have created inequalities among institutions and researchers, thereby affecting the democratization of scientific innovation.

Given the complex and evolving nature of big data's role in science, there is a pressing need to understand its trajectory, impact, and interdisciplinary reach. Bibliometric analysis serves as a powerful tool in this context, enabling researchers to map the intellectual structure, thematic evolution, and research frontiers of the field. By analyzing patterns in scientific publications—such as co-citation networks, keyword co-occurrence, and authorship trends—bibliometric studies offer empirical insights into how big data has transformed the landscape of scientific research across disciplines and time.

Despite the explosive growth in big data-related research, there remains a lack of systematic understanding of how this paradigm has influenced the broader architecture of scientific inquiry. Existing studies tend to be domain-specific or conceptual, often neglecting a comprehensive quantitative assessment of the field’s intellectual development. This fragmentation inhibits our ability to grasp the full scope and interdisciplinary dynamics of the “big data revolution” in science. Consequently, there is a gap in the literature concerning a macro-level overview of the thematic trends, collaborative structures, and influential works that characterize this transformative era. This study aims to conduct a comprehensive bibliometric analysis to explore the role of big data in the ongoing revolution of scientific practices and paradigms.

## METHOD

This study adopts a bibliometric approach to analyze the role of big data in the scientific revolution. Bibliometric analysis offers a systematic and quantitative method for evaluating patterns in scientific literature, allowing researchers to uncover trends, identify influential publications, and map intellectual structures across disciplines. In this study, the bibliometric analysis was conducted using VOSviewer, a widely used software tool for constructing and visualizing bibliometric networks.

The bibliometric data were extracted from the Scopus database, selected for its comprehensive coverage of peer-reviewed literature across diverse scientific fields. The search was conducted using a carefully constructed query to capture relevant publications that explore the intersection of big data and scientific transformation. The search string included terms such as “big data” AND (“science revolution” OR “scientific revolution” OR “scientific paradigm” OR “research transformation”) in the title, abstract, or keywords, ensuring the retrieval of thematically relevant documents. The timespan was limited to the years 2000 to 2024, reflecting the emergence and growth of big data as a scientific paradigm in the 21st century.

The retrieved dataset was downloaded in CSV and RIS formats for processing in VOSviewer. Prior to analysis, the data underwent pre-processing to remove duplicates, incomplete records, and non-research documents such as editorials or errata. Only journal articles, reviews, and conference proceedings were included to maintain the academic rigor and relevance of the analysis. Using VOSviewer (version 1.6.x), three types of bibliometric analyses were conducted. The visualizations produced by VOSviewer include network maps, density maps, and overlay maps. These graphical outputs enabled the identification of research clusters, temporal trends, and the evolution of core topics over time. For instance, the overlay visualization showed how early studies focused on data storage and management, while recent years emphasized ethical considerations and algorithmic transparency. All results from VOSviewer were interpreted contextually, supported by descriptive statistics such as the number of publications per year, average citations per document, and top contributing countries, institutions, and journals. The findings were then triangulated with qualitative insights from the literature to provide a well-rounded understanding of the field's development.

