

Received: 12-03-2022 | Revised in Revision: 05-07-2022 | Accepted: 17-09-2022



Does big data ring a bell? Analyzing the understanding of big data at local government in Indonesia

Djoko Sigit Sayogo¹, Sri Budi Cantika Yuli²
*Accounting Department, Faculty of Economics and Business,
Universitas Muhammadiyah Malang, Indonesia¹*
*Economics Department, Faculty of Economics and Business,
Universitas Muhammadiyah Malang, Indonesia²*

DOI: <https://doi.org/10.22219/jibe.v6i01.25292>

Abstract

This paper attempts to portray public managers' perceptions concerning the big data concept using the case of a local government level in Indonesia. This study conducted in-depth semi-structured interviews with sixteen (16) top-level executives at the agencies and offices of the Regency of Bojonegoro, Indonesia, to gather the data. Our findings demonstrate that most of the executives from the agencies and offices in the Regency of Bojonegoro cannot fully explain what constitutes the big data indicators. However, the majority recognized big data from the volume perspective. Although, several executives mixed up big data with the infrastructure needed to manage the data in terms of storage and preservation. As such, this confusion could presumably also manifest in their comprehension of using data for decision-making. The lack of trust in data, lack of understanding of big data, and overconfidence in personal insights have driven the decision-makers to forego data and rely on intuition and experiences in making decisions.

Keywords: big data; data-driven decision-making; trust on data

Introduction

Government efforts to develop and execute smart, efficient, and more responsive decisions to current social problems are further enhanced by advances in information and communication technologies (ICTs), data, and information (Scholl & Scholl, 2014). The use of data to support a response to the complex social issue has increasingly gained governments' interest (Akter et al., 2019). For that, government worldwide has progressively democratized data to generate and maximize value creation through effort such as open government data (van Oijen, Ubaldi, & Welby, 2019). The advancement of wireless transmission, cloud computing, network communication, smart mobile devices, and social media results in the rapidly increasing data volume in various structures and forms (Zhou, Fu, & Yang, 2016). The proliferation of social media in supporting social interactions daily also raises the sheer volume and richness of human-generated data. The data is predominantly unstructured/semi-structured and its constitutes 95% of all data available (Gandomi & Haider, 2015; Olshannikova et al., 2017). The increased volume and variety of data arguably provide the government with immense opportunities to find an effective solution to growing social issues such as politics (Aradau & Blanke, 2016), better traffic control (Saldana-Perez & Moreno-Ibarra, 2016), or other pressing societal problems.

The phenomenon of vast and increasing volume and variety of data generated through digital technologies and data science innovation is often encapsulated in the term big data. Big data very frequently being defined from the size dimension; in fact size often become the only dimension to define big data (Gandomi & Haider, 2015). Thus, some studies associates big data as large and intricate data that is beyond the capability of existing data processing application to analyze (Wolfert et al., 2017). Following the hype, various studies focused more on instigating the way to use and analyze big data (Olshannikova et al., 2017).

¹E-mail: dsayogo@umm.ac.id

²E-mail: cantika@umm.ac.id

Copyright (c) 2022 Djoko Sigit Sayogo, Sri Budi Cantika Yuli



Creative Commons License

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Many scholars dedicated on studying how the utilization of big data provide an answer to the current challenges facing industry or government (see Fleming et al., 2018; Aradau & Blanke, 2016; Saldana-Perez & Moreno-Ibarra, 2016). Other studies concentrate on finding solution on the inadequacy of current data processing applications from many domains, including technology perspectives (see for instance Liu, Liu & Ansari, 2014; Hirzel et al., 2013), psychology (see for instance Cheung & Jak, 2016; Cao, Meyer-Lindenberg, & Schwarz, 2018), and other domains.

On the other hand, discussion on the concept of big data is fragmented (Gandomi & Haider, 2015); moreover less attention has been available on clarifying the concept of big data (Olshannikova et al., 2017; Mazieri & Soares, 2016). Using the word “vagueness” borrowing from Venkat Krishnamurthy, Director of Product Management at YarcData, Moothy et al. (2017) point to the need to pay more attention to the confusion over the meaning of big data. Majority of the big data definition focuses on the characteristics and technological features; the most popular one such as the “3 V’s” defining big data proposed by Douglas Laney (De Mauro, Greco, & Grimaldi, 2016).

