Available Online at www.e-iph.co.uk
Indexed in Clarivate Analytics WoS, and ScienceOPEN

E-B Environment - Behaviour Proceedings Journal

KICSS2023

Kedah International Conference on Social Science and Humanities
UiTM Kedah (Online), Malaysia, 21-22 June 2023:
2nd International Conference on Business, Finance, Management and Economics (BIZFAME)

e-IPH e-International Publishing House Ltd., United Kingdom

Enlightening the Repercussion of Dark Data Management towards Malaysian SMEs Sustainability

Ahmad Fuzi Md Ajis[1], Jafalizan Md Jali[2*], Isma Ishak[1], Qamarul Nazrin Harun[1]
**Corresponding Author*

[1] School of Information Science, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Johor, Segamat, Malaysia
[2] School of Information Science, College of Computing, Informatics and Media, Universiti Teknologi MARA Cawangan Selangor, Shah Alam, Malaysia

ahmadfuzi@uitm.edu.my, jafalizan@uitm.edu.my, isma242@uitm.edu.my, qamarulnazrin@uitm.edu.my
Tel: +603-7962 2057

Abstract
The sheer volume of dark data impacts the costs for searching and producing appropriate information and imposes a wasted storage cost in operating budgets. Therefore, a grounded theory research was conducted to investigate the dark data phenomenon towards SMEs in Malaysia. Straussian Grounded Theory Methodology was deployed to analyze collected qualitative data to investigate the repercussions of dark data management towards sustainability of Small & Medium enterprises in Malaysia. Consequently, the study found that dark data is a precious asset to leverage and maintain sustainable business, and a model on the repercussions of dark data management was proposed.

Keywords: Dark Data Management; Malaysian SMEs; Dark Data Repercussion Model; Business Sustainability.

DOI: https://doi.org/10.21834/e-bpj.v8iSI15.5070

1.0 Introduction

Although the existence of dark data may not be a significant hindrance to daily data processing, major issues of concern regarding dark data existence lie in the consequences of the data hoarding conventions and the risks it brings together, such as (1) the cost of data hoarding; (2) data safety threats and (3) data quality risk. In the issue of the cost of data hoarding, Veritas (2015) exhibited in their Databerg report that the data hoarding conventions have resulted in the unrelational IT budgeting basis, which strategized on the foundation of data volumes and not data values, which tremendously increased data management cost. This cost was forcedly shouldered by organizations due to the consequences of the myth of data hoarding embraced by organizations, which are (1) more data means more value, (2) storage is free, and (3) all data are equal in value (Veritas, 2015). As a consequence, while more and more data accumulates and resides deep in the repository, rarely being used and underappreciated, such data are exposed to data safety threats, and stolen data would put business owners in a risky situation (Dimitrov et al., 2018). The circumstances may become worse as the data hoarding convention may let dark data immersed within structured and unstructured data challenge the data quality in the sense of accessibility, accuracy, and traceability, which results in faulty data-driven decision-making (Akbar et al., 2018). In the end, these issues would directly influence organizational performance and sustainability (Gimpel, 2020b).

Despite the fact that many academic institutions, healthcare institutions, research institutions, and large businesses have undertaken studies on dark data (Corallo et al., 2021), only some studies have focused on small and medium businesses (Ajis & Baharin, 2019). Given that numerous studies link dark data with corporate efficiency and performance (Berghel, 2007; Neff, 2018; Waide et al., 2017; Schembera & Durán, 2020), it is natural for people to wonder why they are omitted. This gap prompted the researchers to study Malaysian SMEs.

eISSN: 2398-4287 © 2023. The Authors. Published for AMER and cE-Bs by e-International Publishing House, Ltd., UK. This is an open-access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer–review under the responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers), and cE-Bs (Centre for Environment-Behaviour Studies), College of Built Environment, Universiti Teknologi MARA, Malaysia
DOI: https://doi.org/10.21834/e-bpj.v8iSI15.5070

