**Ahlussunnah: Journal of Islamic Education**
https://ejournal.stitahlussunnah.ac.id/index.php/JIE

# Data Collection Methods in the Big Data Era: A Systematic Literature Review on Technical and Ethical Challenges

**Putri Adona[1], Zainur Huda[1], Zulfani Sesmiarni[2]**
[1]Institut Darul Qur'an Payakumbuh, Indonesia
[2]Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia
putriadona@idaqupayakumbuh.ac.id *

**Article Information:**
Received October 15, 2025
Revised November 28, 2025
Accepted December 28, 2025

**Keywords:** *Big data, systematic literature review, technical challenges, ethical issues, Islamic education*

**Abstract**

The Big Data era, characterized by the massive Volume, Velocity, and Variety of information, has revolutionized decision-making processes across various sectors. However, this paradigm shift has created significant methodological gaps, particularly related to population bias and the absence of standardized frameworks for validating non-probabilistic data representations. This study aims to bridge these gaps through a Systematic Literature Review, employing academic documentation and theoretical triangulation to synthesize both the challenges and solutions in the data acquisition phase. The findings identify three dominant data collection methods Web Scraping, Application Programming Interface, and the Internet of Things (IoT) as direct responses to the 3V characteristics of Big Data. Crucial insights reveal a persistent tension between the massive data volume and its validity, further complicated by technical risks (such as application programming interface rate limiting) and legal or ethical concerns (including compliance with Terms of Service and data privacy regulations). Research implementation in this era must therefore adopt a strategic framework, emphasizing essential practices such as personal identifiable information de-identification to ensure privacy rights and the application of Exponential Backoff techniques to overcome application programming interface quota limitations. This review presents a comprehensive synthesis of the pre-analysis phase of Big Data research, underscoring that the integrity and reliability of scientific findings in this era depend heavily on the adoption of rigorous methodological and ethical frameworks.

## INTRODUCTION

The influence of Big Data has become a transformative social phenomenon, generating positive and significant impacts on social welfare and various sectors of society. Over the past decade, the ability to collect, store, and analyze data on a massive scale has revolutionized how organizations and institutions operate across domains such as healthcare, finance, and government (Ibna & Nasution, 2024). Data is now viewed as a strategic asset capable of optimizing public services, enhancing data-driven decision-making, and creating added value through predictive analytics. This

**How to cite:** Adona, P., Huda, Z., Sesmiarni, Z. (2025). Data Collection Methods in the Big Data Era: A Systematic Literature Review on Technical and Ethical Challenges. *Ahlussunnah: Journal of Islamic Education*, *4*(3), 341-349.

**E-ISSN:** 2827-9573
**Published by:** The Institute for Research and Community Service

transformation has triggered a fundamental methodological urgency. As strategic decisions both governmental and corporate become increasingly dependent on conclusions derived from large-scale data analysis (Alfiandi & Ali, 2024), methodological failures in data collection may lead to serious social and economic consequences. Furthermore, the Big Data era inherently fosters interdisciplinary collaboration, requiring data scientists, business analysts, and software engineers to work together to fully leverage this data abundance (Ibna & Nasution, 2024).

In academic literature, Big Data refers to extremely large and complex datasets that exceed the capacity of traditional software to process and analyze effectively (Covey et al., 2024; Karthiyayini & Anandhi, 2024; Zeng & Yu, 2019). This phenomenon is commonly characterized by three key dimensions: Volume, referring to the sheer size of the data; Velocity, describing the speed at which data is generated and processed; and Variety, encompassing the diversity of data types and sources. The Variety dimension acts as a major driver of methodological shifts, as Big Data includes structured (SQL-based), semi-structured, and unstructured data such as text messages, location data, videos, and social media content (Rumata, 2016). Metode pengumpulan data konvensional jelas tidak memadai untuk menangani keragaman data tidak terstruktur yang mendominasi lanskap digital saat ini (Rumata, 2016). Conventional data collection methods are evidently inadequate for handling the predominance of unstructured data in today's digital landscape. Consequently, academic literature increasingly emphasizes the need for advanced research methodologies capable of managing and analyzing such data efficiently, often involving innovative analytical techniques and integrated approaches such as machine learning and artificial intelligence (Ibna & Nasution, 2024).

