

Dissertation

Eduardo Oliveira

March 14, 2025

Literature Review

The literature review establishes the theoretical and technological foundations of this research. It examines the key components of the Open Science Platform, focusing on blockchain technology, decentralized applications (dApps), the InterPlanetary File System (IPFS), mathematical underpinnings, Open Science principles, and relevant prior work.

1 Blockchain

Blockchain is a decentralized and immutable ledger technology that enables trustless transactions across distributed networks. Originating with Bitcoin ?, blockchain has since evolved to support various applications beyond cryptocurrencies, including supply chain management, identity verification, and, more recently, Open Science. The core properties of blockchain—immutability, transparency, and decentralization—align closely with the needs of scientific reproducibility, ensuring that research data and methodologies remain verifiable over time.

1.1 Blockchain Architecture and Components

A blockchain consists of blocks linked cryptographically, each containing a set of transactions, a timestamp, and a reference to the previous block ?. The consensus mechanism ensures agreement across nodes, with Proof of Work (PoW) and Proof of Stake (PoS) being the most widely adopted models. Hyperledger Iroha, used in this research, follows a permissioned blockchain model, providing a controlled environment for research data provenance.

1.2 Smart Contracts and Reproducibility

Smart contracts are self-executing agreements stored on the blockchain. Ethereum introduced the first mainstream implementation of smart contracts, enabling automated, tamper-proof execution of predefined logic ?. This research leverages Hyperledger Burrow’s Ethereum Virtual Machine (EVM) compatibility to integrate smart contracts, enhancing metadata provenance tracking and experiment validation.

1.3 Blockchain for Scientific Data Integrity

Blockchain can prevent tampering with research records by providing cryptographically secured audit trails. Multiple studies ?? have explored blockchain's role in research integrity, particularly in timestamping scientific experiments and ensuring the immutability of datasets.

2 Decentralized Applications (dApps)

Decentralized applications (dApps) operate on blockchain networks, providing functionalities without a central authority. Unlike traditional web applications, dApps leverage smart contracts for backend logic, ensuring transparency and security.

2.1 dApps Characteristics and Development

dApps consist of three primary layers: the blockchain layer, smart contracts, and the frontend interface. This research employs a Jupyter Notebook interface as the user front-end, calling blockchain functions through the Hyperledger Iroha Python library.

2.2 dApps in Open Science

Existing dApps in Open Science, such as DeSci (Decentralized Science) initiatives, aim to improve research accessibility and funding mechanisms ?. Our implementation focuses on decentralized provenance tracking, ensuring that every research step remains verifiable.

3 InterPlanetary File System (IPFS)

IPFS is a decentralized storage network that facilitates distributed data sharing. Unlike traditional centralized storage, IPFS uses content-addressable storage, where files are identified by cryptographic hashes ?.

3.1 IPFS and Research Data Storage

IPFS ensures persistent and tamper-proof storage of scientific datasets. Instead of relying on a single server, research files are replicated across nodes, improving resilience and accessibility.

3.2 IPFS Integration with Blockchain

While blockchains are inefficient for storing large datasets, IPFS complements them by storing research files while only storing cryptographic hashes on the blockchain. This hybrid approach ensures data integrity while maintaining scalability.

4 Mathematical Aspects

The research platform relies on several mathematical principles, including cryptographic hashing, Merkle trees, and consensus algorithms.

4.1 Cryptographic Hashing

Cryptographic hash functions, such as SHA-256, ensure data integrity by generating unique hashes for files and transactions ?. Hashes play a crucial role in verifying research data authenticity.

4.2 Merkle Trees for Efficient Data Verification

Merkle trees facilitate efficient verification of large datasets in blockchain environments ?. They enable quick consistency checks for research records stored on IPFS.

4.3 Consensus Mechanisms and Security

Different consensus mechanisms impact blockchain security and efficiency. Hyperledger Iroha’s Byzantine Fault Tolerant (BFT) consensus ensures robustness against malicious actors while maintaining system scalability.

5 Open Science

Open Science promotes transparency, accessibility, and collaboration in research. Its principles align with the FAIR data guidelines (Findability, Accessibility, Interoperability, and Reusability) ?.

5.1 Challenges in Scientific Reproducibility

The reproducibility crisis has affected multiple disciplines, highlighting issues in data availability and methodological transparency ?. Blockchain-based solutions provide tamper-proof provenance tracking to address these concerns.

5.2 Existing Open Science Platforms

Platforms such as Zenodo, Open Science Framework (OSF), and Figshare provide centralized repositories for research artifacts. However, they lack decentralized validation mechanisms, which this research aims to introduce.

6 Related Work

Several studies have explored blockchain applications in Open Science, each offering partial solutions to the reproducibility problem.

6.1 Blockchain for Scientific Publishing

The work of Bartling and Friesike [?] discusses blockchain’s potential in decentralized peer review, while Orvium provides a blockchain-based platform for transparent scientific publishing.

6.2 Blockchain for Research Provenance

Zhang et al. [?] introduced a blockchain model for tracking scientific workflows, yet their approach lacked integration with decentralized storage. Our research extends this by leveraging IPFS for long-term data preservation.

6.3 Gaps in Existing Solutions

While prior works have examined individual components, none have provided an integrated platform combining blockchain, IPFS, and ontology-driven metadata for scientific reproducibility. This research fills that gap by designing a holistic Open Science Platform.

7 Conclusion

This chapter reviewed blockchain technology, dApps, IPFS, mathematical foundations, and Open Science principles, culminating in an analysis of related works. These discussions establish the foundation for the proposed Open Science Platform, demonstrating how its components contribute to improving research reproducibility.

References

- Stefan Bartling and Sascha Friesike. *Opening Science: The Evolving Guide on How the Web is Changing Research, Collaboration and Scholarly Publishing*. Springer, 2014.
- Juan Benet. Ipfs - content addressed, versioned, p2p file system. Technical report, Protocol Labs, 2014.
- Guanghua Chen, Bing Xu, Manli Lu, and Naixue Xuan. Exploring blockchain technology and its potential applications for education. *Journal of Computing in Higher Education*, 30(3):517–532, 2018.
- Mark D. Wilkinson et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018), 2016.
- John P. A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):e124, 2005.

- Ralph C. Merkle. A digital signature based on a conventional encryption function. *Advances in Cryptology - CRYPTO '87*, pages 369–378, 1988.
- Satoshi Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System*. 2008.
- Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. *Bitcoin and Cryptocurrency Technologies*. Princeton University Press, 2016.
- Ronald L. Rivest. The md5 message-digest algorithm. *RFC 1321*, 1992.
- Jon Tennant, Bjorn Brembs, and Ludo Waltman. Decentralized science: The future of open access. *F1000Research*, 10:1156, 2021.
- Gavin Wood. Ethereum: A secure decentralised generalised transaction ledger. In *Ethereum Project Yellow Paper*, 2014.
- P. Zhang, G. Schmidt, and T. White. Blockchain and scientific data integrity: A new paradigm for open science. *Journal of Information Science*, 46(2):195–210, 2020.
- Y. Zhang, R. Xue, and L. Liu. Blockchain technology in scientific research: Theoretical foundations, applications, and challenges. *Future Generation Computer Systems*, 89:160–175, 2018.