UTS ePRESS

Construction Economics and Building

Vol. 26, No. 2
2026

Citation: Tambwe, O., Aigbavboa, C., Akinradewo, C. 2026. Influence of Data Longevity on Data Management in the South African Construction Industry. *Construction Economics and Building*, 26:2, 1–21. https://doi.org/10.5130/s7p98j97

ISSN 2204-9029 | Published by UTS ePRESS | https://epress.lib.uts.edu.au/journals/index.php/AJCEB

ARTICLES (PEER REVIEWED)

Influence of Data Longevity on Data Management in the South African Construction Industry

Ornella Tambwe[1,*], Clinton Aigbavboa[1], Opeoluwa Akinradewo[2]

[1] CIDB Centre for Excellence & Sustainable Construction Management and Leadership in the Built Environment, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa

[2] Department of Quantity Surveying and Construction Management, University of the Free State, Bloemfontein, South Africa

Corresponding author: Ornella Tambwe, tambweorny@gmail.com

DOI: https://doi.org/10.5130/s7p98j97

Article History: Received 29/03/2025; Revised 16/01/2026; Accepted 20/01/2026; Published 05/06/2026

Abstract

Construction industry stakeholders increasingly rely on social media and information and communication technology tools for communication and business operations. To ensure a smooth construction process and support future needs, such as renovations, it is essential to preserve the information generated throughout a project's life cycle. This study examines the influence of data longevity on data management in the South African construction industry. A quantitative research approach was used, collecting data from construction professionals in Gauteng through an online questionnaire survey. Respondents included quantity surveyors; architects; civil, mechanical, and electrical engineers; and construction and project managers. The findings indicate that key factors influencing data longevity in data management include the use of immutable storage, continuous data protection technologies, durable record-keeping devices, and knowledge of data longevity standards. Moreover, the study emphasises the importance of construction stakeholders developing a thorough understanding of data longevity to ensure effective data management throughout a project's execution. By addressing challenges such as data loss due to obsolescence, cyber threats, and environmental factors, this research contributes to establishing best practices for preserving construction project data, ensuring long-term accessibility and compliance.

DECLARATION OF CONFLICTING INTEREST The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. **FUNDING** The authors thank the Global Excellence and Stature (GES) for financial support.

As the construction industry increasingly relies on digital data, prioritising data longevity is essential for maintaining project integrity, enhancing collaboration, and fostering future innovations. It is recommended that professionals receive training and educate all project members on data longevity principles and their implementation.

Keywords

Construction Industry; Data; Data Longevity; Data Preservation; Data Durability

Introduction

Nowadays, projects are becoming increasingly complex, with growing pressure to complete them within specified timeframes and budgets. This is driven by competitive pressures and rising stakeholder expectations for productivity and innovation (Van Besouw & Bond-Barnard, 2021; Zhang & Wang, 2026). In such a competitive landscape, effective data management is critical because it directly influences and promotes project success (Leppikorpi, 2018). According to M-Files (2019), information is paramount to the successful operation of a construction business. This is because organisations continually strive to enhance their information management practices, as poor document and data management can reduce productivity and incur high costs. Malla et al. (2024) put forward that digitalisation has positively transformed business practices by enabling more efficient and strategic operations. This includes the use of digital tools to manage raw material supply chains, the adoption of smart technologies for construction site management and operations, including data management, and the application of RFID systems to improve the tracking and traceability of construction resources. Pentury et al. (2025) supported the previous statement by stating that technology, as an enabler of digitisation, is also a key factor supporting the Sustainable Development Goals by improving all organisational resource use. However, despite the benefits, the M-Files (2019) report highlighted that 39% of respondents find it challenging and time-consuming to locate the information they need within their systems and repositories. This often results in the frustrating and costly recreation of documents, reducing overall efficiency and costing construction companies millions of dollars. Furthermore, to date, the preservation of heritage buildings remains a challenge due to frequent data file loss and vulnerabilities in the digital capture process (Olorunnishola et al., 2025). According to Ma (2014), Parhami (2019), and Liu et al. (2020), data longevity can ensure data availability, supporting improved accessibility, reliability, and permanence of information. With robust backup systems, data longevity safeguards critical information, enabling organisations to access it whenever needed, even in the long term.

Perera et al. (2020), Khalil & Stravoravdis (2022), and Malla et al. (2024) explained that building information modelling was introduced to overcome communication challenges related to interdisciplinary and interoperability issues among various stakeholder groups. This is particularly important in heritage and historic building projects, which involve a more diverse set of stakeholders across a broader range of disciplines than typical construction projects. A key consideration in these projects is the long-term preservation of digital documentation, as such records are often required to be retained for future reference (Khalil & Stravoravdis, 2022). However, maintaining these records is complicated by the wide variety of software tools used and the rapidly changing software market. Digital documentation stored in current formats may become incompatible with future software developments, posing accessibility challenges. Olorunnishola et al. (2025) noted that, to date, the preservation of heritage buildings remains a challenge due to the frequent loss of data files and vulnerabilities in the digital capture process. Beetz et al. (2013) highlighted the importance of preserving cultural heritage, implementing safety measures, and facilitating the reuse of design and engineering knowledge. The authors also emphasised the need for long-term preservation of information about built environment artefacts to ensure the legal accountability of stakeholders, including designers and engineers. Khalil et al. (2021) further emphasised the importance of data longevity by highlighting the value of old sketches and architectural drawings in cataloguing and surveying heritage structures. These documents can serve as foundational tools for creating a digital twin of a structure, providing insights into the building's original design and highlighting any modifications made over time. Such information is crucial for preserving the history and integrity of heritage buildings. This study assesses the influence of data longevity on data management.

Parhami's (2019) paper discussed data longevity and storage media in the context of the growing volume of digital data. It highlighted the challenges of data decay, format obsolescence, and the need for strategies to ensure data integrity over time. Various storage technologies, their lifespans, and methods for preventing data loss were examined. Rothenberg's (1995; 1999) paper emphasised the risks of media decay, format obsolescence, and hardware/software incompatibility. It warned that valuable historical, scientific, and governmental records may become unreadable over time. The author emphasised the importance of proactive digital preservation strategies to ensure future accessibility. None of these studies has explored data longevity in the South African construction industry. This study addresses that gap by focusing on its unique characteristics. Additionally, it contributes to the body of knowledge by enabling future researchers and industry professionals to build on past findings without unnecessary repetition. In construction, this ensures that historical project records, material specifications, and best practices remain accessible for continuous improvement and reference.

Background of data longevity

Dolgin (2018) and Kansa et al. (2020) explained that the concepts of longevity and ageing are explored across fields such as archaeology and zooarchaeology, as well as in life expectancy. Goh et al. (2019) noted that professionals across sectors are increasingly interested in ageing and longevity. These sectors encompass medicine, research, engineering, biotechnology, and data analytics. This growing interest is driven by the potential to improve organisational performance through better decision-making across various fields, including drug development in medicine. Kansa et al. (2020) highlighted advances in research data management and reproducible research techniques. These advances have created new opportunities for researchers to capture and publish their primary data in various datasets. These datasets must be preserved over time for future research and for future generations. Parhami (2019) emphasised the growing importance of data management, integrity, and preservation, as data become crucial in both social and commercial sectors. This is because physical storage media have limited lifespans and are prone to deterioration. Additionally, data storage formats often become obsolete over time. Therefore, an active strategy to ensure the reliability and longevity of stored data is essential in any data management plan.

In Rothenberg's (1995; 1999) "grandmother scenario", set in 2045, the grandmother's grandchildren discover a 1995 letter and a compact disc read-only memory (CD-ROM) in the attic of her unpurchased home. The letter states that the disc contains information about her unearned money. Excited but unfamiliar with CDs, the grandchildren face the challenge of finding a compatible CD drive and running the necessary software to read the contents. This scenario highlights the uncertainty surrounding access to old digital documents, which are replacing paper records. It raises concerns about the longevity of digital records, as they are more fragile than paper documents (Rothenberg, 1995; 1999). Blum and Beyer (2019) noted that digital permanence is an issue that society must address, as the longevity of data is crucial to avoid data loss for future generations. Parhami (2019) stressed that optical storage media, such as CDs, can deteriorate easily because the substance on which data are stored can degrade over time. To extend the lifespan and reduce data deterioration, these digital optical media must be stored under proper environmental conditions. Shahani et al. (2005) identified several forms of CD deterioration, including spots on the metal layer caused by water condensation, chemical degradation that erases stored data, and the eventual deterioration of the disc's metal layer over time.

