ILLUMINATING DARK DATA: A SYSTEMATIC LITERATURE REVIEW OF ITS CHARACTERISTICS AND MANAGEMENT STRATEGIES

Jafalizan Md Jali[1*], Mohamad Khairul Duan[1*], Ahmad Fuzi Md Ajis[1]

[1]*Center of Information Studies College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, UiTM Puncak Perdana Campus, 40150 Shah Alam, Selangor, Malaysia*

[1]*Center of Information Studies College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Cawangan Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*

[1]*Center of Information Studies College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, UiTM Cawangan Johor, 85000 Segamat, Johor, Malaysia*

[*]Corresponding Author

Email: khairulduan@uitm.edu.my

Received: 9 May 2025
Accepted: 10 June 2025

ABSTRACT

Dark data refers to information collected by organisations that remains underutilised, often due to its unstructured nature. This article presents a systematic literature review (SLR) to define dark data comprehensively and examine its characteristics and management strategies. Despite the growing recognition of its potential value, a universally accepted definition remains absent, with scholars offering varied interpretations. The review analysed 182 articles before narrowing the scope to 81 relevant publications that highlight the difficulties in categorising and accessing dark data. Key characteristics identified include unused, ignored, and unmeasured, which are often generally in the organisation's repositories. This study further explored the implications of dark data on compliance, cybersecurity risks, and organisational efficiency. By addressing the existing literature gaps, this research aims to provide a clearer understanding of dark data and propose frameworks for its effective management. The findings are particularly relevant for any organizations in Malaysia, providing a foundation for future research and practical strategies to optimise data utilisation while mitigating associated risks. Ultimately, this work underscores the need to recognise and manage dark data as a critical component of modern data governance strategies.

Keywords: Dark data, data management, data utilisation, Systematic Literature Review, unstructured data

1.0 INTRODUCTION

Unused and ignored data often result from improper data organisation and management. The distinction between structured and unstructured data does not prevent them from remaining unused and stored within the owner's repository. However, Imdad et al. (2020) noted that unstructured formats are a primary cause of datasets being left unused. The absence of clear meaning in unstructured data such as images, videos, and audio requires further extraction, which often leads to the data being neglected. Nevertheless, its unknown potential value often leads organisations to retain it for more extended periods. George et al. (2023) agreed that dark data remains dormant within an organisation despite its potential intrinsic value. Extracting meaning or context from such data presents a critical dilemma, raising questions about its continued storage, limited accessibility, and uncertain value. Zhou (2020) observed that many social platforms store image data, such as albums and chat images, as standalone resources, separated from core business operations. Over time, these vast collections of image data often become dark data, preserving historical records that cannot be deleted. In the manufacturing industry, Angelo et al. (2023) noted that data is often stored with the intention of future use, yet its immediate value is frequently overlooked. These datasets, although collected, processed, and stored, ultimately remain unused and are referred to as dark data (Gartner, 2017). Unfortunately, a formal definition of dark data has yet to be established.

Many scholars and researchers have published various definitions of dark data from different perspectives and viewpoints. Some scholars have emphasised that dark data refers to information collected during an organisation's routine operations but rarely, if ever, analysed or used to support informed business decisions. Much of this data becomes buried within vast and unorganised repositories. Munot et al. (2019) stated that dark data refers to information obtained and stored by companies for potential future research purposes, which then remains underutilised due to limitations in time and computational resources. Therefore, collected data must be processed, examined, and interpreted to be effectively utilised. Dark data is also referred to as "data exhaust" because it often comprises information that is overlooked despite its potential value to the organisation. Moreover, portions of this data that lack current value may still significantly drain resources, particularly in terms of wasted digital storage space (Martin, 2016).

Dark data is present in any information created or used through mobile storage devices such as tablets, mobile phones, and laptops. Similarly, Veritas (2015) proposed a databerg as in figure 1 to represent the holistic view of dark data. An iceberg analogy is commonly used to explain the concept of dark data. Approximately 20% of an iceberg is visible above the surface, representing data that is actively used and accessible to organisations and users. The remaining 80%, hidden below the surface, symbolises data that may hold significant potential value but remains unseen. This hidden portion of data is often kept for backup, historical reference, or in case it is needed in the future. While some publications have discussed dark data, there is still a lack of literature that defines it in a comprehensive and theoretical manner. This gap has prompted researchers to conduct a systematic literature review (SLR) on the subject of dark data.

