ASSESSMENT OF DATA QUALITY DIMENSIONS INFLUENCING BIG DATA ANALYTICS ROLE IN SUSTAINABLE DEVELOPMENT GROWTH PERFORMANCE

Nur Khairiah Muhammad*
Nor Hasni Osman
Nurul Azita Salleh

School of Technology Management and Logistics
Universiti Utara Malaysia
06010 Sintok, Kedah, Malaysia

*Corresponding author: nurkhairiahmd@gmail.com.my

Abstract : *Many nations are increasingly interested in the value of massive amounts of data, driven by the growing importance of Big Data Analytics (BDA) in today's competitive landscape. The adoption of Big Data Analytics enables organizations to strategically enhance their operations efficiencies, gain a competitive edge, and sustain long-term growth. In the context of Sustainable Development Growth (SDG), ensuring high data quality becomes even more critical, as decisions based on inaccurate or incomplete data can lead to suboptimal outcomes and potentially adverse environmental or social impacts. Prior research on Big Data Analytics has ignored data quality in favor of adding more big data attributes – referred to as Vs (volume, variety, velocity, etc.) Poor quality, outdated, and incomplete data can result in inadequate decision-making. Therefore, the primary aim of this study is to explore the importance of Data Quality Dimensions (DQD) and Big Data Analytics adoption can influence the effective impact measurement and data collection for the success of SDGs. Key variables from research literature reviews were incorporated into the research framework for this study. This study used quantitative method of cross-sectional survey to data professional practitioners and management board (senior and middle managers) that involved in Big Data Strategies within Malaysia. By introducing novel insights in the realm of Big Data Analytics, this study contributes to the body of literature and serves as a valuable resource for future scholars and industry practitioners who wish to investigate BDA solutions associated to performance of SDGs.*

Keywords: *Big Data Analytics (BDA), Data Quality Management, Data Quality Dimension (DQD), Sustainable Development Growth (SDG)*

2024 JGBSE
Article Received : 06 January 2024, Revised: 13 January 2024, Accepted 20 January 2024

Introduction

In the era of digital transformation, the proliferation of data has reached unprecedented levels, leading to the emergence of "Big Data." Coined by Doug Laney (2001), big data encapsulates the exponential growth and accessibility of structured, semi-structured, and unstructured data from diverse sources such as social media, sensors, and digital transactions. This overflow of data presents both opportunities and challenges across various sectors, including business, healthcare, and academia, owing to its vast volume, velocity, and variety. Big data holds the promise of revolutionizing decision-making processes, driving innovation, and enhancing efficiency. However, as organizations struggle with the complexities of big data, understanding its challenges becomes imperative for leveraging its full potential.

The complexity generated by big data created a range of situations in which companies are finding it difficult to develop a strategy that will allow them to leverage the data to extract business values and provide insights that will give them a competitive edge. Organizational ramifications: the operation's profitability and productivity rates would have dropped from 5% to 6%. (McAfee & Brynjolfsson, 2012). The organizations missed the chance to determine the primary effects on cutting business expenses and business strategy data. The organizations also missed the chance to determine the potential effects on cutting business expenses and to obtain business strategy data. As a result, businesses face the risk of jeopardizing the standard of services rendered and the effectiveness of corporate decision-making (Davenport et al., 2012).

With current trends, the globe is currently undergoing an industrial development cycle that involves automating and integrating artificial intelligence into industries and facilities. Big Data comes in many forms; structured and unstructured, much of it is imprecise, inconsistent, confusing, or ambiguous, which increases the risk of making incorrect decisions (Ijab et al., 2019; Janssen et al., 2017; Juddoo, 2016; Onyeabor Grace Amina, 2018). According to (Taleb et al., 2018) , 80% of the data generated is thought to be unstructured. Indeed, evaluating the quality of unstructured data is a tedious task and laborious. Working with a lot of unstructured data is one of the biggest obstacles to implement big data analytics. (Cai & Zhu, 2015). The need for high-quality data is essential if firms are to devote more time to analytics techniques rather than putting in the tedious job of cleaning and extracting the data.

