Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2018 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

6-26-2018

Enablers And Inhibitors Of Effective Use Of Big Data: Insights From A Case Study

Feliks Sejahtera
The University of Queensland, f.sejahtera@uq.edu.au

Wei Wang
The University of Queensland, w.wang9@uq.edu.au

Marta Indulska
Business School, The University of Queensland, m.indulska@business.uq.edu.au

Shazia Sadiq
The University of Queensland, shazia@itee.uq.edu.au

Enablers and Inhibitors of Effective Use of Big Data: Insights from a Case Study

Completed Research Paper

Feliks Prasepta Sejahtera
School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia QLD 4072; & Industrial Engineering Department, Unika Atma Jaya, Jakarta 12930
f.sejahtera@uq.edu.au;
feliks.prasepta@atmajaya.ac.id

Wei Wang
School of Information Technology and Electrical Engineering, The University of Queensland
St Lucia QLD 4072
w.wang9@uq.edu.au

Marta Indulska
UQ Business School
The University of Queensland
St Lucia QLD 4072
m.indulska@business.uq.edu.au

Shazia Sadiq
School of Information Technology and Electrical Engineering, The University of Queensland
St Lucia QLD 4072
shazia@itee.uq.edu.au

Abstract

Big data has attracted significant attention in recent years due to its potential. While many organisations have access to big data, there is a lack of evidence and guidance on effective use of big data. Information systems research has explored effective use in a variety of contexts. However, it is yet to specifically consider the unique characteristics of big data. This paper presents the results of an empirical study that aimed to identify significant enablers and inhibitors of effective use of big data using an exploratory case study as a research method. We found adequate system capabilities, established the culture of collaboration, and good working attitude to be the key enablers and poor data quality, lack of data understanding, data silos, lack of time, lack of cost-benefit analysis, lack of top management support, and lack of technical skills to be the key inhibitors.

Keywords: Effective Use, Big Data, Enablers, Inhibitors, Definition

Introduction

Big data is characterised by the so-called Vs, namely Volume (vast amounts of data generated), Variety (different type of data), Velocity (speed of data generated), Veracity (messiness of the data) and Value (turning data into useful insights) (Anuradha 2015). The last V – value – is to some extent dependent on how effectively organizations can use big data. The effective use of big data includes the complete technological stack, including raw data processing, storage, ways of managing data, and analytics (Merino et al. (2016)) and information systems that support the use of big data. The extent of big data projects success will be determined by how effectively an organization implementing big data.

Burton-Jones and Grange (2012) define effective use as using a system in a way that helps attain goals. Existing information systems research has explored effective use in a variety of contexts. However, it is yet to specifically consider the unique characteristics of big data. Yet, effective use of big data is said to have the potential to transform economies and deliver a new wave of growth (Henke et al. 2016; Manyika et al. 2011). At the same time, however, research indicates that while most organisations have access to big data, they don't necessarily have the capacity to use it effectively (LaValle et al. 2013). Accordingly, understanding the enablers and inhibitors that influence effective use of big data is a necessary step towards providing guidance for organisations, such that the full potential and value of big data can be realised.

Prior research on effective use in the context of information systems has focused at the system level i.e. effective use of information systems (Agarwal et al. 2010; Boudreau and Seligman 2006; LeRouge et al. 2007; Pavlou et al. 2008). For example, Burton-Jones and Grange (2012) proposed a theory of the effective use of information systems. Their work is specifically based on representation theory but recognises that many other factors could drive effective use (e.g., intention to use, organisation culture, etc). Similarly, the Technology Acceptance Model (TAM) adapted social-psychological/behavioural theory and introduced perceived ease of use and perceived usefulness as antecedents of information systems effectiveness. UTAUT (The Unified Theory of Acceptance and Use of Technology) was developed on the basis of social cognitive theory and diffusion of innovations theory. While these studies shed light on effective use of information systems in a conventional enterprise setting, they do not consider big data specifically and the unique challenges related to the big data 'Vs'. Thus, we argue that the body of knowledge on the effective use of information systems needs to be revisited in the context of big data.

Although, historically, each of these characteristics has been tackled to varying levels of success, it is their collective impact that presents a unique set of challenges, quite distinct from previous information systems research. For example, the database research community has a long history of contributing to query performance for large data (Abadi et al. 2006; Candea et al. 2011; Wang et al. 2015) and data stream processing (Abadi et al. 2003; Babcock et al. 2002; Jiang et al. 2014) that addresses volume and velocity issues. More recently, scale-out architectures have been developed to store vast amounts of data. To handle velocity of the data, cloud infrastructures that allow bursting for additional computation power have been developed for real-time data analysis. Similarly, there are several contributions towards handling the variety of the data, such as unstructured, graph, spatial, microblog and text data. However, organisations are still struggling to manage and harness value from big data (Segarra et al. 2016; Wamba et al. 2017; Zeng and Glaister 2017). Recent studies indicate that most big data initiatives fail to deliver on their expectations (Marr 2015) and that the failure is often not due to technology but due to problems associated with organising and managing the socio-technical complexity of big data projects.

