Modelling Dark Data Management Framework: a Grounded Theory

Ahmad Fuzi Md Ajis[1], Isma Ishak[2], Qamarul Nazrin Harun[3]

[1,2,3] *Faculty of Information Management, Universiti Teknologi MARA Cawangan Johor, Malaysia.*
[1]*ahmadfuzi@uitm.edu.my,* [2]*ismai242@uitm.edu.my,* [3]*qamarulnazrin@uitm.edu.my*

Abstract

The phenomenon of Big Data has no longer become a new issue for business players as the evolution of the big data characteristics from the classical 3Vs (volume, velocity, and variety) expanded to become the 42Vs. Interconnected electronic devices and IoT tremendously contribute to the increasing incidents of Big Data, which are not only experienced personally but also by businesses. While the awareness of dark data is still lacking and the evidence of publication merely exhibits an approach to the dark data phenomenon, this creates a huge gap in the subject matter clearly. Therefore, the study was proposed to investigate dark data management by Malaysian SMEs for business operations. The core of the study does not compensate the SMEs for the better way of managing dark data, but rather focuses on investigating its current circumstances. The study employed a qualitative approach to investigate the major purpose of the research and propose to analyze the data using Grounded Theory Methodology (GTM). As a result, the identification of the valuable dark data residing in the enterprise repository can be defined appropriately, providing insight on the construction of the management of dark data framework.

Keywords— Dark data, dark data management framework, grounded theory, Malaysia.

Introduction

Some of the impacts of dark data are that it makes distributed and duplicated content widespread without explicit oversight and weakens security. In addition, the sheer volume of dark data impacts the costs for searching and producing appropriate information and imposes a wasted storage cost in operating budgets. Meanwhile, nonevidence-based decisions have always been made by the administration of an organization, which in the end would jeopardize the whole organization's operation because of ignoring the policies and circulars, which are referred to as "dark data" in this case [5], [10].

The search for dark data publications was extended to look for any research regarding dark data management in a Malaysian context. Surprisingly, only one paper appears to be published from Malaysia. Accordingly, it was assumed that Malaysian awareness pertaining to dark data management is very minimal, yet they are experiencing the circumstances of having dark data in their storage, even though many researchers emphasize that dark data significantly influences better business intelligence, increased competitive advantages, and minimizes enterprise risks [4], [10], [12], [13], [16]. Therefore, this study was proposed to investigate on how dark data being managed by the SMEs for business operation to construct the dark data management framework.

Literature Review

Analyzing published literature assisted scholars in developing an understanding of the epistemology of dark data. Scholars, Scopus, and Web of Science were used to locate existing literature in the topic under study. The literature searches encountered a significant obstacle, since publications published during the recent five years indicated a scarcity of articles. As a result, the search ranges for the literature were widened to include any

publication date that had the word "dark data" in the article.

The search results seem to be more compelling, since a total of 7,020 hits were discovered. Unfortunately, only 56 journal papers and 17 non-academic pieces were included in the study after removing comparable material that was indexed in all databases, omitting non-English literature, and excluding full-text articles. A analysis of those publications showed three significant findings about the present status of dark data: (1) the vague definition of dark data; (2) the causes of dark data; and (3) a restricted strategy to dark data management.

Due to a dearth of study on the dark data issue, the volatility of dark data definitions focuses on searchability, which encompasses metadata and classification [8], [21]; and usability [11]. The researchers observed that the factors affecting the incidence of dark data are restricted to the process of guaranteeing data quality, which includes data accessibility, correctness, and traceability [14].

While dark data is defined as vital data that exists outside of what is typically recorded and analysed [14], the strategy to managing dark data has been proven to be quite restricted. Only a few academics and industry participants have argued that dark data may be adequately managed by developing a data lake infrastructure utilising schematic approach [26]; data management consists of four phases: identification, categorization, controls, and continual monitoring [4].

