Current Challenges of Big Data Quality Management in Big Data Governance: A Literature Review

Yunusa Adamu Bena[1] Roliana Ibrahim[2] Jamilah Mahmood[3]

[1] Faculty of Engineering, Kebbi State University of Science and Technology, Aliero.
bena@graduate.utm.my
[2] Faculty of Computing, Universiti Teknologi Malaysia.
roliana@utm.my
[3] Faculty of Computing, Universiti Teknologi Malaysia
Jamilah.mahmood@utm.my

Abstract. Advent of big data has changed the way data-driven organizations collect, process and analyses the data. However, considering how data is generated from heterogenous sources, which gives the ample amount of available information to be generated dramatically. This has forced the data-driven organizations to invest more in technologies that will determine the quality of their data through an effective data governance process. However, the lingering issues on the challenges of big data quality management in big data governance remains an alarming issue that requires much attention from the researchers and practitioners, because studies on this area are still in its infancy stage. Therefore, the main objective of this study is to identify the current challenges of data quality management in big data governance from previous studies and address them accordingly. To achieve this objective, we conducted a literature review on the recent studies that address these challenges. This study has also examined the available data quality dimensions used by the individual studies. Moreover, a total of 41 recent studies published between (2020-2023) were utilized by this review. The results of our findings including the authors recommendations on these challenges, we proposed a framework for big data quality management based on ISO 8000-61 for data quality reference model, ISO 8000-62 for organizational process maturity assessment, and "Plan-Do-Check-Act" approach for big data quality assessment. Subsequently, we proposed affordability, Operationability, and data literacy to be included in the big data quality dimensions.

Keywords: Big Data, Big Data Quality Management, Big Data Quality Dimensions, Big Data Quality Challenges, Big Data Governance.

1   Introduction

The advent of big data has created more data quality management challenges [1] . Because there is a significant need to assess the quality of the collected data before taking a decision [2]. Similarly, the data that does not fulfil the quality standards will lead to inaccurate organizational decision-making [3]. Nowadays researchers, practitioners and analyst are focusing more on identifying the major challenges of big data quality

management and possible causes. According to Mohammad [4], veracity as one of the characteristics of big data, is among the challenges that involves big data quality challenges, therefore if data fails its quality standards, the entire analytical results will be meaningless and may results in creating inaccurate decision-making. Moreover, studies on theoretical aspect of big data quality management in data governance still requires more attention. Because data-driven organizations are complaining of being distracted and discouraged from extracting the values from the big data due to the quality challenges [5]. Furthermore, based on the recent survey conducted by [6] , about 80-87% of the big data projects could not generate a sustainable output due to the issues of data quality management. Therefore, the fact about data nowadays is that the more you invest in data quality management technology, the more your organization will have a clear and transparent decision-making process. When the data quality is not properly managed, it can lead to poor inadequate, inconsistent, and erroneous decisions-making which can further escalate the computational cost and decline in profits [7].

To address the aforementioned issues, a literature review on the recently published studies were conducted to explore the researchers and practitioners' perception regards to the possible root causes of these challenges and recommendations on how to mitigate these challenges. Since the domain area of this study is on big data, all the studies from different contexts were included. Hence, most of the data-driven organizations do generate their data from heterogenous sources. Similarly, the data quality dimensions proposed or used by the individual authors has also been examined, so as to enable our study to identify the currently used dimensions and their limitations. According to [8], it is important to distinguished between the traditional and big data quality dimensions, because it will guide the researchers to determine whether they are applicable or not applicable to big data. The study has yet integrated the aspect of big data governance, because according to DAMA-DMBOK [36], (i.e., Data Management Body of Knowledge), data governance is the core component of data management that accommodated other 9 disciplines, which data quality management is among the components.

This paper is organized with the following structures: The introduction section that highlights the problems, aim, and objectives of this paper. The literature review that presented the results of our findings, the proposed framework for data quality management in big data governance, and lastly the conclusion that contains the limitations and recommendations of this research.