Despite the exponential growth in data volume, scholars highlight substantial methodological gaps. The abundance of data does not necessarily guarantee higher quality or representativeness compared to conventionally collected data (Rumata, 2016). A critical challenge lies in the inadequacy of sampling frames, raising fundamental questions about population representation such as whether active users on a specific social media platform can accurately represent the general population (Rumata, 2016). This issue is referred to as the Population Gap. Furthermore, a Methodological Gap exists due to the lack of standardized frameworks to validate representations and eliminate bias from non-probabilistic data. Another specific gap relates to the temporal dimension of Big Data analysis; neglecting time dynamics leads researchers to miss social or market changes that are essential to the Velocity characteristic of Big Data (Mach-Król & Hadasik, 2021).

This study aims to bridge these gaps through a Systematic Literature Review. Previous Systematic Literature Reviews have predominantly focused on analytical aspects of Big Data, such as data modeling associated with the Variety dimension and NoSQL databases (column-oriented, document-oriented, graph, and key-value) (Martinez-Mosquera et al., 2020; Tosi et al., 2024). While these studies have helped identify emerging research trends, the novelty of this study lies in providing a comprehensive synthesis of the pre-analysis phase, specifically data collection methods in the Big Data era. This research directly links the characteristics of Big Data (3V) with specific methodological choices Web Scraping, Application Programming Interface (API), and the Internet of Things (IoT) while offering a conceptual framework to address both technical risks (e.g., application programming interface rate limiting) and legal/ethical risks (e.g., compliance with Terms of Service and the Computer Fraud and Abuse Act that are rarely discussed together in a single methodological study.

Therefore, this research is crucial as the paradigm shift in data-driven research demands that scholars adopt fundamentally new data acquisition techniques. Without rigorous methodological guidelines on these non-conventional techniques,

researchers risk compromising the validity of their findings due to poor data quality or encountering unintended legal consequences from violating Terms of Service or cyber laws. This study provides a framework for risk mitigation, ensuring that Big Data research while potentially transformative across disciplines is conducted ethically, reliably, and with high scientific integrity.

As a Systematic Literature Review, this study offers a structured and replicable synthesis of empirical evidence (Granados-Duque & García-Perdomo, 2021). The findings are highly relevant for practitioners, policymakers, and research ethics committees (Institutional Review Boards). Moreover, they serve as a foundation for developing updated ethical and methodological guidelines aligned with the realities of Big Data, such as managing application programming interface rate limitations and addressing privacy concerns arising from IoT-based data collection (Pinto et al., 2024).

## METHODS

The type of research used in this study is a Systematic Literature Review. The Systematic Literature Review method was chosen because it allows researchers to examine journals, articles, books, or other documents systematically, with each step following established procedures and rules (Andriani, 2022). Systematic Literature Review is an essential tool for accurately and faithfully summarizing evidence, as it employs explicit and systematic methods to identify, select, and critically assess relevant research, as well as to collect and analyze data from the studies included in the review (Granados-Duque & García-Perdomo, 2021). In the context of Big Data methodology, which is both interdisciplinary and rapidly evolving, the use of Systematic Literature Review ensures that the understanding presented regarding data collection methods is supported by empirical evidence that has been gathered and evaluated transparently (Tosi et al., 2024). This approach also helps to mitigate the risk of narrative bias, which often occurs in conventional literature reviews (Covey et al., 2024; Karthiyayini & Anandhi, 2024; Purssell & McCrae, 2020; Shen & Smith, 2024).