## RESULT

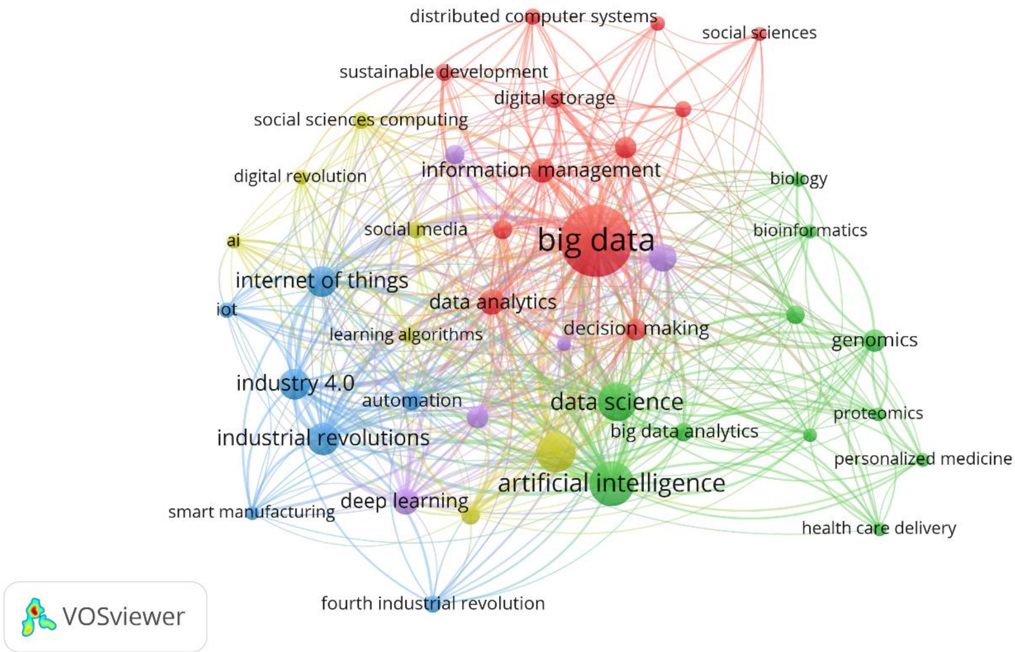


Figure 1. Network Visualization
Source: Data Analysis

The network visualization displays the co-occurrence of keywords related to "big data" in the context of the scientific revolution. At the center of the map is the term "big data", which is the most prominent and frequently occurring keyword, signifying its central role in the scholarly discourse. The large red node indicates a high frequency and strong connectivity with numerous other terms. Its position at the core reflects the integrative and cross-disciplinary nature of big data, which interacts with diverse fields including data analytics, artificial intelligence, genomics, social sciences, and information management. Surrounding the central node are distinct thematic clusters, each color-coded to reflect related keyword groupings. The red cluster (upper region) is oriented toward social sciences and information systems, including terms such as "information management," "social media," "social sciences," and "digital storage." This cluster illustrates how big data has become integral in areas that manage and interpret human-generated information, highlighting the sociotechnical dimensions of data science.

The green cluster (right side) is closely tied to biological and medical sciences, with key terms such as “genomics,” “bioinformatics,” “proteomics,” “personalized medicine,” and “health care delivery.” This indicates the crucial role of big data in life sciences, especially in supporting the data-intensive nature of precision medicine and systems biology. The strong linkages between "artificial intelligence," "data science," and these biomedical terms suggest a convergence of computational and biological research efforts. The blue cluster (left side) is associated with industrial and technological domains, including terms like “internet of things (IoT),” “industry 4.0,” “industrial revolutions,” “automation,” and “smart manufacturing.” These keywords reflect the integration of big data into cyber-physical systems and advanced manufacturing processes, underscoring its transformative role in the Fourth Industrial Revolution. The presence of “deep learning,” “ai,” and “learning algorithms” within this cluster shows how predictive analytics and machine learning are enhancing operational efficiency in industrial contexts. The yellow nodes scattered across the network act as connectors between various domains, linking topics such as “digital revolution” and “sustainable development” with both social and technical disciplines. These bridging terms point to the interdisciplinary nature of big data, as it is increasingly applied to address global challenges, policy development, and sustainable innovation.