The value of data quality is prominently important (X. L. Meng, 2021). As mentioned by Wamba et al. (2019), it is intriguing fact that Big Data Analytics implementations have to face the lack of data quality control mechanisms applied prior to data usage. Lack of high quality data is a common pitfall to Data Analytics adoption (X. L. Meng, 2021). For example, in the realm of Big Data utilization, the varying scales of maintaining data quality, along with interconnectedness, variations of data attributes contribute to increased complexity in data. Conducting analytical work becomes a multi-step and iterative process that includes data exploration, data cleansing, transformation, building analytical models, interpret and visualize. If the underlying analytical performs poorly, it adds latency to the process (Zikopoulos et al., 2012).

Decision-making depends on data, which also serve as the foundation for accountability. Ensuring data quality is crucial in the context of sustainable development, since decision-making processes depend on the availability of accurate and trustworthy data .Effective policy creation, monitoring, and evaluation become nearly impossible in the absence of high-quality data that provides the appropriate information on the right things at the right time (Lucivero, 2020). According to Schoenherr & Speier-Pero (2015), lack of data quality control would hamper the data analytics adoption and affect in providing accurate management decisions. The quality of data is made more difficult in the situations by the abundance of data and the variety of decision-making duties. Decision-makers are compelled to adapt to changing conditions since data is available anywhere. Likewise, in the realm of social development and poverty alleviation, inaccurate data pertaining to economic growth, healthcare access, school attainment, and income distribution can hinder the implementation of successful policies and initiatives that cater to marginalized communities (WEF, 2017).

Several studies have examined the potential of data quality for Big Data Analytics Adoption; nevertheless, certain problems regarding the factors that influence the dimensions of data quality are still unresolved (Haryadi et al., 2016; Wahyudi et al., 2018). Moreover, research on Big Data Analytics and data quality is still in its early stages and has not yet reached a satisfactory degree of maturity (Taleb et al., 2018). That leads to the fact that a thorough investigation into data quality is desperately needed to identify the most crucial aspects for Big Data Analytics Adoption implementation.

The Role of Big Data Analytics in Sustainable Development Growth (SDG)

Big Data Analytics enables the processing and analysis of vast amounts of data from diverse sources, offering insights into trends, patterns, and correlations relevant to sustainable development. By harnessing the power of advanced analytics techniques, policy makers within countries that participated in SDGs agenda can identify opportunities and challenges in areas such as economic growth environmental conservation, social equity.

The pursuit of high level of data quality within Big Data Analytics capabilities in the measurement of sustainability in global development has become a paramount objective, transcending traditional paradigms of economic growth and societal progress. Sustainable development, as articulated in the Brundtland Report (1987) highlights the necessity of meeting current needs without sacrificing the capacity of future generations to meet their own needs. This concept has evolved into a guiding principle informing policy agendas, business strategies, and academic discourse worldwide.

In 2015, the United Nations Sustainable Development Goals (SDGs) introduced a comprehensive blueprint for action which comprises 17 SDGs objectives and 169 metrices to end poverty, save the environment, and guarantee everyone's prosperity. The urgency of sustainable development arises from interconnected environmental, social, and economic challenges facing humanity today, including climate change, biodiversity loss, poverty, inequality, and social injustice. These complex issues require a holistic and integrated approach to development, emphasizing equity, resilience, and long-term viability (Rockström et al., 2009; Nation, 2015). Achieving these ambitious goals by 2030 demands collaborative efforts from governments, businesses, civil society, and academia, highlighting the need for interdisciplinary collaboration and innovative solutions (Nation, 2015).

Managing Data Quality

With the growth of big data, data quality administration has become a crucial area of study. Even little mistakes can result in inaccurate insights, process inefficiencies, revenue losses, and an inability to comply with industry and governmental regulations in the absence of proper data quality management. The testing tasks include securing organized huge data that is complex from various sources and successfully integrating it. When there is minimal data, it can be verified manually by making an inquiry and using Extract, Transform, Load (ETL) or Extract, Load, Transform (ELT). It is hard to gather, clean, coordinate, lastly get the required excellent data inside a viable time allotment. Since the amount of unstructured data in large datasets will provide an exceptional opportunity to transform unstructured data into organized categories and conduct additional data processing (Cai & Zhu, 2015). Indeed, data quality management remains a major challenge for the current data processing techniques.