Parhami (2019) highlighted that data longevity is closely linked to managing issues such as data decay, storage failure, and media lifespan. Data longevity is a concern across different types of storage media, including CDs, magnetic storage, and flash-based storage (Rothenberg; 1999; Shahani et al., 2005; Choi et al., 2017). Lunt et al. (2012) and Goda and Kitsuregawa (2012) identified three main types of data storage: solid state [including dynamic random access memory (DRAM), static random access memory (SRAM), and flash drive], optical discs [such as digital versatile discs (DVDs) and compact discs (CDs), and Blu-ray)], and magnetic storage (including hard disc drives and tapes). However, none of these storage options, with the notable exception of certain types of DVDs, are designed to store data permanently (Lunt et al., 2012). Lunt et al. (2012, 2013) further noted that computer storage has always been temporary, except for magnetic core storage. Although disc drives were first introduced in the 1960s and have since evolved, they continue to provide short-term data storage, with data lifespans estimated to be 1 to 7 years. Flash memory, the most common non-volatile solid-state storage today, has a data lifespan of 1 to 12 years, depending on memory density. One of the options for long-term recordable data storage is optical discs, which can last between 1 and 25 years, depending on various factors (Lunt et al., 2012, 2013). From Crofts' (2008) perspective, a significant amount of digital data has been lost due to the degradation of the magnetic tapes. Storage media degradation is a crucial issue, but so are rapidly changing storage technologies and evolving file formats. The fast pace of technological change makes it difficult to rely on any specific codec, compression format, or digital format as a stable means of preserving screen heritage, such as movies and motion pictures.

Ma (2014) and Parhami (2019) emphasised that data are always stored on physical media, whether locally or in the cloud. The authors noted that such data are vulnerable to format obsolescence, physical degradation, and electrical deterioration. Therefore, supplementing physical storage with a logical organisation is critical to ensure data longevity and protect its integrity. For instance, Liu et al. (2020) highlighted that large-scale data centres and cloud storage systems often rely on thousands of hard drives for primary data storage. Although a single hard drive rarely fails each year, having multiple hard drives increases the risk of hardware failure. Such failures can result in service interruptions and permanent data loss. To mitigate these issues, many organisations have transitioned to solid-state drives (SSDs), which use flash memory for long-term data storage. However, the higher density of flash memory has reduced chip-level reliability (Meza et al., 2015). As a result, SSD failures in data centres can cause downtime and, in some cases, permanent data loss.

Kansa et al. (2020) pointed out that data entry is often optimised for speed and consistency using codes. If a code sheet becomes separated from the dataset, the risk of data loss increases, which can shorten the lifespan of the data. Moreover, many researchers use custom alphanumeric codes for their projects, which are often modified or expanded over time. These changes can cause the dataset to deviate from the original code sheet, posing an additional challenge to the longevity of the data. To ensure data longevity, storage longevity must also be considered, as storage systems are the containers where data reside (Lunt et al., 2013; Meza et al., 2015; Liu et al., 2020). As more valuable information is stored digitally in the 21st century, it is increasingly crucial to design storage systems that can maintain data durability for decades across all sectors, including the construction industry. Khalil and Stravoravdis (2022) and Tanga et al. (2022a) highlighted that construction projects involve numerous temporary teams that disband once a project is completed, thereby transferring data responsibility to different individuals across different organisations. Additionally, because various organisations use diverse systems and follow different standards, and because assets last for decades, maintaining data longevity is complex, challenging the continuity and accessibility of information about the asset life cycle. Data longevity in the construction industry is broadly understood through data governance theory, which focuses on organisational structures, stewardship roles, data quality controls, and accountability mechanisms that regulate how information is created, stored, protected, used, and reused across its life cycle (DAMA International, 2017; Abraham et al., 2019; Tanga et al., 2022b).

In other words, data governance focuses on how choices are made regarding data and how processes and individuals must behave in relation to it (DAMA International, 2017). It can reduce risks such as data inconsistency and misuse, which are major threats to long-term data preservation. In this way, it transforms how data are managed, turning a purely technical challenge into an organisational capability that sustains data value beyond the lifespan of individual projects. Information life cycle management frameworks further complement data governance theory by providing a systematic approach to managing project data from inception through to long-term storage and retention. This includes structured retention management that preserves data in line with legal, regulatory, and business requirements. It enables organisations to set compliant retention periods, keep data accessible for audits and operational needs, and securely dispose of data when they are no longer required (El Arass& Souissi, 2018; Sharma & Vaid, 2024). The authors added that legal holds also prevent the deletion of information needed for ongoing cases, protecting critical records. Secure storage and rule-based controls further safeguard data from premature loss or unauthorised access.

Other pillars of data longevity extend beyond data storage by guiding governance of data and assessing risk management, as well as organisational practices. This encompasses standards such as the Open Archival Information System (OAIS), CoreTrustSeal, ISO 31000, and ISO 16363 (Frank, 2020; Lin et al., 2020). The OAIS provides a widely recognised reference model that guides the long-term preservation and accessibility of digital information (Lee, 2010). Building on this foundation, the CoreTrustSeal focuses on certifying trustworthy data repositories by promoting reliability, integrity, and sustainable data stewardship practices. ISO 16363 extends the OAIS framework by offering measurable criteria for assessing and auditing the trustworthiness of digital repositories (Frank, 2022). Complementing these preservation-focused standards, ISO 31000 addresses risk management by enabling organisations to systematically identify, assess, and mitigate threats that may compromise long-term data preservation (Frank, 2020; Lin et al., 2020). Malla et al. (2024) added that data longevity in the construction industry depends on ISO 19650, which sets clear standards for the consistent production, organisation, and management of information throughout the construction project life cycle. These standards encourage the organised use of a common data environment, ensuring that data are appropriately classified, versioned, and safely stored, thereby making traceability and retrieval over time easier (Malla et al., 2024; Ye et al., 2024).

TECHNOLOGIES AND METHODS FOR DATA LONGEVITY

The preservation of material in digital form has advanced significantly, and many practitioners worldwide have access to established techniques, frameworks, standards, and software (Beetz et al., 2013). The various data longevity technologies and methods include the following.

THE USE OF MILLENNIA DISCS AND DISASTER MANAGEMENT SERVICES

To achieve data longevity, construction firms should combine high-quality storage media with effective disaster recovery and management practices. Archival-grade solutions such as Millennia Discs (M-Discs) are designed to preserve data for 1,000 years by permanently etching information into the recording layer, making them more resilient than conventional optical discs (Lunt et al., 2012; 2013; Parhami, 2019). Nonetheless, this type of storage alone is insufficient, as digital data remain vulnerable to both physical and cyber-related risks. Integrating disaster management services complements data longevity efforts by reducing risk; strengthening preparedness, response, and recovery planning; and ensuring data availability (Sakurai & Murayama, 2019). Furthermore, adopting information and communication technology (ICT)-enabled disaster planning frameworks helps organisations coordinate information, protect data assets, and recover critical records following disruptive events, thereby strengthening overall data longevity (Frank, 2020).

REGULAR DATA BACKUP AND OPTIONAL DATA INSURANCE COVERAGE

Data backup involves copying data to a separate device, system, or service to prevent permanent data loss and enable recovery in the event of issues such as system failure or calamities (Khatri & Brown, 2010; Filecoin, 2022). This is critical for ensuring data longevity, as a backup allows lost data to be recovered, thereby maintaining its permanence for years, depending on the lifespan of the chosen backup storage medium. For example, on-site disc-based backup systems offer faster backup and recovery, use more reliable media than tapes, and isolate client faults by storing multiple versions of data. However, these systems remain vulnerable to hardware, software, and administrative failures (Kotla et al., 2007). Bhalerao and Pawar (2017) and Feng (2022) highlighted that cloud storage has emerged as a popular solution for data backup due to its scalability and availability. Nevertheless, the large data volumes requiring backup in cloud environments often create performance and storage capacity challenges. These issues are mitigated through techniques such as deduplication, which eliminates redundant data within the same backup system. To further protect data, including those stored in the cloud, Ma (2014) proposed an optional personal data insurance coverage. In this model, users pay a monthly or yearly insurance premium based on the declared value of their data. If a data item is deemed lost and unrecoverable by the cloud storage provider, the data owner can file a claim. Once the loss is verified, the cloud service provider (acting as the insurance carrier) compensates the user (stakeholder) at the insured value. This approach provides users with a form of risk management against data loss in construction projects, ultimately supporting the longevity of data.

