ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse53922020.pdf
https://doi.org/10.30534/ijatcse/2020/53922020

Big Data Analytics Application Model Based on Data Quality Dimensions and Big Data Traits in Public Sector

Muslihah Wook[1], Zam Zarina Abdul Jabar[1], Muhammad Hakiem Halim[1], Noor Afiza Mat Razali[1], Suzaimah Ramli[1], Nor Asiakin Hasbullah[1], Norulzahrah Mohd Zainuddin[1]
[1]Department of Computer Science, Universiti Pertahanan Nasional Malaysia, Malaysia
muslihah@upnm.edu.my
zam.zarina.aj@gmail.com
hakiem2397@gmail.com
noorafiza@upnm.edu.my
suzaimah@upnm.edu.my
asiakin@upnm.edu.my
norulzahrah@upnm.edu.my

ABSTRACT

Big data analytics (BDA) represents a new technological paradigm with its ability to extract valuable knowledge from high amounts of data. Exploring the effect of data quality dimensions (DQD) and big data traits (BDT) on BDA application is a relatively new research trend that has not been featured in the existing literature. Thus, this study was conducted to build a new model by integrating the DQD and BDT to examine the BDA application in the context of the Malaysian public sector. This study proposes that the DQD should include intrinsic, contextual, representational, and accessibility dimensions, while variety, validity, and veracity should be considered as the main characteristics of BDT. For this purpose, this study employed theory analysis on on-going research related to DQD and BDT, along with the development of the BDA application model. The proposed model would create new research fields related to DQD and BDT in the BDA domain. Finally, in line with the Public Sector Big Data (DRSA) platform initiated by the Malaysian Administrative Modernisation and Management Planning Unit (MAMPU), the new model is expected to provide a benchmark for the development and application of BDA for the Malaysian public sector.

Key words: Big data, Big data analytics, Big data traits, Data quality dimensions, Public sector.

1. INTRODUCTION

In recent years, many academic researchers, industry practitioners, and government organisations have attempted to effectively utilise big data analytics (BDA) [1]. BDA can be conceptualised as a suite of data management and analytical techniques for handling large and complex data sets [2]. The volume of data related to the public sector has grown dramatically in the past years and it is expected to increase in the coming years due to the increased use of innovative technologies [3]. Several scholars opined that the application of BDA in the public sector could promote government transparency, accountability, and responsiveness to citizens' demands [4], [5]. The number of studies related to the application of BDA in public sector is growing (e.g., [4]–[10]). However, a more in-depth view of such application in the public sector is still required [11].

Previous studies have reported the quality of data derived from BDA as being unsatisfactory, as most people would argue about the accuracy, completeness, and consistency (to name a few) of the data [12]–[15]. Inaccurate, incomplete, and inconsistent data can create serious problems, which could lead to incorrect decisions made by organisations, thus causing them to lose money [14]. Reference [16] pointed out that there is an urgent need to conduct in-depth research into data quality to determine which elements are more important in the context of BDA application. Although several theories or models have been proposed for understanding the problem of data quality, such as the resource based theory (RBT), the organisational learning theory (OLT), the firm performance (FPER), and the data quality framework (DQF), these theories or models were not addressed in previous research studies for BDA application, particularly in the context of the Malaysian public sector. Furthermore, most of these theories and models do not fit into the application of information system, such as BDA, since they are focusing more on service quality compared to data quality [17].

Meanwhile, numerous characteristics or traits have been proposed for big data, such as 3Vs [18], 4Vs [19], 5Vs [20], [21], 6Vs [22], 7Vs [23], 9Vs [24], 10Vs [25], 11Vs [26], and 17Vs [27]. However, there is a lack of uniform consensus regarding the core of big data traits (BDT). This is because comprehensive analysis and research of the integration of quality and BDT in BDA application are still lacking [28], [29]. Although the dimensions of data quality and BDT have been viewed as separated domains, the authors in [26] found that these two domains are intertwined and closely related. To address the problems with BDA application in the public sector and to address the knowledge gaps in the existing literature, this study examined the determinants that could influence this technology. To achieve this objective, two main research questions were asked: (1) What are the determinants of data quality that could potentially influence BDA application by the public sector?; and (2) Which of the BDT could potentially influence the quality of data in the public sector? The conceptual model was based on the data quality dimensions (DQD) framework by [26] and [30]. This framework is a good starting point for exploring determinants that could influence BDA application in the public sector. The DQD framework suggests four main dimensions of IT application, namely, intrinsic, contextual, representational, and accessibility. This study then extended the model by examining critical BDT that could influence the DQD. Although there are diverse Vs of BDT, this study had only focused on the most significant ones.