1.1 Literature review

The advent of big data and its four (4) main characteristics (i.e., Volume, Velocity, Variety, and Veracity), requires modern and advanced technologies and techniques for ensuring the data quality when capturing, processing, storing, distribution and analysis of an organization's data [6]. Similarly, considering the proliferation of data due to the rapid replication of nowadays technologies, data-driven organizations are now becoming keen to invest in data quality management tools and technologies [5]. As the study on this context appears to be new to the researchers and practitioners, there are no much contributions on the definitions of big data quality, thus majority of the studies commonly used "Fitness for Use", which was first coined by Juran [9]. However, some

scholars consider the definition to be based on the context or area of application. For example, Fadlallah [7] defined big data quality as the process of assessing the data quality level in the context of big data. While [10] considered big data quality as the ability of a data to meet users' requirements.

The Impact of Big Data Quality Management. Addressing the challenges and impact of big data quality management can never be overemphasized, because it has a tremendous influence on the viability of modern organizations [11,12]. Nowadays, modern organizations are becoming more conversant with the impact of big data quality management as one of the ways for improving the organizations decision-making, because data quality is considered as a fundamental components in the generic life cycle of big data processing in data warehouses [5]. According to [7], big data quality management is a flexible solution that enables a supports for big data quality assessment in different domains areas, and it facilitates the organizational planning and execution of big data quality assessment, and efficient big data governance at any phase in the big data life cycle. Among its impact to the modern organizations, it brings about significant value to organizational processes if supported and managed by a good level of data quality, which will result to a better organizational decision-making and competitive advantage, optimizes organizational decision-making processes at strategic, tactical and operational levels in the organization [13]. Also help in ensuring successful management of the Internet of Things (IoT) and smart connected products data [10]. Similarly, big data quality management offers a quantifiable analysis on the accuracy and reliability of data [14], contribute to the improvement of organizational decision-making [1], help in identifying the instances of data quality deficiencies, and segregate data of high quality from worse quality[8].

Summary of Studies and Issues Addressed on Big Data Quality Management. To identify the recent studies of the study's domain area, this study solely limited on the recent studies published between (2020-2023). The rationale behind is to explore the recent issues addressed in this area. The table below presents the 36 studies used for achieving the objective of this study.

**Table 1.** Studies on Challenges in Big Data Quality Management.

Authors (2)	Issues Addressed
Elouataoui [14]	Data quality management system that supports large scale data quality detection and repairs.
Zibak [15]	Threat intelligence quality dimensions.
The [16]	Sensor data quality problems.
Goknil [17]	Data quality in data-centric CPS/IoT industrial applications.
Byabazaire [2]	How the aspects of data quality affect the performance of big data model and how this affects subsequent stages.
Li [23]	Quality-aware utilization of spatial IoT data.

Table 2. (*Continued*).

Authors (2)	Issues Addressed
Hossen [11]	Challenges of maintaining big data quality in e-business.
Wook [19]	Traits of big data and the dimensions of data quality.
Xiao [20]	Challenges on integration of big data management and quality management.
Zhang [21]	Data traceability method to improve data quality in a big data environment.
Molinari [12]	The importance of data quality in the context of big data Healthcare.
El-wahab [22]	Analysis for improving the big data quality in the enterprise context.
Li [23]	Quality-aware utilization of spatial IoT data.
Faroukhi [24]	The importance of considering data quality and security requirements in Big Data projects.
Hattawi [25]	The challenges in measuring the quality of big data.
Alkatheeri [26]	Impact of big data quality and management.
Wang [27]	Evaluating and detecting the big quality management issues.
Ramasamy [8]	Big data quality dimensions and recent challenges on quality management in the organizations.
Simard [28]	Method to classifying data quality for decision-making under uncertainty.
Talha [29]	Challenges of merging and integrating access control policies in Big Data
Fadlallah [30]	Limitations of existing solutions supporting the big data quality assessment process.
Fadlallah [7]	Existing context-aware data quality assessment solutions.
Reda [13]	Data quality dimensions metrics, tools, and frameworks for data quality assessment.
Chirkova [3]	Challenges of assessing big data quality.
Zou [31]	Big data quality assessment model for unstructured data.
Alwan [32]	Data quality management challenges in large-scale cyber-physical systems, and data quality challenges in smart cities.
Gu [5]	Data quality management system that supports large scale data quality detection and repairs.
Zhang [33]	Data quality management on the internet of things (IoT).
Kim [10]	Organizational process maturity model for IoT data quality management.
Mohammad [4]	Approach for improving data quality from big data.
Yee [34]	Measurement and prediction of data quality in big data.
Cisneros [35]	Scaling big data quality-aware query processing.
Feliks [1]	Factors affecting the effective use of big data.
Escobar [6]	Big data challenges in manufacturing industries.