AWARENESS OF DATA LONGEVITY STANDARDS AND BREACH OF DUTY PUNISHMENT

A significant challenge in achieving long-term data storage is the lack of clear standards for classifying the vast array of digital records (Boyda, 2013). As new standards and criteria for evaluating diverse types of digital evidence are introduced, archivists gain access to more tools and resources to help decide what should be preserved and what can be discarded. This restores the critical decision-making role of archivists in the preservation process. According to Digitale Bewaring (2003) and Singh et al. (2020), it is essential for employees to be educated about the specific standards adopted by their organisation and to understand their responsibilities regarding data preservation. Negligence in fulfilling these responsibilities is a serious issue. If an employee breaches their duty of care for data, the situation will be assessed based on whether their actions were considered reasonable at the time. Courts or organisational leadership will review evidence, including the employee's actions, witness statements, and technological records (such as camera footage), to determine whether the issue was due to an honest mistake or negligence. If found guilty of negligence, the employee may face legal consequences (Singh et al., 2020). This accountability encourages employees to handle data responsibly and adhere to security measures designed to support data longevity.

DATA SECURITY MEASURE ADOPTION AND CONTINUOUS USAGE OF DATA SECURITY TECHNOLOGIES

Chandrasena (2022) explained that there are serious concerns regarding the security of systems, particularly in the context of the growing volume of global online business transactions. These concerns stem from harmful vulnerabilities within such systems, which pose a serious risk to businesses. Despite the potential benefits of conducting transactions online, some businesses have been slow to adopt new technologies due to fears of losing investments and compromising confidentiality. Additionally, security flaws are a significant deterrent for both individuals and businesses when engaging with e-commerce. For instance, a single widely publicised security flaw can damage a company's reputation, have a significant impact on the e-commerce industry, and jeopardise its long-term viability (Chandrasena, 2022; Tanga et al., 2022a). Therefore, establishing a continuous data security system will enable businesses to conduct their day-today transactions with ease and adopt various technologies with a minimised fear of being attacked by cybercriminals. Various security tools include encryption, intrusion detection, anti-malware and antivirus scanners, regular updates, authentication, and artificial intelligence (Efimova et al., 2021; Zhang & Wang, 2026).

RESOURCE AVAILABILITY AND EDUCATION AND TRAINING OF EMPLOYEES

Lin et al. (2020) stressed that as ICT becomes more widespread, reliance on electronic data and the systems that facilitate access to it has increased. For these systems to be effective, they must gain the trust of users by demonstrating that they can manage data responsibly and securely. However, achieving data longevity with these ICT tools is challenging due to the growing threat of cybercrimes (Tambwe et al., 2023). Digitale Bewaring (2003), Boyda (2013), and Shaban et al. (2022) argued that governments and businesses must recognise the importance of digital longevity and take proactive steps to address it. This includes allocating funds, developing smart systems for data longevity, and creating policies, rules, and processes to safeguard data. Public awareness, investment in technical resources, cultural shifts, and ongoing education and training are also essential. For this to be effective, employees must understand the importance of these policies and be willing to follow them. This will only happen if the policies do not impede their daily tasks and if the tools provided make their work easier (Digitale Bewaring, 2003). Therefore, digital longevity goes beyond technology; it involves comprehensive planning, education, and organisational change.

USAGE OF DURABLE AND IMMUTABLE STORAGE AND REGULAR STORAGE MAINTENANCE

The use of durable and immutable storage supports long-term data retention by providing media with extended physical lifespans. Durable options include Blu-ray, M-Discs, and cartridges (Lunt et al., 2013; Parhami, 2019), while immutable storage systems include blockchain, Filecoin, and write once read many (WORM) discs (Shahani et al., 2005; Goda & Kitsuregawa, 2012; Perera et al., 2020; Tao et al., 2022). Additionally, Shahani, et al. (2005) highlighted that the Association for Intelligent Information Management developed an error-monitoring system for WORM and CD optical media, with support from the National Institute of Standards and Technology, to extend the life of data. Nonetheless, longevity still depends on the regular maintenance of both hardware and software storage to prevent deterioration and obsolescence. This will allow stakeholders to keep construction project data without constraints.

AWARENESS OF DATA RISKS AND ADOPTION OF DATA CLEANSING AND INTEGRITY CHECKS

According to Tanga et al. (2021), data risks threaten the longevity of project data through technical, human, and environmental vulnerabilities. While cyberattacks dominate discussions about the fourth industrial revolution (4IR), physical access issues, human error, inadequate backup systems, and environmental hazards remain equally damaging (Cains et al., 2021; Tanga et al., 2022a, b). Additionally, if an organisation has weak control methods, such as poor training or unmanaged hardware, it creates additional vulnerabilities that shorten data life (Cains et al., 2021). Therefore, construction project members should be aware of all types of data risks. Furthermore, Ganti and Sarma (2002) and DAMA International (2017) stressed that one way to diminish data risks is to use data cleansing, as it improves long-term usability by detecting and correcting errors, ensuring that data conform to established standards. Adding regular integrity checks to data cleansing is vital, as it prevents loss, tampering, or leakage, thereby supporting availability and longevity (He et al., 2010; Grillenberger & Romeike, 2017; Kim & Lee, 2021). However, implementing effective integrity management in construction also requires baseline data security controls.

INTENTIONAL AND CONTROLLED DATA DUPLICATION

Zhao et al. (2014) suggested that the rapid growth of Internet-based services, including email, blogging, social networking, search, and e-commerce, has significantly altered web users' behaviour and patterns. For instance, many activities that people used to do physically are disappearing at a high rate. Activities such as buying and selling goods, and maintaining constant online communication with friends and family have changed (Kemp, 2019). Most of the generated data need to be stored for easy management and reference. Data storage, as part of data management, can be costly; therefore, methods to reduce costs are highly needed (Feng, 2022). One way to save space and money is to adopt various methods for removing duplicates; however, removing all duplicates from the system weakens data availability (Paulo & Pereira, 2014; Kaur et al., 2018). To solve this, it is advisable to opt for intentional and controlled data duplication or minimum necessary redundancy (DAMA International, 2017).

HARDWARE EMULATION

Rechert, et al. (2016) stated that hardware emulation is an important strategy for digital preservation, enabling original software to run on future systems without modification. By recreating the behaviour of legacy hardware, emulation helps ensure continued access to older construction-related software and digital project assets (Digitale Bewaring, 2003; Rechert et al., 2016). Although effective at extending usability, it requires ongoing monitoring and updated workflows to manage emerging risks. Figure 1 provides a summary of all the data longevity methods and technologies discussed here.

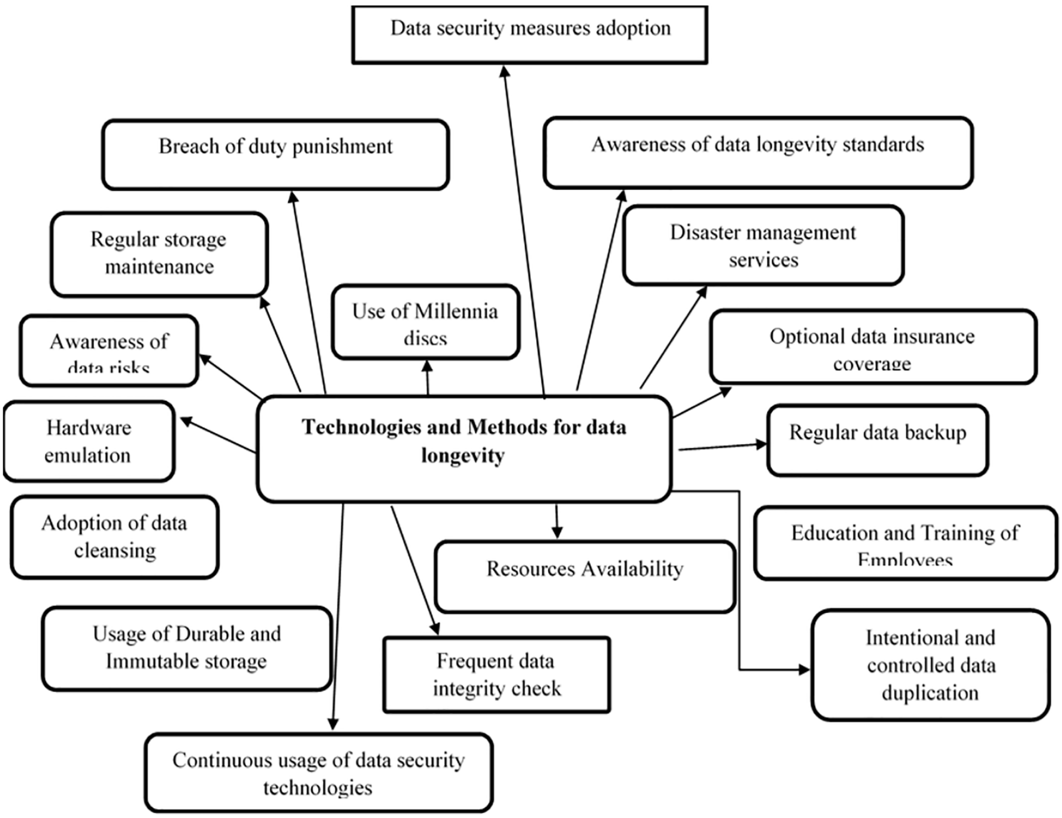

Figure 1. Summary of data longevity methods and techniques. *Source: Author's creation (2025)*

