What is Effective Use of Big Data? The Consensual Definition of Effective Use of Big Data

Feliks P Sejahtera Surbakti

1Industrial Engineering Department, Atma Jaya Catholic University of Indonesia, Jakarta, 12930, Indonesia
feliks.prasepta@atmajaya.ac.id

Abstract. Big Data has attracted significant attention in recent years due to its potential. The use of Big Data is believed to be another source of competitive advantage for organizations. A leading organization in the future is an organization that can use Big Data effectively. However, what constitutes effective use of Big Data remains elusive and evidence suggests that most Big Data projects fail to deliver on their expectations and that executives are slow to adopt Big Data. Existing Engineering Management and Information Systems research have explored effective use in a variety of contexts; however, it is yet to specifically consider the unique characteristics of Big Data. Accordingly, there is a need for rigorous study on what the core of the concept of effective use of Big Data. The aim of the study is to understand better the effective use of Big Data. A multiple case studies approach is followed in the investigation. Eight organizations from private and public organizations were investigated. Data were gathered primarily using semi-structured interviews and through document analysis. The study results show the core of the concept of effective use of Big Data.

Keywords: Big Data. Effective Use, Definition. Qualitative Research

1. Introduction

The effective use of Big Data has the ability to realize data into actionable insights. This paper builds understanding on what constitute to effective use of Big Data. Study shows that although most organizations have access to Big Data, they do not have the capacity to use it effectively [1, 2]. Indeed, evidence suggests that organizations struggle with how to use Big Data effectively [3, 4]. Moreover, another study show that although executives understand the need for the realization of Big Data for increasing profits, less than a quarter of them believe their organization makes value from Big Data use [5]. In spite of the failures as well as slow progress, the opportunities of Big Data is promising and widely acknowledged, for example [6] highlights that Big Data can deliver a new wave of productive growth and change economies.

How to use Big Data effectively is a complex phenomenon. So far, the effective use of Big Data is rarely covered in literature. It is still unclear what effective use of Big Data is. While the study of effective use should give direction in such a context, Theory of Effective Use [7] was established on the foundation of common Information Systems and Management, which may not contain the unique attributes of Big Data. Developing a consensual definition of effective use of Big Data is a necessary step towards providing guidance to organizations such that the worth of Big Data can be better realized. This research is aligned with [8], who states that academicians can make important
contributions to academic and practice, by “introducing new constructs” and “by better conceptualising existing constructs”, which is what this study does with how to use Big Data effectively.

The objective of this study is to define the core of the concept of effective use of Big Data in public and private sectors context. Effective use in the Information Systems and Management discipline is defined as the degree to which the use of a system increases achievement of the goals an organization has set concerning the system [7]. While there is agreement on this definition in the discipline, there is a lack of understanding of effective use in the context of Big Data. Consensus, however, is important. Ronda-Pupo and Guerras-Martin [9] argue that the level of consensus on a definition of a concept shown by a scientific community can be used as a measure of the development of a body of knowledge. Big Data has instead developed rapidly and arguably in a disorganized manner, that such a universally accepted formal statement denoting what is the meaning of the effective use of Big Data does not exist. Even though theory and definition of effective use in the Information Systems and Management discipline have been proposed, the author shows that the knowledge developed on effective use of Information Systems and Management needs to be revisited in the Big Data context.

The rest of the paper is organized as follows. In the following section, I explain my method to undertake the formulation of definition and proposing a core of concept and definition of effective use of Big in the result and discussion section. Finally, I conclude the paper with a summary of the contribution and discussion of related limitations.

2. Methods
In this study, an exploratory case studies method is selected because it allows for a collection of rich descriptive data and is suitable for presenting a unique case in research areas where there is lack of theory [10]. To understand how to use big data effectively, data are gathered primarily using semi-structured interviews and through document and video analysis. The semi-structured interviewing technique was chosen as the main data collection technique in this case study approach to enable the researcher to ask for more detail, probe an issue, and go back and forth among important points and request further explanation. Interviews are suitable to explore topics and explain other findings [11].

It is important to choose the right case organization and have available to the researcher data from the case organization when conducting case studies. As stated earlier private and public sectors as chosen the case organizations. This decision was made as Big Data provides organizations with opportunities that are still largely unexploited. However, without a better understanding of the complex issues involved in real-word settings, the effective use Big Data remains elusive. Therefore, the public and private organizations offered an excellent opportunity to study, in which using Big Data is critical to their competitiveness and success.