Dark data has been defined from a variety of views, as Corallo and colleague [6] analysed 22 papers relevant to the definition of dark data from a variety of areas and perspectives. They undertook a thorough study of the literature in order to support the formulation of a dark data definition specific to the manufacturing sector. The definition of dark data appears to fall under the properties of searchability [15], accessibility [7], unknown existence [17], uncategorized and ignored data [14], which are influenced by their

formats and result in unused data being hoarded over time [26].

While few publications address the features and characteristics of dark data, the definition gap based on sound theory remains unresolved and unexplained by real-world experiences. There were several scholarly articles and research organizations that tackled dark data. However, from a non-academic perspective, only large organizations have published white papers and research articles on the existence of dark data, its management, and the potential uses [6].

Dark data occurrences in the field of big data sparked an eager investigation into the data's effect and current state on businesses. Numerous research businesses from across the globe did dark data research in attempt to decipher the mystery surrounding dark data and the advantages it presented to the global big data community. As a consequence, doing dark data research through the lens of Malaysia's small and medium-sized firms will fill the hole.

In a nutshell, the literature study reveals that hidden values lurk in large unstructured data, which presents the opportunity for hidden, abandoned, and underused prospective data, colloquially referred to as "dark data," to devour storage space in organizational repositories. According to scholars, its presence would be damaging to the development and survival of businesses.

On the other hand, a convincing definition of dark data has not been established, and a number of dark data definitions have been discovered, each expressing a unique interpretation of the dark data analogy. Due to the inability to detect the presence of dark data, the provisional definition generates considerable misunderstanding among data owners over how to plan for dealing with such data. Furthermore, given the scarcity of publications demonstrating the advantages of dark data in reality, the risk of dark data's presence seems to lie in emasculating the

untapped potential of dark data rather than disclosing the leverage it provides.

Although just a few research firms discovered the existence of dark data and more than half of it was recognized as such from data holdings, the identification of dark data was still based on the principles of underused stored data assets. Indeed, dark data notion encompasses more than data hoarding. As a result, although reaping the benefits of dark data remains a trade secret of large corporations, this phenomenon has produced a massive gap in dark data use, particularly for small businesses and public sectors. As a result, more inquiry is necessary to close the research gap and expose the shadowy data management methods.

Methodology

A qualitative research technique was used to begin the investigation since there is minimal evidence supporting the hypothesis of dark data, which led to the study adopting a constructivist or interpretivist philosophical stance. SMEs were chosen as a sample in this study because publication of dark data research in journals or white papers is typically dominated by large firms or research entities such as Fortune 500 companies in the United States or proprietary research firms such as Veritas and Ipsos, but none were discovered involving SMEs, particularly in Malaysia.

Malaysian SMEs

According to the Bank Negara Malaysia (BNM), enterprises in Malaysia are defined based on economic activities that are divided into categories and sectors based on sales turnover or employment figures. Generally speaking, Malaysian businesses may be divided into two types of general categories; (1) Manufacturing and (2) Services & Others. Manufacturing enterprises are those that turn raw materials into finished goods via the use of chemicals or physical processes. Manufacturing is defined as a firm having a sales turnover of less than RM50 million or full-time employees of less than 200 people (whichever is lower). Meanwhile, business sectors distinguish between services and others, with services

including everything from retail to wholesale to hotels and restaurants to financial intermediation. Other services include everything from manufacturing to distribution to entertainment to financial intermediation to research and development (R&D) to logistics, warehouses, engineering, and other manufacturing-related services like research and development (R&D). In addition, primary agriculture, construction, and mining and quarrying are all examples of economic activity classified as others.

Malaysian SMEs have seen a growth in their number of establishments over the previous decade, with a total of 645,136 new businesses being established in 2011 and 1,151,119 in 2020. As a group, they generated 38.3 percent of the nation's Gross Domestic Product (GDP), which is considered the second-largest contribution and is seen as very essential in the country's nation-building programme. The national budget of Malaysia has budgeted RM1.9 billion to encourage the growth of SMEs in the country by 2021.

Those enterprises that fall outside of the National SME Development Council's (NSDC) definition of a small and medium-sized enterprise (SME) are classified as large firms and do not fall under the SMEs cluster [22].