Decision support systems in organisations are moving from model-driven to data-driven. In model-driven decision support systems, users and analysts interact primarily with a predefined (mathematical) model and its results. Model-driven approaches have been highly successful in helping to solve well-defined and structured problems (for example, what-if-analysis). Meanwhile, in data-driven approaches, users (data scientists) interact primarily with the data. This approach promises to solve unstructured, exploratory and so-called wicked problems, by identifying relations or patterns in data to gain insights and foresights for decision-making. Given the complexity of the methods and the size of the data on which they are applied, there are understandably some cultural issues within organisations that can become barriers in adopting a data-driven approach to decision making (Davenport and Patil 2012; McAfee et al. 2012). According to IBM (LaValle et al. 2013), 1 in 3 business leaders do not trust the information they use for decision making. Indeed, one of big data characteristics is veracity, which means that a lack of data quality and trustworthiness might influence the credibility of insight generated. Furthermore, there is a belief that a machine cannot replace the instinct and capability of humans (Chui et al. 2016). Decision makers may need to learn how to supplement human insights with actionable insights from big data analytics. Such resistance and barriers may result in widening of the business-technology divide - an enduring problem that has been the target of years of research in information systems, and hence needs to be studied in the specific context of big data initiatives.

Based on the above arguments, the effective use of big data in an organisation is a complex phenomenon. Understanding the technical, social and business enablers and inhibitors that are influencing effective use of big data is a necessary step towards developing a theory of the effective use of big data and, thus, providing guidance to organisations such that the value of big data can be better realised. Accordingly, in this study, we explore the case of a data-intensive project that meets the characteristics of big data. The study aims to identify, in an empirical manner, enablers and inhibitors that might influence effective use of big data. Based on a case study research design, we identify and describe 15 enablers and inhibitors that influence effective use of big data. We also use the study to distil a definition of effective use of big data to facilitate a more conscious usage of the term and more coherent development of research on this subject.

This paper is organized as follows. In the next section, we report on the current body of knowledge related to effective use of information systems. We then present the methodology that constitutes a description of the case study setting, data coding and analysis, and approach to identify enablers and inhibitors from a case study. We identify and describe 15 enablers and inhibitors for effective use of big data and discuss the outcomes of the study in the results and discussions section. We conclude by summarising the findings and outlining future research opportunities.

Related Work

A substantial number of prior studies have focused on developing models and theory relating to the use and effective use of information systems and technology. TAM is one such well-tested model of information technology use (Davis et al. 1989). TAM is based on two theories - TRA (Theory of Reasoned Action) and the expectancy-value model. It suggests that information system usage is determined by two major variables: perceived usefulness and perceived ease of use. The model was further enhanced by Venkatesh et al. (2003), with the development of the UTAUT (Unified Theory of Acceptance and Use of Technology). UTAUT posits that usage behaviour is influenced by several factors, including performance expectancy, effort expectancy, social influence, facilitating conditions, gender, age, experience and voluntariness of use. TAM and UTAUT are widely recognized as the underpinning theories of information technology use in IS research.

However, to achieve designed objectives or goals, systems must be used effectively, not just used. The Theory of Effective Use (TEU) (Burton-Jones and Grange 2012) is a natural progression from studying use, as in TAM and UTAUT, to the study of effective use. TEU is founded on representation theory (Burton-Jones and Grange 2012). It proposes two levels of effective use, *viz*. the nature and the drivers of effective use. Adaptation actions and learning actions are identified as the major antecedents (drivers) of effective use. Trieu (2013) extended TEU in the context of business intelligence (BI) systems, by studying the impact of enterprise architecture maturity stages on effective use. In the big data context, Merino et al. (2016) conducted a study to explore whether data quality levels were sufficient for intended use. Accordingly, they proposed a model to assess the level of Quality-in-Use of big data. This research, to the best of our knowledge, is the latest attempt to develop TEU further, and to extend the body of knowledge on information systems and technology use.

Effective use of big data has received attention from practitioners and researchers in many industries. For example, studies in the healthcare sector have explored effective use of big data in ambulatory care (Thorpe and Gray 2015) and Neonatal Intensive Care Unit (Vijayalakshmi et al. 2015). The use of big data for organisational efficiency and effectiveness has also been investigated (Hazen et al. 2014).

Tallon (2013) in his paper has addressed enablers and inhibitors of big data and analytics use on government aspects. Moreover, several journal articles forecast future challenges and trends related to the use of big data, emphasising the importance of exploiting its benefits (Assunção et al. 2015), (Kambatla et al. 2014), (Chen and Zhang 2014) and (Huser and Cimino 2015). While these articles outline benefits of effective use of big data, they do not explore the enablers and inhibitors of making effective use of big data that including all related aspects.

The lack of studies and lack of core knowledge on enablers and inhibitors that impact effective use of big data is an impediment to the success of projects involving new and large datasets. Such enablers and inhibitors may stem from technology, human sources, and organisational capabilities and from the data itself. Thus, developing a comprehensive model of effective use of big data requires substantial theoretical development, which, in turn, requires the identification of enablers and inhibitors as a prerequisite. We aim to contribute towards this goal by identifying the enablers and inhibitors that have the potential to impact effective use of big data and formulate the definition of effective use of big data.

Research Method

This study uses an exploratory case study as a research strategy. The aim of using an exploratory study is to extract detailed and reliable perspectives of the issues being investigated. In this case, data are gathered primarily using semi-structured interviews and through document analysis. The semi-structured interviewing technique was chosen as the main data collection technique to enable the researcher to ask for more detail, probe an issue, and go back and forth among important points and request further explanation. Accordingly, an open-ended semi-structured interview protocol was prepared.

As the objective of the study was to iteratively identify, develop and saturate emerging concepts, the selection of participants was purposive (Alemu et al. 2017). Since the focus was on concepts and categories rather than descriptions (i.e. conceptualisation rather than representative sampling), the profile of participants (such as personal details, educational background, gender, age, etc.) was not collected and, hence, could not be analysed.