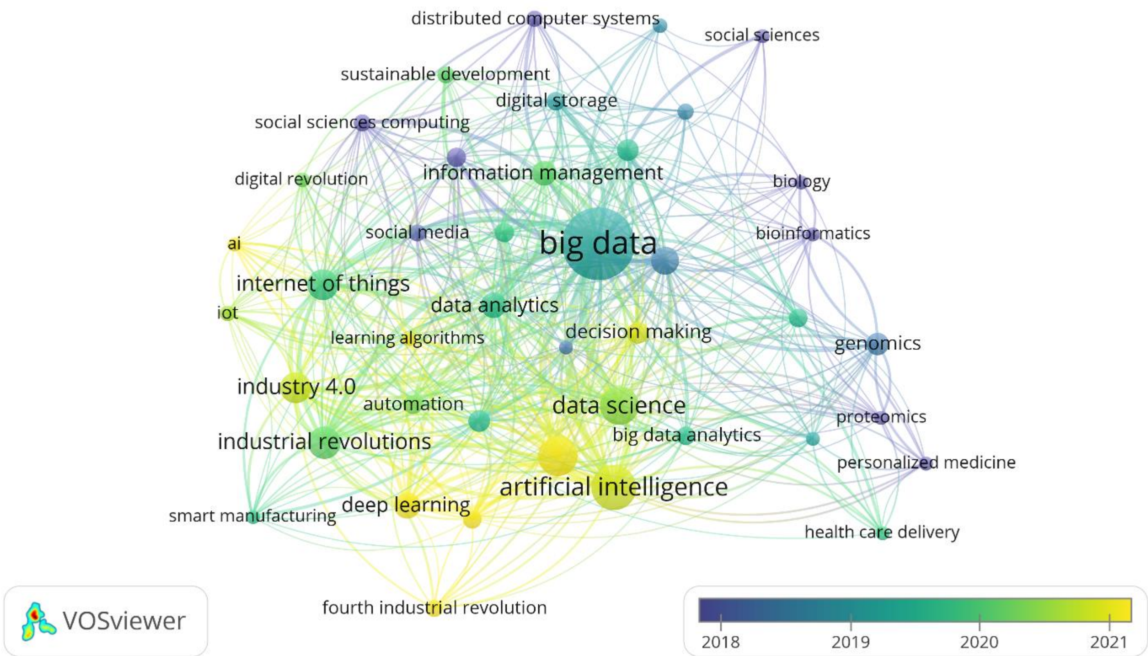


Figure 2. Overlay Visualization
Source: Data Analysis

The overlay visualization from VOSviewer shows the temporal evolution of keywords related to big data within scientific literature, using color gradients to indicate the average publication year associated with each term. The color scale ranges from purple (2018) to yellow (2021), allowing us to observe which topics are more established and which are emerging. The central term “big data,” along with “data analytics,” “information management,” and “decision making,” appears in darker teal hues, suggesting that these foundational concepts gained prominence earlier in the timeline (around 2018–2019). These areas form the core of the discourse around big data, representing mature and widely explored themes.

In contrast, newer and emerging trends are indicated by the yellow-colored nodes, especially in the lower part of the network. Terms like "artificial intelligence," "deep learning," "smart manufacturing," and "fourth industrial revolution" show more recent average publication dates (closer to 2021). This suggests a shifting research focus toward the integration of big data with intelligent systems, automation, and next-generation industrial applications. These developments are in line with the acceleration of digital transformation in various sectors and reflect the increasing role of big data in enabling intelligent decision-making systems and autonomous operations. Meanwhile, keywords located in the upper right quadrant, such as "genomics," "proteomics," "bioinformatics," and "personalized medicine," are tinted in purple and blue, indicating earlier incorporation of big data into the life sciences domain.

Table 1. Most Cited Article

| Citations | Author and Year | Title |
|---|---|---|
| 2320 | [11] | Data mining with big data |
| 2309 | [12] | Industry 4.0: A survey on technologies, applications and open research issues |
| 1600 | [13] | Big Data, new epistemologies and paradigm shifts |
| 1028 | [14] | Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management |
| 720 | [15] | Computer age statistical inference: Algorithms, evidence, and data science |
| 470 | [16] | Towards an operator 4.0 typology: A human-centric perspective on the fourth industrial revolution technologies |
| 462 | [17] | 50 Years of Data Science |
| 383 | [18] | The service revolution and the transformation of marketing science |
| 312 | [19] | From big data to precision medicine |
| 285 | [20] | State of Industry 5.0—Analysis and Identification of Current Research Trends |