Fig 1. : *Databerg (Veritas, 2015)*

2.0 LITERATURE REVIEW

Dark data has been defined from various perspectives. Corallo et al. (2021) reviewed 22 publications, including both academic and non-academic sources, that addressed definitions of dark data across different fields and viewpoints. Their SLR aimed to support the development of a definition specific to the manufacturing industry. The establishment of dark data definition appears to involve several properties, including searchability (Kambies et al., 2017), accessibility (Banafa, 2015), unknown existence (Lugmayr et al., 2017) and data that is uncategorised or ignored (Intel, 2018). These characteristics are often influenced by data formats, which lead to the data being unused but still retained over time (Trajanov et al., 2018). However, the epistemological depth of dark data definitions has yet to be thoroughly explored.

Hand (2020) published a book in which he defined dark data by identifying various types based on his professional experience and observations across multiple fields. Some of these types were drawn from real-world events and published examples. The classifications are as follows:

- DD- Type 1: Data We Know Are Missing
- DD- Type 2: Data We Do not Know Are Missing
- DD- Type 3: Choosing Just Some Cases
- DD- Type 4: Self- Selection
- DD- Type 5: Missing What Matters
- DD- Type 6: Data Which Might Have Been
- DD- Type 7: Changes with Time
- DD- Type 8: Definitions of Data
- DD- Type 9: Summaries of Data
- DD- Type 10: Measurement Error and Uncertainty
- DD- Type 11: Feedback and Gaming
- DD- Type 12: Information Asymmetry

- DD- Type 13: Intentionally Darkened Data
- DD- Type 14: Fabricated and Synthetic Data
- DD- Type 15: Extrapolating Beyond Your Data

The 15 types of dark data described by Hand (2020) remain insufficient to fully reveal the complexity of dark data. These classifications primarily define dark data from a "missing" perspective, referring to data that is missing in meaning, missing from collection or inclusion, or missing from the awareness of its creators, users, and recipients.

The definition of dark data should be considered beyond the paradigm of its possession or discovery, even though these aspects form part of the dark data landscape. Although Corallo et al. (2021) proposed that the key concepts relevant to defining dark data in the manufacturing industry include data description (cataloguing), capturing activity, and advancements in analytical tools, sources, and data formats, the definition remains insufficiently explored. Therefore, the researchers suggested that the definition of dark data should be examined through the following paradigm:

- its usage activity status, categorised into active and inactive conditions
- duration of the data retention period
- ownership and controllability, including aspects of capturing, storing, and managing the data
- awareness and discoverability
- treatment and existence of the data.
- Varying shades of the dark data

3.0 METHODOLOGY

The SLR was conducted with the guidance of an established framework for conducting and reporting systematic reviews (Moher et al., 2009). A search strategy was then formulated. The search coverage aimed to cover as much published literature as possible and focussed on articles indexed in databases such as Emerald Insight, ScienceDirect, Scopus, and Web of Science. Thus, no publication filter was applied. The exact phrase "dark data" was used to initiate the literature search.