The Current Challenges of Data Quality Management in Big Data Governance.
After summarizing the studies and issues addressed by the authors in table 1, This section presents the findings on the current challenges of big data quality management in big data governance. However, the results of our findings were presented based on the authors perception and context of the study. Many researchers have contributed immensely to see how the existing challenges of big data quality can be mitigated. It is obvious that big data governance deals with organization, security, management, and utilizing techniques and technologies that guarantee data is correct, consistent, and available to authenticated users [23]. As stated earlier in the introduction, data governance is the core component of data management, that accommodates the other 9 disciplines as depicted in figure 1.
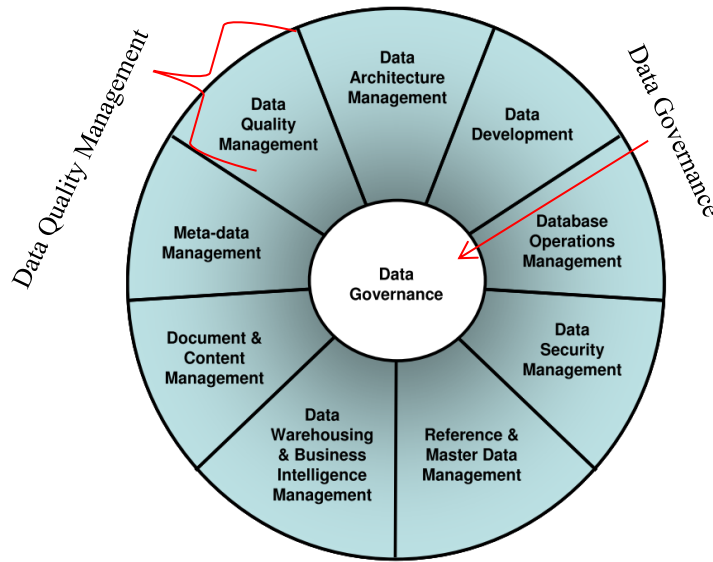
Fig. 2. Data Governance as the core component of data management source: (DAMA Dama-International [36].

As mentioned earlier, in order to identify the challenges of big data quality, the aspect of data governance has to be put in place. Because the issues of big data quality can never be solved without identifying the challenges from data governance. In this regard many scholars contributed to identifying some of the challenges from data governance that contributed to the major challenges of big data quality, as shown in table 2.

Table 2. Current Challenges of Big Data Quality Management in Big Data Governance

Challenge (s)	Description	Reference
Lack of stakeholder's engagement.	Financing the big data governance and data quality management programme are still lacking by the organizational stakeholders, which leads to the resistance to adoption and ineffective implementation.	[37] and [38]
Big data governance and data quality management literacy.	Data-driven organizations are lacking awareness on big data governance and data quality literacy programme.	[38, 37,28]
Big data quality dimensions.	Lack of unified standards, technologies, and guidelines on using big data quality dimensions	[14,16,17]
Resistance to change	Many organizations cannot cope with the dynamic and rapid replication of nowadays technologies.	[38, 37,22]
Limited ability to enforce policies and compliance.	Issues on enforcing government and organizational policies in an organization with multiple departments.	[38, 37,10]
Lack of expertise	Due to the advent of AI, IoT and machine learning technologies, many organizations are lacking expertise and operational knowledge for managing the organization's data.	[28, 20,17]
ISO standardization	Lack of well-defined ISO standards and automated tools and technologies for evaluating and validating the processes of big data quality management	[14,16,17]
AI, IoT, SCPs, and Sensors.	Electromagnetic interferences, high temperature, loss of connections, and geographical locations can lead to the failure of data quality.	[17] [11]
Heterogeneity	The heterogenous nature of big data can lead to the compromise of integrity of data.	[40, 20, 22, 25, 21]
Unclear methodological approaches.	Lack of clear and transparent process of evaluating the relevancy of organizational data.	[1,41, 8]
Privacy and security	Issues of information leakage and data compromise within and outside the organization remains the major challenges of big data quality.	[29, 26, 25,21, 40]
Big data quality management framework.	Lack of data quality frameworks that contains processes and guidelines on how quality of data will be maintained.	[20, 14, 35, 29, 25]