A total of 12 participants from the eight organizations were interviewed, as summarized in Table 1. Eisenhardt [12] suggests from 4 to 10 is ideal number of cases. The interviews averaged 60 minutes in duration, with a total time of 781 minutes. Based on the interviews data set, the author found that saturation happened within the twelfth interviews. This empirical evidence is supported by Guest, Bunce [13] who argued within the twelve interviews, new themes emerged infrequently as analysis continued.

| Participant | Role                                | Experience | Industry  |
|-------------|-------------------------------------|------------|-----------|
| Participant 1 | Research Manager                 | >2 years   | Resources |
| Participant 2 | Artificial Intelligence Manager    | >7 years   | Resources |
| Participant 3 | Risk and Information Management Manager | >3 years   | Finance   |
| Participant 4 | Engagement Manager                | >2 years   | Logistics |
| Participant 5 | CEO and Founder                  | >4 years   | Start-Up  |
| Participant 6 | Data Analyst                      | >2 years   | Health    |

Table 1. Interview Participants
The data analysis process was conducted iteratively as encouraged by the grounded theory methodology. The coding started with a process referred to as microanalysis [14], that is, a line-by-line analysis of each semi-structured interview transcript to identify initial codes was carried out. The author used NVivo 12 to organise and analyse the data to generate and label nodes/codes following the guidelines from Saldaña [15].

3. Result and Discussion

This section provides the research study’s findings assembled through a general inductive approach to qualitative research. Triangulation of the 12 participants was an important strategy that facilitated any inclination towards researcher bias. The analysis was anchored by following research question: What is the meaning of effective use of Big Data?

Further in-depth analysis produced “units of meaning” used to reconstruct participants narratives into a cohesive yet agile statement about the meaning of effective use of Big Data, without losing the essence of narratives in their entirety. Burton-Jones and Grange [7] put forth the concept of effective use in Information Systems and Management context, which refers to the use of a system “in a way that helps attain the goals for using the system” (p.2) and propose adaptation and learning actions as the drivers of effective use. Their definition of use draws on information theory to fully integrate the nature and purpose of an information system, and thus is inclusive of not only the system, but also of the user and the tasks. From this definition, there are three keywords for effective use: “in a way”, “using the system” and “attain the goals”, and the ultimate measure of effective use is to attain the goals (G).

On the other hand, De Mauro, Greco [16] analysed the most frequent keywords included in Big Data related articles and considering their mutual relationships and identified four top research themes in current literature, namely: 1. Driver 2. Information, 3. Technology, 4. Methods. To define effective use of Big Data, the author classified the definition into 5 themes that derived from Burton-Jones and Grange [7] and De Mauro, Greco [16]: Driver (D), Information (I), Technology (T), Methods (M), and Goal (G).

The appendix summarises the meaning of effective use of Big Data that come from empirical evidence. By looking to the definition that presented by participants and the whole interviews, the author proposes that the core of the concept of effective use of Big Data can be claimed by:

• Define the problem to describe effective Big Data initiatives should start from clear goals and problem driven instead of fuzzy objectives and data driven and show the value to the organization; (D)
• Using insights from information to describe the transformation from information to insights as the main pathway Big Data is impacting organizations and society; (I)
• Build Big Data infrastructure and data analytics to describe Big Data initiatives should be use driven instead of technology driven, Big Data technology and analytics as support mechanisms not the goals; (T)
• Scalability and sustainability mean that the organizations must be able to choose the scope of the projects and get the data, powered by a sustainable growing process from performance and financial perspectives; (M)

| Participant | Role                        | Experience | Industry |
|-------------|-----------------------------|------------|----------|
| 7           | Director, Health Analytics  | >15 years  | Health   |
| 8           | Project Manager             | >10 years  | Health   |
| 9           | Analytics Architect         | >6 years   | Sports   |
| 10          | Sport Scientist             | >10 years  | Sports   |
| 11          | Data Scientist              | >3 years   | Energy   |
| 12          | Head of Data & Analytics    | >10 years  | Finance  |
• Make solution to describe effective use of Big Data means solve the problem. (G)

The author argues the definition of effective use of Big Data as follows:

“Effective use of Big Data refers to harnessing value from Big Data by defining the problem, showing the value to the organization, and making solution by using insights from data and build Big Data infrastructure and data analytics by considering scalability and sustainability.”