1.1 Problem Statement

Kambies et al. (2017) reviewed an analysis of the International Data Corporation about the expansion of the digital universe from 2013 to 2020. It was expected that the digital universe (digital data) would reach 44 zettabytes, which is equivalent to 44 trillion gigabytes of data. While data is massively accumulated, the cost of indexing and storing the data to allow data searchability would increase tremendously and burden the data owners' budget. Up to 90% of this data is unstructured, which creates a high relativity of dark data existence (Martin, 2016). 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 (Commvault, 2014; Martin, 2016; Veritas, 2017). These circumstances have put the data owners in a crucial dilemma as to deciding what data to keep and what data to put on hold, which, in the end, will become the safest decision due to the potential future benefit of the data hoarded. This was proven by Splunk (2019), whereby statistically, from 1,300 global business leaders, 60% declared that half or more data in an organization were found to be dark data being kept, while much of it was not understood to exist.

These statistics have provided the primary concern of this study involving data handling issues, as discussed earlier, whereby data hoarding could attract significant risk to the enterprise data, including the cost of data hoarding, data safety threats, and data quality risk. Therefore, this study was executed to investigate the repercussions of dark data management on the sustainability of Malaysian Small & Medium Enterprises (SMEs).

1.2 Context and Scope

The study does not cover the dark data management field as a whole but discovers the dark data phenomenon and its management from the perspective of Malaysian SMEs using a qualitative approach. The core of the study needs to compensate the SMEs for a better way of managing dark data. Instead, it focuses on investigating its current circumstances. The study investigated the repercussions of the dark data phenomenon on SMEs by analyzing the practices of data management to recognize dark data experience based on generic data context. Finally, the study intends to establish the repercussion model of dark data management among Malaysian SMEs.

2.0 Literature Review

The literature searches deployed to look for any research regarding dark data management in Malaysia regardless of their business size, either SMEs or large companies. However, limited papers appeared to be published in Malaysia. Accordingly, it was assumed that Malaysian awareness of 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 influence better business sustainability and intelligence, increased competitive advantages, and minimized enterprise's risks (Hitachi, 2013; Commvault, 2014; Kevin et al., 2016; HighQuest Solution, 2016; Hand, 2020).

Unused and ignored data is the result of the inexistence of proper data organization and management. The distinction between structured and unstructured data does not prevent unused data from residing within the owner's repository. However, Liu et al. (2019) and Imdad et al. (2020) mentioned that the causal factor of data is unused datasets due to the unstructured format of data. The missing meaning of the unstructured data that lies deep within it, which requires to be extracted, such as images, videos, and audio, makes the data to be left unused, but its unknown potential value makes them to be kept longer. This data meaning or context extraction provides a critical dilemma on its prolonged existence, inaccessibility, and unclear value of stored data, which questioned its conservation and its worthiness of mining. These data were collected, processed, and stored, yet they failed to be utilized, termed as the Dark Data (Gartner, 2014). Unfortunately, the formal definition of Dark Data is yet to be established.

Many scholars and researchers publish a variety of Dark Data definitions based on various perspectives and viewpoints. However, the definition of dark data has yet to achieve its global definition. Few scholars emphasized that dark data is information collected as a function of an organization's normal operations but rarely or never analyzed or used to make intelligent business decisions. Most of it gets buried within a vast and unorganized collection of other data assets. Dark data can also be referred to as "data exhaust" because most of the information is considered overlooked information, even though that data has valuable input to the organization. And the portions that aren't of value can be a significant drain on resources, including wasted digital storage space (Martin, 2016).

Dark data exists for any information users and creators as they use any mobile storage devices such as tablets, mobile phones, and laptops. An analogy of an iceberg is an excellent example of how to explain dark data. Approximately 20% of the iceberg would be visible is regarded as the data that are actively used and visible to the organizations and users. Surprisingly, the bottom part of the iceberg, which is the remaining 80% of it, could reside with great opportunity for the organizations and users. However, they are hidden and unexposed, usually kept for reasons such as backup, heritage, and just in case the data is needed in the future.