The primary data collection method employed in this Systematic Literature Review study is documentation, namely the examination of available academic documents such as local and international journals, as well as articles related to the implementation of Big Data in research (Andriani, 2022). The data collection protocol follows a systematic sequence of steps. These steps include the use of specific keywords in the search form across academic database repositories (for instance, journal indexing platforms). After the initial identification, a process of collecting selected journals was carried out to answer the research questions and to ensure the quality and availability of the required information. To enhance the internal validity of the synthesis, a Quality Assessment process was implemented. This Quality Assessment process aims to screen documents and ensure that only the most credible and relevant articles are considered in the analysis, thereby reducing the risk of selection bias.

The data analysis in this study employs triangulation techniques to enhance the validity and reliability of the findings (Engkizar et al., 2023, 2025). In the context of qualitative research or literature reviews, data triangulation may involve combining literature review results and secondary data (Collier, 2020; Engkizar et al., 2023; 2025; Lemon & Hayes, 2020; Poggi, 2024; Wee et al., 2021). Specifically, Theory Triangulation was applied, which requires researchers to exercise expert judgment when comparing empirical findings from the collected data with various relevant theoretical perspectives. This triangulation is particularly important when discussing Big Data methodologies, as empirical findings derived from Big Data are often susceptible to platform bias (Engkizar et al., 2022; Rumata, 2016). By comparing the

findings on data collection methods with established theories in social or technological research, triangulation ensures that the results are not based on a single data source or methodological approach (Engkizar et al., 2025). This process is expected to produce a deeper and more comprehensive understanding, thereby increasing public trust and acceptance of the research outcomes.

## RESULT AND DISCUSSION

### Classification of Big Data Sources and Types

Big Data is primarily defined by the challenges arising from its three core characteristics: the massive Volume of data, the rapid Velocity at which it is generated, and the diverse Variety of its forms. Among these, data Variety plays a critical role in determining the choice of data collection methods. Big Data can generally be classified into three main types: structured data (for instance, SQL-based datasets ready for direct processing), semi-structured data (structured data that is not organized within conventional database models), and unstructured data (uncategorized information such as text messages, videos, or social media content) (Rumata, 2016). The selection of acquisition methods whether Web Scraping, application programming interface, or Internet of Things (IoT) is a direct response to these 3V characteristics. Specifically, Web Scraping and application programming interface techniques primarily address the Variety dimension by handling structured and unstructured data from the web, while IoT methods are designed to manage Velocity, as they enable the generation of real-time data streams.

### Data Exploration through Web Scraping

Web Scraping serves as a fundamental and essential process for collecting Big Data datasets, which often constitute the backbone of large-scale data analysis, Machine Learning (ML), and Artificial Intelligence (AI) model training. This technique allows academic researchers to monitor market trends or conduct large-scale studies by acquiring publicly available content from the internet. The technical implementation of web scraping varies depending on the complexity of the targeted web pages (vanden Broucke & Baesens, 2018; Zhao, 2022). It can range from simple text pattern matching to the use of XPath locators, particularly for pages with minimal JavaScript markup. While its flexibility makes web scraping a powerful tool for data acquisition, it also introduces inherent challenges related to data quality and legal compliance, which must be carefully managed.

### Data Integration via Application Programming Interface

The use of Application Programming Interface provides a more structured and formalized channel for collecting Big Data, especially from social media platforms or business data sources. Large-scale public data, such as Google search trends, can be retrieved using application programming interfaces. In the business domain, the development of rest application programming interfaces has facilitated centralized social media integration, enabling small and medium enterprises (SMEs) to efficiently collect customer reviews and analyze engagement data from multiple platforms. The main advantage of using application programming interfaces for research lies in their ability to deliver relatively clean and structured datasets. However, this method also has notable limitations it is entirely dependent on the availability, policies, and technical constraints set by the platform providers (such as social media companies), which are often subject to sudden changes or access restrictions.

