SUNTEXT REVIEWS

SunText Review of Case Reports & Images
ISSN: 2766-4589

Open Access
Research Article
Volume 5:1

The Application of Blockchain Technology in the Field of Electronic Document Management: A Case Study of the ARCHANGEL Project

Peng S*

School of History and Public Management of Yancheng Teachers University, Jiangsu, Yancheng, China

***Corresponding author:** Peng S, School of History and Public Management of Yancheng Teachers University, Jiangsu, Yancheng, China; E-mail: songp@yctu.edu.cn

Received date: 31 December 2023; **Accepted date:** 02 January 2024; **Published date:** 08 January 2024

Citation: Peng S (2024) The Application of Blockchain Technology in the Field of Electronic Document Management: A Case Study of the ARCHANGEL Project. SunText Rev Case Rep Image 5(1): 119.

DOI: https://doi.org/10.51737/2766-4589.2024.119

Abstract

The ARCHANGEL project represents a pivotal advancement in the domain of blockchain technology, focusing on the establishment of a secure, efficient, and reliable framework for preserving electronic archival information. At the core of its operational ethos—"not to detect counterfeit but to substantiate the genuine"—resides its guiding principle. This initiative provides the public with robust archive verification services, enhancing the trustworthiness of information preserved within its system. By conducting an in-depth analysis of ARCHANGEL's technological foundations and scholarly aspects, as well as a comprehensive understanding of its blockchain execution model, we examine the project's potential impact on the future integration of blockchain methodologies in managing electronic records. This includes developing a respected infrastructure, increasing public trust in archival systems, and promoting a decentralized and collaboratively managed structure. Simultaneously, we undertake a critical evaluation of the project's current limitations, acknowledging the need for continuous research and improvement in this innovative and dynamic field.

Keywords: Blockchain; Electronic archive; Archive trust; Decentralization

Introduction

In the contemporary era, marked by the rapid progression of informatization, the impact on archival practices extends beyond the straightforward digitization of analog information resources. This paradigm shift is manifesting as a holistic re-envisioning of digital electronic archive management. The burgeoning prevalence of information technology has precipitated an exponential increase in both the volume and diversity of native electronic documents. Nevertheless, governing these documents introduces a plethora of challenges and technical intricacies that extend beyond the scope of conventional archival principles and methodologies. The dissociation between the mediums of electronic documents and the information they contain, combined with the fluid nature of digital data, allows for the potential modification of content throughout its lifecycle—from creation and transmission to storage and retrieval [1]. Such alterations hold the promise of unlocking latent value across myriad domains, yet they also raise concerns about the integrity of electronic documents. The daunting task of verifying whether files have been altered, deleted, or lost during their lifecycle engenders skepticism about their reliability as archival records. In response to these challenges, blockchain technology has emerged as a formidable countermeasure to reinforce the security, integrity, and authenticity of electronic documents [2]. Its intrinsic characteristics—such as immutability, transparency, traceability, and enhanced security—are foundational to addressing these concerns [3]. The ARCHANGEL project, spanning from June 2017 to June 2019, exemplified a pioneering collaboration led by the University of Surrey, along with partners such as the UK's National Archives, to harness the strengths of blockchain. The initiative aimed to safeguard the integrity and truthfulness of electronic document content over extended periods, ensuring that metadata and archival materials remain unchanged, thereby enhancing trust in digital archive administration at national, societal, and public levels [4]. At the heart of ARCHANGEL's philosophy was the commitment to ascertain authenticity, not merely to detect forgeries. The project developed a prototype