2. LITERATURE REVIEW

2.1 Big Data Analytics

There is no global consensus on the definition of big data. Various stakeholders (e.g., academia, industry, media, and public interest) provide diverse and contradictory definitions [31]. According to [32], the term ‘big data’, minus the term ‘analytics’, represents an enormous amount of data, whilst the term ‘analytics’, minus the term ‘big data’, indicates simple applications of mathematical and statistical analysis tools. Nowadays, ‘big data’ and ‘analytics’ are merged together to represent the advanced analytical techniques used to make better decisions to benefit organisations. According to [33], ‘big data’ is often associated with the term ‘analytics’ and commonly known as ‘big data analytics’, which refers to the ability to extract information from data using various techniques, such as artificial intelligence, mathematics, statistics, and other techniques to support the decision making process. Although most studies use ‘big data’ and ‘big data analytics’ (BDA) interchangeably, this paper will use BDA as this term clearly reflects the specific concept of big data. Reference [34] defined BDA as the plethora of advanced digital techniques designed to identify trends, detect patterns, and unveil unforeseen knowledge related to human behaviour from high amount of data.

Recognizing BDA’s ability for analysing high amount of data, actionable ideas can be delivered for measuring performance, thus leading to competitive advantages in any organisations [35], [36]. However, a recent report showed that only 25% of organisations that use BDA had significantly improve their performance [37]. This phenomenon is mainly caused by the growth of unstructured data in the digital world, which constitute 90% of big data [38]. Similarly, authors in [34] argued that there is an increasing amount of various data, ranging from structured (numeric) to unstructured data (numbers, text, audio, images, video, etc.) that are collected from a multiplicity of platforms (e.g., social media, network sensors, and Internet of Things [IoT]).

Unlike traditional data analysis in which data types are typically structured, BDA is complex since it has to deal with various types of unstructured data [39]. However, BDA is expected to execute on the combination of both unstructured and structured data. Nowadays, various analytical techniques are available, including data mining, visualization, statistical analysis, and machine learning algorithms to reveal solutions for both structured and unstructured data. Numerous studies have tackled this area by enhancing pre-existing techniques, proposing new ones, and testing different combinations of various algorithms and technologies [40]–[42].

2.2 Application of Big Data Analytics in Public Sector

The public sector collects and generates massive amounts of data through their daily activities. These data sets cover a range of agencies including education, health, transportation, environment, and agriculture. The effectiveness of the services offered by the public sector can be seen through improved services to the citizens, which would require less resources to provide quality services [8]. According to [43], these improvements can be achieved by applying BDA, as it enables the public sector to make information-based or fact-based decisions, create new solutions, reduce waste, and plan intelligently for the future [5]. The application of BDA in the public sector could even enhance their ability to serve their citizens in addressing national issues related to basic needs (food, clothing, and shelter), quality education, and increased employment rate [44].

Despite the potential value of BDA, the public sector seems to be falling behind the private sector [7]. This is due to the challenges between the promise of BDA and the real-world BDA practice in the public sector [6]. Additionally, the public sector uses BDA to promote public good, whereas the private sector uses BDA to gain profits [44]. The private sector has been using BDA in their daily activities. For example, Walmart, Sears, and Amazon use BDA to better understand customers’ behaviour through their buying decisions. Another example can be seen among prominent social media companies, such as Google, Twitter, Facebook, and eBay that have created intelligent business models by assessing users’ behaviour, preferences, and information or product request. Following the private sector's move to embrace BDA, the application of this technology in the public sector is mainly to increase their efficiency, effectiveness, transparency, and accountability. Although several scholars have discussed the obstacles and threats of using BDA in public sectors, some applications have been successfully implemented, such as smart governance [10], smart cities [45], public transportation [46], and open data [47].

The promise of BDA application has led the Malaysian government to acknowledge the potential use of this technology as one of the national agenda [48]. The government has appointed the Malaysia Administrative Modernization and Management Planning Unit (MAMPU) to implement the BDA project for the public sector in 2015 [49]. In 2013, Multimedia Super Corridor (MSC) Malaysia has approved RM24 million as allocation for BDA pilot projects involving four public sector agencies [48]. Nowadays, the application of BDA in the Malaysian public sector is still not fully established and many challenges still need to be overcome [50]. Various BDA challenges have been highlighted, such as data acquisition and metadata, data quality, data storage, data sharing and transfer, scalability, data analysis, data querying and indexing, data uncertainty, data privacy, security and ethics, and data visualisation [51]. However, this study has only focused on data quality because before embarking on BDA application, it is necessary to obtain sufficient assurance about the quality of the large data that will be used [16], particularly in the Malaysian public sector.

2.3 Data Quality Dimensions

Data are critical resources in major applications within organisations, thus the quality of data is essential for managers or practitioners to solve related problems. Data quality is defined as the wellness and appropriateness of data to meet the requirements of organisations [16]. It is also defined as the degree to which the data fits its use [14], [30]. These definitions underline the view that data quality is not only related to the data collection itself, but also to the use of data [52]. Reference [30] divided data quality into four dimensions to better comprehend the perspective of a data user. These dimensions are intrinsic (i.e., the correctness of data), accessibility (i.e., the ease of obtaining the data), contextual (i.e., the quality context of the data), and representational (i.e., the showing of data in a clear manner). Each dimension has several elements used as unique measurements of data quality. For example, accuracy and believability are considered as elements of the intrinsic dimension, relevance and completeness are considered as elements of the contextual dimension, understandability and interpretability are considered as elements of the representational dimension, and accessibility, access security, and ease of operations are considered as elements of the accessibility dimension. Table 1 provides the data quality elements according to their dimensions.