Research methodology

This study employed a quantitative approach to assess the influence of data longevity on data management within the South African construction industry. The targeted population included professionals within organisations registered with the Construction Industry Development Board (CIDB) located in South Africa. Gauteng province was chosen as the study area because of the high volume of construction activities, organisations, and professionals involved in both the public and private sectors (CIDB, 2020). The initial target population was 3,630, as indicated by the CIDB (2020) construction monitor. Using Yamane's sample size equation at a 95% confidence level and a ±5% margin of error, a calculated sample size of 360 was attained for the study. A closed-ended electronic questionnaire was adopted for easy dissemination. The questionnaire contained a cover letter detailing the objectives of the study and ensuring participants' anonymity and their rights in participating in the survey. The study was approved by the Ethics and Plagiarism Committee of the Faculty of Engineering and the Built Environment at the University of Johannesburg (approval number UJ-FEBE-FEPC-01074). The questionnaire was divided into sections, with the first exploring the background information of the respondents. The second section sought answers to the influence of data longevity on data management in construction projects in South Africa. This section was assessed using a 5-point significance scale as suggested by Vagias (2006). Fellows and Liu (2008) have previously emphasised the need to pilot a survey instrument to test if the questions in a questionnaire are intelligible and unambiguous. Thus, the questionnaire was piloted among a select group of construction professionals. The final survey instrument was then modified based on the input received from the pilot study and distributed to the identified respondents. Using a random sampling technique that ensured that each individual and group had an equal chance of selection (Eiselen et al., 2007), 200 valid responses were received and deemed suitable for data analysis. This represented a response rate of 55.6% and was considered adequate for further analysis. Figure 2 gives an overview of the research workflow adopted.

The analysis of the data gathered on the background information was conducted using frequency and percentage. In contrast, the mean item score (MIS) was used to rank the respondents' ratings of the influence of data longevity on data management. Where two or more variables had the same MIS, the one with the lowest standard deviation (SD) was ranked first, as suggested by Pallant (2007). Since the number of factors assessed was considerably large, exploratory factor analysis (EFA) was employed to identify these underlying similarities and regroup similar factors into more manageable subscales. Moreover, EFA was used to understand the interactions between variables and evaluate internal consistency, revealing any constraints not immediately visible in direct analysis (Pallant, 2007). In conducting EFA, the factorability of the data gathered was first tested. To determine this factorability, the Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) and Bartlett's test of sphericity (BTS) were employed with a 0.6 cut-off for KMO and a significant p-value of <0.05 for BTS as proposed in existing studies (Tabachnick & Fidell, 2007; Pallant, 2020). The KMO, BTS, and communality results all indicated that the data gathered were factorable; as such, EFA was conducted using principal component analysis with direct oblimin rotation. The choice of direct oblimin rotation was premised on the need to allow correlation between latent factors, in contrast to varimax, which constrains correlations to zero (Eiselen et al., 2007). The reliability of the data was also tested using Cronbach's alpha (α) test, and an α-value of 0.980, which is above the 0.7 threshold, was derived.

Findings and discussion

BIOGRAPHICAL DATA RESULTS

The analysis of respondents' professional affiliations confirmed that all participants were well qualified to complete the questionnaire. The largest group (29.5%) held a master's degree, followed by those with

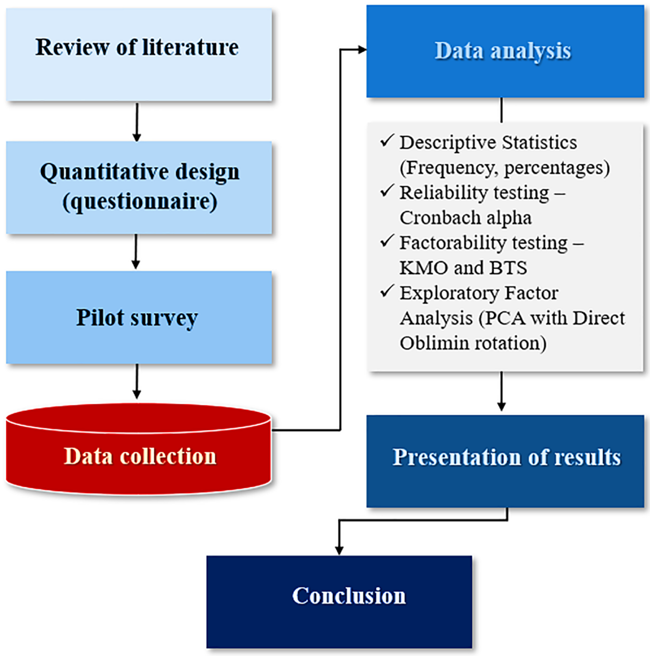

Figure 2. Overview of research workflow. *Source: Author's creation (2025)*

an honours degree (21.5%), doctoral holders (14%), individuals with a post-matric certificate or diploma (14%), and those with a bachelor's degree (14%). In terms of professional roles, engineers made up 31% of respondents, including mechanical engineers (9.5%), electrical engineers (8%), and civil engineers (13.5%). Construction managers represented 23%, construction project managers 15.5%, quantity surveyors 20.5%, and architects 10%. Furthermore, all participants had relevant work experience ranging from 1–5 years to over 20 years, ensuring that their responses were well-informed and aligned with the study's objectives.

MEAN ITEM SCORE AND STANDARD DEVIATION FOR THE INFLUENCE OF DATA LONGEVITY ON DATA MANAGEMENT IN THE CONSTRUCTION INDUSTRY

Based on a literature review, 13 variables were identified as data longevity attributes in data management in the construction industry. These variables were ranked by respondents and subjected to descriptive factor analysis to facilitate the interpretation of the data. The results are displayed in Table 1. According to the participants, "use of immutable storage" was ranked first, with a mean item score of 4.480 and a standard deviation of 0.672 (MIS = 4.480; SD = 0.672); "continuous usage of data protection technologies" (MIS = 4.435; SD = 0.684) and "usage of durable record-keeping devices" (MIS = 4.425; SD = 0.676) were among the first three factors. In contrast, "adoption of data-cleansing methodologies" (MIS = 4.260; SD = 0.732) and "ensuring data remain accessible during a disaster" (MIS = 4.230; SD = 0.546) were ranked among the last two factors.

The objective was set to evaluate the influence of data longevity on data management. According to the results of the descriptive analysis, the use of immutable storage is ranked the highest, confirming the submission of Parhami (2019), who emphasised that immutable storage refers to data storage systems where once data are written, they cannot be altered or deleted, ensuring protection against tampering or accidental deletion. Its high ranking is due to the construction sector's ever-increasing contractual, regulatory, and liability-driven requirement for auditable records, especially for stakeholder dispute resolution, collaboration, and long-term asset operation (Liu et al., 2023). Common implementations include WORM (Goda & Kitsuregawa, 2012) and blockchain-based storage systems (Perera et al., 2020; Tao et al., 2022). Previous

Table 1. Mean item score and standard deviation for data storage attributes for data management in the construction industry.

