Dissertation

Eduardo Oliveira

February 19, 2025

Abstract

The scientific community faces a critical challenge: failed reproductions and replications, which undermine the validity and reliability of research findings. The lack of effective mechanisms to ensure reproducibility leads to decreased accuracy, increased costs, and diminished confidence in scientific discoveries. To address this issue, we propose a decentralized approach leveraging blockchain technology, smart contracts, and the Inter Planetary File System (IPFS) to enhance the adoption of Open Science principles. This research investigates how these technologies can facilitate data sharing and foster reproducibility across diverse scientific disciplines. By constructing an artifact that emulates the functions of an Open Science platform, we explore the potential benefits and technical challenges of integrating these technologies, aiming to contribute to a more transparent, reliable, and trustworthy scientific ecosystem. The findings from this study provide insights into how decentralized technologies can address current gaps in reproducibility and contribute to overcoming the limitations of existing solutions.

1 Introduction

The reproducibility crisis in science arises when researchers fail to replicate the results of a study, even when using the same methods and materials. This issue is prevalent across scientific disciplines, where replication is essential for validating findings, advancing knowledge, and maintaining public trust in research.

Concerns over transparency and reproducibility have intensified in recent years. Studies indicate that a significant proportion of published research does not withstand rigorous scrutiny when retested, raising doubts about the reliability of scientific knowledge. This crisis results in wasted resources, misdirected efforts, and diminished confidence in research findings. Despite various efforts to improve reproducibility, such as better reporting standards and Open Science initiatives, significant challenges remain in ensuring that research findings are consistent, transparent, and verifiable.

Several factors contribute to this challenge, including flawed study designs, data quality issues, unreliable measurement tools, and inconsistent research practices. Poorly designed studies may overlook critical variables, while inadequate measurement tools can produce misleading results. Furthermore, incomplete reporting or the omission of key methodological details can hinder independent verification of findings. Even existing Open Science practices, while valuable, face limitations in scalability, consistency, and automation, often requiring extensive manual effort to achieve reproducibility.

A promising approach to addressing these issues is the adoption of decentralized technologies that enhance transparency, accessibility, and collaboration. Open Science advocates for the sharing of research outputs such as publications, datasets, software, and methodologies to enable verification and reuse. However, existing systems and tools are often fragmented, not fully transparent, or prone to manipulation. Decentralized technologies such as blockchain, IPFS, and smart contracts offer innovative solutions to these problems by providing secure, transparent, and automated means of data sharing, validation, and record-keeping. These technologies are ideally suited for enabling reproducibility by ensuring the immutability of records, enhancing trust in data, and promoting more consistent sharing practices across diverse research fields.

To address the challenges of reproducibility, this study proposes a structured artifact designed to provide a transparent framework for documenting, validating, and disseminating research outputs. This artifact will support a wide range of materials, including reports, protocols, datasets, images, and videos. By ensuring secure and accessible record-keeping, the artifact aims to enhance methodological rigor and enable independent verification of scientific findings.

2 Research Context

The scientific community has long recognized the importance of reproducibility in ensuring the validity and accuracy of research findings. Reproducibility refers to the ability of researchers to replicate another researcher's study or experiment, with minimal changes, to verify its results and conclusions [**?**]. However, despite widespread efforts to address this issue, failed reproductions continue to be a persistent problem across various scientific domains, which not only compromises the credibility of individual studies but also undermines our collective understanding of scientific phenomena.

Studies have shown that reproducibility rates vary widely across disciplines, but even among those fields with a strong tradition of rigor, replication rates often fall short. For instance, a study published in the journal Science found that only 22

Several interrelated factors contribute to these poor rates of reproducibility. Methodological flaws, such as sampling biases or inadequate control groups, often result in difficulties replicating findings. Additionally, a lack of transparency in research practices—such as insufficient reporting on methods, materials, and results—hinders independent verification and replication efforts. Furthermore, limited resources, including small research teams or restricted funding, can prevent the rigorous experimentation necessary to ensure high reproducibility levels.

The need for transparency and openness in scientific research has led to the development of the Open Science movement. Open Science is a broad initiative that advocates for the transparency, accessibility, and collaboration of research processes and outputs. Its tenets include making research data, publications, software, and methodologies publicly accessible, enabling independent verification and reuse of scientific work. Open Science encourages practices such as publishing raw data, sharing code, and providing comprehensive methodological descriptions that allow others to reproduce and extend findings.