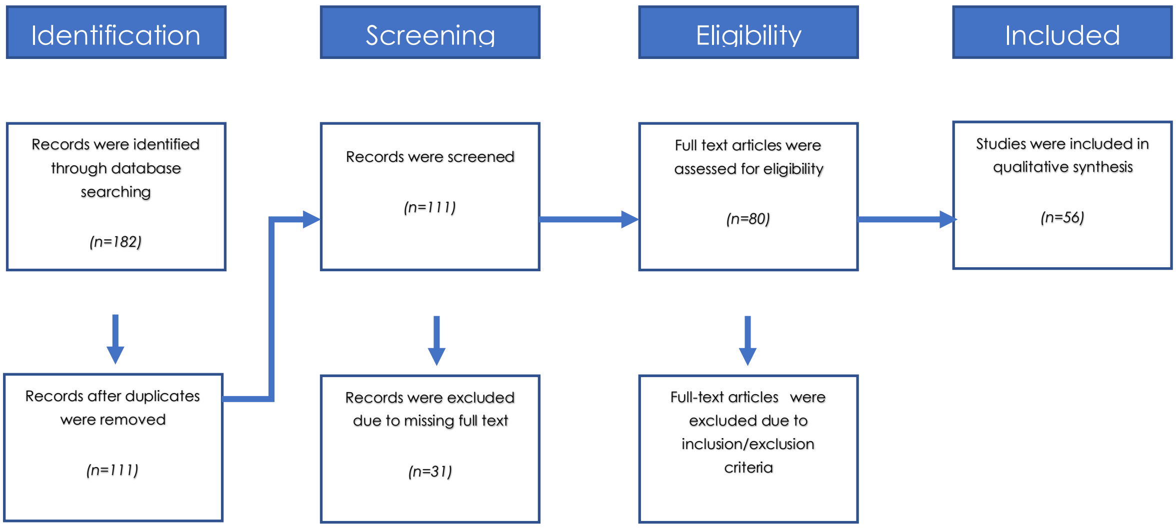

Fig. 2: *Filtering process adapted from Moher (2009).*

The inclusion/exclusion criteria considered the relevance of article content to the area of study, which is dark data, and required English as the medium of delivery, as the researchers are not proficient in other languages. The review focussed on identifying patterns in the definitions of the dark data phenomenon, its causal factors, and strategies used to manage dark data. The selected papers were read and analysed to identify publications that mentioned the term dark data. Relevant passages were then analysed and categorised using Atlas.ti to address the objectives of the SLR.

4.0 FINDINGS

The search returned a total of 182 journal articles. After the removal of duplicates, only 111 articles remained. These results were then screened to obtain full-text articles, during which 31 publications were excluded while the remaining 81 were assessed for eligibility. Each passage could be assigned to more than one category, and by the end of the coding process, 205 text passages from 56 articles were successfully coded.

Gartner's definition of dark data was frequently cited in the literature analysed. It defines dark data as the information assets that organisations collect, process, and store during regular business activities but generally fail to utilise for other purposes (for example, analytics, business relationships, or direct monetisation). A review of the 56 articles indicated that definitions of dark data were often derived from the characteristics of the data itself. Codes representing these characteristics were conceptualised from the text passages and subsequently categorised as exhibited in Table 1.

Unused data is widely acknowledged by most scholars as a key characteristic. These data remain unanalysed or unexploited yet are still retained by the owner. Some data are excluded from publication, preventing their dissemination and potential use. Furthermore, dark data is also defined as ignored data that is hoarded, neglected, and left unprocessed. Some are underused, unmeasured, untapped, or unvalued, and their continued storage may increase operational costs and reduce the efficiency of storage systems, especially in digital environments. In the case of tangible data, excessive storage may overwhelm human capacity for control and management. Some scholars further define dark data as information that is hidden and difficult to locate, particularly when embedded in unstructured formats such as image files. In some cases, the data is completely lost or remains unseen.