service based on Distributed Ledger Technology (DLT), which facilitates a participatory network environment where data can be accessed, replicated, and synchronized across a communal database without the need for a centralized authority [5]. Each participant retains a copy of the ledger, protected by public and private keys, as well as digital signatures. Ledger entries are harmonized and secured by a unique cryptographic hash, making them verifiable, audit-worthy historical records within the network. Data generated, preserved, and disseminated by network participants is consistently traceable and verifiable. In this architecture, unauthorized alterations are exceedingly difficult, enhancing the reliability of DLT-mediated information and promoting an environment of trust and confidence among participants. Blockchain comprises multiple technologies rather than a single solution, with DLT being a prominent example—a decentralized register maintained across various entities' data blockchains. The foundational elements of blockchain include peer-to-peer networking, distributed consensus protocols, and asymmetric cryptographic techniques [6]. These components work together to ensure data decentralization, distributed storage, data existence and integrity verification, traceability, and permanence. Blockchain operates based on several core technical principles: (1) A peer-to-peer network infrastructure that utilizes multiple nodes for synchronized services, duplication, and ledger maintenance. (2) Record access and validation depend on distributed consensus protocols that ensure the authentication and integrity of all network records. (3) Asymmetric cryptographic techniques create a secure data blockchain. When committing electronic documents to a repository, the system uses a private key for hashing, while users employ a corresponding public key for decryption, eliminating the need for key exchanges and thus enhancing the security and confidentiality of electronic files. Decentralization in blockchain refers to the dispersion of record-keeping responsibilities across all network nodes, rather than centralization within a single entity. While many studies conflate DLT with blockchain, the ARCHANGEL project asserts that the distinction between them is subtle yet significant—blockchain is a secure, decentralized type of ledger technology, whereas distributed ledger encompasses a broader range of data infrastructure technologies, including blockchain [7]. For clarity and academic precision, 'blockchain' will be the preferred term in this discourse, though it is crucial to differentiate between the two when discussing DLT specifics.

Research Content of the ARCHANGEL Project

Establishing a Blockchain network

The ARCHANGEL initiative has led the development of a blockchain network, involving a broad consortium of Archival and Memory Institutions (AMIs), including some national archives. The blockchain's interdisciplinary and transnational character extends across various regions, reflecting ARCHANGEL's goal to increase participation beyond traditional archival entities to include non-traditional AMIs such as news agencies and digital public archives [8]. ARCHANGEL's operational model requires a minimum of seven active participants to effectively support the blockchain's architecture. Although, theoretically, a simple majority could modify the blockchain, the significant inclusion of authoritative national archival institutions serves as a safeguard, reinforcing the blockchain's integrity and reducing the risk of data distortion due to their respected professional status and trustworthiness [9]. Moreover, ARCHANGEL utilizes the asymmetric encryption capabilities inherent in blockchain technology, aligning with stringent information control standards to strengthen data security. During the selection of a suitable blockchain platform, ARCHANGEL outlined two conceptual models reflecting the "public chain" and "private chain" aspects of blockchain technology. Archival institutions have the flexibility to choose a model that aligns with their needs and the intended openness of the blockchain platform. One model suggests a conditionally open, publicly accessible chain through "licensed ledger-keeping," allowing any individual or organization to participate or disengage autonomously [10]. This model permits participants to review database copies, contribute to digital information maintenance, and verify information content fidelity, though the addition of new records to ARCHANGEL is limited to authorized entities. Alternatively, a controlled alliance chain might be established among various archival and memory institutions across different countries and disciplines, ensuring stability through regulated membership and permissions. Due to their legal status and professional conduct, public archival institutions, and reputable non-public entities, are less likely to experience withdrawal or managerial disruptions, providing a stable consortium blockchain. After extensive research and empirical trials, ARCHANGEL chose the former model, creating a blockchain aimed at dependable digital archive management. Blockchain technology is celebrated for its principle of "decentralization." However, ARCHANGEL's "licensed ledger-keeping" strategically leverages Distributed Ledger Technology (DLT), granting organizers control and, to some extent, sacrificing decentralization to achieve necessary data permissions and access controls for regulatory compliance. Within the ARCHANGEL blockchain, the National Archives have exclusive rights to append, preserve, and update digital information, preventing other participants from altering data, and thus enhancing the chain's stability and security. The consensus mechanism is fundamental to blockchain, necessary for granting specific participant permissions and maintaining ledger uniformity. It is the algorithmic core of the blockchain, providing

distributed consensus capability. Known consensus protocols include Proof of Work (POW), Proof of Authority (POA), and Proof of Importance (POI) [11]. ARCHANGEL initially adopted POW during its research, testing, and development stages, but later transitioned to explore the POA approach. In POA, authorized network nodes vote or delegate block management. POA suits conditional public chains like ARCHANGEL, matching a network of recognized and authoritative participants, unlike the more democratic POW.

Establishing a comprehensive archive verification process

ARCHANGEL's cornerstone is its verification mechanism, ensuring files remain unchanged or undeleted during storage. The process begins with hashing the archive upon induction and recording the hash value on the blockchain. Authorized users can later verify a file's originality by comparing its current hash value with the initial record. The verification procedure includes seven steps:

Format Identification: Digital archival tools identify the file's format (e.g., PDF, DOC).