Source: Scopus, 2025

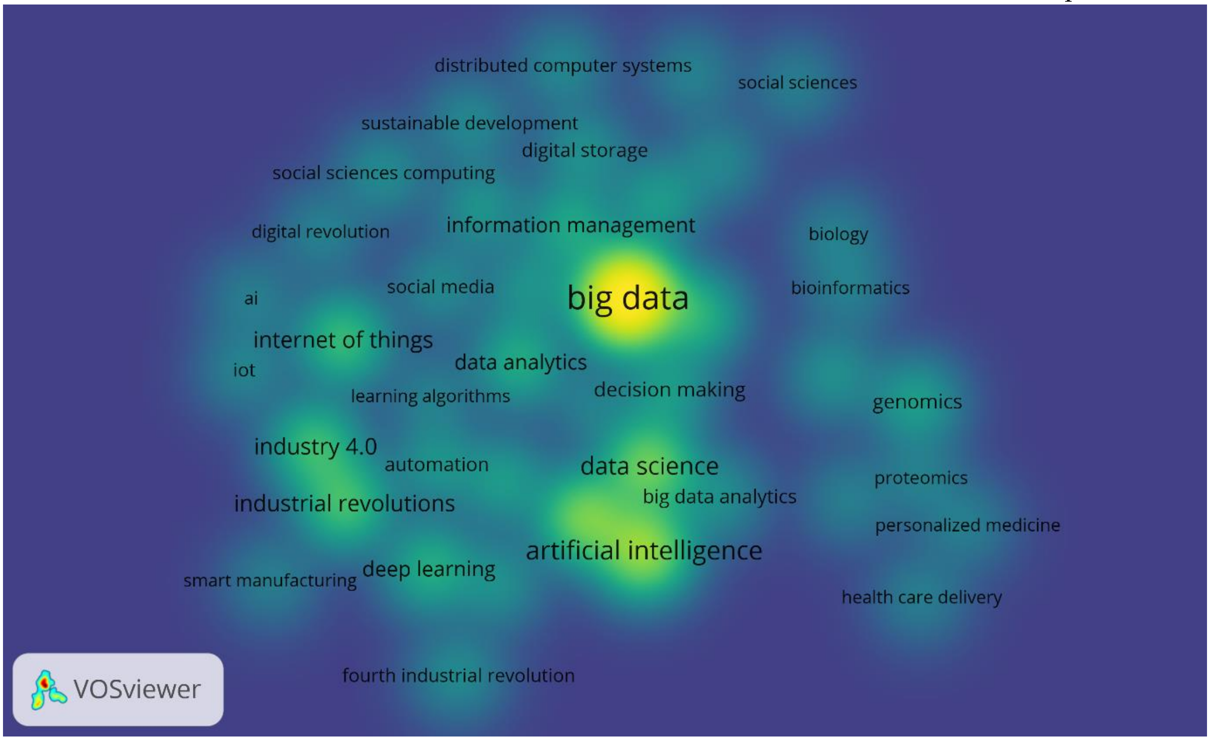

Figure 3. Density Visualization
Source: Data Analysis

The density visualization highlights the frequency and intensity of keyword occurrences related to big data in scientific literature. In this heatmap, brighter yellow areas represent terms that appear most frequently and are most central to the literature, while green and blue areas indicate moderate to lower occurrences. At the center, the term “big data” stands out with the highest density, confirming its role as the dominant theme and intellectual anchor in the field. Closely associated terms such as “data analytics,” “artificial intelligence,” “data science,” and “information management” also exhibit high densities, suggesting that these concepts form the core cluster of discourse surrounding the application of big data in science.

Peripheral areas with cooler green and blue tones—such as “genomics,” “bioinformatics,” “personalized medicine,” “industry 4.0,” and “internet of things”—indicate emerging or domain-specific applications of big data. These terms, while not as frequent as the central cluster, show meaningful connectivity and specialization in subfields like healthcare, biology, and industrial automation. The spread of density across disciplines signifies the interdisciplinary diffusion of big data technologies, supporting their transformative role in both theoretical and applied scientific research. This visualization effectively reinforces the idea that while big data remains the epicenter, its influence radiates widely across diverse scientific revolutions.

## DISCUSSION

The role of big data in reshaping the landscape of scientific research has become increasingly evident over the past decade, and the findings of this bibliometric study offer a comprehensive view of its growing influence across multiple disciplines. Through the analysis of keyword co-occurrences, temporal trends, and research densities, this study illustrates the paradigmatic shift that big data has triggered in modern science. As visualized in the network, overlay, and density maps generated via VOSviewer, big data has established itself not merely as a tool but as a fundamental framework influencing how science is conceived, conducted, and applied.

The centrality of the term “big data” in the network visualization reinforces its role as a foundational concept in the contemporary research ecosystem. It is the nexus connecting diverse domains such as artificial intelligence (AI), data science, genomics, industry 4.0, and information management. The magnitude of this central node underscores its interdisciplinary importance and reflects the extensive volume of literature that positions big data as a central concern. As a result, big data is not siloed within data engineering or computer science alone, but rather interwoven into the fabric of biomedical sciences, industrial innovation, social sciences, and environmental studies.