4. Conclusion

The analysis of twelve participants provides the first empirically derived picture of the emerging organizational construct that is supposed to foster effective use of Big Data in the organization. It has revealed several insights of theme for the definition of effective use of Big Data. Each of theme is described with a comprehensive set of explanation that help to develop an effective use of Big Data and further research such an organizational construct.

The theoretical contribution of this study is the definition of effective use of Big Data, which is both empirically grounded and theory informed. The study provides better understanding of the essential concept of effective use of Big Data. The theory on Information Systems and Management in Big Data context is scarce. Furthermore, effective use is not well captured in the Engineering Management and Information Systems literature (17) and towards context-specific of Big Data, rather than general theories. The study is one of the first effort to document the effective use of Big Data. We, therefore, contribute to study on effective use of Big Data by providing concrete concept for the organizational effective use of Big Data. The author further adds to effective use in Information Systems and Management research by identifying the core concept and definition of effective use in Big Data context. These contributions build an important foundation for future study.

The selection of the sample and the ways of gathering data exposes the study to certain limitations. Overall, two different types of limitation apply and shall be addressed in future research. First, for restrictions of interviewee availability, only one interview was conducted for most of the organizations. A more diverse perspective from users of the organizations and from different roles within the organizations may reveal a more differentiated picture. Future research should consider a case study approach and analyse the organization from different stakeholder perspectives.

Second, this study focuses on organizations that are early adopters of the Big Data use, the major organizations that have already established Big Data use initiatives and are addressing the issues of effective use. However, the author cannot say with certainty that these options are the only viable or most effective ones at this point. A broader study of organizations that want to realize value from Big Data within the organization using different approaches may reveal a better understanding of the effective use of Big Data.

Acknowledgements
I gratefully acknowledge the funding sources that made my PhD work possible; LPDP (Indonesia Endowment Fund for Education).

5. References

[1] Günther WA, Mehrizi MHR, Huysman M, Feldberg F. 2017. Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, 12(3), 191-209.

[2] Sejahtera F, Wang W, Indulska M, Sadiq S, editors. 2018. Enablers and Inhibitors of Effective Use of Big Data: Insights From A Case Study. Pacific Asia Conference on Information Systems (PACIS), Yokohama, Japan.

[3] Partners N. 2019. Big Data and AI Executive Survey 2019.

[4] Surbakti FPS, Wang W, Indulska M, Sadiq S. 2019. Factors influencing effective use of big data: A research framework. Information & Management In Press.
[5] Smart C, De Meyer-Heydenrych C, editors. 2018. The Use of Big Data in Marketing: An Emerging Market Financial Services Industry Perspective. The 2018 Annual Conference of the Emerging Markets Conference Board, Wits Business School.

[6] Mikalef P, Pappas IO, Krogstie J, Giannakos M. 2018. Big data analytics capabilities: a systematic literature review and research agenda. Information Systems and e-Business Management, 16(3), 547-78.

[7] Burton-Jones A, Grange C. 2012. From use to effective use: a representation theory perspective. Information Systems Research, 24(3), 632-58.

[8] Barki H, Paré G, Sicotte C. 2008. Linking IT implementation and acceptance via the construct of psychological ownership of information technology. Journal of Information Technology, 23(4), 269-80.

[9] Ronda-Pupo GA, Guerras-Martín LÁ. 2012. Dynamics of the evolution of the strategy concept 1962–2008: a co-word analysis. Strategic Management Journal, 33(2), 162-88.

[10] Yin RK. 2017. Case study research and applications: Design and methods. (London: Sage publications)

[11] Saunders M, Lewis P, Thornhill A. 2009. Research methods for business students. Pearson education.

[12] Eisenhardt KM. 1989. Building theories from case study research. Academy of management review, 14(4), 532-50.

[13] Guest G, Bunce A, Johnson L. 2006. How many interviews are enough? An experiment with data saturation and variability. Field methods, 18(1), 59-82.

[14] Strauss A, Corbin J. 1990. Basics of qualitative research techniques: Grounded Theory Procedures and Techniques. (Newbury Park, CA: Sage publications)

[15] Saldaña J. 2015. The coding manual for qualitative researchers. (London: Sage).