1. Hash Value Extraction: A hash algorithm calculates the file's unique, fixed-length hash value.
2. Record Storage: The blockchain logs the hash value, file name or GUID, hash calculation identifier, metadata, and content evidence for future trust management in electronic archives.
3. Original Files and Blockchain: Due to privacy and data volume concerns, the original archive texts are not stored on the blockchain but are linked offline.
4. File Storage: The file is stored securely after processing, with necessary format conversions or data refreshments to maintain integrity.
5. Content Modifications: Changes to archive contents are reflected in updated hash values and metadata on the blockchain.
6. Retrieval and Verification: Users retrieve the stored hash value for comparison, ensuring the archive's authenticity.
7. Audit Trails: The blockchain's traceability logs any unauthorized changes, facilitating discrepancy analysis.

Prototype system development

ARCHANGEL developed a prototype on the Ethereum public test network, chosen for its robustness and acceptance in the blockchain community. The system features a user-friendly interface with two main functions: "Upload" for adding electronic files and generating initial hash values, and "Search" for retrieving and comparing stored hash values with those from the time of archiving. Matching hash values indicate file integrity, confirming its unaltered status during custody.

ARCHANGEL Project Inspiration

Enhancing the authenticity of digital archives

Blockchain Scholars worldwide are exploring blockchain technology's potential to enhance the authenticity of digital archives, meeting the strict demands of electronic file management. The research focuses on several key aspects:

Firstly, integrating procedural registration and metadata encapsulation, complemented by electronic signatures and timestamping technologies, creates a comprehensive authenticity chain. This chain spans the initial monitoring and regulatory oversight, intermediate documentation and recordation, and final audits and tracking stages. Such a multi-faceted approach forms a robust technical defense against tampering with electronic files.

Secondly, the infrastructure of blockchain, combined with consensus mechanisms, securely encapsulates electronic file summaries within its blocks. This establishes a comprehensive system that upholds the authenticity of electronic files at every lifecycle stage.

Thirdly, blockchain alliances' unique numbering systems are used to record essential information about electronic documents on the chain, enhancing their authenticity. Meanwhile, the management of electronic files' authenticity is streamlined by verifying hash values on the blockchain, reducing managerial costs, and establishing a reliable framework to preserve these documents' legal integrity.

Fourthly, blockchain's innate ability to track unauthorized changes, along with its distributed storage model, provides strong resilience against attacks on individual nodes. Additionally, consensus algorithms support the veracity of the information, ensuring the untouchability of electronic files throughout their management cycle.

A holistic examination of digital archives

Management Applications The principles and theories heralded within the industry, particularly those advocating front-end control, are seen as groundbreaking in the digital documentation era. However, practical implementation faces challenges due to the varied management styles, authorities, and responsibilities across different organizational departments. This variability makes it difficult to include electronic document creators in a unified management system. The fundamental unit of a blockchain's structure is the block. The block's header contains critical metadata about electronic files, such as the receiving department, timestamp of reception, and the system that received the file. In contrast, the block's body houses the substantive content of the electronic files, continuing the information chain from one block to the next. As electronic files are received by nodes, their content is stored within the block body, while the

metadata is encapsulated within the block header and interconnected with subsequent blocks, forming a verifiable and traceable chain. By storing electronic archives and their metadata within a single block and merging these components, the issue of disjointed electronic documents and metadata is resolved. Once created, the encapsulated metadata becomes immutable. Additionally, metadata generated post-archiving, like archival location, user information, utilization timelines, and destruction schedules, is continuously integrated within the block header, creating a comprehensive metadata repository for the electronic archive. Blockchain nodes keep detailed records of electronic archives, and any content modifications are verified across the network. Using asymmetric encryption and hash algorithms, data recording and distribution are both transparent and reliable. Blockchain-enabled electronic file management systems support the complete lifecycle management of electronic files, guaranteeing their authenticity, reliability, tamper-resistance, and traceability.