Unstructured data is unquestionably growing rapidly. It is more difficult to find and analyze unstructured data than organized data. Organizations are recognizing the importance of segregating and organizing their data in a manner that allows for efficient analysis, lower the information asymmetry, enhance decision making while upholding business values (Solana-González et al., 2021). Data categorization involves separating different types of data, such as customer information, transaction records, or operational data, into distinct categories or databases. Deficient data created from unstructured data could come about into disregarded openings. Because of the assorted variety, speed and size of data moving through databases, it is to a great degree difficult to find patterns that prompt basic shattering conclusions.

The concept of quality is a multifaceted construct, and in order to fully communicate its dimensions, it must be combined with other aspects (Todoran et al., 2015). Datasets with inherent quality go into the intrinsic category, whereas contextual category draws attention to the task's need that data quality be taken into account. in the given setting. The representational category characterizes the quality of the data in respect to the data display, while the accessibility category highlights the significance of computer systems that facilitate data access (Wahyudi et al., 2018).

There are multiple dimensions in each category that are utilized as particular metrics for data quality. For example, the qualities in the contextual category are timeliness and relevance, while the dimensions in the intrinsic category are objectivity and accuracy. The dimensions under the representational category are interpretability and understandability, whereas the accessibility category's dimensions are ease of use and access security.

Data Quality Dimensions

The relationship of Big Data Quality across the Big Data adoption lifecycle is fundamental to measure the success of Big Data Adoption. The goal of this research is to provide a framework that can be used to manage Data Quality from the beginning to the analytics and visualization stage, hence assisting with decision-making. The definition of acceptable big data quality is heavily influenced by the requirements and kind of applications for Big Data.

Before beginning any Big Data-related project, a high-quality Big Data evaluation is important. This is because it is viable to avoid the significant expenses associated with processing meaningless data early in its lifecycle. Dealing with unstructured, schema-free data gathered from several sources. Additionally, a framework for managing the quality of big data can assemble quality control systems to monitor and guarantee data quality over the course of the big data lifecycle.

According to Wang and Strong (Wang, 1996) data quality dimension is a set of attributes for data quality that each reflect a different facet or idea of data quality. It also defines data quality as appropriate to use. Followings are the for attributes of Data Quality Dimensions:

Intrinsic - Refers to the quality of data on their own, which is often interpreted to mean how closely data values match actual or true values. Good data is by its own nature objective, true, and accurate, and it originates from reliable sources. Dimensions include accuracy, objective, believability, and reputation.

Contextual - Emphasizes the need that data quality be evaluated considering the current task, which is primarily defined as the degree to which the data are relevant (applicable) to the data user's work. Contextual Data Quality centers on the task of the data consumer rather than the representational context. For instance, timely and comprehensive information must be relevant to the consumer in order to be considered contextually appropriate. Value-added, relevance, timeliness, completeness, and the right quantity of data are some of the dimensions.

Representational - Indicates that the system must present data in such a way that it is easy to understand so that the consumer of data is able to interpret the data, which is defined as the degree of comprehensibility and clarity in which the data is presented.

Accessibility - Highlights the significance of systems, which are defined as the degree to which data consumers may access or get it. Moreover, the system needs to be safe.

Because of heterogeneity and incompleteness of data, it might be accepted that the obscure qualities are factually like the known qualities. Be that as it may, this will prompt a skewed outcome. Embracing big data analytics at organization level innovation means to accomplish firm heterogeneity and henceforth afford for high esteem and mindfulness in securing reasonable points of interest.

Importance of Data Quality

Data is the foundation of Big Data Analytics Adoption. Successful of big data adoption and utilization require the timely availability of high-quality data; data consistency guarantees that the same data is sent throughout the company; The extent to which information is available in the organizational repository is known as data completeness (Kwon et al., 2014). Data quality is crucial for data analysis used in decision making; low data quality can be a major obstacle of big data analytics adoption (Wessel, 2016).