An understanding of the users’ perceptions of big data become necessity to achieved the benefits of big data application (Fleming et al., 2018). Misunderstanding of the conception of big data lead to distrust and significantly affect the usage of big data for decision-making (Egan & Haynes, 2018). Although the hype for the quest to define big data concept majority happened in around 2016 to 2017 in the developed countries’ perspectives (see Moothy et al., 2017; Olshannikova et al., 2017), there might be a different story when looking from the developing countries’ viewpoints. As such, this paper attempt to portray the understanding of the public managers at local government level regarding the concept of big data from their perceptions. This study conducted in-depth semi-structured interviews with sixteen (16) top-level executives at the agencies and offices of the Regency of Bojonegoro, Indonesia to gather the data.

This paper consists of 4 sections, including the preceding introduction. Section 2 highlights studies evaluating big data and its dimensions. Section 3 presents a description of the research methodology. Section 4 discusses the results, findings, discussion of the findings, and conclusion.

Literature Review

Many scholars argue that the definition of big data is currently fragmented and undecided (Gandomi & Haider, Olshannikova et al., 2017; Mazieri & Soares, 2016; Moothy et al., 2017). Size becomes the first characteristics commonly presented when asked about the definition of Big Data (Kitchin, 2014; Gandomi & Haider, 2015). However, big data is more than just the amalgamation of massive data. For instance, the government has been collecting massive data through the national census (Kitchin, 2014). There are different inherent characteristics of big data compared to traditional and small data (Kitchin, 2014). De Mauro, Greco, and Grimaldi (2016) propose three classifications of big data definition based on: a) the attributes of data, b) the technological needs, and c) the social impact.

Big data Early studies focus on the characteristics of data in defining the concept of big data. The commonly known definition is the “3 Vs.” of big data proposed by Douglas Laney from Gartner Inc, underpinning the three dimensions of big data comprising of volume, velocity, and variety (Beyer & Laney, 2012; De Mauro, Greco & Grimaldi, 2016). Volume refers to the abundant and exhaustive data collected through directed, automated, and volunteer systems (Kitchin, 2016). Variety is the heterogeneity of structured, semi-structured, and unstructured data (Cukier, 2010; Gandomi & Haider, 2015). Velocity depicts the rapid rate of data collection processes (Gandomi & Haider, 2015). Many other “V’s” are added to represent the characteristics of big data, including veracity. As the fourth V, veracity represents data quality in terms of data integrity and accuracy that provide trust and confidence to the users trust toward the data usage (Moothy et al., 2017). Other scholar focuses in highlighted the different attributes of data, such as the granularity, diversity, and interconnectivity of data (Kitchin, 2016).

The second group of definitions underlined the technological need to analyze the data by arguing the need to have the complex and massive computing power to analyze big data. For instance, big data is associated with the need for new and innovative forms of processing (Bayer & Laney, 2012) or computational and storage requirements (de Mauro, Greco, & Grimaldi, 2016). The final group focuses on the impact of big data by arguing for the impact of big data as the game-changing strategy for many institutions, including the industry and government (Fleming et al., 2018; Aradau & Blanke, 2016; Saldana-Perez & Moreno-Ibarra, 2016). Aside from the three groups above, another scholar such as Kitchin (2014) emphasizes the continuous collection of granular data as a characteristic of big data.

Despite the many different perspectives in understanding big data definition, most scholars agree on the three data collection processes for big data (Kitchin, 2016). The first is the controlled and directed system in which the purpose, processes, and boundary of data to be collected are specified in advance. The data from the controlled and directed systems are limited in scope and size and was collected through

sampling techniques (Miller, 2010; Kitchin, 2014). Controlled and directed data represent data that the data owner consciously submits through certain mechanisms propagated by an institution, such as government census data or survey data. The second system is an automated system that represents the automatic capturing of data due to certain technologies. The data is often submitted to the mechanism through the consent of the data owner. An example of this type of data is consumer data inputted through the point of sale system (Kitchin, 2014; 2016). Finally, the volunteer system represents the data generated through interactions in social media and computer-mediated communications (Kitchin, 2016). The technologically mediated social interactions through popular social online services resulted in a tremendous amount of data related to people and their behaviors (Olshannikova et al., 2017). The data represents the “digital breadcrumbs” as the side effects of people using and consuming a large amount of digital content and interactions online or through smart devices (Pentland, 2009).