Despite the substantial contributions of Open Science, the current systems and tools supporting these practices remain fragmented and often lack the scalability and consistency required for widespread adoption. As such, significant challenges in ensuring reproducibility persist, particularly when dealing with large, complex datasets, proprietary information, or sensitive research materials.

Given these challenges, the integration of decentralized technologies such as blockchain, IPFS, and smart contracts presents a potential solution to address the limitations of current Open Science frameworks. Blockchain technology offers a robust mechanism for ensuring the immutability and integrity of records, making it easier to track data provenance and verify results over time. IPFS provides a scalable, decentralized infrastructure for storing large datasets, ensuring they remain accessible and tamper-proof. Smart contracts, meanwhile, automate the verification and validation process, making it more efficient and secure. Together, these technologies enable a more transparent and accessible research ecosystem that enhances the reproducibility of scientific findings across various domains.

3 Motivation

The scientific community faces a critical challenge in ensuring the integrity and validity of research findings. Reproducibility—the ability to replicate another researcher's study or experiment with minimal changes—forms the cornerstone of scientific progress. However, the failure to reproduce results has become a pervasive issue, undermining the credibility of individual researchers, academic institutions, and entire fields of study. This problem not only hampers the advancement of knowledge but also raises significant concerns regarding the reliability and transparency of scientific research across a variety of disciplines.

Despite ongoing efforts to improve reproducibility, challenges persist due to deep structural issues across many scientific fields. The increasing reliance on data-driven approaches, computational modeling, and high-throughput experimentation has created an environment where results are often difficult to verify or replicate. Furthermore, existing publication systems, funding mechanisms, and peer-review processes have proven ineffective in addressing these reproducibility concerns, exacerbating the problem across both natural and interdisciplinary sciences.

The Open Science movement provides a foundational framework for enhancing transparency and collaboration in research. By advocating for the open sharing of research outputs, including publications, datasets, software, and methodologies, Open Science encourages practices such as data sharing, open peer review, and standardized protocols. These practices foster trust and enable independent verification, crucial for improving reproducibility.

However, the adoption of Open Science faces several barriers, particularly regarding the scalability of data sharing and the automated validation of research findings. While Open Science principles promote transparency, existing tools often fall short in ensuring the integrity and efficient verification of research, especially in complex, data-intensive fields.

This dissertation proposes a decentralized solution to further enhance Open Science by integrating blockchain technology, IPFS, and smart contracts into scientific workflows. Blockchain guarantees the creation of immutable records, ensuring that research data and methodologies remain verifiable and unaltered over time. IPFS addresses the need for decentralized, tamper-proof storage of large datasets, ensuring easy access and secure sharing. Smart contracts automate the validation and verification of research, reducing human error and improving efficiency.

By incorporating these decentralized technologies, this research aims to foster a more secure, transparent, and reproducible scientific ecosystem. The goal is not just to address the current challenges of reproducibility but to strengthen the broader Open Science framework by providing the tools necessary to ensure data integrity, facilitate collaboration, and support independent verification across scientific domains.

4 General Objectives

This research seeks to contribute to the development of a more transparent, collaborative, and reproducible scientific ecosystem. By tackling the central challenges related to reproducibility and trust in scientific research, we intend to create a platform that promotes secure and transparent data sharing, mitigates the risk of errors or data manipulation, and encourages collaboration among researchers from diverse scientific disciplines. This platform is designed to strengthen the reliability of research outcomes, ensuring that they are easily verifiable and replicable by others, thus advancing the adoption of Open Science principles contributing to the advancement of Open Science principles.

5 Specific Objectives

To realize the general objective, the research will focus on the following specific goals:

1. **Develop a distributed application platform for secure data sharing and collaboration:** Enabling seamless interaction among researchers

across various scientific domains. By offering an accessible, encrypted, and auditable system for sharing data, methods, and results, this platform will promote the broader adoption of Open Science principles and foster interdisciplinary collaboration.

2. **Evaluate the effectiveness of the proposed decentralized model in promoting reproducibility:** Employing both experimental and simulation-based methods. This objective focuses on testing the platform's capability to facilitate successful replications of scientific experiments, thereby contributing to a more rigorous, reliable, and trustworthy scientific process.