Because of heterogeneity and incompleteness of data, it might be accepted that the obscure qualities are factually like the known qualities. Be that as it may, this will prompt a skewed outcome. The more extravagant an organization's asset, the more organization can receive new asset in agile way (Lieberman & Montgomery, 1998). Embracing big data analytics at organization level innovation means to accomplish firm heterogeneity and henceforth afford for high esteem and mindfulness in securing reasonable points of interest. Due to potential significance of data quality management to the success of Big Data Analytics adoption, this parameter in included in the research study to determinate the relationship between Data Quality Management with Big Data Analytic adoption.

In the context of Sustainable Development Growth, the reliability and accuracy of data are essential prerequisites for meaningful analysis and interpretation in big data analytics. Poor data quality, characterized by errors, inconsistencies, and incompleteness, can undermine the validity of analytical results and lead to flawed decision-making. Therefore, ensuring data quality is paramount to unlock the full potential of big data analytics for measuring sustainable development growth.

Enhancing the accuracy of current data is essential to facilitate nations in developing evidence-based strategic choices. Although there are national and international standards for evaluating the quality of data, their use in the assessment of particular SDG indicators is still in its infancy (X. Meng, 2019, 2021; Nilashi et al., 2023; Shanmugam et al., 2023). The quality of data is also impacted by incomplete data, which makes it more difficult to monitor SDG performance accurately and make organizational decisions. Any analysis technique may also be unable to correctly forecast SDG performance in the event that the data are incomplete. These situations give rise to a debate on whether the evaluation of a country's performance is impacted by the scarcity of data. In the absence of data, particularly high-quality data, sustainable development is bound to fail. Therefore, the absence of quality data could provide a significant obstacle to the process of evaluating the SDGs' performance.

Research Framework

Grounded on the theoretical framework of the model: Technology-Organization-Environment (TOE) framework, the research framework for this study was developed incorporating key variables derived from a review of the research literature on Data Quality perspective. Table 1 shows the research articles pertaining to Data Quality.

Year	Author	Title
2023	Shanmugam, D. B. Dhilipan, J. Prabhu, T. Sivasankari, A. Vignesh, A	The Management of Data Quality Assessment in Big Data Presents a Complex Challenge, Accompanied by Various Issues Related to Data Quality
2021	Meng, Xiao-li	Enhancing (publications on) data quality: Deeper data mining and fuller data confession
2019	Meng, Xiao-li	Data Science: An Artificial Ecosystem
2018	Wahyudi, Agung Kuk, George Janssen, Marijn	A Process Pattern Model for Tackling and Improving Big Data Quality
2018	Taleb, Ikbal Serhani, Mohamed Adel Dssouli, Rachida	Big Data Quality Assessment Model for Unstructured Data
2017	Janssen, Marijn van der Voort, Haiko Wahyudi, Agung	Factors influencing big data decision-making quality
2016	Haryadi, Adiska Fardani Hulstijn, Joris Wahyudi, Agung Van Der Voort, Haiko Janssen, Marijn	Antecedents of big data quality: An empirical examination in financial service organizations
2015	Todoran, Ion George Lecornu, Laurent Khenchaf, Ali Le Caillec, Jean Marc	A methodology to evaluate important dimensions of information quality in systems
2013	Kaisler, Stephen Armour, Frank Espinosa, J. Alberto Money, William	Big Data: Issues and Challenges Moving Forward
2013	Wielki, Janusz	Implementation of the Big Data concept in organizations – possibilities, impediments and challenges

Table 1. Summary of Data Quality Research Articles

Fostering sustainable development growth requires the successful integration of data quality considerations into the overall data lifecycle management process, under the Technology-Organization-Environment (TOE) model. By prioritizing data quality across technological infrastructure, organizational practices, and environmental compliance, organizations can enhance decision-making capabilities, drive innovation, and contribute significantly to the achievement of sustainable development goals in the long term (Nilashi et al., 2023). Figure 1 illustrates the integration of Data Lifecycle Management practices associated to Technology-Organization-Environment (TOE) framework, and the importance of data quality to be prioritized.