Table 1: *Dark data characteristics identified by scholars*

Researchers	Concepts or Dark Data Characteristics	Codes
Angelo, Maria, Vito, Mariangela & Manuela (2021); Cecilia, Cano, Morales-García, Llanes,& Imbernón (2020); Easterday, et al. (2018); Hernández, Arcas-Túnez, Muñoz, & Cecilia (2018); Akbar, Al-Mutahr & Nazeh (2018); Krüger & Marshall (2017); Björnmalm, Faria, & Caruso (2016); Angelo, Maria, Vito, Mariangela & Manuela (2021); Hawkins, Huie, Almeida, Chen, & Ferguson (2020); Brooks, Bryan Heidorn, Stahlman, & Chong (2016); Patil & Siegel (2009); Heidorn (2008); Illig (2008)	not analysed data; unconsidered for publication; undisseminated data; unexploited collected data; unpublished data, data that did not make the cut to be published;	unused data
Angelo, Maria, Vito, Mariangela & Manuela (2021); Munot, Mehta, Mishra & Khanna (2019); Easterday, et al. (2018); Gimpel (2020b); Hand (2020); Gimpel & Alter (2021); Shukla, Manjunath, Saxena, R., Mondal & Lodha (2015); Voulgaridis, Angelopoulos & Chatzimisios (2020); Akbar, Al-Mutahr & Nazeh (2018); Purss, et al. (2015)	hoarded data; neglected data; neglected valuable potential information; unprocessed data; underused data; unmeasured data; untapped data; unused data; unvalued data	ignored data
Berghel (2007); Neff (2018); Angelo, Maria, Vito, Mariangela & Manuela (2021); Waide, Brunt & Servilla (2017); Schembera & Durán (2020); Brooks, Bryan Heidorn, Stahlman, & Chong (2016); Berghel, Hoelzer & Sthultz (2008); Illig (2008); DiMatteo (2021); Heidorn (2008); O'Shea (2008)	covert data; hardly located data; hidden data; hidden in unstructured data; invisible data; lost data; unable to be located; unseeable data; unseen data	hardly located data

Schembera & Durán (2020); Gimpel & Alter (2021); Angelo, Maria, Vito, Mariangela & Manuela (2021); Kolesnichenko, et al. (2017)	incompatible data format; ineffective used data; mismatched data standard; unrecognized data	unusable data
Angelo, Maria, Vito, Mariangela & Manuela (2021); Hand (2020); Gimpel (2020b); Schembera & Durán (2020)	disintegrated use of data; missing data; partially visible data; unstructured data; virtually non-existent data	missing data
Easterday, et al. (2018); Angelo, Maria, Vito, Mariangela & Manuela (2021); Stahlman, Bryan Heidorn, & Steffen (2018); Heidorn, Stahlman, & Steffen (2018); Gimpel (2020a)	unarchived data; unavailable data; uncurated data; unsaved data; unstored data	lost data
Gimpel (2020a); DiMatteo (2021); Angelo, Maria, Vito, Mariangela & Manuela (2021)	unknown data; unknown data context or value being kept; unknown data on performance comparison; unknown data which cannot quantified or qualified; unknown modified status of data	unknown data
Angelo, Maria, Vito, Mariangela & Manuela (2021); Schembera & Durán (2020); Gimpel (2020b); Easterday, et al. (2018); Shukla, Manjunath, Saxena, Mondal, & Lodha (2015)	cannot be categorized; not carefully labeled data; unorganized data; untagged data	uncategorised data
Schembera & Durán (2020); Munot, Mehta, Mishra & Khanna (2019); Angelo, Maria, Vito, Mariangela & Manuela (2021); Gimpel (2020b)	missing metadata data; uncatalogued data; unclassified data; unreachable data	inaccessible data
Angelo, Maria, Vito, Mariangela & Manuela (2021); Schembera & Durán (2020); Lugmayr, Stockleben, Scheib, & Mailaparampil (2017); Hodgkinson, K., Rezgui, A. (2013)	uncaptured data; uncollected data; unaware data	wild data
Angelo, Maria, Vito, Mariangela & Manuela (2021); Björnmalm, Faria, & Caruso (2016); Gimpel (2020a); Schembera & Durán (2020); Gimpel (2020b)	inaccessible collected data; unindexed data	unindexed data
Gimpel (2020a)	wrongly organised data	scattered data

Damaged or unusable data is another characteristic of dark data, often resulting from incompatible data formats due to mismatched data standards, which lead to ineffective usage and render the data unrecognisable. Besides, missing data is also a key aspect, typically caused by the disintegrated use of data collected from different tools and stored in varying formats. Partially visible data adds to the complexity of data integration. More critically, the loss of data with known value and potential benefits is concerning. Such loss is often due to mishandling of data, including the absence of proper data curation, archiving, and storage.

Other features of dark data have also been characterised by scholars in terms of its unknown conditions and circumstances, where the context and value of the data are unknown to the keeper. This is further intensified by its inability to be quantified and qualified by any means. On the other hand, dark data is also characterised by the data handling practices of the keeper. Its existence is often the result of inactive processes related to data categorisation, cataloguing, classification, categorisation, collection, indexing, and data organisation.