Case Study Setting

While there is a tendency for individuals to equate big data simply with a large volume of data, it is in fact variety, not volume or velocity, which drives big data initiatives and investments (Bean 2016). A 2016 survey of over fifty executives in large organizations (Partners 2016) suggests that variety trumps volume and velocity when it comes to big data success. Organizations continue to report that variety is the primary technical driver behind big data investments (40%), with volume (14.5%) and velocity (3.6%) lagging well behind. Organizations are seeking to integrate more sources of data, including new sources as well as legacy sources. Accordingly, we chose a project that collected data for all research data throughout an Australian university - the aim of the project was to present research-related information to fulfil government requirements. The project focused on a wide range of data sources, including structured and unstructured data (such as text files, email, website and scientific data). The selected case involved the management and analysis of over 15 million tuples and 8000 entities. The veracity of data in the selected case was also of concern because of major challenges in relation to data accuracy and trustworthiness.

Data Collection

Primary data collection was undertaken using interview protocols, supported by reviews of related documents and systems, where available. During the interview, participants were asked 4 categories of questions, as follows:

- Setting: The characteristics, purposes, benefits and examples of successful use of big data.
- Perception: The perceived and actual value of big data.
- Experiences: Challenges and support mechanisms of effective use of big data.

- Opinions: Meaning of big data and effective use of big data.

We further created two versions of the above questions, varied slightly for business vs technical users, and administered them depending on the type of participant involved: business user and technical staff. Five participants from the single organisation were interviewed. These participants include 2 technical staff and 3 business users. Interviews were conducted between July 2017 and August 2017. All interviews were digitally recorded. The interviews averaged 60 minutes in duration, with a total time of 334 minutes. All interviews were transcribed to facilitate analysis, as were the related interview notes made by three researchers involved in the interviews. Table 1 provides an overview.

Table 1. Summary of Interview Data

Participant	**Role**	**Gender**	**Date and duration of interview**
Participant A	Project Director	Female	26-07-2017, 1h07min
Participant B	Manager of Data Services	Male	31-07-2017, 1h11min
Participant C	Project Officer	Female	01-08-2017, 1h13min
Participant D	Director, Planning and Business Intelligence	Female	07-08-2017, 1h19min
Participant E	Director, Research Analysis and Operations	Female	07-08-2017, 44min

Data Coding and Analysis

The coding started with a process referred to as microanalysis (Strauss and Corbin 1990). That is, a line-by-line analysis of semi-structured interview to identify initial codes. This research follows the three steps of Saldaña (2015), to generate and label codes: First, review the characteristics of the research question(s). Second, identify coding methods (such as in vivo, selective coding) consistent with the research question(s). Third, assign codes (using coding methods) that best represent the relevant information in the data. We used NVivo 11 to organise and analyse the data because it is useful for moving data easily from one code to another and to document the data as it is analysed (Miles and Huberman 1994).

The coding process was conducted using a dual coder approach. The first researcher coded the 213 pages of transcripts using open coding. Table 2 provides some exemplary coding. The second researcher then verified all the codes by checking the codes and the transcripts to ensure the context and the code names are related. Three outcomes were possible in this 2nd coder verification approach: text doesn't match the code, text matches a different code, or text doesn't match any code. A discussion to reach consensus took place to finalize the code's name. The coding process produced 15 enablers and inhibitors.

Table 2. Exemplary Coding

Quote	Code
"…income data is particularly missing, [there are] all sorts of manual processes involved in attaching income record to projects, it's not straightforward at all..."	Poor Data Quality
"…not understanding what the data was collected for originally or why it was assigned for classification […] people - they have not been involved in that process, they can just make sort of the assumptions, the assumptions are the enemy of meaningful insight from data…"	Lack of Data Understanding

Identification of Enablers and Inhibitors

Enablers of effective use of big data are factors that contribute to strengthening the attainment of the big data initiative goals. Inhibitors of effective use of big data, on the flip side, are defined as factors which contribute to weakening the attainment of the big data initiative goals. The enablers and inhibitors are identified by examining their positive or negative effects on the effective use of big data. While they can be two sides of the same coin, in this paper we report them in the positive or negative sense depending based on their perception by the participants in the case study at hand.

We identified the most significant enablers and inhibitors using a weighted sum of three criteria. We use the weighted sum method as this method is commonly used in multi-criteria decision analysis and ranking techniques (Triantaphyllou 2000). The first criterion is the frequency of occurrence of the codes in the transcripts. The greater the number of occurrences of the codes, the stronger, or more prominent the enabler or inhibitor. The second criterion is the number of interviewees that mention the underlying enabler or inhibitor. The greater the number of interviewees mentioning the enabler or inhibitor, the more prominent it is. The third criterion relates to whether the enabler or inhibitor is explicitly mentioned by interviewees as being significant, critical or important (it receives a score of 1 if so, and 0 otherwise). We assign a weight of 1, 3, and 5, respectively, to each of the three criteria and sum the result to obtain the score for ranking, following the formula: score = 1 * (number of occurrence) + 3 * (number of interviewees that mention the factor) + 5 * (1 if important, 0 if is not). For example, if an enabler occurs 3 times, and is mentioned by 2 interviewees and is indicated as important, then: score = 1 * 3 + 3 * 2 + 5 * 1 = 14.