Table 1: Data Quality Elements and Dimensions [52]

Data Quality Dimensions	Data Quality Elements
Intrinsic	Accuracy
	Objectivity
	Believability
	Reputation
Contextual	Value-added
	Relevancy
	Timeliness
	Completeness
	Appropriate amount of data
Representational	Interpretability
	Understandability
	Concise representation
	Consistent representation
Accessibility	Accessibility
	Access security
	Ease of operations

Research on data quality for BDA application in fields related to the public sector are limited. Therefore, this study has reviewed available literature on data quality for BDA application in other fields as well. As shown in Table 2, 22 studies had focused on data quality elements for BDA application from 2013 to 2019. These studies were in various fields, namely, industrial firms [13], [14], [53], market research [12], [15], [54], transportations [55]–[57], business [58], [59], supply chain [60], health [61], [62], and public sector [63]. Table 2 also lists 12 data quality elements that have been extracted from the available literature, namely, completeness, accuracy, consistency, timeliness, accessibility, understandability, believability, format, relevancy, quantity, correctness and security which reveal high frequencies of reference at 21, 19, 16, 14, 8, 8, 6, 6, 5, 4, 3 and 3 respectively.

Table 2: Data Quality Elements for BDA Application Extracted through Literature Analysis

No.	Data Quality Elements	Reference
1.	Completeness	[12]–[15], [53]–[70]
2.	Accuracy	[12]–[15], [54]–[56], [58]–[63], [65]–[70]
3.	Consistency	[14], [15],
4.	Timeliness	[12], [13], [15], [54]–[56], [58]–[61],[63], [67]–[69]
5.	Accessibility	[14], [61], [63], [65], [67]–[70]
6.	Understandabilit y	[14], [59], [63], [65], [67]–[70]
7.	Believability	[14], [58], [59], [61], [63], [68]
8.	Format	[12], [13], [54], [57], [63], [68]

9.	Relevancy	[58], [61], [63], [66], [68]
10.	Quantity	[55], [56], [58], [61]
11.	Correctness	[58], [64], [67]
12.	Security	[14], [59], [67]

2.4 Big Data Traits

Big data has many characteristics or traits that were introduced by various scholars. Initially, author in [18] introduced three main traits or 3Vs, known as 'volume, velocity, and variety' of big data. Over the years, more Vs were added to big data, such as 4Vs [19], [71], 5Vs [20], [21], 6Vs [22], 7Vs [23], 9Vs [24], 10Vs [25], 11Vs [26], and 17 Vs [27]. Table 3 lists several studies on big data traits (BDT). Some of them attempted to provide the maximum number of traits for big data, such as in [27], while others classified the traits in different categories. For instance, the work of [24] grouped 9Vs into five categories, namely, collecting data (veracity, variety), processing data (velocity, volume), integrity of data (validity, variability, volatility), visualisation of data (visualisation), and worth of data (value). Meanwhile, [26] classified 11Vs into the aforementioned data quality categories (intrinsic, accessibility, representational, and contextual).

Table 3 also shows that most of these studies agree that volume, velocity, and variety are the main traits of big data. Volume refers to the size of the data set that organisations are trying to harness to improve decision making. Velocity is defined as the rate at which data are generated, processed, analysed, and visualised, while variety refers to the diversity of data types [51]. Apart from these 3Vs, several scholars proposed veracity, value, validity, variability, volatility, and visualisation as the new traits of big data [22]–[27]. Veracity means the accuracy, completeness, and quality of data, which relates to the reliability and consistency of data types. Value describes the potential benefits that organisations would receive, while validity is defined as the correctness of data for its intended use and variability means the change of data flow or format in some duration of time [51]. Volatility refers to the life duration of a data set that is valid to be used and stored [25], and visualisation is a way to make all the data comprehensible in a manner that is easy to understand and read [24]. Meanwhile, the other Vs, namely, viscosity, venue, vocabulary, vagueness, verbosity, voluntariness, versatility, viability, visibility, vast resources, and virality showed considerably lesser meaning to what BDT is essentially about.

3. METHODOLOGY

The methodology of this study consisted of two phases. The first phase was focused on identifying related determinants in data quality from theory analysis. The second phase involved the development of the proposed BDA application model. Figure 1 illustrates a summary of the research activities involved in this study.