Businesses that have registered with the Company Commission of Malaysia (CCM) and are thus eligible to be classified as SMEs in Malaysia were used as samples for the research. Companies registered with CCM can be verified by searching for their company name on the SSM e-Info portal. Unregistered businesses, on the other hand, were excluded from the study because they were not officially recognized as valid and had already violated the Registration of Business Act 1956 (as amended). Thus, 18 Malaysian SMEs were selected as cases for the research using theoretical sampling and regarded as expert samples based on their expertise and practices in data processing. This expert sampling technique was used to elicit expert responses

throughout the data gathering process. For this study, expert samples were drawn from company owners who have specialized knowledge in the subject area. They are recognized as experts in the field due to their expertise and practices in enterprise data handling; they are prominent in the company's data handling processes; and they are active in analyzing their data to improve the company's performance. Because the goal is to densify existing emerging categories through range and variation in dimension and fill in the gaps in emerging theory by looking into new emerging categories, data from the analysis provide information on what to sample next. In theoretical sampling, data from the analysis provide information on what to sample next [3], [27]. In grounded theory, Strauss and Corbin [2] identified theoretical sampling stages that occurred in tandem with the phases of coding analysis that took place which happened during Open Coding, Axial Coding and Selective Coding.

Despite the fact that registration of a company is one of the requirements, the definition of SMEs in Malaysia allows for the exclusion of enterprises that are not relevant to the research. Large businesses such as factories, insurance providers, and any other organization that does not meet the criteria of a small and medium-sized enterprise (SME) were eliminated from the study and deemed to be unsuitable since they were not included in the population that was chosen. Some Malaysian businesses outsourced or hired personnel to deal with the company's data management responsibilities; however, these entities were excluded from participating in the survey because they are referring to the needs and requirements determined by business owners, and because certain procedures or actions taken are sometimes unable to be explained in detail, they were excluded from participating in the survey. The sampling approach used to assure reputable and dependable sources of information for a research project has a significant impact on the selection of data providers. The non-probability sampling approach was used in this study to collect data and resources in order to achieve the goals of the study. The non-probability sampling procedure was carried out in three steps in order to minimize data collection conflicts between the various stages.

The data collection method was semi-structured interviews supervised by an interview guide to keep the interview on course, which included several open-ended questions to elicit information on dark data from the respondents' experience and viewpoint [19]. The data were transcribed and analyzed using the Grounded Theory Methodology (GTM). The GTM coding processes were used to the transcribed interviews in three stages: open coding, axial coding, and selective coding. The process of study design is shown in detail in Fig. 1. The numbers in the graphic correspond to the stages of the research process.

Content analysis is the foundation qualitative analysis used in this study as the researcher analyzed the manuscript document of the interview transcription before engaging with data analysis process. These manuscripts of data were recorded from the story told by the theoretical samples which purposively selected cases to initiate the research and to achieve the study objectives. During the analysis, codes and concepts were assigned to phenomenon called as the coding process and then those codes and concepts compared to each other using theoretical comparison known as constant comparative analysis which involving both cross-case pattern analysis and cross-case pattern analysis. Besides, during the GTM analysis, the researcher was required to analyze the data inductively to free himself from any preconceived ideas and concepts to ensure the categories established is grounded by the data. At this point deductive analysis was avoided to protect the data analysis results from any existing theories or knowledge. During the coding process the researcher experienced both indigenous concept analysis known as the in-vivo codes and analyst generated concept which used to provide categories to represent

phenomenon experienced by the respondents as transcribed in the manuscript. Another challenge that the researcher has to face during the analysis is to distinguish the concepts and categories into distinct topic by interpreting meaning and its implication. The concepts and categories were also distinguished using Indigenous Typologies and Analyst Constructed Typologies to accommodate the requirement of the GTM approach and finally answering the research questions.