Table 1. Composition of the Interviewees

No.	Department	Position	Number of Interviewees
1	Department of Library and Information Service	Head of Department	1
2	Department of Information and Communication	Head of Department	1
3	Department of Education	Head of Department	1
4	Department of Civil Registry Services	Head of Department	1
5	Department of Trade and Micro Enterprise	Head of Department	1
6	Department of women's empowerment, child protection, and family planning	Head of Department	1
7	Department of Industry and Labor	Head of Department	1
8	Department of Social Services	Head of Department	1
9	Agency for Development and Regional Planning	Head of Department	1
10	Agency for Finance and Asset Management	Head of Department	1
11	Department of Village and Community Empowerment	Head of Department	1
12	Agency for Regional Investment	Head of Department	1
13	Department of Culture and Tourism	Head of Department	1
14	Agency for Personnel and Trainings	Head of Department	1
15	Agency for Regional Disaster Management	Head of Department	1
16	Department of Animal Husbandary and Fisheries	Head of Department	1
Total			16

The proponents argue that big data offer terrific possibilities shifting the scarcity of data into a data-rich society resulting in various benefits. The large and heterogeneous data has large potentials to improve productivity, competitiveness, and efficiency for the private sectors (Kitchin, 2016). Likewise, big data hold the promise of innovative ways to solve complicated social issues facing the government. Such as using big data mining to improve public health (Pentland, Lazer, Brewer, & Heibeck, 2009) or in effectively managing the pandemic (Harrison & Pardo, 2020), or bettering the traffic governance (Saldana-Perez & Moreno-Ibarra, 2016).

As such, the extraordinary potential value of big data can be unlocked only when it is used, among others, to drive decision-making (Gandomi & Haider, 2015). Using the insight gathered from the analysis of a vast amount of data as the basis of its decision is beneficial for an organization (Akter et al., 2019), particularly in administrative and policy settings (Mandinach, 2012; Hwang, Nam, & Ha, 2021). Brynjolfsson, Hitt, & Kim (2011) demonstrated the economic benefits of adopting data-driven decision-making for the business through econometric evidence. Decisions based on data is also critical in crises and

high uncertainty environment such as healthcare (Kennedy, 2009). Grounding policy on data produces more transparent, faster, and precise policy actions that could improve the efficiency and effectiveness of the government's operations. The government could build a better trusting relationship with the public by establishing their decisions based on rigorous and accurate data (Matheus et al., 2018; Harrison & Sayogo, 2014). On the contrary, lack of data-driven insight that leads to imprecise policy decisions and actions could result in widening public distrust (Matheus et al., 2018).

Data availability and advancement in data science do not guarantee data usage in the decision-making process (Marsh, Paine, & Hamilton, 2006). Studies point to the various challenges facing public officials in adopting data-driven decision-making to support their works. The existence of resources, compliance with technologies, and skills are among the critical factors affecting data-driven decision-making (Brynjolfsson & McElheran, 2016). The need for high-quality, correctness, timeliness, and costs of using and analyzing data restrict the ability to use data-driven decision-making (Kennedy, 2009). The need for cross-disciplinary skills to analyze and use data poses another challenge for data-driven decision-making (Towe et al., 2020); as well as the data variety, availability, and quality (Witjas-Paalberends et al., 2018)

Research Method

The study conducted in-depth semi-structured interviews to identify and analyze the perceptive understanding of the officials at the local government level in Indonesia concerning big data. In-depth interviews with sixteen (16) top-level public managers of agencies and offices at the Regency of Bojonegoro, Indonesia, were conducted to gather the empirical data (see table 1). All interviews were recorded and transcribed as a whole to obtain rich empirical data. Interviews were transcribed and analyzed following an inductive logic and using grounded theory techniques (Strauss & Corbin, 1997). The analytical processes conducted by the research team were iterative.