Data longevity attributes	Mean item score (MIS)	Standard deviation (SD)	Rank
Use of immutable storage	4.480	0.672	1
Continuous usage of data protection technologies	4.435	0.684	2
Usage of durable record-keeping devices	4.425	0.676	3
Knowledge of data longevity standard	4.420	0.637	4
Awareness of construction data risks	4.410	0.666	5
Subscription to data insurance	4.410	0.710	5
Sustainability of repository management	4.395	0.708	7
Frequent project data integrity checks	4.355	0.679	8
Education of employees on data longevity	4.340	0.613	9
Hardware emulation for data preservation	4.330	0.737	10
Intentional and controlled data duplication	4.290	0.706	11
Adoption of data cleansing methodologies	4.260	0.732	12
Ensuring data remain accessible during a disaster	4.230	0.546	13

studies of construction data management have shown that firms prioritise immutable and tamper-resistant storage, largely because it supports regulatory compliance, improves traceability, and ensures the reliable documentation of assets throughout long operational lifespans (Perera et al., 2020;Liu et al., 2023).

According to respondents, the second most influential factor in the longevity of construction data was the continuous use of data protection technologies. This ranking reflects heightened awareness of cybersecurity risks in increasingly digitalised construction environments, where data breaches can slow project progress (Tambwe et al., 2025). Chandrasena (2022) highlighted significant security concerns about potentially harmful systems, which pose risks to global Internet business transactions and deter many businesses from adopting online technologies. Addressing these security flaws is crucial to maintaining confidentiality and preventing reputational damage (Efimova et al., 2021; Tanga et al., 2022a). A comparable study by Bernardo et al. (2024) observed that many organisations prioritise security-oriented technologies over governance-based measures because of their immediate and visible benefits. Moreover, the use of durable record-keeping devices was also identified as a factor affecting data longevity in the built environment, with an SD of 4.425 based on the survey results. According to Lunt (2013), durable data storage has existed for a long time and continues to exist today, as seen in materials such as ceramics, parchment, papyrus, and gold. However, with the introduction of the 4IR, these materials are now useless. Consequently, storage such as Millennia Discs (Parhami, 2019) and Blu-ray (Goda & Kitsuregawa, 2012) has emerged as viable solutions. Previous studies by Lunt et al. (2012) and Lunt et al. (2013) similarly suggested that although durable storage is very important, its implementation is often constrained by capacity limitations and integration challenges with contemporary digital systems.

In addition, both knowledge of data longevity standards and awareness of construction data risks are ranked among the top five factors that influence construction data longevity in data management, aligning with the opinions of authors such as Frank (2020) and Cains et al. (2021). Their relatively high ranking suggests growing recognition that technical solutions alone are insufficient to eliminate risks without incorporating standards and risk-awareness frameworks (Tambwe et al., 2025). Boyda (2013) stressed that significant challenges in long-term digital storage stem from the lack of standardised criteria for managing vast digital records. This lack of standards complicates the organisation, preservation, and accessibility of digital data over time, affecting data integrity, security, and the ability to make informed decisions about data retention and disposal strategies. Frank (2020) supported this view, emphasising that standards play a huge role in addressing these challenges. According to Abraham et al. (2019), in comparison with the study by Olapade et al. (2019), which underemphasised data governance-related factors, the relatively high ranking observed in this study suggests a gradual shift towards more holistic data management approaches in construction. Furthermore, respondents ranked data cleansing methodologies and disaster data accessibility as having a lower influence on data management. This is because data cleansing occurs at specific project stages, while disaster accessibility measures are rarely activated. Although vital for long-term resilience, stakeholders prioritise practices with more immediate impacts on asset life cycle and performance management (Tambwe et al., 2025).

AN EXPLORATORY FACTOR ANALYSIS FOR THE INFLUENCE OF DATA LONGEVITY ON DATA MANAGEMENT

All factors with eigenvalues greater than one are positioned along the steep slope of the plot, whereas those with eigenvalues below one appear along its gradual decline. Table 2 outlines the number of influencing factors and their corresponding eigenvalues. Based on these results, only two clusters satisfy Kaiser's criterion, with eigenvalues exceeding one. These clusters were therefore considered for the factor analysis. Table 3 presents the results of the EFA clustering, which was performed using the oblimin rotation method due to the interconnections among the variables. As shown in Table 3, the 13 variables related to the impact of data longevity on construction data management are categorised into two clusters. These two-factor clusters emerged from the EFA, which was applied to identify correlation patterns within the dataset. The pattern matrix illustrates the linear combinations of the measured variables. The discussion of the three extracted factors is presented subsequently.

Cluster 1

"Data robustness and preservation" contained the following variables: "Hardware emulation for data preservation" (86.5%)", "adoption of data cleansing methodologies" (75.0%), and "usage of durable record-keeping devices" (71.9%) were the three first factors, while "education of employees on data longevity" (54.6%), "frequent project data integrity checks" (49.4%), and "intentional and controlled data duplication" (44.7%) were the three last factors.

Digitale Bewaring (2003) stressed that hardware emulation plays a major role in data preservation, significantly influencing data management by enabling the replication of real-world hardware behaviour, thereby keeping legacy data accessible despite technological obsolescence. This approach will protect data from loss while supporting system recovery and faster debugging. In the construction industry, sustained access to preserved project data supports long-term asset performance by enabling informed maintenance planning, life cycle decision-making, and performance monitoring of built assets during the operation and maintenance phases. Parhami (2019) and Lunt et al. (2012; 2013) further highlighted that ensuring long-term data storage requires high-quality media, which can last for centuries due to their specialised recording technology. This improves life cycle continuity by preserving asset information needed for end-of-life decision-making. Such long-lasting storage solutions are particularly valuable for construction assets with extended service lives, where historical design and maintenance records remain critical long after project completion. Additionally, 75% of respondents in a questionnaire survey indicated that data cleansing

Table 2. Total variance explained for data longevity.

Total variance explained							
Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings[a]
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total
1	5.459	41.989	41.989	5.459	41.989	41.989	5.254
2	1.078	8.291	50.280	1.078	8.291	50.280	2.469
3	0.908	6.986	57.265				
4	0.832	6.399	63.664				
5	0.733	5.635	69.299				
6	0.684	5.260	74.559				
7	0.591	4.546	79.105				
8	0.572	4.401	83.506				
9	0.531	4.087	87.593				
10	0.479	3.686	91.279				
11	0.430	3.311	94.590				
12	0.365	2.805	97.395				
13	0.339	2.605	100.000				
Extraction method: principal component analysis.							

[a] When components are correlated, the sums of squared loadings cannot be added to obtain a total variance.

methodologies also have a major impact on data management. This aligns with Ganti and Sarma (2022) and DAMA International (2017), who explained that data cleansing helps maintain high-quality data by identifying and correcting errors, thereby reducing misunderstandings among project stakeholders. With the adoption of digital twins in the built environment, high-quality, cleansed data are essential, as inaccurate or inconsistent datasets can compromise the reliability of digital replicas used for asset monitoring and predictive maintenance (Khalil et al. 2021; McKinsey & Company, 2023). By enhancing data usability over time, data cleansing supports not only operational efficiency but also the long-term accuracy of digital asset representations.

Beyond data cleansing, durable storage media and immutable storage solutions also influence data management, as noted by 71.9% and 67.9% of respondents, respectively. Technologies such as Blu-ray, M-Discs, and cartridges, alongside blockchain, Filecoin, and WORM discs, improve data security and longevity (Lunt, 2013; Tao et al., 2022). This is particularly important for heritage buildings and historically significant infrastructure, where durable and immutable storage mechanisms are important, as they preserve original design documentation, restoration records, and conservation histories that must remain accessible across generations. In particular, immutable storage methods such as blockchain and WORM discs prevent

Table 3. Pattern matrix for data longevity.

Pattern matrix[a]		
	Component	
	1	2
Hardware emulation for data preservation	0.869	
Adoption of data cleansing methodologies	0.750	
Usage of durable record-keeping devices	0.719	
Use of immutable storage	0.679	
Continuous usage of data protection technologies	0.662	
Awareness of construction data risks	0.655	
Sustainability of repository management	0.633	
Subscription to data insurance	0.605	
Education of employees on data longevity	0.546	
Frequent project data integrity checks	0.494	
Intentional and controlled data duplication	0.447	
Knowledge of data longevity standards		0.762
Ensuring data remain accessible during a disaster		0.731
Extraction Method: principal component analysis. Rotation method: oblimin with Kaiser normalisation.		

[a] Rotation converged in 7 iterations.

tampering, thereby supporting regulatory requirements across construction, finance, and other industries (Shahani et al., 2005; Goda & Kitsuregawa, 2012; Lunt et al., 2013; Parhami, 2019; Perera et al., 2020; Tao et al., 2022).