3. **Design and implement a decentralized application for data storage:** by leveraging blockchain technology, IPFS and smart contracts to ensure data integrity and security. This objective ensures that research data is stored securely and immutably, addressing challenges related to data transparency and protection.

6 Research Hypothesis

This research aims to explore how a decentralized artifact built using blockchain technology, the InterPlanetary File System (IPFS), and smart contracts can contribute to solving the reproducibility challenges in scientific research. The central research question guiding this study is: *How can blockchain technology, smart contracts, and IPFS enhance data sharing and improve reproducibility in scientific research?*

We hypothesize that the proposed artifact, by integrating these decentralized technologies, will facilitate more reliable, transparent, and auditable data sharing practices. Through the use of blockchain for immutable record-keeping, IPFS for secure and decentralized data storage, and smart contracts for automating verification processes, the artifact will foster improved collaboration among researchers, enhance the transparency of methods and results, and support the independent replication of experiments.

The hypothesis of this research is that the decentralized platform, when fully implemented, will contribute significantly to improving the reproducibility of scientific research. This will be evaluated by assessing the artifact's effectiveness in enabling the sharing of data, methods, and results in a secure, transparent, and auditable manner.

7 Methodology

Our methodology will employ a Design Science Research (DSR) approach, which involves generating, testing, and refining solutions to real-world problems through iterative design and experimentation. The goal of this research is to develop a decentralized application that improves scientific research collaboration, reproducibility, and transparency. The application will provide a secure, transparent, and auditable platform for researchers to share data, methods, and results.

The methodology will consist of several key steps. First, we will begin with conceptual modeling. This involves developing models that define the system architecture, data structures, and user interfaces, laying the groundwork for the application's design. The second phase will focus on prototyping and testing, during which prototypes of the system will be developed. These prototypes will then be tested with a small group of users to gather feedback, which will inform subsequent iterations of the design.

After testing and refining the prototype, the final system will undergo evaluation and validation. This phase will assess the effectiveness of the system in meeting its goals, focusing on factors such as usability, performance, and its impact on reproducibility and transparency in scientific research. Evaluation methods may include user testing and performance assessments, ensuring that the final solution addresses the challenges identified in the research context.

The Design Science Research methodology guiding this study adheres to key principles. The first principle is innovation, where the research is committed to creating a novel solution that effectively addresses the identified challenges. The second principle,improvement, emphasizes the advancement of existing practices in decentralized data storage and sharing, ultimately enhancing the rigor and reliability of scientific work. Lastly, the principle of validation ensures that the final system undergoes rigorous testing and evaluation to demonstrate its effectiveness in promoting reproducibility and transparency across scientific domains.

8 Summary of Contributions

This research presents a structured approach to addressing reproducibility challenges in scientific research through the development of a decentralized artifact integrating blockchain technology, the InterPlanetary File System (IPFS), and smart contracts. By embedding these technologies into a transparent and auditable framework, the study contributes to advancing Open Science practices and fostering trust in scientific findings. The key contributions of this work are as follows.

First, the research introduces a decentralized application designed to enhance the transparency, accessibility, and verification of research data. By leveraging blockchain's immutability, IPFS's decentralized storage, and smart contracts' automation capabilities, the proposed system enables secure data sharing while ensuring the integrity of research outputs. This platform aligns with Open Science principles, allowing researchers to disseminate their methodologies and results in a verifiable manner, thereby mitigating reproducibility issues.

Second, this study evaluates the artifact's effectiveness in supporting reproducibility through a structured validation process. The research employs Design Science Research (DSR) to iteratively design, implement, and assess the platform, ensuring its practicality and impact within scientific workflows. Through experimental testing and simulation-based evaluations, the study examines how decentralized technologies contribute to reducing inconsistencies in research outcomes.

Additionally, this dissertation builds upon prior work by the author, including the published paper **On the Use of Blockchain Technology to Improve the Reproducibility of Preclinical Research Experiments**, presented at the 25th International Conference on Enterprise Information Systems. This prior research laid the foundation for exploring blockchain's role in scientific reproducibility, and the current study extends these insights by integrating additional technologies and refining their application within Open Science frameworks.

By addressing core limitations in current research sharing practices, this work offers an decentralized approach to ensure transparency, accountability, and long term accessibility of scientific knowledge. The findings provides a practical solution to one of the most pressing challenges in modern research: the reproducibility crisis.