Characterization of dark data based on its features and properties assist researchers to recognize and define dark data from its silos. Its emergence is considered significant for data keepers as it helps them to manage or even avoid the occurrences of dark data. While some may accept dark data as a harmful substance in the data repository (Hand, 2020; Munot et al. (2019); Lugmayr et al., 2017), some are leveraging them for betterment (Almeida et al., 2021; Gimpel & Alter, 2021; Hawkins et al., 2020; Gimpel, 2020a). Benefiting from dark data requires the organization to manage and deal with dark data during pre-existence and/or post-existence.

3.0 Methodology

The study employed Qualitative Research to investigate the significant purpose of the research. Specifically, Grounded Theory Methodology (GTM) was applied according to Corbin and Strauss (1998) of systematic grounded theory procedures to analyze the collected data. In grounded theory study, research is not initiated based on a theory or framework to be proved, yet it studies the area and allows relevant theory to emerge (Strauss & Corbin, 1998). Figure 1 displays the process of the GTM procedure of the study.

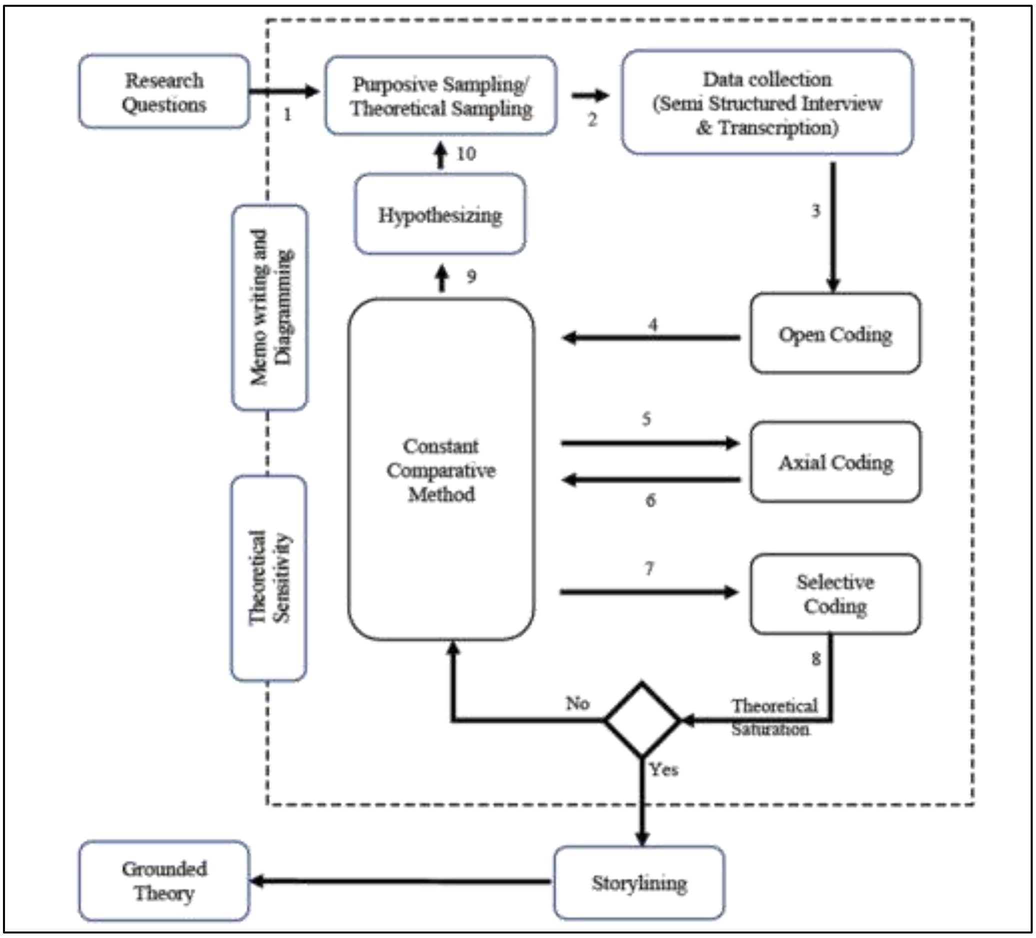

Fig. 1: Grounded Theory Methodology Research Design (Ajis et al., 2022)