Memo writing, diagramming, and field notes were analyzed in this study and were deemed to be crucial components that aid in the elicitation of theory using the grounded theory technique. Finally, story lining was utilized to connect all categories and ideas by presenting the problem of dark data management among Malaysian SMEs.
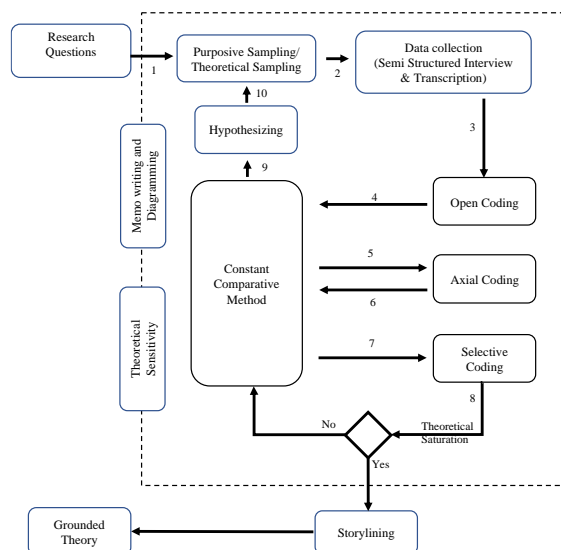

Fig. 1: Grounded Theory Methodology

Findings

Findings of the study indicated that all business owners were employing Dark Data Lifecycle Management (DDLM), which entails appointing a data expert or caretaker and doing data caretaking or data stewarding duties. Not only did business owners utilize DDLM to manage the data in the repository, but also to avoid the emergence of dark data. As seen in Fig. 2, DDLM employs a six-stage lifecycle approach to dark data management.

Specializing Caretaking

A specialized caretaker is assigned to carry out Dark Data Lifecycle Management, which is the first step in DDLM's inception and progression (DDLM). This was overcome by study conducted by Schembera & Duran [20] and Splunk [24], which concluded that the work of dark data management must be carried out by particular persons who are knowledgeable and capable of performing dark data management functions. According to the findings of the survey, the majority of business owners held the bulk of data management duties since they are the individuals who are most acquainted with the overall functioning and decision-making of the company.

Collecting

The second stage of collection is the next step in the DDLM cycle. During business operations, it refers to the activity of procuring data, in which business owners are involved in the process of capturing dark data across multiple platforms, generating data from transactions, gathering data for specific business purposes, and storing the data in a data storage facility. Because of the effects of black data, which refers to data that has been obtained but has gone missing or been destroyed, the respondents' data collection operations were efficiently structured via data procurement and data preserving measures. The research conducted by Kevin et al. [16] and Zhang et al. [28] also revealed that the method of gathering dark data should be used as the first step in order to detect dark data that is already there.
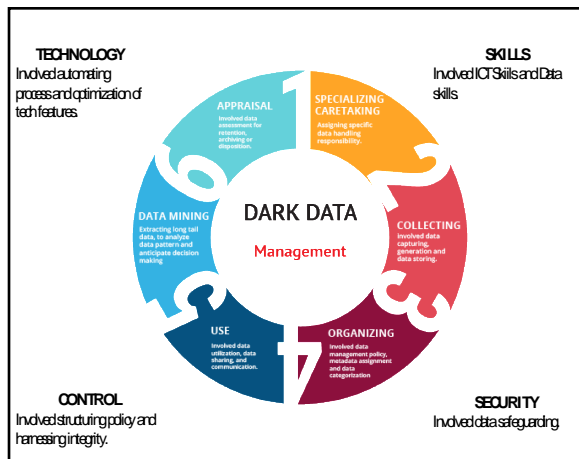
Fig. 2: Dark Data Management Framework

Organizing

As a result of respondents' difficulties in accessing and retrieving dark data after the collection stage, we created a data organization system to address this issue. Because of this disarray, it has been shown that the vast majority of micro-sized organizations have met undisclosed data (black data) that has caused damage to their company's operations and performance. A consequence of this was that the participants added metadata to the stored data (such as categories, headers, and descriptions) in order to make the dark data retrievable and to reduce instances of the dark data phenomenon. A few activities such as cataloguing and categorizing were identified as techniques to organizing dark data in order to improve data accessibility and retrievability, according to the findings of the study. It was the primary goal of organizing dark data, as described by Almeida et al. [2], to make it findable, accessible, interoperable, and reusable, otherwise known as FAIR. This goal could not be achieved by uncataloged data [6] and unclassified data [2],[21].