Result

This study conducted a case study encompassing in-depth interviews with the executives of agencies and offices in the regency of Bojonegoro. This paper presents the preliminary efforts to critically analyze the executives' awareness about big data from the subjective opinion of the interviewees. These represent the indicator of big data understanding from the perceptions of the direct users. The interviewees were asked to specify their recognition of Big Data from their perspective. The understanding of big data from the subjective perceptions of the interviewees was measured based on the definition of Big Data from Kitchin & McArdle (2016), focusing on the three V's of volume, velocity, and variety of data.

Inferring from the word "big" in big data, many interviewees easily associate big data with the volume of data, asserting that big data represent "a very large and comprehensive data and the data is real and up-to-date." However, many of the interviewees fail to see the correlation of big data with various data and instead relate big data with data uniformity. The interviewees perceived that big data represents the collection of data with a similar format. As stated by an interviewee, "Big data means one integrated data, the format is uniform and usable by anybody." Furthermore, all of the interviewees failed to recognize the importance of unstructured data. They pointed out that big data sources come only from structured data managed by government agencies and offices. The interviewees asserted the notion that "big data means all data related to the existing data in each agency and office such that the agencies and offices are responsible for updating the data."

Many of the interviewees confused big data with a database. They mostly relate big data with the accumulation and integration of data from different agencies and offices into a single repository. One of the interviewees even equated big data with a data bank and stated that "big data is a large data, it is a collection of various data, is a data bank." Similarly, other interviewees mixed the notion of big data with tools to analyze data. One of the interviewees asserted that big data represents "all data within the scope of the agencies and offices in the Regency of Bojonegoro into [one] application." Another interviewees specifically refer big data as tools instead of data and stated that "Big data is not large data but more like tools for any agencies or offices to fulfill their data needs."

Likewise, the interviewees were mindful of the importance of speed to collect data. The majority of the interviewees assert the influence of velocity in collecting data on the decision-making process. However, their understanding of velocity refers to the speed of updating the data managed by the government instead of the ability to collect structured and unstructured data rapidly. As asserted by one of the interviewees, "...speed is our main priority for decision-making. We need to aim for updating data every hour daily, you know, because you never know when the leader needs to make a decision. it happened suddenly disregard of time, so data has to be available with a snap of a finger."

The interview results demonstrate that experiences and intuition predominantly influence the current practices for making a decision. Only the upper echelon understood the need for data collection and used and was not yet fully understood by all levels and personnel at the agencies and offices. Therefore, the interview results indicated that the decision-makers rarely or never 100% use only data to base their decision. They supplemented the decision with the experiences, their own experiences, or benchmarking on other regencies or cities' experiences. One of the interviewees specified, "more than 50% experiences and the rest often data, real data off course [including] experiences from other regencies or cities. Experience is a must for the basis of decision-making."

The interviewees affirmed the use of experience to identify data relevance and choose which data to use for the decision. As acknowledged by one of the interviewees, "...from that experience, ya...not all data is used to make a decision. Yes, some data is important, but we feel not affect the decision; we choose only data that matter." The interviewees asserted that the issue's urgency would further influence how they will use experience and institution. The very urgent issue mainly will be dealt with using intuition; as stated by one of the interviewees, "intuition for urgent matter requires prompt action, we use intuition for things like that about more than 50% using intuition the rest experiences and supplemented by data."

Conclusions

Our findings demonstrate that most of the executives from the agencies and offices in the Regency of Bojonegoro cannot fully explain what constitutes the big data indicators. However, they understand certain substances of big data. In the sense that the majority recognized big data from the volume perspective. Although, several executives mixed up big data with the infrastructure needed to manage the data in terms of storage and preservation. As such, this confusion could presumably also manifest in their comprehension of data-driven decision-making. The findings indicate the persistence in using experience and intuition instead of using data to support decision-making. The findings concur with Egan & Haynes (2018). The lack of trust in data and overconfidence in personal insights have driven the decision-makers to forego data and rely on intuition and experiences. Our case indicates the lack of trust also stem from a lack of understanding of what constitutes big data and its benefits

Acknowledgment

This work was supported by the Ministry of Education, Culture, Research, and Technology of Indonesia under the Grant No. 007/AMD-SP2H/LT-MULTI-PDPK/LL7/2021. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Ministry of Education, Culture, Research, and Technology of Indonesia.

