Exposing the Organisational Damage Caused by Dark Data and how to Prevent it

Graham Chant
Stewart Barr and Associates, Adelaide, Australia

graham.chant@sb-a.com.au

Abstract: Dark data refers to data that organisations routinely collect, process, and store but do not use for decision-making, analysis, or value creation. Although often overlooked, dark data has become a significant organisational concern due to the volume, velocity, and heterogeneity of contemporary information ecosystems. Its presence creates four interrelated clusters of organisational problems. First, it undermines strategic and operational decision-making by obscuring potentially valuable insights and distorting evidence bases. Second, it creates operational inefficiencies, including duplicated data, uncontrolled storage growth, and technical debt. Third, dark data introduces compliance, security, and ethical risks because organisations may unknowingly retain sensitive, regulated, or high-risk information without adequate oversight. Finally, it reflects and reinforces organisational and cultural barriers, such as weak stewardship, siloed practices, and low data literacy, which prevent coordinated and responsible data use. This study investigates how these problems can be reduced or minimised through coordinated operational activities rather than isolated, technology-centric interventions. A comprehensive literature review methodology was employed, using structured search, screening, and thematic synthesis across peer-reviewed publications and scholarly frameworks in data governance, information management, organisational risk, and knowledge management. This approach ensured systematic coverage of relevant domains and enabled the integration of theoretical and empirical perspectives. The review findings demonstrate that dark data cannot be eliminated entirely, but its organisational impact can be substantially mitigated by four categories of coordinated action: (1) data discovery and cataloguing to improve visibility and accountability; (2) lifecycle control and data-quality management to reduce redundancy and unmanaged growth; (3) organisational enablement to strengthen stewardship, literacy, and incentives; and (4) governance-by-design architectural practices that embed rules, controls, and monitoring directly into data platforms and pipelines. Together, these activities provide a coherent framework for reducing dark data and structuring future empirical research on dark data governance.

Keywords: Dark data, Cataloguing, Lifecycle management, Governance

1. Introduction

Organisations increasingly rely on data to support decision-making, innovation, and operational efficiency. Yet, despite significant investments in data platforms and analytics capabilities, a substantial proportion of organisational data remains unused, poorly understood, or effectively invisible. This phenomenon is commonly described as dark data, information that is collected, processed, and stored during routine operations but not meaningfully exploited for organisational purposes (Hand, 2020; Martin, 2016; Gartner Inc, 2013). Estimates suggest that 80–90% of organisational data is dark, much of it unstructured, siloed, or unknown (Duarte, 2026; Eryurek et al., 2021). While dark data is often discussed as a missed opportunity for value creation, a growing body of literature suggests that it also creates significant and multifaceted organisational damage.

Dark data is data that are collected, processed, and stored but not used. They can be conceptualised as latent organisational knowledge that remains unactivated due to gaps in knowledge management (KM) processes and governance. Foundational KM literature argues that knowledge only generates value when it is effectively created, shared, and applied within organisational contexts (Nonaka & Takeuchi, 1995; Alavi & Leidner, 2001). From this perspective, dark data reflects a breakdown across the knowledge lifecycle: data are generated but not contextualised, stored but not discoverable, and available but not mobilised for decision-making. Data governance frameworks further reinforce this connection, emphasising that the usability of data depends on metadata, stewardship, quality controls, and lifecycle management (DAMA International, 2017; ISO/IEC 38505-1, 2017). In line with the knowledge-based view of the firm, which positions knowledge as the primary source of competitive advantage (Grant, 1996), dark data represents underutilised knowledge capital that can be transformed into actionable insight through KM practices supported by analytics and artificial intelligence (AI). Accordingly, dark data is both a symptom of weak KM and governance, and a strategic opportunity: when effectively governed and integrated into KM processes, it can enhance decision quality, operational efficiency, and innovation.

Rather than representing a benign surplus, dark data has been shown to generate four interrelated clusters of organisational problems: (1) strategic and decision-making problems, (2) operational and efficiency problems, (3) compliance, security, and ethical risks, and (4) organisational and cultural problems. These impacts extend beyond technical inefficiencies, affecting how organisations make sense of their environment, allocate

resources, manage risk, and coordinate collective action. Understanding these problem clusters is therefore a prerequisite for developing effective prevention and mitigation strategies.