Although the features and characteristics of dark data have been discussed in various publications, a definition of dark data based on proper theory remains vague and insufficiently explained through real-world phenomena. The majority of the reviewed publications discussed dark data from an academic perspective, and most research was conducted by research-based organisations. In contrast, from a non-academic perspective, only large firms have published white papers and research articles on the existence of dark data, its management, and potential benefits (Angelo et al., 2021). Therefore, conducting dark data research from the perspective of any organizations in Malaysia could address the existing gap in the literature.

5.0 DISCUSSION

Dimitrov et al. (2018) regard the impact of dark data as a risk that could harm business enterprises. The existence of dark data within an enterprise's repository may interfere with legal compliance related to data management regulations and legislation. In Malaysia, for example, the establishment of the Personal Data Protection Act (PDPA) prohibits data from being retained indefinitely. Data storage beyond seven years must be justified, and unreasonable excuses for retaining data can result in penalties.

Dimitrov et al. (2018) also explained that dark data is often deeply embedded and rarely used or recognised. Unlike currently used data, its existence increases vulnerability to cybersecurity threats, including breaches of personal data and data theft. This places business owners in a highly risky position. Stolen dark data can jeopardise intellectual property rights and compromise trade secrets. At the same time, the leakage of such neglected data may result in a loss of business intelligence if accessed by malicious actors. Security breaches and data theft can damage an enterprise's reputation and harm profitability. Moreover, businesses that fail to invest in exploring and managing their dark data may lose valuable opportunities for improvement. These include increasing productivity, analysing staff and consumer behaviour to enhance services and profitability, and avoiding future liabilities. At worst, a business that ignores the presence of dark data will remain continuously exposed to its risks and miss out on its potential value.

The impact of dark data on the data quality. The existence of dark data, which is embedded within both structured and unstructured datasets, presents challenges to the accessibility of unstructured data. Unstructured data often possess valuable information in the form of images, videos, and audio, which are difficult to retrieve using standard computer-based keyword processing methods. Unfortunately, this type of unstructured data contributes to the growth of dark data and may include redundant, obsolete, and trivial (ROT) content (IBM, 2016). The resulting obstruction to data accessibility can jeopardise the accuracy of data analysis and could lead to faulty data-driven decision-making. Hence, It is essential to manage and organise such data effectively, as proper utilisation can enhance the quality of decision-making and improve the precision of business insights.

Commvault, an IT company that has addressed the impact of dark data, reported the costs that business entities may incur due to its presence. Storing dark data within internal repositories requires substantial storage capacity, which significantly increases the cost of storage management. The issue worsens with the expansion of file sharing among employees using remote devices such as mobile phones, where data and documents can be uploaded by any authorised staff member, further contributing to the accumulation of dark data.

Enterprises may seek to implement data management and analytics strategies to extract value from the data with increasing awareness of the vast amount of data stored. However, as the volume and complexity of the data grow, the associated processing costs also increase. These costs could potentially have been minimised if data governance had been implemented at an early stage of the business. Expenses also rise with the introduction of data extraction technologies used to mine dark data. as well as the need to train staff not only to recognise the presence of dark data but also to apply appropriate technologies in managing it effectively.

Moreover, a case study on smart grid electricity meters found that dark data becomes highly valuable when used to profile customer behaviour and demands, as well as to minimise power failures (Imdad et al., 2020). In clinical settings, oversight by medical professionals is essential for managing dark data effectively and uncovering previously unknown information (Kim, 2024). Besides, there is an urgent need to implement advanced computational solutions to identify patterns in big data and manage dark data for healthcare analysis (Letafati et al..2023). Dark data can also provide strategic advantages in business by supporting the development of profitability-focused strategies. According to Moumeni et al. (2021), unmanaged and uncategorised dark data can significantly increase the time and costs associated with finding, reviewing, and analysing information during the discovery process.