Use

Furthermore, the collection and structuring of data facilitates the consumption or use of the data collected. People may apply for jobs and engage with one another both within and outside the firm if they have access to and can get their data from the system. Using dark data to aid in decision-making and performance measurement, according to the results, included utilizing the information to make decisions and measure performance. Dark data was also utilized to retrieve data that had been corrupted or rendered inaccessible, and it was shared among authorized users in order to improve data transparency. These signals were also emphasized by Gimple [8], who identified four approaches to use dark data to produce corporate value, including enhancing return on assets, reengaging consumers, updating product offerings, and exploring new business models. Furthermore, Hands [10] said that dark data (data passwords) were also employed to safeguard the data from being accessed by unauthorized individuals, according to the author.

Data Mining

Using dark data requires data analysis and required company owners to acquire additional data. The results show that data mining activities need additional data extraction. Data mining allowed business executives to get a thorough picture of their operations by examining data trends. Concurrently, the participants analyzed dark data to foresee the future and develop the firm. In this attempt, the firm owner seeks for data not covered by the standard data lifecycle management (DLM). It is a new addition to the DLM. Scholars and research articles on dark data management seldom discussed data mining. Few technologies have been built to support Deepdive, Winover, and Datumize [1], [14], [28]. However, the quality of data value evaluation and contextual extraction made dark data extraction characteristics restricted. The research concluded that data mining should be part of the dark data management process rather than a separate approach.

Appraisal

Finally, a decision is made on whether the dark data will be kept or disposed of. It's common for business owners and caretakers to use data mining as a first step in identifying and valuing any darker data they discover over the course of their work. Auditing and weeding procedures were used to verify the validity and accuracy of the dark data, and the value of the dark data was evaluated as a consequence. Data that was no longer needed or relevant was then

permanently deleted, freeing up room for fresh data to accumulate.

The results also show that the DDLM is an attempt to provide a comprehensive analytical recommendation for dark data management solutions. These six stages were supported by the enabling attributes consists of:

1) Technology

It's all about the technology used to process the dark data. The data shows that technology was employed to automate data processing and technological characteristics were adjusted to meet the business demands for dark data processing. As stated by Gimpel and Alter [9], companies should invest in dark data technology and link up with that technology to begin reaping the rewards of dark data analysis.

2) Skills

Data suggests that anybody dealing with dark data should have a working knowledge of ICT equipment, as well as data literacy and data management abilities. Dark data processing may be sped up with the use of ICT equipment rather to being done by hand, which takes a long time. Data scientists and computer engineers are needed to deal with the dark data, according to Schembera and Duran [20], who developed the Scientific data officer position.

3) Security

There is evidence to suggest that data security is concerned with preventing the loss of data as a result of its accessibility, usage, or interoperability. Researchers found that restricting access to data, constantly monitoring data protection to uncover security gaps and treating data catastrophe planning as a critical contingency plan all contributed to the safety of dark data. For the restoration of critical business activities and the continuance of operations until the system can be restored to its pre-disaster condition, these techniques provide adequate data protection [18], [23].

4) Control

According to results, control is a technique through which all data caretaking procedures are linked and adhere to a standard and structured procedure. The policy and the integrity culture are the two most important

aspects of control. As a result, the control mechanism's success rate would be skewed by the absence of one or both of the components. With the policy compliance and integrity culture, dark data may be repressed since it manages all data caretaking operations, and this control characteristic has already been won by Commvault (2014) and Kevin et al. (2016).