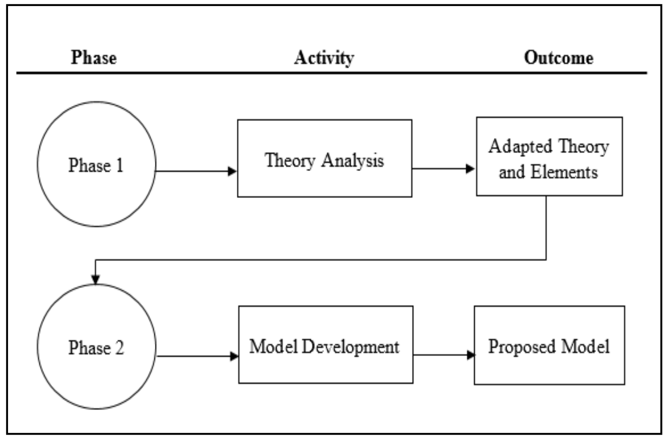

Figure 1: Research Activities

The following subsections discuss the details of the two phases of the research methodology:

3.1 Theory Analysis

This study has initially selected theories or models based on their implementation in the context of data quality assessment for BDA application. In this phase, four well-known theories were identified and analysed, namely, the resource-based theory (RBT), organisational learning theory (OLT), firm performance (FPER), and data quality framework (DQF). Theories that have similar names were refined, namely, resource-based view and knowledge-based view were refined to RBT. Table 4 lists the elements that have been extracted from these theories, with their respective references.

The table shows that a total of 22 elements have been extracted from these theories. However, the names of some elements have been refined due to the redundancy of their meaning and description. These elements are format and presentation, currentness and timeliness, relevancy and fitness, quantity and volume, believability and credibility, and understandability and readability. Subsequently, 18 distinct elements were identified from these theories, namely, accuracy, completeness, consistency, currentness, compliance, format, security, privacy, use behaviour, believability, accessibility, relevancy, quantity, value-added, reputation, trustworthiness, understandability, and redundancy.

3.2 Model Development

Based on the literature review and theory analysis in the previous sections, this study proposes a BDA application model based on the integration of DQD and BDT, as depicted in Figure 2. This model was developed to incorporate critical DQD, which have an impact on BDA application. It is worth mentioning that different applications can have different requirements since not all DQD are always relevant [55]. Therefore, this study has chosen the dimensions that are commonly accepted and widely used in BDA application. Each dimension has its own corresponding quality elements. In this way, the intrinsic, contextual, representational, and accessibility dimensions can be used to evaluate BDA application. The intrinsic dimension is objective to the correctness of the data, which is composed of two elements, namely, accuracy and believability. Contextual dimension relies on the context within which the data is used and this study considered two elements, namely, completeness and timeliness. The representational dimension refers to the representation of data in a clear manner and this study considered consistency and understandability as its elements. Lastly, the accessibility dimension means the ease of obtaining data and concerns over the importance of computer system to facilitate accessing and storing of data. Access security and ease of operations are considered as the elements of the accessibility dimension for this study.

Table 3: Big Data Traits

Big Data Traits	**3Vs** [1]	**4Vs 1C** [2]	**4Vs** [3]	**5Vs** [4]	**5Vs** [5]	**6Vs** [6]	**7Vs** [7]	**9Vs** [8]	**10Vs** [9]	**11Vs** [10]	**17Vs 1C** [11]
Volume	/	/	/	/	/	/	/	/	/	/	/
Velocity	/	/	/	/	/	/	/	/	/	/	/
Variety	/	/	/	/	/	/	/	/	/	/	/
Veracity		/		/	/	/	/	/	/	/	/
Value			/	/	/		/	/	/	/	/
Validity						/	/	/	/	/	/
Volatility						/	/	/	/	/	/
Variability								/	/	/	/
Visualisation								/	/		/
Complexity		/									/
Viscosity									/		/
Venue											/
Vocabulary											/
Vagueness											/
Verbosity											/
Voluntariness											/
Versatility											/
Viability										/	
Visibility										/	
Vast resources										/	
Virality											/

Table 4: Theory or Model

Elements	**Theory/Model**	**References**
Accuracy Completeness Currentness Format	Resource Based Theory (RBT)	[12], [13], [15], [53], [72]
Compliance Security Privacy	Organisational Learning Theory (OLT)	[14]
Consistency Use behaviour	Firm Performance (FPER)	[72]
Accuracy Completeness Consistency Believability Currentness Accessibility Relevancy Quantity Value-added Reputation Trustworthiness Understandability Redundancy	Data Quality Framework (DQF)	[14], [59], [60], [70]

The developed BDA application model has also incorporated the critical BDT, which could impact DQD. Based on Table 3, there is no single universal consensus on what constitutes as the core BDT. Different scholars have proposed new traits, which may interfere with the main issues related to big data [73]. The lack of a consensual BDT core means that many scholars have challenged and ignored previous traits and proposed new ones [74], [75]. Although most of the previous studies have agreed that volume, velocity, and variety are related to big data, the existence of other traits has hampered the understanding of how big data is changing the quality of data. Reference [26] investigated 11 traits, namely, volume, velocity, variety, variability, veracity, validity, visibility, volatility, viability, value, and vast resources within four dimensions of data quality that included 16 elements. However, the vast amount of BDT may not be comprehensive enough to capture the DQD of BDA application [15]. These traits should be able to rigorously measure the exact dimensions of data quality. Therefore, this study has focused on variety, validity, and veracity since these traits were found to be more important for most DQD [26].