Building on the identification of four categories of governance action, this study conceptualises these activities as a coordinated governance capability and examines their combined effect on mitigating the organisational damage caused by dark data. To examine how dark data generates organisational damage and how it can be mitigated, this study is guided by the following research question. How does an integrated data governance capability, combining discovery, lifecycle management, organisational enablement, and governance-by-design architecture, reduce the strategic, operational, compliance, and cultural impacts of dark data?

1.1 Strategic and Decision-making Problems

Dark data undermines organisational decision-making by obscuring potentially valuable insights and distorting evidence bases. Hand (2020) argues that unknown or unused data creates systematic blind spots, leading organisations to make strategic decisions based on partial, biased, or incomplete information. In this sense, dark data does not merely represent unused assets but actively shapes what organisations can and cannot know.

Research on data-driven strategy further demonstrates that selective visibility of data influences managerial sense-making and strategic direction. Constantiou and Kallinikos (2015) show that when only certain datasets are visible or trusted, organisational strategies become anchored to what is measurable rather than what is meaningful. Empirical studies of dark data reinforce this point, linking unused or poorly governed data to missed opportunities for innovation, learning, and competitive advantage (Marumolwa & Marnewick, 2025; Smuts & Van der Merwe, 2025). As a result, dark data contributes directly to weakened strategic foresight and degraded decision quality.

1.2 Operational and Efficiency Problems

Beyond strategic concerns, dark data generates substantial operational inefficiencies. Unused and unmanaged data increases storage costs, complicates system maintenance, and contributes to technical debt through duplication and uncontrolled growth. George (2023) demonstrates that dark data imposes tangible operational burdens, including unnecessary infrastructure expenditure and increased complexity in data processing environments.

Studies in data and information governance highlight that these inefficiencies are not simply technical failures but consequences of weak lifecycle control and governance practices. Khatri and Brown (2010) argue that without clearly defined decision rights and accountability, data assets proliferate without coordination, leading to redundancy and rework. Similarly, Marumolwa and Marnewick (2025) identify operational inefficiency as a recurring outcome of dark data accumulation, noting that organisations often expend resources managing data that delivers no corresponding value. In this way, dark data erodes operational performance while remaining largely invisible in traditional efficiency metrics.

1.3 Compliance, Security, and Ethical Risks

Dark data also creates significant compliance, security, and ethical risks because organisations may not know what data they hold, where it resides, or how it is used. Hand (2020) highlights that unknown data assets frequently include personal, sensitive, or regulated information, exposing organisations to legal and regulatory consequences. When data is not catalogued or classified, organisations struggle to meet retention, deletion, audit, and disclosure obligations.

Information security governance research reinforces this concern. Von Solms and von Solms (2006) demonstrate that unmanaged information assets expand the organisational attack surface and undermine effective security control cycles. From a policy perspective, the OECD (2022) emphasises that trust in data-driven systems depends on stewardship, visibility, and control, conditions that are fundamentally incompatible with large volumes of dark data. Consequently, dark data represents not only unrealised value but also latent organisational liability, particularly in environments subject to privacy, ethical, and security constraints.

1.4 Organisational and Cultural Problems

Finally, dark data reflects deeper organisational and cultural dysfunctions. Research consistently shows that weak stewardship, siloed practices, and low data literacy contribute to the persistence of unused and unmanaged data. Benfeldt, Persson & Madsen (2019) frame data governance as a collective-action problem,

demonstrating that data remains poorly governed when responsibilities are unclear and incentives are misaligned.

Ongena et al. (2023) further show that data literacy is a critical organisational capability influencing whether data is meaningfully interpreted, shared, and reused. When employees lack the skills, authority, or motivation to engage with data governance processes, dark data accumulates by default. In this sense, dark data is not merely a technical artefact but a manifestation of organisational culture, reflecting how responsibility, knowledge, and accountability are distributed across the organisation (Smuts & Van der Merwe, 2025).

1.5 Research Purpose

Taken together, these four problem clusters demonstrate that dark data causes widespread organisational damage that extends well beyond missed analytical opportunities. This paper examines how dark data produces strategic, operational, compliance-related, and cultural harm, and synthesises existing research to identify coordinated governance practices capable of preventing or mitigating these effects. By reframing dark data as a systemic organisational problem rather than a purely technical issue, the study aims to advance both theoretical understanding and practical approaches to dark data prevention.