Awareness of construction data risks is another critical factor affecting large-scale data management (Cains et al., 2021). The Information Technology Authority (2017) and Tambwe et al. (2023) emphasised that recognising these risks allows organisations to implement effective strategies, such as fire prevention services, staff training forums, encryption, and blockchain, ensuring long-term data integrity. Furthermore, 65.5% of survey participants highlighted the importance of continuous data protection technologies in data management. Several authors, including Chandrasena (2022), Tanga et al. (2022a), and Efimova et al. (2021), have explained that encryption, intrusion detection systems, anti-malware tools, and artificial intelligence-driven defences are essential for safeguarding construction data against evolving threats. These continuous security measures are also fundamental to digital twin technologies, which depend on uninterrupted, secure data flows to accurately reflect asset conditions in real-time. Olorunnishola et al. (2025) explained that digital twin technologies serve as a digital archive, preserving detailed records of the original structure that can aid in restoration efforts in the event of damage. Thus, digital twins are considered a pillar of data management, encompassing ongoing security practices in the construction industry. The regular maintenance of both hardware and software is essential to ensure continued data accessibility and prevent obsolescence. Sustaining a data repository also requires securing adequate funding, ongoing technical support, and adapting to technological advancements to maintain functionality and relevance (Digitale Bewaring, 2003). Failure to meet the requirements and necessities may undermine long-term asset performance and limit the effectiveness of digital asset management systems. In addition to regular update maintenance, Ma (2014) proposed data insurance as a means of mitigating losses, allowing users to file claims and receive compensation if insured data become unrecoverable, a view shared by 60.5% of survey respondents.

Furthermore, the increasing reliance on ICT for data management highlights the need for investments in security, policy development, and staff training to support long-term data preservation (Lin et al., 2020; Shaban et al., 2022). Tambwe et al. (2023) stressed that educating employees about data risks and equipping them with the necessary security tools build confidence and enhance data protection throughout project execution. This education strengthens organisational capacity to sustain reliable data environments that support asset life cycle management, heritage conservation, and advanced digital applications. Kim and Lee (2021) agreed with respondents who believed that frequent project data integrity checks and controlled data duplication have a positive impact on data management, ensuring consistent backups. These practices enhance not only data quality but also life cycle resilience by providing accurate information for maintenance planning and predictive asset management. However, while deduplication enhances storage efficiency, excessive removal of redundant data may compromise availability, making intentional and controlled duplication necessary (Paulo and Pereira, 2014; DAMA International, 2017; Kaur et al., 2018). Balancing efficiency and availability promotes trustworthy datasets, which are essential for long-term asset performance analysis and the good functioning of digital twins.

Cluster 2

"Data continuity assurance" contained the following variables: "knowledge of data longevity standard" (76.2%) and "ensuring data remain accessible during a disaster" (73.1%).

Understanding data longevity standards is crucial for effective data management, ensuring that information remains intact and accessible over time. Organisations such as ISO provide guidelines on best practices for data storage, preservation, and retrieval (Kaur et al., 2018; Frank, 2020; Lin et al., 2020; Malla et al., 2023). In the construction industry, the sustained availability of high-quality data supports long-term asset performance by enabling continuity of operational knowledge and informed decisions throughout the asset's life cycle. In heritage construction, preserving data is paramount for retaining historically significant information, ensuring conservation and compliance with regulations. Additionally, maintaining data accessibility during disasters, such as system failures or natural calamities, requires robust backup strategies, redundancy measures, and disaster recovery plans (Khatri and Brown, 2010; Filecoin, 2022). These resilience measures are equally critical for digital twin technologies, which depend on trustworthy data flows to maintain functional virtual representations of physical assets over time (Olorunnishola et al., 2025). Adhering to longevity standards and ensuring accessibility during crises enhance the continuity of data operations, thereby reducing the risk of data corruption (Sakurai & Murayama, 2019). Alongside employee education on longevity standards, these practices promote a resilient and sustainable data system (Digitale Bewaring, 2003).

Implications of findings and workflow of data longevity in construction

The study's findings can significantly enhance data management in construction projects by promoting the adoption of effective data longevity strategies. Stakeholders can leverage these insights to enhance performance by optimising data storage and reducing costs. Ensuring data longevity offers multiple advantages, including informed decision-making through access to historical data, improved maintenance and operations, regulatory compliance through accurate record-keeping, cost efficiency through precise budgeting, and facilitating the transfer of information among stakeholders. Educators can utilise these findings to raise awareness of the importance of data longevity in construction and to develop training programmes that equip students and professionals with the skills to implement these strategies effectively. Policymakers can use the study's insights to inform regulations and guidelines that mandate the adoption of data longevity practices, ensuring long-term data preservation in the industry. Ultimately, the research contributes to the existing body of knowledge by identifying key factors for achieving effective data longevity, thereby serving as a foundation for future studies on the efficacy of various data longevity strategies in construction projects. Figure 3 illustrates the workflow for data longevity in the construction sector, outlining the steps required to successfully implement data longevity, from planning and awareness to data governance and compliance.

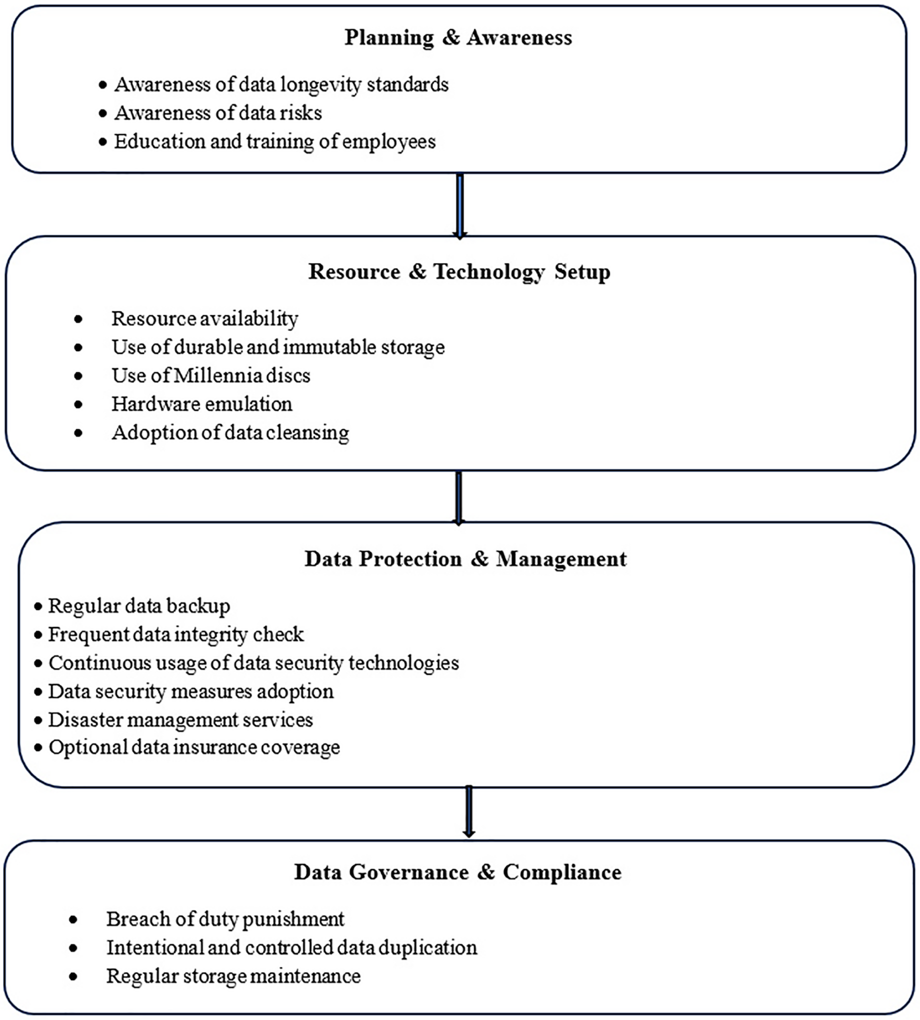

Figure 3. Workflow of data longevity in construction.