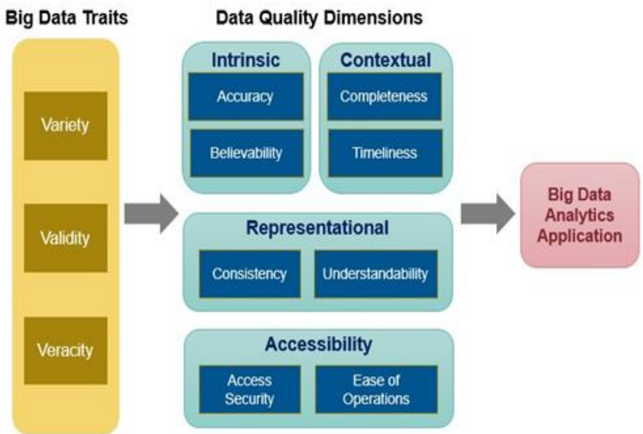

Figure 2: Proposed Conceptual Model for BDA Application

5. CONCLUSION

This present study has introduced a conceptual model that integrates DQD and BDT for examining BDA application in the context of the Malaysian public sector. DQD comprises of intrinsic (accuracy and believability), contextual (completeness and timeliness), representational (consistency and understandability), and accessibility (access security and ease of operations) dimensions. Meanwhile, BDT consists of the most significant characteristics of big data, namely variety, validity, and veracity. In the near future, this study will conduct a data collection process from several agencies in the Malaysian public sector. Structural equation modelling will be applied to analyse the collected data in order to explore the significant determinants of the proposed conceptual model. Overall, it is hoped that this model may offer a comprehensive explanation of BDA application, particularly in the Malaysian public sector.

ACKNOWLEDGEMENT

The authors would like to thank the Centre for Research and Innovation Management, Universiti Pertahanan Nasional Malaysia for supporting this research under the Short-Term Research Grant, code UPNM/2018/GPJP/2/TK/5.

REFERENCES

1. M. K. Saggi and S. Jain. **A survey towards an integration of big data analytics to big insights for value-creation**, Inf. Process. Manag., vol. 54, no. 5, pp. 758–790, Sep. 2018. https://doi.org/10.1016/j.ipm.2018.01.010
2. S. Verma, S. S. Bhattacharyya, and S. Kumar. **An extension of the technology acceptance model in the big data analytics system implementation environment**, Inf. Process. Manag., vol. 54, no. 5, pp. 791–806, Sep. 2018.
3. F. Haneem, N. Kama, N. Taskin, D. Pauleen, and N. A. Abu Bakar. **Determinants of master data management adoption by local government organizations: An empirical study**, Int. J. Inf. Manag., vol. 45, pp. 25–43, Apr. 2019.
4. C. Fredriksson, F. Mubarak, M. Tuohimaa, and M. Zhan. **Big data in the public sector: A systematic literature review**, *Scand. J. Public Adm.*, vol. 21, no. 3, pp. 39–61, 2017.
5. K. Hardy and A. Maurushat. **Opening up government data for big data analysis and public benefit**, Comput. Law Secur. Rev., vol. 33, no. 1, pp. 30–37, Feb. 2017. https://doi.org/10.1016/j.clsr.2016.11.003
6. K. C. Desouza and B. Jacob. **Big Data in the public sector: Lessons for practitioners and scholars**, Adm. Soc., vol. 49, no. 7, pp. 1043–1064, Aug. 2017.
7. B. Klievink, B.-J. Romijn, S. Cunningham, and H. de Bruijn. **Big data in the public sector: Uncertainties and readiness**, Inf. Syst. Front., vol. 19, no. 2, pp. 267–283, Apr. 2017.
8. R. Munné. **Big Data in the Public Sector**, in *New Horizons for a Data-Driven Economy*, J. M. Cavanillas, E. Curry, and W. Wahlster, Eds. Cham: Springer International Publishing, 2016, pp. 195–208.
9. N. Rogge, T. Agasisti, and K. De Witte. **Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art**, Public Policy Adm., vol. 32, no. 4, pp. 263–281, Oct. 2017.
10. M. N. I. Sarker, M. Wu, and M. A. Hossin. **Smart governance through big data: Digital transformation of public agencies**, in 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, 2018, pp. 62–70. https://doi.org/10.1109/ICAIBD.2018.8396168
11. V. Weerakkody, K. Kapoor, M. E. Balta, Z. Irani, and Y. K. Dwivedi. **Factors influencing user acceptance of public sector big open data**, Prod. Plan. Control, vol. 8, no. 11–12, pp. 891–905, 2017.
12. S. Akter, S. F. Wamba, and S. Dewan. **Why PLS-SEM is suitable for complex modelling? An empirical illustration in big data analytics quality**, Prod. Plan. Control, vol. 28, no. 11–12, pp. 1011–1021, Sep. 2017.
13. N. Côrte-Real, P. Ruivo, and T. Oliveira. **Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?**, Inf. Manage., vol. 57, no. 1, p. 103141, Jan. 2020.
14. M. Ghasemaghaei and G. Calic. **Can big data improve firm decision quality? The role of data quality and data diagnosticity**, Decis. Support Syst., vol. 120, pp. 38–49, May 2019. https://doi.org/10.1016/j.dss.2019.03.008
15. S. F. Wamba, S. Akter, and M. de Bourmont. **Quality dominant logic in big data analytics and firm performance**, Bus. Process Manag. J., vol. 25, no. 3, pp. 512–532, Jun. 2019.
16. F. I. Salih, S. A. Ismail, M. M. Hamed, O. Mohd Yusop, A. Azmi, and N. F. Mohd Azmi. **Data Quality Issues in Big Data: A Review**, in *Recent Trends in Data Science*