The remainder of the paper is structured as follows. Section 2 describes the methodology that was adopted for this paper, section 3 provides an analysis of the results and discussion, and section 4 concludes with implications and directions for future research.

2. Methodology

This study adopted a comprehensive, structured literature review methodology to identify, synthesise, and integrate scholarly work on dark data, data governance mechanisms, and organisational risk. The objective of the review was to develop a coherent conceptual framework that explains: (1) the organisational problems created by dark data, and (2) the coordinated governance activities that reduce or mitigate these problems. A structured evidence synthesis approach was selected to address the fragmented and multi-disciplinary nature of dark data research. Dark data spans information systems, organisational studies, data governance, security, and data architecture. Accordingly, multiple academic databases were used to ensure coverage of conceptual, empirical, and technical contributions.

Searches were conducted across Science Direct database, Web of Science, Scopus, ProQuest and Google Scholar. Key words found in the title and or abstract of published articles were included in the inquiry. A combination of Boolean operators and controlled vocabulary was used. Examples include:

- “dark data” AND governance
- “unused data” OR “hidden data” OR “orphaned data”
- “data lifecycle management” AND organisation*
- “data stewardship” AND governance
- “data visibility” AND risk
- “data contracts” AND architecture

Searches were limited to English-language publications from the last 16 years.

Studies were included if they:

- addressed dark data, unused data, or ungoverned information
- examined data governance, information governance, or governance-by-design
- discussed organisational, strategic, operational, or compliance impacts
- were peer-reviewed or recognised scholarly works
- contained conceptual, empirical, or technical contributions

Studies were excluded if they:

- were purely commercial or vendor-promotional
- focused exclusively on algorithms without organisational context
- lacked methodological clarity or theoretical contribution
- were duplicates or non-scholarly commentary

The initial search returned a broad set of records. Duplicates were removed. Titles and abstracts were screened for relevance using the inclusion criteria. Full-text review followed for all potentially relevant papers.

During full-text analysis, papers were assessed for conceptual depth, methodological clarity, and relevance to dark data and governance.

A thematic synthesis approach was used to extract findings. Extracted data included:

- definitions and conceptualisations of dark data
- descriptions of organisational problems
- governance mechanisms proposed
- technical and architectural enablers
- empirical findings on impacts or effectiveness

Codes were grouped into higher-order themes, resulting in:

Four problem clusters.

- Strategic & decision-making problems
- Operational & efficiency problems
- Compliance, security & ethical risks
- Organisational & cultural risks

Four governance response clusters.

- Discovery & cataloguing
- Lifecycle discipline
- Organisational enablement
- Governance-by-design architecture

At each stage of data collection and extraction, the review process focused on systematically examining the extant literature within the relevant thematic and disciplinary domains. The analysis sought to identify the dominant theoretical perspectives, key constructs, empirical methods, and contextual settings represented in prior studies (Paul and Criado, 2020). Attention was given to areas of conceptual fragmentation, methodological limitation, and underexplored research questions. By synthesising these patterns across the literature, the review aimed to identify substantive research gaps and to develop a structured agenda for future research grounded in those gaps. The research method used is illustrated in Figure 1.