### Real-time Data Collection from the Internet of Things (IoT) and Sensors

The use of Internet of Things (IoT) and sensor-based data collection represents the most prominent example of the Velocity and Volume characteristics of Big Data. IoT systems, through the integration of physical devices, software, and sensors, enable large-scale and distributed real-time data acquisition. This sensor data, which is often structured, can be processed using Big Data technologies to generate

strategic insights across sectors such as healthcare, finance, and business. Although IoT-based data collection provides unparalleled real-time operational insights, it also introduces several challenges most notably the high infrastructure costs required to store and process massive data streams, as well as critical privacy and security concerns, particularly when the collected information contains personal or sensitive data (Pinto et al., 2024).

A comparison of these three major data collection methods Web Scraping, application programming interface integration, and IoT-based acquisition along with their respective characteristics and challenges, is presented in the following table.

**Table 1. Comparison of Big Data Data Collection Methods in Research**

| Method | Main Data Characteristics | Academic Advantages | Key Challenges |
|---|---|---|---|
| Web Scraping | Unstructured, Semi-Structured (High Volume) | Data flexibility, Access to massive public content. | Legal issues (ToS, Copyright), IP blocking risk, Noisy or unclean data. |
| Application Programming Interface | Structured, Semi-Structured (High Velocity) | Clean data, Authorized access, Easy automation. | Rate limiting, Usage quotas, Platform dependency. |
| IoT / Sensor | Real-time, Structured (Extreme Velocity and Volume) | Real-time operational insights, Large-scale data collection. | Infrastructure cost, Privacy and data security issues |

**Access Restrictions and Resource Management**

One of the major technical barriers in academic Big Data collection lies in the implementation of rate limiting and quota management by application programming interface providers. These mechanisms are designed as defensive strategies to protect system infrastructure from excessive requests, ensure consistent availability, and prevent Denial of Service (DoS) attacks. For instance, the U.S. Patent and Trademark Office (USPTO) restricts metadata retrieval to five million calls per week. While such measures are crucial for maintaining system stability, they directly constrain academic research that requires high data volumes within a short period to capture the desired Velocity characteristic of Big Data. Researchers without access to premium application programming interfaces or adequate computational resources (for example, to overcome bulk download limitations) are often forced to narrow their research scope, leading to significant challenges in achieving the full analytical potential of Big Data. Thus, the need for high application programming interface availability inadvertently results in strict quota and call limitations that restrict the ambition and scalability of research efforts.

**Ethical and Legal Challenges in Web Scraping**

Ethical and legal challenges have become increasingly complex, particularly following global privacy scandals such as the Cambridge Analytica case, which prompted many social media platforms to heavily restrict or even shut down application programming interface access. As an inevitable consequence, web scraping has often been perceived as a "necessary evil" for acquiring empirical digital data. However, this approach carries substantial legal risks. The legality of web scraping depends heavily on the methods employed and whether the practice violates a website's Terms of Service (ToS). Violations of ToS especially when involving personal data collection or causing server overload may constitute unauthorized access and potentially breach the Computer Fraud and Abuse Act (CFAA) in the United States. This creates an ethical–methodological dilemma where researchers may feel compelled to employ legally ambiguous methods simply to obtain the public data required for large-scale Big Data analysis. Therefore, researchers must exercise great caution and uphold ethical integrity in all data acquisition practices.

**Table 2. Ethical and Legal Risk Mitigation Strategies in Big Data Collection**

| Risk Challenge | Risk Description | Recommended Mitigation Strategies |
|---|---|---|
| ToS/CFAA Violation | Unauthorized or excessive access to servers. | Always comply with robots.txt and platform-imposed rate limits; use proper User Agents. |
| Data Privacy Issues | Collection of personal or identifiable information without consent. | Apply anonymization and aggregation techniques, and focus only on data explicitly permitted for collection; comply with local data privacy regulations. |
| Data Quality (Cleaning) | Scraped data often contains noise and unwanted text elements. | Implement text pattern matching and rigorous data cleaning procedures before analysis. |