The challenges presented in table 2, were identified from the recent studies. However, there are some authors whose argued that, when identifying the challenges of big data quality management in big data governance, there is need for focusing more on quality

dimensions [19][2] and [14]. As such, this study has yet identified the currently used dimensions on big data quality management, so as to cover the entire possible challenges in the domain area.

Big Data Quality Dimensions. Big data quality dimensions are the measurement or characteristics that researcher and practitioners use in assessing the quality of data within a big data environment [8,5,14,23]. Moreover, these dimensions are considered critical when managing and analyzing big data quality in an organization, because they serve as yardstick for ensuring the data's trustworthiness and fitness when making informed decisions. Figure 2 presents the big data quality dimensions used by the authors and are utilized based on the context of the studies.

Fig. 2. Data Quality Dimensions and their frequencies.

Based on the identified data quality dimensions from the literature, the results show that accuracy, completeness, consistency, reliability, and validity are the known to be the right dimensions to be considered when evaluating the quality of data. However, these dimensions can be best described as the commonly used dimensions for evaluating both traditional and modern data quality management.

However, to be able to address the challenges of big data quality management in big data governance, there is need for incorporating the big data governance challenges into big data quality dimensions. As such our study has proposed and recommended three (3) additional big data quality dimensions, which according to our findings will improve the process of big data quality management and mitigate the current challenges. The additional big data quality dimensions are as follows:

1. *Affordability:* Ability to invest and deploy in the tools and technologies on big data quality management.

2. *Data literacy:* Educating, sensitizing, and updating the big data steering committee with the recent tools and technology paradigm on big data quality management.
3. Operationability: Technical and operational knowledge on AI, IoT, smart connected products (SCPs) and other big data quality assessment tools.

Because most of the modern organizations that are dealing with big data could not afford to deploy big data quality management tools, while some of the data governance teams are not conversant with the modern technology paradigms.

Authors recommendations on how mitigate the current challenges. Many contributed to highlighting some recommendations that can be used to mitigate the existing challenges of big data quality management in big data governance. According to [17, 40, 24,8,29] the existing challenges can be mitigated through developing a holistic framework that includes components with data monitoring and visualization techniques, data cleaning, data repair and data quality standards. While in an organization dealing with AI, IoT, and smart connected products, [18, 12, 11, 22], recommended on deploying artificial intelligence and machine learning technologies, and developing a framework with different layers on how to tracks device locations, network connection for data transmission, and monitoring data streams. Similarly, [20] argued that the only solution to the existing challenges of big data quality management in data governance is proving a data literacy programme across the organization. Moreover, Wang [27] argued that applying parallelization and entity detection algorithms is one of the solution for overcoming the current challenges of big data quality management. Though, Simard [28] seen using classification method to represent the state of data quality is the best way to mitigate these challenges. According to [4], solving the veracity nature of big data will be the initial solution to big data quality management in big data governance, because it affects the correctness and reliability data quality. Though, Cisneros [35] and Feliks [1]argues that integrating query processing techniques with data quality management methodologies can be the best option.