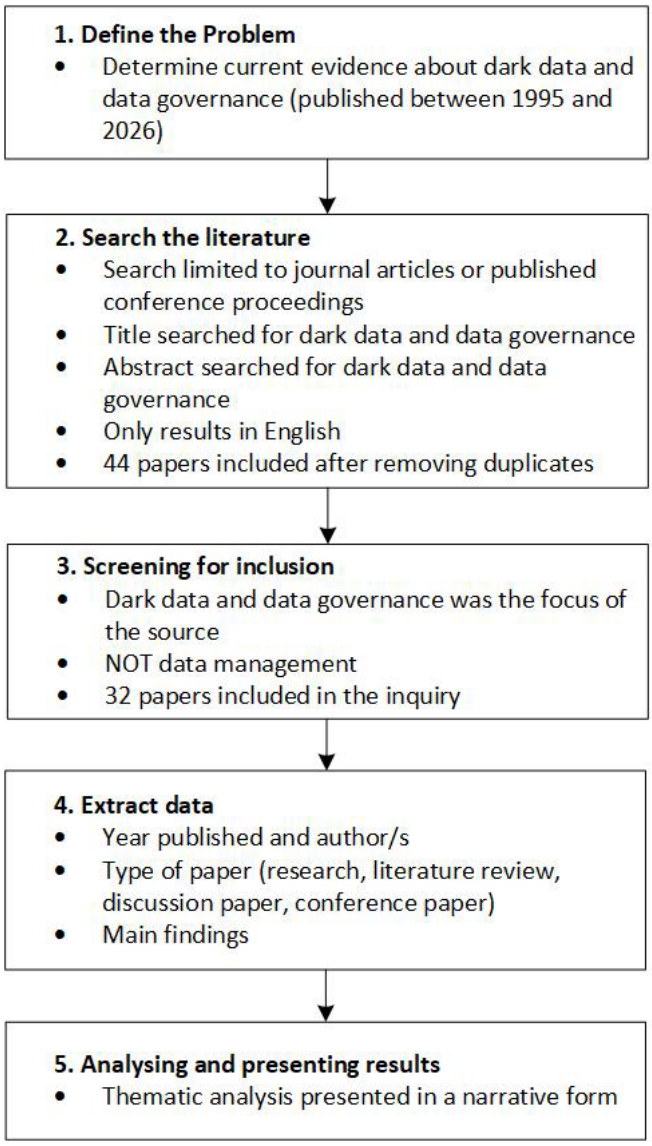

Figure 1: Flow chart of Research Method

3. Results and Discussion

The literature review demonstrates that dark data cannot be eliminated entirely. Digital operations inevitably generate residual and context-specific data. However, the organisational damage caused by dark data can be substantially mitigated through four coordinated categories of action: discovery and cataloguing, lifecycle and data-quality control, organisational enablement, and governance-by-design architecture. Within each category, specific operational activities translate governance principles into practice. These activities move dark data reduction from abstract policy to actionable implementation. These mechanisms function collectively as an integrated governance capability rather than as isolated interventions.

3.1 Data Discovery and Cataloguing: Improving Visibility and Accountability

Dark data persists when organisations lack systematic visibility into their data assets. Discovery and cataloguing therefore represent the first operational layer of mitigation.

3.1.1 Implement enterprise data catalogues and metadata management

A foundational requirement for reducing dark data is systematic data discovery and cataloguing. Dark data persists primarily because organisations lack visibility into what data exists, where it resides, and who is responsible for it. Hand (2020) argues that data cannot generate value, or be governed, if it remains unknown. Similarly, empirical research by Marumolwa and Marnewick (2025) identifies fragmented storage and weak metadata discipline as primary drivers of dark data accumulation. Metadata includes the date and time stamps

of when the data was created and modified, a string containing the full directory and ownership, and descriptive information on the content of the data (Schembera, 2019).

Architectural research further demonstrates that discovery depends on structural consolidation. Harby and Zulkernine (2025) show that unified data lakehouse environments enhance metadata completeness and lineage tracking, reducing the fragmentation that enables dark data to proliferate. Together, these studies indicate that discovery and cataloguing are not optional enhancements but foundational governance mechanisms for improving accountability and preventing informational opacity.

Implementing enterprise data catalogues formalises visibility by linking technical assets to business meaning. Metadata management, ownership attribution, and glossary alignment are widely recognised as foundational governance capabilities (Alhassan, Sammon, & Daly, 2016; Otto, 2011). Without metadata standards and asset registries, data becomes orphaned and opaque. Harby and Zulkernine (2025) further show that unified data lakehouse environments improve metadata consistency and lineage tracking, reinforcing that catalogues must be embedded within platform architecture.

3.1.2 Conduct data inventories, lineage mapping, and system audits

Data governance scholarship reinforces that visibility requires formalised structures. Khatri and Brown (2010) emphasise the importance of defined decision rights and accountability over data assets, while Alhassan, Sammon, and Daly (2016) identify metadata management and asset inventory as core governance activities.