The purposive sampling method was conducted in phases involving the selection of expert sampling, snowballing sampling, and inductive grounded and emergent theory sampling (Patton, 2014). According to Patton (2014), a mixture of these approaches provides a quality selection of samples. Therefore, the sample selection started with one purposive sample of Small and Medium Enterprises business owners in Malaysia, which was chosen as the expert sample based on their expertise and practices of Data Management, dominance in the Data Management procedure, and involvement in company performance analysis. This approach is required to provide accurate data for the study. Further, snowballing samples were chosen to ensure quality expert sample selection, as the referrals are someone known by the earlier expert sample. Afterward, theoretical sampling is deployed to densify emergent categories by collecting data to analyze events rather than individuals or respondents. It directs who to sample next, what to explore, and where (Straus & Corbin, 1998). The Theoretical Sampling Guide Matrix, modified from Lehmann (2011), has sequenced and sped up the process of theoretical sampling for this study. Collectively, only 18 companies from all three business sizes, microenterprises, small enterprises, and medium enterprises, were involved as the data completed its saturation (Strauss & Corbin, 1998; Urquhart, 2013). The demographic description helped densify emerging categories through range and variation in properties and dimensions.

Semi-structured open-ended interviews with respondents were captured on video and audio to start the study and uncover dark data concepts and categories. After each interview, recorded data were transcribed, analyzed, and coded to extract the central concept. Constant comparative analysis was used to establish the hierarchical structure of emerging categories and to find saturation before engaging with the following interview (Urquhart, 2013). The analysis comprised open, axial, and selective coding. Open coding was used to open up the data and find emerging categories by dividing it up and performing constant comparative analysis. Then, relationships of emerging categories were established during the axial coding using the coding paradigm (Strauss & Corbin, 1998). Finally, the selective coding phase identified the central category of dark data management. These three coding phases were executed iteratively and went through the hypothesizing stage repetitively for each sample analyzed, which happened a lot during data categorization and data clustering. The hypothesizing stage provides a foundation to enable what to be investigated further and which concepts and categories can complete its saturation (Strauss & Corbin, 1998; Urquhart, 2013; Patton, 2014).

In order to ensure the correct coding procedure, the researcher's theoretical sensitivity must be in place. Theoretical sensitivity is a pre-requisite substance that could be obtained during the precoding or mid-coding process to facilitate the coding process. It is also used to prevent preconception ideas from being forced onto the data. The researcher obtains theoretical sensitivity through literature, professional and personal experience, and knowledge gained during the analytical process. Memo writing, diagramming, and field notes were utilized in the analysis and regarded as essential substances that support the elicitation of theory in the grounded theory method in this research. Atlas.ti application was used to facilitate the analysis process in the management of concepts and categories along with their relationships. In the end, storylining was used to integrate all categories and concepts by describing the phenomenon of dark data and dark data management among Malaysian SMEs to answer the research questions.

Reliability and validity of the study are also significant concerns, whereby the study's validity was established based on credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1984). Prolonged involvement with respondents with little virtual meeting monitoring established research credibility. Triangulation and peer debriefing also ensure data trustworthiness. Due to study limitations, data was triangulated across Micro, Small, and Medium Enterprises (MSMEs). Negative case analysis was also chosen and analyzed to show theory development variation and research norms. This time, a few replies were unfavorable. For assessing referential adequacy, several data recording means were utilized. After analysis, selected respondents were called again for member verification to determine the response to the findings. The study's transferability is limitedly describing the dark data phenomenon and management based on Malaysian SMEs' experience and practices, but its results can be further tested using the quantitative approach for robust generalizability. The research's reliability is further assessed by comparing prior participants' experiences and practices to those of the current participants. Finally, confirmability was used to measure research correctness by requiring facts to be provided unmodified by the inquirer's value judgments.