[16] De Mauro A, Greco M, Grimaldi M, editors. 2015. What is big data? A consensual definition and a review of key research topics. AIP conference proceedings, AIP.

[17] Burton-Jones A, Volkoff O. 2017. How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. Information Systems Research, 28(3), 468-89.

Appendix

| Participant | Effective Use of Big Data | D | I | T | M | G |
|-------------|---------------------------|---|---|---|---|---|
| 1           | *I think effective use of Big Data in the field of my work is about making solutions for business problems by completing prediction models and optimizations algorithms* | ✔ | ✔ | ✔ |
| 2           | *From business perspective, it means the scalability at first glance, so first of all, scalability to be able to choose what we were doing globally and to be able to get all the data that you're being provided by the sensory so far and all this new data that is coming from large device and be able to scale all this stuff but also sustainability because so I used to work as database professional so and I know that a lot of figures and row we were already able to scale lots of data using regular databases but the problem is that it was not kind of sustainable because the performance we started to decrease a large and also the cost is increasing the large. So, I think when we talk about Big Data, mainly what comes in my head is scalability, and on this scalability, which is powered by a sustainable growing process from a financial perspective but also for performance perspective* | ✔ | ✔ |
| Participant | Effective Use of Big Data                                                                 | D | I | T | M | G |
|------------|------------------------------------------------------------------------------------------|---|---|---|---|---|
| 3          | *I think its related to the first answer, basically bottom-up, it’s not top-down, bottom-up, what is the value, what is the potential benefit, and start from there,* because once you start with that, and then you can gather, okay I need this data, and it is the feasibility okay, I need this data. So, a lot of approaches basically getting the data first, and think about this case later, but I believe in use cases first and then get the data and then **build your infrastructure** based on whatever you need | ✓ |   |   |   | ✓ |
| 4          | *I think personally it has to come down to using a Big Data approach gives you some measure of uplift and whatever you're trying to do compared to a different approach, or no approach may be, some level of uplift in some level of measurable success because I think that's you know, this is doing analysis for analysis say, but then there's doing analysis to achieve an outcome, and ultimately you want to achieve an outcome that you can measure compare it to some benchmark.* |   |   |   |   | ✓ |
| 5          | *Always more efficient to start with the end in mind. Then move back to see what data the organization need to collect, and how would the organization remove the noise and do desired transformations* |   | ✓ |   |   | ✓ |
| 6          | *The mission of the company is to make solar energy much more accessible and much more useful, so do things like initial introduction [...] and helping save the customer money at the same time. So that's the effective use of the data, if we can take all this usage data and take price data, like tariff data, and also our knowledge about where the people live, and we can do things like put the customers together in a big virtual power plant, we can achieve our mission directions, we can save the customers money and it can be a win-win all around* |   |   |   |   | ✓ |
| 7          | *Ultimately in our environment, effective use of Big Data is what we can help to inform the development of health policy and shape the decisions of health of ministry, and he or she can feel that the decision that they take withstanding difficult requirements, sometimes that may override, the decision they make firmly on evidence* | ✓ |   |   |   | ✓ |
| 8          | *I'm not sure have a 100% answer for that one, I guess, I think effective use of Big Data is no not really that different from effective use of regular data is, making sure that being able to get to help you answer the questions that you want answered* | ✓ |   |   |   | ✓ |
| 9          | *Make a difference in the real world. So, the impact on the health service of Australians. It changes how Australians get their health services, how long they live, how they able to maintain their health. That's the context why we are doing data analytics in the health context* |   |   |   |   | ✓ |
| 10         | *Effective use of Big Data, mm, I don’t know the answer for that. I think it is making better use, not just, you bringing data, to the purpose of generate more value out of it so it’s like an asset so yeah you’re gonna have effective use of Big Data it needs to be able to be using way that adds value, otherwise all you’re doing is* |   |   |   |   | ✓ |
So effective use of Big Data means *influencing change; making decisions*. Well, it’s just using *visualization tools*. It’s using *information*, which is updated regularly with minimal input from humans, but it is pulling out the necessary information; live dashboards, tools which people can use without having to look at the data themselves. *And that’s how you’re able to visualise things, and to understand it, and to make change, and to effect change.*

It’s that making sure that we are *solving problems*, making sure that you can *show the value to the organization*, and *understanding the business need first*, and those *problem statements first*.