One key insight emerging from the analysis is the close relationship between big data and artificial intelligence. The strong co-occurrence and high density of terms such as “artificial intelligence,” “deep learning,” and “data analytics” point to a synergistic relationship where AI techniques are employed to extract meaning from massive datasets. This relationship has become especially critical in domains like personalized medicine, where AI models trained on big data have enabled predictive diagnostics, patient-specific treatment plans, and real-time clinical decision support systems [7]. Similarly, in industrial applications, the combination of big data and AI underpins predictive maintenance, process automation, and smart manufacturing, all of which are hallmarks of the Fourth Industrial Revolution [8].

In the life sciences and healthcare, keywords like “genomics,” “bioinformatics,” “proteomics,” and “health care delivery” emphasize how big data has transformed these traditionally data-scarce disciplines into data-intensive sciences. For example, genome sequencing technologies now generate terabytes of data that require sophisticated computational tools to analyze. The co-evolution of big data and bioinformatics has enabled researchers to identify complex genetic patterns, develop targeted therapies, and advance personalized medicine [9]. The temporal visualization further reveals that many of these applications gained momentum earlier (circa 2018), indicating that healthcare was among the first sectors to fully embrace the scientific potential of big data.

The overlay map provided additional temporal insights, illustrating how the focus of big data research has shifted over time. Earlier literature (2018–2019) clustered around traditional data concerns—such as "data management," "decision making," and "social sciences"—while more recent topics (2020–2021), indicated by yellow nodes, reveal a pivot toward AI-centric, automation-driven, and industry-focused themes. This shift is reflective of the ongoing digital transformation, where newer technologies such as smart manufacturing, IoT, and AI are being infused with big data capabilities to unlock novel industrial and economic potentials. These changes suggest a growing maturity and diversification of the big data paradigm.

Another significant aspect uncovered in the visualizations is the growing attention to societal and ethical dimensions of big data. Keywords such as "sustainable development," "digital revolution," and "social media" point to an increasing awareness of the broader implications of data-driven science. Big data is not only a technical instrument but also a cultural and ethical force. Its use in surveillance, political analysis, public opinion shaping, and even algorithmic bias presents profound challenges for democracy and equity [10]. Hence, alongside the technical evolution, a parallel strand of research is developing to critically examine the social ramifications of data-intensive practices.

The density visualization reinforces the thematic clustering seen in the network maps. The brightest region—dominated by "big data," "artificial intelligence," "data analytics," and "data science"—reflects the concentrated scholarly attention on these interrelated themes. The peripheral yet visible regions—such as "internet of things," "industry 4.0," "health care delivery," and "social sciences computing"—indicate emerging and specialized areas of application, suggesting that big data is actively diffusing into both domain-specific and interdisciplinary territories. This dispersion pattern is a key marker of scientific revolutions, as described by [21], where a novel paradigm gradually permeates and redefines established disciplines.

Collaboration trends, while not explicitly visualized in the presented figures, are often inferred from dense co-occurrence networks. The clustering of related keywords from various sectors—academic (e.g., "biology," "decision making"), industrial (e.g., "smart manufacturing," "automation"), and societal (e.g., "social media," "sustainable development")—points toward a Penta Helix model of scientific engagement, where academia, government, business, civil society, and media co-produce knowledge in a shared ecosystem. This suggests that the big data revolution in science is not only interdisciplinary but also multi-stakeholder in nature, relying on broad cooperation to realize its transformative potential.

Despite its promise, the bibliometric landscape also highlights the challenges and limitations in the current state of big data research. There is a tendency for highly cited topics to cluster around technologically dominant regions such as AI and genomics, potentially marginalizing other fields like humanities or traditional social sciences. Moreover, the absence of critical terms related to ethics, data privacy, and governance in the high-density zones may point to an underrepresentation of normative frameworks in mainstream big data science. This signals the need for a more balanced and inclusive research agenda that integrates technical excellence with ethical responsibility.

## CONCLUSION

This bibliometric study demonstrates that big data has become a central pillar in the transformation of modern scientific paradigms, reshaping methodologies, accelerating discovery, and fostering interdisciplinary integration. The analysis reveals that big data is deeply intertwined with emerging technologies such as artificial intelligence, deep learning, and the Internet of Things, while also playing a critical role in sectors like genomics, industrial automation, and healthcare. Temporal and density visualizations highlight its expanding influence and the evolution of research priorities over time, indicating a shift from foundational data management toward intelligent and autonomous systems. However, the study also identifies gaps in addressing ethical, social, and governance dimensions, suggesting the need for a more holistic approach to data-driven science. Overall, big data stands not only as a tool but as a transformative force driving a new era of scientific revolution—one that demands both technological innovation and critical reflection.