**De-identification of personally identifiable information and application programming interface reset management**

Addressing the technical and ethical challenges in Big Data collection requires strategic implementation, particularly in managing sensitive data and utilizing application programming interfaces. In the ethical context, the concept of "data reset" is operationally implemented through de-identification of Personally Identifiable Information. De-identification is an essential process that involves removing, masking, or replacing sensitive personally identifiable information such as names, email addresses, or genderfrom raw datasets. The primary goal is to minimize the risk associated with handling personal data while maintaining the usefulness of the data for analytical purposes. Common de-identification techniques include redaction or masking (concealing parts of the data), tokenization or pseudonymization (replacing real identities with artificial tokens), and bucketing (grouping specific values into broader categories). A carefully executed de-identification process which includes auditing, personally identifiable information detection, method selection, and result validation effectively enforces the data subject's right to be forgotten, as mandated by modern data privacy regulations.

From a technical perspective, the management of application programming interface quota resets plays a crucial role in large-scale research. application programming interface rate limits are typically measured per minute (requests per minute/RPM), per token (tokens per minute/TPM), or per day (requests per day/RPD), with quota resets usually scheduled at midnight Pacific Time or weekly, for instance, at midnight UTC on Sundays. When these limits are exceeded, the application programming interface returns an HTTP 429 (Too Many Requests) status code, often accompanied by a Retry-After header indicating when the client may attempt the next request. Implementation strategies to handle these constraints include exponential backoff, in which the system waits for an initial delay period (e.g., one second), retries the request, and doubles the delay after each failure until success or the maximum number of retries is reached. Additionally, monitoring reset headers such as X-RateLimit-Reset or Retry-After is essential to determine precisely when the quota will be reset, ensuring that the data collection workflow remains efficient and sustainable.

Data quality serves as a fundamental prerequisite for the validity of research findings. Although the volume of data collected through Big Data techniques can be immense, methods such as web scraping often yield unstructured data that contain irrelevant text, requiring an intensive data cleaning process. More fundamentally, one of the greatest challenges in Big Data research lies in the limitations of sampling frameworks, which can introduce significant representational bias (Rumata, 2016). Failure to validate the representativeness of non-probabilistic data often drawn from self-selected samples of specific platform users may lead to misleading conclusions,

even when advanced analytical techniques are applied. Therefore, researchers must carefully address representation issues by critically analyzing the inherent platform bias and aligning the dataset with the research context to ensure that findings remain credible and representative.

**Hybrid Approaches and Data Triangulation**

To address methodological complexity, adopting a hybrid approach that integrates Big Data methods with conventional research techniques is highly recommended. This approach effectively combines the scale (Volume and Velocity) provided by massive quantitative datasets with the depth and contextual understanding (Veracity) offered by qualitative methods. To enhance validity, research should employ data triangulation techniques. For instance, findings derived from Big Data analysis such as those obtained via application programming interfaces or market data scraping can be validated and enriched with qualitative data from in-depth interviews conducted with practitioners or policy executives. This synthesis ensures that research does not merely identify what happens based on large-scale data but also understands why it happens within its qualitative context, thereby producing a more comprehensive and balanced understanding.

## CONCLUSION

This Systematic Literature Review found that data collection methodologies in research during the Big Data era have become polarized into three main pathways Web Scraping, application programming interfaces, and IoT each representing a technical response to the characteristics of Volume, Velocity, and Variety. The key finding highlights a significant tension between massive data Volume and its Validity, further exacerbated by the Population Gap caused by limitations in non-probabilistic sampling frameworks. Therefore, research in this era must master both technical risk management such as navigating application programming interface rate limiting and legal or ethical risks, including compliance with Terms of Service and data privacy ethics, to ensure that research remains sustainable and legally sound. Compliance implementation should include the de-identification of Personally Identifiable Information to protect privacy rights and the adoption of Exponential Backoff techniques to manage application programming interface quota limitations.