Proposed Framework for Data Quality Management in Big Data Governance. Many studies contributed to highlighting some recommendations that can be used to mitigate the current challenges of big data quality management in big data governance. However, this study considered the recommendations made by [17, 40, 24,8,29]. Therefore, this study proposed a big data quality management framework that integrates the components of big data governance. The proposed framework is developed based on ISO 8000-61, for data quality management process reference model and ISO 8000-62 for organizational process maturity assessment. Because these ISO standards help to carry out data quality assessment based on the documented methods when assessing the quality of organization's information on data. Similarly, the ISO 8000 family helps in specifying the processes that are required for big data quality management. Hence, it serves as reference model for assessing the maturity of an organization's data quality management capability and involves organization's stakeholders in the data quality

management processes. Furthermore, the "Plan-Do-Check-Act" approach is also utilized in the proposed framework as shown in the figure below.
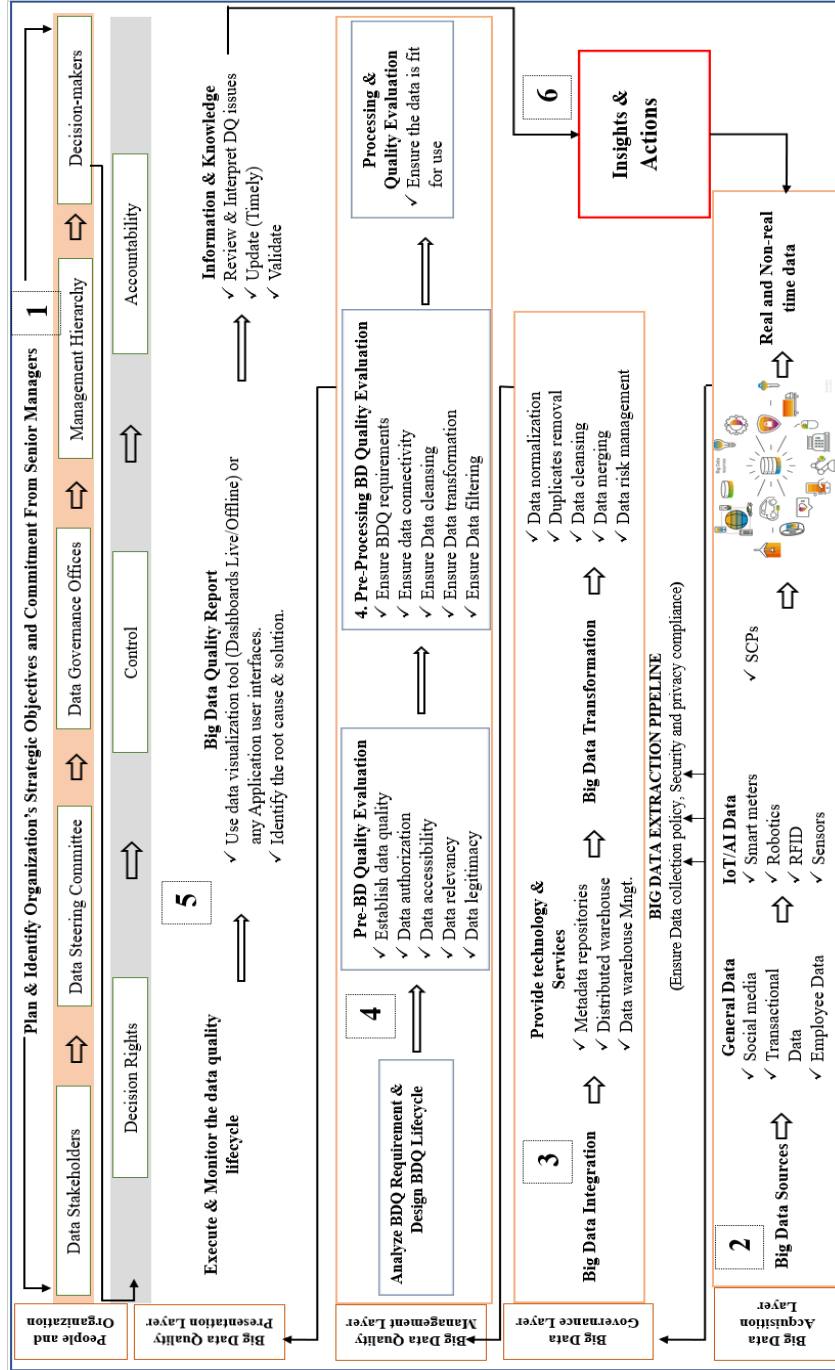
Fig. 3. Proposed Framework for Big data quality management (Authors own creation).

Description of the Proposed Big Data Quality Management Framework. The proposed framework contains four (4) layers and six (6) major activities, and here is the detailed description of the components.