and Soft Computing, vol. 843, F. Saeed, N. Gazem, F. Mohammed, and A. Busalim, Eds. Cham: Springer International Publishing, 2019, pp. 105–116.

17. P. H. S. Panahy, F. Sidi, M. A. Jabar, H. Ibrahim, and A. Mustapha. **A framework to construct data quality dimensions relationships**, Indian J. Sci. Technol., vol. 6, pp. 4422–4431, 2013.
18. D. Laney. **3D Data Management: Controlling Data Volume, Velocity and Variety**, Gartner Blog Network, 06-Feb-2001. [Online]. Available: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. [Accessed: 20-Aug-2019].
19. V. S. Thiyagarajan and K. Venkatachalapathy. **Isolating values from big data with the help of four V's**, Int. J. Res. Eng. Technol., vol. 4, no. 1, pp. 132–135, 2014.
20. H. J. Hadi, A. H. Shnain, S. Hadishaheed, and A. H. Ahmad. **Big data and five V's characteristics**, Int. J. Adv. Electron. Comput. Sci., vol. 2, no. 1, p. 8, 2015.
21. Ishwarappa and J. Anuradha. **A brief introduction on big data 5Vs characteristics and hadoop technology**, Procedia Comput. Sci., vol. 48, pp. 319–324, 2015. https://doi.org/10.1016/j.procs.2015.04.188
22. A. T. Yassin. **Analyzing 6Vs of big data using system dynamics**, in The 2nd Scientific Conference of the College of Science, 2014, p. 9.
23. M. A. Khan, M. F. Uddin, and N. Gupta. **Seven V's of big data understanding big data to extract value**, in Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education, Bridgeport, CT, USA, 2014, pp. 1–5.
24. S. S. Owais and N. S. Hussein. **Extract five categories CPIVW from the 9V's characteristics of the big data**, Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 3, pp. 254–258, 2016.
25. N. Khan, M. Alsaqer, H. Shah, G. Badsha, A. A. Abbasi, and S. Salehian. **The 10 Vs, issues and challenges of big data**, in Proceedings of the 2018 International Conference on Big Data and Education - ICBDE '18, Honolulu, HI, USA, 2018, pp. 52–56.
26. A. Wahyudi, A. Farhani, and M. Janssen. **Relating big data and data quality in financial service organizations**, in *Challenges and Opportunities in the Digital Era*, vol. 11195, S. A. Al-Sharhan, A. C. Simintiras, Y. K. Dwivedi, M. Janssen, M. Mäntymäki, L. Tahat, I. Moughrabi, T. M. Ali, and N. P. Rana, Eds. Cham: Springer International Publishing, 2018, pp. 504–519.
27. P. S. Arockia, S. S. Varnekha, and K. A. Veneshia. **The 17 V's of big data**, Int. Res. J. Eng. Technol., vol. 4, no. 9, pp. 328–333, 2017.
28. O. Müller, I. Junglas, J. vom Brocke, and S. Debortoli, **Utilizing big data analytics for information systems research: challenges, promises and guidelines**, Eur. J. Inf. Syst., vol. 25, no. 4, pp. 289–302, Jul. 2016. https://doi.org/10.1057/ejis.2016.2
29. V. C. Storey and I.-Y. Song. **Big data technologies and management: What conceptual modeling can do**, Data Knowl. Eng., vol. 108, pp. 50–67, Mar. 2017.
30. R. Y. Wang and D. M. Strong. **Beyond accuracy: What data quality means to data consumers**, J. Manag. Inf. Syst., vol. 12, no. 4, pp. 5–33, 1996.
31. I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, J. Krogstie, and M. Mäntymäki, Eds., **Digital Transformation for a Sustainable Society in the 21st Century**, vol. 11701. Cham: Springer International Publishing, 2019.
32. N. R. Sanders. **How to use big data to drive your supply chain**, Calif. Manage. Rev., vol. 58, no. 3, pp. 26–48, May 2016.
33. A. Belhadi, K. Zkik, A. Cherrafi, S. M. Yusof, and S. El Fezazi. **Understanding big data analytics for manufacturing processes: Insights from literature review and multiple case studies**, Comput. Ind. Eng., vol. 137, p. 106099, Nov. 2019. https://doi.org/10.1016/j.cie.2019.106099
34. M. Favaretto, E. De Clercq, and B. S. Elger. **Big data and discrimination: perils, promises and solutions. A systematic review**, J. Big Data, vol. 6, no. 1, p. 12, Dec. 2019.
35. P. Mikalef, M. Boura, G. Lekakos, and J. Krogstie. **Big data analytics and firm performance: Findings from a mixed-method approach**, J. Bus. Res., vol. 98, pp. 261–276, May 2019.
36. S. F. Wamba, A. Gunasekaran, S. Akter, S. J. Ren, R. Dubey, and S. J. Childe. **Big data analytics and firm performance: Effects of dynamic capabilities**, J. Bus. Res., vol. 70, pp. 356–365, Jan. 2017.
37. Y. Hyun, R. Hosoya, and T. Kamioka. **The implications of big data analytics orientation upon deployment**, in Proceedings of the 6th International Conference on Information Technology: IoT and Smart City, Hong Kong, 2018, pp. 42–48. https://doi.org/10.1145/3301551.3301566
38. U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, **Critical analysis of big data challenges and analytical methods**, J. Bus. Res., vol. 70, pp. 263–286, Jan. 2017.
39. X. Zhang and S. Xiang, **Data Quality, Analytics, and Privacy in Big Data**, in *Big Data in Complex Systems*, vol. 9, A. E. Hassanien, A. T. Azar, V. Snasael, J. Kacprzyk, and J. H. Abawajy, Eds. Cham: Springer International Publishing, 2015, pp. 393–418.
40. L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren. **Information security in big data: privacy and data mining**, IEEE Access, vol. 2, pp. 1149–1176, 2014.
41. C. Liu, C. Yang, X. Zhang, and J. Chen. **External integrity verification for outsourced big data in cloud and IoT: A big picture**, Future Gener. Comput. Syst., vol. 49, pp. 58–67, Aug. 2015.
42. A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih. **Big data technologies: A survey**, J. King Saud