As a study employing the Systematic Literature Review method, this research has inherent limitations. The primary weakness lies in the potential risk of selection bias, as the analysis outcomes depend entirely on the predetermined quality assessment criteria and keyword parameters used to filter documents. Although the Systematic Literature Review protocol was rigorously followed, the possibility of publication bias namely the tendency to publish studies with significant or positive results may still influence the availability of reviewed literature. Furthermore, since this study is based on secondary data synthesis, the findings may not fully reflect the real-time implementation and operational challenges of Big Data research, which can only be deeply explored through primary case studies or field data collection.

Based on these findings and limitations, future research should focus on developing a more explicit methodological framework to address the representation gap in Big Data. This may include standardized guidelines on how to integrate non-representative data obtained from social media platforms with representative data from traditional sources. Moreover, further empirical studies are needed to explore the infrastructural cost implications both for IoT sensor data storage and bulk application programming interface data processing as well as to examine legal compliance challenges faced by academics when employing high-capacity data collection methods. Finally, subsequent research should directly test how the characteristics of Big Data (Volume, Velocity, and Variety) can be optimally utilized to address specific research problems while developing mitigation mechanisms for

the methodological limitations that arise.

## REFERENCES


Alfiandi, A., & Hapzi Ali. (2024). Pengaruh Big Data, Pengambilan Keputusan dan Strategis Pemerintah terhadap Kesejahteraan Sosial. *JIM: Jurnal Ilmu Multidisplin*, *2*(4), 384–390. https://doi.org/10.38035/jim.v2i4.430

Andriani, W. (2022). Penggunaan Metode Sistematik Literatur Review dalam Penelitian Ilmu Sosiologi. *Jurnal PTK Dan Pendidikan*, *7*(2). https://doi.org/10.18592/ptk.v7i2.5632

Collier, J. E. (2020). Data Screening, Assessing Reliability, Validity, and Identification. In *Applied Structural Equation Modeling Using AMOS* (pp. 17–35). Routledge. https://doi.org/10.4324/9781003018414-2

Covey, B., Theriault, J., O'Blenis, P., & Hart, D. (2024). Big Data Processing Platform for Large Acoustic Datasets and Complex Data Pipelines: Leveraging Cutting Edge Open Source Software to Build Scalable Cost Effective Solutions. *OCEANS 2024 - Halifax*, 1–6. https://doi.org/10.1109/OCEANS55160.2024.10754471

Engkizar, E., Jaafar, A., Hamzah, M. I., Fakhruddin, F. M., Oktavia, G., & Febriani, A. (2023). Changes in Students' Motivation to Memorize the Quran: A Study at Quranic Higher Education Institutions in Indonesia. *International Journal of Islamic Studies Higher Education*, *2*(3), 240–258. https://doi.org/10.24036/insight.v2i3.240

Engkizar, E., Jaafar, A., Hamzah, M. I., Langputeh, S., Rahman, I., & Febriani, A. (2025). Analysis Problems of Quranic Education Teachers in Indonesia: A Systematic Literature Review. *International Journal of Islamic Studies Higher Education*, *4*(2), 92–108. https://doi.org/10.24036/insight.v4i2.232

Engkizar, E., Jaafar, A., Muslim, H., Mulyadi, I., & Putra, Y. A. (2025). Ten Criteria for an Ideal Teacher to Memorize the Quran. *Journal of Theory and Research Memorization Quran*, *1*(1), 26–39. https://joqer.intischolar.id/index.php/joqer

Engkizar, E., Jaafar, A., Taufan, M., Rahman, I., Oktavia, G., & Guspita, R. (2023). Quran Teacher: Future Profession or Devotion to the Ummah? *International Journal of Multidisciplinary of Higher Education (IJMURHICA)*, *6*(4), 633–644. https://doi.org/10.24036/ijmurhica.v6i4.321

Engkizar, E., Sarianti, Y., Namira, S., Budiman, S., Susanti, H., & Albizar, A. (2022). Five Methods of Quran Memorization in Tahfidz House of Fastabiqul Khairat Indonesia. *International Journal of Islamic Studies Higher Education*, *1*(1), 54–67. https://doi.org/10.24036/insight.v1i1.27