4.0 Findings

Suppressing the dark data was found to be the consequence as the researchers found that it is impossible for business owners to eliminate the existence of the dark data permanently rather than dealing with the limited occurrences of the phenomenon with the execution of the dark data management strategies. Suppressing dark data refers to the consequences of the business owners' effort to minimize the accumulation of dark data, which comprises the actual and expected outcomes of the strategies proposed in Figure 2. Both benefits and drawbacks were also found to be experienced by the business owners while suppressing the dark data. Part of suppressing dark data outcomes is lighting the dark data, securing competitiveness, mitigating threats, and affecting resources.

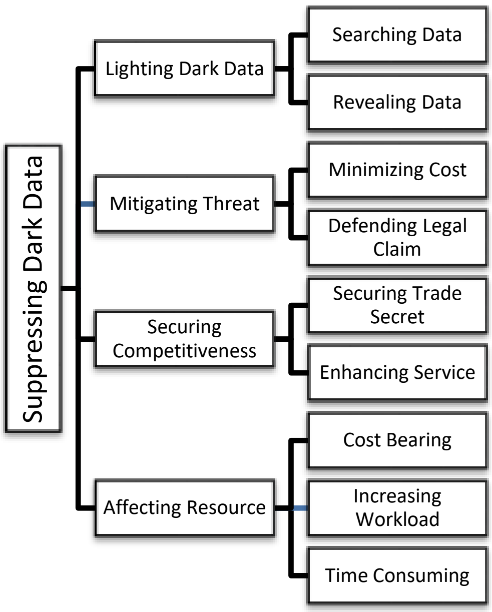

Fig. 2: Proposed Dark Data Management Repercussion Model

Execution of the interaction and strategies towards the dark data phenomenon leads to the repercussion or consequence of dark data management, which suppresses the occurrences of dark data. Suppressing dark data involved lighting the darkness of data, mitigating threats, and securing competitiveness. Dealing with dark data within the context of business sustainability urges optimization of available data for the benefit of the business. Therefore, the discoverability of dark data is one of the significant concerns, whereby the process of seeking information reveals the existence of dark data during business activities. It involves processes of discovering available information and data application by lighting up the dark data. This assists in mitigating the risk of data loss threats and securing competitiveness by enhancing services and securing trade secrets. On the other hand, suppressing dark data affects resources. Since dark data management is a cost-bearing initiative, the workload of the data caretakers increases and consumes a lot of time. Finally, the success-ability of dark data suppression relies on the practical implementation of dark data management.

Dark data suppression is the primary purpose of why the dark data is being managed as to reduce its impact towards surviving the business. Suppressing dark data is referred to as the process of diminishing the occurrences of dark data phenomenon (Schembera, 2019) from being experienced by the business owners, as it was indicated as impossible to eliminate due to the possibility of occurrences being unpredictable as the business still operated. Major suppressing dark data impact is to enhance data quality in the aspect of data accessibility, accuracy, and data auditability (Akbar et al., 2018). Schembera (2019) suggested by enriching metadata using technical, descriptive, process, and domain-specific metadata, dark data can be reduced, and dark data FAIR attributes are plausible. Moreover, managed dark data provides a sufficient asset to mitigate the risk of minimizing data loss and securing business competitiveness by enhancing service delivery and protecting trade secrets.

Although suppressing dark data has a beneficial impact, the findings indicated it also influences complications toward successful dark data management. Suppressing dark data would affect enterprise resources, whereby suppressing dark data initiatives required capital investments in hiring data specialist personnel procuring data handling technologies and data storage facilities. This was indicated by Veritas (2015), who claimed in their Databerg report that redundant, obsolete, and trivial (ROT) data could cost organizations expensive storage fees. Therefore, dark data storage fees might not be excluded from this matter. Other than that, the data caretaker's workload will be increased with the task of dark data management procedures as dark data may reside within business operation data derivatives, and those procedures consume a lot of time.