Comprehensive data inventories and lineage mapping enhance traceability and accountability. Khatri and Brown (2010) emphasise the importance of clearly defined data ownership and decision rights, which depend on asset identification. Lineage tracking supports auditability and quality assurance, particularly where multiple transformations occur. Without catalogues, lineage tracking, and ownership assignment, datasets become orphaned and eventually invisible.

From a risk perspective, von Solms and von Solms (2006) argue that effective governance requires visibility into information flows to maintain control cycles. Periodic system audits ensure that previously catalogued assets remain governed rather than drifting into obscurity.

3.1.3 Apply automated discovery techniques (machine learning (ML) classification and natural language processing (NLP) for unstructured content)

Manual discovery processes struggle to scale in environments rich in unstructured data. Automated classification, entity extraction, and metadata inference using ML and NLP enhance scalability and consistency. Bhoite (2025) demonstrates how AI can assist in automating schema enforcement and data contract generation, illustrating the shift toward intelligent governance tooling. Such techniques enable organisations to surface previously hidden datasets and assess their sensitivity or utility.

Together, these activities operationalise visibility and transform dark data from “unknown” to “governable.”

3.2 Lifecycle Control and Data-quality Management: Reducing Redundancy and Unmanaged Growth

Discovery alone does not prevent dark data accumulation. Lifecycle discipline ensures that data remains controlled, relevant, and proportionate.

3.2.1 Establish rules for creation, validation, retention, archival, and disposal

Beyond discovery, effective mitigation of dark data requires disciplined lifecycle management. Dark data accumulates when organisations lack consistent processes governing data creation, validation, retention, archival, and disposal. George (2023) demonstrates that unmanaged data growth produces measurable operational costs, while Marumolwa and Marnewick (2025) identify uncontrolled accumulation as a recurring pattern across organisations.

Lifecycle governance requires explicit policies governing how data enters, moves through, and exits the organisation. Otto (2011) frames lifecycle management as central to data governance maturity. Otto (2011) conceptualises data governance as requiring formalised policies that regulate data definition, quality, and use across its lifecycle. Khatri and Brown (2010) similarly argue that governance must allocate decision authority over data standards and lifecycle rules to prevent duplication and decay. Ghaffari Heshajin et al. (2024) demonstrate in a governance review that classification, retention scheduling, and disposal controls are core components of information governance frameworks.

3.2.2 Reduce redundant, obsolete, or trivial (ROT) data

Redundant obsolete and trivial (ROT) data is data identified as redundant, or duplicate data, obsolete, no longer having business value, and trivial data with little or no business value (Dimitrov, 2018). ROT data needs to be proactively minimized by securely deleting it on a regular basis (Dimitrov, 2018). Uncontrolled data accumulation increases operational complexity and cost. George (2023) identifies dark data as a source of infrastructure inefficiency, while Marumolwa and Marnewick (2025) link unmanaged growth to governance failure. Systematic ROT reduction programmes operationalise lifecycle rules and limit technical debt.

From a risk perspective, von Solms and von Solms (2006) show that unmanaged information assets weaken organisational control cycles and expand exposure to security and compliance threats. Retention, classification, and disposal practices are therefore not merely operational housekeeping tasks; they are governance controls that limit the accumulation of redundant, obsolete, or unmanaged data.

3.2.3 Introduce versioning and provenance tracking

Version control and provenance metadata ensure that data remains traceable and trustworthy. Without systematic provenance tracking, datasets degrade in reliability and reusability. Wang et al (2015) describe provenance as the lineage and processing history of a data product, and the record of the processes that led to it. Von Solms and von Solms (2006) reinforce that traceability is central to maintaining control cycles, while governance scholarship consistently links provenance to quality assurance and auditability (Alhassan et al., 2016).

Lifecycle and quality management reduce dark data by preventing uncontrolled growth and ensuring that data remains fit for purpose throughout its existence.

3.3 Organisational Enablement: Strengthening Stewardship, Literacy, and Incentives

Governance mechanisms require human capability and organisational alignment.

3.3.1 Build data literacy across business units

Technical controls alone are insufficient to prevent the persistence or reaccumulation of dark data. Organisational capability and culture play a decisive role. Benfeldt, Persson, and Madsen (2019) conceptualise data governance as a collective-action problem, showing that unclear roles and misaligned incentives lead to data hoarding and disengagement from stewardship responsibilities.