Defining Effective use of Big Data

To enhance understanding of effective use of big data, we also aimed to synthesise from the semi structured interview data a definition for effective use of big data. To ensure consistency of views, we first discussed what big data is by asking participants "*What does big data mean to you?*" Following this discussion, we explored through asking "*What does effective use of big data mean to you?*" what the participants considered to constitute effective use of big data.

Results and Discussion

Enablers and Inhibitors

In the following, we describe the identified enablers and inhibitors and present them from the most significant to the least significant based on the earlier described selection/scoring criteria in research method section, namely: number of codes, number of interviewees and the perceived importance. We identify four enablers (E): (Adequate System Capabilities, Established Culture of Collaboration, Good Working Attitude, and Champions) and 11 inhibitors (I): (Poor Data Quality, Lack of Data Understanding, Data Silos, Lack of Time, Lack of Cost-Benefit Analysis, Lack of Top Management Support, Lack of Technical Skills, Lack of Data Management Process, Lack of Right People, Lack of Clear Goal, Poor Data Privacy and Security), as shown in Table 3 together with their relative ranking. In the following we offer summaries of each enabler and inhibitor, providing examples from our interview data.

Table 3. Identified Significant Enablers and Inhibitors

#	Name	# of codes (Weight=1)	# of interviewees (Weight=3)	Importance (Weight=5)	Score	Ranking
1	Poor Data Quality (I)	41	5	Yes	61	1
2	Lack of Data Understanding (I)	31	4	Yes	48	2
3	Adequate System Capabilities (E)	24	4	Yes	41	3
4	Data Silos (I)	22	3	Yes	36	4

#	Name	# of codes (Weight=1)	# of interviewees (Weight=3)	Importance (Weight=5)	Score	Ranking
5	Lack of Time (I)	18	4	Yes	35	5
6	Established Culture of Collaboration (E)	17	4	Yes	34	6
7	Lack of Cost-Benefit Analysis (I)	15	4	Yes	32	7
8	Good Working Attitude (E)	17	3	Yes	31	8
9	Lack of Top Management Support (I)	21	3	No	30	9
10	Lack of Technical Skills (I)	10	4	Yes	27	10
11	Lack of Data Management Processes (I)	20	2	No	26	11
12	Lack of Right People (I)	9	2	Yes	20	12
13	Lack of Clear Goal (I)	12	2	No	18	13
14	Champions (E)	9	2	No	15	14
15	Poor Data Privacy and Security (I)	4	1	No	7	15

Poor Data Quality (I)

The term data quality is defined as fitness for use (Juran 1992). Data quality is a cross-disciplinary and often domain specific problem or one that is dependent on the context due to the importance of fitness for use (Sadiq et al. 2011). Data quality of big data should include a focus on usability (Bertino and Ferrari 2018; Cai and Zhu 2015; Jianzhong and Xianmin 2013). Data quality issues impede the effective use of big data and have become a significant inhibitor, as indicated by interviewees in our study, e.g. "*It's combination of things, hmm… I would say top of these is data quality…*" Data quality problems may have many forms, including: data that is not formatted properly and does not include unique identifies, data duplicates, missing data, misclassification, and poor control of data quality at point of entry, resulting in 'garbage in garbage out'.

Lack of Data Understanding (I)

Lack of data understanding relates to users not being aware of the purpose for which data was collected originally, which may result in them making incorrect assumptions about the data, leading to poor quality insights. For example, one interviewee indicated: *"…not understanding what the data was collected for originally or why it was assigned for classification […] people - they have not been involved in that process, they can just make sort of the assumptions, the assumptions are the enemy of meaningful insight from data…"* Researchers and practitioners suggest providing a 'user-oriented' view of metadata during data collection that can effectively improve data understanding (Di Giovanni et al. 2009). To improve data understanding and engagement of the users, Dimara et al. (2017) suggests using narratives in the context of data analysis and communication.

Adequate System Capabilities (E)

To ensure effective use of big data, system capabilities must be augmented to meet the characteristics of big data that exceed a normal system's capabilities. System capabilities are gradually challenged due to the scalability and performance considerations purely owing to changing nature of data (Mohanty et al. 2013). The studied organisation also created a system to satisfy the dynamic user needs. "*So, we have built some great systems off the back of that…*" The interviewee emphasized that system capabilities are most important/critical support for effective use of big data.

Data Silos (I)

Data silos are a collection of datasets with no specific links to any other data (Jennings and Finkelstein 2010). The data in the case studied was fragmented, spread across systems that were not integrated and belonged to various departments: "*We do our best, but, um, there's certainly a lot of challenges around, um, bringing customer information to the right area. You know, making sure it's aligned with the right area of the organisation…*" Data silos can be tackled by implementing good data integration practices. Data integration is a process used to combine data from disparate sources into meaningful and valuable insights. It allows different data types (such as datasets, documents and tables) to be merged by users, organisations and applications, for use as personal or business process and/or functions.