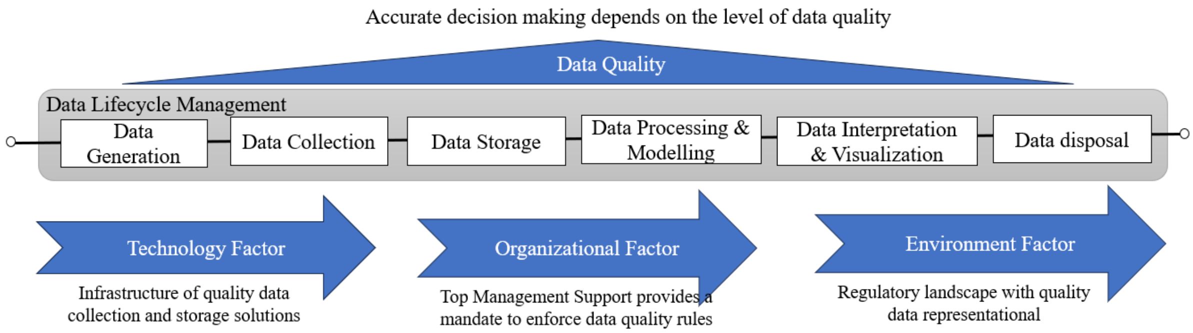

Figure 1. Integration of Data Lifecyle Management with TOE framework, in relation to Data Quality aspect.

Conclusion

The purpose of this study is to evaluate the significance of data quality in the context of Big Data Analytics role with the perspective of Sustainable Development Performance (SDG) performance. Without accurate and reliable data, effective monitoring, and evaluation become exceedingly challenging. The absence of quality data not only impedes the evaluation of SDG performance but also undermines efforts towards long term result monitoring. Therefore, it is imperative to prioritize data quality across all stages of the data lifecycle, integrating it within the Technology-Organization-Environment (TOE) framework. This holistic approach ensures that data quality considerations are embedded in technological infrastructure, organizational practices, and environmental compliance, thus facilitating informed decision-making, driving innovation, and contributing significantly to sustainable development growth. The quantitative method employed in this study, through cross-sectional surveys targeting data professionals and management boards involved in Big Data Strategies in Malaysia, provides valuable insights into the realm of Big Data Analytics for Malaysia. Such research contributes to the existing literature and serves as a vital resource for scholars and industry practitioners alike, offering guidance for advancing BDA solutions in support of SDG attainment.

References

Brundtland, G. H. (1987). Our Common Future ('The Brundtland Report'): World Commission on Environment and Development. *The Top 50 Sustainability Books*, 52–55. https://doi.org/10.4324/9781351279086-15

Brundtland, G. H. (1987). Our Common Future ('The Brundtland Report'): World Commission on Environment and Development. *The Top 50 Sustainability Books*, 52–55. https://doi.org/10.4324/9781351279086-15

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*, 2. https://doi.org/10.5334/dsj-2015-002

Davenport, T. H., Barth, P., & Bean, R. (2012). How ' Big Data ' is Different. *MIT Sloan Management Review*, *54*(1), 22–24.

Fosso Wamba, S., Akter, S., & de Bourmont, M. (2019). Quality dominant logic in big data analytics and firm performance. *Business Process Management Journal*, *25*(3), 512–532. https://doi.org/10.1108/BPMJ-08-2017-0218

Haryadi, A. F., Hulstijn, J., Wahyudi, A., Van Der Voort, H., & Janssen, M. (2016). Antecedents of big data quality: An empirical examination in financial service organizations. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 116–121. https://doi.org/10.1109/BigData.2016.7840595

Ijab, M. T., Salwana, E., Surin, M., & Nayan, N. M. (2019). *Conceptualizing Big Data Quality Framework From a Systematic*. 25–37.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Juddoo, S. (2016). Overview of data quality challenges in the context of Big Data. *2015 International Conference on Computing, Communication and Security, ICCCS 2015, October*. https://doi.org/10.1109/CCCS.2015.7374131

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387–394. https://doi.org/10.1016/j.ijinfomgt.2014.02.002

Laney, D. (2001). META Delta. *Application Delivery Strategies*, *949*(February 2001), 4. https://doi.org/10.1016/j.infsof.2008.09.005