Data literacy influences whether data is interpreted, reused, and governed responsibly. Ongena, van de Wetering, and van Offenbeek (2023) provide empirical evidence that data literacy is a measurable organisational competence influencing data interpretation, performance and responsible use. Without literacy, governance tools remain underutilised and dark data persists through disengagement.

3.3.2 Align incentives toward shared data use and stewardship

Benfeldt, Persson, and Madsen (2019) conceptualise data governance as a collective-action problem: without aligned incentives and clear stewardship roles, data remains siloed. Aligning incentives reduces hoarding behaviour and improves shared accountability.

Smuts and Van der Merwe (2025) further position dark data as a knowledge management failure, arguing that organisations often lack the cultural mechanisms necessary to convert available information into actionable insight. Without clear stewardship roles, aligned incentives, and embedded literacy practices, data governance initiatives struggle to translate policy into practice.

3.3.3 Develop capabilities for analytics, AI, and insight generation

Smuts and Van der Merwe (2025) position dark data as a knowledge-management failure. Developing analytics and AI capability encourages purposeful reuse and discourages data neglect. Insight generation refers to the process through which analytical methods transform raw data and analytical results into integrated understanding that supports informed decision-making (Tan & Chan, 2016; Sarker, 2021). Where insight generation becomes embedded in organisational practice, unused data is more likely to be surfaced and evaluated.

Organisational enablement therefore acts as a stabilising mechanism, ensuring that discovery and lifecycle controls are enacted consistently rather than circumvented.

3.4 Governance-by-design Architecture: Embedding Rules into Platforms and Pipelines

The final domain shifts governance from reactive oversight to embedded prevention.

3.4.1 Move towards unified data platforms (data lakes / lakehouses)

The final category of coordinated action involves embedding governance directly within technological architectures. Traditional governance approaches have often relied on retrospective audits, manual reviews, and policy compliance processes. However, research in data governance increasingly demonstrates that governance mechanisms are more effective when implemented through automated controls and embedded within data platforms and system architectures. By operationalising governance rules directly in data pipelines, metadata systems, and access controls, organisations can shift from reactive oversight to preventive governance-by-design (Otto, 2011; Abraham et al., 2019; Gupta, Wallace, & Baskerville, 2018).

Harby and Zulkernine (2025) show that lakehouse architectures integrate metadata, governance, and storage layers, reducing fragmentation. Architectural consolidation improves visibility and control, limiting the structural conditions that allow dark data to accumulate.

3.4.2 Embed governance directly into pipelines (data contracts, policy-as-code)

Harby and Zulkernine (2025) show that unified data architectures improve governance enforceability by consolidating metadata, access control, and lineage mechanisms. Bhoite (2025) advances this further by demonstrating how AI-driven data contracts can enforce schema validation, quality thresholds, and usage rules at ingestion, preventing non-compliant or unmanaged data from entering the system.

Bhoite (2025) demonstrates how AI-driven data contracts enforce schema, quality, and usage constraints at ingestion. Embedding validation and compliance checks directly into pipelines operationalises governance at scale and reduces reliance on manual audits.

3.4.3 Use AI-assisted tools for hidden pattern detection and data utility assessment

From a governance theory perspective, von Solms and von Solms (2006) emphasise continuous control cycles, suggesting that effective oversight must be embedded into operational systems rather than applied sporadically. Governance-by-design shifts dark data mitigation from reactive clean-up to preventative enforcement.

AI-assisted monitoring tools enable continuous assessment of dataset usage, duplication, and sensitivity. By detecting unused or anomalous assets, organisations can prioritise remediation efforts. This approach aligns with the continuous control cycle described by von Solms and von Solms (2006).

Governance-by-design architecture ensures that dark data mitigation is preventative rather than reactive.

By embedding validation, monitoring, and compliance rules directly into pipelines and platforms, organisations reduce reliance on manual intervention and limit the re-emergence of unmanaged datasets.