Lack of Time (I)

This inhibitor is specific to big data projects with a tight project schedule, as was the case in the project we studied. The duration of this project was relatively short, with a critical deadline that had to be met due to regulatory requirements. Four of our 5 interviewees emphasized this inhibitor influencing effective use of big data. They indicated that there was not enough time to collect and analyse all the data and also to build the required system to facilitate their analysis: "*That was very rushed, and we had a very short period of time…*"

Established Culture of Collaboration (E)

A good culture of collaboration is reflected in a group of people with different functional expertise coming together to complement their skills and work toward a common goal – e.g. business professionals working effectively together with data analysts and IT professionals to gain insights from data (Kiron and Shockley 2011). Four interviewees highlighted the strong culture of collaboration in the organisation as the most predominant enabler in the effective use of big data: "*The culture of collaboration and common purpose that I think our organisation has been quite good at…*" One interviewee indicated that the culture of collaboration was been deeply embedded in the project team: "*There is a rare exception, where someone didn't act very collaboratively and so, that sort of behaviour been very rare…*"

Lack of Cost-Benefit Analysis (I)

Cost-Benefit Analysis is the process of evaluating a planned action by determining what net value it will have for the organisation. A cost-benefit analysis needs to performed to measure the potential long-term benefits of adopting big data to support data-driven decision making and communicate findings to non-technical stakeholders (Skourletopoulos et al. 2017). The organisation in the studied case invested in expensive data sources but then found that not all purchased data was needed/used. "*It isn't giving you an answer that's useful in any business sense. But it's also about the cost of it all, you know, it's the cost-benefit stuff…*"

Good Working Attitude (E)

Working attitude is a tendency to respond positively or negatively towards a certain situation (Cacioppo et al. 1997). From this study, we identified several indictors of working attitudes that can enable big data projects: committed, focused, hard-working, good morale, efficient, ethical, and trustworthy: "*... everybody was committed to do as best as we could within the period of time, so everybody worked incredibly hard, and there was really good morale with the all of the team to get all of the data in, by certain time…*"

Lack of Top Management Support (I)

Top management support is the willingness of top management to provide necessary resources and authority for project success (Jang and Lee 1998). The studied organisation faced a problem with lack of top management support, in terms of providing more investment in infrastructure: "*…they have to spend money on the right infrastructure because big data can actually be very useful…*" Top management needs to address correctly defining and appropriating the necessary resources to ensure big data project *success*.

Lack of Technical Skills (I)

Technical skills are the knowledge and abilities needed to accomplish analytical, scientific or technology-related duties, as well as other tasks relating to big data tools or processes. The problem faced by the studied organisation was not the lack of tools but rather problems with utilising them corectly, especially in the context of analytics. "*Those sorts of tools that are out there, […] our ability to use them is limited…*" Organizations in order to leverage data through big data analytics need human capital with high level of technical skills.

Lack of Data Management Process (I)

According to Kiron et al. (2012), adequate data management processes need to be established to create an environment in which effective use big data is enabled. In the studied organization, the interviewees identified lack of data management process, from data capture, storage, transformation through to provision of results. For example: "…*in terms of the data capture, there are some limitations*…"

Lack of Right People (I)

The availability of the right people is important to the big data project success (Gao et al. 2015). However, involving the right people is often difficult. The studied organization faced this problem: "*That's another thing that was quite complex in doing in this project. How to ensure we had the right staff…*" Another interviewee pointed out: "*So we will advertise for people, and we will rarely get more than one appointable…"*

Lack of Clear Goal (I)

McAfee et al. (2012) argues that a clear goal is a prerequisite for an organization to be a top performer in their use of big data. The studied organization suffered from a lack of clear goals in terms of their big data project and lacked a detailed and well planned step-by-step project map. "*The goal was never defined upfront. We didn't realise I think from the very beginning […] we didn't know that we were going to have to build a system…"*

Champions (E)

A champion is often a manager who actively promotes IS solutions (Guinan et al. 1998). In the case of the studied project and organization, there was a manager who developed a reliable and easy to use system to support data management. "*So as far as I'm concerned, the number one support mechanism that I think the organization can't do without is [name of person] …*" Another interviewee mentioned the same name again "*…but there's a key individual in this organization, [name of person] …"*

Poor Data Privacy and Security (I)

Based on (Halaweh and El Massry 2017) empirical study, data privacy and security is one of the factors that affects the success or failure of big data implementation in organization. Our study reinforces this finding. The studied organization had a problem with privacy issue "*I mean, there's obviously privacy issues…"* and storing data in secure places. "*I think the organization has to make sure that secure storage for the working data and secure storage for the data once they finished…*"

Defining Effective Use of Big Data

Having outlined the identified inhibitors and enablers, we now move on to explore how our participants perceived big data and its effective use.

To recap, effective use in the information systems discipline is defined as the degree to which the use of an information system increases achievement of the goals an organisation has set concerning the system (Burton-Jones and Grange (2012). Therefore, effective use is said to be attained if an organisation realises the intended outcomes of system use, such as increasing customer satisfaction or decreasing delivery time. This definition considers the multilevel nature of system usage (individual, group, and organisational level) and assumes that "the relevant goal for effective use is simply whatever end-point the system is intended to achieve". Furthermore, 'goal achievement' has objective qualities that can be quantitatively measured to assess performance. For example, if a company deploys a system to reduce the time of delivery product, the system is used effectively if the time for delivering the product decreases.

While there is agreement on this definition in the information systems 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 (2012) 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 progress of a discipline and knowledge. Big data has instead evolved quickly, and arguably in a disorderly manner, so that such a universally accepted formal statement denoting what is the meaning of the effective use of big data does not exist.

Through our case study, we explored what effective use of big data meant to the project participants. Their summarised responses as shown in Table 4.