Discussion and Conclusion

Dark data refers to valuable assets that are held inside or outside of an institution's repository and storage facility and are difficult or impossible to locate, are unavailable, and are not used due to their unavailability. However, their potential worth would have an impact on the accuracy of data analysis, which would be used to assist strategic planning and decision making in a commercial firm's decision-making process.

Dark Data Lifecycle Management was identified as a strategy that would contribute to the suppression of dark data while also allowing for the mitigation of the risk and effect of dark data. This effective implementation of DDLM is made possible by the enabling attributes, which are comprised of technology, skills, security, and management control. As a result, even if dark data could not be completely eradicated, lowering the frequency of dark data lowered data storage consumption, cost of keeping, and handling, while simultaneously increasing company profitability and improving operation performance.

From the perspective of the study's practical application, the results may pique the attention of corporate executives who want to interact with dark data by using the idea of dark data management and utilizing them for their own profit. By utilizing the DDLM method, business executives will be able to predict future strategic planning and learn about their abandoned valuable assets, referred to as dark data. Aside from that, the establishment of a Dark Data Management Framework will assist those company executives who are new to the process of managing their dark data assets in

gaining a comprehensive understanding of dark data. It is anticipated that the application of theories derived from the study will assist business leaders in obtaining more accurate performance measurement, reducing a significant amount of time spent searching for data, and avoiding lost data trails or business data history as a result of the dark data phenomenon. Furthermore, the implementation of dark data management theory helps business executives to predict future strategic planning for their organizations while also stimulating commercial development.

The models that have been built may also be of interest to software developers who are working on the creation of computer applications for the suppression of the dark data phenomena in businesses. It is possible to utilize the framework of dark data management as a basis for the application by automating dark data mining and value assessment methods, which reduces the time required compared to when these activities are carried out in the traditional manner. Another advantage of this technology is the ability to analyze large amounts of data while still generating precise patterns and analyses.

The findings of the research may be of interest to the information professional community since they offer them with fresh insights into dark data and may serve as a conceptual foundation for future exploration. Additionally, the black data issue has been observed to be encountered by a variety of healthcare providers, engineering industries, and government institutions, among other entities. The study's created theories are based on the behaviors of small and medium-sized enterprises (SMEs) and dark data, but it would be fascinating to see whether they might be applied in other sectors.

In conclusion, the dark data hoarded inside the repository will no longer pose a danger to the data owners but will instead serve as a valuable asset that can be used to leverage and reinforce

competitive advantage, as well as to maintain a profitable company.

Although the study completed with the construction of Dark Data Management framework, however it has been built over the limitation of the study. Therefore, as the use of theoretical sampling, the samples were narrowed down to Malaysian SMEs, and only a small number of samples were explored in order to answer the study's research objectives. In future research, in order to evaluate the robustness of the theory established as a consequence of the research findings, conducting a wider statistical survey approach in order to create generalizability of the theory such as quantitative technique may be deemed appropriate for testing the idea in order to determine its generalizability. Furthermore, future research could extend the validation of the results by embarking in other disciplines with different group of respondents.

REFERENCES

1. Akbar, L. S, Al-Mutahr, K. & Nazeh, M. (2018). Aligning IS/IT with Business Allows Organizations to Utilize Dark Data. International Journal of Innovative Technology and Exploring Engineering. 8(2), 80 - 85.
2. Almeida, C. A. et al. (2021). Excavating FAIR Data: the Case of the Multicenter Animal Spinal Cord Injury Study (MASCIS), Blood Pressure, and Neuro-Recovery. Neuroinformatics. doi: 10.1007/s12021-021-09512-z
3. Charmaz, K. (2006). Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. London: Sage
4. Commvault (2014). 5 Ways to Illuminate your dark data. US: Commvault Systems.
5. Cook, C. N., Hockings, M., & Carter, R. (2010). Conservation in the dark? The information used to support management decisions . Front Ecol Environ, 181-186.
6. Corallo, A. et al. (2021). Understanding and Defining Dark Data for the

Manufacturing Industry. IEEE Transactions on Engineering Management,

7. DiMatteo, S. (2021). Cloud Services: Deftly Deal With Dark Data. Chemical Processing. 83(3), 40-48.

8. Gimpel, G. (2020). Dark data: the invisible resource that can drive performance now. Journal of Business Strategy

9. Gimpel, G. & Alter, A. (2021). Benefit From the Internet of Things Right Now by Accessing Dark Data. IT Professional. 23(2), 45-49.