Lieberman, M. B., & Montgomery, D. B. (1998). First-mover (dis) advantages: Retrospective and link with the resource-based view. *Strategic Management Journal*, *1125*(June), 1111–1125. https://doi.org/10.1002/(SICI)1097-0266(1998120)19:12<1111::AID-SMJ21>3.3.CO;2-N

Lucivero, F. (2020). Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives. *Science and Engineering Ethics*, *26*(2), 1009–1030. https://doi.org/10.1007/s11948-019-00171-7

McAfee, A., & Brynjolfsson, E. (2012). Big Data. The management revolution. *Harvard Buiness Review*, *90*(10), 61–68. https://doi.org/10.1007/s12599-013-0249-5

Meng, X. (2019). *Data Science : An Artificial Ecosystem*. *1*, 1–6.

Meng, X. L. (2021). Enhancing (publications on) data quality: Deeper data minding and fuller data confession. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *184*(4), 1161–1175. https://doi.org/10.1111/rssa.12762

Nation, U. (2015). *GLOBAL SUSTAINABLE DEVELOPMENT REPORT*. 1–69. ditjenppi.menlhk.go.id

Nilashi, M., Keng Boon, O., Tan, G., Lin, B., & Abumalloh, R. (2023). Critical Data Challenges in Measuring the Performance of Sustainable Development Goals: Solutions and the Role of Big-Data Analytics. *Harvard Data Science Review*, *5*(3). https://doi.org/10.1162/99608f92.545db2cf

Onyeabor Grace Amina, A. T. (2018). *Big Data and Data Quality*. *3*(1), 1–12. https://doi.org/10.1007/978-3-319-62461-7_1

Rockström , Johan; Steffen, Will; Noone, Kevin; Persson, A. (2009). Purification of M. leprae isolated from human skin biopsies. *Planetary Boundaries: Exploring the Safe Operating Space for Humanity*, *54*(3), 475–476.

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics*, *36*(1), 120–132. https://doi.org/10.1111/jbl.12082

Shanmugam, D. B., Dhilipan, J., Prabhu, T., Sivasankari, A., & Vignesh, A. (2023). The Management of Data Quality Assessment in Big Data Presents a Complex Challenge, Accompanied by Various Issues Related to Data Quality. *Research Highlights in Mathematics and Computer Science Vol. 8*, *April*, 78–91. https://doi.org/10.9734/bpi/rhmcs/v8/18858d

Solana-González, P., Vanti, A. A., García Lorenzo, M. M., & Bello Pérez, R. E. (2021). Data mining to assess organizational transparency across technology processes: An approach from it governance and knowledge management. *Sustainability (Switzerland)*, *13*(18). https://doi.org/10.3390/su131810130

Taleb, I., Serhani, M. A., & Dssouli, R. (2018). Big Data Quality Assessment Model for Unstructured Data. *Proceedings of the 2018 13th International Conference on Innovations in Information Technology, IIT 2018*, 69–74. https://doi.org/10.1109/INNOVATIONS.2018.8605945

Todoran, I. G., Lecornu, L., Khenchaf, A., & Le Caillec, J. M. (2015). A methodology to evaluate important dimensions of information quality in systems. *Journal of Data and Information Quality*, *6*(2). https://doi.org/10.1145/2744205

Wahyudi, A., Kuk, G., & Janssen, M. (2018). A Process Pattern Model for Tackling and Improving Big Data Quality. *Information Systems Frontiers*, *20*(3), 457–469. https://doi.org/10.1007/s10796-017-9822-7

Wang, R. Y. (1996). Beyond accuracy: What data quality meadans to data consumers. *Journal of Management Information Systems*, *12*(4), 5–34. https://doi.org/10.1080/07421222.1996.11518099

WEF. (2017). The Inclusive Growth and Development Report. *World Economic Forum*, *September*, 75. http://www3.weforum.org/docs/Media/WEF_Inclusive_Growth.pdf

Wessel, M. (2016). How Big Data Is Changing Disruptive Innovation. *Harvard Business Review*, 1–6. https://hbr.org/2016/01/how-big-data-is-changing-disruptive-innovation

Zikopoulos, P. C., DeRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D., & Giles, J. (2012). Harness the Power of Big Data. In *Osborne/McGraw-Hil*. https://doi.org/10.1007/s13398-014-0173-7.2