References

- Akter, S., Bandara, R., Hani, U., Wamba, S. F., Foropon, C., & Papadopoulos, T. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. *International Journal of Information Management*, 48, 85-95.
- Aradau, C., & Blanke, T. (2017). Politics of prediction: Security and the time/space of governmentality in the age of big data. *European Journal of Social Theory*, 20(3), 373-391.
- Beyer, M. A., & Laney, D. (2012). The importance of "big data": A definition. Gartner. G00235055.
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance?. Available at SSRN 1819486.
- Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133-39.
- Cao, H., Meyer-Lindenberg, A., & Schwarz, E. (2018). Comparative evaluation of machine learning strategies for analyzing big data in psychiatry. *International journal of molecular sciences*, 19(11), 3387.
- Cheung, M. W. L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/meta-analyze approach. *Frontiers in psychology*, 7, 738.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*.
- Egan, D., & Haynes, N. C. (2018). Manager perceptions of big data reliability in hotel revenue management decision making. *International Journal of Quality & Reliability Management*. 36(1), 25-39.

- Fleming, A., Jakku, E., Lim-Camacho, L., Taylor, B., & Thorburn, P. (2018). Is big data for big farming or for everyone? Perceptions in the Australian grains industry. *Agronomy for Sustainable Development*, 38(3), 1-10.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Harrison, T. M., & Sayogo, D. S. (2014). Transparency, participation, and accountability practices in open government: A comparative study. *Government information quarterly*, 31(4), 513-525.
- Hirzel, M., Andrade, H., Gedik, B., Jacques-Silva, G., Khandekar, R., Kumar, V., ... & Wu, K. L. (2013). IBM streams processing language: Analyzing big data in motion. *IBM Journal of Research and Development*, 57(3/4), 7-1.
- Hwang, S., Nam, T., & Ha, H. (2021). From evidence-based policy making to data-driven administration: proposing the data vs. value framework. *International Review of Public Administration*, 1-17.
- Kennedy, D. (2009). Tough times call for data-driven decisions: but finding the right data sources requires ingenuity and resourcefulness. *Healthcare Financial Management*, 63(3), 102-104.
- Kitchin, R. (2016). Big data. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, 1-3.
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 2053951716631130.
- Liu, J., Liu, F., & Ansari, N. (2014). Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop. *IEEE network*, 28(4), 32-39.
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71-85.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making Sense of Data-Driven Decision Making in Education: Evidence from Recent RAND Research. Occasional Paper. Rand Corporation.
- Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, 37(3), 101284.
- Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big data: Prospects and challenges. *Vikalpa*, 40(1), 74-96.
- Olshannikova, E., Olsson, T., Huhtamäki, J., & Kärkkäinen, H. (2017). Conceptualizing big social data. *Journal of Big Data*, 4(1), 1-19.
- Pentland, A. (2009). Reality mining of mobile communications: Toward a new deal on data. *The global information technology report 2008–2009*, 1981.
- Pentland, A., Lazer, D., Brewer, D., & Heibeck, T. (2009). Using reality mining to improve public health and medicine. In *Strategy for the Future of Health* (pp. 93-102). IOS Press.
- Saldana-Perez, A. M. M., & Moreno-Ibarra, M. (2016). Traffic analysis based on short texts from social media. *International Journal of Knowledge Society Research (IJKSR)*, 7(1), 63-79.
- Towe, R., Dean, G., Edwards, L., Nundloll, V., Blair, G., Lamb, R., ... & Manson, S. (2020). Rethinking data-driven decision support in flood risk management for a big data age. *Journal of Flood Risk Management*, 13(4), e12652.
- Witjas-Paalberends, E. R., van Laarhoven, L. P. M., van de Burgwal, L. H. M., Feilzer, J., de Swart, J., Claassen, E., & Jansen, W. T. M. (2018). Challenges and best practices for big data-driven healthcare innovations conducted by profit–non-profit partnerships—a quantitative prioritization. *International Journal of Healthcare Management*, 11(3), 171-181.
- Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, 215-225.