10. Hand, D. J., (2020). Dark Data: Why What You Don't Know Matters. USA: Princeton University Press.

11. Heidorn, P. B., Stahlman, G. R. & Steffen, J. (2018). Astrolabe: Curating, Linking, and Computing Astronomy's Dark Data. The Astrophysical Journal Supplement Series. 236(3), 1-12.

12. HighQuest Solution (2016). Dark Data Making your Organisation data-enabled? Retrieved on April 4th, 2019 from https://doczz.net/doc/8987800/white-paper-dark-data-making-your-organisation-data

13. Hitachi (2013). Big Data - Shining the light on enterprise dark data (EDD). Retrieved April 15, 2019 from https://www.hitachivantara.com/en-us/resources.html

14. Intel (2018). Datumize and Intel transform dark data into operational insight for manufacturing and logistics. Retrieved 18 February 2022 from https:// www.intel.sg/ content/ dam/ www/ public/ us/ en/documents/solution-briefs/datumize-dark-data-in-manufacturing- and-logistics- solution- brief.pdf

15. Kambies, T. et al. (2017). Dark analytics: Illuminating opportunities hidden within unstructured data. Retrieved on16 Feb 2022. from: https: //www2.deloitte.com/insights/us/en/focus/tech-trends/2017/dark-data-analyzing-unstructured- data.html

16. Kevin, N. M., et. al. (2016). Dark data: Business Analytical tools and Facilities for

illuminating dark data. Scientific Research Journal. 4, 1-10.

17. Lugmayr, A., Stockleben, B., Scheib, C., Mailaparampil, M.A. (2017) Cognitive big data: survey and review on big data research and its implications. What is really "new" in big data?. Journal of Knowledge Management, 21(1), pp.197-212. Retrieved from doi: https://doi.org/10.1108/JKM-07-2016-0307

18. Meilani, D., Arief, I. & Habibitullah, M. (2019). Designing Disaster Recovery Plan of Data System for University. IOP Conference Series: Materials Science and Engineering. 697. 012028. 10.1088/1757-899X/697/1/012028.

19. Patton, M. Q., (2014). Qualitative research and analysis method. (4th ed.). California, SAGE.

20. Schembera, B. & Duran, J. M. (2020). Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer. Philosophy & Technology. 33, 93–115

21. Shukla, M. et al. (2015). POSTER: WinOver Enterprise Dark Data. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15). Association for Computing Machinery, New York, NY, USA, 1674–1676. DOI:https://doi.org/10.1145/2810103.2810131

22. SMECorp (2020). SME Definition. Retrieved from https://www.smecorp.gov.my/index.php/en/policies/2020-02-11-08-01-24/sme-definition

23. Soufi, H. R., Torabi, S. A. & Sahebjamnia, N. (2018). Developing a Novel Quantitative Framework for Business Continuity Planning," Int. J. Prod. Res., pp. 1–22

24. Splunk (2019). The state of dark data. Retrieved on July 3rd, 2020 from https://www.splunk.com/en_us/form/the-state-of-dark-data.html

25. Strauss, A., & Corbin, J. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage

26. Trajanov, D., Zdraveski, V., Stojanov, R. & Kocarev, L. (2018). Dark Data in Internet of Things (IoT): Challenges and Opportunities. Proceedings of the 7th Small Systems Simulation Symposium 2018, Niš, Serbia, 12th-14th February 2018.

27. Urquhart, C. (2013). Grounded Theory for Qualitative Research: A Practical Guide. Sage Publications: London.

28. Zhang, C., Shin, J. Re, C., Cafarella, M. & Niud, F. (2016). Extracting databases from dark data with deepdive. Proceedings of International Conference on Management of Data. 847–859.