Granados-Duque, V., & García-Perdomo, H. A. (2021). Systematic review and meta-analysis: Which pitfalls to avoid during this process. *International Braz J Urol*, *47*(5), 1037–1041. https://doi.org/10.1590/S1677-5538.IBJU.2020.0746

Ibna, A. Z., & Nasution, M. I. P. (2024). Implikasi Penggunaan Basis Data dalam Era Big Data. *Jurnal Sains Student Research*, *2*(4), 255–265.

Karthiyayini, J., & Anandhi, R. J. (2024). To Analyze the Various Machine Learning Algorithms That Can Effectively Process Large Volumes of Data and Extract Relevant Information for Personalized Travel Recommendations. *SN Computer Science*, *5*(4), 336. https://doi.org/10.1007/s42979-024-02667-x

Lemon, L., & Hayes, J. (2020). Enhancing Trustworthiness of Qualitative Findings: Using Leximancer for Qualitative Data Analysis Triangulation. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2020.4222

Mach-Król, M., & Hadasik, B. (2021). On a certain research gap in big data mining for customer insights. *Applied Sciences (Switzerland)*, *11*(15). https://doi.org/10.3390/app11156993

Martinez-Mosquera, D., Navarrete, R., & Lujan-Mora, S. (2020). Modeling and

management big data in databases-A systematic literature review. *Sustainability (Switzerland)*, *12*(2). https://doi.org/10.3390/su12020634

Pinto, G. P., Donta, P. K., Dustdar, S., & Prazeres, C. (2024). A Systematic Review on Privacy-Aware IoT Personal Data Stores. *Sensors*, *24*(7), 1–23. https://doi.org/10.3390/s24072197

Poggi, C. (2024). Mammographic Findings, Recalls and Triangulation. In *Breast Imaging Techniques for Radiographers* (pp. 115–123). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63314-0_13

Purssell, E., & McCrae, N. (2020). Reviewing Qualitative and Quantitative Studies and Mixed-Method Reviews. In *How to Perform a Systematic Literature Review* (pp. 113–121). Springer International Publishing. https://doi.org/10.1007/978-3-030-49672-2_9

Rumata, V. M. (2016). The Opportunities and Challenges of The Big Data Implementation in Social Science Reseacrh: A Literature Review. *Jurnal Penelitian Komunikasi Dan Opini Publik*, *20*(1), 123337.

Shen, Y., & Smith, T. (2024). Qualitative research in finance: a systematic literature review. In *How to Conduct Qualitative Research in Finance* (pp. 17–45). Edward Elgar Publishing. https://doi.org/10.4337/9781803927008.00009

Tosi, D., Kokaj, R., & Roccetti, M. (2024). 15 years of Big Data: a systematic literature review. *Journal of Big Data*, *11*(1), 1–45. https://doi.org/10.1186/s40537-024-00914-9

vanden Broucke, S., & Baesens, B. (2018). From Web Scraping to Web Crawling. In *Practical Web Scraping for Data Science* (pp. 155–172). Apress. https://doi.org/10.1007/978-1-4842-3582-9_6

Wee, C., Hwang, J., Kwon, E., & Shin, K. (2021). Reliability and Validity Study of Family Supportive Supervisor Behavior(FSSB). *The Korean Data Analysis Society*, *23*(4), 1965–1982. https://doi.org/10.37727/jkdas.2021.23.4.1965

Zeng, J., & Yu, H. (2019). Effectively Unified optimization for Large-scale Graph Community Detection. *2019 IEEE International Conference on Big Data (Big Data)*, 475–482. https://doi.org/10.1109/BigData47090.2019.9005481

Zhao, B. (2022). Web Scraping. In *Encyclopedia of Big Data* (pp. 951–953). Springer International Publishing. https://doi.org/10.1007/978-3-319-32010-6_483