1. *Big Data Acquisition Layer.* Is considered as the first layer that describes the big data sources of data generated by an organization. The big data sources may be from the general data which include general data (e.g., data from social media, employees, or transaction), IoT/AI, (e.g., Sensors, RFID, Robotics, and smart connected products). Moreover, these data could be in real-time or otherwise.
2. *Big Data Governance Layer.* Before data is collected or generated by an organization, it has to be ensured that government and organization's policies, security, and privacy are complied with before commencing the integration processes. This layer involves top-level managers and big data governance teams, to deliberate on the deployment of tools, techniques, and technologies for data quality management.
3. *Big Data Quality Management Layer.* The activities in this layer are to ensure that big data quality management requirement is put in place, (i.e., tools and technologies). The layer analyses the big data quality (BDQ) requirements and design of data lifecycle, pre-big data quality evaluation, pre-processing of quality evaluation, processing, and evaluation.
4. *Big Data Quality Presentation Layer.* The activities in this layer are to monitor the data quality management lifecycle using the data visualization tool (e.g., business intelligence dashboards). This will help in identifying the root cause of any unintended data and enable the data governance team to interpret and review the issues, there by validating and updating data quality report which will help in organization decision-making.

However, activity one (1) and six (6) are considered as strategic managers with strategic organizational objectives and make decisions from knowledge or insights obtained from the data governance team. Moreover, the proposed framework is designed with iterative activities, with the arrows indicating the flow and connection of each activity in the big data quality management lifecycle.

Conclusion. As discussed earlier, the study aimed at identifying the current challenges of big data quality management in big data governance from the recent scholarly articles published between (2020-2023). However, the paper is divided into virtually two domain areas (i.e., big data quality and big data governance). Though, this study considered both as one entity. Big data governance is the core component in data management as depicted in figure 2, while data quality management is a subset or one of the activities in data management. Therefore, ignoring the aspect of data governance when identifying the challenges of data quality management is just like providing a solution to unidentified problems. This study has identified the challenges that contributed to the current issues of data quality management from different contexts. Similarly, the commonly used data quality dimensions were identified from the previous study, which

has enabled this study to propose three (3) additional big data quality dimensions that we considered much significant in addressing the existing challenges of big data quality management. Furthermore, a framework for big data quality management is proposed by this study, based on the previous authors recommendations. The proposed framework can be used in mitigating the identified challenges from the previous studies. Moreover, the proposed framework was developed based on ISO 8000-61 for data quality process reference model and ISO 8000-62 for organizational process maturity assessment, and "Plan-Do-Check-Act" approach for big data quality assessment. In order to encourage the research in this area, we recommend the future research to focus on the implementation of our proposed framework in organizations dealing with big data, or data-driven organizations, and utilize the additional big data quality dimensions proposed by this study.

References

1. F. P. S. Surbakti, W. Wang, M. Indulska, and S. Sadiq, "Factors influencing effective use of big data: A research framework," *Inf. Manag.*, vol. 57, no. 1, p. 103146, 2020, doi: 10.1016/j.im.2019.02.001.

2. J. Byabazaire, G. M. P. O. Hare, and D. T. Delaney, "End-to-End Data Quality Assessment Using Trust for Data Shared IoT Deployments," *IEEE Sens. J.*, vol. 22, no. 20, pp. 19995–20009, 2022, doi: 10.1109/JSEN.2022.3203853.

3. R. Chirkova, J. Doyle, and J. Reutter, "Ensuring data readiness for quality requirements with help from procedure reuse," *J. Data Inf. Qual.*, vol. 13, no. 3, 2021, doi: 10.1145/3428154.

4. B. Mohammad, W. Alzyadat, and M. Al-fayoumi, "An Approach to Improve Data Quality from Big Data Aspect by Sensitive Cost and Time," pp. 22–26, 2020, doi: 10.1109/ICICS49469.2020.239526.

5. R. Gu *et al.*, "SparkDQ: Efficient generic big data quality management on distributed data-parallel computation," *J. Parallel Distrib. Comput.*, vol. 156, pp. 132–147, 2021, doi: 10.1016/j.jpdc.2021.05.012.