Poor management of dark data may result in security vulnerabilities, regulatory non-compliance, inefficiencies, and resource wastage. Ultimately, this can undermine the credibility of data-driven decisions and strategies. Gimpel (2020) emphasised the importance of educating all employees about dark data until they recognise its potential value to the organisation. For example, consider a computer game company. While analysing strategies to boost profits, experts and consultants might overlook the influence of consumer gender in product sales. A simple review could reveal that over 80 percent of mind game players are girls, whereas around 65 percent of strategic games are purchased by boys. Surprisingly, the company may not yet realise that girls could generate higher profits than boys. In this case, the company's dark data was gender information that it had not previously recognised or utilised (Faghih et al.,2021).

The utilisation and management of dark data have shown tangible benefits. Commvault, for instance, successfully reduced storage costs by 70 percent, minimised risk and liability through automated data deletion and implemented improved data classification to enhance access control and security. Jackson and Hodgkinson (2022) developed a model to explore how digitalisation aids in decision-making processes. However, they also noted that it could limit in-depth discussions, reduce knowledge reuse, and potentially contribute to the growth of dark data while increasing the digital carbon footprint.

Schembera and Durán (2019) also proposed the establishment of a role known as Scientific Data Officer (SDO) to tackle the issues associated with dark data. This role involves detecting, organising, and overseeing dark data within organisations to ensure it is not neglected, underutilised, or mismanaged. Akbar et al. (2018) agreed that employing proper Information Technology and Information Systems for dark data management may create profitable opportunities for businesses, although they acknowledged that the risk of unsuccessful implementation remains.

6.0 CONCLUSION

The study contributes to addressing a gap in existing research, where definitions and characteristics of dark data have been assigned in various ways by researchers and scholars. The findings help identify the core characteristics that define dark data. However, the study limited to the all publications indexed in four databases as mentioned in the methodology section with coverage upon the field of dark data. Surprisingly only 81 selected publication were analyzed after filtration of relevant publications due to exclusion criteria. Veritas (2015) analogised the dark data using the concept of a "databerg," illustrating how dark data resides within organisational datasets. The tip of the iceberg (databerg) represents "bright data," which is regularly and actively used by the data owner, while the submerged portion symbolises dark data, which holds potential value but remains hidden. However, based on the study's analysis, the databerg analogy is considered inaccurate and does not fully capture the complexity of dark data. The bottom of the iceberg only represents a portion of dark data, which continues to grow as information is stored for "just in case" reason. The scope of dark data extends beyond what is depicted at the bottom of the iceberg. Similarly, the widely cited Gartner (2014) definition of dark data is also viewed as incomplete in its description of dark data.

Consequently, based on the review, the study defines dark data as "knowledge, information or data assets within or beyond data repositories with unknown potential value, lacking findability, accessibility, interoperability, and usability due to data crises and faulty data caretaking procedures. Additionally, data has become a scientifically and technologically valuable commodity of significant public interest. Therefore, it is imperative that data be reusable in research and accessible to the entire scientific community (Schembera & Durán, 2020). Kim (2024) also emphasised that understanding and managing dark data is crucial for the effective utilisation of big data, particularly given its size and diversity. Experts need strategic insight to interpret dark data, a view supported by Moumeni et al. (2021), who noted that enhancing the value of dark data can reduce costs, minimise the risk of data breaches, and improve productivity.

ACKNOWLEDGEMENT

This research was funded by Universiti Teknologi MARA Cawangan Johor under the Geran Bestari Fasa 1/2023.

REFERENCES

Angelo, M. (2021). Characteristics of dark data: A systematic review. J*ournal of Data Management, 12*(3), 45-67.

Banafa, A.(2015). The impact of dark data on organisational performance. *International Journal of Information Systems, 10*(2), 100-115.

Berghel, H. (2007). The dark side of data: Understanding dark data and its implications. *Communications of the ACM, 50*(8), 28-30.

Corallo, A., Crespino, A. M., Del Vecchio, V., Lazoi, M., & Marra, M. (2021). Understanding and defining dark data for the manufacturing industry. *IEEE Transactions on Engineering Management, 70(*2), 700-712. doi: 10.1109/TEM.2021.3051981

Dimitrov, D.(2018). Dark data as a risk factor in business enterprises: Implications for compliance and cybersecurity. *Journal of Business Ethics*, *149*(4), 789-802.