### REFERENCES


[1] I. Song dan Y. Zhu, "Big data and data science: what should we teach?," *Expert Syst.*, vol. 33, no. 4, hal. 364–373, 2016.

[2] G. George, E. C. Osinga, D. Lavie, dan B. A. Scott, "Big data and data science methods for management research," *Academy of Management Journal*, vol. 59, no. 5. Academy of Management Briarcliff Manor, NY, hal. 1493–1507, 2016.

[3] L. M. Ghiringhelli, J. Vybiral, S. V Levchenko, C. Draxl, dan M. Scheffler, "Big data of materials science: critical role of the descriptor," *Phys. Rev. Lett.*, vol. 114, no. 10, hal. 105503, 2015.

[4] S. González-Bailón, "Social science in the era of big data," *Policy & internet*, vol. 5, no. 2, hal. 147–160, 2013.

[5] Y. Demchenko, P. Grosso, C. De Laat, dan P. Membrey, "Addressing big data issues in scientific data infrastructure," in *2013 International conference on collaboration technologies and systems (CTS)*, IEEE, 2013, hal. 48–55.

[6] S. Halford dan M. Savage, "Speaking sociologically with big data: Symphonic social science and the future for big data research," *Sociology*, vol. 51, no. 6, hal. 1132–1148, 2017.

[7] C. C. P. Snijders, U. Matzat, dan U.-D. Reips, "' Big Data': big gaps of knowledge in the field of internet science," *Int. J. internet Sci.*, vol. 7, no. 1, hal. 1–5, 2012.

[8] J. Fan, F. Han, dan H. Liu, "Challenges of big data analysis," *Natl. Sci. Rev.*, vol. 1, no. 2, hal. 293–314, 2014.

[9] H. Guo, L. Wang, F. Chen, dan D. Liang, "Scientific big data and digital earth," *Chinese Sci. Bull.*, vol. 59, hal. 5066–5073, 2014.

[10] W. Wang dan E. Krishnan, "Big data and clinicians: a review on the state of the science," *JMIR Med. informatics*, vol. 2, no. 1, hal. e2913, 2014.

[11] X. Wu, X. Zhu, G.-Q. Wu, dan W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, hal. 97–107, 2013.

[12] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *J. Ind. Inf. Integr.*, vol. 6, hal. 1–10, 2017.

[13] R. Kitchin, "Big Data, new epistemologies and paradigm shifts," *Big data Soc.*, vol. 1, no. 1, hal. 2053951714528481, 2014.

[14] M. A. Waller dan S. E. Fawcett, "Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management," *Journal of Business logistics*, vol. 34, no. 2. Wiley Online Library, hal. 77–84, 2013.

[15] B. Efron dan T. Hastie, *Computer age statistical inference, student edition: algorithms, evidence, and data science*, vol. 6. Cambridge University Press, 2021.

[16] D. Romero *et al.*, "Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies," in *proceedings of the international conference on computers and industrial engineering (CIE46), Tianjin, China*, 2016, hal. 29–31.

[17] D. Donoho, "50 years of data science," *J. Comput. Graph. Stat.*, vol. 26, no. 4, hal. 745–766, 2017.

[18] R. T. Rust dan M.-H. Huang, "The service revolution and the transformation of marketing science," *Mark. Sci.*, vol. 33, no. 2, hal. 206–221, 2014.

[19] T. Hulsen *et al.*, "From big data to precision medicine," *Front. Med.*, vol. 6, hal. 34, 2019.

[20] A. Akundi, D. Euresti, S. Luna, W. Ankobiah, A. Lopes, dan I. Edinbarough, "State of Industry 5.0—Analysis and identification of current research trends," *Appl. Syst. Innov.*, vol. 5, no. 1, hal. 27, 2022.

[21] T. Kuhn, "A communicative theory of the firm: Developing an alternative perspective on intra-organizational power and stakeholder relationships," *Organ. Stud.*, vol. 29, no. 8–9, hal. 1227–1254, 2008.