Conclusion and recommendations

This study aimed to assess the influence of data longevity on data management in the construction industry. To achieve this objective, a quantitative research method was adopted. A questionnaire survey was distributed to professionals in the South African construction industry, and their responses were analysed. The findings revealed that data longevity influences data management by promoting the use of immutable storage, continuous data protection technologies, durable record-keeping devices, knowledge of data longevity standards, and awareness of construction data risks. To fully benefit from data longevity, stakeholders must carefully implement the measures and strategies outlined in this study throughout their projects. Incorporating data longevity strategies into daily construction operations significantly enhances data preservation for future use in construction projects. The findings further highlight the impact of data longevity on construction data management, encouraging stakeholders to maintain a strong focus on data storage practices and to proactively mitigate risks that threaten data longevity throughout the project life cycle. This proactive approach ensures that critical information remains accessible and secure, supporting informed decision-making and overall project success, including retrofitting and renovation. To enhance data longevity in the construction industry, it is recommended that professionals receive training and that all project members be educated on relevant principles and their practical application. Additionally, collaboration with universities is essential for researching emerging data longevity practices, particularly in addressing local challenges such as limited resources or regulatory complexities. Encouraging these institutions to develop case studies, pilot projects, and frameworks will facilitate the direct application of these practices within the construction sector. This study employed a quantitative method, which limited the understanding of participants' insights. Additionally, confirmatory factor analysis (CFA) was not performed, leaving the generalisability of the factor structure untested. Future studies should adopt a mixed-methods approach to complement statistical findings with qualitative insights and apply CFA across regions to validate factor stability. Moreover, research could explore how emerging technologies, such as artificial intelligence, blockchain, and digital twins, support data longevity in the construction sector.

TERMINOLOGY DEFINITIONS

Data longevity: ensures that data remain retrievable and usable through preservation strategies by maintaining accessibility, as well as managing its lifespan.

Data security: protects data from threats, including loss and unauthorised access, to ensure the confidentiality and integrity of data.

References

Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Beetz, J., Dietze, S., Berndt, R., & Tamke, M. (2013). Towards the long-term preservation of building information models. In *Proceedings of the CIB W78 2013: 30th International Conference on Applications of IT in the AEC Industry* (pp. 209–217). Tsinghua University.

Bernardo, B. M. V., São Mamede, H., Barroso, J. M. P., & dos Santos, V. M. P. D. (2024). Data governance and quality management——Innovation and breakthroughs across different fields. *Journal of Innovation & Knowledge*, *9*(4), 1–34. https://doi.org/10.1016/j.jik.2024.100598

Bhalerao, A., & Pawar, A. (2017). A survey on data deduplication for efficiently utilising cloud storage for big data backups. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 933–938). IEEE. https://doi.org/10.1109/ICOEI.2017.8300844

Blum, R., & Beyer, B. (2019). Achieving digital permanence. *Communications of the ACM*, *62*(5), 36–42. https://doi.org/10.1145/3303872

Boyda, J. (2013). On loss in the 21st century: Digital decay and the archive. *Spectator*, *33*(2), 6–10.

Cains, M. G., Flora, L., Taber, D., King, Z., & Henshel, D. S. (2021). Defining cyber security and cyber security risk within a multidisciplinary context using expert elicitation. *Risk Analysis*. https://doi.org/10.1111/risa.13687

Chandrasena, S. (2022). Data preservation and risk management in management information systems. *South Florida Journal of Development*, *3*(1), 1459–1479. https://doi.org/10.46932/sfjdv3n1-112

Choi, W., Arjomand, M., Jung, M., & Kandemir, M. (2017). Exploiting data longevity for enhancing the lifetime of flash-based storage class memory. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, *1*(1), 1–26. https://doi.org/10.1145/3084458

Construction Industry Development Board (CIDB). (2020). *Construction monitor*. https://www.cidb.org.za/wp-content/uploads/2021/04/Construction_Monitor_Contractor_Development-July-2020.pdf

Crofts, C. (2008). Digital decay. *The Moving Image*, *8*(2), 33–35. https://doi.org/10.1353/mov.0.0015

DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge*. Technics Publications.

Digitale bewaring. (2003). *From digital volatility to digital permanence: Preserving email* https://pages.ischool.utexas.edu/galloway/spring2019/INF392K2019/volatility-permanence-email-en.pdf

Dolgin, E. (2018). Longevity data hint at no natural limit on lifespan. *Nature*, *559*(7712), 14–15. https://doi.org/10.1038/d41586-018-05582-3

Efimova, O. V., Igolnikov, B. V., Isakov, M. P., & Dmitrieva, E. I. (2021). Data quality and standardization for effective use of digital platforms. In *2021 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)* (pp. 282–285). IEEE. https://doi.org/10.1109/ITQMIS53292.2021.9642876

Eiselen, R., Uys, T., & Potgieter, T. (2007). *Analysing survey data using SPSS 13* (3rd ed.). Johannesburg: Statkon, University of Johannesburg.

El Arass, M., & Souissi, N. (2018). Data lifecycle: From big data to smart data. In *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)* (pp. 80–87). IEEE. https://doi.org/10.1109/CIST.2018.8596547

Fellows, R. R., & Liu, A. (2008). *Research methods for construction* (3rd ed.). Wiley-Blackwell.

Feng, D. (2022). *Data deduplication for high-performance storage systems*. Springer Nature. https://doi.org/10.1007/978-981-19-0112-6

Filecoin (2022) *Filecoin for large data*. https://filecoin.io/blog/posts/filecoin-for-large-data/

Frank, R. D. (2020). The social construction of risk in digital preservation. *Journal of the Association for Information Science and Technology*, *71*(4), 474–484. https://doi.org/10.1002/asi.24247

Frank, R. D. (2022). Risk in trustworthy digital repository audit and certification. *Archival Science*, *22*(1), 43–73. https://doi.org/10.1007/s10502-021-09366-z

Ganti, V., & Sarma, A. D. (2022). *Data cleaning*. Springer Nature.

Goda, K., & Kitsuregawa, M. (2012). The history of storage systems. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1433–1440. https://doi.org/10.1109/JPROC.2012.2189787

Goh, W. W. B., Thalappilly, S., & Thibault, G. (2019). Moving beyond the current limits of data analysis in longevity and healthy lifespan studies. *Drug Discovery Today*, *24*(12), 2273–2285. https://doi.org/10.1016/j.drudis.2019.08.008

Grillenberger, A., & Romeike, R. (2017). Key concepts of data management: An empirical approach. In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research* (pp. 30–39). https://doi.org/10.1145/3141880.3141886

He, Q., Li, Z., & Zhang, X. (2010). Data deduplication techniques. In *2010 International Conference on Future Information Technology and Management Engineering* (pp. 430–433). IEEE. https://doi.org/10.1109/FITME.2010.5656539

Kansa, S. W., Atici, L., Kansa, E. C., & Meadow, R. H. (2020). Archaeological analysis in the information age: Guidelines for maximizing the reach, comprehensiveness, and longevity of data. *Advances in Archaeological Practice*, *8*(1), 40–52. https://doi.org/10.1017/aap.2019.36

Kaur, R., Chana, I., & Bhattacharya, J. (2018). Data deduplication techniques for efficient cloud storage management: A systematic review. *The Journal of Supercomputing*, *74*, 2035–2085. https://doi.org/10.1007/s11227-017-2210-8

Kemp, S. (2019). *Digital trends 2019: Every single stat you need to know about the internet*. https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-Internet/

Khalil, A., & Stravoravdis, S. (2022). Digital building data longevity and interoperability challenges in the documentation of heritage buildings. In *9th International Workshop 3D-ARCH : 3D Virtual Reconstruction and Visualization of Complex Architectures* (pp. 283–289). https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-283-2022

Khalil, A., Stravoravdis, S., & Backes, D. (2021). Categorisation of building data in the digital documentation of heritage buildings. *Applied Geomatics*, *13*, 29–54. https://doi.org/10.1007/s12518-020-00322-7

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148–152. https://doi.org/10.1145/1629175.1629210

Kim, W. B., & Lee, I. Y. (2021). Survey on data deduplication in cloud storage environments. *Journal of Information Processing Systems*, *17*(3), 658–673.

Kotla, R., Alvisi, L., & Dahlin, M. (2007). SafeStore: A durable and practical storage system. In *USENIX Annual Technical Conference* (pp. 129–142).

Lee, C. A. (2010). Open archival information system (OAIS) reference model. In *Encyclopedia of Library and Information Sciences* (3rd ed., pp. 4020–4030). https://doi.org/10.1081/E-ELIS3-120044377

Leppikorpi, T. (2018). *Utilizing information systems in inter-organizational collaboration*. Master's dissertation, University of Tampere.