Faghih, N. (2021). Entrepreneurial motivation index: Importance of dark data. *Journal of Global Entrepreneurship Research*, *11*(1), 15-27. https://doi.org/10.1007/s40497-021-00277-y.

Gartner. (2017). *The concept of dark data and its implications for organisations*. Gartner

George, A. S., Sujatha, V., George, A. H., & Baskar, T. (2023). Bringing light to dark data: A framework for unlocking hidden business value. *Partners Universal International Innovation Journal*, *1*(4), 35-60.

Gimpel, G. (2020). Dark data: The invisible resource that can drive performance now. *Journal of Business Strategy*,(ahead-of-print). https://doi.org/10.1108/jbs-02-2020-0046.

Gimpel, H., & Alter, S. (2021). Understanding the characteristics and challenges of dark data in organisations. *Information Systems Research, 32*(2), 300-315.

Hand, D. J. (2020). D*ark data: A guide to understanding and managing unused information*. Cambridge University Press.

Imdad, M.(2020). The hidden value of dark data: Challenges and opportunities for organisations. *Data Science Journal, 19*(1), 1-15.

Intel Corporation. (2018). *The implications of dark data for business operations and strategy*. Intel White Paper.

Jackson, T. W., & Hodgkinson, I. R. (2022). Keeping a lower profile: How firms can reduce their digital carbon footprints. *Journal of Business Strategy, 44*(6),363-370. https://doi.org/10.1108/jbs-03-2022-0048.

Kambies, K. (2017). Searchability and accessibility issues in dark data management. *Journal of Information Technology, 34*(4), 345-359.

Kim, H. (2024). Dark data in real-world evidence: Challenges, implications, and the imperative of data literacy in medical research. J*ournal of Korean Medical Science, 39*(9). https://doi.org/10.3346/jkms.2024.39.e92.

Krüger, J., & Marshall, P. (2017). Data exhaust: Understanding the implications of unutilised information in organisations. *Journal of Data Science, 15*(2), 123-132.

Letafati, M., & Otoum, S. (2023, September). Global differential privacy for distributed metaverse healthcare systems. In *2023 International Conference on Intelligent Metaverse Technologies & Applications (iMETA)* (pp. 01-08). IEEE.

Lugmayr, A. (2017). The unknown existence of dark data: Challenges for organisations in managing unstructured information. *Journal of Knowledge Management, 21*(3), 456-471.

Martin, J. (2016). *The costs associated with dark data: An overview for organisations*. Harvard Business Review.

Moher, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Med, 6*(7).

Moumeni, L., Slimani, I., El Farissi, I., Saber, M., & Belkasmi, M. G. (2021). Dark data as a new challenge to improve business performances: Review and perspectives. *2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)*, 216-220. https://doi.org/10.1109/icdata52997.2021.00049

Munot, K., Mehta, N., Mishra, S., & Khanna, B .(2019). Importance of dark data and its applications*. IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, (pp. 1-6).doi: 10.1109/ICSCAN.2019.8878789.

Neff, G. (2018). Data as a resource: The implications of dark data on organisational strategy and decision-making processes. *Information Society, 34*(5), 345-359.

O'Shea, C. (2008). Dark data and its impact on business intelligence: An emerging concern for organisations today?.*Business Intelligence Journal, 13*(2), 22-30.

Schembera, B., & Durán, J. M. (2019). Dark data as the new challenge for big data science and the introduction of the scientific data officer. Philosophy & Technology, 33(1), 93-115. https://doi.org/10.1007/s13347-019-00346-x

Schembera, B., & Durán, 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.

Shukla, R.(2015). Dark data in the digital age: Implications for compliance and security management.*Journal of Cybersecurity and Privacy, 1*(2), 45-60.

Trajanov, D.(2018). Hoarding or harnessing? The dual nature of dark data in organisational contexts.*Information Systems Frontiers, 20*(4), 789-804.

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

Zhou, K. (2020). A framework for image dark data assessment. *World Wide Web*, *23*(3), 2079-2105. https://doi.org/10.1007/s11280-020-.