6. C. A. Escobar, M. E. McGovern, and R. Morales-Menendez, "Quality 4.0: a review of big data challenges in manufacturing," *J. Intell. Manuf.*, vol. 32, no. 8, pp. 2319–2334, 2021, doi: 10.1007/s10845-021-01765-4.

7. H. Fadlallah *et al.*, "Context-aware big data quality assessment: a scoping review," *J. Data Inf. Qual.*, Jun. 2023, doi: 10.1145/3603707.

8. A. Ramasamy and S. Chowdhury, "Big Data Quality Dimensions: A Systematic Literature Review," *J. Inf. Syst. Technol. Manag.*, vol. 17, 2020, doi: 10.4301/s1807-1775202017003.

9. Juran, "Quality Management and Juran's Legacy," no. June, pp. 665–677, 2007, doi: 10.1002/qre.

10. S. Kim *et al.*, "Organizational process maturity model for IoT data quality management," *J. Ind. Inf. Integr.*, vol. 26, 2022, doi: 10.1016/j.jii.2021.100256.

11. Hossen et.al, "A S tudy on the A spects of Q uality of B ig D ata on O nline B usiness and R ecent T ools and T rends T owards C leaning D irty D ata," no. August, pp. 209–213, 2020, doi: 10.1109/ICSGRC49013.2020.9232648.

12. A. Molinari, "The quality concerns in health care Big Data," pp. 302–305, 2020.

13. O. Reda, I. Sassi, A. Zellou, and S. Anter, "Towards a data quality assessment in big data," *ACM Int. Conf. Proceeding Ser.*, pp. 91–96, Sep. 2020, doi: 10.1145/3419604.3419803.

14. W. Elouataoui, I. El Alaoui, and S. El Mendili, "An Advanced Big Data Quality Framework Based on Weighted Metrics," 2022.

15. A. Zibak, C. Sauerwein, and A. C. Simpson, "Threat Intelligence Quality Dimensions for Research and Practice," *Digit. Threat. Res. Pract.*, vol. 3, no. 4, 2022, doi: 10.1145/3484202.

16. H. Y. Teh, A. W. Kempa-Liehr, and K. I. K. Wang, "Sensor data quality: a systematic review," *J. Big Data*, vol. 7, no. 1, pp. 1–49, 2020, doi: 10.1186/s40537-020-0285-1.

17. A. Goknil *et al.*, "A Systematic Review of Data Quality in CPS and IoT for Industry 4.0," *ACM Comput. Surv.*, vol. 55, no. 14 S, 2023, doi: 10.1145/3593043.

18. C. Liu, P. Nitschke, S. P. Williams, and D. Zowghi, "Data quality and the Internet of Things," *Computing*, vol. 102, no. 2, pp. 573–599, 2020, doi: 10.1007/s00607-019-00746-z.

19. M. Wook, N. A. Hasbullah, N. M. Zainudin, Z. Zarina, and A. Jabar, "Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling," *J. Big Data*, 2021, doi: 10.1186/s40537-021-00439-5.

20. Y. Xiao, J. Huang, and M. Zhu, "Opportunities and Challenges Brought by Digital Quality Management to Big Data Technology," *2022 Int. Conf. Intell. Manuf. Ind. Big Data*, no. 2019063, pp. 89–94, 2022, doi: 10.1109/ICIMIBD58123.2022.00027.

21. G. Zhang, "A data traceability method to improve data quality in a big data environment," pp. 290–294, 2020, doi: 10.1109/DSC50466.2020.00051.

22. M. G. A. El-wahab, "Dimensionality for Big Data Quality : A Review," vol. 3, no. 2, pp. 1–10, 2021.

23. H. Li, H. Lu, C. S. Jensen, B. Tang, and M. A. Cheema, "Spatial Data Quality in the Internet of Things: Management, Exploitation, and Prospects," *ACM Comput. Surv.*, vol. 55, no. 3, 2022, doi: 10.1145/3498338.

24. A. Z. Faroukhi, "Big Data Value Chain : A Unified Approach for Integrated Data Quality and Security," 2020, doi: 10.1109/ICECOCS50124.2020.9314391.