4. Conclusion and Future Research

This study is based on a comprehensive literature review, which synthesises theoretical, empirical, and technical contributions across data governance, information management, and organisational risk. As with any literature-based analysis, several limitations should be acknowledged. First, the review is constrained by the availability of published research; empirical studies on dark data practices remain comparatively scarce, and many practical implementations are documented only in industry reports rather than peer-reviewed literature. Second, the evidence concerning governance-by-design mechanisms, such as data contracts and automated controls, is still emerging, with limited longitudinal analysis of their effectiveness. Third, organisational and cultural dimensions of dark data remain difficult to measure, and existing studies rely heavily on self-reported behaviours rather than observed practice.

This study examined how an integrated data governance capability, encompassing discovery, lifecycle control, organisational enablement, and governance-by-design architecture, mitigates the organisational damage caused by dark data. The findings confirm that dark data generates consistent harms across strategic, operational, compliance, and cultural domains, affecting decision quality, efficiency, risk exposure, and organisational data practices.

The results show that mitigation does not occur through isolated interventions but through the coordinated operation of governance activities. Discovery and cataloguing improve visibility and accountability; lifecycle control reduces redundancy and unmanaged growth; organisational enablement strengthens stewardship and data literacy; and governance-by-design embeds rules into platforms and pipelines, enabling preventative control. These mechanisms operate as an integrated capability, with mutually reinforcing effects. Visibility enables control, control limits reaccumulation, enablement sustains adoption, and architectural embedding ensures consistency. Where any dimension is weak, dark data persists.

This study contributes an integrated framework linking dark data to organisational damage and governance response. It demonstrates that while dark data cannot be eliminated, its impact can be substantially mitigated through coordinated governance capability.

In doing so, the findings directly answer the research question by showing that integrated governance capability reduces dark data impacts through the combined and reinforcing effects of visibility, control, enablement, and embedded architectural enforcement.

Future research should pursue empirical validation of the framework presented in this study. Case studies and ethnographic approaches could illuminate how dark data arises within workflows, how governance processes are enacted in practice, and how organisational culture shapes the success or failure of interventions. Quantitative studies could develop measurement instruments for dark data extent, governance capability, and organisational outcomes, supporting the development of structural models. Technical research could further evaluate the performance, limitations, and adoption barriers of automated governance mechanisms such as AI-driven data contracts and pipeline enforcement tools. Finally, future work should explore how governance maturity interacts with data architecture, organisational structures, and regulatory environments to influence dark data accumulation and mitigation.

Ethics Declaration: I did not need an ethical clearance for the research referred to in this paper.

AI Declaration: I did not use an AI tool in the development of this paper.

References

Abraham, R., Schneider, J. and vom Brocke, J. (2019) “Data governance: A conceptual framework, structured review, and research agenda”. *International Journal of Information Management*, 49, 424–438.

Alavi, M., & Leidner, D. E. (2001) “Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues”. MIS Quarterly, 25(1), 107–136. https://doi.org/10.2307/3250961

Alhassan, I., Sammon, D. and Daly, M. (2016) “Data governance activities: an analysis of the literature”*, Journal of Decision Systems*, 25:sup1, 64-75, DOI:10.1080/12460125.2016.1187397.

Benfeldt, O., Persson, J.S. and Madsen, S. (2019) “Data Governance as a Collective Action Problem”, *Information Systems Frontiers*, 22(2), 299-313. https://doi.org/10.1007/s10796-019-09923-z.

Bennett, S. (2017) “What is information governance and how does it differ from data governance?” *Governance Directions, 69(8), 462–467. https://search.informit.org/doi/10.3316/informit.070201793736360.*

Bhoite, H. (2025) *AI-Driven Generation of Data Contracts in Modern Data Engineering Systems.* arXiv:2507.21056v1, https://doi.org/10.36227/techrxiv.174776182.24947749/v1.

Constantiou, I.D. and Kallinikos, J. (2015) “New games, new rules: big data and the changing context of strategy”. *Journal of Information Technology*, 30 (1). pp. 44-57. ISSN 0268-3962.

Dimitrov, W., Сярова, C. and Petkova, L.G. (2018) *Types of dark data and hidden cybersecurity risks*, DOI: 10.13140/RG.2.2.31695.43681.

DAMA International. (2017). DAMA-DMBOK: Data management body of knowledge (2nd ed.), Technics Publications, LLC., Denville, New Jersey, USA

Duarte, F. (2026). “Amount of Data Created Daily (2026)”, *Exploding Topics*, https://explodingtopics.com/blog/data-generated-per-day.