Univ. Comput. Inf. Sci., vol. 30, no. 4, pp. 431–448, Oct. 2018.
43. W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg. **Debating big data: A literature review on realizing value from big data**, J. Strateg. Inf. Syst., vol. 26, no. 3, pp. 191–209, Sep. 2017. https://doi.org/10.1016/j.jsis.2017.07.003
44. N. Z. Zainal, H. Hussin, and M. N. M. Nazri. **Big data initiatives by governments -- issues and challenges: A review**, in 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M), Jakarta, Indonesia, 2016, pp. 304–309.
45. E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi. **Applications of big data to smart cities**, J. Internet Serv. Appl., vol. 6, no. 1, p. 25, Aug. 2015.
46. A. H. Johar and H. Khalid. **Big data analytics adoption and implementation in public transportation: The gap in practise**, Open Int. J. Inform., vol. 7, no. Special Issue, 2019.
47. M. Janssen, D. Konopnicki, J. L. Snowdon, and A. Ojo. **Driving public sector innovation using big and open linked data (BOLD)**, Inf. Syst. Front., vol. 19, no. 2, pp. 189–195, Apr. 2017. https://doi.org/10.1007/s10796-017-9746-2
48. R. H. R. M. Ali, R. Mohamad, and S. Sudin. **A proposed framework of big data readiness in public sectors**, in Proceedings of The International Conference on Applied Science and Technology (ICAST'16), 2016, p. 020089.
49. K. Anna and K. Nikolay. **Survey on big data analytics in public sector of Russian Federation**, in 3rd International Conference on Information Technology and Quantitative Management, 2015, vol. 55, pp. 905–911.
50. E. A. Wahdain, A. S. Baharudin, and M. N. Ahmad. **Big data analytics in the Malaysian public sector: The determinants of value creation**, in *Recent Trends in Data Science and Soft Computing*, vol. 843, F. Saeed, N. Gazem, F. Mohammed, and A. Busalim, Eds. Cham: Springer International Publishing, 2019.
51. D. Gupta and R. Rani. **A study of big data evolution and research challenges**, J. Inf. Sci., vol. 45, no. 3, pp. 322–340, Jun. 2019.
52. A. Wahyudi, G. Kuk, and M. Janssen. **A process pattern model for tackling and improving big data quality**, Inf. Syst. Front., vol. 20, no. 3, pp. 457–469, Jun. 2018.
53. O. Kwon, N. Lee, and B. Shin. **Data quality management, data usage experience and acquisition intention of big data analytics**, Int. J. Inf. Manag., vol. 34, no. 3, pp. 387–394, Jun. 2014.
54. S. Ji-fan Ren, S. F. Wamba, S. Akter, R. Dubey, and S. J. Childe. **Modelling quality dynamics, business value and firm performance in a big data analytics environment**, Int. J. Prod. Res., vol. 55, no. 17, pp. 5011–5026, 2016. https://doi.org/10.1080/00207543.2016.1154209
55. D. Ardagna, C. Cappiello, W. Samá, and M. Vitali. **Context-aware data quality assessment for big data**, Future Gener. Comput. Syst., vol. 89, pp. 548–562, Dec. 2018.
56. C. Cappiello, W. Samá, and M. Vitali. **Quality awareness for a successful big data exploitation**, in Proceedings of the 22nd International Database Engineering & Applications Symposium on - IDEAS 2018, Villa San Giovanni, Italy, 2018, pp. 37–44.
57. P. Woodall, A. Borek, and A. K. Parlikad. **Data quality assessment: The hybrid approach**, Inf. Manage., vol. 50, no. 7, pp. 369–382, Nov. 2013. https://doi.org/10.1016/j.im.2013.05.009
58. A. Immonen, P. Paakkonen, and E. Ovaska. **Evaluating the quality of social media data in big data architecture**, IEEE Access, vol. 3, pp. 2028–2043, 2015.
59. J. Merino, I. Caballero, B. Rivas, M. Serrano, and M. Piattini. **A data quality in use model for big data**, Future Gener. Comput. Syst., vol. 63, pp. 123–130, 2016.
60. B. T. Hazen, C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer. **Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications**, Int. J. Prod. Econ., vol. 154, pp. 72–80, Aug. 2014. https://doi.org/10.1016/j.ijpe.2014.04.018
61. I. Taleb, R. Dssouli, and M. A. Serhani. **Big data pre-processing: A quality framework**, in IEEE International Congress on Big Data, New York City, NY, USA, 2015, pp. 191–198.
62. I. Taleb, H. T. E. Kassabi, M. A. Serhani, R. Dssouli, and C. Bouhaddioui. **Big data quality: A quality dimensions evaluation**, in Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress, Toulouse, 2016, pp. 759–765.
63. M. T. Ijab, A. Ahmad, R. A. Kadir, and S. Hamid. **Towards big data quality framework for Malaysia's public sector open data initiative**, in *Advances in Visual Informatics*, vol. 10645, H. Badioze Zaman, P. Robinson, A. F. Smeaton, T. K. Shih, S. Velastin, T. Terutoshi, A. Jaafar, and N. Mohamad Ali, Eds. Cham: Springer International Publishing, 2017, pp. 79–87.
64. S.-T. Lai. **An iterative and incremental data quality improvement procedure for reducing the risk of big data project**, J. Softw., vol. 12, no. 12, pp. 945–956, 2017. https://doi.org/10.17706/jsw.12.12.945-956
65. D. Firmani, M. Mecella, M. Scannapieco, and C. Batini. **On the meaningfulness of "big data quality",** Data Sci. Eng., vol. 1, no. 1, pp. 6–20, Mar. 2016.
66. C. Kiefer. **Assessing the quality of unstructured data: An initial overview**, LWDA, p. 12, 2016.
67. J. Gao, C. Xie, and C. Tao. Big data validation and quality assurance -- Issuses, challenges, and needs, in 2016 IEEE Symposium on Service-Oriented System