Liu, D., Wang, B., Li, P., Stones, R. J., Marbach, T. G., Wang, G., Liu, X., & Li, Z. (2020). Predicting hard drive failures for cloud storage systems. In *Algorithms and Architectures for Parallel Processing: 19th International Conference, ICA3PP 2019* (pp. 373–388). Springer. https://doi.org/10.1007/978-3-030-38991-8_25

Liu, H., Han, S., & Zhu, Z. (2023). Blockchain technology toward smart construction: Review and future directions. *Journal of Construction Engineering and Management*, *149*(3), 1–16. https://doi.org/10.1061/JCEMD4.COENG-11929

Lunt, B. M., Linford, M. R., & Davis, R. (2012). *Research on another permanent data storage solution*. https://library.imaging.org//admin/apis/public/api/ist/website/downloadArticle/archiving/9/1/art00005. https://doi.org/10.2352/issn.2168-3204.2012.9.1.art00005

Lunt, B. M., Linford, M. R., Davis, R. C., Jamieson, S., Pearson, A., & Wang, H. (2013). *Toward permanence in digital data storage*. https://www.researchgate.net/publication/263340007_Toward_Permanence_in_Digital_Data_Storage#fullTextFileContent. https://doi.org/10.2352/issn.2168-3204.2013.10.1.art00029

Ma, X. (2014). On the feasibility of data loss insurance for personal cloud storage. In *6th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 14)* (pp. 1–5). Philadelphia, Pennsylvania, USA.

Malla, V., Tummalapudi, M., & Delhi, V. S. K. (2024). Perceptions of built-environment professionals on using ISO 19650 standards for information management. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *16*(1), 04523045. https://doi.org/10.1061/JLADAH.LADR-1076

McKinsey & Company. (2023). *From start-up to scale-up: Accelerating growth in construction technology*. https://www.mckinsey.com//~/media/mckinsey/industries/private%20equity%20and%20principal%20investors/our%20insights/from%20start%20up%20to%20scale%20up%20accelerating%20growth%20in%20construction%20technology/from-start-up-to-scale-up-accelerating-growth-in-construction-technology-vf.pdf

Meza, J., Wu, Q., Kumar, S., & Mutlu, O. (2015). A large-scale study of flash memory failures in the field. *ACM SIGMETRICS Performance Evaluation Review*, *43*(1), 177–190. https://doi.org/10.1145/2796314.2745848

M-Files. (2019). *The state of information management in the construction industry*. https://fts-eu.com//resources/the-2019-intelligent-information-management-benchmark-report/

Olapade, D. T., Shiyanbola, R. E., Ayorinde, I. O., & Olaleye, A. (2019). Property data management practice in an emerging property market: Experience from Lagos, Nigeria. *Journal of Property Research and Construction (JPRC)*, *3*(1), 59–66.

Olorunnishola, A., Zhang, C., Kumar, D., Liu, P., & Denzer, A. (2025). Preservation of heritage buildings through digitisation—Reality capture of the Old Faithful Inn: An overview of the process, potentials, and challenges. In *Computing in Civil Engineering* (pp. 703–713). https://doi.org/10.1061/9780784486122.078

Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS for Windows 15* (3rd ed.) Maidenhead: Open University Press.

Parhami, B. (2019). *Data longevity and compatibility*. https://web.ece.ucsb.edu/~parhami/pubs_folder/parh19f-ebdt-data-longevity-compatib.pdf. https://doi.org/10.1007/978-3-319-77525-8_331

Paulo, J., & Pereira, J. (2014). A survey and classification of storage deduplication systems. *ACM Computing Surveys (CSUR)*, *47*(1), 1–30. https://doi.org/10.1145/2611778

Pentury, C., Karsaman, R. H., Rahman, H., & Rosmansyah, Y. (2025). Application of digitalization and computerization technology in road construction. *Automation in Construction*, *171*, 1–13. https://doi.org/10.1016/j.autcon.2025.106018

Perera, S., Nanayakkara, S., Rodrigo, M. N. N., Senaratne, S., & Weinand, R. (2020). Blockchain technology: Is it hype or real in the construction industry? *Journal of Industrial Information Integration*, *17*, 1–20. https://doi.org/10.1016/j.jii.2020.100125

Rechert, K., Falcao, P., & Ensom, T. (2016). *Introduction to an emulation-based preservation strategy for software-based artworks*. https://www.researchgate.net/publication/311614858_Introduction_to_an_emulation-based_preservation_strategy_for_software-based_artworks

Rothenberg, J. (1995). Ensuring the longevity of digital documents. *Scientific American*, *272*(1), 42–47. https://doi.org/10.1038/scientificamerican0195-42

Rothenberg, J. (1999). *Ensuring the longevity of digital information*. https://www.clir.org/wp-content/uploads/sites/13/ensuring.pdf

Sakurai, M., & Murayama, Y. (2019). Information technologies and disaster management—Benefits and issues. *Progress in Disaster Science*, *2*, 100012–100016. https://doi.org/10.1016/j.pdisas.2019.100012

Shaban, A. I., Farhan, M. A., & Ahmed, S. R. (2022). Building a smart system for preservation of government records in digital form. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1–6). IEEE. https://doi.org/10.1109/HORA55278.2022.9800034

Shahani, C. J., Manns, B., & Youket, M. (2005). *Longevity of CD media: Research at the Library of Congress*. Preservation Research and Testing Division, Washington, DC.

Sharma, C., & Vaid, A. (2024). Leveraging SAP Information Lifecycle Management (ILM): Latest insights and applications. *Zenodo*, *5*(6), 167–173.

Tambwe, O. T., Aigbavboa, C. O., & Akinradewo, O. (2023). Benefits of construction data risks management in the construction industry. *Journal of Engineering, Design and Technology*, *23*(2), 458–476. https://doi.org/10.1108/JEDT-11-2022-0577

Tambwe, O. T., Aigbavboa, C. O., Akinradewo, O. I., & Adekunle, P. A. (2025). Measures to address cyber-attacks in construction project data management processes: A cybersecurity perspective. *IET Information Security*, *1*, 1–11. https://doi.org/10.1049/ise2/7398742

Tanga, O., Akinradewo, O., Aigbavboa, C., & Thwala, D. (2021). Usage of cloud storage for data management in the built environment. In *Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2021 Virtual Conferences on Human Factors in Software and Systems Engineering, Artificial Intelligence and Social Computing, and Energy, July 25–29, 2021, USA* (pp. 465–471). Springer International Publishing. https://doi.org/10.1007/978-3-030-80624-8_58

Tanga, O., Akinradewo, O., Aigbavboa, C., & Thwala, D. (2022a). Cyber-attack risks to construction data management in the fourth industrial revolution era: A case of Gauteng province, South Africa. *Journal of Information Technology in Construction*, *27*, 846–863. https://doi.org/10.36680/j.itcon.2022.041

Tanga, O., Akinradewo, O., Aigbavboa, C., Oke, A., & Adekunle, S. (2022b). Data management risks: A bane of construction project performance. *Sustainability*, *14*(19), 1–20. https://doi.org/10.3390/su141912793

Tao, X., Liu, Y., Wong, P. K. Y., Chen, K., Das, M., & Cheng, J. C. (2022). Confidentiality-minded framework for blockchain-based BIM design collaboration. *Automation in Construction*, *136*(2), 104172–104183. https://doi.org/10.1016/j.autcon.2022.104172

Vagias, W. M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management, Clemson University.

Van Besouw, J., & Bond-Barnard, T. (2021). Smart project management information systems (SPMIS) for engineering projects–Project performance monitoring & reporting. *International Journal of Information Systems and Project Management*, *9*(1), 78–97.

Ye, X., Liu, L., & König, M. (2024). Improving data and process management for digital twins: A smart contract-driven solution integrating CDE and blockchain for the construction industry. In *Computing in Civil Engineering 2024* (pp. 332–340). https://doi.org/10.1061/9780784486122.036

Zhang, Q., & Wang, L. (2026). AI in smart buildings and construction 4.0: Implementation areas and influencing factors in construction organizations. *Automation in Construction*, *181*, 1–17. https://doi.org/10.1016/j.autcon.2025.106623

Zhao, L., Sakr, S., Liu, A., & Bouguettaya, A. (2014). *Cloud data management*. Springer. https://doi.org/10.1007/978-3-319-04765-2