

Dissertation

Eduardo Oliveira

February 25, 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 InterPlanetary 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 implements a decentralized 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 and better reporting guidelines such as PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [8], ARRIVE (Animal Research Reporting of In Vivo Experiments) [9] and FAIR (The FAIR Guiding Principles for scientific data management and stewardship) [12], 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.

A promising approach to addressing reproducibility issues lies in the adoption of Open Science [2], both as a movement and as a methodological framework that fosters transparency, collaboration and accessibility in research. Open Science promotes unrestricted access to research outputs, including publications, datasets, code, methodologies and protocols, enabling verification and reuse while reducing barriers to scientific progress. By advocating for systematic, reproducible and accessible research practices, Open Science enhances accountability and trust within the scientific community. However, despite its transformative potential, its current implementations encounter significant technical and structural limitations, particularly in scalability, consistency and automation. Existing infrastructures often rely on centralized repositories and fragmented systems that lack transparency, are vulnerable to manipulation or demand extensive manual oversight to ensure verifiability.

To overcome these challenges, decentralized technologies offer a robust proposition framework that aligns with the principles of Open Science while addressing its technical constraints. Blockchain technology ensures the immutability and integrity of research records, preventing unauthorized modifications, tampering and traceability. IPFS (InterPlanetary File System) provides decentralized and persistent data storage, eliminating single points of failure and ensuring long-term accessibility of research outputs. Smart contracts facilitate the automated enforcement of predefined research-sharing policies reducing the need for intermediaries and fostering trust in collaborative environments. By integrating these decentralized technologies, it becomes possible to construct an infrastructure that not only upholds Open Science principles but also strengthens reproducibility by delivering secure, transparent and verifiable mechanisms for data sharing and validation across scientific disciplines.

Building upon these foundations, this study proposes the development of a structured artifact designed to provide a transparent framework for documenting, validating and disseminating research outputs. This artifact will accommodate diverse materials, including reports, protocols, datasets, images and videos, ensuring secure and accessible record-keeping. By leveraging decentralized technologies, the proposed artifact targets the enhancement of methodological rigor, support independent verification of scientific findings and ultimately contribute to a more trustworthy and reproducible scientific ecosystem.

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 [11]. 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 PLOS Biology found that only 22% of papers published in top-tier journals were replicable [3]. Similarly, a review of studies on reproducibility in Nature revealed that only 15% of studies reported high levels of reproducibility [5].

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 datasets, publications, code 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 [6], IPFS [1] and smart contracts [10] 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. Furthermore, existing publication systems, funding mechanisms and peer-review processes have proven ineffective in addressing these reproducibility concerns, exacerbating the problem.

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 research outputs, 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 hopes to promote 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 is set 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 facilitates 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) [4] 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 a model 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 and selected.

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 its impact on reproducibility and transparency in scientific research.

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, transparency and auditability.

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 on. 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**[7], 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.

9 Conclusion

The reproducibility crisis remains a significant challenge across scientific disciplines, raising concerns about the reliability and integrity of research findings. Transparency, accessibility and verifiability are essential to ensuring the credibility of scientific knowledge, yet traditional research practices often fall short in addressing these needs. In response to these limitations, this dissertation proposes a decentralized artifact that integrates blockchain technology, the InterPlanetary File System (IPFS) and smart contracts to enhance research reproducibility within the framework of Open Science.

Through a Design Science Research (DSR) approach, this study develops, evaluates and refines a decentralized platform that enables secure and transparent data sharing. By leveraging blockchain’s immutability, IPFS’s decentralized storage and smart contracts’ automation capabilities, the proposed solution addresses key barriers to reproducibility, such as data integrity, methodological opacity and limited accessibility of research outputs. The artifact is designed to facilitate independent verification of scientific findings while fostering collab-

oration across research communities.

The study’s findings contribute to ongoing discussions on Open Science by demonstrating how decentralized technologies can support a more rigorous and trustworthy research environment. By ensuring the preservation and availability of research data and methodologies in an auditable and tamper-proof manner, this work provides a pathway toward more transparent and accountable scientific practices. The evaluation of the artifact through experimental testing and simulation-based modeling highlights its potential to improve reproducibility by offering an alternative to conventional research-sharing mechanisms.

Furthermore, this research builds upon prior work, including the author’s publication **On the Use of Blockchain Technology to Improve the Reproducibility of Preclinical Research Experiments**, extending the discussion by incorporating additional decentralized technologies and refining their application within scientific workflows. By addressing critical shortcomings in existing research dissemination methods, the study underscores the transformative potential of decentralized systems in enhancing trust and reliability in scientific endeavors.

While this research provides a foundational step toward improving reproducibility through decentralized solutions, further work is necessary to refine and expand its applicability across diverse scientific disciplines. Future research should explore the integration of additional security and privacy mechanisms, assess real-world adoption challenges and integration with current scientific dissemination platforms. As Open Science initiatives continue to gain momentum, the adoption of decentralized technologies represents a promising avenue for reshaping the way research is conducted, verified and shared, ultimately strengthening the reproducibility of scientific research.