Engineering (SOSE), Oxford, United Kingdom, 2016, pp. 433–441.

68. L. Cai and Y. Zhu. **The challenges of data quality and data quality assessment in the big data era**, Data Sci. J., vol. 14, no. 0, p. 2, May 2015.
69. N. Abdullah, S. A. Ismail, S. Sophiayati, and S. M. Sam, **Data quality in big data: A review**, Int. J. Adv. Soft Comput. Its Appl., vol. 7, no. 3, p. 12, 2015.
70. C. Batini, A. Rula, M. Scannapieco, and G. Viscusi. **From data quality to big data quality**, J. Database Manag., vol. 26, no. 1, pp. 60–82, Jan. 2015. https://doi.org/10.4018/JDM.2015010103
71. J. Singh and A. Rana. **Exploring the Big Data Spectrum**, Int. J. Emerg. Technol. Adv. Eng., vol. 3, no. 4, pp. 73–76, 2013.
72. S. Ji-Fan Ren, S. F. Wamba, S. Akter, R. Dubey, and S. J. Childe. **Modelling quality dynamics, business value and firm performance in a big data analytics environment**, Int. J. Prod. Res., vol. 55, no. 17, pp. 5011–5026, 2016.
73. S. M. Shahar, N. S. Mohd Satar, and K. A. Abu Bakar. **Exploring service quality dimensions of it shared services in the Malaysian public sector**, Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 1.4, pp. 97–104, Sep. 2019. https://doi.org/10.30534/ijatcse/2019/1581.42019
74. E. M. Saida, E. B. Younès, and H. Nabil. **Towards a reference big data architecture for sustainable smart cities**, Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 1, pp. 820–827, Feb. 2020. https://doi.org/10.30534/ijatcse/2020/118912020
75. A. Bola and T. Qing. **The relationship between unified communications and big data analytics**, Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 1.1, pp. 191–194, 2019. https://doi.org/10.30534/ijatcse/2019/3681.12019