Eryurek, E. et al. (2021) *Data Governance: The Definitive Guide. People, Processes, and Tools to operationalize Data Trustworthiness*, O’Reilly Media Inc., Sebastopol, California, USA.

Gartner Inc. (2013) Innovation Insight: File Analysis Innovation Delivers an Understanding of Unstructured Dark Data, Alan Dayley

George, A.S. et al. (2023) “Bringing Light to Dark Data - A Framework for Unlocking Hidden Business Value”, *Partners Universal International Innovation Journal*, 1(4), 35–60. https://doi.org/10.5281/zenodo.8262384

Ghaffari Heshajin et al. (2024) “A framework for health information governance: a scoping review”, *Health Research Policy and Systems (2024)* 22:109, https://doi.org/10.1186/s12961-024-01193-9.

Grant, R. M. (1996) “Toward a knowledge-based theory of the firm”. Strategic Management Journal, 17(S2), 109–122. https://doi.org/10.1002/smj.4250171110

Gupta, S., Wallace, W. A. and Baskerville, R. (2018) "The effect of automated controls on data governance". *MIS Quarterly Executive*, 17(4), 307–321.

Hand, D.J. (2020) *Dark Data - Why-what-you-don't-know-matters*, Princeton University Press, Princeton, USA.

Harby, A.A. and Zulkernine, F. (2025). "Data Lakehouse: A survey and experimental study", *Information Systems*, Volume 127, 2025, 102460, ISSN 0306-4379, https://doi.org/10.1016/j.is.2024.102460.

Khatri, V. and Brown, C.V. (2010) "Designing data Governance", *Communications of the ACM*, vol. 53 | no. 1, DOI: 10.1145/1629175.1629210.

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

Marumolwa and Marnewick, (2025) "Unveiling-Dark-Data-in-Organisations: Sources-Challenges-and-Mitigation-Strategies", *International Journal of Service Science, Management, Engineering, and Technology*, Volume 16, Issue 1, January-December 2025, DOI: 10.4018/IJSSMET.386167.

Nonaka, I. and Takeuchi, H. (1995) The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, New York.

OECD (2022) "Responding to societal challenges with data: Access, sharing, stewardship and control", *OECD Digital Economy Papers*, No. 342, OECD Publishing, Paris, https://doi.org/10.1787/2182ce9f-en.

Ongena, G., van de Wetering, R. and van Offenbeek, M. (2023) "Data literacy for improving organizational performance: A competence-based approach and multidimensional operationalization". *Government Information Quarterly*, 40(1), 101706. https://doi.org/10.1016/j.giq.2022.101706.

Otto, B. (2011) "Organizing data governance: Findings from the telecommunications industry and consequences for large service providers". *Communications of the Association for Information Systems*, 29, 45–66.

Paul, J. and Criado, A. R. (2020) "The art of writing literature review: What do we know and what do we need to know?" *International business review*, 29, 1010717, https://doi.org/10.1016/j.ibusrev.2020.101717.

Sarker, I.H. (2021) "Data science and analytics: An overview from data-driven smart computing and decision-making". *SN Computer Science*, 2, 377.

Schembera, B. and Durán, J.M. (2019) "Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer". *Philos. Technol.* 33, 93–115 (2020). https://doi.org/10.1007/s13347-019-00346-x.

Smuts, H. and Van der Merwe, A. (2025) "From dark data to insight: The role of knowledge management in promoting digital decarbonisation", *South African Journal of Information Management*, 27(1), a1967, https://doi.org/10.4102/sajim.v27i1.1967.

Tan, S. and Chan, T. (2016) "Defining and conceptualizing actionable insight: A conceptual framework for decision-centric analytics". Proceedings of the Australasian Conference on Information Systems (ACIS). 2015, Adelaide.

von Solms, R. and von Solms, S.H. (2006) "Information Security Governance: A model based on the Direct–Control Cycle". *Computers & Security*, 25(6), 408–412. https://doi.org/10.1016/j.cose.2006.07.005.

Wang, J. et al. (2015) "Big data provenance: Challenges, state of the art and opportunities". 2015 IEEE international conference on big data (Big Data) (pp. 2509-2516), DOI:10.1109/BigData.2015.7364047.