Table 4. Definition of Effective Use of Big Data

Participant	**Definition**
Participant A	Use of big data for *evidence-based decisions*, or to *change behaviour*.
Participant B	Not *gaining insight* from flawed understanding of the data.
Participant C	Not misrepresenting data. The ability to *link stories that we haven't been able to link in the past* and be able to apply them.
Participant D	Giving you *answers that are useful to business* and with appropriate use of resources (cost-benefit analysis).
Participant E	Drawing together a conclusion based on data that is of use or meets our requirements. The data might give you a very interesting story. It is utilizing that information to help *drive the goals of an organisation* – i.e. what we do from here. It is to ensure that if we're setting a strategic agenda, that the data supports it or in the future shows *advancement towards whatever that strategy is*. It is more about using it to set goals that are achievable, but also possibly aspirational based on where you are now.

By reviewing these responses, we argue that the core of the concept of effective use of big data can be expressed by the following notions:

- Evidence-based decision making
- Generation of new insights
- Setting of organisational goals
- Behavioural change

We observe that the 'object' to which effective use of big data should refer in its definition is 'organization goals', but not necessarily towards only achieving them, as in effective use of information systems, but beyond that towards setting new strategies and aspirations and changing behaviour. That is more than meeting goals. Effective use of big data is also the ability to set new goals in an informed manner. We argue that this observation is instrumental in defining a distinctive and authentic definition for effective use of big data.

Concluding Discussion

Effective use of big data in organisations is a complex phenomenon and understanding the technical, organizational and business enablers and inhibitors that contribute to effective use of big data is key to providing organisations with guidance for increased success in big data projects. This study is the first attempt to explore enablers and inhibitors that may influence effective use of big data. In this paper, we report on a case study that investigates the experience of an organisation's use of big data. Through the study, we identified four enablers and 11 inhibitors. We propose these items need to be studied further to provide guidance for companies for their big data projects. To summarise the enablers and inhibitors, and their impact on effective use of big data, we develop a preliminary model that provides a starting point for future research to achieve a more nuanced understanding of effective use of big data. Grounded in our analysis of the data, the conceptual model in Figure 1 is based on the following propositions:

> Proposition 1: Effective use of big data is influenced by right data that can be measured by data understanding, data quality, data management process, and data privacy and security.
>
> Proposition 2: Effective use of big data is influenced by people and organizational support, which can be measured by top management support, culture of collaboration, technical skills, good working attitude, and right people.
>
> Proposition 3: Effective use of big data is influenced by system support, which can be measured by system capabilities and system integration.
>
> Proposition 4: Effective use of big data is influenced by perceived benefit, which can be measured by cost-benefit analysis and clear goals.

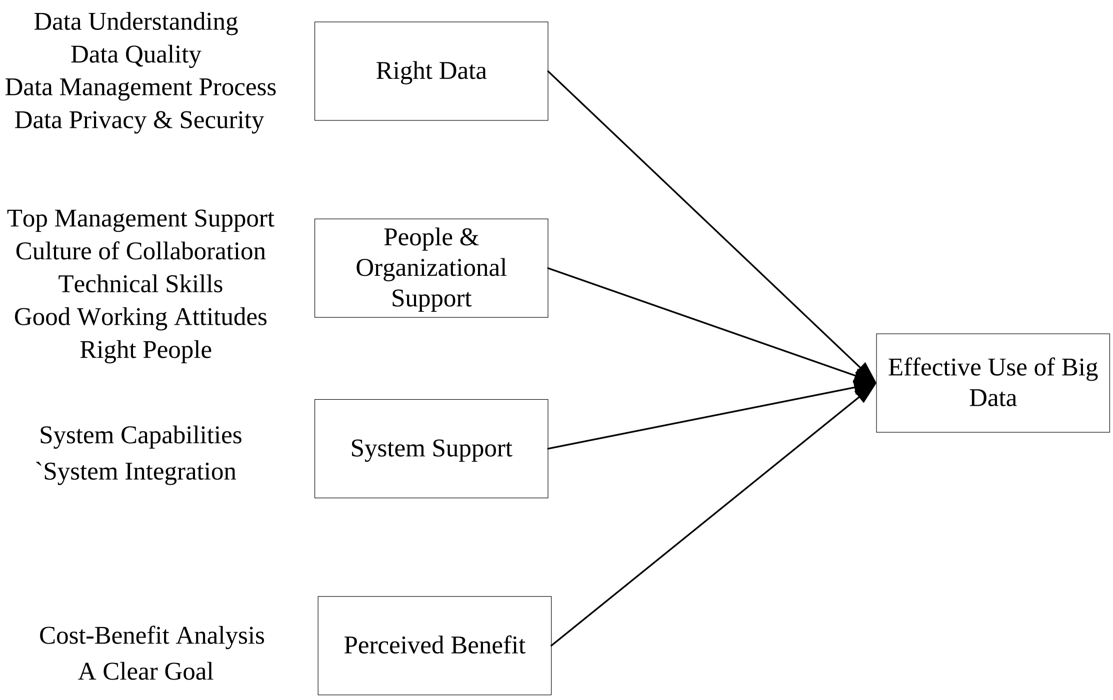

Figure 1. Model of the Enablers of Effective Use of Big Data

Although this study offers an initial step in identifying enablers and inhibitors that impact effective use of big data, there are some limitations in this research. To study effective use of big data, and to ultimately develop a theory of it, sophisticated methodological approaches are needed to build strong evidence of enablers and inhibitors and relationships, together with associated hypotheses or propositions. Substantive theory-building efforts are required to ensure an adequate consideration of the complex problem represented by big data.

Our findings in this paper are based on a single case study. Hence external validity associated with the extent of generalisability of the study requires further consideration. We envision an exhaustive and rigorous future study which will define effective use of big data on the basis of a larger set of empirical evidence, further investigating enablers and inhibitors that influence effective use of big data and examining the relationships between them. Going forward, studying the various enablers and inhibitors through multiple case studies in different organizations, is needed as the basis to refine, evaluate and develop a theory of effective use of big data.