5.0 Discussion

The grounded theory analysis found that providing rich metadata to increase the findability of dark data would not guarantee suppression of dark data as its characteristics are not merely unfound data. Redundant, obsolete, and trivial data, also known as the R.O.T data, has always been regarded by scholars and industries as dark data (Veritas, 2015), but the data analysis suggests differently. The findings indicate that dark data may redundantly exist, but they are not obsolete and trivial. Dark data is not obsolete based on its birthdate, but it becomes old and unused due to the inexistences of data mining and value appraisal activities during caretaking processes. Besides, trivial data should be disregarded from dark data as the data indicates that dark data is vital and beneficial as the utilization of dark data has grown business revenue among the majority of the respondents.

In regard to information professionals, the community suggests that FAIR data is what should be achieved to suppress dark data (Schembera, 2019; Almeida et al., 2021). While the findability (F), accessibility (A), and interoperability (I) are acceptable for fulfilling this hope, the reusability (R) has forsaken the nature of the dark data which exists in black and grey. The analyzed data indicates that usability (U) of data is much more promising than reusability (R) to suppress the dark data as reusability only regards data that is readily owned and exists in the storage, which this study called grey data. Reusability disregards black data as part of dark data, which is not only within the organizational repository but could be anywhere as it also relates to unknown data, disconnected data, and unfound data. Moreover, the study indicates that the dark data management approach creates significant implications for enterprise owners' resources as they are required to spend resources to manage dark data either to avoid the company's downfall or to tap hidden potential data value in their repositories.

6.0 Conclusion and Recommendation

The study's primary purpose is limited to studying the repercussions of dark data management on the sustainability of Malaysian SMEs. The finding of the study is limited to data collection from interviews of 18 enterprise owners based on their experience and realistic practices of dark data management. It is recommended that observation data provide more detailed findings to describe the repercussions of dark data management on the sustainability of business. The study exhibited that the phenomenon of dark data happened within the effort of sustaining the business. Within this survival effort, dark data occurred and was characterized by the dark data shadows, which are black and grey data, which are both differentiated by their findability, accessibility, interoperability, possession, and usability. Although suppressing dark data provide significant benefits to SMEs, it also comes with complication whereby the resources of SMEs is affected as the execution of dark data suppression incurs cost, increase data management workload, and consume much time. As these findings were derived using a qualitative approach, the robustness of the proposed model developed by the study results could be tested using a more extensive statistical survey technique to generate quantitative generalizability. Further, as data is utilized in any field, a universal contextualization of dark data management repercussions could be achieved if future research could extend the validation of the results by embarking in other disciplines with different groups of respondents and studying in different study environments. Future research on areas such as Dark Data Value Assessment and Dark Data Retention and Disposition in the government sector would provide significant implications for the dark data management field and body of knowledge.

Acknowledgement

This research publication was funded by the Universiti Teknologi MARA Cawangan Johor under the fund of Geran Bestari Fasa 1/2022.

Paper Contribution to Related Field of Study

The study findings contribute to the area of information management at two levels: theory and practice. The significant contribution is at the level of the theoretical part of dark data management. The model of dark data management repercussion constructed in this study contributes to the body of knowledge by providing an expansion on the understanding of dark data phenomenon and its consequences. On the other side, the revelation of the theory on dark data and its management provides new insight into how the industry could leverage its dark data and benefit from it.

References

Ajis, A. F. M. & Baharin, S. H. (2019) "Dark data management as frontier of information governance," in Proc. 9th IEEE Symp. Comput. Appl. Ind. Electron., pp. 34-37, doi: 10.1109/ISCAIE.2019.8743915.

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.

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.

Berghel, H., Hoelzer, D., & Sthultz, M. (2008). Chapter 1 Data Hiding Tactics for Windows and Unix File Systems. Advances in Computers. 74, pp. 1-17, doi: 10.1016/s0065-2458(08)00601-3

Commvault (2014). 5 Ways to Illuminate your dark data. US: Commvault Systems.

Corallo, A., Crespino A. M., Vecchio, V. D., Lazoi, M. & Marra, M. (2021). Understanding and Defining Dark Data for the Manufacturing Industry. IEEE Transactions on Engineering Management, vol. 70, no. 2, pp. 700-712, Feb. 2023, doi: 10.1109/TEM.2021.3051981.