Advancing security in digital archive management

The traditional centralized management model relies heavily on central nodes, a significant vulnerability in conventional electronic records management systems. Professional archival institutions are typically responsible for providing access to, and the use of, archives. If such institutions face unexpected challenges like natural disasters, financial instability, or unauthorized data breaches, the resulting risks to the archives are greatly increased. Although remote storage mechanisms are sometimes used to lessen these risks, they are less affected by the physical space limitations that impact traditional paper archives. In sharp contrast, electronic archives stored on electronic devices are not bound by physical storage limitations and are susceptible to manipulation through relatively simple technical means. The decentralized model of blockchain technology disperses archival information across numerous nodes, markedly improving security measures. In this distributed framework, compromising a single node does not threaten the entire network's integrity, thus ensuring the resilience and robustness of electronic file management systems.

The Role of Blockchain Certification Technology in the Construction of Electronic Archives

Enhancing traceability of original electronic records

Blockchain technology, as discussed previously, relies on hash values—a robust mechanism that detects any alterations within a block's data and signals this change to all subsequent blocks. This feature is essential for maintaining the integrity of each data block within the blockchain, effectively guarding against unauthorized amendments, deletions, or destruction. Through its inherent properties, blockchain serves as an exemplary tool for ensuring the completeness, authenticity, and usability of electronic records [12]. The application of blockchain technology guarantees the preservation of informational integrity throughout the entire lifecycle of electronic records management. As a result, stakeholders can trust blockchain to provide incontrovertible traceability of original electronic records, thereby enhancing accountability and ensuring transparency.

Streamlining electronic records management and access electronic archives

As repositories of varied data types within digital environments, extend access beyond a single custodian to include multiple authorized individuals. Blockchain technology advocates for a system where only legally authorized managers can access and alter archived data. Furthermore, it validates the legitimacy of an electronic file within the blockchain only if it remains unchanged. This approach creates a secure environment for electronic records users, insulated from potential third-party interference. It simplifies the management process by eliminating the need for complex verification methods, such as relying on metadata or proprietary authentication protocols, thus enhancing the efficiency of electronic records management and access.

Serving society more effectively

A blockchain-based electronic record management system provides every authorized user with reliable access to electronic records, fostering trust between the public and archival institutions. This trust is crucial for electronic archives to serve society more effectively. Archival institutions can utilize blockchain technology to standardize the management of electronic archives across different units, ensuring their safety, reliability, and systematic organization. Such standardization forms a solid foundation, preparing archives to meet societal needs in future situations effectively. Moreover, integrating blockchain technology into electronic records management aligns the environmental management continuum with the diverse needs of stakeholders. It facilitates the smooth exchange of information and collaborative workflows among electronic records, significantly reducing management overhead and enhancing operational efficiency [13]. The implementation of blockchain technology in this domain not only secures archiving, storage, transmission, authentication, and the integration of value in electronic record information but also serves as an impetus for societal development and progress.

The Unresolved Issues in Practical Applications

Addressing reliability in digital archives

While blockchain technology excels at ensuring the integrity of electronic files throughout the archival process, it does not inherently verify the accuracy and completeness of the content of digital files or their accurate representation of recorded events. For instance, the ARCHANGEL project is focused on preserving file integrity over time but does not address the authentication of file content before archiving. This poses a significant challenge to the goal of creating digital archives that are both reliable and capable of instilling public trust in archival systems. Consequently, research combining blockchain with archival management is still in its nascent stages. Future studies might investigate ways to guarantee the prolonged accessibility and lasting sustainability of blockchain technologies for improved archival verification services, evaluate the potential for blockchain applications to cover the entire lifecycle of electronic records to confirm their authenticity and reliability and consider the feasibility of comprehensive file storage on the blockchain for maintaining their unalterable state.

Ensuring accessibility of digital archives

As mentioned, file format transformation and content modification are common practices in file management. Research into blockchain systems has primarily concentrated on hash computations to secure the trustworthiness of electronic records; however, such research does not tackle the challenge of ensuring long-term accessibility for digital archives. This issue necessitates the integration of additional application-level technologies to realize the goal of permanent archival storage. These technologies include but are not limited to, the implementation of smart contract technologies.

Navigating storage limitations for archive data

The vast amount of data within archives renders blockchain an unsuitable medium for large-scale data storage. Each blockchain node is expected to continuously download, store, and update an ever-growing dataset to maintain network synchronization. Despite improvements in computational speeds, advancements in data storage and transfer capacities have not kept pace. Proposed solutions often involve compromises, such as reducing the volume of data stored or the number of nodes required for transaction verification [14]. Yet, with the continuous increase in the number of digital files, a long-term strategy for data storage capacity, data processing bandwidth, and computational power requires ongoing growth and refinement.