Acknowledgements

The authors gratefully acknowledge: the LPDP (Indonesia Endowment Fund for Education) for supporting this research with a grant to the first author; and the support of the Australian Research Council Discovery Project DP170101172 on Managing Data with High Redundancy and Low Value Density.

REFERENCES

Abadi, D., Madden, S., and Ferreira, M. 2006. "Integrating Compression and Execution in Column-Oriented Database Systems," in: *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. ACM, pp. 671-682.

Abadi, D. J., Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Stonebraker, M., Tatbul, N., and Zdonik, S. 2003. "Aurora: A New Model and Architecture for Data Stream Management," *The VLDB Journal—The International Journal on Very Large Data Bases* (12:2), pp. 120-139.

Agarwal, R., Angst, C. M., DesRoches, C. M., and Fischer, M. A. 2010. "Technological Viewpoints (Frames) About Electronic Prescribing in Physician Practices," *Journal of the American Medical Informatics Association* (17:4), pp. 425-431.

Alemu, G., Stevens, B., Ross, P., and Chandler, J. 2017. "The Use of a Constructivist Grounded Theory Method to Explore the Role of Socially-Constructed Metadata (Web 2.0) Approaches," *Qualitative and Quantitative Methods in Libraries* (4:3), pp. 517-540.

Anuradha, J. 2015. "A Brief Introduction on Big Data 5vs Characteristics and Hadoop Technology," *Procedia computer science* (48), pp. 319-324.

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., and Buyya, R. 2015. "Big Data Computing and Clouds: Trends and Future Directions," *Journal of Parallel and Distributed Computing* (79), pp. 3-15.

Babcock, B., Babu, S., Datar, M., Motwani, R., and Widom, J. 2002. "Models and Issues in Data Stream Systems," in: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, pp. 1-16.

Bean, R. 2016. "Variety, Not Volume, Is Driving Big Data Initiatives," in: *MIT Sloan Rev.*

Bertino, E., and Ferrari, E. 2018. "Big Data Security and Privacy," in *A Comprehensive Guide through the Italian Database Research over the Last 25 Years*. Springer, pp. 425-439.

Boudreau, M.-C., and Seligman, L. 2006. "Quality of Use of a Complex Technology: A Learning-Based Model," *Contemporary Issues in End User Computing* (17), pp. 1-22.

Burton-Jones, A., and Grange, C. 2012. "From Use to Effective Use: A Representation Theory Perspective," *Information Systems Research* (24:3), pp. 632-658.

Cacioppo, J. T., Gardner, W. L., and Berntson, G. G. 1997. "Beyond Bipolar Conceptualizations and Measures: The Case of Attitudes and Evaluative Space," *Personality and Social Psychology Review* (1:1), pp. 3-25.

Cai, L., and Zhu, Y. 2015. "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," *Data Science Journal* (14:2), pp. 1-10.

Candea, G., Polyzotis, N., and Vingralek, R. 2011. "Predictable Performance and High Query Concurrency for Data Analytics," *The VLDB Journal—The International Journal on Very Large Data Bases* (20:2), pp. 227-248.

Chen, C. P., and Zhang, C.-Y. 2014. "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data," *Information Sciences* (275), pp. 314-347.

Chui, J. Manyika, and M. Miremadi. 2016. "Where Machines Could Replace Humans—and Where They Can't (yet)." McKinsey Quarterly.

Davenport, T. H., and Patil, D. 2012. "Data Scientist," *Harvard business review* (90:5), pp. 70-76.

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management science* (35:8), pp. 982-1003.

Di Giovanni, F., Piazza, D., and d'Italia, B. 2009. "Processing and Managing Statistical Data: A National Central Bank Experience," *Bank of Italy*).

Dimara, E., Bezerianos, A., and Dragicevic, P. 2017. "Narratives in Crowdsourced Evaluation of Visualizations: A Double-Edged Sword?," *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*: ACM, pp. 5475-5484.

Gao, J., Koronios, A., and Selle, S. 2015. "Towards a Process View on Critical Success Factors in Big Data Analytics Projects," *AMCIS*, pp. 1-14.

Guinan, P. J., Cooprider, J. G., and Faraj, S. 1998. "Enabling Software Development Team Performance During Requirements Definition: A Behavioral Versus Technical Approach," *Information systems research* (9:2), pp. 101-125.

Halaweh, M., and El Massry, A. 2017. "A Synergetic Model for Implementing Big Data in Organizations," *Operations and Service Management: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*), pp. 281-298.

Hazen, B. T., Boone, C. A., Ezell, J. D., and Jones-Farmer, L. A. 2014. "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," *International Journal of Production Economics* (154), pp. 72-80.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., and Sethupathy, G. 2016. "The Age of Analytics: Competing in a Data-Driven World," *McKinsey Global Institute report*).

Huser, V., and Cimino, J. J. 2015. "Impending Challenges for the Use of Big Data," *International Journal of Radiation Oncology* Biology* Physics*).

Jang, Y., and Lee, J. 1998. "Factors Influencing the Success of Management Consulting Projects," *International Journal of Project Management* (16:2), pp. 67-72.

Jennings, B., and Finkelstein, A. 2010. "Implicit Social Production: Utilising Socially Generated Data by-Products," *International Conference on Business Process Management*: Springer, pp. 359-371.