25. W. Hattawi, S. Shaban, and A. Al Shawabkah, "Recent Quality Models in BigData Applications," *2021 Int. Conf. Inf. Technol.*, pp. 811–815, 2021, doi: 10.1109/ICIT52682.2021.9491629.

26. Y. Alkatheeri, A. Ameen, O. Isaac, A. Al-Shibami, and M. Nusari, "The Mediation Effect of Management Information Systems on the Relationship between Big Data Quality and Decision making Quality," *TEST Eng. Manag.*, vol. 82, no. 2, pp. 12065–12074, 2020.

27. Y. Wang, C. Yu, J. Hou, Y. Zhang, X. Fang, and S. Wu, "Research on the Key Issues of Big Data Quality Management, Evaluation, and Testing for Automotive Application Scenarios," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/9996011.

28. V. Simard, M. Rönnqvist, L. Lebel, and N. Lehoux, "A Method to Classify Data Quality for Decision Making Under Uncertainty," *J. Data Inf. Qual.*, vol. 15, no. 2, 2023, doi: 10.1145/3592534.

29. M. Talha, A. Abou, and E. Kalam, "Big Data between Quality and Security : Dynamic Access Control for Collaborative Platforms," vol. 27, no. 12, pp. 1300–1324, 2021, doi: 10.3897/jucs.77046.

30. H. Fadlallah *et al.*, "BIGQA: Declarative Big Data Quality Assessment," *J. Data Inf. Qual.*, Jun. 2023, doi: 10.1145/3603706.

31. H. Zou *et al.*, "Big Data Quality Assessment Model for Unstructured Data," *2022 IEEE Int. Conf. Big Data (Big Data)*, vol. 3, no. Iciba, pp. 559–563, 2020, doi: 10.1109/BigData55660.2022.10020564.

32. A. A. Alwan, M. A. Ciupala, A. J. Brimicombe, S. A. Ghorashi, A. Baravalle, and P. Falcarin, "Data quality challenges in large-scale cyber-physical systems: A systematic review," *Inf. Syst.*, vol. 105, p. 101951, 2022, doi: 10.1016/j.is.2021.101951.

33. L. Zhang, D. Jeong, and S. Lee, "Data quality management in the internet of things," *Sensors*, vol. 21, no. 17, pp. 1–21, 2021, doi: 10.3390/s21175834.

34. K. Yee, "Big Data Quality Prediction on Banking Applications : Extended Abstract," pp. 791–792, 2020, doi: 10.1109/DSAA49011.2020.00119.

35. S. Cisneros-Cabrera, A. V. Michailidou, S. Sampaio, P. Sampaio, and A. Gounaris, "Experimenting with big data computing for scaling data quality-aware query processing," *Expert Syst. Appl.*, vol. 178, no. July 2020, p. 114858, 2021, doi: 10.1016/j.eswa.2021.114858.

36. D. F. Ayala, *The DAMA Guide to the Data Management Body of Knowledge First Edition The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide) First Edition Susan Earley , Assistant Editor Deborah Henderson , Project Sponsor*.

37. S. Varshney and K. Allen, "Implementing Data Governance," pp. 1–32, 2020.

38. M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy Artificial Intelligence," *Gov. Inf. Q.*, vol. 37, no. 3, p. 101493, 2020, doi: 10.1016/j.giq.2020.101493.

39. I. A. Shea Sharon, "What is Data Security? The Ultimate Guide," *Tech Targets*, pp. 1–24, 2022, [Online]. Available:

https://www.techtarget.com/searchsecurity/Data-security-guide-Everything-you-need-to-know

40. V. Wenz, A. Kesper, and G. Taentzer, "Clustering Heterogeneous Data Values for Data Quality Analysis," *J. Data Inf. Qual.*, vol. 15, no. 3, pp. 1–33, 2023, doi: 10.1145/3603710.

41. F. Gualo, M. Rodriguez, J. Verdugo, and I. Caballero, "The Journal of Systems & Software Data quality certification using ISO / IEC 25012 : Industrial experiences ☆," *J. Syst. Softw.*, vol. 176, p. 110938, 2021, doi: 10.1016/j.jss.2021.110938.