Conclusion

To date, China has not established specific technical standards for blockchain, nor has it set standardized protocols for uploading information onto blockchain platforms. Meanwhile, the field of blockchain technology and its applications is rapidly advancing. To maintain the technology's ongoing relevance and ensure the long-term interpretability of its contents, the development of standardized specifications is crucial. Establishing a blockchain-based archival trust infrastructure requires collaboration among various stakeholders, including those who manage blockchain operations and entities responsible for generating, managing, and utilizing archives. It is essential to define and implement clear management norms, specifying the roles and responsibilities of all parties involved in blockchain applications to avoid issues of ambiguous management duties, overlapping oversight, and regulatory gaps. For example, protocols for file uploading and access policies on blockchain platforms can provide a solid foundation for content verification. Blockchain technology, while innovative, faces particular challenges in the domain of electronic record management. Its core hash operation is highly sensitive to any binary changes, presenting a considerable challenge for electronic records that often undergo preservation-related format changes, limiting its practical utility. To overcome this, the development of specialized, content-aware hash algorithms would be beneficial. Moreover, integrating blockchain with complementary technologies such as format conversion and media migration could enable seamless interaction between different systems. Such integration can mitigate the functional limitations of blockchain and hasten the advancement toward trusted digital archives. Currently, the scope of electronic archives managed by archival institutions is somewhat limited, not fully meeting public expectations. However, as blockchain technology matures and becomes more intertwined with the electronic archives sector, its potential to enhance the accuracy and integrity of these archives is expected to create a unique position within the industry. With the eventual legal recognition of blockchain's evidentiary value in the archival process, a new era of electronic document service providers may emerge. Traditional archival institutions are likely to continue operating as public service entities, possibly transforming the landscape of physical archives. At the same time, individuals will gain the ability to upload files for verification of authenticity and completeness and to conduct searches within digital archives tailored to their specific needs.

References

1. Xiaodong H, Xinrong H. Analysis of the application of blockchain technology in electronic document management. J Archives Construction. 2018; 2: 4-8.
2. Yuenan L. Preliminary exploration of the application of blockchain technology in file archive management. J Zhejiang Archives. 2018; 5: 7-11.
3. Qianqian Y. Analysis of electronic archive trust management model based on blockchain technology: inspiration from the

ARCHANGEL project in the UK. J Archives Res. 2019; 3: 135-140.

4. Jin S, Sixin X, Xiaoke Z. Research on the model of electronic document authenticity guarantee system based on blockchain technology. J Library Information Knowledge. 2019; 6: 111-119.
5. Ying Z. Research on the application of blockchain technology in electronic archive management. J Friends Secretary. 2023; 3: 37-39.
6. Zipeng W. Research on the application prospects of blockchain electronic file management based on multiple cases. J Zhejiang Archives. 2020; 2: 36-39.
7. Yi Z. Blockchain certification: a trusted digital archive - ARCHANGEL project in the UK and its inspiration. J Lantai World. 2020; 2: 16-20.
8. Lida C. Research on the application of blockchain technology in electronic archive management. J Sci Technol Perspectives. 2021; 20: 71-73.
9. Yuenan L, Yunpeng W. Long-term preservation of digital archives based on blockchain: existing explorations and future development. J Archives Communication. 2018; 6: 44-53.
10. Xiaopei Z. Research on electronic archive information security protection based on blockchain. J Archive Manage. 2020; 4: 34-35.
11. Renjie M, Mengyun L. Several issues on the application of blockchain technology in the utilization of archives in China. J Archives Manage. 2020; 4: 29-33.
12. Yunxia N, Wanqing L. Analysis of trusted electronic file protection system based on blockchain. J Archives Construction. 2021; 11: 28-31.
13. Collomosse J, Bui T, Brown A, Sheridan J, Green A, Bell M, et al. ARCHANGEL: Trusted archives of digital public documents. Proceedings of the ACM Symposium on Document Eng. 2018; 1-4.
14. Lemieux VL. A typology of blockchain recordkeeping solutions and some reflections on their implications for the future of archival preservation. [C] 2017 IEEE international conference on big data (Big Data). IEEE. 2017; 2271-2278.