Dimitrov, W., Siarova, S. & Petkova, L. (2018). Types of dark data and hidden cybersecurity risks. DOI: 10.13140/RG.2.2.31695.43681

Gartner (2014). Turning Dark Data into Smart Data: How Email and File Level Analytics Can Lead to Greater Business Value in the Age of Information. Retreived on Apr 4th 2023 from https://www.gartner.com/en/information-technology/glossary/dark-data

Gimpel, G. (2020a). Dark data: the invisible resource that can drive performance now. Journal of Business Strategy, Vol. 42 No. 4, pp. 223-232. https://doi.org/10.1108/JBS-02-2020-0046

Gimple, G. (2020b). Bringing dark data into the light: Illuminating existing IoT data lost within your organization. Business Horizons 63, 519-530, doi: 10.1016/j.bushor.2020.03.009

Gimpel, G. & Alter, A. (2021). Benefit From the Internet of Things Right Now by Accessing Dark Data. IT Professional. 23(2), 45-49, doi: 10.1109/mitp.2020.3025483

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

Hawkins, B. E., Huie, J. R., Almeida, C., Chen, J. & Ferguson, A. R. (2020). Data Dissemination: Shortening the Long Tail of Traumatic Brain Injury Dark Data. Journal of Neurotrauma. 37, 2414–2423, doi: 10.1089/neu.2018.6192

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

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

Imdad, M. et al. (2020). Dark Data: Opportunities and Challenges. International Research Journal of Computer Science and Technology. 1(1), 38-46.

Kambies, T., Roma, P., Mittal, N. & Sharma, S. K. (2017) Dark analytics: Illuminating opportunities hidden within unstructured data. Retrieved on Oct. 16, 2022 from https: //www2.deloitte.com/insights/us/en/focus/tech-trends/2017/dark-data- analyzing- unstructured- data.html

Kevin, N. M., et. al. (2016). Dark data: Business Analytical tools and Facilities for illuminating dark data. Scientific Research Journal. 4, 1-10.

Lehmann, H. (2011). The Dynamics of International Information Systems: Anatomy of a Grounded Theory Investigation. New Zealand: Springer.

Lincoln, Y. S., Guba, E. G. (1984). Naturalistic inquiry. California: Sage.

Liu, Y. et al. (2019). A Framework for Image Dark Data Assessment. In: Shao, J., Yiu, M., Toyoda, M., Zhang, D., Wang, W., Cui, B. (eds) Web and Big Data. APWeb-WAIM 2019. Lecture Notes in Computer Science(), vol 11641. Springer, Cham. doi:10.1007/978-3-030-26072-9_1.

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, doi: 10.1108/JKM-07-2016-0307

Martin, E. J. (2016). Dark Data: Analyzing Unused and Ignored Information. Retrieved on April 4th, 2023 from https://www.thetilt.com/content/dark-data-analyzing-unused-ignored-information

Munot, K., Mehta, N., Mishra, S. & Khanna, B. (2019). Importance of Dark Data and its Applications. 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). 1-6, doi: 10.1109/ICSCAN.2019.8878789.

Neff, E.P. (2018). Dark data see the light. Lab Animal. 47(2), 45-48, doi: 10.1038/laban.1405

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

Schembera, B. (2019). The dark side of data management. Retrieved on April 4th, 2023 from https://www.researchgate.net/publication/355545901

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, doi: 10.1007/s13347-019-00346-x

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

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

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

Veritas (2015). The databerg report: see what others don't, identify the value, risk and cost of your data. California: Veritas.

Veritas (2016). State of Information Governance: 2016 Report. California: Veritas

Veritas, DLT Solutions & GovLoop (2017). Dark Data Management: The Next Frontier for Government Data. Washington: GovLoop

Waide, R.B., Brunt, J.W., & Servilla, M.S. (2017). Demystifying the landscape of ecological data repositories in the United States. BioScience. 67(12), pp.1044-1051.