Jiang, D., Chen, G., Ooi, B. C., Tan, K.-L., and Wu, S. 2014. "Epic: An Extensible and Scalable System for Processing Big Data," *Proceedings of the VLDB Endowment* (7:7), pp. 541-552.

Jianzhong, L., and Xianmin, L. 2013. "An Important Aspect of Big Data: Data Usability [J]," *Journal of Computer Research and Development* (6), p. 006.

Juran, J. M. 1992. *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services*. Simon and Schuster.

Kambatla, K., Kollias, G., Kumar, V., and Grama, A. 2014. "Trends in Big Data Analytics," *Journal of Parallel and Distributed Computing* (74:7), pp. 2561-2573.

Kiron, D., Prentice, P. K., and Ferguson, R. B. 2012. "Innovating with Analytics," *MIT Sloan Management Review* (54:1), pp. 47-62.

Kiron, D., and Shockley, R. 2011. "Creating Business Value with Analytics," *MIT Sloan Management Review* (53:1), pp. 57-70.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., and Kruschwitz, N. 2013. "Big Data, Analytics and the Path from Insights to Value," *MIT sloan management review* (21:2), pp. 20-31.

LeRouge, C., Hevner, A. R., and Collins, R. W. 2007. "It's More Than Just Use: An Exploration of Telemedicine Use Quality," *Decision Support Systems* (43:4), pp. 1287-1304.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute.

Marr, B. 2015. *Big Data: Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance*. Hoboken, New Jersey: John Wiley & Sons.

McAfee, A., Brynjolfsson, E., and Davenport, T. H. 2012. "Big Data: The Management Revolution," *Harvard business review* (90:10), pp. 60-68.

Merino, J., Caballero, I., Rivas, B., Serrano, M., and Piattini, M. 2016. "A Data Quality in Use Model for Big Data," *Future Generation Computer Systems* (63), pp. 123-130.

Miles, M. B., and Huberman, A. M. 1994. *Qualitative Data Analysis* Thousand Oak, CA: Sage Publications.

Mohanty, S., Jagadeesh, M., and Srivatsa, H. 2013. "Application Architectures for Big Data and Analytics," in *Big Data Imperatives*. Springer, pp. 107-154.

Partners, N. 2016. "Big Data Executive Survey 2016: An Update on the Adoption of Big Data in the Fortune 1000," NewVantage Partners, Boston.

Pavlou, P. A., Dimoka, A., and Housel, T. J. 2008. "Effective Use of Collaborative It Tools: Nature, Antecedents, and Consequences," *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*: IEEE, pp. 40-40.

Ronda‐Pupo, G. A., and Guerras‐Martin, L. Á. 2012. "Dynamics of the Evolution of the Strategy Concept 1962–2008: A Co‐Word Analysis," *Strategic Management Journal* (33:2), pp. 162-188.

Sadiq, S., Yeganeh, N. K., and Indulska, M. 2011. "20 Years of Data Quality Research: Themes, Trends and Synergies," *Proceedings of the Twenty-Second Australasian Database Conference-Volume 115*: Australian Computer Society, Inc., pp. 153-162.

Saldaña, J. 2015. *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage Publications.

Segarra, L. L., Almalki, H., Elabd, J., Gonzalez, J., Marczewski, M., Alrasheed, M., and Rabelo, L. 2016. "A Framework for Boosting Revenue Incorporating Big Data," *Journal of Innovation Management* (4:1), pp. 39-68.

Skourletopoulos, G., Mavromoustakis, C. X., Mastorakis, G., Batalla, J. M., Dobre, C., Panagiotakis, S., and Pallis, E. 2017. "Big Data and Cloud Computing: A Survey of the State-of-the-Art and Research Challenges," in *Advances in Mobile Cloud Computing and Big Data in the 5g Era*. Springer, pp. 23-41.

Strauss, A., and Corbin, J. 1990. *Basics of Qualitative Research Techniques: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage publications.

Tallon, P. P. 2013. "Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost," *Computer* (46:6), pp. 32-38.

Thorpe, J. H., and Gray, E. A. 2015. "Big Data and Ambulatory Care: Breaking Down Legal Barriers to Support Effective Use," *The Journal of ambulatory care management* (38:1), pp. 29-38.

Triantaphyllou, E. 2000. "Multi-Criteria Decision Making Methods," in *Multi-Criteria Decision Making Methods: A Comparative Study*. Springer, pp. 5-21.

Trieu, T. V. H. 2013. "Extending the Theory of Effective Use: The Impact of Enterprise Architecture Maturity Stages on the Effective Use of Business Intelligence Systems," *ICIS* Milan, pp. 1649-1659.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS quarterly*), pp. 425-478.

Vijayalakshmi, T., Kumar, V., Gokulraj, J., and Malathy, A. 2015. "Effective Use of Bigdata and Social Media in–Neonatal Intensive Care Unit," *International Journal of Engineering Research and Technology* (4:02 (February-2015)), pp. 442-444.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., and Childe, S. J. 2017. "Big Data Analytics and Firm Performance: Effects of Dynamic Capabilities," *Journal of Business Research* (70), pp. 356-365.

Wang, H., Zheng, K., Zhou, X., and Sadiq, S. 2015. "Sharkdb: An in-Memory Storage System for Massive Trajectory Data," *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*: ACM, pp. 1099-1104.

Zeng, J., and Glaister, K. W. 2017. "Value Creation from Big Data: Looking inside the Black Box," *Strategic Organization* (12), pp. 1-36.