10 Document Structure

This section provides an overview of the structure and organization of this dissertation. By delineating the logical progression of chapters, sections, and their interconnections, the document structure is designed to guide readers through a cohesive narrative that unfolds the research journey, findings, and contributions.

- **Chapter 1: Introduction** – The first chapter establishes the foundational context for this research. It articulates the motivation driving the study, setting the stage for an exploration of decentralized technologies within the landscape of Open Science and reproducibility. The chapter outlines the research context, defines the general and specific objectives, introduces the research hypotheses, and details the methodology employed. Finally, it presents a summary of the dissertation’s contributions and outlines the subsequent chapters that structure the research narrative.
- **Chapter 2: Literature Review** – This chapter examines the relevant literature that informs the research. It explores key concepts such as

blockchain technology, decentralized applications, Open Science principles, IPFS, and smart contracts. Additionally, it synthesizes related works to establish the scholarly foundation against which this research unfolds, providing the necessary referential context that supports the subsequent chapters.

- **Chapter 3: Proposed Decentralized Model** – This pivotal chapter presents the core contribution of the research: the proposed decentralized governance model. It begins with a comprehensive overview, explaining its architecture, data model, technical specifications, and design rationale. A detailed account of the model’s implementation follows, along with an empirical validation process designed to assess its effectiveness. This chapter aligns with Objective 1 by proposing an innovative solution, furthers Objective 2 by ensuring technical feasibility, and directly addresses Objective 3 by evaluating the model’s impact on reproducibility.
- **Chapter 4: Conclusions and Future Work** – The final chapter synthesizes the research journey, encapsulating its findings, insights, and implications. It revisits the research hypotheses in light of empirical results and examines their alignment with the research objectives. A forward-looking perspective is provided through an exploration of potential directions for future research, reinforcing Objective 5 and underscoring the study’s potential to drive further advancements in the field.
- **Appendices and References** – The appendices contain supplementary materials that enhance the dissertation’s main content. These include detailed technical specifications, source code snippets, and additional data that contribute to a comprehensive understanding of the research process. The reference section compiles the scholarly works consulted throughout the study, providing the theoretical and empirical foundation underpinning the research.

References

- [1] Juan Benet. IPFS - Content Addressed, Versioned, P2P File System (DRAFT 3), November 2022. Last accessed on 30 Nov 2022.
- [2] Erin D. Foster and Ariel Deardorff. Open Science Framework (OSF). *jmla*, 105(2), April 2017.
- [3] Leonard P. Freedman, Iain M. Cockburn, and Timothy S. Simcoe. The Economics of Reproducibility in Preclinical Research. *PLOS Biology*, 13(6):e1002165, June 2015.
- [4] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research. *MIS Quarterly*, 28(1):75–105, 2004.

- [5] Story C. Landis, Susan G. Amara, Khusru Asadullah, Chris P. Austin, Robi Blumenstein, Eileen W. Bradley, Ronald G. Crystal, Robert B. Darnell, Robert J. Ferrante, Howard Fillit, Robert Finkelstein, Marc Fisher, Howard E. Gendelman, Robert M. Golub, John L. Goudreau, Robert A. Gross, Amelie K. Gubitz, Sharon E. Hesterlee, David W. Howells, John Huguenard, Katrina Kelner, Walter Koroshetz, Dimitri Krainc, Stanley E. Lazic, Michael S. Levine, Malcolm R. Macleod, John M. McCall, Richard T. Moxley III, Kalyani Narasimhan, Linda J. Noble, Steve Perrin, John D. Porter, Oswald Steward, Ellis Unger, Ursula Utz, and Shai D. Silberberg. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*, 490(7419):187–191, October 2012.
- [6] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2009.
- [7] Eduardo Oliveira, Rafael Z. Frantz, Carlos Molina-Jiménez, Thiago Heck, Sandro Sawicki, and Fabricia Roos-Frantz. On the use of blockchain technology to improve the reproducibility of preclinical research experiments. In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023)*, pages 173–179. INSTICC, SciTePress, 2023.
- [8] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 2021.
- [9] Nathalie Percie du Sert, Amrita Ahluwalia, Syed Alam, Mark T. Avey, Monya Baker, Warren J. Browne, et al. Reporting animal research: Explanation and elaboration for the arrive guidelines 2.0. *PLOS Biology*, 18(7):e3000411, 2020.
- [10] Nick Szabo. Smart contracts, January 1994. Last accessed on 28 Jan 2021.
- [11] Nicole A. Vasilevsky, Matthew H. Brush, Holly Paddock, Laura Ponting, Shreejoy J. Tripathy, Gregory M. LaRocca, and Melissa A. Haendel. On the reproducibility of science: unique identification of research resources in the biomedical literature. *PeerJ*, 1:e148, September 2013.
- [12] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, 2016.